Acknowledgments

The realization of this thesis would not have been possible without the active support provided by some people I would like to thank.

First of all, I am most grateful to my supervisor Professor Eric de Bodt for his support, guidance, suggestion and availability. His enthusiasm and his creativity, his total investment of every moment and his passion of research are not less than the co-authors of this thesis. I am also deeply indebted to Professor Jean-Gabriel Cousin who collaborated with me to deliver a joint article (which appears in the last empirical section of this thesis) published in *Finance*. His advice, ideas, suggestions and econometric competences are extremely valuable to me. I am also very grateful to Professors Michel Levasseur, Luc Bauwens, Léopold Simar and Nihat Aktas who made outstanding and very helpful suggestions on different points of my thesis, and brought me their expertise in their fields. Professors Michel Levasseur, Luc Bauwens, Jacques Hamon and Patrick Roger have accepted to be members of my thesis jury. I thank them for the time they devoted and for the interest they showed for my work.

This thesis would not have been possible without the institutional support of the UCL and without the financial support of the Fonds National de la Recherche Scientifique. I think also of Professors Philippe Grégoire and Armin Schwienbacher who invest a lot of themselves in the "Finance" unit.

Finally, I would warmly thank the support and friendship of many colleagues at UCL. In particular, I would like to mention Hervé Van Oppens, Vassilis Samaras, Jérôme Barbarin, Walid Ben Omrane, Savina Princen, Ilham Riachi, Bertrand Jacob, Gaël Imad Eddine, Samuel de Lhonneux, Olivier Vercruysse, Yannick de Harlez, Giorgio Tesolin, Arlène Derbaix and Stéphane Nassaut.

Table of contents

ACKNOWLEDGMENTS ..... **I**

TABLE OF CONTENTS ..... **III**

TABLE OF FIGURES AND TABLES ..... **VII**

INTRODUCTION **1**

1 CONSTRAINED INVESTORS AND THE COST OF CAPITAL **9**

1.1 LITERATURE 9

1.2 CONVENTIONS 10

1.3 NOTATIONS 11

1.4 TERMINOLOGY 12

1.5 ASSUMPTIONS 13

1.6 MODEL DEVELOPMENT AND MAIN RESULTS 14

1.6.1 Unconstrained investors (benchmark case) *14*

1.6.1.1 Derivation of the optimal matrix of demanded quantities 14

1.6.1.2 Derivation of the equilibrium prices 16

1.6.1.3 Cost of capital 18

1.6.2 Constrained investors *18*

1.6.2.1 Derivation of the optimal matrix of demanded quantities 18

1.6.2.2 Derivation of the equilibrium prices 25

1.6.2.3 Derivation of the expected utility 26

1.6.2.4 Comments on the effects of the deviations 29

1.6.3 Extension with a more realistic supply curve *38*

1.6.4 Addition of a short-selling constraint *39*

1.6.5 Does under-diversification really matter? *42*

1.6.5.1 General discussion 42

1.6.5.2 Relevancy of some empirical settings 43

2 EMPIRICAL ESTIMATION: S&P 500 MEMBERSHIP AND THE COST OF CAPITAL **47**

2.1 INTRODUCTION 47

2.2 THEORETICAL BACKGROUND 50

2.2.1 The « index effect » *50*

2.2.2 Competing hypotheses *53*
2.2.2.1 Imperfect elasticity based hypotheses 54
2.2.2.1.1 The Price Pressure Hypothesis 54
2.2.2.1.2 The Imperfect Substitutes and the Downward-Sloping Demand Curve Hypothesis 56
2.2.2.2 Perfect elasticity based hypotheses 57
2.2.2.2.1 The Liquidity Cost Hypothesis 57
2.2.2.2.2 The Information Content Hypothesis 58
2.2.2.2.3 The Market-Segmentation and Investor Recognition Hypothesis 61
2.2.3 Contribution *63*
2.2.3.1 Positioning of our theoretical model 63
2.2.3.1.1 Discussion about the elasticity of the demand curve 63
2.2.3.1.2 The link between the five hypotheses and our theoretical model 66
2.2.3.2 Positioning of our empirical model 66
2.3 DATA 69
2.3.1 Data sample *69*
2.3.2 Variables definition *73*
2.4 ECONOMETRIC SPECIFICATION 80
2.4.1 Morck and Yang cross-sectional regression *80*
2.4.2 Pooled OLS regression *83*
2.4.3 Endogeneity and self-selection *87*
2.4.3.1 Fixed-effect panel estimation 88
2.4.3.2 Instrumental variables estimation 88
2.4.3.3 Self-selection estimation 89
2.5 RESULTS 93
2.5.1 Morck and Yang cross-sectional regression *93*
2.5.2 Pooled OLS regression *96*
2.5.3 Fixed-effect panel estimation *100*
2.5.4 Probit estimation *101*
2.5.5 Instrumental variables estimation *103*
2.5.6 Self-selection estimation *106*
2.6 CONCLUSION 109

3 NEWS AND THE COST OF CAPITAL IN A UNIVERSE OF CONSTRAINED INVESTORS 113

3.1 INTRODUCTION 113

3.2 LITERATURE 115

3.3 CONVENTIONS, NOTATIONS AND ASSUMPTIONS 116

3.4 MODEL DEVELOPMENT 116

3.4.1 General setting 116

3.4.2 Introduction of two simplifications 117

3.4.2.1 Maximum one signal by asset 118

3.4.2.2 Two signals: micro (on asset *j*) and macro (on market *M*) 118

3.4.3 Determination of the unconditional distribution of prices 121

3.4.4 Determination of the distribution of prices conditional to signals 123

3.4.4.1 Computation of the conditional expectation 124

3.4.4.2 Computation of the conditional variance and covariance 126

3.4.5 Extension from one to several stock and market signals 129

3.4.6 Link between micro and macro signals 130

3.4.7 Final pricing equation 132

3.5 INTERPRETATIONS 134

3.5.1 Impact of micro information on the cost of capital 134

3.5.2 Impact of macro information on the cost of capital 142

3.5.3 Impact of the unconditional variance of firm j on the cost of capital 145

3.5.4 Impact of the unconditional variance of the market on the cost of capital 146

3.5.5 Impact of the unconditional correlation on the cost of capital 146

3.6 CONCLUSION 146

4 EMPIRICAL ESTIMATION OF THE IMPACT OF PUBLIC INFORMATION ON THE CONDITIONAL VOLATILITY OF STOCK RETURNS 149

4.1 THE FRENCH CASE 149

4.1.1 Introduction 149

4.1.2 Information and conditional volatility 153

4.1.2.1 Modeling the conditional volatility 153

4.1.2.2 News arrivals and market reactions 155

4.1.3 Sample selection *156*

4.1.4 Model specification *172*

4.1.4.1 The GARCH model 172

4.1.4.2 The two-state market model (TSMM) 174

4.1.5 Results *177*

4.1.5.1 GARCH framework 177

4.1.5.2 MSR framework: the two-state market model (TSMM) 182

4.1.5.2.1 "Firm by firm" analysis 186

4.1.5.2.1.1 Probit regression: the dependent variable is a dummy 186

4.1.5.2.1.2 GMM regression: the dependent variable is the smooth probability itself 190

4.1.5.2.2 Panel data analysis 193

4.1.5.2.3 Analysis refinement: categorization by topic and by timing 196

4.1.5.3 Informational content and asymmetry issues 201

4.1.6 Conclusion *206*

4.2 THE US CASE 208

4.2.1 Sample Selection *208*

4.2.2 Model Specification *223*

4.2.3 Results *223*

4.2.3.1 GARCH framework 223

4.2.3.2 MSR framework 223

4.2.3.2.1 Panel data analysis with the smooth probability 223

4.2.3.2.1.1 Global analysis 223

4.2.3.2.1.2 Introduction of a time effect 228

4.2.3.2.1.3 Analysis by category 231

4.2.3.2.1.3.1 Categorization by topic 231

4.2.3.2.1.3.2 Categorization by timing 234

4.2.3.2.2 Cross-sectional data analysis with the transition probability 235

4.2.4 Conclusion *237*

CONCLUSION **239**

REFERENCES **242**

APPENDIX A **252**

APPENDIX B **262**

Table of Figures and Tables

Figure 1-1: Impact of the deviation on the equilibrium price and quantities 23
Figure 1-2: Impact of the deviation on the equilibrium with a partially elastic supply 39
Figure 2-1: Impact of a positive deviation on the equilibrium with a partially inelastic demand 64
Figure 2-2: Impact of a positive deviation on the equilibrium with a perfectly elastic demand 64
Figure 2-3: Evolution of the t-stats of the Membership Dummy and the Index Weight over Time .. 96
Figure 3-1: The risk component as a function of deviations and signaling 136
Figure 4-1: Daily number of news releases on the French market from Jan. 1999 to Dec. 2003 159
Figure 4-2: Weekly number of news releases on the French market from Jan. 1999 to Dec. 2003 . 160
Figure 4-3: Total news releases average by day of the week on the French market 161
Figure 4-4: Total number of news releases by company on the French market 167
Figure 4-5: Total number of news releases by industry on the French market 168
Figure 4-6: Total number of news releases by subject on the French market 169
Figure 4-7: Comparison of the volatility with the level of communication for Vivendi 185
Figure 4-8: Coefficients of various regressions with different subsets of news releases 199
Figure 4-9: Daily number of news releases on the US market from Nov. 1995 to Dec. 2005 210
Figure 4-10: Weekly number of news releases on the US market from Nov. 1995 to Dec. 2005 211
Figure 4-11: Total news releases average by day of the week on the US market 212
Figure 4-12: Total news releases average by hour of the day in New-York on the US market 213
Figure 4-13: Total number of news releases by company on the US market 218
Figure 4-14: Total number of news releases by industry on the US market 219
Figure 4-15: Total number of news releases by subject on the US market 220

Table 2-1: Summary of the theoretical reasoning underlying the competing hypotheses 68
Table 2-2: Historical evolution of S&P 500 additions and deletions 70
Table 2-3: Descriptives statistics 71
Table 2-4: Summary of the underlying empirical implications of the competing hypotheses 92
Table 2-5: Cross-Sectional OLS Regression of Tobin's Q on the S&P 500 Membership Dummy 93
Table 2-6: Cross-Sectional OLS Regression of Tobin's Q On S&P 500 Index Weight 93
Table 2-7: How Regression Coefficients on the S&P Dummy (and Weight) Changed over Time 95
Table 2-8: Pooled OLS regression 96
Table 2-9: Pooled OLS regression without *SPxWeight* 99
Table 2-10: Fixed-effect panel estimation 100
Table 2-11: Probit estimation of *SP* on relevant variables 102
Table 2-12: Instrumental variables regression 105
Table 2-13: Instrumental variables regression without *SPxWeight_F* 105
Table 2-14: Self-selection estimation 108
Table 2-15: Self-selection estimation without the dummy *SP* 108

Table 4-1: Factiva classification of companies ("CO" field) on the French market 162
Table 4-2: Factiva classification of industries ("IN" field) 163
Table 4-3: Factiva classification of companies ("NS" field) 164
Table 4-4: Summary statistics on the French market 170
Table 4-5: GARCH model 178
Table 4-6: Estimates of the Market Model (MM) and of the Two-State Market Model (TSMM) .. 183
Table 4-7: Probit regressions 188
Table 4-8: GMM regressions 191
Table 4-9: Panel data analysis 195
Table 4-10: Panel data analysis for various types of news: categories are not mixed 198
Table 4-11: Panel data analysis for various types of news: categories are mixed together 200
Table 4-12: Summary Statistics across volatility regimes on the French market 204
Table 4-13: Informational value and asymmetry issues 205
Table 4-14: Factiva classification of 50 companies ("CO" field) on the US market 215
Table 4-15: Summary statistics on the US market 221
Table 4-16: Panel fixed-effect regression of *Prob* on *D* 225
Table 4-17: Panel fixed-effect regression of *Prob* on four dummies *D(k)* 226
Table 4-18: Panel fixed-effect regression of *Prob* on the number of news releases *NBN* 227
Table 4-19: Panel fixed-effect regression of *Prob* on *NBN* and *Vol* 228
Table 4-20: Panel fixed-effect regression of *Prob* on *NBN* with time effect 230
Table 4-21: Panel fixed-effect regression on various topics of news 231
Table 4-22: Panel fixed-effect regression of *Prob* on category *Reg/Gov Policy* 232
Table 4-23: Panel fixed-effect regression of *Prob* on category *Ownership Change* 233
Table 4-24: Panel fixed-effect regression of *Prob* on category *Contracts* 233
Table 4-25: Panel fixed-effect regression of *Prob* on *NBN* split by timing 234
Table 4-26: Cross-sectional regression of the transition probabilities 236

Introduction

We all have in mind a couple of dramatic examples of how a piece of information issued by some economical or political entity, at first view more or less innocuous, resulted in tremendous consequences for a private company or, worst, for the whole financial market. Between a stock falling by 30% and another soaring by 20%, there are often only an awkward profit warning and a skilled rumor of acquisition. Serge Tchuruk, Alcatel's CEO, still probably remembers the disastrous consequences of his injudicious profit warning of September 17, 1998. At first sight, the news release looked quite innocent: "The operational result of the telecommunications sector (…) will suffer from the large decrease in investments (…) and from the crisis in South-East Asia and in Russia. Alcatel reckons that the operational performance in 1998 will not reach the expected level." Among other things, Alcatel said it expected to make a profit in 1998 of €1.1 billion, down sharply from an earlier forecast of €1.4 billion. The market reaction was terrible: shares of the company lost 38% in one day, involving that not less than €10 billion of Alcatel's market capitalization vanished in a couple of hours. Never had the "Paris bourse" experienced such a dramatic stock crash. In this perspective, a better understanding of the processes and the mechanisms relating the information flow to the stock price (and thus, to the cost of capital of the firm) is of prime importance.

Research in finance tried very early to identify the various determinants of the cost of capital of the firm. The most famous asset pricing model, the CAPM, was developed more than forty years ago, by the parallel work of William Sharpe (1964), Jack Treynor (1961), John Lintner (1965) and Jan Mossin (1966), who themselves built on the earlier work of Harry Markowitz (see Harry Markowitz (1999) for a good summary of portofolio theory before 1960). Under some restrictive assumptions, this model identifies the systematic risk as the main determinant of the price of a stock. In this framework, idiosyncratic risk does not matter. Subsequent studies have put forward the potential limitations of this model and suggested alternative perspectives and additional determinants of the stock price. The purpose of this dissertation is to bring a new contribution to this endless quest.

This quest begins with Section 1, where we develop a simple theoretical model of "asset pricing" that presents mainly two interests. Firstly, it extends the standard CAPM model to a more realistic world where some investors are "constrained" and do not hold (for various reasons) the optimal CAPM quantities. Secondly, it provides a useful theoretical

framework for the subsequent sections of this dissertation. In particular, the apparition of the idiosyncratic variance in the pricing equation will be exploited in Section 3 and Section 4 that investigate the impact of public information on the stock price and its major drivers (future price expectation, variance and covariance). In a pure CAPM world, it would only make sense to examine the impact of the news on the expectation and the covariance, as the idiosyncratic risk can be fully diversified. In a world with deviations from the CAPM, the variance of the price matters again and this additional term may enrich the analysis of the links between information and cost of capital.

The starting point of our theoretical model of Section 1 is the CAPM model that we extent to imperfectly diversified, or constrained, investors. By imperfectly diversified (or constrained) investors, we mean investors who do not hold the optimal quantities they should if they maximized their utility function. For some reasons (they are potentially many of them, like another optimization behavior, strategies of control, …), those agents are in deviations of the predicted optimal holdings. These deviations can be positive and negative and can thus be understood as an overweight or an underweight in a given stock. The main consequence of this extended model is to cause an extra variance term to appear in the pricing equation, besides the usual covariance term of the CAPM. We conclude that if investors are not optimally diversified, and if the aggregate deviation at the market level is substantial, the price is positively linked to the product of the variance of the stock price by the aggregate deviation of investors in this stock.

The idea of "constrained investors" is not new, but authors who have already considered it use a different framework and build on a different intuition. For example, a significant amount of literature ranging from Brennan (1975) to Viard (1995) explore the asset pricing effects of fixed holding costs, which can be a cause of deviations from the CAPM, but not the only one. Our purpose is not to explain the origins of the deviations, but to provide a theoretical setting that can account for them and make some predictions about their main implications.

Apart few exceptions, like Braun and Larrain (2008), most of articles using CAPM are not concerned with the supply side of the market and prefer to assume it to be fixed and constant. In some settings, especially when private information is introduced in the model, the supply is set to be random, otherwise prices could be fully revealing of all public and private information and would lead to trivial results. Since we only consider public information, we follow the traditional way proposed by Fama (1976), Miller

(1977), Li and Fleisher (2004) and Greenwood (2005) and we assume a deterministic constant asset supply, even though we provide later an extension to a supply linear in the price for the sake of completeness.

Finally, our theoretical model also briefly tackles the issue of the constraint on short-sales. Literature is sometimes contradictory on the implication of short-sales constraints on prices: Miller (1977) and Harrison and Kreps (1978) predict that such a constraint drives prices up, because it keeps "pessimistic" investors out of the market, Diamond and Verrechia (1987) argue that prices will be less informative but still are biased upward, whereas Bai, Chang and Wang (2006) show that short-sale constraints may cause asset prices to decrease because less informative prices increases uncertainty. Since we consider the uncertainty (the variance-covariance matrix) as exogeneous and independent of the short-selling constraint, like Miller (1977) for example, we do not pretend to give a complete and thorough analysis on the topic of short-selling, but rather to provide some insight on this issue in the context of our model, with its set of (sometimes limitative) assumptions.

After the development of the theoretical model of Section 1, it would be logical to propose in Section 2 a direct empirical test of the main conclusions of our theoretical model. However, it proved to be difficult, as the theoretical setting and concepts cannot be easily transposed in a straightforward and reliable manner into an empirical context. Therefore, Section 2 has finally evolved into an empirical article about an issue that could be seen as a potential manifestation of a deviation: the index effect. This index effect, largely documented by the literature, consists in the value premium benefiting to the companies members of a widely traded index (typically the S&P 500 index). It can be seen as an example of deviations from two different but complementary perspectives: either because index funds overweight the companies members of the index in their portfolio (leading to a positive deviation), or because they underweight the non-member companies (leading to a negative deviation). This index effect has been broadly investigated by the researchers who have come with several possible explanations, some of them being very close to our deviation theory, some others adopting a totally different approach without being necessary contradictory. Thanks to various econometric specifications in line with Morck and Yang (2002) and Campa and Kedia (2002), our Section 2 corroborates the index value effect claimed in the previous literature, tests more or less directly some of the five competing hypotheses advanced

by researchers on the topic (introduced by Harris and Gurrel (1986), Shleifer (1986), Mikkelson and Partch (1985), Goetzmann and Garry (1986) and Merton (1987)), and explains why some of the results of Morck and Yang (2002) may present some bias. The interest of this Section 2 is mainly to gather and confront all hypotheses proposed by the literature by means of advanced and robust econometrical specifications that allow to control for the endogeneity issue and to ensure more confidence in the reliability of the results.

The challenge of the next sections is to introduce the "information" component in the price equation. Specifically, we want to investigate the pair information-price, by means of a theoretical model and an empirical estimation.

First of all, we have to precise what kind of information we are interested in. Indeed, information can be roughly classified into two broad categories: public and private information. By definition, public information is of common knowledge, while private one is detained solely by one or some individuals and leads to information asymmetry issues (see Easley and O'Hara (2004)). Section 3 and Section 4 of the dissertation will restrict their scope to public information only. The universe of private information calls for different settings, modeling and literature. The reader interested by the modeling of private information on the financial markets will easily find a lot of articles on the topic, from the famous paradox of Grossman and Stiglitz (1980) to the more recent article of Chae (2005), like Ausubel (1990), Hasbrouck (1991) and Easley and O'Hara (2004) among others. Many of them also address the issue of insider trading, a still today very debatable topic in the field of private information (see Van Oppens (2007) for a thorough review on this subject).

In Section 3, we first begin by introducing the topic and the necessity to add this information element in our analysis. We then enrich the theoretical model of Section 1 with the concept of signals. Those signals are the theoretical counterpart of the news releases we will consider in the subsequent empirical Section 4. The purpose of Section 3 is to investigate how the precision and the frequency of microeconomic and macroeconomic signals impact the stock price through its three main factors: the expectation, the variance and the covariance of the vector of future prices in a universe of potentially constrained investors. The development of this extended model, which mainly builds on Lambert, Leuz and Verrechia (2006) and classical results of multivariate analysis, allows us to draw several interesting conclusions about the

existence of an optimal communication frequency for example, and the determinants that make the firm more or less willing to communicate or, on the contrary, to remain opaque. We study in detail the impact of information intensity and quality on the conditional expectation, variance and covariance of the future prices, and we show that the mechanisms governing the relation information – volatility – price are complex and depend heavily on the magnitude of the potential deviation, on the context of the market, on the characteristics of the firm and of the signals concerning her.

Finally, Section 4 aims at empirically testing our theoretical model of Section 3, at least partially, since it investigates in particular the link between information and conditional volatility. There are several reasons why our empirical study focus especially on the "variance" medium, rather than on the "expectation" or "covariance" channels, although our theoretical model covers all of them. First of all, our sample of news is such that it is very difficult to classify our news releases into good and bad news, and such a classification is of prime importance in order to study the link between expectation and the equilibrium price, since the sign of the relation depends heavily on the positive or negative content of the news release. Secondly, even though the link between the conditional covariance and the equilibrium price is of course of prime interest as well, to our knowledge, the dynamic modeling of the conditional covariance has not been studied as much as the one of the conditional variance. Although we plan to investigate the empirical link between information, covariance and prices in the future, it was important to situate this dissertation in the line of known and asserted studies on the topic, which are numerous in the field of the variance modeling. Indeed, several models have been proposed in order to estimate the conditional variance of returns: GARCH models, Markov-switching regression models, implied volatility of options, realized volatility, etc. From those various techniques, we have used the two first in our empirical study of Section 4: a GARCH model and a Markov-switching regression model. The reasons underlying this methodological choice are explained in detail in our empirical Section 4.

The latter restricts its scope to "micro news", i.e. news specific to the firm (profit warnings, M&A, analyst comments, …), as opposed to "macro news", which is information about the market as a whole (unemployment, inflation, …). Indeed, the features of our sample of news releases allows us to identify the company concerned by each micro announcement, and to examine at the firm-level the impact of such micro

news release. We could have studied the effect of macro news on the market as well, like other researchers before us (see Ederington and Lee (1993) for example), but given the nature of our sample, focusing on the micro level would deliver more added value than focusing on the macro level.

Specifically, we conduct two empirical studies on two different samples of Reuters news releases covering 40 companies on the French market and 458 on the US market over a similar period of 5 years. By using GARCH (see Engle (1982) and Bollerslev (1986)) and Markov Switching Regimes (MSR) regressions, we bring significant evidence of the positive relation between the intensity of the information flow and the conditional volatility, already partly documented by Kalev, Liu, Pham and Jarnecic (2004). We also show that this relation depends strongly on the subject and the timing of announcements, as (for example) news concerning regulations and decisions of public authorities seems to increase the most uncertainty among investors, while news announcing next privatizations tends to decrease the uncertainty around the future prospects of the firm.

The choice of a Markov-switching regression model is related to a powerful argument, which is known as the Mixture of Distribution Hypothesis (MDH). Suggested for the first time by Clark (1973), it posits that the joint distribution of daily return and volume can be modeled as a mixture of bivariate normal distributions. Roll (1988) gives empirical support to it, while Ross (1989) provides theoretical evidence. In this context, the MSR model turns out to be the most suitable specification, since it is the closest translation of the mixture of distribution hypothesis into an empirical estimation. Indeed, Chauvet and Hamilton (2005) shows that a simple Markov switching model fits particularly well the data when these are suspected to follow a mixture of distributions.

To sum up, this thesis, revolving around the triad “information – volatility – price”, is organized as follows. The theoretical Section 1 focuses in particular on the “volatility – price” relation and develops a theoretical model extending the standard CAPM model to a world of possibly constrained investors, leading to an extra idiosyncratic variance term in the pricing equation. The empirical Section 2 is more “article-oriented” and investigates an empirical issue where deviation phenomena could be at work: the index effect. Then comes Section 3 that introduces the concept of informative “signals” and develops a theoretical model around the whole triad “information – volatility – price”

on the basis of the model of Section 1. The empirical “article-oriented” Section 4 focuses on one precise aspect of this extended model: the link “micro information” – volatility”. And this dissertation comes to an end with the final conclusion.

1 Constrained Investors and the Cost of Capital

1.1 Literature

As already sketched in the previous general introduction, the basics of the framework of our theoretical model are classical. We start with the CAPM model of Sharpe (1964), Jack Treynor (1961), John Lintner (1965) and Jan Mossin (1966), which has been used for example in the famous article of Grosman and Stiglitz (1980), and many other authors. Like Lambert, Leuz and Verrecchia (2007), we rather opt for the price version of the CAPM (strictly equivalent to the return representation, at least in a two-period setting), given already by Fama (1976). We then introduce the possibility for the investors not to be optimally diversified. Some authors, like Piazzesi, Schneider and Tuzel (2007) have already considered the case of those "constrained" investors, but in a quite different framework and setting, and with a different intuition. Brennan (1975), Goldsmith (1976), Levy (1978), Mayshar (1979) (1981), Leape (1987) and Viard (1995) examine the asset pricing effects of fixed holding costs, which in turn may result in positive or negative deviations from the optimal CAPM holdings. Contrary to them, we do not consider one specific type of deviation due to one specific cause, but we rather introduce exogeneous deviations without saying anything about their causation. They can potentially encompass all elements that the CAPM does not take into account: the fact that investors do not actually optimize their future wealth but present another (unknown) maximization behavior, the fact that there are transaction costs, holding costs, taxes, liquidity shortfalls, or the fact that investors hold only partial information, or present limited rationality, etc. As a matter of fact, whereas we do not focus on the causes of those deviations, we are mainly interested in the consequences of such deviations on the final pricing equation. To our knowledge, nobody has considered yet the under-diversification problem as we do in this dissertation.

Most of articles using CAPM and, more generally, investigating asset pricing issues, as in Fama (1976), Miller (1977), Li and Fleisher (2004) and Greenwood (2005) (some examples among many others), are not really concerned with the supply side of the market and therefore assume the assets supply to be fixed and constant. To our knowledge, there exist only rare attempts to consider supply shocks, as Braun and Larrain (2008) who investigate how IPOs affect the pricing of other assets, although they keep the supply of all assets constant. Since the supply is not what we are the most

interested in, we follow the traditional way: we also suppose a constant supply throughout the main model, even if we propose to relax this assumption in a later short subsection and to introduce a supply linear in the price, to offer a more complete coverage to the reader.

In a homogeneous expectations CAPM model with positive stock supplies, a possible constraint on short-selling does not matter since it is not binding. Indeed, in such a setting, all investors are perfectly identical so that their optimal quantities must be the same. If the supply is strictly positive, they will equally share the supply between them and hold positive amounts of each asset. Since optimal quantities are positive, the constraint is not binding. However, as soon as deviations are introduced, this constraint may become binding and affect asset pricing, because optimal quantities can become negative. Literature has tackled this short-selling issue in a universe of heterogeneous expectations, and most of authors have concluded that short-selling constraints tend to drive prices up, since they keep "pessimistic" investors out of the market (see Miller (1977) and Harrison and Kreps (1978) for example). However, using a rational expectations model, Diamond and Verrecchia (1987) show that constraining short sales reduces the informational efficiency of prices but does bias them upward. And Bai, Chang and Wang (2006) further demonstrate that under rational expectations, short-sale constraints can in fact cause asset prices to decrease. Indeed, since short-selling constraints make pessimistic investors unable to trade, prices are less informative and uncertainty increases. If agents are risk averse, prices are to decrease to compensate this new volatility. Since our model (like Miller (1977)) considers the risk (basically $\mathbf{\Sigma}$) as exogenous, it cannot take such informational effect into account, so that our discussion of the impact of a short-selling constraint in Section 1.6.4 does not pretend to be complete and thorough, but only to give some insight on this issue in the context of our model.

1.2 Conventions

Scalars: *italic* (eg. : b, R, …)

Vectors: **bold lower case** ($\mathbf{w}$, $\mathbf{m}$, …)

Matrices: **BOLD UPPER CASE** ($\mathbf{X}$, $\mathbf{\Sigma}$, …)

Random variable: ~ ($\tilde{\mathbf{p}}$)

1.3 Notations

Those notations are valid in both models of Section 1 and Section 3. But some of them will only be used for the first time in Section 3. Where necessary, we precise whether the variable is expressed in "value" (typically, $ for example), or in "quantity" (typically, number of shares).

N = number of investors on the market;

K = number of risky assets on the market;

$\mathbf{w}^E$ = [N×1] vector of initial wealth endowment (in $);

$\mathbf{w}$ = [N×1] vector of current wealth, after the new strategic allocation (in $);

$\mathbf{m}^E$ = [N×1] vector of initial endowments in the risk-free asset (in $);

$\mathbf{m}$ = [N×1] vector of the current holdings in the risk-free asset, after the new strategic allocation (decision variable) (in $);

$\mathbf{X}^E$ = [N×K] matrix of initial endowments in the risky assets (in number of shares);

$\mathbf{X}$ = [N×K] matrix of risky assets hold at time t, after the new strategic allocation (decision variable) (in number of shares);

$\mathbf{x}_i$ = [K×1] vector of the number of shares detained in the risky assets by investor i (i.e.: the transposed i^{th} row of matrix $\mathbf{X}$), after the new strategic allocation (decision variable) (number of shares);

R_f = gross return of the risk-free asset (i.e.: 1 + risk-free rate);

$\mathbf{p}$ = [K×1] vector of the current equilibrium prices of the K risky assets;

$\tilde{\mathbf{p}}$ = [K×1] vector of the future prices (or, equivalently, pay-offs) of the K risky assets;

$\boldsymbol{\mu}$ = [K×1] vector of the expectations of the future prices $\tilde{\mathbf{p}}$ of the K risky assets;

$\boldsymbol{\Sigma}$ = [K×K] Variance-Covariance matrix of the future prices $\tilde{\mathbf{p}}$ of the K risky assets:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1,j} & \cdots & \sigma_{1,K} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \sigma_{1,j} & \cdots & \sigma_j^2 & \cdots & \sigma_{j,K} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ \sigma_{1,K} & \cdots & \sigma_{j,K} & \cdots & \sigma_K^2 \end{pmatrix}$$

$U(.)$ = investor utility function;

b = investor risk aversion coefficient;

$\mathbf{s}$ = [K×1] vector of the risky assets supply (in number of shares);

s_k = k^{th} element of $\mathbf{s}$, i.e. the supply of the k^{th} asset (in number of shares);

$\bar{\mathbf{s}}$ = [K×1] vector of the average per capita supply of risky assets: $\bar{\mathbf{s}} = \dfrac{\mathbf{s}}{N}$.

$\tilde{p}_M \equiv \mathbf{s}'\tilde{\mathbf{p}}$ is the future market price such that: $\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \right)$

$\tilde{p}_{\underline{M}} \equiv \tilde{p}_M - s_j \tilde{p}_j$ is the future price of the market composed of all assets excluding asset j:

$$\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_{\underline{M}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_j \\ \mu_{\underline{M}} \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_{j,\underline{M}} \\ \sigma_{j,\underline{M}} & \sigma_{\underline{M}}^2 \end{pmatrix} \right)$$

$\tilde{R}_j = \dfrac{\tilde{p}_j - p_j}{p_j}$ is the future return of asset j.

$\tilde{z}_{j,q}$ is the q^{th} signal about future price $\tilde{p}_j$ of asset j: $\tilde{z}_{j,q} \equiv \tilde{p}_j + \sqrt{v_{j,q}}\,\tilde{u}_{j,q}$

with $v_{j,q}$ the variance of the noise of the signal $\tilde{z}_{j,q}$;

$\tilde{u}_{j,q} \sim N(0,1)$.

1.4 Terminology

The theoretical model developed in Section 1 and Section 3 uses two different adjectives to qualify the distribution of future prices (or the moments of this distribution): unconditional and conditional.

By unconditional (distribution, expectation, variance or covariance), we imply "unconditional to signals". It means that the unconditional distribution of future prices is the distribution of prices determined without the knowledge of any signal. This "unconditional distribution" is the one used in Section 1, since the latter develops the model in the absence of signals.

On the contrary, by conditional (distribution, expectation, variance or covariance), we imply "conditional to signals". It means that the conditional distribution of future prices is the distribution of prices determined with the knowledge of micro and macro signals. These signals provide investors with additional information which they can use to update their "guess" of the future prices. It follows that the conditional distribution will of course be more accurate than the unconditional one. This "conditional distribution" is the one used in Section 3, since the latter develops further the model of Section 1 by introducing the signals in the pricing process.

By default, the distribution or its moments are considered to be "unconditional". When they are taken conditional to the signals, the adjective "conditional" will be added.

1.5 Assumptions

- Our model is a discrete model with one period and two dates:
    - Today = current time = beginning of the period
    - Tomorrow = future time = end of the period

  All "variables" of today are observed and known (for ex.: $\mathbf{p}$), while all "variables" of tomorrow are unknown and random, and appear with a tilt: ~ (for ex.: $\tilde{\mathbf{p}}$).
- Each investor i is endowed with initial amount m_i^E of risk-free asset and a vector $\mathbf{x}_i^E$ of risky assets (in number of shares).
- All investors maximize the same CARA utility function (for example, a negative exponential utility function: $U(\mathbf{w}) = -\exp(-b\mathbf{w})$); other utility functions exist, but the CARA utility function is standard in the literature on this topic (Grossman & Stiglitz (1980) and Easley and O'Hara (2004) for example).
- The gross return (R_f) of the risk-free asset is constant through time;

- Investors can be long or short on all the assets without specific constraints but the budget constraint and the market clearing conditions; as long as investors are perfectly diversified and hold the CAPM optimal quantities, presence of short-sale constraints does not matter. But as soon as deviations from the CAPM optimality are introduced, short-sale constraints may become binding, even if expectations of investors are homogeneous.
- The asset supply **s** is fixed: **s**(**p**) = **s**; this assumption will be relaxed to allow the asset supply to increase linearly with the price.
- The probability distribution function (pdf) of the futures prices of the risky assets is multivariate normal and is the same (homogeneous expectations and common knowledge) for all investors: $\tilde{\mathbf{p}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The assumption of Gaussian prices may seem annoying, because it can potentially result in negative prices in some extreme cases. In this respect, log-normal prices could be a nice alternative. However, Gaussian prices allow for explicit closed-form solutions, while log-normal prices do not, at least in this particular setting. To give just two examples, the Laplace transform of a log-normal random variable has no explicit solution, which would stop us to go further than Equation (1-3), unless we postulate directly a mean-variance setup and start with Equation (1-4). However, in Section 3, there would be no explicit (or at least not straightforward) form for the distribution of prices conditional to signals if unconditional prices are not Gaussian. Finally, the use of Gaussian prices is not unusual in the literature: see for example Easley and O'Hara (2004).

1.6 Model development[1] and main results

1.6.1 Unconstrained investors (benchmark case)

1.6.1.1 Derivation of the optimal matrix of demanded quantities

Each investor *i* is initially endowed with an amount m_i^E of risk-free asset and a vector $\mathbf{x}_i^E$ of risky assets (in number of shares), which results in a total initial wealth: $w_i^E = m_i^E + \mathbf{x}_i^{E\prime}\mathbf{p}$. He has the opportunity to change the amounts invested in the risk-

[1] We thank Michel Levasseur for having helped us to set up the initial structure of this pricing model.

free and risky assets and to define a new strategic allocation in order to maximize its future expected utility, provided that his wealth remains constant. The new strategic allocation (m_i, $\mathbf{x}_i$) of each investor i must therefore satisfy the following budget constraint (written here for all investors):

$$\mathbf{w}^E = \mathbf{w} \Leftrightarrow \mathbf{m}^E + \mathbf{X}^E\mathbf{p} = \mathbf{m} + \mathbf{X}\mathbf{p} \tag{1-1}$$

The future wealth is given by:

$$\tilde{\mathbf{w}} = R_f\mathbf{m} + \mathbf{X}\tilde{\mathbf{p}} \tag{1-2}$$

Each investor i is supposed to choose the vector $\mathbf{x}_i$ (the decision variable) that maximizes the expected utility of his future wealth:

$$\text{Max} \quad E(U(\tilde{\mathbf{w}}) = -E\left(\exp(-b\tilde{\mathbf{w}})\right) = -\exp\left\{-b\left[E(\tilde{\mathbf{w}}) - \frac{b}{2}Var(\tilde{\mathbf{w}})\right]\right\} \tag{1-3}$$

w.r.t. $\mathbf{X}$

s.t. (1-1) is satisfied.

Note that the last equality is obtained thanks to the Laplace transform of a Gaussian distribution.

If we incorporate the budget constraint in the objective function, if we replace $\tilde{\mathbf{w}}$ by its expression of Equation (1-2) and if we remember that maximizing -exp(-f(x)) is equivalent to maximizing f(x) because the exponential function is an monotonously increasing function, we get the following general maximization problem:

$$\text{Max} \quad R_f\mathbf{w} + \mathbf{X}(\boldsymbol{\mu} - R_f\mathbf{p}) - \frac{b}{2}diag(\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}') \tag{1-4}$$

w.r.t. $\mathbf{X}$

Or alternatively, under a form specific to each investor i:

$$\text{Max} \quad R_f w_i + \mathbf{x}_i'(\boldsymbol{\mu} - R_f\mathbf{p}) - \frac{b}{2}\mathbf{x}_i'\boldsymbol{\Sigma}\mathbf{x}_i \tag{1-5}$$

w.r.t. $\mathbf{x}_i$

Since all investors have the same utility function (with the same risk aversion coefficient) and the same future price expectations, the vector of the optimal quantities bought of each risky asset will be the same for each **unconstrained** investor *i*:

$$\mathbf{x}_i^U = \boldsymbol{\Sigma}^{-1} \frac{\boldsymbol{\mu} - R_f \mathbf{p}}{b} \text{ for } i = 1,\ldots, N \tag{1-6}$$

and the complete matrix of quantities demanded by all investors is:

$$\mathbf{X}^U = \mathbf{1}_N \left(\frac{\boldsymbol{\mu}' - R_f \mathbf{p}'}{b} \right) \boldsymbol{\Sigma}^{-1} \tag{1-7}$$

where $\mathbf{1}_N$ denotes a [N×1] vector of ones.

1.6.1.2 Derivation of the equilibrium prices

The market clearing condition states that:

$$\mathbf{X}^{U'} \mathbf{1}_N = \mathbf{s} \Leftrightarrow \boldsymbol{\Sigma}^{-1} \frac{\boldsymbol{\mu} - R_f \mathbf{p}}{b} \mathbf{1}_N' \mathbf{1}_N = \mathbf{s} \Leftrightarrow \mathbf{p} = \frac{1}{R_f} \left[\boldsymbol{\mu} - \frac{b}{N} \boldsymbol{\Sigma} \mathbf{s} \right] = \frac{1}{R_f} \left[\boldsymbol{\mu} - b \boldsymbol{\Sigma} \bar{\mathbf{s}} \right] \tag{1-8}$$

where $\bar{\mathbf{s}}$ is the vector of the average per-capita supply of risky assets: $\bar{\mathbf{s}} = \frac{\mathbf{s}}{N}$, and $\frac{b}{N}$ can be interpreted as the inverse of the aggregate risk tolerance of the market: as the risk aversion coefficient decreases and the number of investors increases, the market aggregate risk tolerance increases.

Or alternatively, the price p_j of each asset *j* is given by:

$$p_j = \frac{\mu_j - \frac{b}{N} \sum_{k=1}^{K} \sigma_{j,k} s_k}{R_f} \tag{1-9}$$

This result is standard in asset pricing literature: it shows that the current prices are equal to the expected future prices corrected for the (systematic) risk and discounted at the risk-free rate.

Equation (1-8) can be expressed at the market level, if we define the market price p_M as the total capitalization of the market:

$$p_M \equiv \mathbf{s}'\mathbf{p} = \mathbf{s}' \frac{1}{R_f} \left[\boldsymbol{\mu} - \frac{b}{N} \boldsymbol{\Sigma} \mathbf{s} \right] = \frac{\mu_M - \frac{b}{N} \sigma_M^2}{R_f} \tag{1-10}$$

with
$$\mu_M \equiv E(\tilde{p}_M) \equiv E(\mathbf{s}'\tilde{\mathbf{p}}) = \mathbf{s}'E(\tilde{\mathbf{p}}) = \mathbf{s}'\boldsymbol{\mu}$$
$$\sigma_M^2 \equiv Var(\tilde{p}_M) \equiv Var(\mathbf{s}'\tilde{\mathbf{p}}) = \mathbf{s}'Var(\tilde{\mathbf{p}})\mathbf{s} = \mathbf{s}'\boldsymbol{\Sigma}\mathbf{s}$$

This new concept of market price also allows us to define the covariance between stock j and market prices $\sigma_{j,M}$ and to rewrite the scalar asset pricing Equation (1-9) accordingly:

$$p_j = \frac{\mu_j - \frac{b}{N}\sigma_{j,M}}{R_f} \tag{1-11}$$

with $\begin{pmatrix} \sigma_{1,M} \\ \vdots \\ \sigma_{j,M} \\ \vdots \\ \sigma_{K,M} \end{pmatrix} \equiv \begin{pmatrix} Cov(\tilde{p}_1, \tilde{p}_M) \\ \vdots \\ Cov(\tilde{p}_j, \tilde{p}_M) \\ \vdots \\ Cov(\tilde{p}_K, \tilde{p}_M) \end{pmatrix} \equiv \begin{pmatrix} Cov(\tilde{p}_1, \mathbf{s}'\tilde{\mathbf{p}}) \\ \vdots \\ Cov(\tilde{p}_j, \mathbf{s}'\tilde{\mathbf{p}}) \\ \vdots \\ Cov(\tilde{p}_K, \mathbf{s}'\tilde{\mathbf{p}}) \end{pmatrix}$

$$= Cov(\tilde{\mathbf{p}}, \mathbf{s}'\tilde{\mathbf{p}}) = Cov(\tilde{\mathbf{p}}, \tilde{\mathbf{p}})\mathbf{s} = Var(\tilde{\mathbf{p}})\mathbf{s} = \boldsymbol{\Sigma}\mathbf{s}$$

It follows from (1-11) that it is now possible to relate the aggregate risk tolerance *b/N* to the market price:

$$\frac{b}{N} = \frac{\mu_M - R_f p_M}{\sigma_M^2} \tag{1-12}$$

The aggregate risk tolerance of the market determines the risk premium for market-wide risk.

Plugging (1-12) into (1-8) provides another interesting expression for the asset equilibrium prices:

$$\mathbf{p} = \frac{1}{R_f}\left[\boldsymbol{\mu} - \frac{\mu_M - R_f p_M}{\sigma_M^2}\boldsymbol{\Sigma}\mathbf{s}\right] \tag{1-13}$$

In scalar notations, it is easy to see (plug (1-12) into (1-11)) that we finally obtain the CAPM in terms of prices (see Chapter 8 of Fama (1976)):

$$p_j = \frac{\mu_j - \frac{\mu_M - R_f p_M}{\sigma_M^2}\sigma_{j,M}}{R_f} \tag{1-14}$$

1.6.1.3 Cost of capital

Consistently with Lambert, Leuz and Verrecchia (2007), we define the cost of capital to be the expected return on the firm's stock, i.e.:

$$E(\tilde{R}_j) = E\left(\frac{\tilde{p}_j - p_j}{p_j}\right) = \frac{E(\tilde{p}_j) - p_j}{p_j}$$

$$= \frac{(R_f - 1)\mu_j + \frac{\mu_M - R_f p_M}{\sigma_M^2}\sigma_{j,M}}{\mu_j - \frac{\mu_M - R_f p_M}{\sigma_M^2}\sigma_{j,M}} \tag{1-15}$$

Or also:

$$E(\tilde{R}_j) = \frac{\mu_j}{\frac{\mu_j - \frac{\mu_M - R_f p_M}{\sigma_M^2}\sigma_{j,M}}{R_f}} - 1 = \frac{R_f}{1 - \frac{\mu_M - R_f p_M}{\mu_j \sigma_M^2}\sigma_{j,M}} - 1 \tag{1-16}$$

The cost of capital $E(\tilde{R}_j)$ is increasing in R_f, b and $\sigma_{j,M}$ and decreasing in μ_j and N. Remember that $\frac{\mu_M - R_f p_M}{\sigma_M^2} = \frac{b}{N}$, so that if b and N are fixed, an increase in μ_M or a decrease in R_f or σ_M^2 are automatically compensated by a corresponding increase of the market capitalization p_M.

1.6.2 Constrained investors

1.6.2.1 Derivation of the optimal matrix of demanded quantities

Let us now imagine that investors are not able to detain perfectly diversified portfolios because they are constrained in such a way that their actual detention (say, $x_{i,j}^C$) in asset j deviates from their optimal asset *j* detention (say, $x_{i,j}^*$) by a quantity $\varepsilon_{i,j}$:

$$\mathbf{x}_i^C \equiv \mathbf{x}_i^* + \boldsymbol{\varepsilon}_i \text{ or equivalently } \mathbf{X}^C \equiv \mathbf{X}^* + \mathbf{E} \tag{1-17}$$

$$\text{with } \mathbf{X}^* = \begin{pmatrix} x^*_{1,1} & \cdots & x^*_{1,j} & \cdots & x^*_{1,K} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x^*_{i,1} & \cdots & x^*_{i,j} & \cdots & x^*_{i,K} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x^*_{N,1} & \cdots & x^*_{N,j} & \cdots & x^*_{N,K} \end{pmatrix} \text{ and } \mathbf{E} = \begin{pmatrix} \varepsilon_{1,1} & \cdots & \varepsilon_{1,j} & \cdots & \varepsilon_{1,K} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \varepsilon_{i,1} & \cdots & \varepsilon_{i,j} & \cdots & \varepsilon_{i,K} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ \varepsilon_{N,1} & \cdots & \varepsilon_{N,j} & \cdots & \varepsilon_{N,K} \end{pmatrix}$$

The above equation is a formal definition of the concept of "constrained investors". An investor is constrained in an asset *j* if he does not hold the optimal quantities $x^*_{i,j}$ for this asset but holds a quantity $x^C_{i,j}$ which deviates from the optimal quantity $x^*_{i,j}$ by a number of shares $\varepsilon_{i,j}$. This deviation may have numerous causes: blockholders, familial ownership, index firms, large and famous overweighted firms (hence the existence of positive deviation, i.e. investors detain on average too much of them), small and anonymous firms (hence the existence of a negative deviation, i.e., investors detain on average too less of them), holding costs. The purpose of our model is not to explain the origin of those deviations, but rather to analyze the effect of those deviations on the traditional CAPM equilibrium prices.

Now, how can we exactly understand $x^*_{i,j}$ in Equation (1-17)? In which sense is $x^*_{i,j}$ optimal? At this point, there are different ways to define this $x^*_{i,j}$.

One could define it as the optimal quantities of asset *j* that should hold an investor *i* in a world where everyone holds the optimal quantities $x^*_{i,j}$. Those optimal quantities of reference would then be the ones given by Equation (1-6), and the deviation of investor *i* in asset *j* would be defined as the difference between his actual (possibly constrained) holding $x^C_{i,j}$ in asset j and the optimal holding of reference $x^*_{i,j}$.

However, there is maybe a better "optimal" $x^*_{i,j}$ to choose as a reference. Indeed, when some investors begin to deviate from their optimal portfolio, the prices change, and the optimal quantities to hold change as well, since that investors can take into account the fact that deviations exist on other assets and/or investors to revise their optimal holdings. Therefore, we define $x^*_{i,j}$ as the ***optimal*** quantity of asset *j* that investor *i* should hold to maximize his utility ***given that*** he has chosen his quantities $x^C_{i,k}$ (***not necessarily optimal***) in all other assets $k \neq j$. In other words, investor *i* "observes" the vector of prices **p** prevailing on the market and the vector of **his** current (and possibly

constrained) holdings in all assets $\mathbf{x}^{C}_{i,(exceptj)}$ other than *j* (he does not need to know the holdings of other investors to make his choice). On the basis of this information, he can determine his optimal holdings $x^{*}_{i,j}$ in asset *j*.

$x^{*}_{i,j}$ is thus defined as the $x_{i,j}$ maximizing:

$$\begin{aligned} &\text{Max} \quad R_f w_i + x_{i,j}(\mu_j - R_f p_j) + \sum_{\substack{k=1 \\ k\neq j}}^{K} x^{C}_{i,k}(\mu_k - R_f p_k) - \frac{b}{2}\left(\sigma_j^2 x_{i,j}^2 + 2x_{i,j}\sum_{\substack{k=1 \\ k\neq j}}^{K}\sigma_{j,k}x^{C}_{i,k}\right) \\ &\text{w.r.t.} \quad x_{i,j} \end{aligned} \tag{1-18}$$

The first order condition of this maximization problem is given by:

$$\begin{aligned} &\mu_j - R_f p_j - b(\sigma_j^2 x^{*}_{i,j} + \sum_{\substack{k=1 \\ k\neq j}}^{K}\sigma_{j,k}x^{C}_{i,k}) = 0 \\ &\Rightarrow x^{*}_{i,j} = \frac{\mu_j - R_f p_j - b\sum_{\substack{k=1 \\ k\neq j}}^{K}\sigma_{j,k}x^{C}_{i,k}}{b\sigma_j^2} \end{aligned} \tag{1-19}$$

And the process is repeated for each *j=1,...,K*. Investor *i* has thus *K* maximization problems to solve. Note that in the standard case (see Equation (1-5)), the maximization of the utility w.r.t. all variables is equivalent to *K* maximization problems w.r.t. each variable separately.

The following numerical example will resolve any ambiguity and illustrate how a deviation affects the initial equilibrium situation in terms of prices and optimal quantities.

Let us imagine a market with one risky stock and two investors, Mister One and Mister Two. The gross risk-free rate $R_f = 1$, the risk aversion coefficient $b = 0.5$, the total supply in the risky asset $s = 10$, the expected price of the risky asset $\mu = 200$ and the variance of the risky asset $\sigma^2 = 40$. It follows that the optimal CAPM holdings x_i^U are equal to 5 for *i = 1, 2* and the equilibrium price of the risky asset *p* is equal to 100 (by a straightforward application of Equation (1-9)). This is the initial equilibrium.

Figure 1-1 illustrates the situation graphically. The aggregate demand curve (*D*) of the risky asset and the aggregate supply curve (*S*) of the risky asset are plotted together in order to determine the equilibrium prices and quantities.

The aggregate demand curve equation (*D*) expresses the demanded price (p^D) as a function of the aggregate demanded quantities ($q^D = x_1 + x_2$) and can be obtained by a direct application of Equation (1-9):

$$D \equiv p^D = \frac{\mu - \frac{b\sigma^2}{N} q^D}{R_f} \tag{1-20}$$

As mentioned in the assumptions of Section 1.5, the aggregate supply curve is supposed to be fixed and inelastic:

$$S \equiv q^S = s \tag{1-21}$$

If we replace the parameters of these equations by their values we obtain:

$$D \equiv p^D = 200 - 10q^D \tag{1-22}$$

$$S \equiv q^S = 10 \tag{1-23}$$

The intersection of the demand and supply curves gives the equilibrium price and quantities of the risky asset: $p^D = p^S = 100$ and $q^D = q^S = 10$. This is the point (1) in Figure 1-1.

We may now consider to introduce a deviation. Indeed, investor One may decide (for a number of reasons) to change its position in the risky asset and adopt a deviation $\varepsilon_1 = 2$. His new temporary constraint holding is then 5+2=7 (as long as the prices have not changed yet). His motivations to do so can be various. Such a behavior is suboptimal in a CAPM perspective, since investor One does not maximize anymore the expectation of its future utility. It does not necessarily mean that investor A is irrational. Motivations to deviate from the "optimal" holdings are numerous: alternative optimizing behavior, alternative utility function, holding costs, index tracking, block holding, particular constraints, etc. Deviations here are not synonymous of anomalies.

This sudden additional holding of 2 in the risky asset by investor One provokes a translation of the aggregate demand curve *D* to the right (by two units). The new aggregate demand curve is denoted *D'* and has the following equation:

$$D \equiv p^D = \frac{\mu - \frac{b\sigma^2}{N}(q^D - \varepsilon_1)}{R_f} \tag{1-24}$$

If we replace the parameters of this equations by their values we obtain:

$$D \equiv p^D = 220 - 10q^D \tag{1-25}$$

The new holdings of investor One (5+2) and Two (5) are the result of a temporary situation. Indeed, the excess demand of One provokes an imbalance between the aggregate demand (7 (One) + 5 (Two) = 12) and the global supply (10). This is illustrated by the point (2) Figure 1-1 which is obviously not an equilibrium. Therefore, the price of the risky stock will rise until aggregate demand equals aggregate supply. It can be shown that it is going to increase from 100 to 120 in order to reach the equilibrium (direct application of Equation (1-31) below). Indeed, when the price is equal to 120, the new optimal quantities to hold for investors One and Two are equal to (see Equation (1-19)):

$$x_1^* = x_2^* = \frac{\mu - R_f p}{b\sigma^2} = \frac{200 - 1\times 120}{0.5\times 40} = 4$$

Notice that with one risky asset, the optimal quantities of investors One and Two are equal.

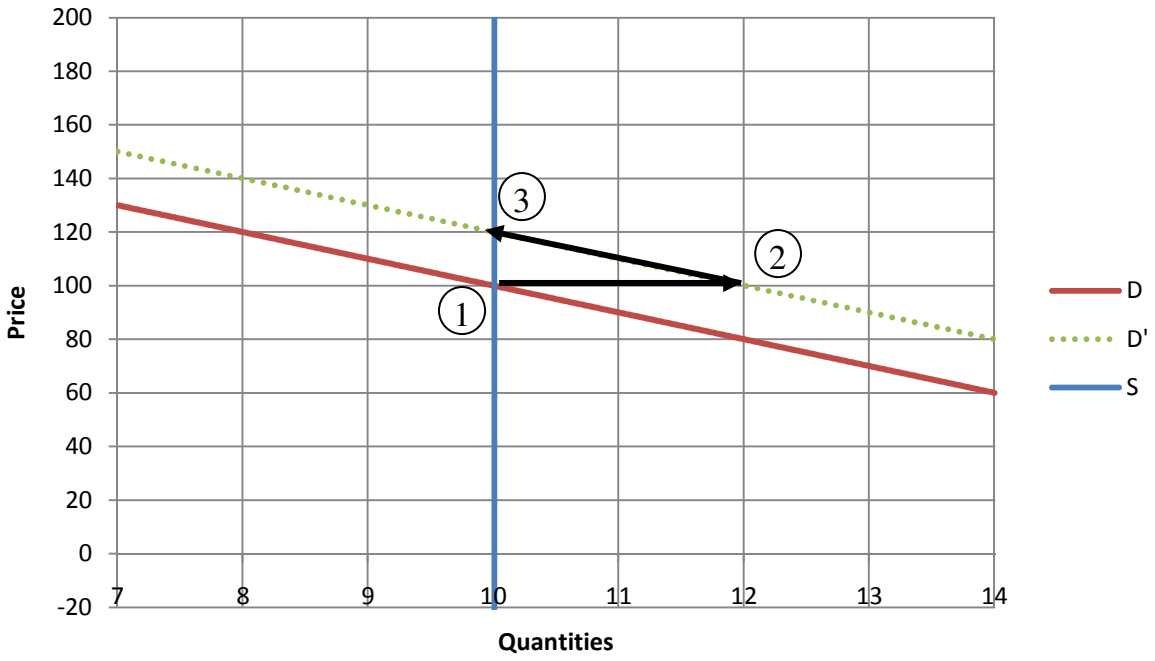
The actual holdings of investors One and Two, x_1^C and x_2^C, after taking the potential deviations into account, will be the following (by application of Equation (1-17)):

$$x_1^C = x_1^* + \varepsilon_1 = 4 + 2 = 6$$
$$x_2^C = x_2^* + \varepsilon_2 = 4 + 0 = 4$$

The aggregate demand ($x_1^C + x_2^C = 6 + 4 = 10$) is now equal to the aggregate supply. The new price of 120 ensures the equilibrium.

It is important to notice that the deviation of investor A is defined with respect to the optimal holdings of investor One given the new price of 120: $\varepsilon_1 = x_1^C - x_1^* = 6-4 = 2$, and **not** with respect to the initial CAPM optimal holdings valid with the former price of 100: $\varepsilon_1 \neq x_1^C - x_1^U = 6-5 = 1$. Indeed, x_1^U was the optimal quantities to hold when the price was 100. As the price is now 120, the optimal quantities to hold have decreased from 5 to 4, and they are thus now equal to x_1^*. It explains why the aggregate deviation can be non-null, as the optimal quantities to hold change as deviations appear.

Figure 1-1: Impact of the deviation on the equilibrium price and quantities



In summary:

1) Initial equilibrium (no deviations):

   $x_1^* + x_2^* + \varepsilon_1 + \varepsilon_2 = x_1^U + x_2^U + \varepsilon_1 + \varepsilon_2 = 5+5+0+0 = s = 10$

   $p = 100$

2) Final equilibrium (with a deviation of 2 for investor One):

   $x_1^C + x_2^C = x_1^* + x_2^* + \varepsilon_1 + \varepsilon_2 = 4+4+2+0 = s = 10$

   The price has increased from 100 to 120.

   The optimal holdings x_i^* have both decreased from 5 to 4.

   The sum of the optimal holdings (4+4) and of the aggregate deviation (2) sum up to the aggregate supply (10). The equilibrium is achieved.

Graphically, the new equilibrium is defined by the intersection of the new demand (*D'*) and supply (*S*) curves: $p^D = p^S = 120$ and $q^D = q^S = 10$. This is the point (3) in Figure 1-1.

We continue the development of the model interrupted at Equation (1-19). By substituting (1-17) into (1-19) (replace $x_{i,j}^*$ by $x_{i,j}^C - \varepsilon_{i,j}$), we get:

$$\boldsymbol{\mu} - R_f \mathbf{p} - b\boldsymbol{\Sigma}\mathbf{x}_i^C = -b\boldsymbol{\varphi}_i \tag{1-26}$$

where $\boldsymbol{\varphi}_i = diag(diag(\boldsymbol{\Sigma}) * \boldsymbol{\varepsilon}_i') = \begin{pmatrix} \sigma_1^2 \varepsilon_{i,1} \\ \vdots \\ \sigma_j^2 \varepsilon_{i,j} \\ \vdots \\ \sigma_K^2 \varepsilon_{i,K} \end{pmatrix}$

One can check that the above first-order condition (1-26) derives in fact from the following maximization problem, which is similar to the unconstrained case except that the variance-covariance matrix is modified in a certain way to take the deviations into account:

$$\text{Max} \quad R_f w_i + \mathbf{x}_i^{C'} \left(\boldsymbol{\mu} - R_f \mathbf{p}\right) - \frac{b}{2} \mathbf{x}_i^{C'} \left(\boldsymbol{\Sigma} - \mathbf{I}_k \bullet \left(\boldsymbol{\varphi}_i \mathbf{1}_N'\right)\right) \mathbf{x}_i^C \tag{1-27}$$

$$\text{w.r.t.} \quad \mathbf{x}_i^C$$

where $\bullet$ stands for an "element-by-element product", so that

$$\boldsymbol{\Sigma} - \mathbf{I}_k \bullet \left(\boldsymbol{\varphi}_i \mathbf{1}_N'\right) = \begin{pmatrix} \sigma_1^2(1-\varepsilon_{i,1}) & \cdots & \sigma_{1,j} & \cdots & \sigma_{1,K} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ \sigma_{1,j} & \cdots & \sigma_j^2(1-\varepsilon_{i,j}) & \cdots & \sigma_{j,K} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ \sigma_{1,K} & \cdots & \sigma_{j,K} & \cdots & \sigma_K^2(1-\varepsilon_{i,K}) \end{pmatrix}$$

The quantities of each asset *j* demanded by each investor *i* are given by:

$$\mathbf{x}_i^C = \boldsymbol{\Sigma}^{-1} \frac{\boldsymbol{\mu} - R_f \mathbf{p}}{b} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\varphi}_i \tag{1-28}$$

Or with all investors in the same matrix:

$$\mathbf{X}^{C} = \mathbf{1}_N \frac{\boldsymbol{\mu}' - R_f \mathbf{p}'}{b} \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}^{-1} \tag{1-29}$$

where $\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1' \\ \vdots \\ \boldsymbol{\varphi}_i' \\ \vdots \\ \boldsymbol{\varphi}_N' \end{pmatrix}$

1.6.2.2 Derivation of the equilibrium prices

The market clearing condition imposes that:

$$\mathbf{X}^{C'} \mathbf{1}_N = \mathbf{s} \Leftrightarrow \boldsymbol{\Sigma}^{-1} \frac{\boldsymbol{\mu} - R_f \mathbf{p}}{b} \mathbf{1}_N' \mathbf{1}_N + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Phi}' \mathbf{1}_N = \mathbf{s}$$

$$\Leftrightarrow \mathbf{p} = \frac{1}{R_f}\left[\boldsymbol{\mu} - \frac{b}{N}(\boldsymbol{\Sigma}\mathbf{s} - \boldsymbol{\varphi})\right] = \frac{1}{R_f}\left[\boldsymbol{\mu} - b(\boldsymbol{\Sigma}\bar{\mathbf{s}} - \bar{\boldsymbol{\varphi}})\right] \tag{1-30}$$

where:

$\boldsymbol{\varphi} = diag(diag(\boldsymbol{\Sigma}) * \boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma_1^2 \varepsilon_1 \\ \vdots \\ \sigma_j^2 \varepsilon_j \\ \vdots \\ \sigma_K^2 \varepsilon_K \end{pmatrix}$ and $\boldsymbol{\varepsilon} = \mathbf{E}' \mathbf{1}_N$ is a [K×1] vector of the <u>aggregate</u> deviation in each asset.

$\bar{\boldsymbol{\varphi}} = diag(diag(\boldsymbol{\Sigma}) * \bar{\boldsymbol{\varepsilon}}) = \begin{pmatrix} \sigma_1^2 \bar{\varepsilon}_1 \\ \vdots \\ \sigma_j^2 \bar{\varepsilon}_j \\ \vdots \\ \sigma_K^2 \bar{\varepsilon}_K \end{pmatrix}$ and $\bar{\boldsymbol{\varepsilon}} = \frac{\mathbf{E}' \mathbf{1}_N}{N}$ is a [K×1] vector of the <u>average</u> deviation in each asset.

In a scalar notation, the equilibrium price of asset j is equal to:

$$p_j = \frac{\mu_j - \frac{b}{N}\left(\sigma_{j,M} - \sigma_j^2 \varepsilon_j\right)}{R_f} \tag{1-31}$$

The most striking feature of this pricing equation in a universe of constrained investors, by comparison with the Equation (1-11) in a world of perfectly diversified investors, is that the idiosyncratic variance of the stock now matters. Indeed, the covariance between the stock and the market is not the only risk component of the equilibrium price anymore but is corrected by an idiosyncratic variance term. And the higher the absolute value of the aggregate deviation, the more sensitive the price of the stock to the specific risk relative to the latter.

The market price is given by:

$$p_M \equiv \mathbf{s}'\mathbf{p} = \mathbf{s}'\frac{1}{R_f}\left[\boldsymbol{\mu} - \frac{b}{N}(\boldsymbol{\Sigma}\mathbf{s} - \boldsymbol{\varphi})\right] = \frac{\mu_M - \frac{b}{N}(\sigma_M^2 - \mathbf{s}'\boldsymbol{\varphi})}{R_f} \tag{1-32}$$

The last equation shows that the market price increases (resp. decreases) as the aggregate deviation of the investors increases (resp. decreases).

1.6.2.3 Derivation of the expected utility

By substituting Equation (1-30) into (1-28), we have that:

$$\mathbf{x}_i^C = \boldsymbol{\Sigma}^{-1}\frac{\boldsymbol{\mu} - R_f\mathbf{p}}{b} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}_i = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\bar{\mathbf{s}} - \bar{\boldsymbol{\varphi}}) + \boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}_i = \bar{\mathbf{s}} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}}) \tag{1-33}$$

And by substituting (1-30) and (1-33) into (1-5) we obtain the expected utility of each investor *i*:

$$\begin{aligned}
&E(U(\tilde{w}_i)) \\
&= -\exp\left\{-b\left[R_f w_i + \mathbf{x}_i^{C\prime}(\boldsymbol{\mu} - R_f\mathbf{p}) - \frac{b}{2}\mathbf{x}_i^{C\prime}\boldsymbol{\Sigma}\mathbf{x}_i^C\right]\right\} \\
&= -\exp\left\{-b\left[R_f w_i + \left(\bar{\mathbf{s}}' + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\right)b(\boldsymbol{\Sigma}\bar{\mathbf{s}} - \bar{\boldsymbol{\varphi}}) - \frac{b}{2}\left(\bar{\mathbf{s}}' + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Sigma}\left(\bar{\mathbf{s}} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})\right)\right]\right\} \\
&= -\exp\left\{-b\left[\begin{array}{l} R_f w_i + b\left(\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}} - \bar{\mathbf{s}}'\bar{\boldsymbol{\varphi}} + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\bar{\mathbf{s}} - (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\varphi}}\right) \\ -\frac{b}{2}\left(\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}} + 2(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\bar{\mathbf{s}} + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})\right) \end{array}\right]\right\}
\end{aligned}$$

$$= -\exp\left\{-b\left[R_f w_i + b\left(\frac{\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}}{2} - \bar{\mathbf{s}}'\bar{\boldsymbol{\varphi}} - (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\varphi}}_t\right) - \frac{b}{2}\left((\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}_i - (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\varphi}}\right)\right]\right\}$$

$$= -\exp\left\{-b\left[R_f w_i + b\left(\frac{\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}}{2} - \bar{\boldsymbol{\varphi}}'\bar{\mathbf{s}} + \frac{\bar{\boldsymbol{\varphi}}'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\varphi}}}{2} - \frac{\boldsymbol{\varphi}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}_i}{2}\right)\right]\right\} \quad (1\text{-}34)$$

The "plus" sign after $R_f w_i$ in (1-34) must not surprise. If we want to keep the trade-off return-risk apparent in the expression, we must stop at the third line of the development:

$$E(U(\tilde{w}_i)) = -\exp\left\{-b\left[\begin{array}{l} R_f w_i + b\left(\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}} - \bar{\mathbf{s}}'\bar{\boldsymbol{\varphi}} + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\bar{\mathbf{s}} - (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\varphi}}\right) \\ -\frac{b}{2}\left(\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}} + 2(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\bar{\mathbf{s}} + (\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}})\right) \end{array}\right]\right\}$$

where we find the familiar $-\frac{b}{2}$ before the risk component of the utility. Note that in a universe of unconstrained investors, it reduces to:

$$E(U(\tilde{w}_i)) = -\exp\left\{-b\left[R_f w_i + b\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}} - \frac{b}{2}\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}\right]\right\} = -\exp\left\{-b\left[R_f w_i + \frac{b}{2}\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}\right]\right\} \quad (1\text{-}35)$$

with $R_f w_i + b\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}$ and $\frac{b}{2}\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}$ the classical return and risk components of the investor utility.

The three last terms of (1-34) decompose the general effect of deviations on the expected utility. Before beginning with their respective interpretations, we remind the intuitive meaning of the two following variables:

$\bar{\boldsymbol{\varphi}}$ is a [K×1] vector; i.e. it consists of one element $\bar{\varphi}_j$ for each stock on the market. It is equal to the average deviation of the whole market in each asset (multiplied by the variance of each asset). Thus, if all investors are in positive deviation on a given asset j for example, $\bar{\varphi}_j$ will be positive for this asset j and will represent the average deviation in this asset (multiplied by the variance). As a consequence, $\bar{\boldsymbol{\varphi}}$ is the same for all investors (it does not depend on index "i").

$\boldsymbol{\varphi}_i$ is a [K×1] vector; i.e it consists of one element $\varphi_{i,j}$ for each stock on the market. It is equal to the deviation of investor i in each asset (multiplied by the variance of each asset). As a consequence, $\boldsymbol{\varphi}_i$ is not the same for all investors (it does depend on index "i"), but each $\boldsymbol{\varphi}_i$ is different for each investor.

We examine now the last three terms of (1-34) one by one and we try to give them an economic intuition:

1) $-\overline{\boldsymbol{\varphi}}'\overline{\mathbf{s}}$ shows that the average overweight (resp. underweight) in a given asset j tends to decrease (resp. increase) the expected utility of all investors, since it amounts to reducing (resp. increasing) the available supply of shares on the market as shown in Equation (1-31), and thus it leads to a price increase (resp. decrease).

2) $+\frac{\overline{\boldsymbol{\varphi}}'\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{\varphi}}}{2}$ shows that the imperfect diversification of certain investors benefits everybody. Everyone takes advantage from the "mistakes" of the others, and this advantage is independent of the sign of the deviation (overweight or underweight). The investors in deviation also benefit from this global effect, but it is clear that the benefit that they withdraw from this general effect will be much smaller in magnitude than the cost they will suffer from their own deviation. This cost appears in the following third term.

3) $-\frac{\boldsymbol{\varphi}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}_i}{2}$ shows that it is never good to be imperfectly diversified (at least according to the particular utility function we chose): the higher the deviation, the higher the loss of utility, all other things being equal. Contrary to effects 1) and 2) which are global and the same for each investor, this effect is individual and particular to each investor i.

Note also some interesting facts, that are maybe particular to this setting (CARA utility function, future wealth maximization behavior, …), but could be different in another framework:

- The model predicts that the higher the volatility on the market (higher variances and covariances in the $\boldsymbol{\Sigma}$ matrix), the higher the investors' utility. It follows

from the fact that $\bar{\mathbf{s}} > 0$ and that the magnitude of $\frac{\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}}{2}$ is likely to be of a larger order than the deviation terms. In a universe of unconstrained investors (i.e. with null deviations), the intuition of such an "anomaly" appears clearly in (1-35): the increased volatility has indeed a more positive impact on the return term $R_f w_{i,t} + b\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}$ (through the current prices decrease) than on the risk counterpart $\frac{b}{2}\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}$. The risk premium that investors demand to compensate a higher volatility more than compensates the loss in utility due to this additional volatility.

- In a world of perfectly diversified investors (i.e. with null deviations), the higher the number of assets *K* and the lower the number of investors *N* (if we accept to consider non-atomistic investors), the higher the utility. This is due to the $+\frac{b}{2}\bar{\mathbf{s}}'\boldsymbol{\Sigma}\bar{\mathbf{s}}$ term in Equation (1-35), since it is increasing in the number of assets and decreasing in the number of investors (provided that the total supply is fixed). When deviations appear, this assertion is likely to hold true, except in some particular cases.
- The higher the initial wealth, the higher the utility; this is of course obvious, but if we think of family ownership for example, it could explain why a family shareholder decides, on one day, to sell off its remaining stake: because its initial wealth has decreased (for dissemination reasons, and so on…) compared to the wealth of the first generation, so that the under-diversification disutility effect becomes so strong that the initial wealth cannot compensate for it anymore.

1.6.2.4 Comments on the effects of the deviations

In such a framework, the following conclusions hold:

- The price of asset *j* does not depend on the deviations in other assets: this is a direct consequence of the fixed and inelastic supply curve;
- Equation (1-31) tells us that the price of asset *j* depends on the global deviation in asset *j*, $\varepsilon_{j,t}$, which is the sum of the individual deviations $\varepsilon_{i,j}$ of each investor *i* in asset *j*: $\varepsilon_j = \sum_{i=1}^{N} \varepsilon_{i,j}$. Consequently, as long as the aggregate

deviation by asset ε_j remains constant, the individual deviations $\varepsilon_{i,j}$ of each investor *i* in each asset *j* can change without any impact on the equilibrium asset prices (but it needs to be compensated by an equivalent variation of the deviation $\varepsilon_{i,j}$ of another investor to keep ε_j constant of course;

- It follows that the only relevant fact here is whether the aggregate demand for asset *j* is constrained ($\varepsilon_j \neq 0$). The fact that a specific shareholder is more or less diversified ($\varepsilon_{i,j}$ close to 0 or not) does not matter in this framework. To state it in a simple way: two investors with null deviations or two investors with respective deviations -2 and +2 cause the same impact on the price of the risky asset, since the aggregate deviation (here, 0) is the same in both cases.
- The effect of a positive (resp. negative) average deviation $\bar{\varepsilon}_j$ in asset *j* is in fact exactly equivalent to a restriction (resp. increase) of the aggregate supply in asset *j* by the same amount $\bar{\varepsilon}_j$.
- Firms with positively constrained shareholders (i.e. such that their $\varepsilon_{i,j} > 0$) have a lower cost of capital than the others. The higher $\bar{\varepsilon}_j$, the lower the cost of capital.

We propose here to perform an in-depth analysis of Equation (1-31), which constitutes the most original result of our model. We especially want to tackle a specific issue: how an increase in the variance of a stock can induce an increase in its price? Economically speaking, this seems totally counter-intuitive and at odds with the traditional risk-return trade-off. But mathematically speaking, this is a direct consequence of Equation (1-31), since the coefficient of σ_j^2, $\frac{b\varepsilon_j}{NR_f}$, is necessarily positive if ε_j is positive. What is the economic intuition behind such a phenomenon?

Equation (1-31) teaches us that the systematic risk is not the only driver of the price, but must be complemented by an extra variance term. The higher the absolute value of the aggregate deviation, the higher is the impact of the specific risk (i.e. the stock variance) on the stock price. However, we need to be careful in the interpretation of this equation.

Firstly, it is difficult to assess a priori the economic significance of the impact of the deviation. Indeed, it is hard to evaluate the relative magnitudes of the covariance term $\sigma_{j,M} = \sum_{k=1}^{K} \sigma_{j,k} s_k$ and the variance term σ_j^2. In the extreme case where there is only one asset on the market, the ratio of the covariance $\sigma_{j,M}$ to the variance σ_j^2 is equal to the supply of asset j s_j. When there is more than one risky asset, the covariance between *j* and the market depends heavily on the signs of the covariances of asset *j* with the other assets.

Secondly, Equation (1-31) seems to suggest that an increase of the stock variance σ_j^2 results in an increase in the stock price p_j, provided that the deviation in this stock ε_j is positive. This is true if we consider that $\sigma_{j,M}$ remains constant when σ_j^2 increases. But we can observe from an explicit development of $\sigma_{j,M}$ that σ_j^2 is a component of $\sigma_{j,M}$, and that assuming $\sigma_{j,M}$ constant when σ_j^2 increases is equivalent to assuming a proportional compensation of the increase of σ_j^2 by the decrease of some covariance $\sigma_{j,k}$ of stock *j* with another stock *k*. In order to safely interpret Equation (1-31), we need to rewrite it and to isolate together all terms that depend somehow on σ_j^2:

$$p_j = \frac{\mu_j - \frac{b}{N}\left(\sum_{\substack{k=1 \\ k\neq j}}^{K} \sigma_{j,k} s_k + \sigma_j^2 \left(s_j - \varepsilon_j\right)\right)}{R_f} = \frac{\mu_j - \frac{b}{N}\left(\sigma_{j,\underline{M}} + \sigma_j^2 \left(s_j - \varepsilon_j\right)\right)}{R_f} \qquad (1\text{-}36)$$

where we define:

$$\tilde{p}_{\underline{M}} \equiv \tilde{p}_M - s_j \tilde{p}_j$$

$$\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_{\underline{M}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_j \\ \mu_{\underline{M}} \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_{j,\underline{M}} \\ \sigma_{j,\underline{M}} & \sigma_{\underline{M}}^2 \end{pmatrix}\right)$$

Equation (1-36) achieves this goal of isolating the terms in σ_j^2, but only if we consider the covariance $\sigma_{j,\underline{M}}$ as constant when stock variance σ_j^2 increases. If we rather consider the correlation $\rho_{j,\underline{M}}$ as constant, we must rewrite it as follows:

$$p_j = \frac{\mu_j - \frac{b\sigma_j}{N}\left(\sum_{\substack{k=1 \\ k\neq j}}^{K}\sigma_k \rho_{j,k} s_k + \sigma_j\left(s_j - \varepsilon_j\right)\right)}{R_f} = \frac{\mu_j - \frac{b\sigma_j}{N}\left(\sigma_{\underline{M}}\rho_{j,\underline{M}} + \sigma_j\left(s_j - \varepsilon_j\right)\right)}{R_f} \quad (1\text{-}37)$$

From Equation (1-37), we conclude that the stock price p_j can be positively related to the stock variance σ_j^2 when the aggregate deviation in the stock ε_j is larger than a certain threshold ε_j^{th}. This critical threshold ε_j^{th} is such that:

$$\frac{\partial p_j}{\partial \sigma_j} = 0 \Leftrightarrow -\frac{b}{NR_f}\left(\sigma_{\underline{M}}\rho_{j,\underline{M}} + 2\sigma_j\left(s_j - \varepsilon_j^{th}\right)\right) = 0$$

$$\Leftrightarrow \varepsilon_j^{th} = s_j + \frac{\sigma_{\underline{M}}}{2\sigma_j}\rho_{j,\underline{M}} \quad (1\text{-}38)$$

When the deviation ε_j is larger than this threshold ε_j^{th}, our theoretical model predicts that the higher the variance, the higher the price. At first view, this is totally counter-intuitive and in strong opposition with the classical trade-off risk-return. We will nonetheless show how this strange phenomenon can in fact receive an easy explanation.

To simplify the issue as much as possible, we will only consider one asset and two investors. In such a setting, the critical threshold ε^{th} is equal to s (since there is only one asset, we omit the index *j*). We will try to understand what happens when the global deviation increases beyond this threshold and why the relationship between price and volatility suddenly changes once this threshold is reached.

Let us assume that investor 1 is in deviation, while investor 2 is holding the optimal quantity of the asset. The utility of investor 2 is given by:

$$U_2 = R_f w_2 + x_2(\mu - R_f p) - \frac{b}{2}x_2^2\sigma^2$$

We are especially interested in the respective impacts of the volatility and the price on the utility of investor 2. In this purpose, we calculate the following partial derivatives:

$$\frac{\partial U_2}{\partial \sigma^2} = -\frac{b}{2} x_2^2$$
$$\frac{\partial U_2}{\partial p} = -R_f x_2$$

Notice that those derivatives consider only the direct effect of an increase of the volatility or the price on the utility, and suppose that the other variables are kept constant for the moment.

Those derivatives teach us that an increase of the volatility has always a negative direct effect on the utility, while an increase of the price has a negative (resp. positive) direct effect on the utility if x_2 is positive (resp. negative). It follows that as long as x_2 is positive, an increase of the stock variance (negative effect on the utility) needs to be compensated by a decrease of the asset price (positive effect on the utility) to keep the utility of investor 2 unchanged. However, if x_2 becomes negative, a increase of the stock variance σ^2 will still lead to a negative effect on the utility, while a decrease of the asset price would cause a negative effect as well on the utility, since x_2 is now negative. Therefore, in such a situation, an increase of the stock variance must necessarily induce an increase of the price to compensate investor 2 for the decrease of utility resulting from the increase of the risk component ($\frac{b}{2} x_2^2 \sigma^2$) of its utility. Usually, in normal conditions, x_2 is positive. But with the introduction of deviations, it can be optimal to hold negative quantities x_2 of the asset. In this case, since the optimal investor is short on the asset, he prefers to have a high price than a low price, since he must sell the asset today to buy it tomorrow. Therefore, the relation linking the stock variance to the price will change when the optimal holdings become negative. As long as the optimal holdings are positive, we have the classical trade-off risk-return: higher variance, lower price. As soon as the optimal holdings become negative, the sign of the relation changes: higher variance, higher price. This is always the trade-off risk-return, but for an investor who decides to go short on the stock.

If investor 1 is in deviation ε_1 and investor 2 is at the optimum, the actual equilibrium holdings of investor 2 are, by application of Equation (1-33):

$$x_2 = \frac{s}{2} + \frac{1}{\sigma^2}(0 - \frac{\sigma^2 \varepsilon_1}{2}) = \frac{s - \varepsilon_1}{2}$$

We can plug this new expression of x_2 in the utility equation of investor 2:

$$R_f w_2 + x_2(\mu - R_f p) - \frac{b}{2} x_2^2 \sigma^2 = R_f w_2 + \left(\frac{s-\varepsilon_1}{2}\right)(\mu - R_f p) - \frac{b}{2}\left(\frac{s-\varepsilon_1}{2}\right)^2 \sigma^2$$

It appears now clearly that, in a setting of two investors and one asset, reaching the critical threshold of the deviation is equivalent to investor 2 holding negative quantities:

$$\varepsilon^{th} = s \Leftrightarrow x_2^{th} = 0$$

When it happens, an increase of the stock variance results in an increase of the price because the sign of the coefficient of p in the utility equation ($-R_f\left(\frac{s-\varepsilon_1}{2}\right)$) changes and, as a consequence, demands an opposite response of the price to an increase of the variance.

In a setting with two investors and two assets, things are more complicated to analyze. The critical threshold is no longer $\varepsilon_j^{th} = s_j$, because an increase of the stock variance has more than the two effects mentioned above, on $\left(\frac{s-\varepsilon_1}{2}\right)(\mu - R_f p)$ and $-\frac{b}{2}\left(\frac{s-\varepsilon_1}{2}\right)^2 \sigma^2$. Covariance effects begin to appear, and complexity increases. If we suppose that investor 1 is in deviation on asset 1 but optimal on asset 2, and that investor 2 holds optimal holdings on both assets, the deviation $\varepsilon_{1,1}$ of investor 1 in asset 1 must be higher than the critical threshold of asset one (by application of Equation (1-38) to a universe with two investors and two assets):

$$\varepsilon_1^{th} = s_1 + \frac{\sigma_{\underline{M}}}{2\sigma_1}\rho_{1,\underline{M}} = s_1 + \frac{s_2\sigma_2}{2\sigma_1}\rho_{1,2}$$

The optimal quantities hold by investor 2 at equilibrium are given by:

$$\begin{pmatrix} x_{2,1} \\ x_{2,2} \end{pmatrix} = \begin{pmatrix} \frac{s_1}{2} \\ \frac{s_2}{2} \end{pmatrix} + \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{1,2}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{1,2} \\ -\sigma_{1,2} & \sigma_1^2 \end{pmatrix} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{\sigma_1^2 \varepsilon_{1,1}}{2} \\ 0 \end{pmatrix} \right)$$

$$= \frac{1}{2} \begin{pmatrix} s_1 - \frac{\sigma_1^2\sigma_2^2\varepsilon_{1,1}}{\sigma_1^2\sigma_2^2 - \sigma_{1,2}^2} \\ s_2 + \frac{\sigma_1^2\sigma_{1,2}\varepsilon_{1,1}}{\sigma_1^2\sigma_2^2 - \sigma_{1,2}^2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} s_1 - \frac{\varepsilon_{1,1}}{1-\rho_{1,2}^2} \\ s_2 + \frac{\sigma_1}{\sigma_2} \frac{\rho_{1,2}\varepsilon_{1,1}}{1-\rho_{1,2}^2} \end{pmatrix}$$

Notice this interesting fact: the holdings of investor 2 in asset 1 are independent of the variances of both stocks. An increase of variance does not modify holdings of asset 1 for both investors. Note also that $x_{2,1}$ will become negative as soon as $\varepsilon_{1,1} \geq (1-\rho_{1,2}^2)s_1$.

In the setting with one asset, the deviation needed to be at least as high as the supply ($\varepsilon_{1,1} \geq s_1$) in order to observe a negative $x_{2,1}$.

The utility expression becomes the following:

$$R_f w_2 + x_{2,1}(\mu_1 - R_f p_1) + x_{2,2}(\mu_2 - R_f p_2) - \frac{b}{2}\left(x_{2,1}^2\sigma_1^2 + 2x_{2,1}x_{2,2}\sigma_1\sigma_2\rho_{1,2} + x_{2,2}^2\sigma_2^2\right)$$

Suppose that the deviation is such that $\varepsilon_{1,1} \geq (1-\rho_{1,2}^2)s_1$, which implies that $x_{2,1}$ is negative. And suppose that the correlation coefficient $\rho_{1,2}$ is non-negative (which is usually the case). When the variance σ_1^2 of stock 1 increases, different effects are at work in this expression.

1) A positive effect through the $x_{2,2}(\mu_2 - R_f p_2)$ term. Indeed, a increase of σ_1^2 results in an increase of $x_{2,2}$ (since $x_{2,2} = \frac{1}{2}\left(s_2 + \frac{\sigma_1}{\sigma_2}\frac{\rho_{1,2}\varepsilon_{1,1}}{1-\rho_{1,2}^2}\right)$) and an increase of $(\mu_2 - R_f p_{2t})$ (since $(\mu_2 - R_f p_2) = \frac{b}{N}\left(s_1\sigma_1\sigma_2\rho_{1,2} + \sigma_2^2 s_2\right)$).
2) A negative effect through the $-\frac{b}{2}x_{2,1}^2\sigma_1^2$ term (it is the classical variance effect we have already met in the one asset setting).

3) A positive effect through the $-\frac{b}{2}2x_{2,1}x_{2,2}\sigma_1\sigma_2\rho_{1,2}$ term, since $x_{2,1,t}$ is negative, and $x_{2,2}\sigma_1$ is an increasing function of σ_1.

4) A negative effect through the $-\frac{b}{2}x_{2,2}^2\sigma_2^2$ term, since $x_{2,2}^2$ is increasing in σ_1.

There is a mix of positive and negative effects of an increase of σ_1 when the optimal investor is short on asset 1 (i.e. $x_{2,1}$). In fact, as long as $x_{2,1}$ is small in absolute value, the positive effect predominates. It follows that this positive effect will be compensated by a negative effect through the term $+x_{2,1}(\mu_1 - R_f p_1)$, and this negative effect can only be achieved by a decrease of the price p_1, since $x_{2,1}$ is negative. In this case, a higher volatility σ_1 results in a lower price p_1, even if investor 2 is short on asset 1.

However, as the deviation of investor 1 increases beyond a certain threshold $\varepsilon_1^{th} = s_1 + \frac{s_2\sigma_2}{2\sigma_1}\rho_{1,2}$ as computed above, the negative effect on this utility of investor 2 of this volatility increase becomes larger than the positive effect. Therefore, the appropriate response is now an increase of the price of asset 1. A higher volatility induces here a higher price.

To sum up, the relation variance – price depends on the magnitude of the deviation.

If $\varepsilon_{1,1} \le (1-\rho_{1,2}^2)s_1$, the optimal holdings of investor 2 are positive and the classical trade-off risk-return prevails: a higher volatility σ_1 will be compensated by a lower price p_1, because the utility of investor 2 is negatively affected by a higher volatility but positively affected by a lower price.

If $(1-\rho_{1,2}^2)s_1 \le \varepsilon_{1,1} \le \varepsilon_j^{th} = s_j + \frac{\sigma_{\underline{M}}}{2\sigma_j}\rho_{j,\underline{M}}$, the optimal holdings $x_{2,1}$ of investor 2 in asset 1 are now negative (apparition of short sales), but since an increase of volatility σ_1 has still more good effects than bad effects on its utility, the adequate price response is still a decrease of price p_1. The difference with the previous case, is that the increase of volatility has good effects on investor 2's utility while the decrease of price has bad

effects (because now investor 2 wants to go short on the asset and is not pleased with a price decrease).

Finally, if $\varepsilon_j^{th} = s_j + \frac{\sigma_{\underline{M}}}{2\sigma_j}\rho_{j,\underline{M}} \leq \varepsilon_{1,1}$, the optimal holdings $x_{2,1}$ of investor 2 in asset 1 are always negative as before, but the increase of volatility σ_1 has more adverse effects on its utility than good ones. Therefore, the compensation changes, and becomes a price increase instead of a price decrease. Indeed, a price increase is something positive for investor 2, since he wants to be short on the asset, so he expects the highest price possible.

Two final remarks are to be done. It is very difficult to say whether the deviation can actually become so high that it gets larger than the critical threshold: $\varepsilon_j^{th} = s_j + \frac{\sigma_{\underline{M}}}{2\sigma_j}\rho_{j,\underline{M}} \leq \varepsilon_{1,1}$. Is it purely theoretical, or may it happen in the reality? If the correlation with the market is positive, the presence of short sales is a necessary but not sufficient condition to the inversion of the classical risk-return trade-off. Is it possible to observe such high levels of deviation on the market, so large that they would invert the traditional negative relation between price and volatility? A priori, this is quite unlikely, at least under normal market conditions. But we may imagine that this kind of paradoxical phenomena could arise in very specific and temporary market circumstances. By the way, even if such critical threshold is never reached, the deviation can still have a potentially significant impact on prices, and on the relation between the variance and the price. Our pricing Equation (1-37) at least gives us two lessons:

1) A higher deviation induces a higher price, ceteris paribus;
2) A higher deviation makes the relation between variance and price less and less negative. In particular theoretical situations, this relation can potentially become positive.

Eventually, it is important to see that short-selling constraints do not change our results, but rather reinforce them. In order to observe our particular effects of inversion of the volatility-price relation, we do not need that an investor actually holds negative quantities of an asset, but that the optimal behaviour is to hold negative quantities, even

if it is impossible to actually do it on the market. The optimal investor 2 of our reasoning above can be purely virtual. He does not need to be real. But effects on prices will be real.

1.6.3 Extension with a more realistic supply curve

Up to here we have assumed that the aggregate supply of each asset *j* was fixed and independent of the price. Even if most of literature assumes fixed asset supply (see Fama (1976), Miller (1977), Li and Fleisher (2004) and Greenwood (2005) to mention only a few of them), this is most probably too restrictive, as Braun and Larrain (2008) show that supply shocks can affect stock prices. We propose here a simple extension of our model with a linear supply curve, without saying anything about the determinants underlying this supply curve:

$$\mathbf{s} = \mathbf{c} + d\mathbf{p} \tag{1-39}$$

In this more general framework, Equation (1-30) becomes:

$$\mathbf{X}^{C'}\mathbf{1}_N = \mathbf{c} + d\mathbf{p} \Leftrightarrow \boldsymbol{\Sigma}^{-1}\frac{\boldsymbol{\mu} - R_f\mathbf{p}}{b}\mathbf{1}_N'\mathbf{1}_N + \boldsymbol{\Sigma}^{-1}\mathbf{E}_t'\mathbf{1}_N = \mathbf{c} + d\mathbf{p}$$

$$\Leftrightarrow \boldsymbol{\Sigma}^{-1}\frac{\boldsymbol{\mu} - R_f\mathbf{p}}{b}N + \boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi} = \mathbf{c} + d\mathbf{p} \Leftrightarrow \frac{N}{b}R_f\mathbf{p} + d\boldsymbol{\Sigma}\mathbf{p} = \frac{\boldsymbol{\mu}}{b}N - \boldsymbol{\Sigma}\mathbf{c} + \boldsymbol{\varphi}$$

$$\Leftrightarrow \left[R_f\mathbf{I}_K + \frac{bd\boldsymbol{\Sigma}}{N}\right]\mathbf{p} = \boldsymbol{\mu} - b\left(\frac{\boldsymbol{\Sigma}\mathbf{c}}{N} - \overline{\boldsymbol{\varphi}}\right)$$

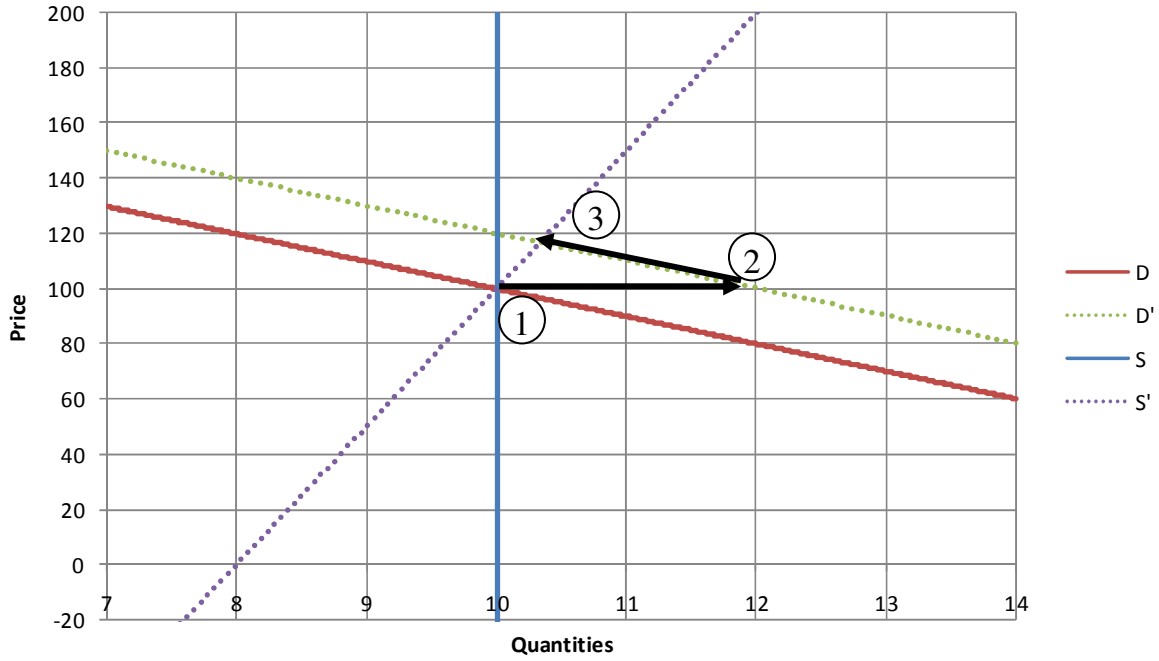
$$\Leftrightarrow \mathbf{p} = \left[R_f\mathbf{I}_K + \frac{bd\boldsymbol{\Sigma}}{N}\right]^{-1}\left[\boldsymbol{\mu} - b\left(\frac{\boldsymbol{\Sigma}\mathbf{c}}{N} - \overline{\boldsymbol{\varphi}}\right)\right] \tag{1-40}$$

Notice that for $d = 0$ (case of the inelastic supply), (1-40) reduces to (1-30).

Note also that in this case, the price of asset *j* is no more independent of the deviations in the other assets.

When the supply is not totally inelastic (i.e. strictly positive), an aggregate positive deviation has still a positive impact on the price, but less than with an inelastic supply. Indeed, the impact of the deviation on the price is partially absorbed by an increase of the quantity supplied. This is schematized by Figure 1-2 on a market with one risky asset.

Figure 1-2: Impact of the deviation on the equilibrium with a partially elastic supply



Therefore, there is a double positive effect on the market capitalization: a positive effect through the increase of the equilibrium price, and a positive one through the increase of the equilibrium exchanged quantities.

1.6.4 Addition of a short-selling constraint[2]

Up to now we have assumed that investors could buy and sell assets without any restrictions, and hold positive as well as negative quantities of assets. The reality is likely to be slightly different. If holding positive quantities of stocks on a liquid market is rarely a problem, even if it does not come without some transaction costs, the detention of negative quantities of stocks necessarily implies some kind of short-selling. Although the market of derivatives may offer an alternative way to achieve the same goal as the pure short-selling operation, we must recognize that short-selling remains less obvious and evident to realize than a classical long position in a stock. Literature has explored different ways to introduce such short-selling constraints, in particular in a universe of heterogeneous expectations, and most of authors have concluded that short-selling constraints tend to drive prices up, since they keep "pessimistic" investors out of

[2] We thank Michel Levasseur for having highlighted this point.

the market (see Miller (1977) and Harrison and Kreps (1978) for example). However, using a rational expectations model, Diamond and Verrecchia (1987) show that constraining short sales reduces the informational efficiency of prices but does bias them upward. And Bai, Chang and Wang (2006) further demonstrate that under rational expectations, short-sale constraints can in fact cause asset prices to decrease. Indeed, since short-selling constraints make pessimistic investors unable to trade, prices are less informative and uncertainty increases. If agents are risk averse, prices are to decrease to compensate this new volatility. Since our model (like Miller (1977)) considers the risk (basically **Σ**) as exogenous, it cannot take such informational effect into account, so that the discussion that we propose hereafter about the impact of a short-selling constraint does not pretend to be complete and thorough, but only to give some insight on this issue in our particular framework.

In the context of our model, the most natural and straightforward way to proceed is to introduce an additional constraint in the maximization problem that imposes the non-negativity of stock quantities hold by the investors: $\mathbf{X}^C \geq 0$.

The original maximization problem now becomes a maximization problem under inequality constraints:

$$\text{Max} \quad R_f w_i + \mathbf{x}_i^{C'}\left(\boldsymbol{\mu} - R_f \mathbf{p}\right) - \frac{b}{2} diag\left(\mathbf{x}_i^{C'}\left(\boldsymbol{\Sigma} - \mathbf{I}_k \bullet \left(\boldsymbol{\varphi}_i \mathbf{1}'_N\right)\right)\mathbf{x}_i^C\right) \tag{1-41}$$

$$\text{w.r.t.} \quad \mathbf{x}_i^C$$

$$\text{s.t.} \quad \mathbf{x}_i^C \geq 0$$

Since the objective function is quasi-concave and the constraints are linear, we can apply the Kuhn-Tucker conditions (they are necessary and sufficient to find the optimum):

$$\boldsymbol{\mu} - R_f \mathbf{p} - b\boldsymbol{\Sigma}\mathbf{x}_i^C = -b\boldsymbol{\varphi}_i - \boldsymbol{\lambda}_i \tag{1-42}$$

$$\boldsymbol{\lambda}_i \bullet \mathbf{x}_i^C = 0$$

$$\mathbf{x}_i^C \geq 0$$

$$\boldsymbol{\lambda}_i \geq 0$$

with $\boldsymbol{\lambda}_i$ the [K×1] vector of the Lagrange's multipliers $\lambda_{i,j}$.

Let us define the following matrices:

- $\boldsymbol{\Lambda}$ the [N×K] matrix of the Lagrange's multipliers $\lambda_{i,j}$;

- $\boldsymbol{\lambda} = \boldsymbol{\Lambda}'\mathbf{1}_N$;

- $\bar{\boldsymbol{\lambda}} = \dfrac{\boldsymbol{\Lambda}'\mathbf{1}_N}{N}$;

The new vector of equilibrium prices is then given by:

$$\mathbf{p} = \frac{1}{R_f}\left[\boldsymbol{\mu} - \frac{b}{N}(\boldsymbol{\Sigma}\mathbf{s} - (\boldsymbol{\varphi} + \frac{\boldsymbol{\lambda}}{b})\right] = \frac{1}{R_f}\left[\boldsymbol{\mu} - b(\boldsymbol{\Sigma}\bar{\mathbf{s}} - (\bar{\boldsymbol{\varphi}} + \frac{\bar{\boldsymbol{\lambda}}}{b})\right] \tag{1-43}$$

Since $\boldsymbol{\lambda}_i \geq 0$, the effect of adding a new constraint on the short-selling operations actually works as a mere positive deviation effect that seems to reinforce positive and mitigate negative deviations $\varepsilon_{i,j}$.

We can derive the equilibrium holdings of the individuals:

$$\mathbf{X}^C = \mathbf{1}_N \frac{\boldsymbol{\mu}' - R_f\left\{\frac{1}{R_f}\left[\boldsymbol{\mu} - b(\boldsymbol{\Sigma}\bar{\mathbf{s}} - (\bar{\boldsymbol{\varphi}} + \frac{\bar{\boldsymbol{\lambda}}}{b})\right]\right\}'}{b}\boldsymbol{\Sigma}^{-1} + \left(\mathbf{E} + \frac{\boldsymbol{\Lambda}}{b}\right)\boldsymbol{\Sigma}^{-1}$$

$$= \mathbf{1}_N\left[\bar{\mathbf{s}}'\boldsymbol{\Sigma} - (\bar{\boldsymbol{\varphi}}' + \frac{\bar{\boldsymbol{\lambda}}'}{b})\right]\boldsymbol{\Sigma}^{-1} + \left(\mathbf{E} + \frac{\boldsymbol{\Lambda}}{b}\right)$$

$$= \mathbf{1}_N\bar{\mathbf{s}}' + \left(\mathbf{E} + \frac{\boldsymbol{\Lambda}}{b} - \frac{\mathbf{1}_N\mathbf{1}_N'}{N}\left(\mathbf{E} + \frac{\boldsymbol{\Lambda}}{b}\right)\right)\boldsymbol{\Sigma}^{-1} = \mathbf{1}_N\bar{\mathbf{s}}' + \left(\mathbf{I}_N - \frac{\mathbf{1}_N\mathbf{1}_N'}{N}\right)\left(\mathbf{E} + \frac{\boldsymbol{\Lambda}}{b}\right)\boldsymbol{\Sigma}^{-1} \geq 0$$

$$\Leftrightarrow \mathbf{x}_i^C = \bar{\mathbf{s}} + \boldsymbol{\Sigma}^{-1}\left[(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}}) + \left(\frac{\boldsymbol{\lambda}_i - \bar{\boldsymbol{\lambda}}}{b}\right)\right] \geq 0 \tag{1-44}$$

And the Lagrange's multipliers:

$$\boldsymbol{\lambda}_i = \bar{\boldsymbol{\lambda}} - b\left[\boldsymbol{\Sigma}\left(\bar{\mathbf{s}} - \mathbf{x}_i^C\right) + \left(\boldsymbol{\varphi}_i - \bar{\boldsymbol{\varphi}}\right)\right] \geq 0 \tag{1-45}$$

It should be remembered that $\lambda_{i,j} > 0$ only if the constraint is binding, that is, if $x_{i,j} = 0$. This is the case when it could be optimal for investor *i* to short-sell asset *j*, but because of the constraint, the investor is bound to hold a non-negative number of shares of asset *j*. In such a situation, $\lambda_{i,j} > 0$, and the price p_j is higher than it would be if short-selling was not restricted.

According to Equation (1-45), the effect of short-selling more likely appears when market risk aversion is low (*b*), stock variance and covariance are low, and dispersion in the matrix of deviations is high. Strong heterogeneity of beliefs is likely to lead to higher incentive to short-sell, and thus higher asset prices if short-selling is restricted in some way.

Equation (1-45) also reveals that short-selling of investor *i* in asset *j* basically happens for two reasons: either the own deviation of investor *i* in asset *j* ($\varepsilon_{i,j}$) is sufficiently low and drives $x_{i,j}$ so down that the constraint becomes binding, or the mean deviation of all investors ($\bar{\varepsilon}_j$) on this asset is high enough to drive the asset price up and discourage investor *i* to go long on this asset.

Finally, remember that the effects predicted by Equation (1-45) strongly depend on our initial hypotheses, especially that the variance-covariance matrix $\mathbf{\Sigma}$ is considered to be exogenous. As mentioned in the literature review, introducing a risk component associated to the short-selling constraint could lead to very different results.

1.6.5 Does under-diversification really matter?

1.6.5.1 General discussion

Equation (1-30) suggests that the impact of under-diversification on the price of a given asset *j* is somehow proportional to $\frac{b\sigma_j^2}{NR_f}$: the higher *N*, the smaller the impact. If we take the limit to infinity for *N*, is there any remaining effect? The answer depends on how the global deviation term moves with *N*. To see this, we first need to rearrange Equation (1-31) under another form:

$$p_j = \frac{\mu_j - \frac{b}{N}\left(\sum_{k=1}^{K}\sigma_{j,k}s_k - \sigma_j^2\varepsilon_j\right)}{R_f} = \frac{1}{R_f}\left[\mu_j - \frac{b}{N}\left(\frac{\sum_{k=1}^{K}\sigma_{j,k}s_k}{\sum_{k=1}^{K}s_k}\frac{\sum_{k=1}^{K}s_k}{K}K - \sigma_j^2\varepsilon_j\right)\right]$$

$$= \frac{1}{R_f}\left[\mu_j - \frac{b}{N}\left(\bar{\sigma}_{j,.}\bar{s}K - \sigma_j^2\varepsilon_j\right)\right] \quad (1\text{-}46)$$

with $\bar{s} \equiv \frac{\sum_{k=1}^{K}s_k}{K}$ and $\bar{\sigma}_{j,.} \equiv \frac{\sum_{k=1}^{K}\sigma_{j,k}s_k}{\sum_{k=1}^{K}s_k}$

We are now able to take the limit to infinity **for both *N* and *K***. Taking only the limit for *N* or *K* would lead to trivial cases. The limit for *N* would give a cost of capital equal to R_f, while the limit to *K* would give zero prices and an infinite cost of capital.

$$\lim_{N,K\to+\infty} p_j = \lim_{N,K\to+\infty}\frac{1}{R_f}\left[\mu_j - \frac{b}{N}\left(\bar{\sigma}_{j,.}\bar{s}K - \sigma_j^2\varepsilon_j\right)\right]$$

$$= \frac{\mu_j - b\bar{\sigma}_{j,.}\bar{s} + b\sigma_j^2\lim_{N,K\to+\infty}\frac{\varepsilon_j}{N}}{R_f} \quad (1\text{-}47)$$

If we forget for a while about the last term, this result is the well-known implication of the CAPM that idiosyncratic risk is fully diversifiable and thus not priced while systematic risk is not fully diversifiable and thus is priced. In a context of imperfectly diversified investors, the last term is non-null and thus deserves some consideration. It can be negligible if *N* increases while the aggregate deviation remains constant. On the other hand, if the aggregate deviation increases at the same pace as *N*, under-diversification matters and is likely to have a real impact on prices.

1.6.5.2 Relevancy of some empirical settings

There are basically at least two ways to test our theoretical conclusions.

A first possibility would be to demonstrate that the idiosyncratic risk is priced in the cross-sectional dimension of stock returns. Several papers have developed asset pricing models that take into account the investors' lack of diversification. Levy (1978), Merton (1987), and Malkiel and Xu (2001) build extensions of the CAPM where the investors,

for some exogenous reason, hold undiversified portfolios. However, their modeling is different from ours in some respects. For example, Malkiel and Xu (2001) consider two classes of investors: the constrained and the unconstrained ones, and they define the constrained ones as investors unable to hold some stocks, i.e. such that their holdings in those stocks are fixed to 0. In this respect, our model is more general since it allows for all possible deviations. In those models, the resulting pricing equation is very close "in spirit" to ours and relates the returns of stocks to their beta with the market and their beta with respect to a market-wide measure of idiosyncratic risk. With or without this theoretical ground, many authors have tried to test econometrically whether idiosyncratic risk matters in the pricing of cross-sectional stock returns. The study of Douglas (1969) was among the earliest to find some link between idiosyncratic variance and stock return. Then came Fama and MacBeth (1973) who rejected the idea that the idiosyncratic variance could impact stock returns by means of a more powerful cross-sectional test. Since then, studies have transposed this topic in other fields of finance (see Lehmann (1990), Bessembinder (1992), King, Sentana and Wadhwani (1994), Falkenstein (1996) and Green and Rydqvist (1997)), and found significant relation between the idiosyncratic risk and bond prices, the equity holdings of mutual funds manager, foreign currency and agricultural futures. But recently, two studies on the topic directly tested the link of idiosyncratic risk on stock returns: Goyal and Santa-Clara (2003) in the time-series dimension and Malkiel and Xu (2004) in the cross-sectional dimension. Both conclude to the significance of the idiosyncratic risk factor in their regressions. However, this point remains controversial, as a recent article of Wei and Zhang (2005) rejects the results of Goyal and Santa-Clara (2003).

A second possibility to test our model would be to check whether the deviation is priced somehow by the market. It is possible to imagine different empirical settings to test this conclusion of our model. In our empirical Section 2, we chose to focus on the "index effect", which refers to the value premium associated to members of some indices. Indeed, many studies have proved that companies belonging to a widely traded index (typically, the S&P 500) present a higher value than the others, all other things being equal. The literature has proposed different explanations for this index effect. Some of them are in line with our "deviation" story, while others propose very different perspectives. We propose to give in Section 2 a general survey of all hypotheses appearing in the literature about the index effect, to discuss how they can partly be linked to our theoretical model of Section 1 and to test them through various

econometric specifications able to control for statistical issues like endogeneity and self-selection.

2 Empirical estimation: S&P 500 membership and the cost of capital

2.1 Introduction

Since the seminal articles of Goetzmann and Garry (1986), Harris and Gurel (1986), Arnott and Vincent (1986), Shleifer (1986) and Woolridge and Ghosh (1986), we know that a stock newly included in (resp. deleted from) the S&P 500 index experiences a substantial price increase (resp. decrease). This "index effect" has been widely confirmed by a vast amount of literature since then, like Jain (1987) and Lamoureux and Wansley (1987) for the oldest and Chen, Noronha and Singal (2004) and Hrazdil and Scott (2007) for the most recent (among many others). According to Chen et al. (2004), not less than 905 additions and 905 deletions have been observed from July 1967 to December 2000, and they lead on average to significant abnormal returns during the 1976 to 2000 period: additions exhibit an average announcement day abnormal return of about 3% between 1976 and 1989 and 5% between 1989 and 2000, while deletions show -8% from 1989 up to 2000. About sixty days after the effective change[3] (i.e. on the window [announcement date; effective date + 60]), the cumulative abnormal return subsequent to an addition still reaches 4% (resp. 6%) over the 1976-1989 (resp. 1989-2000) period, while it becomes insignificant for a deletion. Exact figures may vary from one author to another, but they all conclude in favor of a statistically and significant index effect.

The current issue is therefore not whether such an index effect exists, but what are its origin and causes. In this respect, literature shows no unanimity anymore. Five hypotheses have been proposed by various authors to explain the index effect: the Price Pressure Hypothesis (Harris and Gurel (1986)), the Downward Sloping Demand Curve Hypothesis (Shleifer (1986)), the Liquidity Cost Hypothesis (Mikkelson and Partch (1985)), the Information Content Hypothesis (Jain (1987)) and the Investor Recognition Hypothesis (Merton (1987)). They all predict a significant price increase in case of an addition, but they differ under other aspects: whether the impact is transient or

[3] Before 1989, announcement and effective dates of change were roughly the same. The announcement occurred at the close of a day and the effective change took place just before the market opening of the next day. From 1989, in order to alleviate price pressures, Standard and Poor's modified its policy and pre-announced changes before they effectively occurred. The period between announcement and effective dates varies from one day to about one month.

permanent, whether it is symmetric for deletions, whether the trading volume is affected, and in which way, etc. Many studies have tested those hypotheses and have come with very interesting results. In particular, one should mention here the article of Kaul, Mehrotra and Morck (2000) that manages to isolate the pure "Downard Sloping Demand Curve Hypothesis" by focusing on a pure information-free event. They examine how the merely technical rearrangement of weights in the Toronto Stock Exchange 300 in 1996 caused significant excess returns for the stocks having seen their weights increase. However, empirical evidence in the literature remains partial, because articles often restrict themselves to one specific hypothesis. On the top of that, results are sometimes in contradiction with each other: for instance, some studies provide evidence of temporary effects, while others tend to dismiss such an explanation, as the next section will show it. Finally, to our knowledge, no article has explicitly considered yet the issue of endogeneity in the study of the index effect. In this respect, econometric estimations like the introduction of instrumental variables, the exploitation of the panel structure of the data or the treatment of a suspected self-selection effect could help to address at least partly this issue. As a consequence, there is some room for an additional contribution to the topic.

Our contribution is twofold.

First, we provide a setting global enough to test four out of the five hypotheses proposed by the literature. Specifically, we begin with a close replication of the study of Morck and Yang (2002) who focus on the fact to be in the S&P 500 rather than on the inclusion event itself. They use cross-sectional regressions instead of the classical event study of most of previous authors, and Tobin's q ratio instead of cumulative abnormal returns as a measure of performance. This approach is not intrinsically better than the previous ones, but may give another valuable perspective on the topic, since it focuses on the fact to be in the index rather than on the inclusion event itself, and allows for different kinds of estimation, like a panel study estimation and a self-selection treatment approach. It also offers the opportunity to use alternative dependent variables (the Price-Earning Ratio, the Market-to-Book ratio and the Tobin's q) in order to assess the robustness of the results, while event studies generally focus on the behaviour of the cumulative abnormal returns only. Morck and Yang (2002) bring support for the existence of a significant and positive S&P500 membership premium. However, they essentially focus on their hypothesis that the membership premium is associated to the

growth of index funds through time, and they do not test explicitly for the other hypotheses suggested by previous authors. Therefore we propose ourselves to enrich their initial regression with additional variables able to capture the predictions of the various hypotheses put forward by the literature. We especially test the Price Pressure Hypothesis, the Downward Sloping Demand Curve Hypothesis, the Certification Effect Hypothesis (a particular form of the Information Content Hypothesis) and the Investor Recognition Hypothesis.

Second, we propose three different econometric specifications (in line with Campa and Kedia (2002) who did the same in a different context) in order to treat the endogeneity issue linked to the S&P 500 membership dummy: a fixed-effect panel study, an instrumental variables regression and a self-selection estimation. Indeed, introducing the fact to be member of the S&P 500 membership index as an explicative variable for the performance of the firm (proxied by its Tobin's q), as Morck and Yang (2002) do, potentially raises serious endogeneity concerns. Intuitively, we can suspect these two variables to be related to the same inobservable(s). Some determinants of the performance may well be the same as some determinants of the fact to be a member of the S&P 500 index. In this case, the exogeneity of the S&P 500 membership dummy cannot be ensured anymore and the coefficients of a classical OLS regression are no longer consistent. As a consequence, it calls for more complex econometric specifications like the three ones mentioned above in order to tackle this endogeneity problem.

Our results tend to show the simultaneous existence of both a temporary (short-term) positive effect and a permanent (long-term) positive effect of the inclusion in the S&P 500 index on the firm value. Being a member tends to increase the firm's value, but being a new member (less than one year) increases it even more. The existence of such a transient effect gives some support to the Price Pressure Hypothesis, but the existence of a permanent effect calls for other explanations as well. We show the existence of an asymmetric effect on the short term: performance increases temporarily after an inclusion, but does not decrease after a deletion. This would give some support to a short-term version of the investor Recognition Hypothesis. However, in the long-run, we do not find significant support for the Investor Recognition Hypothesis: awareness of investors seems to fade away as time goes on. Quite surprisingly, we do not find much support in favor of the Downard-Sloping Demand Curve Hypothesis: the weight

of a company in the S&P 500 does not seem to positively influence the firm value, on the contrary, we rather detect a negative relation. Finally, thanks to the self-selection estimation, we show that the S&P 500 membership premium still subsists after the introduction of the Mill's ratio in the regression equation. What is more, the negative sign of the coefficient of this latter variable leads us to reject the Certification Effect Hypothesis. Firms members of the index do not present better fundamentals than others, but benefit from a S&P 500 membership value premium.

The rest of Section 2 is organized as follows. In Section 2.2, we briefly discuss related literature and present the competing hypotheses. Section 2.3 describes the data, sample selection criteria, and preliminary analysis. Section 2.4 discusses the estimation methodology. Section 2.5 presents the results. Section 2.6 concludes.

2.2 Theoretical background[4]

2.2.1 The « index effect »

The index effect is not restricted to the S&P 500 index. Evidence of price pressures has been found in supplementary indices such as S&P 400 and S&P 600 (Jain (1987) and Blitzer and Dash (2004)), the German market for the DAX blue-chip and the DAX mid-cap (Deininger, Kaserer and Roos, 2000), the New-Zealand market (Elayan, Li and Pinfold, 2000), the Canadian market for the TSE 300 index (Masse, Hanrahan, Kushner and Martinello (2000)), the French market for indices CAC 40, SBF 120 and FTSE 100 (Vespro (2006)) and MSCI country indices (Chakrabarti, Huang, Jayaraman and Lee (2002)).

However, this article focuses on the S&P 500 for mainly three reasons.

First, many authors have delivered results on it, so that the index effect on the S&P 500 is very well documented. From Harris and Gurel (1986) to Hradzil and Scott (2007), all studies agree to confirm the existence of a significant index effect on the S&P 500.

Second, some hypotheses proposed by the literature are effective only if the index has a minimum level of popularity and trading. In particular, the hypotheses justifying the

[4] This literature review capitalizes on many different sources. In particular, the literature reviews of Brooks, Kappou and Ward (2004) and Hradzil and Scott (2007) were found especially comprehensive and valuable. The division into "five hypotheses" is borrowed from them. The interested reader is referred to them for any additional information on the vast topic of the index effect.

index effect by the rebalancing of index funds (namely, the Price Pressure Hypothesis and the Downward Sloping Curve Hypothesis) can make sense only if the index is sufficiently tracked by the index funds. This condition is likely to be met by the S&P 500 index, as Woolridge and Ghosh (1986) tell us that "since the return on the S&P 500 index is the most common yardstick for gauging investment performance for money managers, it is also the most common portfolio for indexation purposes". And more recently, according to Morck and Yang (2002), "Goetzmann and Massa (2003) estimate that $80 to $100 billion was formally indexed to the S&P 500 in mid-1998. This works out to about ten percent of the market capitalization of the S&P 500". This is probably underestimated, "because besides the numerous mutual funds indexed to S&P 500, a huge amount of money is informally indexed to the S&P 500 by corporate and public sector pension funds. In addition, many actively managed funds use the S&P 500 as a benchmark. This creates an incentive for their managers to invest money in the S&P index and then deviate from that strategy when they feel they have private information. The result is another tier of less formally indexed investment. These considerations make a precise measurement of the total value of assets indexed to S&P 500 a virtual impossibility", but give some insight of the potentially substantial amount of money tracking the S&P 500. This is not necessarily the case for other indices, since Brooks, Kappou and Ward (2004) write that "other indices that might appear to be quite sensible and reasonable benchmarks in the US market" do not seem to present any significant index effect "because there is much less indexed money tracking them". For example, Beneish and Gardner (1995) do not find significant price or volume effects for the stocks added to the Dow Jones Industrial Average (DJIA), although stocks removed from this index exhibit significant price declines.

Third, we are given an indirect proof of a significant and strong index effect on the S&P 500 by the Standard and Poor's Index Committee itself. In 1989, the latter revised its announcement policy and decided to pre-announce changes an average of five days before they become effective in order to alleviate price pressures during the event. Interestingly, the index effect did not vanish with the new announcement policy of Standard and Poor's. Beneish and Whaley (1996 and 1997), and other authors after them like Lynch and Mendenhall (1997), Wrugler and Zhuravskaya (2002) and Chen, Noronha and Singal (2004), found that the price increase after inclusion was permanent and, surprisingly, that under the new policy, the average price increase was even greater. It is worth noticing that this separation between the announcement day and the

effective date of the composition change allows for testing some of the implications the Efficient Market Hypothesis (Fama (1970, 1991))[5].

Literature has come with various hypotheses to explain the index effect: the Price Pressure hypothesis, the Imperfect Substitutes / Downward-Sloping Demand Curve for Stocks Hypothesis, the Liquidity Cost Hypothesis, the Information Content / Index Member Certification Hypothesis, and the Market Segmentation / Investor Recognition Hypothesis. They mainly differ in the permanent or temporary pattern of the stock price or volume change and the kind of information revealed by the inclusion or the exclusion. This will be the purpose of the next section to describe them.

While those hypotheses are mainly concerned with the effect of index inclusions on price levels, there is also an active stream of the literature showing that the additions of stocks to an index may also affect patterns of comovement. For example, Vijh (1994) finds that at both daily and weekly frequencies, stocks added to the S&P 500 between 1975 and 1989 experience a significant increase in their betas on the value-weighted return of NYSE and AMEX stocks, a close proxy for the value weighted S&P return. If we believe in Standard & Poor's claim that the decision to add a new stock to the index does not signal any opinion on the stock's fundamental value, this result is in contrast to the fundamentals view, which states that any comovement in prices must be due to comovement in fundamentals. However, it would be consistent with a friction- or sentiment-based view, which predicts that an addition to the index leads to a shift in the correlation structure of returns. Indeed, in economies with frictions or with irrational investors, and in which there are limits to arbitrage, comovement in prices is delinked from comovement in fundamentals. Therefore, an addition to the index may affect the beta of the stock without impacting the fundamentals. Barberis, Shleifer and Wurgler (2005) confirm the conclusions of Vijh (1994) and provide new evidence in support of the friction- or sentiment-based theories of comovement. First, by applying a univariate regression analysis similar to Vijh's to the longer sample of data now available, they uncover considerably larger effects. According to them, "while stocks added to the S&P during 1976–1987 experience an average increase in daily S&P beta of 0.067 after inclusion, the average increase in the 1988–2000 period is 0.214. In light of the growing

[5] Indeed, if the inclusion itself does not bring any more information than its previous announcement, and if the whole information of the announcement is quickly reflected in the stock prices, according to the EMH hypothesis, one should not observe any significant price upraise between the announcement and

importance of the S&P 500 as both a category and a habitat, the fact that the effect is not only present but larger in more recent data is especially supportive of the friction- or sentiment-based views" (Barberis and al. (2005), p. 286).

2.2.2 Competing hypotheses[6]

Chen, Noronha and Singal (2004) basically distinguish two theories aiming at explaining stock price changes.

The first one supposes to turn one's back to the long-held assumption that stocks have perfect substitutes and to the perfectly elastic demand that follows from it. Once **imperfectly elastic demand** is accepted, the price is merely determined through the application of the market clearing condition and is situated at the unique intersection of the aggregate demand and supply curves. Two perspectives can then emerge. Either stocks have a **short-term** downward sloping demand curve and any excess demand simply implies a **temporary** price increase. This is the Price Pressure Hypothesis of Harris and Gurel (1986), Wooldridge and Ghosh (1986), Lamoureux and Wansley (1987), Lynch and Mendenhall (1997), Blouin, Raedy and Shackelford (2000) and Malkiel and Radisish (2001). Or stocks have a **long-term** downward sloping demand curve so that any excess demand should cause a **permanent** price increase. Main works in favor of this Downward-Sloping Demand Curve Hypothesis are Shleifer (1986), Beneish and Whaley (1996), Lynch and Mendenhall (1997), Kaul, Mehrotra and Morck (2000), Morck and Yang (2002), Blume and Edelen (2001), Wurgler and Zhuravskaya (2002) and Petajisto (2006).

On the contrary, the second theory persists to assume a **perfectly elastic** stock demand curve and sticks to the fundamental view that the price of a stock is basically the sum of discounted expected future cash flows (i.e. the Fisher equation), with the required rate of return k_s as the actualization rate:

$$p_t = \sum_{s=1}^{T} \frac{E_t[CF_s]}{(1+k_s)^s}$$

---

inclusion days. This is not the purpose of this article to investigate this issue, but the interested reader will find references and a more comprehensive presentation on this topic in Brook et al. (2004).

[6] Table 2-1 provides the reader with a complete summary of the implicit theoretical assumptions and predictions of all competing hypotheses. This table summarizes information from Sections 2.2.2 and 2.2.3.

In such a perspective, observed excess return can only happen if the observed change is stock price is due to a change in expected future cash flows or a change in the discount rate applicable to that stock. The Liquidity Cost Hypothesis (Mikkelson and Partch (1985), Amihud and Mendelson (1986), Edminster, Graham and Pirie (1996), Erwin and Miller (1998), Chordia (2001) and Hegde and McDermott (2003)), the Information Content Hypothesis (Goetzmann and Garry (1986), Jain (1987), Dhillon and Johnson (1991) and Denis, McConnell, Ovtchinnikov and Yu (2003)) and the Investor Recognition Hypothesis (Merton (1987) and Chen, Noronha and Singal (2004)) are consistent with this theoretical positioning.

2.2.2.1 Imperfect elasticity based hypotheses

2.2.2.1.1 The Price Pressure Hypothesis

According to this hypothesis, the index effect would be due to a transient liquidity shortfall likely to happen when index funds operate a major and collective rebalancing. Investors who provide liquidity to the market without any motivation to trade should be compensated by a premium that reflects their extra costs and the risks of these trades. Therefore, prices are expected to temporarily increase (resp. decrease) to convince those passive investors to sell (resp. buy) shares in order to absorb the temporary positive (resp. negative) demand shift. Economically speaking, the demand curve of the individual assets moves to the right for a while, which results in higher price and trading volume, and then comes back to its initial position, with initial equilibrium price and volume. In our Equation (1-31), the Price Pressure Hypothesis translates into a temporary deviation ε_j. It does not affect the parameters (expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$) of the future price distribution since it is an information-free event. The Price Pressure Hypothesis is a relatively old hypothesis, already popularized by Scholes (1972) in his study on secondary offerings. Harris and Gurel (1986) are among the very first authors to provide evidence in favor of this hypothesis in the specific context of the index effect. They use an event study approach and examine cumulative abnormal returns around inclusions and deletions events occurring between 1973 and 1983. The authors do not observe any significant price increase during the 1973-1977 period, probably because index funds were not popular enough at that time. However, on the period ranging from 1978 to 1983, they observe a significant cumulative abnormal

return of 3.13% after inclusion, but just temporary since it is fully reverted after two weeks.

At first sight, the trading volume should also exhibit the same kind of behavior: a temporary increase due to the massive purchase or sale of shares by the index funds, followed by a reversal to equilibrium when the index funds have finished their rebalancing operations. However, Harris and Gurel (1986) find a permanent average trading volume increase. They do not find this result counter-intuitive and justify it by "the presence of funds, not necessarily index funds, which only invest in securities on the S&P 500 list". Contrary to the temporary shift of demand originating from the rebalancing of the index funds, that latter kind of additional demand would be permanent and benefit to all S&P 500 members. This argument already announces the Liquidity Cost Hypothesis that will follow below.

Woolridge and Ghosh (1986) find results totally opposite to the ones of Harris and Gurel (1986). They observe a permanent price increase (resp. decrease) and a temporary volume increase for additions (resp. deletions) over the 1977 to 1983 period. This permanent price effect is confirmed by a contemporaneous study of Arnott and Vincent (1986) who find significant and persistent price increases (resp. declines) for four weeks after the addition (resp. deletion) over the 1980 to 1984 period.

Nevertheless, some later authors will give some hope to the defendants of the Price Pressure Hypothesis. Lamoureux and Wansley (1987) show that prices and volumes reverse about 20 days after the event, and Lynch and Mendenhall (1997) consider index funds to be the main driving force behind the temporary price pressure and they argue that the latter would end once trading volume returned to its normal level. In the same line, Malkiel and Radisish (2001), the most recent paper supportive of this hypothesis, confirm the results of Harris and Gurel (1986), with long horizon event studies on quarterly data explicitly including a linear time trend. However, Massa and Goetzmann (2003), in an analogous study, but using daily data and a conventional econometric methodology, reject a complete reversal.

In the end, this hypothesis remains subject to question, as most other empirical research finds permanent rather than temporary increases. For example, Garry and Goetzmann (1986) and Shleifer (1986) find no reversal while Dhillon and Johnson (1991), Beneish and Whaley (1996), Lynch and Mendenhall (1997), and Kaul, Mehrotra and Morck (2000) find only a partial reversal.

2.2.2.1.2 The Imperfect Substitutes and the Downward-Sloping Demand Curve Hypothesis

If Scholes (1972) is right when he claims that stocks are not "unique works of art" and their demand curves are kept flat by arbitrage between perfect substitutes, the concept of a downward-sloping demand curve for stocks does not make sense. But it is unlikely that specific stock characteristics could be exactly described by a combination of other securities and we know that replicating strategies are rarely perfect. Provided that substitution between assets presents at least some imperfections, the individual stock demand curves may not be flat anymore and could exhibit some kind of downward-sloping shape. It follows that any excess demand (like index funds demand) on the market will push the demand curve to the right and price up until the equilibrium between supply and demand is reached again. As long as this supplementary demand is permanent (which should be the case in the context of index funds rebalancing), no reversal is expected in the long term and abnormal trading activity should be transient until new equilibrium. In terms of our theoretical model of Section 1.6, the effect is similar to the one of the Price Pressure Hypothesis, except that it translates into a **permanent** (instead of temporary) deviation ε_j in our Equation (1-31) Likewise, it does not affect the parameters (expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$) of the future price distribution since it is an information-free event.

This hypothesis implicitly supposes that the primary strategy of the index funds is to replicate the index as closely as possible, while one could imagine other mimicking strategies, like the use of derivatives (mainly swaps) or correlated stocks, in order to reduce the transaction costs of a perfectly replicated portfolio of the 500 assets. This is confirmed by Morck and Yang (2002), when they write that "although indexing can also be accomplished with derivatives, many institutional investors and mutual funds constrain themselves from using derivatives. They must therefore hold the index stocks. Also, because index fund managers are penalized for tracking error, and so *must* hold precisely the stocks in the index they are tracking and no others. Stocks in the index thus do not have close substitutes insofar as far as these index fund managers are concerned".

Shleifer (1986) brings important evidence to this Downward-Sloping Demand Curve hypothesis when he shows that the share price increase at the announcement date of an

index addition is positively related to the shift of the demand curve for stocks. Wurgler and Zhuravskaya (2002) and Petajisto (2006) support this premise as they show that slope of the stocks' long-run demand curves varies with the level of arbitrage risk. In the specific context of index funds, institutional investors and mutual funds have no other choice than holding the index stocks. Index fund managers are indeed penalized for tracking error, and so must hold precisely the stocks in the index they are tracking and no others. It follows that from their point of view, stocks in the index do not have close substitutes, and any position in imperfect substitutes would come with an undesirable additional risk. In an interesting setting, Kaul et al. (2000) examine the redefinition of stocks weights in the index due to public float changes, which is an (a priori) information-free event, and report a significant and permanent price increase around and after the event date. Finally, Morck and Yang (2002) observe that S&P 500 companies present a valuation premium correlated with the growth of indexing, which is consistent with the downward sloping demand curve hypothesis.

One of the central issues, with this hypothesis, is whether changes in investor demand can be large enough to affect individual stock prices. While the above studies have already brought some evidence, another interesting argument has been made concerning this point. Indeed, when examining the correlation between capital inflows and stock prices, an issue of causation arises. Capital inflows might increase investor demand, and so push up prices. The other way around, rising prices might induce "positive feedback" investors to buy equity, as in De Long et al. (1990). Warther (1995) argues that his finding of a simultaneous correlation of returns with inflows is not due to positive feedback traders since he finds no association between previous month returns and current month inflows. However, Edelen and Warner (1999) find evidence of positive feedback using bi-weekly and daily data.

2.2.2.2 Perfect elasticity based hypotheses

2.2.2.2.1 The Liquidity Cost Hypothesis

If an index is followed and tracked by many of the largest index funds, the inclusion of a stock to this index should significantly increase its liquidity, i.e. the investors ability to sell a stock almost immediately at the current price. Mikkelson and Partch (1985) and Amihud and Mendelson (1986) find that an increase in the stock's liquidity can result in

an increase in its price due to lower transaction costs and bid-ask spreads. In the case of an inclusion in the S&P 500 index, this increase in liquidity may be mainly due to higher trading volume or greater information production resulting in reduced information asymmetry. The mechanism at work would be thus the following: addition to the S&P 500 should logically be followed by greater institutional interest, greater trading volume and/or information production, greater liquidity, lower bid-ask spreads, lower required rate of return and, eventually, permanent price increase for the included stock. Edmister, Graham and Pirie (1996) discover permanent price effects attributed to permanently increased liquidity after inclusion and a permanent decrease in trading costs, rejecting the Price Pressure Hypothesis (which anticipates temporary price and volume effects), as well as the Downward Sloping Demand Curve Hypothesis (which anticipates permanent price but temporary volume effects). Erwin and Miller (1998) also find permanent increase in price and volumes and permanent decrease in bid-ask spreads, with the notable exception of stocks with traded options, for which they find no significant changes in price and bid-ask spreads. The existence of two markets sharing the arbitrage profits could explain this specificity.

However, this hypothesis is not uniformly supported. Beneish and Whaley (1996) show that the bid-ask spreads are not permanently reduced following S&P addition, while Kaul et al. (2000) find no significant changes in the bid-ask spreads around information-free event dates that cause changes in stock supply. Eventually, Jain (1987) and Blitzer and Dash (2004) observe price effects of significant magnitudes in other indices that are not as closely followed by large index tracking funds as the S&P 500. Evidence in favor of the liquidity hypothesis is thus quite mixed.

The Liquidity Cost Hypothesis is more difficult to apprehend in our theoretical model since the latter does not consider the transaction cost issue. However, the increase of liquidity predicted by the Liquidity Cost Hypothesis in case of inclusion should lead to a decrease of information asymmetry and help to partly diminish the uncertainty around the stock future cash-flows and thus, its standard deviation σ_j.

2.2.2.2.2 The Information Content Hypothesis

Although S&P's Index Committee claims that changes in the famous index's composition is a pure information-free event, many researchers have a different opinion

and suggest that the S&P 500 is no longer representative of the market as a whole, but rather is reflective of the stock picking ability of Standard and Poor's employees. In this view, any addition to or deletion from the index carries a significant piece of information about future expected cash-flows which is quickly incorporated into price and is solely responsible for the index price effect. In terms of Equation (1-31), it translates into higher expected future price μ_j for the included firm *j*. The price effect is expected to be permanent, whereas the volume increase is expected to be temporary. It already deserves some comments. First of all, note that these symptoms are identical to those predicted by the Downward-Sloping Demand Curve Hypothesis, which makes very difficult to discriminate the two hypotheses. Furthermore, the Information Content Hypothesis is potentially strongly endogenous and lets the causation issue open. Does index inclusion lead to easier access to capital and better operating performance (and thus higher expected future cash flows) due to closer scrutiny and monitoring of management? In this case, index inclusion carries and reveals information about **future** good performance, and we call it ICHa. Or the other way around, does S&P embed some non-public information into their selection process? Here, past and current information about good performance causes index inclusion: this is the Certification of an Index Member Hypothesis, a particular form of the Information Content Hypothesis, and we call it ICHb. The two explanations are not exclusive and it is hard to disentangle one from the other. Goetzmann and Garry (1986) find a long-term drop in the prices of 7 deleted stocks. They suggest that delisting conveys special information that analysts would focus less on these firms in the future. Nevertheless, they observe a price fall previous to the announcement as well, which could indicate less good fundamentals at that moment. The deletion would then simply confirm existing bad news rather than provoking future poor monitoring.

Other authors provide some evidence that inclusion in the index amounts to a "certification of quality" and that this is the ultimate cause of the value increase. For example, Jain (1987) finds abnormal returns for additions into industry indexes that are not used as passive investment benchmarks, and Dhillon and Johnson (1991) show that included firms' bond and call prices rise in step with their stock prices.

Jacques (1988)'s results tend to support the view that inclusion can improve future performance of the firm. He finds indeed about 4% extra return per year (1983-1988) for S&P stocks. This better performance could trigger a positive feed-back loop, since

financial institution may be more willing to lend to firms that are index members, simply because their future performance is expected to be better. In turn, easier access to capital will involve better operating performance, higher expected cash-flows, and again higher realized returns and performance, which closes the loop. Using behavioral arguments, Subrahmanyam and Titman (2001) explain how "feedback from financial market prices to cash flows arises when a firm's nonfinancial stakeholders, for example, its customers, employees, and suppliers, make decisions that are contingent on the information revealed by the price. Complementarities across stakeholders result in cascades, wherein relatively small stock price moves trigger substantial changes in asset values" (p. 2389).

Notice that such an information-based hypothesis could explain the results obtained by large block trades studies, such as Lakonishok, Shleifer and Vishny (1991, 1992) and Chan and Lakonishok (1993, 1995) among others. Purchases or sales by large blockholders would affect the prices because they allow public shareholders to better infer the true value of the firm. It is also consistent with the idea of Jensen and Meckling (1976) that concentrated ownership mitigates corporate governance problems.

However, support for such a hypothesis is far from being unanimous. Wurgler and Zhuravskaya (2000) find that the abnormal returns associated with inclusion in the S&P 500 are larger for stocks that are less likely to have close substitutes. And Kaul et al. (2000) show that the pure mechanical rearrangement of the weights of stocks already present in a widely followed index, without any addition or deletion, still leads to price effects. Such findings are consistent with a demand shift explanation, but rule out categorically any certification or information-based story.

The most recent defendant of the Information Content Hypothesis is Denis, McConnell, Ovtchinnikov and Yu (2003), who find that newly included companies experience significant increases in earnings per share forecasts and improvements in realized earnings. Inclusion of firms in the S&P 500 index seems thus to contain information about their future performance. However, Hradzil and Scott (2007) examine this contention and while they confirm Denis et al. (2003)'s positive earnings news, they show that the latter is not associated with announcement returns, and further that the observed positive earnings news is entirely attributable to large discretionary accruals consistent with managerial manipulation of earnings.

2.2.2.2.3 The Market-Segmentation and Investor Recognition Hypothesis

It's Merton (1987) who is at the source of this interesting behavioural Investor Recognition Hypothesis. According to him, investors are only fully aware of "famous" and commonly traded stocks (typically, S&P 500 member stocks), and have a quite poor knowledge of the rest of the market. As a consequence, they are rather prone to "hold only the stocks they are aware of, so that they will be inadequately diversified and demand a premium—a shadow cost—for the nonsystematic risk they bear"[7]. Chen, Noronha and Singal (2004) brought empirical support to this theoretical hypothesis as they argue that a stock's addition to the S&P 500 index makes the traders conscious of its existence, and as soon as index funds (for example) rebalance their portfolios to integrate the new stock, "the required rate of return on that stock should fall due to a reduction in the shadow cost"[8]. Up to here, predicted price effects are roughly in line with the ones announced by the previous hypotheses, except on one interesting point. This hypothesis, adapted by Chen, Noronha and Singal (2004) to the context of the index effect, is indeed quite original in that it implies an asymmetrical reaction in case of addition or deletion. Since it would not make sense that investors suddenly become unaware of a freshly deleted stock, the increase in the shadow cost of the deleted cost is expected to be significantly lower than the one of the newly included stock. Consistently, their observations of cumulative abnormal returns demonstrate a permanent but asymmetrical effect on the stock price after the index change.

In our theoretical framework, the Investor Recognition Hypothesis is typically an under-diversification issue. When a stock is included in the index, it comes out of the shadow and investors suddenly become (more) aware of it. As a consequence, they choose to rebalance their portfolio and increase the weight of the new stock. Such a process can be understood as passing from a negative deviation (holdings of the stock are under optimum) $\varepsilon_j<0$ to a null or positive deviation $\varepsilon_j\geq 0$, once the stock is recognized. Note that contrary to the deviation ε_j induced by the Downward-Sloping Demand Curve Hypothesis, which is directly linked to index funds rebalancing and thus to the weight of the firm j in the index, the deviation ε_j implicitly induced by the Investor Recognition Hypothesis (stricto sensu) only depends on the fact to be recognized or not by the

[7] Chen, Noronha and Singal (2004), p. 1902.

[8] Ibidem, p. 1902.

investors, and so should be roughly independent of the proportion of the company capitalization in the S&P 500 (proxied later by the variable *SPxWeight*).

Given the many existing competing hypotheses, the topic can appear as quite controversial. However, although some of those hypotheses clearly exclude each other, most of them are rather complementary than exclusive. Moreover, the different if not contradictory results are not so puzzling if we consider that data and methodology used can significantly vary across the numerous empirical studies on this topic.

2.2.3 Contribution

2.2.3.1 Positioning of our theoretical model

2.2.3.1.1 Discussion about the elasticity of the demand curve

We have seen in Section 1 that our theoretical model assumes an inelastic or at least imperfectly elastic supply curve. However, on the demand side, no explicit assumption has been done. This question is worth investigating: does our theoretical model implicitly assume perfect or imperfect elasticity of individual stock demand curve? This is an important issue, because if the demand curve is perfectly elastic, the deviations will have no effect on the equilibrium price and quantities. In order to see why, let us write the equation of the aggregate demand as a function of the price, **D(p)**, which we can deduce in a straightforward manner from Equation (1-30):

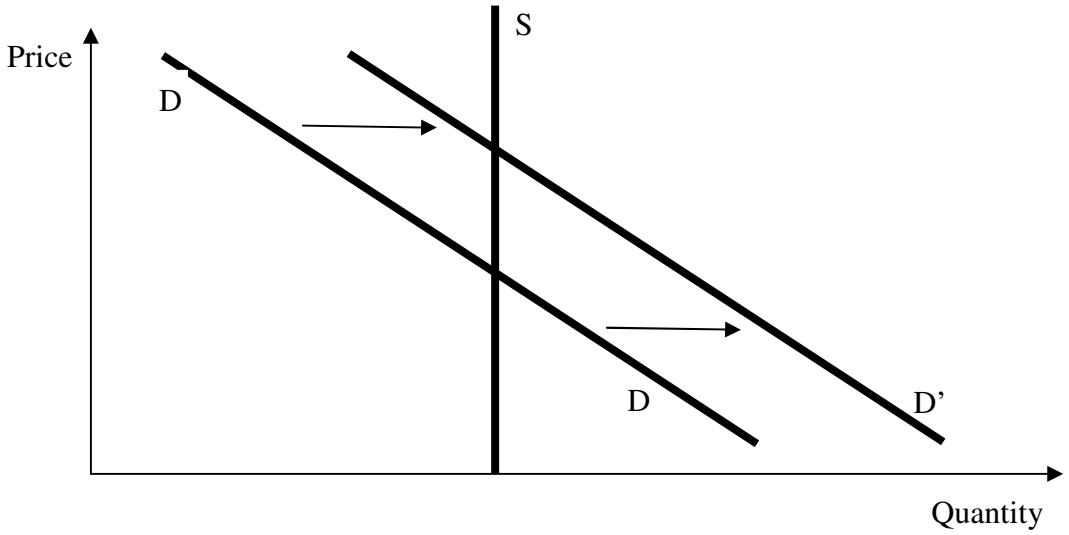
$$\mathbf{D(p)} = \mathbf{X}^{C'}\mathbf{1}_N = \boldsymbol{\Sigma}^{-1}\frac{\boldsymbol{\mu} - R_f\mathbf{p}}{bN} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi} \tag{2-1}$$

where:

$$\boldsymbol{\varphi} = diag(diag(\boldsymbol{\Sigma}) * \boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma_1^2\varepsilon_1 \\ \vdots \\ \sigma_j^2\varepsilon_j \\ \vdots \\ \sigma_K^2\varepsilon_K \end{pmatrix}$$ and $\boldsymbol{\varepsilon} = \mathbf{E}'\mathbf{1}_N$ is a [K×1] vector of the aggregate deviation in each asset.

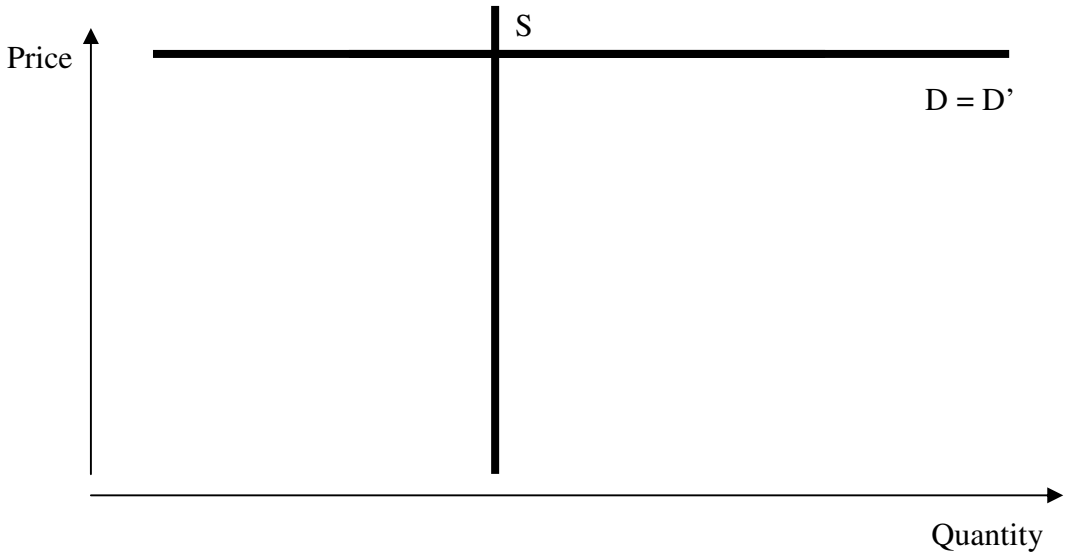
This equation tells us that a positive (resp. negative) global deviation in asset j ε_j translates (through the $\boldsymbol{\varphi}$ term) into a shift of the aggregate demand curve of asset j to the right (resp. left). Whether this shift will have an impact on the equilibrium price and quantities depends on the elasticity of the demand curve. If the aggregate demand curve of asset j is imperfectly elastic (i.e., it is at least slightly downward-sloping), any shift of the aggregate demand curve will have an impact on the equilibrium price, as pictured in Figure 2-1.

Figure 2-1: Impact of a positive deviation on the equilibrium with a partially inelastic demand



On the contrary, if the demand curve is perfectly elastic (i.e. horizontal), a shift of the demand curve leaves the equilibrium price unchanged, as made clear by Figure 2-2.

Figure 2-2: Impact of a positive deviation on the equilibrium with a perfectly elastic demand



The slope of the aggregate demand curve of asset *j* is given by the partial derivative of the price p_j of asset *j* with respect to the aggregate demand quantities D_j of asset *j*. Starting from the above equation of **D(p)**, we first isolate **p** as a function of **D** and we write it in particular for asset *j* (scarlar way):

$p_j = \dfrac{\mu_j - \dfrac{b}{N}\left(\sum_{k=1}^{K}\sigma_{j,k}D_k - \sigma_j^2\varepsilon_j\right)}{R_f}$ where D_k is the aggregate demand quantities for asset *k*.

The partial derivative of the price p_j of asset *j* with respect to the aggregate demand quantities D_j of asset *j* is thus given by:

$$\frac{\partial p_j}{\partial D_j} = \frac{\partial\left[\dfrac{\mu_j - \dfrac{b}{N}\left(\sum_{k=1}^{K}\sigma_{j,k}D_k - \sigma_j^2\varepsilon_j\right)}{R_f}\right]}{\partial D_j} = -\frac{b\sigma_j^2}{NR_f} \tag{2-2}$$

As *N* grows very large, any individual action is drown in the mass and is unlikely to have any significant price impact. However, inclusions are precisely this kind of events that are possibly followed by collective rebalancing of large index funds such that observable price effect may arise. In fact, in our theoretical model, the elasticity of the demand curve must be understood as a relative concept. If deviations are isolated and not widespread in the market, they will be diluted as the number of investors increases. In this case, the demand curve is perfectly elastic in practice. Mathematically speaking, the global deviation $\varepsilon_{j,t}$ does not increase as the number of investors *N* increases so that the resulting effect of the global deviation on the price tends to zero:

$$\lim_{N\to+\infty}\frac{\partial p_j}{\partial \varepsilon_j}\varepsilon_j = \lim_{N\to+\infty}\frac{b\sigma_j^2}{NR_f} * \varepsilon_j = 0 \tag{2-3}$$

However, if those deviations are systemic, collective and largely followed by a broad range of investors on the market, the global deviation ε_j increases at the same pace as the number of investors *N*, and may compensate for the very weak slope of the demand curve $-\dfrac{b\sigma_j^2}{NR_f}$, which decreases at the same pace as *1/N*. In this case, a non-negligible term could appear in the pricing equation and drives prices up. Consequently, the demand cannot be considered as perfectly elastic anymore with respect to the magnitude of the global deviation. Mathematically speaking, the resulting effect of the deviation on the asset price is:

$$\lim_{N\to+\infty}\frac{\partial p_j}{\partial \varepsilon_j}\varepsilon_j = \lim_{N\to+\infty}\frac{b\sigma_j^2}{NR_f} * \varepsilon_j = \lim_{N\to+\infty}\frac{b\sigma_j^2}{NR_f} * o(N) \propto \frac{b\sigma_j^2}{R_f} \tag{2-4}$$

However, if our model does not rule out an almost perfect elasticity explanation in certain cases, the underlying reason for such elasticity is the atomicity of investors (large *N*) rather than perfect substitution between assets. In our mean-variance setting, perfect substitution between assets can only arise if these assets returns are perfectly correlated and exhibit the same probability distribution (i.e. the same mean and variance if they are gaussian). Furthermore, we know that market imperfections and transaction costs make actually this perfect substitution quite doubtful.

2.2.3.1.2 The link between the five hypotheses and our theoretical model

We have seen in Section 2.2.2 that all the 5 explanations proposed by the literature introduce a "shock" into the price equation through different means, namely ε_j (the Price Pressure Hypothesis, the Downward-Sloping Demand Curve Hypothesis and the Investor Recognition Hypothesis), μ_j (the Information Content Hypothesis) or σ_j (the Liquidity Cost Hypothesis) but finally lead to the same consequence: a stock price increase (always permanent, except for the Price Pressure Hypothesis). The Price Pressure Hypothesis, the Downward-Sloping Demand Curve Hypothesis and the Investor Recognition Hypothesis can be translated in terms of deviations into our theoretical model. Consequently, if we find some evidence for those hypotheses, it would bring by the same token strong support to our "under-diversification" thesis and the warranted demand effect we expect to observe for firms members of the S&P 500.

2.2.3.2 Positioning of our empirical model

It is of course too early to unveil our empirical results and comment them in details, but it is worth noticing that if our theoretical model is general enough to integrate most of hypotheses, we have constructed our empirical setting of next Section 2.3 in such a way that some hypotheses may not survive our results. The Price Pressure Hypothesis, Downward-Sloping Demand Curve Hypothesis, ICHb and Investor Recognition Hypothesis are directly tested (at least with respect to some of their implications) in our regression framework. The Liquidity Cost Hypothesis and ICHa are more difficult to check in a straightforward manner, but our results can deliver some clues about their

plausibility. The purpose of the following section is to describe in detail our empirical model.

Table 2-1: Summary of the theoretical reasoning underlying the competing hypotheses

PPH stands for Price Pressure Hypothesis, DSH for Downward-Sloping Demand Curve Hypothesis, LCH for Liquidity Cost Hypothesis, ICH for Information Content Hypothesis (ICHa: performance causes inclusion; ICHb: inclusion causes performance) and IRH for Investor Recognition Hypothesis. D stands for Demand, S for Supply and DS for Downward-Sloping.

					PPH	DSH	LCH	ICH		IRH
								ICHa	ICHb	
THEORY	Underlying asset pricing process:	Market Clearing condition Vs Fisher equation	Price = $D \cap S$		X	X				
			Price = $\sum_t \{ E[CF_t/(1+R_t)] \}$				X	X		X
	Underlying assumptions:	Imperfect stock substitutes Not perfectly elastic D and S curves	If inclusion, price ↑ because	D is Short-Term DS	X					
				D is Long-Term DS		X				
		Perfect stock substitutes Perfectly elastic (flat) D curve		CF_t ↑				X		
				R_t ↓			X			X (asym)
	In our theoretical framework:	- Imperfect stock substitutes (except particular cases) - Not perfectly elastic S curve - No rigid assumption on D curve (elasticity is relative and depends on *N*)		ε_j ↑	X (temp)	X (perm)				X (asym)
				σ_j ↓			X			
				μ_j ↑				X		

2.3 Data

2.3.1 Data sample

Our basis sample begins with all US firms listed on CRSP from September 1976 to December 2005 with SHRCD codes 10, 11, 12 and 18. With such a sample, we are certain to cover all possible candidates firms to an inclusion in the S&P 500, and a representative sample of the US market. Accordingly to common filters used in the literature and in Morck and Yang (2002), we do not include firms in banking and financial industries (i.e. Standard Industrial Classification (SIC) codes ranging from 6000 to 6999 included) as accounting information for these firms are quite specific and not directly comparable to that of other firms[9]. Similarly to Morck and Yang (2002), we delete observations in which the share price, the volume, the number of shares outstanding, the reported earnings, the common equity, the total assets, the earnings per share, the SIC code or the stock return are missing, or negative when the variable should clearly be non-negative (e.g., the price). However, if these variables exist, but entries for research and development spending, advertising spending, short term debt and long term debt are missing, the missing variables are assumed to be nil.

As reported by Morck and Yang (2002), this sample contains some extreme observations, which probably reflect coding errors by Compustat. We therefore winsorize the data at the first (except for the non-negative variables) and 99th percentiles for all important variables. Morck and Yang (2002) showed that such a winsorising allowed to get rid of most of problematic data.

Still in line with Morck and Yang (2002), "we do not include firms smaller" (in terms of "total assets") "than the smallest S&P 500 index firm for a specific month on the grounds that very small firms may not be valued by investors in the same way as larger firms" (p. 14). However, this argument is questionable, since we have also performed all our regressions on an extensive sample including all firms with sufficient financial data, whatever their size, and results have proved to be qualitatively similar. They are available upon request.

[9] In order to assess the sensitivity of our results to the exclusion of the financial sector from the population, we also performed our study on the population including this financial sector. Results proved to be qualitatively similar to the ones presented in the current study. They are available on request.

A company is in the S&P 500 in month *t* if it is in the index the last trading day of that month. Index additions and deletions data until 2000 come from Jeff Wurgler's website, while subsequent data until December 2005 are constructed on the basis of information available on the S&P website.

Table 2-2: Historical evolution of S&P 500 additions and deletions

This table provides the yearly number of additions (*SPadd*) and deletions (*SPdel*) on the S&P 500 index, on the initial raw database, as received from Jeff Wurgler's website and the S&P 500 website. *N_SP* is the number of S&P 500 members in the sample, before the data treatment (suppression of the financial sector, winsorizing, etc.) and after. We also give the number of firms not included in the S&P in our final sample (*N_NoSP*), and the total number of firms in our final sample (*N_Tot*), after treatment.

	Before treatment			After treatment		
Year	SPadd	SPdel	N_SP	N_SP	N_NoSP	N_Tot
1976	6	0	481			
1977	8	0	489	408	1,374	1782
1978	11	0	500	420	1,351	1771
1979	14	14	500	420	1,220	1640
1980	12	12	500	422	1,361	1783
1981	21	21	500	422	1,862	2284
1982	29	29	500	417	1,145	1562
1983	19	19	500	407	1,251	1658
1984	30	30	500	403	1,376	1779
1985	28	28	500	408	1,290	1698
1986	28	28	500	411	1,297	1708
1987	27	27	500	407	1,228	1635
1988	26	26	500	399	1,224	1623
1989	30	30	500	407	1,221	1628
1990	13	13	500	404	1,313	1717
1991	13	13	500	406	1,262	1668
1992	7	7	500	412	924	1336
1993	13	13	500	404	1,531	1935
1994	17	17	500	400	1,025	1425
1995	35	35	500	399	891	1290
1996	24	24	500	394	885	1279
1997	31	31	500	398	829	1227
1998	48	48	500	395	1,013	1408
1999	41	42	499	396	1,142	1538
2000	58	57	500	386	1,342	1728
2001	30	30	500	385	874	1259
2002	24	24	500	376	926	1302
2003	9	9	500	385	1,160	1545
2004	20	20	500	392	1,227	1619
2005	20	20	500	390	1,020	1410

Table 2-2 describes the historical evolution of the number of S&P 500 additions and deletions.

Out of the initial 500 S&P companies,

- about 5% are lost due to missing values;
- between 6% (beginning of the period) and 20% (end of the period) are lost due to the deletion of the financial companies;
- less than 3% are lost due to the winsorizing process (deletion of outliers).

Table 2-3 provides the main descriptive statistics of all variables used in this empirical study. All those variables will be described in detail in the next section.

Table 2-3: Descriptives statistics

This table provides descriptive statistics (minimum, maximum, median, mean, standard deviation, skewness, kurtosis and the number of observations) for all variables used in this empirical study. The observations taken into consideration are the ones obtained after various pre-treatment (winsorizing, deletion of financial industries, and so on). They are those used in the regressions. *Dat* is the date, in format yyyymmdd. *PER* is the Price-Earning ratio, *Ret* the monthly return, *TQ* the Tobin's Q, *MB* the Market-to-Book ratio, *SP* the S&P 500 membership dummy, *SPpast* a dummy = 1 if the firm has been a member of the S&P 500 but is not anymore, *SPadd12* (resp. *SPdel12*) a dummy = 1 if the firm has been included in (resp. removed from) the S&P 500 over the last 12 months, *MktCap* the market capitalization of the firm in $ thousands, *NoIPO* a dummy = 1 if no missing value has been detected in CRSP data (prices, volumes and returns) over the last 12 months, *Liq* the liquidity ratio, *Prc* the price, *RpEarn* a dummy = 1 in case of 4 consecutive quarters of positive earnings, *Lev* the leverage, *Variance* a monthly historical variance of returns computed on the last 12 months, *RD* the % of Research & Development Spending of the firm in its total assets, *Adv* the % of advertising spending of the firm in its total assets, *Assets* the log of the total assets value of the firm, *SalesG* the annual growth of the sales, *Prc_Low* (resp. *Prc_High*) a dummy = 1 if the price is in the lowest (resp. highest) quintile, *SIC2* the sectorial code of Fama and French, *Lev_Ind* (resp. *PER_Ind*, *TQ_Ind* and *MB_Ind*) is the industry-adjusted *Lev* (resp. *PER*, *TQ* and *MB*), *DLiq* is a dummy = 1 if the liquidity ratio *Liq* > 3, *MktCap_SP* is the S&P 500 market capitalization in $ billions, *Weight* = *MktCap*/*MktCap_SP*, *SPxWeight* (=*SP***Weight*) the weight of the firm in the S&P 500 index and *SPxWeightxVar* the product *SP***Weight***Variance*.

Variable name	Min	Max	Median	Mean	StdDev	Skewness	Kurtosis	N_Obs
Dat	19760930	20051230	19900831	19904832	84667.17	0.08	-1.20	619479
PER	0.13	310.42	13.75	20.11	25.61	5.46	39.46	513342
Ret	-0.40	0.62	0.01	0.01	0.12	0.43	2.36	613756
TQ	0.15	11.50	1.22	1.52	0.95	3.76	21.08	617502
MB	0.00	26.45	1.58	2.22	2.27	4.07	24.49	600198
SP	0	1	0	0.2369	0.43	1.24	-0.47	619479
SPpast	0	1	0	0.0151	0.12	7.95	61.21	619479
SPadd12	0	1	0	0.0093	0.1	10.22	102.48	597397
SPdel12	0	1	0	0.0034	0.06	17.02	287.56	597397
MktCap	19.38	6.02E+08	411862	2533468	12134088	17.54	449.72	616129
NoIPO	0	1	1	0.98	0.14	-7.07	47.96	616129
Liq	0.00	69.19	5.86	9.02	10.08	2.47	7.61	611111
Prc	0.01	35500.00	21.38	27.83	192.52	116.38	15781.74	616129
RpEarn	0	1	1	0.62	0.49	-0.47	-1.77	619479
Lev	0.00	0.90	0.28	0.29	0.18	0.50	0.02	614310
Variance	0.00	0.28	0.01	0.01	0.02	5.01	38.26	615573
RD	0.00	0.47	0.00	0.02	0.04	3.40	16.01	619426
Adv	0.00	0.20	0.00	0.01	0.03	3.13	11.55	614226
Assets	3.02	13.53	6.38	6.49	1.52	0.47	-0.01	619479
SalesG	-0.66	3.75	0.10	0.15	0.31	3.87	26.87	599962
Prc_Low	0	1	0	0.05	0.23	3.97	13.73	619479
Prc_High	0	1	0	0.05	0.22	4.15	15.22	619479
SIC2	1	49	31	27.09	12.19	-0.37	-1.02	617618
Lev_Ind	-0.60	0.77	-0.01	0.00	0.17	0.52	0.53	614310
PER_Ind	-261.18	289.36	-3.28	0.00	24.23	5.47	42.22	513280
TQ_Ind	-4.73	10.12	-0.12	0.00	0.84	3.12	19.68	617502
MB_Ind	-13.74	24.00	-0.28	0.00	2.04	3.73	25.50	600198
DLiq	0	1	1	0.93	0.25	-3.37	9.39	611111
MktCap_SP	515.64	12400.56	2197.55	3979.13	3690.499	0.89	-0.79	619479
Weight	0.00	0.08	0.00	0.00	0	13.06	263.98	616129
SPxWeight	0.00	0.08	0.00	0.00	0	13.26	269.13	616129
SPxWeightxVar	0.00	0.00	0.00	0.00	0	26.71	1235.65	614421

2.3.2 Variables definition

This section describes the construction of all variables that will be used in our various specifications of Section 2.4. It is not the primary aim of the current section to explain why and in which regression we use those variables, and their exact function in our empirical framework. This topic is left to Section 2.4.

Most of them are constructed from **annual**[10] Compustat data ("*item*" refers to Compustat elements) while some come from the **monthly** CRSP data base ("*CRSP*" refers to them). Our first set of variables consists of the ones that will be used in the main value equation, i.e. various variables explaining the level of our four proxies of valuation: the annual Price-Earning Ratio (*PER*), the monthly stock Return (*Ret*), the annual Tobin's q (*TQ*) and the annual Market-to-Book ratio (*MB*).

The Price-Earning Ratio is usually interpreted as the anticipation that the investors have of the future earnings growth of the company: the higher this growth, the higher the *PER*. Of course, this interpretation has to be nuanced, as the *PER* may be influenced by many other factors, like the industry, the economic cycle, etc.

The Tobin's q is supposed to indicate whether a firm is overvalued or undervalued. A ratio equal to 1 indicates that the firm is "fairly valued", while a ratio above (resp. below) 1 means that the company is overvalued (resp. undervalued). Again, this ratio is not perfect and must be used with caution, and preferably with other valuation proxies like in this study.

The Market-to-Book ratio has a close interpretation to the Tobin's q. Both ratios are usually correlated. This ratio aims at comparing two estimations of the value of the firm: an anticipated value given by the market capitalization of the company, and an intrinsic value given by its book value. A high Market-to-Book ratio usually indicates a firm with important growth prospects. Note that this ratio is very dependent on the industry.

The proxies we use for all our dependent and explanatory variables are quite standard in the literature. However, Morck & Yang (2002) compute the Tobin's q and especially the leverage (*Lev*) in a different, more sophisticated way. Furthermore, their sample is smaller than ours, since it only contains companies above a minimum size. Our results

[10] With one exception, since **quarterly** Compustat data are used to compute $RpEarn_{j,t}$

are generally very close to theirs, with the noticeable exception of the sign of the leverage coefficient, positive in their setting, and negative in ours. But the leverage is only a control variable and we find qualitatively similar results on our main variables of interest, $SP_{j,t}$ and $SPxWeight_{j,t}$, therefore we do not consider this difference as problematic.

We define our variables as follows:

- $PER_{j,t} = \frac{item199_{j,t}(Fiscal\ Year\ Price)}{item58_{j,t}\ (EPS\ (Basic),Exclude\ Extra.Items)}$ (2-5)
- $Ret_{j,t} = CRSP_M_RET_{j,t}\ (Holding\ Period\ Return)$ (2-6)
- $TQ_{j,t} = \frac{item6_{j,t} - item60_{j,t}(Com.\ Eq.) + item25_{j,t}(Common\ Shr\ Outst.) * item199_{j,t}}{item6_{j,t}(Total\ Assets)}$ (2-7)
- $MB_{j,t} = \frac{item24_{j,t}(Calendar\ Year\ Price) * item25_{j,t}(Common\ Shares\ Outstanding)}{item60_{j,t}(Common\ Equity)}$ (2-8)
- $SP_{j,t}$ is a dummy indicating whether firm *j* is in the S&P 500 index at the end[11] of year *t*
- $SPpast_{j,t}$ is a dummy indicating whether firm *j* has been at least once in the S&P500 index during the period [*0;t-1*] **AND** is not anymore in the S&P 500 at the end of year *t:*

  $$SPpast_{j,t} = \left(1 - \prod_{s=1}^{t-1}\left(1 - SP_{j,s}\right)\right) * \left(1 - SP_{j,t}\right) \quad (2\text{-}9)$$

- $SPadd12_{j,t}$ is a dummy indicating whether firm *j* has been included in the S&P 500 index during year *t*; in other words, $SPadd12_{j,t}$ is set to one if firm *j* is in the S&P 500 at the end of year *t* but was not in the index at the end of year *t*-1:

  $$SPadd12_{j,t} = \left(1 - SP_{j,t-1}\right) * SP_{j,t} \quad (2\text{-}10)$$

- $SPdel12_{j,t}$ is a dummy indicating whether firm *j* has been deleted from the S&P 500 index during year *t*; in other words, $SPdel12_{j,t}$ is set to one if firm *j* is not in the S&P 500 at the end of year *t* but was a member of the index at the end of year *t*-1:

  $$SPdel12_{j,t} = SP_{j,t-1} * \left(1 - SP_{j,t}\right) \quad (2\text{-}11)$$

[11] Precisely, on December 31. When the return is the dependent variable, it is the situation at the end of each month that matters.

To make those different dummies perfectly clear, let us consider the following example:

Firm *j* has been included in the S&P 500 at time t=5 and has been removed at time t=8. The sequences of our dummies would be such that:

t	1	2	3	4	5	6	7	8	9	10
SP	0	0	0	0	1	1	1	0	0	0
SPpast	0	0	0	0	0	0	0	1	1	1
SPadd12	0	0	0	0	1	0	0	0	0	0
SPdel12	0	0	0	0	0	0	0	1	0	0

- $Weight_{j,t} = \frac{CRSP_M_PRC_{j,t}(Closing\ Price) * CRSP_M_SHROUT_{j,t}}{\sum_{k \in S\&P500} CRSP_M_PRC_{k,t} * CRSP_M_SHROUT_{k,t}(Shr\ Outstanding)}$ (2-12)
- $SPxWeight_{j,t} = SP_{j,t} * Weight_{j,t}$ (2-13)
- $RD_{j,t} = \frac{item46_{j,t}(Research\ and\ Development\ Expense)}{item6_{j,t}(Total\ Assets)}$ (2-14)
- $Adv_{j,t} = \frac{item45_{j,t}(Advertising\ Expense)}{item6_{j,t}(Total\ Assets)}$ (2-15)
- $Lev_{j,t} = \frac{item34_{j,t}(Debt\ in\ Current\ Liabilities) + item9_{j,t}(Long-Term\ Debt)}{item6_{j,t}(Total\ Assets)}$ (2-16)
- $Assets_{j,t} = \ln\left(item6_{j,t}(Total\ Assets)\right)$ (2-17)
- $SalesG_{j,t} = NoIPO_{j,t} * \frac{item12_{j,t}(Net\ Sales) - item12_{j,t-1year}}{item12_{j,t-1year}}$ (2-18)
- $Variance_{j,t} = NoIPO_{j,t} * \frac{\sum_{k=t-11months}^{t} \left(Ret_{j,k} - \frac{\sum_{k=t-11months}^{t} Ret_{j,k}}{12} \right)^2}{12}$ (2-19)
- $NoIPO_{j,t} = \prod_{k=t-11months}^{t} 1_{CRSP_M_PRC_{j,k} \& SHROUT_{j,k} \& VOL_{j,k}\ not\ missing}$ (2-20)

The purpose of this variable *NoIPO* is to distinguish firms that are new on the market (i.e. less than one year of continuous quotation) from more mature firms (i.e. more than one year of continuous quotation). It serves three goals:

- Any missing observation in the 12 last monthly returns makes the computation of the variable *Variance* unfeasible. Rather than eliminating all observations with insufficient history, we use this *NoIPO* variable (=1 only if there is no missing values in the monthly price, share outstanding and volume series over the last 12 months) and we logically set *Variance*[12] to 0 when *NoIPO* is null. As a consequence, *NoIPO* will

[12] Strictly speaking, we should write *NoIPOxVariance* instead of *Variance* to emphasize the fact that it is an interaction variable, but we prefer to simply use "*Variance*" in order to alleviate slightly the notations. The same remark holds for *SalesG,* which is also an interaction variable.

capture the effect of the fact to have non-missing data in the recent history, while *Variance* will capture the effect of variance for those observations that have at least one year complete history.

- We proceed the same way for the variable *SalesG*. This variable is yearly, but a missing value of the lagged *SalesG* will most of the time translate into a missing value in the price / share outstanding / volume series. When it is not the case, the observation is deleted.
- This variable *NoIPO* will also be used later in order to know whether IPO is seasoned by 12 months or not. Indeed, financial data are missing before the IPO, so that the existence of 12 months of consecutive financial data ensures us that any IPO is seasoned by 12 months at least.

- $SPxWeightxVariance_{j,t} = SP_{j,t} * Weight_{j,t} * Variance_{j,t}$ (2-21)
- $year_{t,y} =$ 29 annual dummies for each of the 30 years (1976-2005) of the sample, except 1976 to avoid perfect collinearity (dummies are monthly when we use *Ret* as a dependent variable).

We also compute corrected variables for $PER_{j,t}$, $TQ_{j,t}$, $MB_{j,t}$ and $Lev_{j,t}$, which are denoted $PER_Ind_{j,t}$, $TQ_Ind_{j,t}$, $MB_Ind_{j,t}$ and $Lev_Ind_{j,t}$. We want to adjust them for the industry effect. To do so, we substract from the original variable $X_{j,t}$ its industry mean computed without the observation. We construct this new variable, denoted by $X_Ind_{j,t}$, as follows.

The initial variable is $X_{j,t}$. The industry *i(j,t)* is the industry of firm *j* at time *t*. Let us denote by $n_{i,t}$ the number of firms in this industry at time *t* and by $\bar{X}_{i(j,t),t}$ the average *X* of all firms (firm *j* **excluded**) in this industry *i(j,t)* at time *t*. Finally, our new variable $X_Ind_{j,t}$ is the initial *X* of firm *j* minus the average excluding firm *j*: $X_Ind_{j,t} = X_{j,t} - \bar{X}_{i(j,t),t}$.

Thus, the adjusted average *X* is the original *X* minus the average *X* of all other firms in the same industry (excluding the firm in question). If $n_{i,t} = 1$, the observation is dropped.

We present now the second set of variables that will be used as instruments of the fact to be in the S&P 500 (*SP*) in subsequent econometric regressions. We are looking for

variables explicative of the fact to be member of the S&P 500. The S&P 500 committee identifies on the S&P's website[13] several criteria for a firm to be introduced and maintained in the index, which can be useful in order to determine relevant instruments for the our $SP_{j,t}$ dummy. For each of those criteria, we try to find the best proxy.

1) ***US*** *companies.*

   Since our sample is restricted to US companies only, this condition is de facto fulfilled and cannot lead to any instrument.

2) ***Market capitalization*** *of $4 billion or more for the S&P 500, $1 billion - $4 billion for the S&P MidCap 400 and $300 million - $1 billion for the S&P SmallCap 600. Theses ranges are reviewed from time to time to assure consistency with market conditions.*

   We decide to keep our Market capitalization variable continuous. Indeed, we can reasonably suspect the ranges of acceptance given by the S&P guidelines to be most likely not stable through time from 1976 to 2005, since the market capitalization of firms have followed on average an increasing trend over this period. Therefore, the current thresholds that are adequate today were probably too high in the seventies and the eighties.

   It is of course the variable we expect to be the most predictive of the probability of being an index firm. It is computed as follows:

   $$MktCap_{j,t} = CRSP_M_PRC_{j,t} * CRSP_M_SHROUT_{j,t} \quad (2\text{-}22)$$

   Since this variable is likely to present (and actually it is the case) a distribution with a heavy right tail, we choose to take its natural logarithm:

   $$MktCaplog_{j,t} = \ln(MktCap_{j,t})$$

3) *Initial public offerings (****IPO****) should be seasoned for 6 to 12 months before being considered for addition to indices.*

   $NoIPO_{j,t}$ has already been defined above and is used in the probit to ensure that IPO is seasoned for at least 12 months. It is equal to one when financial data for price, volumes and share outstanding are non-missing for the last 12 months.

[13] http://www2.standardandpoors.com/spf/pdf/index/SP_US_Indices_Methodology_Web.pdf

4) *Adequate **liquidity** and reasonable per-share **price** (the ratio of annual dollar value traded to market capitalization should be 0.3 or greater). Very low stock prices can affect a stock's liquidity.*

We apply directly the S&P definition of the liquidity: "*the ratio of annual dollar value traded to market capitalization*". Concretely, the denominator (the market value) of the liquidity ratio is computed as the average of the 12 last monthly market values in order to be comparable to the numerator, the annual dollar value traded, which is computed on the same twelve-month window:

$$Liq_{j,t} = \frac{1\sum_{k=t-11months}^{t} CRSP_M_PRC_{j,k} * CRSP_M_VOL_{j,k}}{\frac{1}{12} * \sum_{k=t-11months}^{t} CRSP_M_PRC_{j,k} * CRSP_M_SHROUT_{j,k}} \quad (2\text{-}23)$$

We then compute the dummy according to the threshold of 0.3 given by the S&P guidelines:

$$DLiq_{j,t} = 1_{Liq_{j,t} \geq 0.3}$$

Concerning the price, we create two dummies capturing the possible adverse liquidity impacts of extremely low or high prices. More generally, those dummies are also meant to capture any effects specific to those stocks. For example, penny stocks may present higher specific volatility than the average because some funds cannot trade them.

If the price is in the lowest (resp. highest) quintile of the price distribution, $Prc_Low_{j,t}$ (resp. $Prc_High_{j,t}$) is set to one, and otherwise it is set to 0:

$$Prc_Low_{j,t} = 1_{CRSP_M_PRC_{j,t} < q_{5\%}(CRSP_M_PRC)} \quad (2\text{-}24)$$

and $$Prc_High_{j,t} = 1_{CRSP_M_PRC_{j,t} > q_{95\%}(CRSP_M_PRC)} \quad (2\text{-}25)$$

5) *Financial Viability, usually measured as four consecutive quarters of positive as-reported earnings. **As-reported earnings** are GAAP Net Income excluding discontinued operations and extraordinary items.*

We construct a dummy indicating whether the firm experienced four consecutive quarters of positive, as reported earnings:

$$RpEarn_{j,t} = \prod_{k=t-11months}^{t} 1_{item69_{j,k} - item26_{j,k} \geq 0} \quad (2\text{-}26)$$

Note that *item69* (= Net Income (or Loss)) and *item26* (= Extraordinary Income and Discontinued Operations) come from the quarterly data of Compustat.

6) *Another measure of financial viability is a company's balance sheet* ***leverage****, which should be operationally justifiable in the context of both its industry peers and its business model.*

Company's balance sheet leverage $Lev_{j,t}$ is computed as described above. Because it must be considered relatively to industry peers, we correct it for industry effect by substracting the mean of the industry as explained above:

$$Lev_Ind_{j,t} = Lev_{j,t} - \overline{Lev}_{i(j,t),t} \quad (2\text{-}27)$$

7) ***Public float*** *of at least 50% of the stock.*

Public float is not easily accessible, and it is expected to be quite correlated with the $DLiq_{j,t}$ variable. We choose not to use it.

8) ***Operating company*** *and not a closed-end fund, holding company, partnership, investment vehicle or royalty trust. Real Estate Investment Trusts are eligible for inclusion in Standard & Poor's US indices.*

Our sample only comprises firms whose securities are ordinary common shares and Real Estate Investment Trusts (SHRCD # 10, 11, 12 and 18 from CRSP). Consequently, it excludes securities such as Certificates, ADRs, SBIs and Closed-end Funds. This ensures that the 9th condition on operating company and real estate investment is met.

2.4 Econometric Specification[14]

Our empirical model pursues a double objective. In a first step, we want to check whether there is any premium attached to the S&P 500 membership. In a second step, provided that the existence of a membership premium has been confirmed in the first step, we test most of hypotheses proposed by the literature by means of various and robust econometric approaches to ensure the reliability of our results. Since we are interested in the impact of the potential permanent deviation associated to the S&P membership, we need a setting different from a simple event study focusing on the event window. Many studies have provided evidence of abnormal returns around the inclusion date, but our focus is more general since we want to determine whether the fact to be a component of the S&P 500 results in a permanent higher price all other things being equal. We therefore closely follow the empirical setting of Morck and Yang (2002) who give evidence of such S&P 500 membership value and try to confirm their results as a first step. Another advantage of such a setting appears in the second step. Indeed, some of the variables needed to disentangle the various hypotheses are difficult to adapt to an event study framework and are more easily used in our panel regression framework.

2.4.1 Morck and Yang cross-sectional regression

We start with the basic equation of Morck and Yang (2002) (with adapted notations). It is a cross-sectional regression, which means that one regression is performed for each year. Therefore, for a given regression, the index *t* (denoting the year of the observation) is fixed and the index *j* (denoting the company) varies. There are thus in this setting as many observations as companies. Note that the specification will change in Equation (2-30) and following: in those future regressions, all observations will be pooled together and both the indices *t* and *j* will vary.

$$TQ_{j,t} = \beta_0 + \sum_{i=1}^{I} \gamma_{i,t} ind_{i,j,t} + \delta_t (SP_{j,t} \; or \; SPxWeight_{j,t})$$

[14] Table 2-4 provides the reader with a complete summary of the "translation" in our empirical framework of all competing hypotheses. Expected and actual effects on our regression coefficients obtained for each hypothesis are also indicated. Note that this table summarizes information from Sections 2.4 and 2.5. It follows that some elements of this table can only be fully understood after a reading of those latter sections.

$$+\beta_{1,t}RD_{j,t} + \beta_{2,t}Adv_{j,t} + \beta_{3,t}Lev_{j,t} + \beta_{4,t}Assets_{j,t} + u_{j,t} \quad (2\text{-}28)$$

Where:

- $ind_{i,j,t}$ is a dummy indicating whether firm *j* is in industry *i* in year *t*;
- all other variables have been defined in the previous section.

We generally use β for the coefficient of a control variable and δ for the coefficient of a variable of interest. Morck and Yang (2002) propose two variables of interest: the membership dummy $SP_{j,t}$ and the weight of the firm in the index $SPxWeight_{j,t}$. Since the S&P 500 index is value-weighted[15], this latter variable may indeed bring more information than the former. Note that Morck and Yang (2002) do not use these two variables in the same regression but successively in two separate regressions.

$TQ_{j,t}$ corresponds to the famous and widely used Tobin's Q, a measure of the valuation of the firm. Other valuation proxies exist and we propose to use three of them in alternative specifications, in order to check the robustness of our results: the PER (price earnings ratio) $PER_{j,t}$, the stock return $Ret_{j,t}$ (on a monthly basis) and the Market-to-Book ratio $MB_{j,t}$. In the following we will use a generic $q_{j,t}$ to denote the dependent variable of our main regression, knowing that this $q_{j,t}$ can stand alternatively for $PER_{j,t}$, $Ret_{j,t}$, $TQ_{j,t}$ or $MB_{j,t}$.

If *PER*, *TQ* and *MB* are common proxies of the value of the firm, the relevance of *Ret* may seem apriori slightly less straightforward. In fact, if the inclusion in the S&P 500 is actually valued by the market, the expected return is expected to decrease for those firms members of the S&P 500, what should in turn increase their price. Our variable Ret is supposed to capture the ex-post returns that are the consequence of this price increase.

[15] Notice that since 2005, the S&P 500 weights are not based on the capitalization of the firm anymore, but on its floating. Our period ranges from 1976 to 2005, so that our data are only marginally affected by this transition. Nevertheless, to ensure the robustness of our results, we performed the same study on the population without the observation from the year 2005. Results prove to be almost identical. They are available on request. Note that it would be interesting in a later study to investigate the impact of this transition (which is a purely information-free event) on the prices of the S&P 500 members.

Equation (2-28) contains some dummies $ind_{i,j,t}$ in order to control for possible industry specificities. While Morck and Yang (2002) categorize firms into industries according to the three first digits of their SIC code, we prefer to opt for the Fama and French industry classification[16], which is more parsimonious (49 categories for Fama and French VS 447 categories for three-digit SIC) and avoid to get many industries containing only one firm. Once the categories have been defined, industry effects can be controlled in two ways:

1) By the addition of one dummy for each industry (minus one, to avoid perfect collinearity problems);
2) By substracting from the dependent variable $q_{j,t}$ its industry mean computed without the observation. The details of this computation are given above in Section 2.3.2

According to Morck and Yang (2002), this second approach is econometrically preferable to simple fixed effects if some industries contain very few firms. Moreover, it allows us to keep the number of control variables tractable. However, we do not apply this transformation to the stock return $Ret_{j,t}$ since it makes less sense in this case. The final dependent variables are thus the following: $PER_Ind_{j,t}$, $Ret_{j,t}$ $TQ_Ind_{j,t}$ and $MB_Ind_{j,t}$ and will be designated under the generic $q_Ind_{j,t}$ variable in the subsequent equations in order to simplify the notations. Note that we also apply this transformation to the leverage $Lev_{j,t}$ of the probit, which becomes $Lev_Ind_{j,t}$ (reasons to do so have been developed earlier in Section 2.3.2).

Similarly to Morck and Yang (2002), we are now able to perform one cross-sectional regression for each time period (i.e., here, for each year, except for the monthly $Ret_{j,t}$) and to check that our results are consistent with theirs. We write this equation as follows:

$$q_Ind_{j,t} = \beta_{0,t} + \delta_t\left(SP_{j,t}\ or\ SPxWeight_{j,t}\right)$$
$$+\beta_{1,t}RD_{j,t} + \beta_{2,t}Adv_{j,t} + \beta_{3,t}Lev_{j,t} + \beta_{4,t}Assets_{j,t} + u_{j,t} \quad (2\text{-}29)$$

[16] Freely downloadable on Kenneth French website:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

2.4.2 Pooled OLS regression

Equation (2-29) was a preliminary specification aiming at partly replicating the results of Morck and Yang (2002) and confirming the existence of a membership premium increasing through time. Our real contribution starts with the next developments. Our purpose is to find explanations for the index effect and to discriminate between the different hypotheses proposed by the literature. We therefore bring the following adjustments to Equation (2-29).

First of all, we introduce simultaneously in the regression the two main variables of interest of Morck and Yang, $SP_{j,t}$ and $SPxWeight_{j,t}$. Indeed, these two variables do not exactly embed the same effects. If $SP_{j,t}$ captures the fact to be a member of the S&P 500, $SPxWeight_{j,t}$ gives the additional information of the weight of the firm in the S&P 500 on the top of that[17]. The concomitant use of $SP_{j,t}$ and $Weight_{j,t}$ may allow us to discriminate between the Information Content Hypothesis (and, at a lesser extent, the Liquidity Cost Hypothesis and Investor Recognition Hypothesis) on the one hand, and the Downward-Sloping Demand Curve Hypothesis on the other hand. Indeed, the premium anticipated by the Information Content Hypothesis should apply equally to all firms in the index, since the certification effect of ICHb or the better future performance consecutive to the inclusion supposed by ICHa are only linked to the (binary) fact to be in the index, whatever the weight. On the contrary, because it advocates a causal relation between the excess demand subsequent to the index funds rebalancing and the observed price increase, the Downward-Sloping Demand Curve Hypothesis expects a strong association between the weight in the index and the value premium. Liquidity Cost Hypothesis and Investor Recognition Hypothesis are some kind of intermediary cases. Strictly speaking, the simple fact to be a member of the S&P 500 should be enough to trigger the improved liquidity effect (Liquidity Cost Hypothesis) or the investor recognition (Investor Recognition Hypothesis), but some reinforcement of those effects due to the weight in the index is not incompatible with those hypotheses.

[17] As the deviation ε_j of our theoretical model is supposed to represent the excess demand faced by stock j, this variable $SPxWeight_{j,t}$ is the best proxy we can a priori imagine for this deviation ε_j, since the weight of a stock in the S&P 500 is directly linked to the excess demand coming from the index funds whose main goal is to replicate the index as close as possible. In an earlier version, we had chosen the abnormal volume on the inclusion day as another proxy of the deviation ε_j. Unluckily, this proxy is by construction only available for firms that have been included in the index after 1976, which could drastically reduce the number of relevant observations. For this reason, the coefficient was only marginally significant, even though it showed the expected positive sign.

Furthermore, the use of $SPxWeight_{j,t}$ alone in a regression is subject to some potential drawbacks. Indeed, such a variable potentially embeds three effects: the one of $SP_{j,t}$ alone, the one of $Weight_{j,t}$ alone and the one of the interaction between $SP_{j,t}$ and $Weight_{j,t}$. It makes the subsequent interpretation particularly delicate, as a significant coefficient of $SPxWeight_{j,t}$ could be potentially due to three different sources, as long as the variables $SP_{j,t}$ and $Weight_{j,t}$ are not introduced in the same regression to control for their individual contributions. That is why we introduce $SP_{j,t}$, $Weight_{j,t}$ and $SPxWeight_{j,t}$ altogether in our regression equation. However, if such a specification solves an issue, it creates potentially another problem, i.e. a collinearity issue. $Weight_{j,t}$ and $SPxWeight_{j,t}$ are indeed strongly correlated. Nevertheless, we know that by construction interaction variables are intrinsically dependent in some manner on their constituent variables, and we consider this correlation as logical. We will often perform regressions with and without the interaction variable as robustness checks in order to control whether the coefficients of the other variables remain stable whatever the specification. And indeed, it is what we usually observe.

As a second change with respect to the Mork and Yang (2002) specification, we introduce a supplementary dummy, $SPpast_{j,t}$, in order to directly test the asymmetry effect advocated by the Investor Recognition Hypothesis. According to the latter, the fact to have been once in the past in the index **but not today anymore** ($SPpast_{j,t}$) should matter at least as much as the fact to be currently in the index ($SP_{j,t}$). If Investor Recognition Hypothesis is true, the coefficient of $SPpast_{j,t}$should be positive, significant and of the same magnitude as the one of $SP_{j,t}$. Note that $SPpast_{j,t}$ depends on $SP_{j,t}$, since

$$SPpast_{j,t} = \left(1 - SP_{j,t}\right) * (Dummy =$$
$$1 \; if \; the \; firm \; has \; been \; at \; least \; once \; in \; the \; S\&P \; in \; the \; time \; range = [0; t-1]).$$

Thirdly, we add two short-term variables $SPadd12_{j,t}$ and $SPdel_{j,t}$ in order to disentangle short-term and long-term effects of an addition to (or deletion from) the index. These dummies are set to one if stock *j* was added to (resp. deleted from) the index for the 12 months previous to time *t*. If the impact of an inclusion (or exclusion) is entirely permanent, $SPadd12_{j,t}$ and $SPdel_{j,t}$ should not have any impact on the firm

valorization. On the contrary, if the impact is temporary (less than 12 months), they should capture the whole effect.

Fourthly, we introduce the one-year historical variance of the firm $Variance_{j,t}$ and an interaction variable between $SPxWeight_{j,t}$ and the latter. We name it $SPxWeightxVariance_{j,t}$. The adjunction of such a variable makes sense as Equation (1-31) exhibits an interaction product between the variance of the firm σ_j^2 and the firm j total deviation ε_j. However, the interpretation of this variable and of its coefficient is delicate. The model of Section 1 is a static two-period model expressing a link between the price of a firm and (among others) the variance of this price. On the contrary, the current empirical study deals with time-series data and employs the variance of returns rather than the variance of the prices. Therefore, the purpose of $Variance_{j,t}$ and $SPxWeightxVariance_{j,t}$ is more to account for possible volatility effects and to gain possible insights about how the idiosyncratic variance of a stock can impact the "excess value" of the firm than to really test a theoretical relationship that is difficult to directly transpose to the current empirical setting.

Fifthly, we add two supplementary control variables: the annual sales growth $SalesG_{j,t}$ and a dummy $NoIPO_{j,t}$ indicating whether the firm has at least 12 consecutive months of valid CRSP data. The first one $SalesG_{j,t}$ is added because it has proved to bring significant explanatory added value in our regressions and in other regressions of the same kind in the literature (see Campa and Kedia (2002) for example)[18]. The second one is needed because the computation of $Variance_{j,t}$ necessitates 12 months of returns. Rather than deleting all observations that do not meet this condition of 12 consecutive monthly returns, we prefer to introduce this dummy $NoIPO_{j,t}$ and to interact the variance with this dummy to compute $Variance_{j,t}$ and $SPxWeightxVariance_{j,t}$, so that the two latter are equal to 0 when $NoIPO_{j,t}$ is null. In the same logic, the sales growth is interacted with $NoIPO_{j,t}$ to deliver our $SalesG_{j,t}$ variable. The *NoIPO* variable captures all the problematic observations with missing data, while the coefficients of *Variance* and *SalesG* will be influenced only by observations with sufficient historic data.

[18] Note that we had initially introduced the dividend yield as well (for the same reasons as the sales growth), but as this variable has not proved to be significant in any of our regressions, we have decided to remove it.

Finally, since we are not especially interested in the evolution of the membership premium through time, we prefer to run one pooled regression rather than one cross-sectional regression for each year. However, by doing so, a new issue arises. According to Petersen (2008), in panel data sets, i.e. data sets with both time and cross-sectional dimensions, "the residuals may be correlated across firms or across time, and OLS standard errors can be biased"[19]. "One way empirical finance researchers can address two sources of correlation is to parametrically estimate one of the dimensions (e.g., by including dummy variables). Since many panel data sets have more firms than years, a common approach is to include dummy variables for each time period (to absorb the time effect) and then cluster by firm (Lamont and Polk, 2001; Anderson and Reeb, 2004; Gross and Souleles, 2004; Sapienza, 2004; and Faulkender and Petersen, 2006). If the time effect is fixed (…), the time dummies completely remove the correlation between observations in the same time period"[20]. That is why we add yearly dummies $year_{t,y}$ (monthly dummies when *Ret* is the dependent variable) to control for possible time effects on the firms valuation ratios (note that Campa and Kedia (2002) proceed exactly the same way). The firm effect will be treated explicitly later, in the fixed-effect panel estimation of Section 2.4.3.1. The use of this latter technique is justified, since "a significant minority of the papers in the survey" of Petersen (2008) "include firm dummies in their regressions. (…) Once the firm effects are included, the OLS standard errors are unbiased"[21] provided that the firm effect is constant over time.

The final pooled OLS regression is thus the following:

$$
\begin{aligned}
&q_Ind_{j,t} \\
&= \beta_0 + \beta_1 q_Ind_{j,t-1} + \delta_1 SP_{j,t} + \beta_2 Weight_{j,t} + \delta_2 SPxWeight_{j,t} + \delta_3 SPpast_{j,t} \\
&+ \delta_4 SPadd12_{j,t} + \delta_5 SPdel12_{j,t} + \beta_3 RD_{j,t} + \beta_4 Adv_{j,t} + \beta_5 Lev_{j,t} + \beta_6 Assets_{j,t} \\
&+ \beta_7 SalesG_{j,t} + \beta_8 NoIPO_{j,t} + \beta_9 Variance_{j,t} + \delta_6 SPxWeightxVariance_{j,t} \\
&+ \sum_{y=1}^{Y} \theta_y year_{t,y} + u_{j,t}
\end{aligned}
\tag{2-30}
$$

[19] Petersen (2008), p. 1

[20] Ibidem, p. 24

[21] Ibidem, p.30

2.4.3 Endogeneity and self-selection

However, such an econometrical specification potentially hides a serious flaw. The decision of being included in the index is likely to be strongly endogenous. We must consider the hypothesis that the firms included in the index are not a random sample of firms. If the decision to introduce or maintain a company in the index is correlated with the relative value of the firm, $SP_{j,t}$ will be correlated with the error term $u_{t,j}$ in the above equation. The OLS estimate of the coefficient of $SP_{j,t}$ will, therefore, be biased. As a consequence, the current specification is no more valid and needs to be adapted. We assume now that the inclusion decision is determined by:

$$SP^*_{j,t} = \psi Z_{j,t} + \mu_{j,t} \tag{2-31}$$

$SP_{j,t} = 1$ if $SP^*_{j,t} > 0$

$SP_{j,t} = 0$ if $SP^*_{j,t} < 0$,

where $SP^*_{j,t}$ is an unobserved latent variable, $Z_{j,t}$ is a set of firm characteristics that affect the decision of inclusion, and $\mu_{j,t}$ is an error term. The correlation between $SP_{j,t}$ and $u_{j,t}$ in Equation (2-28) (or (2-29)) will arise when

1) some of the exogenous variables in the inclusion Equation (2-31), $Z_{j,t}$, affect the firm's relative value $\frac{V_{j,t}}{A_{j,t}}$ but are not included as regressors in the value Equation (2-30);
2) or the errors $u_{j,t}$ and $\mu_{j,t}$ are correlated.

In both cases, the OLS coefficients will be biased.

In analogy with Campa and Kedia (2002) who apply the same procedure to study the diversification discount, we use three different techniques to control for the correlation between $SP_{j,t}$ and $u_{j,t}$ in Equation (2-30) and come up with an unbiased estimator of δ_1.

2.4.3.1 Fixed-effect panel estimation

First, we take advantage of the panel structure of our data set and use a two-way[22] fixed-effect estimator in Equation (2-30), assuming that all the unobserved heterogeneity that leads to the correlation between the error terms is constant over time. Indeed, if our endogeneity problem is due to unobservable variables that drive both the dependent variable (*TQ*, for example) and the endogenous variable *SP*, and if those unobservable variables are constant through time (but varying in the cross-sectional dimension), a fixed effect panel regression will account for it and solve this particular endogeneity issue. Since the number of firms in the sample is very large (around 11,500), the current statistical software cannot manage a direct two-way fixed-effect panel estimation because of the excessive number of dummies needed for both the time dimension (29 if yearly dummies, 351 if monthly dummies) and the cross-sectional dimension (11,500). Hopefully, alternative techniques exist. We follow Greene (1997), p. 616, who proposes to subtract the time-series mean of each observation in order to get rid of the cross-sectional fixed-effect and to perform an ordinary pooled OLS regression on the resulting demeaned variables. Coefficients are consistent, but a slight correction of the standard errors is needed to take into account the actual number of degrees of freedom implicit to the two-way fixed-effect estimation.

2.4.3.2 Instrumental variables estimation

Second, we try to find adequate and relevant instruments for the membership dummy $SP_{j,t}$, in order to first regress the latter on that set of instruments in a first probit step and then run an OLS on Equation (2-30) in a second step where we replace the membership dummy with its fit of the first step, $SP_F_{j,t}$:

$$
\begin{aligned}
&q_Ind_{j,t} \\
&= \beta_0 + \beta_1 q_Ind_{j,t-1} + \delta_1 SP_F_{j,t} + \beta_2 Weight_{j,t} + \delta_2 SPxWeight_F_{j,t} \\
&+ \delta_3 SPpast_F_{j,t} + \delta_4 SPadd12_F_{j,t} + \delta_5 SPdel12_F_{j,t} + \beta_3 RD_{j,t} + \beta_4 Adv_{j,t} \\
&+ \beta_5 Lev_{j,t} + \beta_6 Assets_{j,t} + \beta_7 SalesG_{j,t} + \beta_8 NoIPO_{j,t} + \beta_9 Variance_{j,t} \\
&+ \delta_6 SPxWeightxVariance_F_{j,t}
\end{aligned}
$$

[22] If the specification is dependent only on the cross section to which the observation belongs, such a model is referred to as a model with one-way effects. A specification that depends on both the cross section and the time series to which the observation belongs is called a model with two-way effects.

$$+\sum_{y=1}^{Y}\theta_y year_{t,y} + u_{j,t} \quad (2\text{-}32)$$

Note that $SPxWeight_F_{j,t}$, $SPpast_F_{j,t}$, $SPadd12_F_{j,t}$, $SPdel12_F_{j,t}$ and $SPxWeightxVariance_F_{j,t}$ are concerned as well by the fitting because $SP_{j,t}$ enters their construction.

The challenge is to find good instruments, that is, variables uncorrelated with the error $u_{j,t}$ in Equation (2-32) but correlated with the inclusion dummy $SP_{j,t}$. The S&P 500 committee identifies on S&P's website several criteria for a firm to be introduced and maintained in the index, which can be useful in order to determine relevant instruments for the dummy. Variables issued from this set of criteria have been detailed in the previous Section 2.3.2. They lead to the following pooled probit equation:

$$SP_{j,t} = \psi_0 + \psi_1 MktCaplog_{j,t} + \psi_2 NoIPO_{j,t_3} + \psi_3 DLiq_{j,t} + \psi_{4,low} Prc_Low_{j,t}$$
$$+\psi_{4,high} Prc_High_{j,t} + \psi_5 RpEarn_{j,t} + \psi_6 Lev_Ind_{j,t} + \mu_{j,t} \quad (2\text{-}33)$$

Contrary to regressions (2-29), (2-30) and (2-32) where variables were annual (except when *Ret* is the dependent variable), the variables used in the pooled probit Equation (2-33) are monthly. $SP_F_{j,t}$ is estimated on a monthly basis and then filtered in order to keep only the December observation of each year before being plugged in Equation (2-32).

Since it is a two-step procedure, the OLS standard errors of the second step are not correct. In order to assess the significance of our coefficients, we opt for a non-parametric bootstrap approach. Basic bootstrap and bootstrap-t were both applied, and they give identical results.

2.4.3.3 Self-selection estimation

The third method to control for endogeneity is to control for the self-selection of firms that are included using Heckman (1979)'s two-stage procedure. For a comprehensive review of the self-selection issue, we refer the interested reader to Li and Prabhala (2005).

For convenient reasons and without any loss of generality, we consider a simplified version of Equation (2-30) where we have removed the lagged dependent variable, all year dummies and variables of interest except $SP_{j,t}$:

$$q_Ind_{j,t} = \beta_0 + \delta SP_{j,t} + \beta_1 RD_{j,t} + \beta_2 Adv_{j,t} + \beta_3 Lev_{j,t} + \beta_4 Assets_{j,t}$$
$$+\beta_5 SalesG_{j,t} + \beta_6 NoIPO_{j,t} + \beta_7 Variance_{j,t} + u_{j,t} \quad (2\text{-}34)$$

We estimate expected firm value conditional on the firm being included as $E\left(q_Ind_{j,t}\middle|SP_{j,t}=1\right) = \beta X_{j,t} + \delta + E(u_{j,t}|SP_{j,t}=1)$, where $X_{j,t}$ stands for the constant and the control variables of the regression. Assuming that the errors in Equations (2-34) and (2-31), $u_{j,t}$ and $\mu_{j,t}$, follow a bivariate normal distribution $N_2\left(\binom{0}{0}, \begin{pmatrix}\sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & 1\end{pmatrix}\right)$, we have

$E\left(u_{j,t}\middle|SP_{j,t}=1\right) = \rho\sigma_u\lambda_1(\psi Z_{j,t})$, where $\lambda_1\left(\psi Z_{j,t}\right) = \frac{\varphi(\psi Z_{j,t})}{\Phi(\psi Z_{j,t})}$,

and $\varphi(.)$ and $\Phi(.)$ are, respectively, the density and cumulative distribution functions of the standard normal. Similarly, the expected value conditional on the non-member firm is $E\left(q_Ind_{j,t}\middle|SP_{j,t}=0\right) = \beta X_{j,t} + E(u_{j,t}|SP_{j,t}=0)$. In this case,

$E\left(u_{j,t}\middle|SP_{j,t}=0\right) = \rho\sigma_u\lambda_2(\psi Z_{j,t})$ where $\lambda_2\left(\psi Z_{j,t}\right) = \frac{-\varphi(\psi Z_{j,t})}{1-\Phi(\psi Z_{j,t})}$.

The difference in the value of member and non-member firms is given by

$$E\left(q_Ind_{j,t}\middle|SP_{j,t}=1\right) - E\left(q_Ind_{j,t}\middle|SP_{j,t}=0\right) = \delta + \rho\sigma_u \frac{\varphi(\psi Z_{j,t})}{\Phi\left(\psi Z_{j,t}\right)\left(1-\Phi\left(\psi Z_{j,t}\right)\right)} \quad (2\text{-}35)$$

The right-hand side of the above equation is what is estimated by the OLS coefficient of $SP_{j,t}$ in (2-30). This estimated coefficient, using OLS, will, therefore, be biased downward (upward) if the correlation of the error terms, ρ, is negative (positive).

Specifically, and in line with Heckman's two-step procedure, we first estimate Equation (2-33) using a probit to get consistent estimates of ψ denoted by $\hat{\psi}$. These are then used to get estimates of λ_1 and λ_2, the correction for self-selection. In the second step, we estimate δ by estimating

$$q_Ind_{j,t} = \beta X_{j,t} + \delta SP_{j,t} + \delta_\lambda\left[\lambda_1\left(\hat{\psi}Z_{j,t}\right) * SP_{j,t} + \lambda_2\left(\hat{\psi}Z_{j,t}\right) * (1 - SP_{j,t})\right]SP_{j,t} + \eta_{j,t}$$

$$= \beta X_{j,t} + \delta SP_{j,t} + \delta_\lambda \lambda + \eta_{j,t} \quad (2\text{-}36)$$

where $\delta_\lambda = \rho\sigma_u$. The sign of δ_λ is thus entirely determined by the sign of ρ, the correlation between the two error terms.

If we explicit all our variables, this regression becomes:

$$\begin{aligned} & q_Ind_{j,t} \\ & = \beta_0 + \beta_1 q_Ind_{j,t-1} + \delta_1 SP_{j,t} + \beta_2 Weight_{j,t} + \delta_2 SPxWeight_{j,t} + \delta_3 SPpast_{j,t} \\ & + \delta_4 SPadd12_{j,t} + \delta_5 SPdel12_{j,t} + \beta_3 RD_{j,t} + \beta_4 Adv_{j,t} + \beta_5 Lev_{j,t} + \beta_6 Assets_{j,t} \\ & + \beta_7 SalesG_{j,t} + \beta_8 NoIPO_{j,t} + \beta_9 Variance_{j,t} + \delta_6 SPxWeightxVariance_{j,t} \\ & + \delta_\lambda Lambda + \sum_{y=1}^{Y} \theta_y year_{t,y} + u_{j,t} \end{aligned} \quad (2\text{-}37)$$

Like for the instrumental variables regression, because we cannot use the OLS standard errors, we apply a bootstrap procedure in order to correctly assess the significance of our coefficients.

Table 2-4: Summary of the underlying empirical implications of the competing hypotheses

PPH stands for Price Pressure Hypothesis, DSH for Downward-Sloping Demand Curve Hypothesis, LCH for Liquidity Cost Hypothesis, ICH for Information Content Hypothesis (ICHa: performance causes inclusion; ICHb: inclusion causes performance) and IRH for Investor Recognition Hypothesis. Lambda is the correction for self-selection variable.

				PPH	DSH	LCH	ICH		IRH	Confirmed?
							ICHa	ICHb		
EMPIRICS	Inclusion case:	Price increase is	temporary	X						
			permanent		X	X	X		X	
		Volume increase is	temporary	X	X		X		X	
			permanent			X				
	Deletion case:	Price effect is	symetric to inclusion: price drops	X	X	X	X			
			asymetric to inclusion: no price change						X	
	In our empirical framework:	We expect the coeff of	SP to be > 0 and significant		X	X	X		X	YES
			SPxWeight to be > 0 and significant		X					NO
			SPpast to be > 0 and significant						X	Rather NOT
			SPadd12 to be > 0 and significant	X						YES
			SPdel12 to be < 0 and significant	X						NO
			Lambda to be > 0 and significant					X		NO
			SPxWeightxVar to be > 0 and significant		X					Mixed
	On the basis of our results, global evidence in favor of this hypothesis is			Mixed	Poor	Still open	Still open	Inexistent	Mixed	

2.5 Results

2.5.1 Morck and Yang cross-sectional regression

Table 2-5: Cross-Sectional OLS Regression of Tobin's Q on the S&P 500 Membership Dummy

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	TQ_Ind											
Main X	SP											
Year	1978				1988				1998			
TYPE	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF
Intercept	0.160	3.675	***		0.317	3.407	***		0.885	4.118	***	
SP	0.082	3.286	***	1.608	0.188	3.736	***	1.759	0.596	7.465	***	1.866
SPxWeight												
RD	1.635	3.772	***	1.063	1.195	2.477	**	1.109	2.576	3.956	***	1.128
Adv	0.062	0.212		1.048	1.117	2.010	**	1.020	2.948	2.643	***	1.021
Lev	-0.400	-7.242	***	1.102	-0.648	-6.977	***	1.120	-0.830	-5.636	***	1.146
Assets	-0.016	-1.993	**	1.551	-0.036	-2.381	**	1.699	-0.116	-3.829	***	1.757
N	1,802				1,743				1,509			
AdjR2	0.056				0.058				0.106			
F	22.467		***		22.339		***		36.872		***	

Table 2-6: Cross-Sectional OLS Regression of Tobin's Q On S&P 500 Index Weight

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	TQ_Ind											
Main X	SPxWeight											
Year	1978				1988				1998			
TYPE	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF
Intercept	0.144	3.487	***		0.216	2.412	**		0.945	4.801	***	
SP												
SPxWeight	11.049	3.594	***	1.202	18.872	1.889	*	1.402	150.206	10.622	***	1.413
RD	1.631	3.766	***	1.062	1.338	2.775	***	1.102	2.348	3.677	***	1.125
Adv	0.194	0.676		1.024	1.264	2.273	**	1.015	2.275	2.069	**	1.027
Lev	-0.412	-7.541	***	1.081	-0.687	-7.438	***	1.101	-0.925	-6.488	***	1.113
Assets	-0.010	-1.464		1.190	-0.013	-0.971		1.377	-0.108	-4.148	***	1.350
N	1,802				1,743				1,509			
AdjR2	0.057				0.052				0.138			
F	22.915		***		20.145		***		49.236		***	

Table 2-5 and Table 2-6 partially and respectively reproduce Tables 7 and 8 of Morck and Yang (2002). They are the direct application of Equation (2-29), i.e. one cross-

sectional regression for each year of the sample. Since Morck and Yang (2002) presented results for 1978, 1988 and 1998, for the reason they judged them to be representative, we do the same for the sake of comparison. Our results are qualitatively similar to theirs: most of coefficients have same sign and show identical evolution over time. Differences only appear in control variables coefficients: *Adv* and *Assets* are sometimes less significant in our setting than in the Morck and Yang setting, and *Lev* has the opposite sign. These slight contradictions are most probably explained by the differences in the variables definition: in particular, the way they construct *Lev* and *TQ* is a little different from ours. We prefer to stick to the common standard definition found in the literature for those variables in order to keep the methodology standard and easily reproducible.

Those differences do not matter, as long as our point was to reproduce their main results and to reach conclusions identical to them: there exists a premium for the index membership and this premium increases with time as shown in Table 2-7 and Figure 2-3, with a decrease over the latter years. Morck and Yang (2002) suggest that the membership premium evolution and the growth of indexing are in fact strongly associated: the former increases along with the latter over almost the whole period, and declines over the last years as the indexing decreases slightly as well.

We systematically give the Variance Inflation Factor (VIF) of each variable in order to check the collinearity among variables, which turns out to be a potential issue in some of our regressions. According to Belsley (1991), "no means have yet been found for determining a meaningful boundary between values for the VIF that can be considered high and those that can be considered low. This issue is clearly related to that of determining when a correlation is close enough to unity to denote collinearity. Cutoffs of 7 or 10 have been suggested for the VIF, but these suggestions have a complete ad hoc basis"[23]. Other collinearity checks have been performed, like the computation of the condition index number and the proportion of explained variance associated to each variable. According to Belsley (1991), "weak dependencies are associated with condition indexes around 5-10, whereas moderate to strong relations are associated with condition indexes of 30-100"[24]. If a condition index number is high and two variables explain a significant variance proportion of it (typically, more than 50%), these two

[23] Belsley (1991), p. 28.

[24] Ibidem, p. 28.

variables are likely to present collinearity problems. VIF and condition index number are complementary diagnostics and usually lead to the same conclusions. Therefore, we only report VIF diagnostics in order to save place, but the other collinearity tests are naturally available on request.

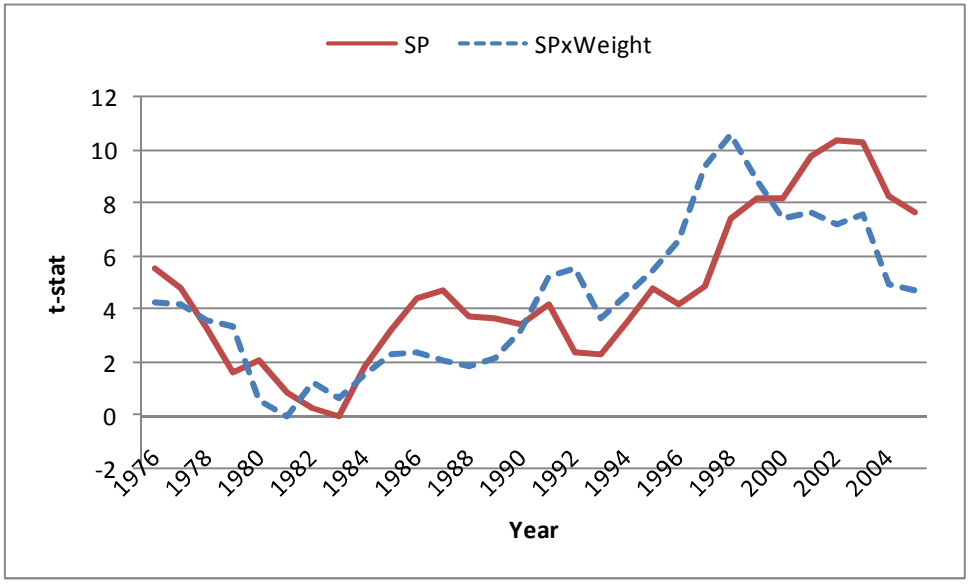
In the case of the cross-sectional regressions of Morck and Yang (2002), the VIF are very satisfactory and do not raise any concern.

Table 2-7: How Regression Coefficients on the S&P Dummy (and Weight) Changed over Time

Slope is the coefficient of the variable and *T-stat* is the student statistic. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%).

DepVar	TQ_Ind					
Main X	SP			SPxWeight		
Year	Slope	t-stat		Slope	t-stat	
1976	0.176	5.536	***	17.365	4.285	***
1977	0.131	4.830	***	14.023	4.186	***
1978	0.082	3.286	***	11.049	3.594	***
1979	0.040	1.620		11.983	3.365	***
1980	0.066	2.057	**	3.131	0.642	
1981	0.045	0.867		0.064	0.008	
1982	0.009	0.259		6.155	1.288	
1983	-0.001	-0.028		4.916	0.701	
1984	0.084	1.848	*	11.079	1.580	
1985	0.125	3.231	***	15.042	2.358	**
1986	0.205	4.407	***	22.120	2.422	**
1987	0.256	4.763	***	21.983	2.117	**
1988	0.188	3.736	***	18.872	1.889	*
1989	0.165	3.643	***	20.740	2.196	**
1990	0.178	3.453	***	30.929	3.214	***
1991	0.246	4.220	***	56.611	5.272	***
1992	0.171	2.364	**	79.497	5.575	***
1993	0.163	2.315	**	55.374	3.705	***
1994	0.214	3.561	***	58.658	4.635	***
1995	0.282	4.810	***	66.056	5.473	***
1996	0.294	4.235	***	90.663	6.644	***
1997	0.373	4.890	***	134.514	9.397	***
1998	0.596	7.465	***	150.206	10.622	***
1999	0.734	8.222	***	133.989	8.903	***
2000	0.847	8.216	***	131.369	7.481	***
2001	0.870	9.788	***	110.976	7.721	***
2002	0.804	10.404	***	93.934	7.270	***
2003	0.660	10.310	***	86.043	7.627	***
2004	0.625	8.301	***	69.926	4.980	***
2005	0.580	7.655	***	69.811	4.724	***

Figure 2-3: Evolution of the t-stats of the Membership Dummy and the Index Weight over Time



2.5.2 Pooled OLS regression

Table 2-8: Pooled OLS regression

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	PER_Ind				Ret				TQ_Ind				MB_Ind			
TYPE	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF
Intercept	3.221	2.105	**		0.019	6.772	***		0.492	13.835	***		1.299	11.385	***	
y_lag	0.023	17.499	***	1.003	-0.035	-28.294	***	1.213	0.483	182.219	***	1.093	0.058	51.727	***	1.051
SP	2.318	5.857	***	2.440	0.006	11.262	***	2.367	0.265	27.818	***	2.402	0.981	33.184	***	2.400
Weight	1,108.337	3.730	***	**45.100**	10.389	26.602	***	**40.529**	240.994	32.144	***	**40.886**	750.931	32.629	***	**41.551**
SPxWeight	-1,079.969	-3.709	***	**43.439**	-9.605	-25.051	***	**38.988**	-209.040	-28.431	***	**39.243**	-633.447	-28.039	***	**39.978**
SPpast	-0.133	-0.115		1.231	0.001	0.557		1.255	-0.046	-1.959	*	1.255	-0.273	-3.538	***	1.266
SPadd12	3.580	3.201	***	1.050	0.004	2.914	***	1.056	0.391	13.546	***	1.044	0.836	9.438	***	1.045
SPdel12	5.802	2.108	**	1.208	-0.007	-2.350	**	1.229	0.100	1.843	*	1.226	0.632	3.663	***	1.241
RD	28.053	7.168	***	1.168	-0.011	-2.505	**	1.158	0.958	11.855	***	1.156	3.103	12.350	***	1.164
Adv	3.921	0.962		1.047	0.013	2.586	***	1.044	0.426	4.268	***	1.044	3.699	11.906	***	1.045
Lev	6.910	8.970	***	1.240	-0.006	-6.835	***	1.214	-0.351	-21.009	***	1.225	0.640	11.884	***	1.223
Assets	-1.039	-7.537	***	3.641	-0.002	-13.878	***	3.438	-0.098	-29.927	***	3.524	-0.307	-30.320	***	3.536
SalesG	0.054	0.119		1.082	-0.005	-11.336	***	1.061	0.119	12.887	***	1.077	0.663	23.217	***	1.057
NoIPO	-1.529	-1.155		1.039	0.000	-0.149		1.055	-0.011	-0.382		1.035	-0.383	-3.929	***	1.035
Variance	98.915	9.081	***	1.280	0.159	17.096	***	1.253	-2.162	-12.658	***	1.235	-8.547	-15.191	***	1.247
SPxWeightxVar	-50.990	-2.160	**	1.023	-0.025	-1.005		1.033	0.651	1.047		1.021	-1.856	-0.969		1.023
N	36,507				556,694				46,134				45,107			
AdjR2	0.017				0.178				0.502				0.135			
F	15.760		***		342.484		***		1083.766		***		165.249		***	

As explained in Section 2.4.3, we already know that a mere OLS regression is unlikely to deliver consistent results in case of endogeneity. However, we choose to present the results of such a specification in order to use it as a benchmark for the subsequent

specifications (fixed-effect panel study, instrumental variables and self-selection treatment) that will explicitly tackle the endogeneity issue. Moreover, an OLS regression may already give us some first insight about the relations between the variables, even though we must be cautious when we are to interpret the results.

Application of Equation (2-30) delivers Table 2-8. We pool all observations together and do not use the panel structure of data (contrary to the later panel fixed-effect regression). We consider four dependent variables *PER_Ind* (that we will denote hereafter simply by *PER* for notational convenience), *Ret*, *TQ_Ind* (denoted hereafter by *TQ*) and *MB_Ind* (denoted herafter by *MB*) and six variables of interest: *SP*, *SPxWeight*, *SPpast, SPadd12, SPdel12* and *SPxWeightxVar*. Results obtained for *TQ* and *MB* are very similar. In the case of *MB*, all coefficients of the Morck and Yang variables have the expected sign (generally positive, except *Assets*). The same holds true for *TQ*, with the exception of *Lev* that we have already mentioned above. The coefficients obtained through the regressions on *PER* generally exhibit the same sign as the sign observed for *TQ* and *MB*, and it is conform to the intuition, since *PER* is a proxy of firm value like *TQ* and *MB*. We note that the coefficients of the *PER* regression are less significant for some variables (*SPpast*, *Adv* and *SalesG*) and that the coefficient of *Variance* is negative, while we would have rather expected a positive sign, like in the *TQ* and *MB* regressions. Finally, *Ret* usually shows similar coefficients as well (as expected), with the noticeable exceptions of *SPpast*, *RD* and *SalesG*.

The positive and significant coefficient of *SP* in the four regressions (*PER*, *Ret*, *TQ* and *MB*) seems to confirm the idea that being a member of the S&P 500 index affects positively the value of the firm. This impact is unlikely to be strictly limited to the inclusion period, since the variable *SPadd12* controls for this possibility. However, the coefficient of this latter variable is positive and significant, and in the case of the *PER* regression even larger than the one of *SP*, at least in this pooled OLS regression. Therefore, short-term and long-term index effects seem to coexist. It follows that the Price Pressure Hypothesis may still constitute a substantial part of the explanation of the index effect, at least with respect to the short-term effect (over one year). The additional long-term effect we detect, however, requires another explanation.

The negative and significant coefficient of *SPxWeight* is puzzling, and seems at odds with the Downward-Sloping Demand Curve Hypothesis. As a reminder, this hypothesis posits that the higher the weight of a company in the S&P 500 (*SPxWeight*), and the

higher its valuation ratio (here proxied by *PER*, *TQ* and *MB*). The reason would come from the excess demand created by the index funds that try to replicate the index as closely as possible. The negative sign of *SPxWeight* tends to reject such an explanation. In fact, it is the *Weight* of a company, irrespective of its membership to the index, which seems to matter: the higher the ratio of the company market capitalization to the S&P 500 market capitalization (i.e. *Weight*), and the higher our valuation proxy of the firm (*PER*, *Ret*, *TQ* and *MB*), all things being equal. It is consistent with the size effect of Banz (1981) that predicts a lower risk premium for large capitalizations. But the negative sign of *SPxWeight* teaches us that the fact to be in the index decreases this positive size effect on the valuation: companies member of the S&P 500 benefit less from this *Weight* effect (i.e. the "size" effect) than an equivalent company of the same size not member of the index.

This result casts some doubts on the conclusion of Morck and Yang (2002), who inserted the variable *SPxWeight* alone in their regression (without *SP* and *Weight*) and deduced from its positive and significant sign (see Table 2-6) that the large demand of indexed funds was the cause of the index effect. In fact, our results show that the positive coefficient of *SPxWeight* alone is not due to the interaction effect *SPxWeight* itself, but to the individual effects of *SP* on the one hand and *Weight* on the other hand that are embedded in *SPxWeight* when it is used alone. Consequently, the reason of the higher valuation of the companies heavily weighted in the S&P 500 (= high *SPxWeight*) is not their large weight in the S&P 500, but rather the fact that they are in the S&P 500 (*SP*) and their ratio of the market capitalization to the one of the S&P 500 (*Weight*), whatever they are included or not in the index.

Whereas the Investor Recognition Hypothesis expects the coefficient of *SPpast* to be significantly positive, we do not observe it, and even the contrary in some cases. Companies that were in the S&P 500 in the past but are not members anymore do not seem to present higher valuation ratios than non-member companies. It is too early to already reject the Investor Recognition Hypothesis, but this result at least suggests that the Investor Recognition Hypothesis is not the principal cause of the index effect, even if it may play some role that we do not detect here.

The coefficient of *SPdel12* has to be interpreted cautiously. A quick look at the descriptive statistics of Table 2-3 reveals that this dummy is equal to one for only 0.34% of all observations, while SPadd12 is equal to one in 0.93% of the cases. At first

sight, this can seem surprising, because the number of additions should be obviously equal to the number of deletions. But we know that a deletion can occur for many different reasons, including merge, acquisition, liquidation, etc. All those firms eliminated from the market just a few days or weeks after the deletion from the index result in missing financial data, and as a consequence they cannot be used in the regression. This "survival" bias and the low number of observations for this dummy may alter the results. In our pooled OLS regressions, we find a positive and relatively significant coefficient. This is not evidence in favor of the Price Pressure Hypothesis, which expected a significantly negative coefficient. We will discuss this issue in more details later, with the additional information of the subsequent regressions.

SPxWeightxVar does not show any positive and significant coefficient and thus does not confirm the prediction of our theoretical final asset pricing Equation (1-31). Next regressions will bring additional information in this respect.

Table 2-9: Pooled OLS regression without *SPxWeight*

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	PER_Ind				Ret				TQ_Ind				MB_Ind			
TYPE	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF
Intercept	2.506	1.651	*		0.013	4.644	***		0.356	10.024	***		0.898	7.866	***	
y_lag	0.023	17.502	***	1.003	-0.034	-27.498	***	1.212	0.491	184.676	***	1.081	0.059	52.210	***	1.050
SP	1.605	4.639	***	1.865	0.000	-0.772		1.829	0.135	16.014	***	1.848	0.585	22.328	***	1.852
Weight	23.867	0.454		1.415	0.775	10.705	***	1.391	31.521	22.530	***	1.399	116.533	27.430	***	1.392
SPpast	-0.168	-0.145		1.231	0.000	0.171		1.254	-0.057	-2.371	**	1.255	-0.300	-3.866	***	1.266
SPadd12	3.768	3.372	***	1.048	0.006	3.821	***	1.054	0.420	14.447	***	1.043	0.936	10.483	***	1.044
SPdel12	5.955	2.163	**	1.208	-0.007	-2.288	**	1.229	0.107	1.952	*	1.226	0.654	3.754	***	1.241
RD	28.966	7.414	***	1.163	-0.003	-0.752		1.152	1.115	13.708	***	1.150	3.572	14.128	***	1.159
Adv	3.967	0.973		1.047	0.014	2.731	***	1.044	0.430	4.267	***	1.044	3.724	11.883	***	1.045
Lev	6.444	8.477	***	1.206	-0.009	-10.888	***	1.184	-0.418	-25.036	***	1.201	0.397	7.397	***	1.192
Assets	-0.814	-6.575	***	2.936	0.000	-3.183	***	2.773	-0.057	-19.251	***	2.845	-0.183	-19.897	***	2.858
SalesG	0.124	0.275		1.080	-0.005	-10.192	***	1.059	0.126	13.576	***	1.076	0.699	24.286	***	1.055
NoIPO	-1.551	-1.172		1.039	0.000	-0.310		1.055	-0.010	-0.335		1.035	-0.386	-3.922	***	1.035
Variance	98.912	9.079	***	1.280	0.156	16.754	***	1.253	-2.263	-13.137	***	1.235	-8.852	-15.601	***	1.247
SPxWeightxVar	-50.101	-2.123	**	1.023	-0.016	-0.671		1.033	0.800	1.275		1.021	-1.371	-0.710		1.022
N	36,507				556,694				46,134				45,107			
AdjR2	0.017				0.177				0.494				0.120			
F	15.802		***		341.292		***		1071.556		***		147.888		***	

Note that the coefficient of *NoIPO* is negative but not always significant, which may seem to indicate that newly quoted stocks benefit from a slight value premium in the first months of their introduction in the stock market.

SalesG is a highly significant control variable (except in the *PER* regression) and thus deserves its introduction in the equation.

Finally, it is worth observing that the coefficient of *Variance* is negative and significant in the case of *TQ* and *MB*. It is consistent with our theoretical Equation (1-31), where the sign of the σ_j^2 term is negative. However, the positive sign of this variable in the *PER* regression is difficult to explain.

As anticipated, the introduction of *Weight* and *SPxWeight* altogether as explicative variables causes a severe collinearity issue, as revealed by the VIF around 40 for these variables. Therefore Table 2-9 presents the same regression without *SPxWeight* in order to check the stability of the coefficients. All VIF present now reasonable levels and all coefficients are stable in comparison with the ones of Table 2-8, with the noticeable exception of the coefficient of *SP* in the *Ret* regression, which is now negative and insignificant. Besides this fact, the collinearity between *Weight* and *SPxWeight* does not seem to affect significantly the stability and thus the way we can interpret our results.

2.5.3 Fixed-effect panel estimation

Table 2-10: Fixed-effect panel estimation

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor and *N* is the number of observations of the regression. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	PER_Ind_d				Ret				TQ_Ind				MB_Ind			
TYPE	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF	Slope	t-stat		VIF
y_lag	0.008	6.290	***	1.002	-0.057	-45.483	***	1.224	0.308	109.542	***	1.064	0.030	33.752	***	1.049
SP	0.898	1.223		1.955	0.011	10.736	***	2.039	0.338	20.528	***	2.063	0.961	20.488	***	2.041
Weight	681.072	1.428		**33.584**	25.877	36.475	***	**31.721**	341.759	31.619	***	**30.848**	934.900	30.407	***	**32.227**
SPxWeight	-493.742	-1.046		**33.743**	-23.325	-33.165	***	**31.829**	-292.390	-27.239	***	**30.973**	-791.851	-25.966	***	**32.381**
SPpast	-1.284	-0.886		1.512	-0.003	-1.921	*	1.635	-0.062	-2.226	**	1.658	-0.267	-3.262	***	1.646
SPadd12	2.020	1.930	*	1.103	0.002	1.233		1.098	0.222	9.092	***	1.089	0.408	5.921	***	1.091
SPdel12	5.788	2.308	**	1.219	-0.007	-2.250	**	1.223	0.135	3.037	***	1.220	0.611	4.678	***	1.235
RD	65.295	5.876	***	1.025	0.030	2.528	**	1.017	0.275	1.477		1.019	4.403	8.284	***	1.019
Adv	5.927	0.643		1.050	0.010	0.811		1.044	-0.331	-1.664		1.044	-0.385	-0.669		1.046
Lev	16.687	13.243	***	1.091	0.011	7.036	***	1.070	-0.392	-16.120	***	1.070	1.772	24.153	***	1.080
Assets	-1.203	-3.962	***	3.973	-0.017	-41.226	***	3.355	-0.238	-38.100	***	3.462	-0.743	-40.904	***	3.588
SalesG	-2.821	-5.846	***	1.061	-0.005	-8.646	***	1.071	0.182	20.566	***	1.080	0.609	24.063	***	1.062
NoIPO	0.727	0.494		1.054	-0.008	-6.420	***	1.080	-0.092	-3.148	***	1.044	-0.448	-5.014	***	1.045
Variance	30.473	2.426	**	1.205	0.426	37.288	***	1.180	-0.847	-4.858	***	1.150	-4.743	-8.870	***	1.161
SPxWeightxVar	-53.579	-2.426	**	1.031	-0.019	-0.732		1.034	0.754	1.436		1.027	1.370	0.908		1.029
N		35,462				556,611				45,038				44,016		

Table 2-10 (based on a fixed-effect panel estimation of Equation (2-30)) tends to confirm results of Table 2-8. Some coefficients lose significance, like *RD* and *Adv*, and important variables like *SP*, *Weight* and *SPxWeight* become insignificant in the *PER* regression. No substantial other change is to notice. The coefficient of *SPpast* is still

negative and in contradiction with the Investor Recognition Hypothesis, *SP* is still positive even if it loses some significance, while *Weight* and *SPxWeight* remain globally unchanged.

The coefficient of *SPadd12* is still positive and significant, and its magnitude is relatively close to the one of *SP*. It tends to confirm the simultaneous presence of short-term inclusion effects, which would be due to the price pressures of the Price Pressure Hypothesis, and long-term membership effect, which could be explained by all the "permanent effects" hypotheses: the Downward-Sloping Demand Curve Hypothesis, Liquidity Cost Hypothesis, Information Content Hypothesis and Investor Recognition Hypothesis.

The positive and even slightly significant coefficient of *SPdel12* is in line with the results from the pooled OLS regression. Furthermore, this coefficient is of the same magnitude as the one of *SPadd12*. Again, we need to be cautious, because this dummy is not totally reliable. Whether this coefficient is significantly positive is uncertain, but we can reasonably deduce that it is not negative. It follows that the Price Pressure Hypothesis alone is maybe not sufficient to explain the two short-term effects we observe, because the non-negativity of the coefficient of *SPdel12* is in contradiction with it.

VIF present slightly lower levels than in the pooled OLS regression, but they remain extremely high for *Weight* and *SPxWeight*, as expected.

2.5.4 Probit estimation

In order to perform the instrumental variables estimation of Section 2.5.5 and the self-selection estimation of Section 2.5.6, we first need to perform a preliminary probit regression of the (suspected) endogenous variable *SP* on a set of relevant instrumental variables (as explained in Sections 2.4.3.2 and 2.4.3.3). The results of this probit estimation are given in Table 2-11 and deliver some interesting facts.

As expected, the coefficient of *MktCaplog* is positive and the most significant of all explicative variables by far. It's the main driver of the probability of being a member of the S&P 500 index.

Table 2-11: Probit estimation of *SP* on relevant variables

Slope is the coefficient of the variable, *T-stat* is the student statistic, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression and *Log Likelihood* is the log likelihood of the probit regression. ClassRate is the percentage of correct classification of all observations, where we fixed the threshold at 0.5, so that an observation with a predicted probability (= the "fit") larger (resp. lower) than 0.5 is classified as a member (resp. non-member) of the S&P 500. ClassRateSP0 (resp. ClassRateSP1) gives the percentage of correct classification for the subset of observations that are non-members (resp. members) of the S&P 500. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10.

DepVar	SP			
Type	Slope	tstat		VIF
Intercept	-9.084	-262.368	***	
MktCaplog	0.544	338.104	***	1.310
NoIPO	0.162	4.764	***	1.360
DLiq	0.866	40.156	***	1.452
Prc_Low	0.007	0.366		1.148
Prc_High	0.121	13.414	***	1.101
RpEarn	0.076	16.016	***	1.163
Lev_Ind	-0.142	-10.255	***	1.106
N		606,195		
Log Likelihood		-229,187		
ClassRate		83.70%		
ClassRateSP0		94.28%		
ClassRateSP1		50.16%		

NoIPO is positive and significant, which is consistent with the S&P criteria of a minimum elapsed time of 6 to 12 months between the introduction of the firm in the stock market and its inclusion in the index.

DLiq is positive and significant, which is again in line with the minimum liquidity ratio requirement of 0.3 of the S&P guidelines. It is even the second most significant variable.

Although S&P guidelines suggest that a low price could decrease the probability of being included in the index, we do not find evidence of such an effect here. Nevertheless, one must keep in mind that S&P guidelines indicate the conditions that a non-member of the S&P 500 must theoretically meet to be considered for inclusion, while our probit regression rather investigates whether those requirements continue to be met after the inclusion, for any company member of the S&P 500. At this point, we do not conclude that a lower price does not decrease the probability for a non-member

to be included in the S&P 500 today, but that on average firms in the S&P 500 do not seem to present higher prices than the others, after having controlled for various factors.

On the contrary a high price significantly increases the probability to be in the S&P 500. But S&P 500 guidelines do not mention such a factor.

As expected, firms with four last quarters of positive reported earnings (proxied by *RpEarn*) are more likely to be part of the index than companies having experienced a loss at least once for the last year. It seems in total agreement with the S&P guidelines, which demand (in theory) 4 consecutive quarters of positive earnings to be considered as a candidate for inclusion.

Firm leverage (*Lev_Ind*) seems to have a negative impact on the probability to be in the index. Again, such a result appears in agreement with the S&P requirement of a reasonable leverage.

Globally, those results incline us to conclude that the criteria requested by the S&P committee for a firm to be considered for inclusion continue to be met later on, after the inclusion.

2.5.5 Instrumental variables estimation

Table 2-12 present the results of a pooled OLS regression where *SP* has been previously regressed (probit) on a set of instrumental variables and then replaced by its fitted value from the first probit step in the second OLS step. Results are consistent with those obtained with the OLS and the fixed-effect panel regressions. Only slight differences arise, which we comment hereafter.

SPpast is not significantly negative anymore, but still not significantly positive either. This does not bring any support to the Investor Recognition Hypothesis.

RD loses significance, while the coefficient of *Adv* becomes negative. Those control variables were already only marginally significant in the panel regression.

Interestingly, *SPxWeightxVar* is now positive and significant for all *PER*, *Ret*, *TQ* and *MB* regressions, while it was insignificant or even negative in the previous specifications, especially in the OLS regression. It is consistent with the predictions of our theoretical model.

The colineary issue is more severe than ever. Indeed, *SP_F* is a very strong positive function of *MktCaplog* because of the high significance of the latter in the first probit step, and *SPxWeight* is by construction driven by *MktCap*. Therefore we perform again a regression excluding *SPxWeight* in order to check the stability of our coefficients.

Results of Table 2-13 (where *SPxWeight_F* has been excluded from the regressors) are consistent with the previous ones, with one noticeable change: the coefficient of *Weight* in the *PER* regression is now significantly negative, which is probably due to the "Weight" effect embedded by the *SP_F* variable. Otherwise, the same comments as before apply.

Table 2-12: Instrumental variables regression

Slope is the coefficient of the variable, *p_ols* is the OLS p-value of the coefficient, *p_boot* is the bootstrap p-value of the coefficient, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10. The bootstrap p-value (*p_boot*) is not given for the dependent variable *Ret*: the high number of observations for this monthly regression leads to unaffordable computational time.

DepVar	PER_Ind				Ret			TQ_Ind				MB_Ind			
TYPE	Slope	p_ols	p_boot	VIF	Slope	p_ols	VIF	Slope	p_ols	p_boot	VIF	Slope	p_ols	p_boot	VIF
Intercept	7.066	0.000	0.001		0.050	0.000		1.159	0.000	0.001		3.280	0.000	0.001	
y_lag	0.023	0.000	0.001	1.004	-0.038	0.000	1.220	0.447	0.000	0.001	1.143	0.052	0.000	0.001	1.057
SP_F	6.252	0.000	0.001	4.725	0.043	0.000	4.568	1.132	0.000	0.001	4.700	3.696	0.000	0.001	4.648
Weight	4,436.043	0.000	0.001	**394.550**	37.397	0.000	**342.599**	729.589	0.000	0.001	**343.650**	2,256.121	0.000	0.001	**349.210**
SPxWeight_F	-4,523.368	0.000	0.001	**379.299**	-36.593	0.000	**327.257**	-712.551	0.000	0.001	**328.500**	-2,191.114	0.000	0.001	**334.074**
SPpast_F	0.255	0.860	0.849	1.020	0.002	0.158	1.026	0.045	0.087	0.068	1.030	-0.005	0.952	0.943	1.026
SPadd12_F	1.273	0.149	0.146	1.621	0.020	0.000	1.628	0.254	0.000	0.001	1.620	0.287	0.000	0.001	1.622
SPdel12_F	1.610	0.007	0.013	1.615	-0.002	0.006	1.613	0.061	0.000	0.001	1.598	0.166	0.000	0.001	1.604
RD	22.221	0.000	0.001	1.215	-0.049	0.000	1.198	0.056	0.476	0.476	1.197	0.379	0.119	0.136	1.205
Adv	-1.888	0.669	0.649	1.231	-0.045	0.000	1.221	-0.916	0.000	0.001	1.210	-0.474	0.137	0.126	1.215
Lev	9.118	0.000	0.001	1.379	0.011	0.000	1.360	-0.026	0.120	0.127	1.348	1.751	0.000	0.001	1.368
Assets	-1.959	0.000	0.001	5.868	-0.010	0.000	5.423	-0.260	0.000	0.001	5.633	-0.796	0.000	0.001	5.582
SalesG	0.046	0.923	0.925	1.162	-0.005	0.000	1.177	0.158	0.000	0.001	1.182	0.707	0.000	0.001	1.160
NoIPO	-2.366	0.072	0.070	1.029	-0.006	0.000	1.081	-0.116	0.000	0.001	1.028	-0.671	0.000	0.001	1.025
Variance	105.034	0.000	0.001	1.311	0.238	0.000	1.307	-0.443	0.008	0.009	1.290	-3.159	0.000	0.001	1.297
SPxWeightxVar_F	37,578.055	0.000	0.001	1.783	41.260	0.000	1.689	3,548.881	0.000	0.001	1.778	9,580.309	0.000	0.001	1.777
N	36,507				556,694			46,134				45,107			
AdjR2	0.019				0.184			0.548				0.215			
F	17.611				354.627			1,301.176				289.061			

Table 2-13: Instrumental variables regression without *SPxWeight_F*

Slope is the coefficient of the variable, *p_ols* is the OLS p-value of the coefficient, *p_boot* is the bootstrap p-value of the coefficient, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10. The bootstrap p-value (*p_boot*) is not given for the dependent variable *Ret*: the high number of observations for this monthly regression leads to unaffordable computational time.

DepVar	PER_Ind				Ret			TQ_Ind				MB_Ind			
TYPE	Slope	p_ols	p_boot	VIF	Slope	p_ols	VIF	Slope	p_ols	p_boot	VIF	Slope	p_ols	p_boot	VIF
Intercept	6.855	0.000	0.002		0.049	0.000		1.115	0.000	0.001		3.192	0.000	0.001	
y_lag	0.023	0.000	0.001	1.003	-0.035	0.000	1.212	0.461	0.000	0.001	1.118	0.053	0.000	0.001	1.056
SP_F	6.670	0.000	0.001	4.685	0.047	0.000	4.507	1.206	0.000	0.001	4.649	3.953	0.000	0.001	4.584
Weight	-171.147	0.010	0.014	2.286	0.111	0.221	2.214	3.893	0.021	0.023	2.229	24.690	0.000	0.001	2.263
SPpast_F	0.017	0.991	0.991	1.019	0.000	0.934	1.025	0.006	0.815	0.808	1.028	-0.126	0.152	0.145	1.024
SPadd12_F	3.020	0.000	0.003	1.391	0.033	0.000	1.423	0.508	0.000	0.001	1.422	1.082	0.000	0.001	1.419
SPdel12_F	2.740	0.000	0.001	1.408	0.006	0.000	1.440	0.212	0.000	0.001	1.433	0.647	0.000	0.001	1.432
RD	22.316	0.000	0.001	1.215	-0.048	0.000	1.198	0.095	0.234	0.209	1.197	0.433	0.079	0.084	1.205
Adv	-6.462	0.136	0.139	1.183	-0.081	0.000	1.176	-1.608	0.000	0.001	1.165	-2.680	0.000	0.001	1.168
Lev	8.680	0.000	0.001	1.364	0.008	0.000	1.344	-0.080	0.000	0.001	1.337	1.541	0.000	0.001	1.352
Assets	-1.644	0.000	0.001	5.178	-0.008	0.000	4.786	-0.212	0.000	0.001	4.947	-0.654	0.000	0.001	4.959
SalesG	0.645	0.155	0.156	1.093	-0.001	0.214	1.084	0.237	0.000	0.001	1.111	0.975	0.000	0.001	1.077
NoIPO	-2.796	0.033	0.027	1.025	-0.011	0.000	1.065	-0.170	0.000	0.001	1.025	-0.850	0.000	0.001	1.022
Variance	105.941	0.000	0.001	1.310	0.248	0.000	1.306	-0.361	0.032	0.049	1.290	-2.916	0.000	0.001	1.297
SPxWeightxVar_F	35,444.686	0.000	0.002	1.778	24.078	0.013	1.684	3,002.681	0.000	0.001	1.768	8,324.883	0.000	0.001	1.771
N	36,507				556,694			46,134				45,107			
AdjR2	0.018				0.182			0.536				0.194			
F	17.357				351.847			1,268.775				259.269			

Up to now, almost all our regressions have exhibited positive and significant coefficients for our short-term dummies *SPadd12* and *SPdel12*, the coefficient of the former being usually larger than the coefficient of the latter. As already said before, the short-term positive effect of an inclusion in the index can be explained by the Price Pressure Hypothesis, but the possible short-term positive effect of a deletion from the index needs a totally different explanation, that none of the current hypotheses proposed by the literature seem able to give. This is pure conjecture, but a possible argument would be the following: although we have not found very strong support for the Investor Recognition Hypothesis until now with the *SPpast* variable, a variant of it could help us to explain this puzzling short-term effect of a deletion. We could imagine that the investor recognition of stocks is only temporary, and works the best in the weeks and months following an addition or a deletion. The announcement of a newly included or excluded stock would have as a consequence the same "recognition" effect on the investors: they would make them aware of the existence of those stocks, whether they are added or deleted. This temporary interest in those stocks would result in (temporary) higher demand and, as a consequence, higher prices. Therefore, if we conjugate the Price Pressure Hypothesis and a short-term variant of Investor Recognition Hypothesis, it could at least partly explain what is happening: the Price Pressure Hypothesis predicts a positive coefficient for *SPadd12* and a negative one for *SPdel12*, while Investor Recognition Hypothesis predicts positive coefficients for both of them. A mix of those two explanations would then result in a strong positive coefficient for *SPadd12* and a weaker positive coefficient for *SPdel12*.

2.5.6 Self-selection estimation

Table 2-14 presents the results of the self-selection estimation. The negative and significant coefficient of the Mill's ratio (*Lambda*) is very interesting. This indicates the prevalence of self-selection and suggests that characteristics that make firms enter the S&P 500 index are negatively correlated with the firm value. It is a powerful argument against ICHb, since the latter posits that an inclusion in the S&P 500 signals and reveals good fundamentals for the newly included firm, while this negative *Lambda* tells a totally opposite story. As a consequence, the value premium is entirely attributable to the index effect (see the very strong *SP* coefficient), and not at all to ex ante better

hidden fundamentals of included firms that a so-called superior S&P picking ability would suddenly reveal to the market.

Table 2-14: Self-selection estimation

Slope is the coefficient of the variable, *p_ols* is the OLS p-value of the coefficient, *p_boot* is the bootstrap p-value of the coefficient, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10. The bootstrap p-value (*p_boot*) is not given for the dependent variable *Ret*: the high number of observations for this monthly regression leads to unaffordable computational time.

DepVar	PER_Ind				Ret			TQ_Ind				MB_Ind			
TYPE	Slope	p_ols	p_boot	VIF	Slope	p_ols	VIF	Slope	p_ols	p_boot	VIF	Slope	p_ols	p_boot	VIF
Intercept	4.937	0.002	0.002		0.056	0.000		1.213	0.000	0.001		3.514	0.000	0.001	
y_lag	0.023	0.000	0.001	1.003	-0.041	0.000	1.221	0.443	0.000	0.001	1.154	0.052	0.000	0.001	1.057
SP	5.394	0.000	0.001	**13.853**	0.067	0.000	**12.597**	1.564	0.000	0.001	**13.195**	5.090	0.000	0.001	**12.728**
Weight	809.493	0.009	0.011	**48.634**	4.002	0.000	**43.550**	117.485	0.000	0.001	**43.574**	334.434	0.000	0.001	**44.582**
SPxWeight	-830.800	0.006	0.008	**45.903**	-4.384	0.000	**41.003**	-108.508	0.000	0.001	**41.020**	-292.074	0.000	0.001	**42.011**
SPpast	-0.038	0.974	0.974	1.232	0.003	0.011	1.256	0.005	0.820	0.813	1.257	-0.121	0.099	0.104	1.267
SPadd12	3.392	0.002	0.006	1.053	0.000	0.780	1.058	0.306	0.000	0.001	1.047	0.516	0.000	0.001	1.048
SPdel12	5.898	0.032	0.032	1.209	-0.006	0.047	1.229	0.146	0.005	0.010	1.227	0.788	0.000	0.001	1.241
RD	25.956	0.000	0.001	1.193	-0.043	0.000	1.177	0.271	0.000	0.001	1.176	0.932	0.000	0.001	1.183
Adv	2.799	0.494	0.485	1.053	-0.008	0.126	1.049	0.056	0.555	0.559	1.047	2.441	0.000	0.001	1.049
Lev	7.608	0.000	0.001	1.318	0.008	0.000	1.302	-0.094	0.000	0.001	1.296	1.596	0.000	0.001	1.309
Assets	-1.343	0.000	0.001	5.006	-0.008	0.000	4.616	-0.225	0.000	0.001	4.826	-0.705	0.000	0.001	4.757
SalesG	-0.105	0.817	0.801	1.092	-0.009	0.000	1.077	0.071	0.000	0.001	1.084	0.445	0.000	0.001	1.071
NoIPO	-1.817	0.170	0.172	1.043	-0.010	0.000	1.077	-0.139	0.000	0.001	1.039	-0.761	0.000	0.001	1.038
Variance	101.514	0.000	0.001	1.285	0.228	0.000	1.273	-0.582	0.000	0.004	1.261	-3.240	0.000	0.001	1.272
SPxWeightxVar	-53.413	0.024	0.028	1.024	-0.063	0.010	1.034	-0.443	0.456	0.304	1.022	-5.402	0.003	0.014	1.023
Lambda	-2.077	0.000	0.001	**10.006**	-0.041	0.000	9.107	-0.858	0.000	0.001	9.418	-2.739	0.000	0.001	9.125
N	36,507				556,694			46,134				45,107			
AdjR2	0.017				0.184			0.547				0.222			
F	15.700				353.911			1,266.404				292.681			

Table 2-15: Self-selection estimation without the dummy *SP*

Slope is the coefficient of the variable, *p_ols* is the OLS p-value of the coefficient, *p_boot* is the bootstrap p-value of the coefficient, *VIF* is the Variance Inflation Factor, *N* is the number of observations of the regression, *AdjR2* is the adjusted R^2 and *F* is the F-test. One star (resp. two and three) indicates a significance level of 10% (resp. 5% and 1%). A *VIF* in bold indicates a VIF higher than 10. The bootstrap p-value (*p_boot*) is not given for the dependent variable *Ret*: the high number of observations for this monthly regression leads to unaffordable computational time.

DepVar	PER_Ind				Ret			TQ_Ind				MB_Ind			
TYPE	Slope	p_ols	p_boot	VIF	Slope	p_ols	VIF	Slope	p_ols	p_boot	VIF	Slope	p_ols	p_boot	VIF
Intercept	1.608	0.282	0.287		0.012	0.000		0.247	0.000	0.001		0.443	0.000	0.001	
y_lag	0.023	0.000	0.001	1.003	-0.035	0.000	1.214	0.492	0.000	0.001	1.077	0.059	0.000	0.001	1.048
Weight	901.203	0.003	0.006	**48.502**	5.030	0.000	**43.469**	133.979	0.000	0.001	**43.533**	403.714	0.000	0.001	**44.515**
SPxWeight	-851.304	0.004	0.008	**45.896**	-4.345	0.000	**41.003**	-102.144	0.000	0.001	**41.014**	-284.599	0.000	0.001	**42.010**
SPpast	-0.428	0.710	0.704	1.227	-0.002	0.090	1.250	-0.116	0.000	0.001	1.250	-0.509	0.000	0.001	1.261
SPadd12	4.162	0.000	0.002	1.038	0.010	0.000	1.044	0.534	0.000	0.001	1.033	1.339	0.000	0.001	1.032
SPdel12	5.766	0.036	0.037	1.208	-0.007	0.016	1.229	0.091	0.095	0.102	1.226	0.593	0.001	0.002	1.241
RD	30.141	0.000	0.001	1.153	0.000	0.954	1.143	1.260	0.000	0.001	1.141	4.152	0.000	0.001	1.149
Adv	5.648	0.164	0.163	1.038	0.027	0.000	1.035	0.761	0.000	0.001	1.037	4.868	0.000	0.001	1.037
Lev	6.174	0.000	0.001	1.187	-0.009	0.000	1.167	-0.446	0.000	0.001	1.184	0.244	0.000	0.001	1.171
Assets	-0.685	0.000	0.001	2.478	0.000	0.136	2.358	-0.035	0.000	0.001	2.392	-0.088	0.000	0.001	2.403
SalesG	0.106	0.816	0.820	1.085	-0.006	0.000	1.068	0.105	0.000	0.001	1.081	0.638	0.000	0.001	1.063
NoIPO	-1.371	0.300	0.305	1.039	0.000	0.824	1.054	0.000	0.994	0.990	1.035	-0.345	0.000	0.003	1.035
Variance	97.658	0.000	0.001	1.280	0.163	0.000	1.255	-2.251	0.000	0.001	1.237	-8.946	0.000	0.001	1.249
SPxWeightxVar	-49.663	0.035	0.032	1.023	-0.025	0.300	1.033	0.692	0.270	0.155	1.021	-1.608	0.407	0.284	1.023
Lambda	0.924	0.000	0.001	1.763	-0.005	0.000	1.711	-0.013	0.025	0.023	1.714	0.014	0.433	0.429	1.721
N	36,507				556,694			46,134				45,107			
AdjR2	0.017				0.179			0.494				0.114			
F	15.291				342.909			1,048.400				136.325			

Note that *SP* and *Lambda* are by construction very correlated, and the large VIF attached to them prove it. It is a well known "drawback" of the simultaneous introduction of the correction for self-selection variable (*Lambda*) along with the treatment dummy (*SP*). Therefore we have to remain cautious when interpreting those results. However, a comparison of the coefficients of the control variables of this regression with the pooled OLS specification shows good stability, and let us believe that this additional collinearity issue is unlikely to affect significantly the precision of our coefficients. Furthermore, when the p-values are so small, collinearity should not be considered as a serious issue. Nonetheless, if we perform the same regression without *SP* as a robustness check (see Table 2-15), the sign of Lambda becomes inconclusive: positive for *PER*, negative for *TQ* and insignificant for *MB*. But the "validity" of such a specification excluding the most important variable of the regression is difficult to assess.

2.6 Conclusion

This whole section aimed at investigating the S&P 500 index effect, i.e. the value premium attached to companies that belong to the S&P 500, and at empirically testing the five hypotheses that have been proposed by the literature: the Price Pressure hypothesis, the Imperfect Substitutes / Downward-Sloping Demand Curve for Stocks Hypothesis, the Liquidity Cost Hypothesis, the Information Content / Index Member Certification Hypothesis (ICHa if the inclusion in the index causes the better performance, and ICHb if the better performance causes the inclusion), and the Market Segmentation / Investor Recognition Hypothesis.

Our study builds mainly on the respective works of Morck and Yang (2002) and Campa and Kedia (2002). The former have investigated the relationship between the growth of index funds and the significance of the index effect. They show that the value premium increases in parallel with the growth of indexing, and suggest that it is the huge demand of index funds for the stocks composing the index that is responsible for the observable higher value (proxied by the Tobin's q) for these stocks. As a first step, we replicate and confirm most of their results, based on simple OLS regressions, on a sample of monthly data ranging from September 1976 to December 2005 comprising all non-financial firms of the CRSP database with sufficient data and a market capitalization at least

equal to the market capitalization of the smallest company of the S&P 500. As a second step, we introduce additional interest and control variables in the regressions in order to test the largest possible set of hypotheses, and we address what we believe to be some misspecification issues. We also broaden the set of the dependent variables that are supposed to proxy for the firm value from one (the Tobin's q) to three (the Price-Earning Ratio, the Tobin's q and the Market-to-Book ratio) in order to make our results more robust. As a third step, because we have good reasons to suspect a potential endogeneity problem in the regressions of Morck and Yang (2002), we apply to our own problematic the methodology that Campa and Kedia (2002) applied in another context (namely, the value premium of diversification), and we perform successively fixed-effect panel, instrumental variables and self-selection regressions, which are different and complementary ways to treat (at least partly) an endogeneity issue. All those regressions lead us to confirm the existence of a permanent index effect found by most authors in this stream of literature. They also allow us to confirm or reject some hypotheses proposed by the literature.

The negative sign in all our regressions of our *SPxWeight* variable, which stands for the weight of a company in the S&P 500 index, is bad news for the defendants of the Downward-Sloping Demand Curve Hypothesis, since the latter posits that the higher the weight of a company in the S&P 500, and the higher the shift of demand and thus the value premium for the stock of this company. We observe indeed a positive weight effect on the firm value, but this is true for all companies, whether they belong to the index or not, and this effect is stronger for the non-members than for the members. It is hardly compatible with a Downward-Sloping Demand Curve explanation.

The negative sign of *Lambda* in our self-selection treatment regression casts also some serious doubts on the ICHb, i.e. the certification effect theory, according to which better fundamentals and performance would trigger the decision of inclusion in the S&P 500. The value premium of the S&P 500 members is not due to possibly better fundamentals of these firms, but rather to the pure advantage to be in the index. In fact, the characteristics leading to the inclusion have an adverse effect on the firm value. But the index effect more than compensates this adverse effect and leads to the observable higher value of the S&P 500 members.

Evidence in favor of the classical version of the Investor Recognition Hypothesis is weak as well. The negative sign of *SPpast* rejects the idea that a former member of the

S&P 500 (but not a member anymore) still benefits from a value premium. However, the Price Pressure Hypothesis and a **short-term** version of the Investor Recognition Hypothesis could be useful to explain the coefficients of our short-term variables *SPadd12* and *SPdel12*. Indeed, the Price Pressure Hypothesis may explain why a newly (less than 12 months) included stock benefits from a premium (because of temporary price pressures) more than "older" members, as suggested by the positive and significant sign of *SPadd12*. The positive sign of *SPdel12* is naturally at odds with a Price Pressure Hypothesis explanation, but could be explained by a short-term version of the Investor Recognition Hypothesis: as a stock is deleted from the index, it experiences a regain of attention from the investors, which can translate in a temporary additional interest for this stock. But at this point, this largely remains pure conjecture.

Our results do not allow to clearly accept or reject the Liquidity Cost Hypothesis and ICHa, because we did not test them directly.

Eventually, they give some weak and partial support to our theoretical model since the instrumental variables regressions confirm our prediction that the higher the variance of the stock, the stronger the positive impact of a positive deviation on the firm value (through the positive coefficient of *SPxWeightxVar*). But all other regressions deliver insignificant or sometimes contradictory results in this respect, so that it is preferable not to draw any definitive conclusions at this point.

3 News and the Cost of Capital in a universe of Constrained Investors

3.1 Introduction

The purpose of this section is to introduce the concept of news in the model developed in Section 1. In particular, we want to examine how news disclosure can impact the cost of capital through the various determinants of our pricing Equation (1-31):

$$p_j = \frac{\mu_j - \frac{b}{N}\left(\sigma_{j,M} - \sigma_j^2 \varepsilon_j\right)}{R_f}$$

We will consider here both firm-specific news (like profit warnings, merger & acquisition announcements, analyst comments, earnings, etc.) and macro-economic news, although the empirical estimation of Section 4 will restrict the analysis to the former kind of announcements. Public news brings to the market more or less valuable and precise information about the fundamentals and the future prospects of the firm. Those signals allow investors to learn about the true distribution of stock prices, to update their expectations and to decrease the uncertainty around future cash-flows. As a consequence, on the basis of this new information, investors will use the distribution of the future prices conditional to the signals, instead of the unconditional distribution of prices, since the former embeds now more information than the latter. Until now, Equation (1-31) has expressed the current price as a function of the unconditional expectation μ_j, variance σ_j^2 and covariance $\sigma_{j,M}$. If we state it explicitly, we obtain Equation (3-1):

$$p_j = \frac{E(\tilde{p}_j) - \frac{b}{N}\left(Cov(\tilde{p}_j, \tilde{p}_M) - Var(\tilde{p}_j)\varepsilon_j\right)}{R_f} \tag{3-1}$$

Micro and macro signals z_j and z_M constitute an information set $\mathbf{z}$ available to any investor and consequently leads to the following conditional distribution of $\tilde{\mathbf{p}}$:

$$\tilde{\mathbf{p}} \mid (z_j, z_M) \sim N\left(E(\tilde{\mathbf{p}} \mid z_j, z_M), Var(\tilde{\mathbf{p}} \mid z_j, z_M)\right)$$
$$\Leftrightarrow \tilde{\mathbf{p}} \mid \mathbf{z} \sim N\left(E(\tilde{\mathbf{p}} \mid \mathbf{z}), Var(\tilde{\mathbf{p}} \mid \mathbf{z})\right)$$

where we define $\mathbf{z}$ as the set of all micro and macro signals available to all investors:

$$\mathbf{z} \equiv \left\{ z_j , z_M \right\}$$

We can of course also consider several signals: Q_j micro signals and Q_M macro signal:

$$\mathbf{z} \equiv \left\{ z_{j,1},...,z_{j,q},...,z_{j,Q_j} , z_{M,1},...,z_{M,q},...,z_{j,Q_M} \right\}$$

Given this new set of information, the pricing equation can be rewritten as follows, on the basis of the new conditional moments of the price distribution instead of the unconditional ones:

$$p_j = \frac{E\left(\tilde{p}_j \mid \mathbf{z}\right) - \frac{b}{N}\left(Cov\left(\tilde{p}_j, \tilde{p}_M \mid \mathbf{z}\right) - Var\left(\tilde{p}_j \mid \mathbf{z}\right)\varepsilon_j\right)}{R_f} \tag{3-2}$$

The purpose of this section is to determine the conditional expectation $E\left(\tilde{p}_j \mid \mathbf{z}\right)$, variance $Var\left(\tilde{p}_j \mid \mathbf{z}\right)$ and covariance $Cov\left(\tilde{p}_j, \tilde{p}_M \mid \mathbf{z}\right)$ and to examine how the number and the quality of signals can improve or deteriorate the cost of capital. We provide some insight about what should be an optimal communication strategy in order to minimize the resulting cost of capital of the firm, and how such a strategy may depend on the potential aggregate stock deviation ε_j. We show that the answer to these questions is not as obvious as it can appear at first sight as that more communication does not unambiguously and necessarily imply lower cost of capital. Nevertheless, provided that the unconditional distribution of returns is stable through time, it is possible to demonstrate that a new signal has the unambiguous effect to decrease the conditional variance and covariance of the price, especially if it is precise.

Our empirical study of Section 4 will focus on the link between information arrival and the stock volatility in a dynamic context. It will allow us to show that the number of announcements largely explains the well-known phenomena of conditional volatility persistence (which triggered the historical development of GARCH models) and that a higher frequency of announcements on a stock is associated with a higher probability of high volatility regime on this stock (in a Markov-switching regression framework).

3.2 Literature

The starting point is the model developed in the previous chapter. The way we introduce the effect of news on prices is borrowed from Lambert, Leuz and Verrecchia (2007). The purpose of this section is "only" to enrich our theoretical model with the notion of signals, to develop equations and to show how news (i.e. signals) can affect the cost of capital in different and various ways. Therefore, to avoid any redundancy, a thorough literature review on the topic is left to Section 4. We limit the subsequent review to a short summary of the theoretical review of Lambert, Leuz and Verrechia (2006) on the topic. We give in brackets the page reference when we quote this paper.

As observed by those authors in their introduction, "there is surprisingly little theoretical work on the hypothesized link" (p. 2) between better information disclosure and a lower cost of capital. Some authors have already tried to investigate the impact of information on the cost of capital in single-firm settings, or settings where cash flows across firms are uncorrelated (see Diamond and Verrecchia (1991), Baiman and Verrecchia (1996), Easley and O'Hara (2004) and Verrecchia (2001) for a survey). They deliver interesting insights, as they suggest an indirect link between disclosure and firms' cost of capital based on market liquidity and adverse selection in secondary markets. But they fail to take into account the distinction between specific and systematic risks and the potential forces of diversification, to cite only a few important elements.

In a multi-firm settings, many authors have worked on the "estimation risk" in finance. In this literature, information signals are typically modeled as arising from a historical time-series of return observations. Barry and Brown (1985) and Coles, Loewenstein and Suay (1995), among others, investigated this field, and found that the betas of the "high information" securities are lower than they would be in the lower information case. However, as noticed by Lambert, Leuz and Verrecchia (2007), "these studies do not address the question of how an individual firm's disclosures can influence its cost of capital within an unequal information environment" (p. 3).

"Rather than restricting attention to signals that are historical observations of returns", the paper of Lambert, Leuz and Verrecchia (2007) (and our own model after them) "uses a more conventional information-economics approach in which information signals are related to realized or future cash flows. With this approach", they "allow for

more general changes in the information environment, and" they "are able to prove much stronger results" (p. 3).

3.3 Conventions, notations and assumptions

Conventions and notations have already been detailed in Section 1.2 and Section 1.3. Assumptions are the same as in Section 1.5, with one more: we assume that signals are normally distributed and independent.

3.4 Model development

3.4.1 General setting

Similarly to Lambert, Leuz and Verrecchia (2007), we introduce $\tilde{z}_{j,q}$, the q^{th} signal about future price $\tilde{p}_j$ of asset *j:*

$$\tilde{z}_{j,q} \equiv \tilde{p}_j + \sqrt{v_{j,q}}\,\tilde{u}_{j,q}$$ [25]

with $\tilde{p}_j$ defined as in the previous chapter (= vector of future prices $\tilde{\mathbf{p}} \sim N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$);

$v_{j,q}$ the variance of the noise of the signal $\tilde{z}_{j,q}$;

$\tilde{u}_{j,q} \sim N(0,1)$.

There are of course other possible ways to model signals, but as already mentioned above, Lambert, Leuz and Verrecchia (2007) argue that "rather than restricting attention to signals that are historical observations of returns", they prefer to "use a more conventional information-economics approach in which information signals are related to realized or future cash flows. With this approach," they "allow for more general changes in the information environment, and" they "are able to prove much stronger results" (p. 3). Furthermore, "this formulation is consistent with the way information is modeled in virtually all conventional statistical inference problems (see DeGroot,

[25] We had also considered to introduce an additive term $\eta_{j,q}$ in the equation as well, to account for possible positive or negative bias, but since this parameter is known by investors, this would have no effect on the resulting conditional distribution of prices. A possible way to introduce such a bias would be to make a difference between the true distribution of the signals and the distribution believed by the investors.

1970). It is also consistent with virtually all papers in the vast noisy rational expectations literature in accounting and finance (see Verrecchia, 2001, for a review)".

Note also that this definition of a signal is the only addition we make to our theoretical model of Section 3. Besides this new element, all the rest remains rigorously identical.

Intuitively, if $v_{j,q} = 0$, the signal is infinitely precise and is sufficient to get complete knowledge in order to perfectly predict the future price $\tilde{p}_j$. Indeed, observing this infinitely precise signal is then equivalent to observing the future price, and we can write: $\tilde{p}_j \left| z_{j,q} = \tilde{p}_j \right| p_j = p_j$. There is no uncertainty about the future price anymore.

On the contrary, if $v_{j,q} \to \infty$, the signal is too noisy and does not bring any valuable information about the future price anymore. In this extreme case, the conditional distribution of $\tilde{p}_j$ will be identical to the unconditional one.

We suppose $Cov(\tilde{u}_{j,q}, \tilde{u}_{k,r}) = 0$ and $Cov(\tilde{u}_{j,q}, \tilde{p}_{k,t+1}) = 0 \ \ \forall j,k = 1,...,K$, $q = 1,...,Q_j$ and $r = 1,...,Q_k$ since the number of signals can be different across firms. Introducing non-zero correlations is possible but gives much more complex expressions. The reader interested in such extension may find valuable to explore the topic of sectorial factors and the models of the Chamberlain-Rotschild (1983) kind.

It is now theoretically possible to compute the new moments of the conditional distribution of asset prices (conditional to all signals in the economy):

- $E\left(\tilde{p}_j \mid z_{1,1},...,z_{1,q},...,z_{1,Q_1},...,z_{j,1},...,z_{j,q},...,z_{j,Q_j},...,z_{K,1},...,z_{K,q},...,z_{K,Q_K}\right)$;
- $Var\left(\tilde{p}_j z_{1,1},...,z_{1,q},...,z_{1,Q_1},...,z_{j,1},...,z_{j,q},...,z_{j,Q_j},...,z_{K,1},...,z_{K,q},...,z_{K,Q_K}\right)$;
- $Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_{1,1},...,z_{1,q},...,z_{1,Q_1},...,z_{j,1},...,z_{j,q},...,z_{j,Q_j},...,z_{K,1},...,z_{K,q},...,z_{K,Q_K}\right)$

3.4.2 Introduction of two simplifications

However, conditioning on so many various signals does not come without a high computation cost. In order to make these moments more legible and tractable, we bring the two following adaptations.

3.4.2.1 Maximum one signal by asset

Firstly, because summations and indices make the above notations a little heavy, we choose to consider a maximum of one signal by asset (i.e., $Q_j <= 1, \forall j$). Therefore, the new signal received by investors about asset *j* is $\tilde{z}_j$:

$$\tilde{z}_j \sim N(\mu_j, \sigma_{j,j} + v_j) : \tilde{z}_j \equiv \tilde{p}_j + \sqrt{v_j}\tilde{u}_j$$

instead of the Q_j signals $\tilde{z}_{j,q}$:

$$\tilde{z}_{j,q} \sim N(\mu_j, \sigma_{j,j} + v_{j,q}) : \tilde{z}_{j,q} \equiv \tilde{p}_j + \sqrt{v_{j,q}}\tilde{u}_{j,q}$$

Note that you could also consider this new signal $\tilde{z}_j$ as an aggregation of all the information provided by the Q_j signals $\tilde{z}_{j,q}$. The "global" signal $\tilde{z}_j$ must then be such that the information given by $\tilde{z}_j$ is exactly the same (not more, not less) as the one given by the Q_j signals $\tilde{z}_{j,q}$. In other words, conditioning on $\tilde{z}_j$ or on the Q_j signals $\tilde{z}_{j,q}$ should give strictly identical results. In this case, the variance of v_j is equal to $\left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)^{-1}$ (that is to say, the precision of the aggregate signal $\tilde{z}_j$ is equal to the sum of the precisions of the individual signals). We will see formally in a subsequent section how it is possible to handle more than one signal by asset.

The new moments to compute are now the following:

- $E\left(\tilde{p}_j \mid z_1,...,z_j,...,z_K\right)$;
- $Var\left(\tilde{p}_j \mid z_1,...,z_j,...,z_K\right)$;
- $Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_1,...,z_j,...,z_K\right)$.

3.4.2.2 Two signals: micro (on asset *j*) and macro (on market *M*)

Secondly, since conditioning on signals of all *K* assets of the economy leads to complex computations (because it requires the inversion of a $[K \times K]$ matrix), we restrict our conditioning to two kinds of signals: a firm-specific micro signal $\tilde{z}_j$ and an aggregate

market macro signal $\tilde{z}_M$ and we will compute the moments of interest in Equation (3-2) conditional to those two signals $\tilde{z}_j$ and $\tilde{z}_M$: the conditional expectation $E(\tilde{p}_j \mid z_j, z_M)$, variance $Var(\tilde{p}_j \mid z_j, z_M)$ and covariance $Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M)$.

As above, $\tilde{z}_j \equiv \tilde{p}_j + \sqrt{v_j}\tilde{u}_j$ is a specific-micro signal about future asset *j* payoff while $\tilde{z}_M$ is a macro signal about the future market price:

$\tilde{z}_M \equiv \tilde{p}_M + \sqrt{v_M}\tilde{u}_M$

Technically, we will consider the macro signals $\tilde{z}_M$ as exogenously given. However, we can try to have a better intuitive understanding of those macro signals. We can see $\tilde{z}_M$ as a signal about the future market price inferred from all micro signals $\tilde{z}_k : k \neq j$. We do not want to use $\tilde{z}_j$ in the collection of micro signals that will produce $\tilde{z}_M$, in order to avoid any redundancy and the apparition of an extra-correlation term between $\tilde{u}_j$ and $\tilde{u}_M$ in the subsequent computations. Indeed, if $\tilde{z}_j$ was partly used to generate $\tilde{z}_M$, the noise of $\tilde{z}_j$ would likely contaminate $\tilde{z}_M$ so that the assumption $Cov(\tilde{u}_j, \tilde{u}_M) = 0$ would be much less sustainable. From our perspective, it is a bit like if the stock market universe was separated in two "worlds":

- the world of company "*j*", which is our company of interest, and this firm generates signals $\tilde{z}_j$;
- "the rest of the world", i.e. the world of all other companies *k* in the market, *k=1...K,* except *j*. Those companies generate signals about the market, $\tilde{z}_M$. In fact, all those companies generate different signals on the market, so that we should write something like $\tilde{z}_M^k$ rather than $\tilde{z}_M$. However, as we wish to focus on only one signal of each type (micro and macro), we will omit this index in the rest of this analysis.

At this stage, for the sake of generality and simplicity, we consider $\tilde{z}_M$ as purely exogenous and we do not say explicitly how $\tilde{z}_M$ is related to micro signals $\tilde{z}_k : k \neq j$. However, in a subsequent section, we will explore one way to make this signal $\tilde{z}_M$ partly endogenous to micro signals $\tilde{z}_k : k \neq j$ bringing information on the assets in the economy. For the moment, we simply assume that all signals $\tilde{z}_k$ emitted by companies

other than *j* say something (even very noisy) about the future market price $\tilde{p}_M$ and can be transformed into a market signal $\tilde{z}_M$:

$$\tilde{z}_k \equiv \tilde{p}_k + \sqrt{v_k}\tilde{u}_k \Rightarrow \tilde{z}_M \equiv \tilde{p}_M + \sqrt{v_M}\tilde{u}_M$$

Our two simplifications can be roughly summarized as such:

$$\underbrace{\tilde{z}_{j,1},\ldots,\tilde{z}_{j,q},\ldots,\tilde{z}_{j,Q_j}}_{\tilde{z}_j},\underbrace{\underbrace{\tilde{z}_{1,1},\ldots,\tilde{z}_{1,q},\ldots,\tilde{z}_{1,Q_1}}_{\tilde{z}_1},\ldots,\underbrace{\tilde{z}_{k,1},\ldots,\tilde{z}_{k,q},\ldots,\tilde{z}_{k,Q_k}}_{\tilde{z}_k},\ldots,\underbrace{\tilde{z}_{K,1},\ldots,\tilde{z}_{K,q},\ldots,\tilde{z}_{K,Q_K}}_{\tilde{z}_K}}_{\tilde{z}_M}$$

with

$$\left(\tilde{z}_{1,1},\ldots,\tilde{z}_{1,q},\ldots,\tilde{z}_{1,Q_1},\ldots,\tilde{z}_{k,1},\ldots,\tilde{z}_{k,q},\ldots,\tilde{z}_{k,Q_k},\ldots,\tilde{z}_{K,1},\ldots,\tilde{z}_{K,q},\ldots,\tilde{z}_{K,Q_K}\right) \supseteq \left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right) \supset \tilde{z}_M$$

Note that if the signals $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$ are constructed in such a way that they synthesize signals $\left(\tilde{z}_{1,1},\ldots,\tilde{z}_{1,q},\ldots,\tilde{z}_{1,Q_1},\ldots,\tilde{z}_{k,1},\ldots,\tilde{z}_{k,q},\ldots,\tilde{z}_{k,Q_k},\ldots,\tilde{z}_{K,1},\ldots,\tilde{z}_{K,q},\ldots,\tilde{z}_{K,Q_K}\right)$, then the "$\supseteq$" becomes a "=". Moreover, $\tilde{z}_M$ is a subset of $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$ because information on the components of the market permits to infer information on the market itself, while the opposite is not true (provided that $\tilde{z}_M$ is constructed as a by-product of the $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$). However, if $\tilde{z}_M$ encompasses also some purely macro information like interest or inflation rates changes, which is unlikely to be contained in the micro-signals $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$, then $\tilde{z}_M$ would not be totally included in $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$ anymore. In this case, it would be worth distinguishing the two kinds of signals impacting the market:

1) Purely macro signal $\tilde{z}_M$ is not contained in $\left(\tilde{z}_1,\ldots,\tilde{z}_k,\ldots,\tilde{z}_K\right)$ and occurs directly under the following form:

   $\tilde{z}_M \equiv \tilde{p}_M + \sqrt{v_M}\tilde{u}_M$

2) Market signal $\tilde{z}_M$ inferred from individual micro signals on assets in the economy is included in $\left(\tilde{z}_1,\ldots,\tilde{z}_j,\ldots,\tilde{z}_K\right)$ and results from the transformation of $\tilde{z}_k$ into $\tilde{z}_M$:

   $\tilde{z}_k \equiv \tilde{p}_k + \sqrt{v_k}\tilde{u}_k \Rightarrow \tilde{z}_M \equiv \tilde{p}_M + \sqrt{v_M}\tilde{u}_M$

The potential link between v_M and v_k will be investigated in a subsequent section. For the moment, we consider $\tilde{z}_M$ as exogenously given.

In summary, we have that:

$$\underbrace{\underbrace{\tilde{z}_{j,1},\ldots,\tilde{z}_{j,q},\ldots,\tilde{z}_{j,Q_j}}_{\tilde{z}_j},\underbrace{\underbrace{\tilde{z}_{1,1},\ldots,\tilde{z}_{1,q},\ldots,\tilde{z}_{1,Q_1}}_{\tilde{z}_1},\ldots,\underbrace{\tilde{z}_{k,1},\ldots,\tilde{z}_{k,q},\ldots,\tilde{z}_{k,Q_k}}_{\tilde{z}_k},\ldots,\underbrace{\tilde{z}_{K,1},\ldots,\tilde{z}_{K,q},\ldots,\tilde{z}_{K,Q_K}}_{\tilde{z}_K},\underbrace{\tilde{z}_{M,1},\ldots,\tilde{z}_{M,q},\ldots,\tilde{z}_{M,Q_M}}_{\tilde{z}_M}}_{\tilde{z}_M}}$$

Assuming that investors do not have a full knowledge of the information contained in the economy but restrict their attention to one particular stock and to the signals driving the market does not seem completely unrealistic. When an investor has to make a decision about an investment in a given asset, he is unlikely to carefully review all the information concerning all other assets of the economy, but will most likely focus on the signals existing about the concerned asset and the macro signals about the general context and the current market environment.

3.4.3 Determination of the unconditional distribution of prices

Before finding the conditional distribution of $(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M)$, we first have to determine the unconditional joint distribution of $(\tilde{p}_j, \tilde{p}_M, \tilde{z}_j, \tilde{z}_M)$.

This is achieved by applying a standard result of the multivariate statistical analysis. We thank Professor Luc Bauwens for having provided us with help and references about this result in particular. The interested reader will find more details on this specific topic in any introductory book to multivariate statistical analysis, like Anderson (2003).

We enounce now the general property of multivariate normal distributions which allows us, under very specific conditions, to infer the joint distribution of two normal random variables from their marginal and conditional distributions[26]:

[26] The text of Property 3-1 and Property 3-2 is directly drawn from the slides of the course of Professor Luc Bauwens "Advanced Econometrics". Slight adaptations have been made to be consistent with our own notations.

Property 3-1

If $\tilde{\mathbf{x}}_2|\mathbf{x}_1 \sim N_{n_2}(\boldsymbol{\pi}_0 + \boldsymbol{\Pi}\mathbf{x}_1, \boldsymbol{\Omega})$, where $\boldsymbol{\pi}_0(n_2 \times 1)$, $\boldsymbol{\Pi}(n_2 \times n_1)$ and $\boldsymbol{\Omega}$ ($n_2 \times n_2$ and PD) are matrices of constants, and $\tilde{\mathbf{x}}_1 \sim N_{n_1}(\boldsymbol{\xi}, \boldsymbol{\Phi})$, where $\boldsymbol{\xi}(n_1 \times 1)$ and $\boldsymbol{\Phi}$ ($n_1 \times n_1$ and PD) are matrices of constants, then

$\tilde{\mathbf{x}} = \begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{pmatrix} \sim N_n(\mathbf{m}, \mathbf{M})$ where $n = n_1 + n_2$,

$\mathbf{m} = \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\pi}_0 + \boldsymbol{\Pi}\boldsymbol{\xi} \end{pmatrix}$ and $\mathbf{M} = \begin{pmatrix} \boldsymbol{\Phi} & \boldsymbol{\Phi}\boldsymbol{\Pi}' \\ \boldsymbol{\Pi}\boldsymbol{\Phi} & \boldsymbol{\Omega} + \boldsymbol{\Pi}\boldsymbol{\Phi}\boldsymbol{\Pi}' \end{pmatrix}$

We can apply Property 3-1 to our specific setting because we know that:

1) $\tilde{\mathbf{p}} \sim N_K(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ thanks to our initial assumption of Section 1.5 about the unconditional distribution of prices;
2) $\tilde{z}_j \equiv \tilde{p}_j + \sqrt{v_j}\tilde{u}_j$ and $\tilde{z}_M \equiv \tilde{p}_M + \sqrt{v_M}\tilde{u}_M$ thanks to our initial assumption of Section 3.4.1 about the form of signals.

From 1), we deduce that the unconditional distribution of $(\tilde{p}_j, \tilde{p}_M)$ is:

$\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix}\right)$ (see Equations (1-10) and (1-11) for a reminder, if necessary)).

From 2), we deduce that the distribution of the signals conditional to the prices is:

$\begin{pmatrix} \tilde{z}_j \\ \tilde{z}_M \end{pmatrix} \Big| \begin{pmatrix} p_j \\ p_M \end{pmatrix} \sim N_2\left(\begin{pmatrix} p_j \\ p_M \end{pmatrix}, \begin{pmatrix} v_j & 0 \\ 0 & v_M \end{pmatrix}\right)$.

We can now directly apply the above Property 3-1 to our particular case with:

$\tilde{\mathbf{x}}_1 = \begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \end{pmatrix}$, $\tilde{\mathbf{x}}_2 = \begin{pmatrix} \tilde{z}_j \\ \tilde{z}_M \end{pmatrix}$, $\boldsymbol{\pi}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\boldsymbol{\Pi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{\Omega} = \begin{pmatrix} v_j & 0 \\ 0 & v_M \end{pmatrix}$, $\boldsymbol{\xi} = \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix}$,

$\boldsymbol{\Phi} = \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix}$ and $n_1 = n_2 = 2$.

The joint unconditional distribution of the prices and the signals is therefore the following:

$$\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \\ \tilde{z}_j \\ \tilde{z}_M \end{pmatrix} \sim N_4 \left(\mathbf{m} = \begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \end{pmatrix} = \begin{pmatrix} \mu_j \\ \mu_M \\ \mu_j \\ \mu_M \end{pmatrix}, \mathbf{M} = \begin{pmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} & \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \\ \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} & \begin{pmatrix} \sigma_j^2 + \nu_j & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 + \nu_M \end{pmatrix} \end{pmatrix} \right)$$

Notice that $\mathbf{m_1} = \mathbf{m}_2$, $\mathbf{M_{11}} = \mathbf{M_{12}} = \mathbf{M_{21}}$ and that all anti-diagonal elements of $\mathbf{M_{11}}, \mathbf{M_{12}}, \mathbf{M_{21}}$ and $\mathbf{M_{22}}$ are equal to $\sigma_{j,M}$, which will make the subsequent computations more tractable. This nice property is a direct consequence of our assumption of independence between the noise $\tilde{u}_j$ of the signal $\tilde{z}_j$ and the noise $\tilde{u}_M$ of the market $\tilde{z}_M$. Note, however, that if noises are independent, signals are not! Indeed, their price components can be correlated.

3.4.4 Determination of the distribution of prices conditional to signals

Multivariate statistic provides us with another nice property of the multivariate normal distributions. In fact, Property 3-2 is the reciprocal of Property 3-1. It permits to infer the marginal and conditional distributions from the joint distribution of two variables, under certain conditions.

Property 3-2

Let us partition $\tilde{\mathbf{x}}$, $\mathbf{m}$ and $\mathbf{M}$ as follows:

$\tilde{\mathbf{x}} = \begin{pmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{pmatrix}$, $\mathbf{m} = \begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \end{pmatrix}$ and $\mathbf{M} = \begin{pmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{pmatrix}$ where $\mathbf{M_{11}}(n_1 \times n_1)$ and $\mathbf{M_{22}}(n_2 \times n_2)$.

If $\tilde{\mathbf{x}} \sim N_n(\mathbf{m}, \mathbf{M})$,

Marginal: $\tilde{\mathbf{x}}_2 \sim N_{n_2}(\mathbf{m}_2, \mathbf{M}_{22})$;

Conditional: $(\tilde{\mathbf{x}}_1 | \tilde{\mathbf{x}}_2 = \mathbf{x}_2) \sim N_{n_1}(\mathbf{m}_{1.2}, \mathbf{M}_{11.2})$ where

$\mathbf{m}_{1.2} = E(\tilde{\mathbf{x}}_1 | \mathbf{x}_2) = \mathbf{m}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2)$,

$$\mathbf{M}_{11.2} = Var(\tilde{\mathbf{x}}_1 | \mathbf{x}_2) = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}.$$

The application of Property 3-2 to our concrete case is straightforward, with $\tilde{\mathbf{x}}_1 = \begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \end{pmatrix}$ and $\tilde{\mathbf{x}}_2 = \begin{pmatrix} \tilde{z}_j \\ \tilde{z}_M \end{pmatrix}$. We finally obtain the distribution of prices conditional to signals:

$$\begin{pmatrix} \tilde{p}_j \\ \tilde{p}_M \end{pmatrix} \Big| \begin{pmatrix} z_j \\ z_M \end{pmatrix} \sim N_2 \left(\begin{array}{l} \mathbf{m}_{1.2} = \begin{pmatrix} E(\tilde{p}_j \mid z_j, z_M) \\ E(\tilde{p}_M \mid z_j, z_M) \end{pmatrix} = \mathbf{m}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\left(\begin{pmatrix} z_j \\ z_M \end{pmatrix} - \mathbf{m}_2 \right), \\ \mathbf{M}_{11.2} = \begin{pmatrix} Var(\tilde{p}_j \mid z_j, z_M) & Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M) \\ Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M) & Var(\tilde{p}_M \mid z_j, z_M) \end{pmatrix} = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21} \end{array} \right)$$

If we develop the conditional expectation and covariance matrices $\mathbf{m}_{1.2}$ and $\mathbf{M}_{11.2}$, we get:

$$\mathbf{m}_{1.2} = \mathbf{m}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\left(\begin{pmatrix} z_j \\ z_M \end{pmatrix} - \mathbf{m}_2 \right) = \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \begin{pmatrix} \sigma_j^2 + \nu_j & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 + \nu_M \end{pmatrix}^{-1} \left[\begin{pmatrix} z_j \\ z_M \end{pmatrix} - \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} \right]$$

$$= \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \frac{1}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \begin{pmatrix} \sigma_M^2 + \nu_M & -\sigma_{j,M} \\ -\sigma_{j,M} & \sigma_j^2 + \nu_j \end{pmatrix} \left[\begin{pmatrix} z_j \\ z_M \end{pmatrix} - \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} \right]$$

$$= \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \frac{1}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} \begin{pmatrix} \sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2 & \sigma_{j,M}\nu_j \\ \sigma_{j,M}\nu_M & \sigma_M^2(\sigma_j^2 + \nu_j) \end{pmatrix} \left[\begin{pmatrix} z_j \\ z_M \end{pmatrix} - \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} \right]$$

$$= \begin{pmatrix} \mu_j + \dfrac{(\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2)(z_j - \mu_j) + \sigma_{j,M}\nu_j(z_M - \mu_M)}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} \\ \mu_M + \dfrac{(\sigma_M^2(\sigma_j^2 + \nu_j) - \sigma_{j,M}^2)(z_M - \mu_M) + \sigma_{j,M}\nu_M(z_j - \mu_j)}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} \end{pmatrix}$$

3.4.4.1 Computation of the conditional expectation

If we develop the first element of this last matrix, we get the conditional expectation of asset *j* price:

$$E(\tilde{p}_j \mid z_j, z_M)$$
$$= \mu_j + \frac{(\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2)(z_j - \mu_j)}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} + \frac{\sigma_{j,M}\nu_j(z_M - \mu_M)}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}$$

$$= \mu_j + \frac{(z_j - \mu_j)}{\frac{\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2 + \nu_j(\sigma_M^2 + \nu_M)}{\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}} + \frac{\sigma_{j,M}\nu_j(z_M - \mu_M)}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} \frac{\nu_j^{-1}\nu_M^{-1}}{\nu_j^{-1}\nu_M^{-1}}$$

$$= \mu_j + \frac{\nu_j^{-1}(z_j - \mu_j)}{\nu_j^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\left(\sigma_M^2 + \frac{1}{\nu_M^{-1}}\right)}}} + \frac{\nu_M^{-1}(z_M - \mu_M)\sigma_{j,M}}{1 + \sigma_j^2\nu_j^{-1} + \sigma_M^2\nu_M^{-1} + (\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2)\nu_j^{-1}\nu_M^{-1}} \quad (3\text{-}3)$$

The impact of a new signal $\tilde{z}_j$ on the conditional expectation of the price of stock j $E(\tilde{p}_j \mid z_j, z_M)$ depends closely on the variance and covariance parameters. In order to make these various effects more obvious, we can reformulate the above Equation (3-3) as follows:

$$E(\tilde{p}_j \mid z_j, z_M)$$

$$= \mu_j + \frac{(z_j - \mu_j)}{1 + \frac{1}{\nu_j^{-1}\sigma_j^2\left(1 - \frac{\rho_{j,M}^2}{\left(1 + \frac{1}{\nu_M^{-1}\sigma_M^2}\right)}\right)}} + \frac{(z_M - \mu_M)\rho_{j,M}}{\left(\frac{1}{\sigma_j} + \sigma_j\nu_j^{-1}\right)\left(\frac{1}{\sigma_M\nu_M^{-1}} + \sigma_M\right) - \sigma_j\sigma_M\rho_{j,M}^2\nu_j^{-1}} \quad (3\text{-}4)$$

The following elements tend to reinforce the impact of a new stock signal z_j (resp. market signal z_M) on the conditional expectation of the price of stock j[27]:

- A high (resp. low[28]) precision of the stock signal ν_j^{-1};
- A low (resp. high) precision of the market signal ν_M^{-1};
- A high (resp. $\frac{1}{\nu_j^{-1}\left(1 - \frac{\rho_{j,M}^2}{\left(1 + \frac{1}{\nu_M^{-1}\sigma_M^2}\right)}\right)}$[29]) unconditional stock variance σ_j^2;

[27] Remember that, ceteris paribus, a increase in the conditional expectation also implies an increase in the stock price itself, as predicted by Equation (3-2) for example

[28] This is more obvious in Equation (3-3) since $\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2 \geq 0$

- A low (resp. $\frac{1}{\nu_M^{-1}\left(1-\frac{\rho_{j,M}^2}{\left(1+\frac{1}{\nu_j^{-1}\sigma_j^2}\right)}\right)}$[30]) unconditional market variance σ_M^2;
- A low (resp. high) unconditional correlation (in absolute value) with the market $\rho_{j,M}$.

By symmetry, we get the following result for the second element of the above matrix, the conditional expectation of the market price:

$$E\left(\tilde{p}_M \mid z_j, z_M\right)$$
$$= \mu_M + \frac{\nu_M^{-1}\left(z_M - \mu_M\right)}{\nu_M^{-1} + \frac{1}{\sigma_M^2 - \frac{\sigma_{j,M}^2}{\left(\sigma_j^2 + \frac{1}{\nu_j^{-1}}\right)}}} + \frac{\nu_j^{-1}\left(z_j - \mu_j\right)\sigma_{j,M}}{1 + \sigma_j^2\nu_j^{-1} + \sigma_M^2\nu_M^{-1} + \left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)\nu_j^{-1}\nu_M^{-1}}$$

3.4.4.2 Computation of the conditional variance and covariance

We compute now the conditional variance-covariance matrix:

$$\mathbf{M}_{11.2} = \begin{pmatrix} Var\left(\tilde{p}_j \mid z_j, z_M\right) & Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M\right) \\ Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M\right) & Var\left(\tilde{p}_M \mid z_j, z_M\right) \end{pmatrix} = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{M}_{21}$$
$$= \mathbf{M}_{11} - \mathbf{M}_{11}\left(\mathbf{M}_{11} + \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}\right)^{-1}\mathbf{M}_{11} = \mathbf{M}_{11}\left[\mathbf{I} - \left(\mathbf{M}_{11} + \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}\right)^{-1}\mathbf{M}_{11}\right]$$
$$= \mathbf{M}_{11}\left[\mathbf{I} - \left(\mathbf{M}_{11} + \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}\right)^{-1}\left(\mathbf{M}_{11} + \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix} - \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}\right)\right]$$

---

[29] Indeed, the coefficient of z_M in Equation (3-4) is a concave function of σ_j^2 and σ_M^2, so that there exists one σ_j^2 and one σ_M^2 which maximize the impact of z_M on the conditional expectation, while null and infinite values of σ_j^2 and σ_M^2 minimize the latter.

[30] Idem previous footnote.

$$= \mathbf{M}_{11}\left(\mathbf{M}_{11} + \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}\right)^{-1}\begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix} = \begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix}\begin{pmatrix} \sigma_j^2 + \nu_j & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 + \nu_M \end{pmatrix}^{-1}\begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}$$

$$= \frac{1}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}\begin{pmatrix} \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_M^2 \end{pmatrix}\begin{pmatrix} \sigma_M^2 + \nu_M & -\sigma_{j,M} \\ -\sigma_{j,M} & \sigma_j^2 + \nu_j \end{pmatrix}\begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}$$

$$= \frac{1}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}\begin{pmatrix} \sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2 & \sigma_{j,M}\nu_j \\ \sigma_{j,M}\nu_M & \sigma_M^2(\sigma_j^2 + \nu_j) - \sigma_{j,M}^2 \end{pmatrix}\begin{pmatrix} \nu_j & 0 \\ 0 & \nu_M \end{pmatrix}$$

$$= \frac{1}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}\begin{pmatrix} \nu_j[\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2] & \sigma_{j,M}\nu_j\nu_M \\ \sigma_{j,M}\nu_j\nu_M & \nu_M[\sigma_M^2(\sigma_j^2 + \nu_j) - \sigma_{j,M}^2] \end{pmatrix}$$

If we work with each element separately, we obtain:

$$Var(\tilde{p}_j \mid z_j, z_M) = \frac{\nu_j[\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2]}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2} = \frac{\nu_j}{\frac{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}{\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}}$$

$$= \frac{\nu_j}{1 + \frac{\nu_j(\sigma_M^2 + \nu_M)}{\sigma_j^2(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}} = \frac{1}{\nu_j^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\sigma_M^2 + \nu_M}}} = \frac{1}{\nu_j^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\sigma_M^2 + \frac{1}{\nu_M^{-1}}}}} \tag{3-5}$$

By symmetry, we have of course that:

$$Var(\tilde{p}_M \mid z_j, z_M) = \frac{1}{\nu_M^{-1} + \frac{1}{\sigma_M^2 - \frac{\sigma_{j,M}^2}{\sigma_j^2 + \frac{1}{\nu_j^{-1}}}}}$$

And the computation of the covariance delivers:

$$Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M) = \frac{\sigma_{j,M}\nu_j\nu_M}{(\sigma_j^2 + \nu_j)(\sigma_M^2 + \nu_M) - \sigma_{j,M}^2}$$

$$= \frac{\sigma_{j,M}}{(\sigma_j^2\nu_j^{-1} + 1)(\sigma_M^2\nu_M^{-1} + 1) - \sigma_{j,M}^2\nu_j^{-1}\nu_M^{-1}} \tag{3-6}$$

Notice that on the basis of this new conditional variances and covariance, we can also rewrite the conditional expectation as follows:

$$
\begin{aligned}
&E(\tilde{p}_j \mid z_j, z_M) \\
&= \mu_j + \nu_j^{-1}(z_j - \mu_j) Var(\tilde{p}_j \mid z_j, z_M) \\
&+ \nu_M^{-1}(z_M - \mu_M) Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M)
\end{aligned} \tag{3-7}
$$

From Equation (3-5), we directly deduce that:

1) The conditional variance is positively related to the unconditional variance σ_j^2. Intuitively, a higher variance prior to the signal leads to a higher variance posterior to the signal, ceteris paribus.
2) On the contrary, the conditional variance is negatively related to the unconditional market variance σ_M^2. Indeed, if the latter is null, the covariance between the market and the stock price is null as well, and any signal about the market is unable to provide any valuable information about stock *j*. On the contrary, as the market volatility increases, the information value of its signals increases and can help to decrease the stock conditional variance. This effect appears more clearly if we decompose the unconditional covariance in the variance equation as follows:

$$
Var(\tilde{p}_j \mid z_j, z_M) = \frac{1}{\nu_j^{-1} + \dfrac{1}{\sigma_j^2 - \dfrac{\sigma_{j,M}^2}{\sigma_M^2 + \dfrac{1}{\nu_M^{-1}}}}} = \frac{1}{\nu_j^{-1} + \dfrac{1}{\sigma_j^2 \left(1 - \dfrac{\rho_{j,M}^2}{1 + \dfrac{1}{\sigma_M^2 \nu_M^{-1}}} \right)}}
$$

   It is clear that if the correlation is maintained constant, an increase of the unconditional market volatility leads to a decrease in the stock *j* conditional volatility.
3) As intuitively expected, the conditional variance is negatively related to the precision of the signal ν_j^{-1}. More precise specific signals decrease the specific risk attached to the stock.
4) It is also negatively related to the market signal ν_M^{-1}, provided that the unconditional covariance between stock *j* and the market $\sigma_{j,M}$ is non null. Intuitively, some of the information contained in the market signal also applies to stock *j* through the covariance mechanism.

5) The conditional variance is negatively correlated to the unconditional correlation (in absolute value) between the stock *j* and the market. Intuitively, the higher the correlation (in absolute value), and the more the investors interested in stock *j* can benefit from the signals coming from the market to infer something about the distribution of the price of stock *j*.

From Equation (3-6), i.e. the conditional covariance equation:

$Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M\right) = \frac{\sigma_{j,M}}{\left(\sigma_j^2 \nu_j^{-1} + 1\right)\left(\sigma_M^2 \nu_M^{-1} + 1\right) - \sigma_{j,M}^2 \nu_j^{-1} \nu_M^{-1}}$, we directly deduce that:

1) The relation between the conditional covariance on the one hand and unconditional variances σ_j^2 and σ_M^2 on the other hand is not monotonous, but a first increasing then decreasing concave function. Please see footnote 16 for detailed explanation.
2) As intuitively expected, and already mentioned by Lambert, Leuz and Verrecchia (2007) under a slightly different form, the conditional covariance is negatively related to the precisions of the stock and market signals ν_j^{-1} and ν_M^{-1}. More precise signals decrease systematic risk.
3) The conditional covariance is positively correlated to the unconditional correlation between the stock *j* and the market.

3.4.5 Extension from one to several stock and market signals

It is possible to relax the assumption of Section 3.4.2.1 of maximum one signal by asset in order to consider Q_j signals on stock *j* and Q_M signals on the market (for example, the number of signals received during a trading day). Appendix A gives a proof for the following adapted equations:

$$
\begin{aligned}
&E\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right) \\
&= \mu_j + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\left(z_{j,q} - \mu_j\right)\right) Var\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right) \\
&+ \left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\left(z_{M,q} - \mu_M\right)\right) Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right) \qquad (3\text{-}8)
\end{aligned}
$$

$$Var(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{M,1},...,z_{M,q},...,z_{M,Q_M})$$

$$= \frac{1}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\sigma_M^2 + \frac{1}{\sum_{q=1}^{Q_M} v_{M,q}^{-1}}}}} \tag{3-9}$$

$$Cov(\tilde{p}_j, \tilde{p}_M \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{M,1},...,z_{M,q},...,z_{M,Q_M})$$

$$= \frac{\sigma_{j,M}}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_M^2\sum_{q=1}^{Q_M} v_{M,q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\right)\left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)} \tag{3-10}$$

From (3-8), (3-9) and (3-10), we see that an increase in the number of signals Q_j (resp. Q_M) has an effect similar to an increase of the precision of the signals $v_{j,q}^{-1}$ (resp. $v_{M,q}^{-1}$).

3.4.6 Link between micro and macro signals

We have explained above that our market signal $\tilde{z}_M$ actuality encompasses two different kinds of signals:

1) A purely and exogenous macro signal $\tilde{z}_{M,q}$ defined as: $\tilde{z}_{M,q} \equiv \tilde{p}_{M,q} + \sqrt{v_{M,q}}\tilde{u}_{M,q}$. Such a signal concerns directly the market and is not the by-product of a signal $\tilde{z}_{k,q}$ on an individual asset;
2) A market signal $\tilde{z}_{M,q}$ which results in fact from the transformation of a $\tilde{z}_{k,q}$ into $\tilde{z}_{M,q}$:

   $\tilde{z}_{k,q} \equiv \tilde{p}_{k,q} + \sqrt{v_{k,q}}\tilde{u}_{k,q} \Rightarrow \tilde{z}_{M,q} \equiv \tilde{p}_M + \sqrt{v_{M,q}}\tilde{u}_{M,q}$

The idea underlying this last transformation is that a signal about the future price of an individual asset *k* is also likely to say something valuable about the future price of the market. If computations were tractable, we would not have to do this transformation, and we would condition the price of stock *j* on signals $\tilde{z}_{j,q}$, on the (real) macro signals

$\tilde{z}_{M,q}$ and on all individual signals $\tilde{z}_{k,q}$ of all assets *k* different from *j*. But we need to reduce the size of the matrix to invert from *[K+1]×[K+1]* to *2×2*, that is why we consider stock *k* signals as market signals.

A natural question arises: how the parameters of the "artificial" inferred signal $\tilde{z}_{M,q}$ depend on the parameters of the initial signal $\tilde{z}_{k,q}$? Specifically, we impose that $v_{M,q}$ is such that $Var(\tilde{p}_M \mid z_{k,q}) = Var(\tilde{p}_M \mid z_{M,q})$, that is, we impose at least that the equivalent $z_{M,q}$ has an effect similar to $z_{k,q}$ on the resulting variance of the conditional distribution of the market price. In other words, we want that conditioning on $z_{k,q}$ or on the artificial $z_{M,q}$ deliver exactly identical results. Note that we could have imagined alternative conditions, but this one seems naturally interesting to explore.

Mathematically, we have (we abandon the index "*q*" for convenience):

$$Var(\tilde{p}_M \mid z_{M,q}) = Var(\tilde{p}_M \mid z_{k,q})$$

$$\Rightarrow \frac{1}{v_M^{-1} + \frac{1}{\sigma_M^2}} = \sigma_M^2 - \frac{\sigma_{k,M}^2}{\sigma_k^2 + \frac{1}{v_k^{-1}}} \Rightarrow v_M^{-1} + \frac{1}{\sigma_M^2} = \frac{1}{\sigma_M^2 - \frac{\sigma_{k,M}^2}{\sigma_k^2 + \frac{1}{v_k^{-1}}}}$$

$$\Rightarrow v_M^{-1} = \frac{1}{\sigma_M^2 - \frac{\sigma_{k,M}^2}{\sigma_k^2 + \frac{1}{v_k^{-1}}}} - \frac{1}{\sigma_M^2} \Rightarrow v_M^{-1} = \frac{\sigma_k^2 v_k^{-1} + 1}{\sigma_M^2 + \sigma_k^2 \sigma_M^2 v_k^{-1} - \sigma_{k,M}^2 v_k^{-1}} - \frac{1}{\sigma_M^2}$$

$$\Rightarrow v_M^{-1} = \frac{\sigma_{k,M}^2 v_k^{-1}}{\sigma_M^2 \left[\sigma_M^2 + \sigma_k^2 \sigma_M^2 v_k^{-1} - \sigma_{k,M}^2 v_k^{-1}\right]} \Rightarrow v_M^{-1} = \frac{1}{\sigma_M^2 \left[\frac{1}{v_k^{-1} \rho_{k,M}^2 \sigma_k^2} + \frac{1}{\rho_{k,M}^2} - 1\right]}$$

$$\Rightarrow v_{M,q} = \sigma_M^2 \left[\frac{1}{\rho_{k,M}^2}\left(\frac{v_{k,q}}{\sigma_k^2} + 1\right) - 1\right] \qquad (3\text{-}11)$$

From this expression, it follows that the variance $v_{M,q}$ of the equivalent market signal is positively related to the variance $v_{k,q}$ of the initial individual signal and to the market unconditional variance σ_M^2, and is negatively related to the unconditional variance σ_k^2 of asset *k* and to the absolute unconditional correlation $\left|\rho_{k,M}\right|$ between this asset and the market.

3.4.7 Final pricing equation

If we plug (3-8), (3-9) and (3-10) in Equation (3-2) (extended to Q signals), we obtain:

$$p_j = \frac{\left\{\begin{array}{l} E\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right) \\ -\frac{b}{N}\left(\begin{array}{l} Cov\left(\tilde{p}_j,\tilde{p}_M \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right) \\ -Var\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right)\varepsilon_j \end{array}\right) \end{array}\right\}}{R_f}$$

$$= \frac{1}{R_f}\left[\begin{array}{l} \mu_j + \dfrac{\sum_{q=1}^{Q_j} v_{j,q}^{-1}\left(z_{j,q}-\mu_j\right)}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \dfrac{1}{\sigma_j^2 - \dfrac{\sigma_{j,M}^2}{\sigma_M^2 + \dfrac{1}{\sum_{q=1}^{Q_M} v_{M,q}^{-1}}}}} + \dfrac{\sigma_{j,M}\sum_{q=1}^{Q_M} v_{M,q}^{-1}\left(z_{M,q}-\mu_M\right)}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_M^2\sum_{q=1}^{Q_M} v_{M,q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\right)\left(\sigma_j^2\sigma_M^2-\sigma_{j,M}^2\right)} \\ -\frac{b}{N}\left(\begin{array}{l} \dfrac{\sigma_{j,M}}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_M^2\sum_{q=1}^{Q_M} v_{M,q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\right)\left(\sigma_j^2\sigma_M^2-\sigma_{j,M}^2\right)} - \dfrac{\varepsilon_j}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \dfrac{1}{\sigma_j^2 - \dfrac{\sigma_{j,M}^2}{\sigma_M^2 + \dfrac{1}{\sum_{q=1}^{Q_M} v_{M,q}^{-1}}}}} \end{array}\right) \end{array}\right]$$

(3-12)

However, this final equation presents two potential flaws for subsequent interpretation.

Firstly, σ_M^2 contains implicitly a $s_j^2\sigma_j^2$ term, since stock *j* is part of the whole market. Therefore, an increase in σ_j^2 implies, all other things being equal, an increase in σ_M^2. For the sake of rigorous interpretation, it would be better to have no relation between the variance of the stock and the variance of the market.

Secondly, the conditional covariance between the stock and the market contains the conditional variance of the stock as well. An increase of the conditional variance of the

stock has consequently two effects, one direct impact through the last term with the deviation and one indirect impact through the conditional covariance with the market. Again, it would be most convenient to disentangle clearly the covariance and the variance effect of an increase of the conditional variance.

That is why we adapt our equations in order to work with the market without stock *j* rather than with the whole market. We define the price of the market excluding stock *j* as in Section 1.3 and Section 1.6.2.4:

$$\tilde{p}_{\underline{M}} = \tilde{p}_M - s_j \tilde{p}_j$$

And the signal on the price of this modified market:

$$\tilde{z}_{\underline{M},q} \equiv \tilde{p}_{\underline{M},q} + \sqrt{v_{\underline{M},q}}\,\tilde{u}_{\underline{M},q}$$

And we condition now on these signals on the modified market rather than on the whole market[31].

$$p_j = \frac{\left\{\begin{array}{l} E\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right) \\ -\frac{b}{N}\left(\begin{array}{l} Cov\left(\tilde{p}_j,\tilde{p}_{\underline{M}} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right) \\ +Cov\left(\tilde{p}_j,s_j\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right) \\ -Var\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right)\varepsilon_j \end{array}\right) \end{array}\right\}}{R_f}$$

$$= \frac{\left\{\begin{array}{l} E\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right) \\ -\frac{b}{N}\left(\begin{array}{l} Cov\left(\tilde{p}_j,\tilde{p}_{\underline{M}} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right) \\ +Var\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j},z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right)\left(s_j-\varepsilon_j\right) \end{array}\right) \end{array}\right\}}{R_f}$$

[31] Concerning the signals initially "*k*" and transformed into macro market signals, assuming that those signals give information on the modified market price rather than on the whole market price makes some sense. However, for the pure macro signals, we could have kept them as signals on the whole market. But it would have obliged us to condition on three types of signals: *j*, whole market and modified market, which would have brought additional complexity in the formulas without giving much interesting supplementary insight.

$$= \frac{1}{R_f}\left[\mu_j + \frac{\sum_{q=1}^{Q_j} v_{j,q}^{-1}\left(z_{j,q}-\mu_j\right)}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2 + \frac{1}{\sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1}}}}} + \frac{\sigma_{j,\underline{M}} \sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1}\left(z_{\underline{M},q}-\mu_{\underline{M}}\right)}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_{\underline{M}}^2 \sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right)}\right.$$

$$\left. - \frac{b}{N}\left(\frac{\sigma_{j,\underline{M}}}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_{\underline{M}}^2 \sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right)} + \frac{s_j - \varepsilon_j}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2 + \frac{1}{\sum_{q=1}^{Q_{\underline{M}}} v_{\underline{M},q}^{-1}}}}} \right)\right] \quad (3\text{-}13)$$

The motivation of the use of a signal on the whole market except *j* rather than on the whole market is purely technical. The purpose is simply to avoid too many interdependencies between the parameters of the pricing equation above, in order to facilitate the subsequent interpretations.

3.5 Interpretations

3.5.1 Impact of micro information on the cost of capital

Two effects must be distinguished: the first concerns the expectation term, while the second concerns the risk term, which consists itself of a covariance term and a variance term.

The impact of the quantity (Q_j) and quality ($v_{j,q}^{-1}$) of signals on the price conditional expectation depends of course on the realization of the signal itself, $z_{j,q}$. If the realization of the signal $z_{j,q}$ is larger than the unconditional expectation μ_j (i.e. $z_{j,q} - \mu_j > 0$), this will be perceived as good news by the investors. In that case, there is an obvious incentive for the company to increase the quantity and the quality of its communication. On the contrary, if the signal is lower than the unconditional

expectation, the opposite prevails. Indeed, the precision of a signal reinforces its impact on the cost of capital.

On the other hand, when the market emits positive signals, the company should rather decrease the precision of its own communication (especially if there is bad news to announce), in order to increase its covariance with the market and to benefit in a larger measure of the good market signaling, and the other way around of course.

If the impact of signaling on the conditional expectation is rather intuitive and straightforward, the corresponding effect on the risk component of the pricing equation is not necessarily so obvious. As the number (Q_j) and/or the precision ($v_{j,q}^{-1}$) of signals released by firm *j* increases, it is clear that the conditional covariance term decreases since σ_j^2, $\sigma_{\underline{M}}^2$ and $\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2$ are all non-negative. The better the communication, the lower the cost of capital (with respect to the covariance term only).

If some deviation $\varepsilon_{j,t}$ from the optimality exists in the aggregate demand of stock *j*, another term appears, i.e. the product of the conditional variance of asset *j* and the global deviation in this asset. Like the conditional covariance, the conditional variance is negatively affected by the number of signals. If the deviation is negative (investors underweight asset *j* in their portfolios on average), the covariance and variance effects sum up and an increase of the number of signals will have a maximal impact on the cost of capital. On the contrary, if the deviation is large enough, the variance effect can become so weak that it may turn out to be opposite to the covariance effect, so that the resulting effect of an increase of the frequency or the quality of signaling on the cost of capital is undetermined and depends on the other parameters. All other things being equal, a large firm member of the S&P 500 index (for example) has a priori less incentive to communicate than a small firm not member of this index, since the first one is likely to present some positive significant deviation ε_j, according to empirical results of Section 2.5.

Notice that in such a framework, an optimal communication frequency does not exist. If the firm has a negative or only slightly positive deviation, its interest is always to communicate as much as it can. If the firm has a large positive deviation, it depends whether the covariance effect is stronger than the variance effect. If yes, the firm has interest to communicate as much as it can; otherwise, it should not communicate at all in order to minimize its cost of capital. But in any case, an increase in the deviation of

an asset diminishes the incentive for this company to communicate much and well, in the sense that a positive deviation has for main consequence to reduce the risk penalty on the price of this asset. And good communication aims precisely at decreasing the magnitude of this risk penalty. It follows that, at any rate, communication efforts are marginally less profitable for companies with a high positive deviation on their stock.

Figure 3-1: The risk component as a function of deviations and signaling

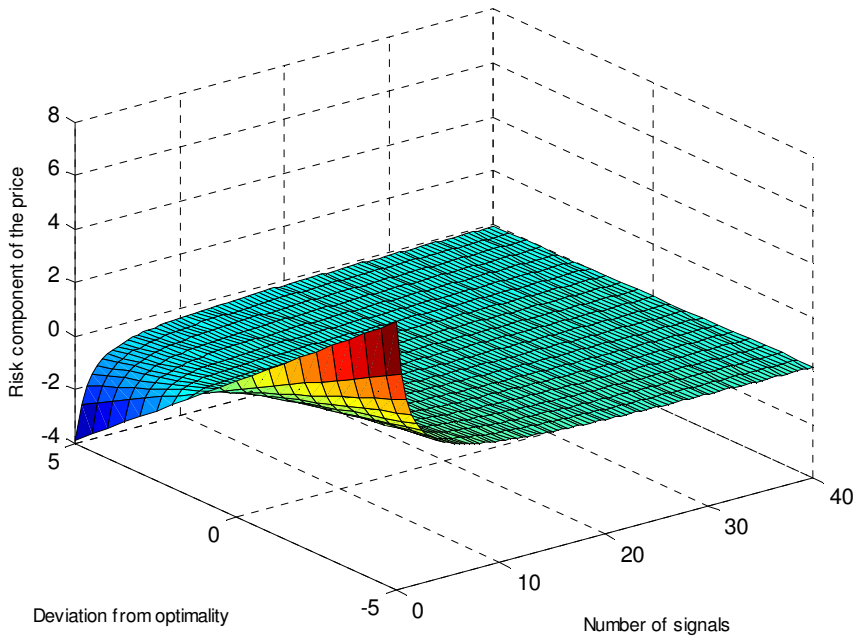


Figure 3-1 shows that there exists a strictly positive "deviation threshold" $\varepsilon_{j,th}$ (graphically situated around 0) above (resp. under) which it is optimal for the firm not to communicate at all (resp. to infinitely communicate) in order to minimize the risk component of its price. By the "risk component" we mean the two terms of the conditional covariance and variance in the above equation of the equilibrium stock price p_j:

$$Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right)$$
$$+Var\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_{\underline{M}}}\right)\left(s_j-\varepsilon_j\right)$$
$$=\frac{\sigma_{j,\underline{M}}}{1+\sigma_j^2\sum_{q=1}^{Q_j}v_{j,q}^{-1}+\sigma_{\underline{M}}^2\sum_{q=1}^{Q_{\underline{M}}}v_{\underline{M},q}^{-1}+\left(\sum_{q=1}^{Q_j}v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_{\underline{M}}}v_{\underline{M},q}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)}+\frac{s_j-\varepsilon_j}{\sum_{q=1}^{Q_j}v_{j,q}^{-1}+\frac{1}{\sigma_j^2-\frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2+\frac{1}{\sum_{q=1}^{Q_{\underline{M}}}v_{\underline{M},q}^{-1}}}}}$$

In order to obtain the above figure, we fix the parameters as follows:

$\sigma_{j,\underline{M}}=10$, $\sigma_j^2=1$, $v_{j,q}^{-1}=1$, $\sigma_{\underline{M}}^2=100$, $v_{\underline{M},q}^{-1}=0.01$, $Q_{\underline{M}}=10$ and $s_j=1$.

Q_j and ε_j are our two variables and they respectively range from 0 to 40 and -5 to 5.

The deviation threshold observed on the figure can be mathematically determined by deriving the price with respect to the global precision of the successive signals on stock j $v_j^{-1}=\sum_{q=1}^{Q_j}v_{j,q}^{-1}$. We also set $\left(\sum_{q=1}^{Q_{\underline{M}}}v_{\underline{M},q}^{-1}\right)$ to be equal to $v_{\underline{M}}^{-1}$ for notational convenience.

We have:

$$\frac{\partial p_j}{\partial v_j^{-1}}=-v_j^2\frac{\partial p_j}{\partial v_j}=\frac{bv_j^2}{NR_f}\frac{\partial\left(\frac{\sigma_{j,\underline{M}}}{1+\sigma_j^2v_j^{-1}+\sigma_{\underline{M}}^2v_{\underline{M}}^{-1}+v_j^{-1}v_{\underline{M}}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)}-\frac{s_j-\varepsilon_j}{v_j^{-1}+\frac{1}{\sigma_j^2-\frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2+\frac{1}{v_{\underline{M}}^{-1}}}}}\right)}{\partial v_j}$$

Remember that the conditional variance and covariance can be expressed under another shape (see Equations (3-5) and (3-6) in their first equality) such that they have the same

denominator:

$$\frac{\partial p_j}{\partial v_j^{-1}} = \frac{bv_j^2}{NR_f} \frac{\partial\left(\frac{\sigma_{j,\underline{M}} v_j v_{\underline{M}}}{\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2} + \frac{v_j\left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]}{\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2}\left(s_j - \varepsilon_j\right)\right)}{\partial v_j}$$

$$= \frac{bv_j^2}{NR_f} \frac{\partial\left(\frac{v_j\left\{\sigma_{j,\underline{M}} v_{\underline{M}} + \left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]\left(s_j - \varepsilon_j\right)\right\}}{\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2}\right)}{\partial v_j}$$

$$= \frac{bv_j^2}{NR_f} \frac{\begin{array}{l}\left\{\sigma_{j,\underline{M}} v_{\underline{M}} + \left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]\left(s_j - \varepsilon_j\right)\right\}\left\{\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right\} \\ - v_j\left\{\sigma_{j,\underline{M}} v_{\underline{M}} + \left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]\left(s_j - \varepsilon_j\right)\right\}\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right)\end{array}}{\left[\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]^2}$$

$$= \frac{bv_j^2}{NR_f} \frac{\left\{\sigma_{j,\underline{M}} v_{\underline{M}} + \left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]\left(s_j - \varepsilon_j\right)\right\}\left\{\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right\}}{\left[\left(\sigma_j^2 + v_j\right)\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]^2}$$

$$= \frac{b}{NR_f}\left\{Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right) + \left(s_j - \varepsilon_j\right) Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)\right\} Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)$$

$$= \frac{bVar\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}{NR_f}\left\{Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right) + \left(s_j - \varepsilon_j\right) Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)\right\} \qquad (3\text{-}14)$$

This last result is remarkable since it shows that the impact of (more and / or better) communication on the asset price is marginally decreasing, since the derivative can be written as follows:

$$\frac{\partial p_j}{\partial v_j^{-1}} = \frac{b}{NR_f} \frac{\left\{\sigma_{j,\underline{M}} v_{\underline{M}} + \left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]\left(s_j - \varepsilon_j\right)\right\}\left\{\sigma_j^2\left(\sigma_{\underline{M}}^2 + v_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right\}}{\left[\left(\sigma_j^2\sigma_{\underline{M}}^2 + \sigma_j^2 v_{\underline{M}} - \sigma_{j,\underline{M}}^2\right) v_j^{-1} + \sigma_{\underline{M}}^2 + v_{\underline{M}}\right]^2}.$$

The absolute value of the above derivative is obviously decreasing in v_j^{-1} since $\left(\sigma_j^2\sigma_{\underline{M}}^2 + \sigma_j^2 v_{\underline{M}} - \sigma_{j,\underline{M}}^2\right)$ is necessarily positive. If the derivative is positive and if delivering signals is costly, there must exist some optimal communication level where the marginal benefit equals the marginal cost of signaling.

It follows from Equation (3-14) that the higher the supply of asset j s_j and the conditional covariance between asset j and the market $Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)$, the larger

the decrease of cost of capital[32] induced by additional or better communication. On the contrary, any positive (resp. negative) deviation in stock j ε_j reduces (resp. reinforces) this positive effect and, if the former is large enough (i.e., larger than a definite threshold that we will determine below), more communication could even result in a price decrease (and consequently an increase of the cost of capital). In such an extreme case, the company has no incentive to communicate at all, and its "optimal" level of communication is null. But this situation is a priori unlikely to occur in practice.

While the risk aversion of investors b tends to increase the absolute magnitude of the marginal effect, the number of investors N and the risk-free rate are likely to reduce the incentive to communicate. Except in some particular cases, the conditional variance of the stock $Var\left(\tilde{p}_j \mid \tilde{z}_j, \tilde{z}_{\underline{M}}\right)$ reinforces the absolute marginal effect of signaling.

We are now able to determine the threshold $\varepsilon_{j,thj}$ such that any deviation above this threshold provides a negative incentive to communicate.

$$\frac{bVar\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}{NR_f v_j^{-1}}\left\{Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right) + \left(s_j - \varepsilon_{j,thj}\right)Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)\right\} = 0$$

$$\Rightarrow \varepsilon_{j,thj} = s_j + \frac{Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)}{Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}$$

$$= s_j + \frac{\sigma\left(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)}{\sigma\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}\rho\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right) \qquad (3\text{-}15)$$

Typically, when the correlation between asset *j* and the rest of the market is positive, the deviation $\varepsilon_{j,thj}$ must at least reach the supply of asset *j* to have a chance to revert the sign of the marginal effect of communication. If we want to get a better intuition of the last term of (3-15), we can rewrite it as follows:

[32] Remember that the higher the current equilibrium price, the lower the cost of capital (see Equation (1-15)).

$$\varepsilon_{j,thj} = s_j + \sqrt{\frac{Var\left(\sum_{\substack{k=1 \\ k\neq j}}^{K} s_k \tilde{p}_k \mid z_j, z_{\underline{M}}\right)}{Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}}\rho\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)$$

$$= s_j + \sum_{\substack{k=1 \\ k\neq j}}^{K} s_k \sqrt{\frac{Var\left(\dfrac{\sum_{\substack{k=1 \\ k\neq j}}^{K} s_k \tilde{p}_k}{\sum_{\substack{k=1 \\ k\neq j}}^{K} s_k} \mid z_j, z_{\underline{M}}\right)}{Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}}\rho\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)$$

$$= s_j + s_j \frac{\sum_{\substack{k=1 \\ k\neq j}}^{K} s_k}{s_j} \sqrt{\frac{Var\left(\tilde{\overline{p}}_{ind-j} \mid z_j, z_{\underline{M}}\right)}{Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}}\rho\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)$$

$$= s_j\left[1 + \frac{1}{weight_{j/\underline{M}}}\frac{\sigma_{ind-j}}{\sigma_j}\rho\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)\right]$$

Where $weight_{j/\underline{M}}$ is the ratio of the supply of asset j to the supply of the rest of the market (the "weight" in the market) and σ_{ind-j} is the standard deviation of the value-weighted market index, except *j*.

Note that the deviation threshold can also be expressed as a function of total market parameters (provided that we condition on a slightly different set of information: signals z_j and z_M):

$$\varepsilon_{j,thj} = \frac{Cov\left(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M\right)}{Var\left(\tilde{p}_j \mid z_j, z_M\right)} = \frac{s_j}{weight_{j/M}}\frac{\sigma_{ind}}{\sigma_j}\rho\left(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M\right)$$

This formulation is more elegant, but the drawback is that σ_{ind} is not independent of σ_j.

It follows that the higher the weight of stock *j* in the market, and the lower the critical threshold of $\varepsilon_{j,thj}$. Since large firms are precisely those that are overweighted by institutional investors and index funds in their portfolio (see the membership value of S&P 500 firms), we expect the incentive to communicate to be especially low, and maybe negative, for those firms.

In that latter case, when the deviation reaches the threshold, the risk component of the price cancels out and we have simply:

$$p_j = \frac{E\left(\tilde{p}_j \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{\underline{M},1},...,z_{\underline{M},q},...,z_{\underline{M},Q_M}\right)}{R_f}$$

If the deviation grows higher, the price becomes increasing in its risk component.

It is difficult to say whether such high deviations exist in the reality. Empirical results of Section 2.5 give some support to the hypothesis that deviations can have substantial effect on prices. But they do not allow us to conclude that deviations could be so high that they could provoke in some extreme cases some kind of inversion of the famous risk-return trade-off: more risk would induce less return.

If we want to explicit in details the different components of this threshold, we can write:

$$\varepsilon_{j,thj} = s_j + \frac{Cov\left(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}\right)}{Var\left(\tilde{p}_j \mid z_j, z_{\underline{M}}\right)}$$

$$= s_j + \frac{\sigma_{j,\underline{M}} \nu_j \nu_{\underline{M}}}{\left(\sigma_j^2 + \nu_j\right)\left(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2} \frac{\left(\sigma_j^2 + \nu_j\right)\left(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2}{\nu_j\left[\sigma_j^2\left(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2\right]}$$

$$= s_j + \frac{\sigma_{j,\underline{M}} \nu_{\underline{M}}}{\sigma_j^2\left(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}\right) - \sigma_{j,\underline{M}}^2} = s_j + \frac{\sigma_j \sigma_{\underline{M}} \rho_{j,\underline{M}} \nu_{\underline{M}}}{\sigma_j^2\left[\sigma_{\underline{M}}^2 + \nu_{\underline{M}} - \sigma_{\underline{M}}^2 \rho_{j,\underline{M}}^2\right]}$$

$$= s_j + \frac{\sigma_{\underline{M}} \rho_{j,\underline{M}}}{\sigma_j\left[1 + \sigma_{\underline{M}}^2 \nu_{\underline{M}}^{-1}\left(1 - \rho_{j,\underline{M}}^2\right)\right]} \qquad (3\text{-}16)$$

An interesting characteristic in the above equation is that the deviation threshold is independent of the precision of the firm communication.

Notice that if $\rho_{j,\underline{M}}$ is positive, $\varepsilon_{j,thj}$ is positive as well, so that firms experiencing a negative deviation should always communicate as much as possible, since their deviation is necessarily lower than the threshold. If the deviation is positive, it may be optimal for the firm (especially if communication is costly) to remain as opaque as possible. It is more likely to be the case if the threshold is low, and thus more particularly when:

- the correlation $\rho_{j,\underline{M}}$ with the rest of the market is low;
- the stock *j* variance σ_j^2 is high (provided that the correlation is positive);
- the number $Q_{\underline{M}}$ and the precision $v_{\underline{M},q}^{-1}$ of the signals released by the rest of the market are high (provided that the correlation is positive);

- the market variance $\sigma^2_{\underline{M}}$ is high (resp. low) if $\sigma^2_{\underline{M}}$ is above (resp. under) a certain threshold $\left((1-\rho^2_{j,\underline{M}})v^{-1}_{\underline{M}}\right)^{-1}$ (provided that the correlation is positive).

3.5.2 Impact of macro information on the cost of capital

The impact of $Q_{\underline{M}}$ and $v^{-1}_{\underline{M},q}$ is symmetric to the effect discussed above of Q_j and $v^{-1}_{j,q}$. A better market signaling will reduce the impact of firm *j* signals $z_{j,q}$ on its cost of capital and reinforce the weight of the market signals $z_{\underline{M},q}$. However, the sensibility of the cost of capital of firm *j* to market signaling closely depends on the magnitude of the volatility of the market $\sigma^2_{\underline{M}}$ and the absolute correlation $\rho_{j,\underline{M}}$ of the latter with asset *j*.

The impact $Q_{\underline{M}}$ and $v^{-1}_{\underline{M},q}$ on the risk component of the cost of capital is identical to the one of Q_j and $v^{-1}_{j,q}$. More and better information about the market decreases both the covariance and the variance of stock *j*. As a consequence, the final effect of a better market signaling will be a reduction in the cost of capital as long as the deviation is not too large (i.e. lower than a critical threshold that will be determined below). The higher $Q_{\underline{M}}$ and $v^{-1}_{j,q}$, the lower the covariance (and thus the cost of capital) and the variance (and thus the lower (resp. higher) the cost of capital if the deviation is not too large (resp. very large)). Small firms or companies with diffuse ownership and not member of heavily traded indices benefit a priori more from general informational transparency than the others.

If we are interested in the marginal effect of market signaling, we can derive the price of stock *j* with respect to the global precision of the market signaling $v^{-1}_{\underline{M}}$:

$$\frac{\partial p_j}{\partial v^{-1}_{\underline{M}}} = \frac{bv^2_{\underline{M}}}{NR_f} \frac{\partial\left(\frac{\sigma_{j,\underline{M}} v_j v_{\underline{M}}}{\left(\sigma^2_j + v_j\right)\left(\sigma^2_M + v_{\underline{M}}\right) - \sigma^2_{j,\underline{M}}} + \frac{v_j\left[\sigma^2_j\left(\sigma^2_{\underline{M}} + v_{\underline{M}}\right) - \sigma^2_{j,\underline{M}}\right]}{\left(\sigma^2_j + v_j\right)\left(\sigma^2_{\underline{M}} + v_{\underline{M}}\right) - \sigma^2_{j,\underline{M}}}\left(s_j - \varepsilon_j\right)\right)}{\partial v_{\underline{M}}}$$

$$= \frac{bv_j v^2_{\underline{M}}}{NR_f} \frac{\partial\left(\frac{\left[\sigma_{j,\underline{M}} + \sigma^2_j\left(s_j - \varepsilon_j\right)\right] v_{\underline{M}} + \left(\sigma^2_j\sigma^2_{\underline{M}} - \sigma^2_{j,\underline{M}}\right)\left(s_j - \varepsilon_j\right)}{\left(\sigma^2_j + v_j\right)\left(\sigma^2_{\underline{M}} + v_{\underline{M}}\right) - \sigma^2_{j,\underline{M}}} \right)}{\partial v_{\underline{M}}}$$

$$= \frac{b\nu_j v_{\underline{M}}^2}{NR_f} \frac{\left[\sigma_{j,\underline{M}} + \sigma_j^2(s_j - \varepsilon_j)\right]\left[(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2\right]}{-\left\{\left[\sigma_{j,\underline{M}} + \sigma_j^2(s_j - \varepsilon_j)\right]\nu_{\underline{M}} + (\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2)(s_j - \varepsilon_j)\right\}(\sigma_j^2 + \nu_j)}$$

$$= \frac{b\nu_j v_{\underline{M}}^2}{NR_f} \frac{\left[\sigma_{j,\underline{M}} + \sigma_j^2(s_j - \varepsilon_j)\right]\left[(\sigma_j^2 + \nu_j)\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right] - (\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2)(s_j - \varepsilon_j)(\sigma_j^2 + \nu_j)}{\left[(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2\right]^2}$$

$$= \frac{b\nu_j v_{\underline{M}}^2}{NR_f} \frac{\sigma_{j,\underline{M}}\left[(\sigma_j^2 + \nu_j)\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right] + \sigma_{j,\underline{M}}^2(s_j - \varepsilon_j)\nu_j}{\left[(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2\right]^2}$$

$$= \frac{b\sigma_{j,\underline{M}}\nu_j v_{\underline{M}}^2}{NR_f} \frac{(\sigma_j^2 + \nu_j)\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2 + \sigma_{j,\underline{M}}(s_j - \varepsilon_j)\nu_j}{\left[(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2\right]^2}$$

$$= \frac{bCov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}{NR_f}\left\{Var(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}) + Cov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})(s_j - \varepsilon_j)\right\} \qquad (3\text{-}17)$$

As $\left|\frac{\partial p_j}{\partial v_j^{-1}}\right|$ was decreasing in v_j^{-1}, we have here that the absolute value of the above derivative with respect to $v_{\underline{M}}^{-1}$ is obviously decreasing in $v_{\underline{M}}^{-1}$ since $(\sigma_j^2\sigma_{\underline{M}}^2 + \sigma_j^2\nu_{\underline{M}} - \sigma_{j,\underline{M}}^2)$ is necessarily positive. However, there is no optimal market communication level for the firm, since the latter does not pay for the signaling of the rest of the market. If the derivative is positive, the more the rest of the market communicates, the better it is for the firm.

It follows from (3-17) that the higher the supply s_j and the variance $Var(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})$ of asset *j*, the larger the decrease of cost of capital induced by additional or better communication. On the contrary, any positive deviation in stock *j* ε_j reduces this positive effect and, if the former is large enough (i.e., larger than a definite threshold that we will determine below), more communication could even result in a price decrease (and consequently an increase of the cost of capital). In such an extreme case, the company suffers from the communication of the rest of the market.

If the risk aversion of investors b tends to increase the absolute magnitude of the marginal effect (positive or negative), the number of investors N and the risk-free rate R_f are likely to reduce the incentive to communicate. Except in some particular cases,

the covariance $Cov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})$ between the stock and the market prices reinforces the absolute marginal effect of signaling.

We are now able to determine the threshold $\varepsilon_{j,th\underline{M}}$ such that any deviation above this threshold causes the cost of capital of the firm to be adversely affected by any market communication:

$$\frac{bCov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}{NR_f}\left\{Var(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}}) + Cov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})(s_j - \varepsilon_{j,th\underline{M}})\right\} = 0$$

$$\Rightarrow \varepsilon_{j,th\underline{M}} = s_j + \frac{Var(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}{Cov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}$$

$$= s_j + \frac{\sigma(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}{\sigma(\tilde{p}_j \mid z_j, z_{\underline{M}})}\frac{1}{\rho(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})} \qquad (3\text{-}18)$$

Equation (3-18) is almost identical to Equation (3-15), so that the same comments apply, with one noticeable exception: $\left|\varepsilon_{j,th\underline{M}}\right|$ is a decreasing function of the correlation, while $\left|\varepsilon_{j,thj}\right|$ is a positive function of the correlation (provided that the latter is positive). Interestingly, $\left|\varepsilon_{j,th\underline{M}} - s_j\right|$ is always larger than $\left|\varepsilon_{j,thj} - s_j\right|$.

Note also that $\varepsilon_{j,thM}$ (i.e. the threshold in a context of signals on the total market, including firm *j*) is in fact equal to the inverse of the conditional beta of the firm:

$$\varepsilon_{j,thM} = \frac{Var(\tilde{p}_M \mid z_j, z_M)}{Cov(\tilde{p}_j, \tilde{p}_M \mid z_j, z_M)} = \beta_j^{-1}$$

If we want to explicit in details the different components of this threshold $\varepsilon_{j,th\underline{M}}$, we can write:

$$\varepsilon_{j,th\underline{M}} = s_j + \frac{Var(\tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}{Cov(\tilde{p}_j, \tilde{p}_{\underline{M}} \mid z_j, z_{\underline{M}})}$$

$$= s_j + \frac{\nu_{\underline{M}}\left[\sigma_{\underline{M}}^2(\sigma_j^2 + \nu_j) - \sigma_{j,\underline{M}}^2\right]}{(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2}\frac{(\sigma_j^2 + \nu_j)(\sigma_{\underline{M}}^2 + \nu_{\underline{M}}) - \sigma_{j,\underline{M}}^2 \nu_j}{\sigma_{j,\underline{M}}\nu_j\nu_{\underline{M}}}$$

$$= s_j + \frac{\sigma_{\underline{M}}^2(\sigma_j^2 + \nu_j) - \sigma_{j,\underline{M}}^2}{\sigma_{j,\underline{M}}\nu_j}$$

$$= s_j + \frac{\sigma_{\underline{M}}\left[1+\sigma_j^2 v_j^{-1}\left(1-\rho_{j,\underline{M}}^2\right)\right]}{\sigma_j \rho_{j,\underline{M}}} \tag{3-19}$$

An interesting characteristic in the above equation is that the deviation threshold is independent of the precision of the market communication.

Notice that if $\rho_{j,\underline{M}}$ is positive, $\varepsilon_{j,th\underline{M}}$ is positive as well, so that firms experiencing a negative deviation always benefit from the market communication, since their deviation is necessarily lower than the threshold. If the deviation is positive, market communication may have an adverse impact on the price of the firm. It is more likely to be the case if the threshold is low, and thus more particularly when:

- the correlation $\rho_{j,\underline{M}}$ with the rest of the market is high;
- the stock variance σ_j^2 is close to $\left((1-\rho_{j,\underline{M}}^2)v_j^{-1}\right)^{-1}$ (provided that the correlation is positive).
- the number $Q_{\underline{j}}$ and the precision $v_{j,q}^{-1}$ of the signals released by the firm are low (provided that the correlation is positive);
- the market variance $\sigma_{\underline{M}}^2$ is low (provided that the correlation is positive).

3.5.3 Impact of the unconditional variance of firm *j* on the cost of capital

As it has already been said before, a high unconditional variance σ_j^2 tends to increase the conditional variance of the stock and to reinforce the impact of signals $z_{j,q}$ while an unconditional variance σ_j^2 close to $\frac{1}{v_j^{-1}\left(1-\frac{\rho_{j,M}^2}{\left(1+\frac{1}{v_M^{-1}\sigma_M^2}\right)}\right)}$ tends to increase the conditional covariance and to reinforce the impact of markets signals $z_{M,q}$.

The final impact on the cost of capital will also depend on the magnitude of the deviation. The combination of a high initial volatility and a high positive deviation

(larger than the threshold) may be a powerful incentive to prefer opacity to transparency.

3.5.4 Impact of the unconditional variance of the market on the cost of capital

As it has already been said before, a high unconditional market variance $\sigma^2_{\underline{M}}$ tends to decrease the conditional variance of the stock and to reduce the impact of signals $z_{j,q}$ while an unconditional variance $\sigma^2_{\underline{M}}$ close to $\dfrac{1}{\nu_{\underline{M}}^{-1}\left(1-\dfrac{\rho^2_{j,\underline{M}}}{\left(1+\dfrac{1}{\nu_j^{-1}\sigma_j^2}\right)}\right)}$ tends to increase the conditional covariance and to reinforce the impact of markets signals $z_{M,q}$.

The final impact on the cost of capital will also depend on the magnitude of the deviation.

3.5.5 Impact of the unconditional correlation on the cost of capital

As it has already been said before, a high unconditional correlation between stock *j* and the market tends to increase the conditional covariance and to reinforce the impact of markets signals z_M, while a high absolute value of the unconditional correlation tends to decrease the conditional variance of the stock and to reduce the impact of signals z_j.

The final impact on the cost of capital will also depend on the magnitude of the deviation.

3.6 Conclusion

From the model developed above, it appears that the impact of signaling on the cost of capital turns out to be not as straightforward as we could have believed it at first sight. This is especially true in a universe of constrained investors, where agents may deviate,

for any reason, from their “optimal” CAPM holdings. The final equations deliver many interesting insights, and more particularly two facts are worth noticing:

Even when communication decreases the risk component of the price, which is usually the case when the deviation is not too large, there must exist some optimal communication level if the communication is costly. Indeed, the marginal benefit of signaling is decreasing with the intensity and precision of the information disclosed. The optimum is such that the marginal benefit equals the marginal cost of signaling.

In some extreme cases, one may observe a totally counter-intuitive phenomenon. If the deviation is larger than a critical threshold, the firm could have no incentive to communicate anymore, and could find optimal to opt for opacity rather than transparency. At any rate, even if this threshold is not reached, the incentive to communicate can strongly vary from one firm to another, depending on its own covariance and variance characteristics, and each firm should determine the communication policy that is the most adequate to its proper situation.

Our model has been built in a static two-period framework. In a dynamic framework, predictions may be different. For example, the forthcoming empirical estimation gives evidence that more communication is also generally associated with higher returns volatility. At first sight, this may seem contradictory with the conclusion of this section, but one must not forget that the theoretical and empirical settings are not directly comparable, since the first one models prices in a two-period framework, while the second one focuses on returns in a dynamic framework. Furthermore, the next section will also show that all announcements do not produce identical effects, because the timing and the topic of the news matter. Such elements can hardly be taken explicitly into account in our theoretical model.

4 Empirical estimation of the impact of public information on the conditional volatility of stock returns

This empirical section aims at investigating the link between news intensity and conditional volatility. It first focuses on a "French" sample of the 40 companies belonging to the CAC 40 index at the time of study. Literature and econometrical techniques are developed in detail and results are extensively commented. Subsequently to this French study, which has been published in Cousin and de Launois (2006), we have conducted a similar study on a wider sample of 500 companies belonging to the S&P 500 at the time of study. Since literature and underlying econometrics are identical in both French and US applications, we will not repeat them in the latter, but rather restrict the US section to comments specific to this sample and the results attached to it.

4.1 The French Case

4.1.1 Introduction

The relation between information flow and asset prices behavior is one of the key issues of modern finance. How, why and under which conditions do financial news impact stock price volatility, count among the classical (and still debatable) questions in this research area. The clear knowledge and the good understanding of such important issues are of first importance for listed companies. The behavior of stock prices around profit warning announcements is a typical (and probably extreme) example of these phenomena. Serge Tchuruk, Alcatel's CEO, still probably remembers the disastrous consequences of his injudicious profit warning of September 17, 1998. At first sight, the news release looked quite innocent: "The operational result of the telecommunications sector (…) will suffer from the large decrease in investments (…) and from the crisis in South-East Asia and in Russia. Alcatel reckons that the operational performance in 1998 will not reach the expected level." Among other things, Alcatel said it expected to make a profit in 1998 of €1.1 billion, down sharply from an earlier forecast of €1.4 billion. The market reaction was terrible: shares of the company lost 38% in one day, involving that not less than €10 billion of Alcatel's market capitalization vanished in a couple of hours. Never had the "Paris bourse" experienced such a dramatic stock crash.

That kind of news hardly makes only one victim. Sector spillover effects of such announcements have been quite common. On September 16, 2004, after Celestica, a large-cap electronics manufacturer, cut its outlook for the third quarter, the whole sector of makers of electric parts and equipment as well as providers of related services dropped: Power-One tumbled 6.4%, Plexus fell 5.4%, Littlefuse shed 3.3% and Benchmark Electronics lost 10% on the New York Stock Exchange. The communications-technology sector was affected the same way, since Stratex Networks dropped 4.4%, Mindspeed Technologies lost 3.3% and MRV Communications fell 3.2%. Financial communication may thus strongly impact stock return, volume and volatility of any company. Such a point deserves to be studied and analyzed carefully.

Our primary objective is to investigate the relation between public information and stock volatility. To what extent does the discontinuous frequency of news releases affect the volatility of the daily stock returns? And what are the factors strengthening or weakening this relation? These are precisely the main questions that we aim to answer in this section. They are crucial from both a practical and an academic perspective, since recent academic works have put into light the importance of developing a better understanding of the determinants of return volatility. Easley and O'Hara (2001) bring some evidence that the less transparent and disclosing a company is, the higher its cost of capital will be. Such a conclusion already offers some insight of how significant the influence of news releases (in quality and in quantity) on stock prices can be. Campbell, Lettau, Malkiel and Xu (2001) find that the proportion of firm-specific variance in the total stock variance has steadily been increasing for the last thirty years and thus provides us with another good reason to investigate the relation between firm-specific news and idiosyncratic volatility. As a matter of fact, one may wonder whether the regular increase in idiosyncratic variance detected by Campbell and al. (2001) could possibly be linked with the obvious contemporaneous increase in the financial communication intensity. Our results seem to support such an assertion. Eventually, Goyal and Santa Clara (2003) detect a positive relation between idiosyncratic variance and expected stock return. For idiosyncratic risk is not perfectly diversifiable and investors are assumed risk-averse, an increase in the firm-specific part of the stock volatility can induce an increase in the cost of capital[33].

[33] However, this point remains controversial. In obvious contradiction with Goyal and Santa Clara (2003), Wei and Zhang (2005) claim that idiosyncratic risk does not matter.

Since we focus on the stock volatility, an adequate modelisation of its dynamics is obviously of prime importance. If GARCH models count undoubtedly among the most famous conditional volatility models, alternative frameworks exist and deserve to be considered. The Markov switching regression (MSR) approach is one of them: this rather original approach avoids the classical drawbacks of the GARCH models and allows us to indirectly test the Mixture of Distribution Hypothesis (MDH), which argues that the stock volatility is directly related to the information frequency.

Although GARCH and MSR models give useful and quite reliable estimates of the conditional stock variance, they still remain reduced form models not dedicated to explain observed changes in stock volatility and to point out the core factors driving its behavior. In our quest for an innovative explicative model of the stock volatility, the MDH may well be of great help, since it suggests explicitly that the stock volatility is closely linked to the information frequency. Therefore it largely justifies the introduction of the "daily number of press releases on a stock" as the most appropriate explanatory variable in the basic equation of our intended explicative model.

With regard to the existing literature, our contributions are of three kinds: they concern the data used as well as the methodology adopted and the results obtained.

In order to test empirically whether information flows drive stock return volatility, we apply our econometrical model to all stocks pertaining to the CAC40 index from January 1, 1999, to December 31, 2003. Datastream provides us with the series of daily stock prices, while Factiva gives us access to a particularly rich and advantageous news sample composed of all time-stamped societal and industrial Reuters news releases concerning all companies selected above. This kind of financial communication counts undoubtedly among the most up to date and relevant public information that investors can collect when trading on the market. Furthermore, since we are only interested in the impact of information intensity on the idiosyncratic stock variance, we restrict our investigation to firm-specific and industrial-specific news and chose not to introduce any market or macroeconomic news in the sample (as it will be explained below, we will control for possible market and macroeconomic shocks on stock variance through controlling for the market variance in the regression equation). Finally, for we want to test the link between the daily number of firm-specific news releases and the corresponding stock volatility, it is necessary to know which company (or companies) is (or are) concerned by each news release. Our sample provides us with such an

interesting item, among many others. It allows us to refine the classical aggregate market-level analysis into a more subtle and accurate firm by firm study.

Our methodology differs in many respects from the previous studies on the topic mentioned in Section 4.1.2. We analyze the impact of news on volatility at the firm-level, i.e. company by company, and not at the market-level, i.e. impact of news on the index. It allows us to capture the entire "firm-specific effect" contained in the news which is obviously lost at the aggregate level. We also employ conditional volatility measures (instead of unconditional volatility measures) in order to take the volatility persistence effect into account. We first replicate on a daily basis the GARCH approach of Kalev, Liu, Pham and Jarnecic (2004) with our specific data set at the firm-level. We then opt for a market model with two (Markov switching) regimes of volatility, which is more robust and general than the GARCH model used by Kalev and al. (2004) because it needs less strong assumptions on the conditional volatility process. Moreover, such a MSR model allows us to explicitly take advantage of the central Roll (1988)'s idea[34] of a mixed return distribution ("with news" and "no news") driven by news arrivals and to test the relevance of such an intuition. The subsequent econometric methodology consists basically of two steps: estimating the daily probabilities of being in a high volatility state thanks to the developed two-state market model and regressing these obtained probabilities on some possibly explicative variables (like the daily number of news wires) in order to understand the determinants driving the stock returns volatility. We also perform a panel analysis to better judge of the overall significance and robustness of our results. Like Kalev and al. (2004), we control for the daily market volume of trades though we find it to be insignificant. Unlike Kalev and al. (2004), we control for factors such as the market conditional volatility and the sector conditional volatility, which turn out to be highly significant. We also choose to perform the same regressions on different news subsets selected on the basis of various criteria, like the subject or the timing of the news release. In a last and deeper analysis, we investigate the asymmetry issue and we propose a measure of the informational content of the news releases.

Our main finding is that public information flows strongly affect the stock returns distribution. In the GARCH framework, our results are consistent with those of Kalev and al. (2004). The news frequency seems to be largely responsible for the persistence

[34] The MDH in fact.

volatility effect, and its impact on the volatility level is substantial. In the MSR model framework, we find a positive and significant impact of the daily frequency of information arrivals on the probability to be in a state of high volatility for each of the 40 companies considered. The subsequent model for panel data confirms this result and demonstrates that this impact is marginally decreasing. The regressions on various data subsets offer very interesting results: news released during trading hours seems to have a much stronger impact than news released during non-trading hours. In terms of subject, press releases concerning *regulation and government policy*, *earnings projections* (especially the subset of *analyst comments and recommendations*) and *mergers and acquisitions* (in order of decreasing importance) appear to have the greatest impact on the stock volatility. Eventually, the investigation of the asymmetry issue and the development of a news informational content proxy lead us to conclude that one good and one bad announcement with identical informational content affect the conditional volatility the same way, so that the classical asymmetry effect reported by the literature simply disappears.

This section is organized as follows. Section 4.1.2 provides a brief review of the literature about the relation between information and conditional volatility. Section 4.1.3 describes the used data sets: stock prices and Reuters news wires. Section 4.1.4 presents the GARCH and the Markov switching regimes models used in the subsequent empirical part. Section 4.1.5 displays and comments results. Section 4.1.6 concludes and provides some perspectives for future research.

4.1.2 Information and conditional volatility

4.1.2.1 Modeling the conditional volatility

In the literature, the time-varying nature of the volatility is a well-known empirical fact. Presence of conditional heteroscedasticity is well documented in most financial time-series. As a natural consequence, several authors have tried to account for this volatility persistence effect in alternative econometric frameworks like ARCH (see Engle (1982)) and GARCH[35] (see Bollerslev (1988)) approaches. By expressing the stock idiosyncratic variance at time *t* as a linear function of the past stock idiosyncratic

[35] ARCH and GARCH stand respectively for autoregressive conditional heteroscedasticity and generalized autoregressive conditional heteroscedasticity (models).

variances and innovations (at *t-1*, *t-2*, etc.), these models provide conditional estimates of the stock volatility while taking the persistence volatility effect into account. Markov switching regression models (MSR) appear as another interesting way to capture conditional heteroscedasticity evolutions. In such a non-linear dynamic framework, stock returns distribution is supposed to switch from a low to a high volatility regime (and vice-versa) according to some fixed transition probabilities. MSR models exhibit a more general structure than the GARCH models, since they impose fewer assumptions on the functional form of the conditional volatility. Moreover, GARCH models may sometimes present some annoying convergence problems, while MSR models are generally more reliable in this respect. Eventually, GARCH models do not rest on a clear theoretical justification (they are sort of reduced form models), whereas MSR models are implicitly based upon such a powerful argument, which is known as the Mixture of Distribution Hypothesis (MDH). To our knowledge, it appears for the first time in Clark (1973) (but not with this specific denomination), who develops a subordinated stochastic process model where the distribution of price change is subordinate to the normal distribution of the information flow (the "directing process"). Although we will not use in this study the subordinated process model, it is important to mention this seminal work of Clark (1973) since his hypothesis has many testable empirical implications interesting for us. Indeed, the MDH posits that the joint distribution of daily return and volume can be modeled as a mixture of bivariate normal distributions. Specifically, they are contemporaneously dependent on an underlying mixing variable that represents the flow of information. As a consequence, the variance of returns at a given interval is expected to be proportional to the rate of information arrival at the market. About fifteen years later, Roll (1988) shows empirically that the sample variance and the kurtosis for example can reveal something about the probability of information and the difference between the information-related distribution and the non-information-related distribution of returns, which is totally consistent with the MDH. One year later, Ross (1989) brings some convincing theoretical evidence supporting the MDH and demonstrates that in an arbitrage-free economy, the volatility of prices is directly related to the rate of flow of information to the market. An avowed objective of our study is precisely to test empirically that mixture of distribution hypothesis and to check whether the stock returns distribution is driven by the frequency of information arrivals on the Paris stock Exchange (Euronext).

In this context, the MSR model turns out to be the most suitable specification, since it is the closest translation of the mixture of distribution hypothesis into an empirical estimation. Chauvet and Hamilton (2005) shows that a simple Markov switching model fits particularly well the data when these are suspected to follow a mixture of distributions. In our case, Davies's test confirms that a Markov switching model with two regimes must be preferred to a simple market model. We could consider more than two regimes, but in absence of strong and annoying outliers, the addition of a third regime does not bring anything valuable and will only increase the complexity of the model.

4.1.2.2 News arrivals and market reactions

If many authors tried to establish the reality of the positive relation between flows of information and stock (or market) volatility, the empirical evidence of such a positive relation is everything but overwhelming, probably because of the difficulty to find good empirical proxy of information arrivals and of the poor volatility estimates that were traditionally used. The difficulty to precisely measure information arrivals appears in the variety of proxies employed by previous empirical studies on the topic. Berry and Howe (1993) use the number of daily newspaper headlines and earnings announcements and Ederington and Lee (1993) investigate the importance of macroeconomic news, whereas Mitchell and Mulherin (1994) employ the number of specific stock market announcements in order to test the impact of the rate of information on the market volatility. « However, the use of unconditional volatility measures such as absolute daily market returns in these studies often generates weak or inconclusive results regarding the news–volatility relation »[36]. Indeed, such a proxy is known to be quite noisy, and the presence of conditional heteroscedasticity in the returns time-series may significantly alter the quality of the results. It is precisely to account for this well-known phenomenon of conditional heteroscedasticity that Kalev and al. (2004) choose to test the relation between firm-specific announcements (as a proxy for information flows) and volatility on the Australian Stock Exchange in a GARCH framework. Their analysis reveals a positive and significant impact of the arrival rate of the selected news variable on the conditional variance of stock returns, even after controlling for the potential

36 Kalev, Liu, Pham and Jarnecic (2004), p. 1442

effects of trading volume and high opening volatility. They find the respective coefficients of these two last control variables to be positive and significant, except on a daily basis. Eventually, they split all their press releases into different categories according to their subject, since Andersen (1996) argues that different types of news have different stochastic arrival processes, so that we may legitimately expect them to have a different impact on the conditional stock volatility.

It is worth noticing that some authors have worked specifically on the French market, like Lardic and Mignon (2003) and Lespagnol and Teïteltche (2005).

4.1.3 Sample selection

Our firms' universe is composed of the 56 (40 after deletion of acquired, merged or vanished firms) CAC40 index stocks that appear at least once in the CAC40 index between January 1999 and December 2003. It provides us with a broad and representative sample of the French Stock Market since it accounts for not less than 70% of the French stock market value. The total market value of the index on November 11, 2003 was €722 billion. The mean (median) company market value was €18 billion (€13 billion). For all firms included in this index, we use daily closing prices from January 1, 1999 to December 31, 2003. The three largest sectors in the index were the financials (22.29%), resources (oil) (18.27 %) and non cyclical consumers' goods (11.82%). We use as a market portfolio the CAC40 index. The data are obtained from Datastream, accessed at the university of Lille 2.

All press releases were collected through the Factiva data base web site[37]. Factiva is a Dow Jones and Reuters company that provides professionals with business news and information sources like The Wall Street Journal or The Financial Times for example. Our sample is composed of all corporate and industrial news wires concerning firms pertaining to the CAC40 index and recorded by Reuters from January 1, 1999 to December 31, 2003. For each news wire, we have got the following fields: the accession number (AN), headline (HD), word count (WC), publication date (PD) and time (ET), source name (SN) and code descriptor (SC), lead paragraph (LP), company (concerned by the news release) code (CO), industry code (IN), subject code (NS), region code (RE), Dow Jones codes (DJIC) and descriptors (DJID), information provider codes

[37]http://www.factiva.com/

(IPC) and Reuters codes (RBBCM). Company, industry and subject codes are the main fields of interest, since they allow us to classify each news release in sector and in topic. One piece of news can of course concern several companies, industries and subjects.

To avoid any redundancy and duplicate announcements that do not bring any additional information value, we restrict the sample to news released by Reuters only. For the same reason, we eliminate all news releases with the same headlines and lead paragraphs. Eventually, after having excluded news items with one missing field or more, we are left with 76341 financial communications over the whole five-year-period.

Figure 4-1 (resp. Figure 4-2) displays the evolution over time of the total daily (resp. weekly) number of news wires. At first sight, the absence of any clear trend indicates that the news time-series is quite stationary[38]. It could seem a bit singular, as we may have expected the frequency of news arrivals to gradually increase over the years. This particular feature of our data presents the advantage to eliminate the risk of spurious results due to a possible simultaneous increase over time of the stock volatility. Another interesting fact is that there is some kind of unexplained sinusoidal cycle (perhaps completely artificial and virtual, without any real meaning), with 2000 and 2002 as higher news frequency years and 1999, 2001 and 2003 as lower news frequency years. Not surprisingly at all, the Christmas-New Year period experiences a news arrivals rate much below the average.

Figure 4-3 focuses on a possible and expected weekly seasonality in the data and shows indeed that the average number of news announcements released during the week-end is much lower than the one of the other weekdays. Since the stock exchange is closed during the week-end, this seasonality is not problematic: news released during the week-end will simply be aggregated to news published on the next Monday. It can also be noticed that the number of announcements released on each weekday is roughly the same.

Fields like company, industry and subject deserve to be examined in detail. Table 4-1, Table 4-2 and Table 4-3 provide an extensive overview of what these fields precisely content. They respectively give the Factiva classification of companies, industries and topics.

[38] We have checked this stationarity hypothesis more formally through the unit root test of Dickey-Fuller that rejects the hypothesis of unit root at a 0.1% level. So we can conclude that the series is stationary.

Figure 4-1: Daily number of news releases on the French market from Jan. 1999 to Dec. 2003

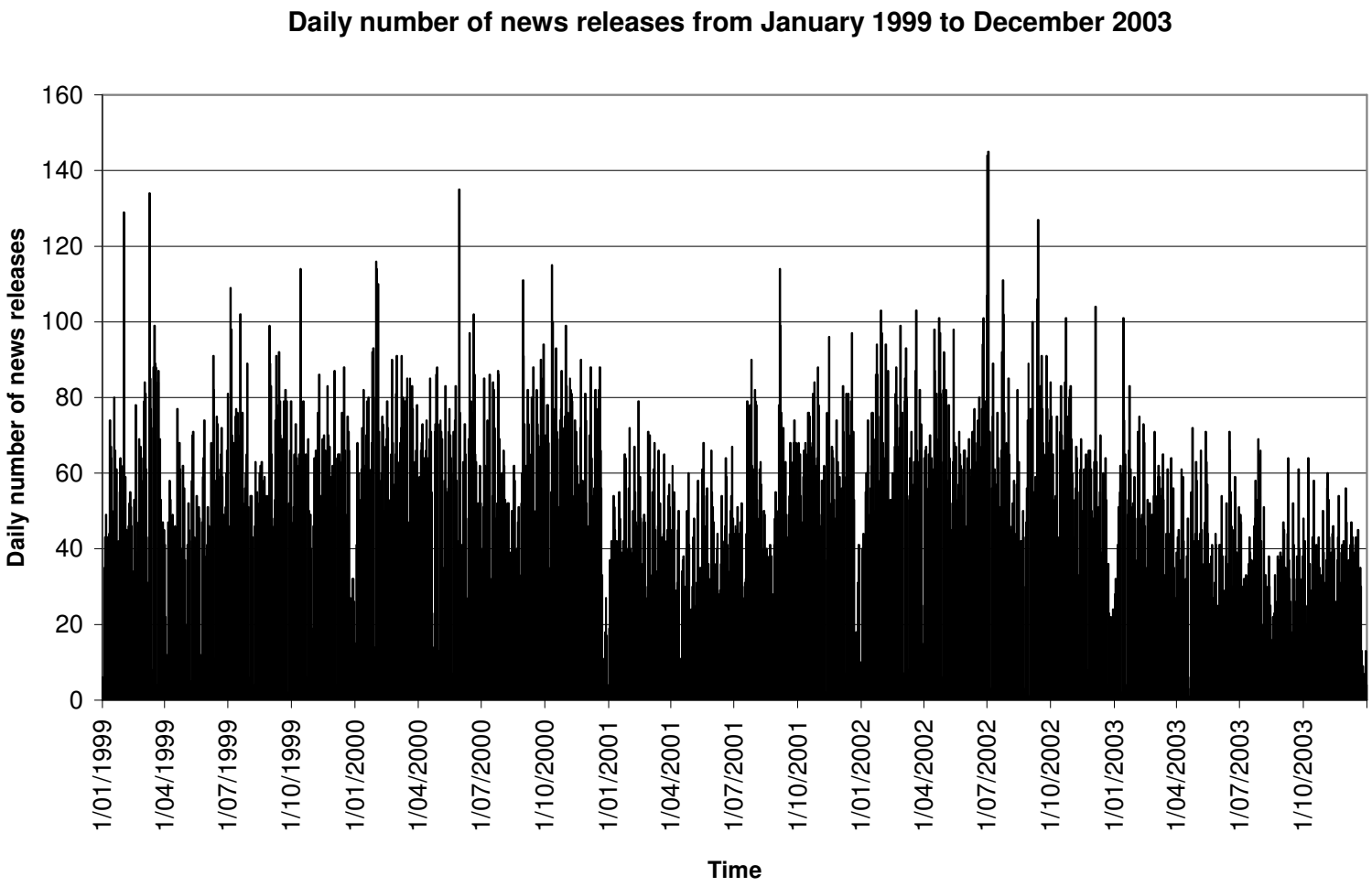


Figure 4-2: Weekly number of news releases on the French market from Jan. 1999 to Dec. 2003

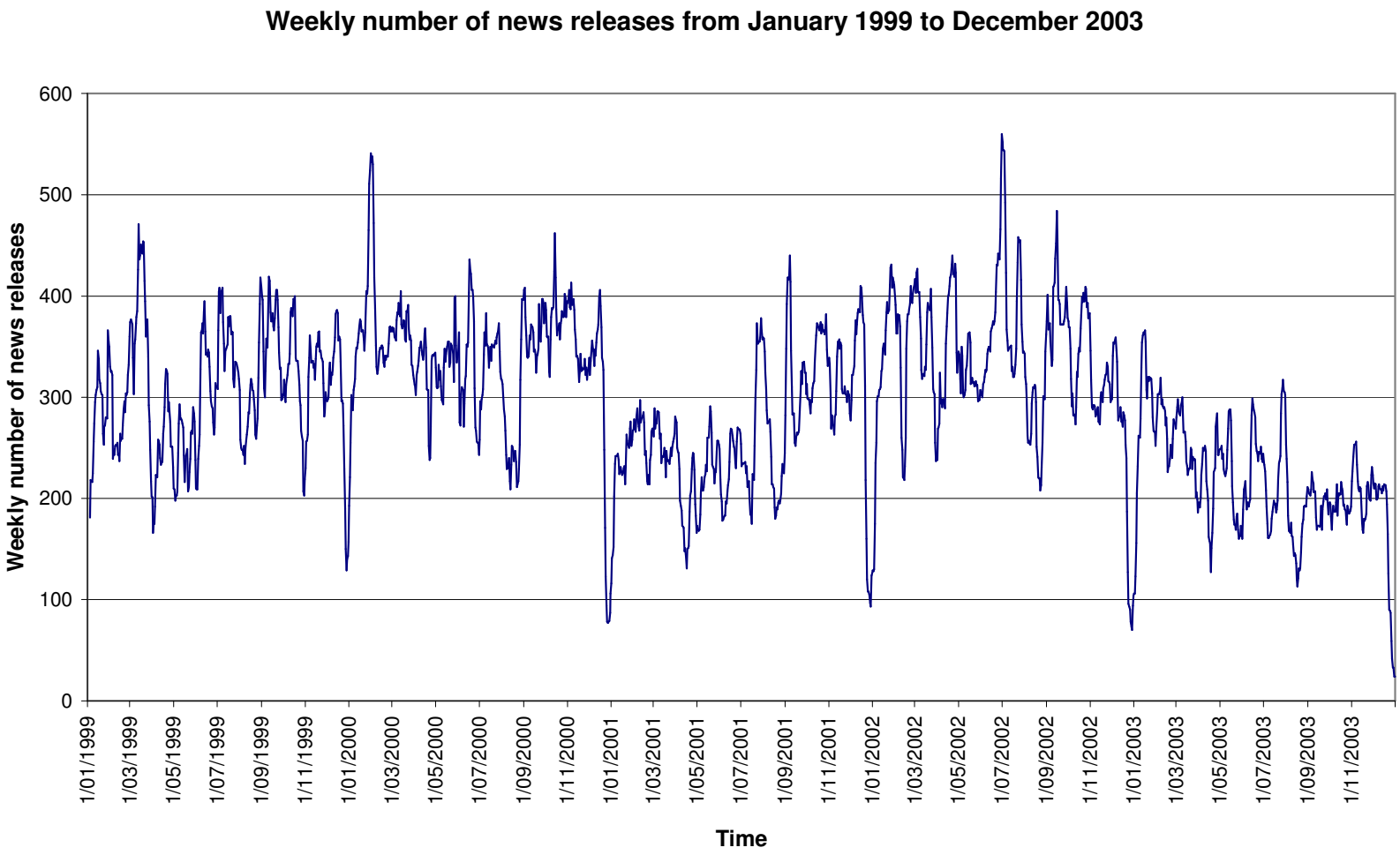


Figure 4-3: Total news releases average by day of the week on the French market

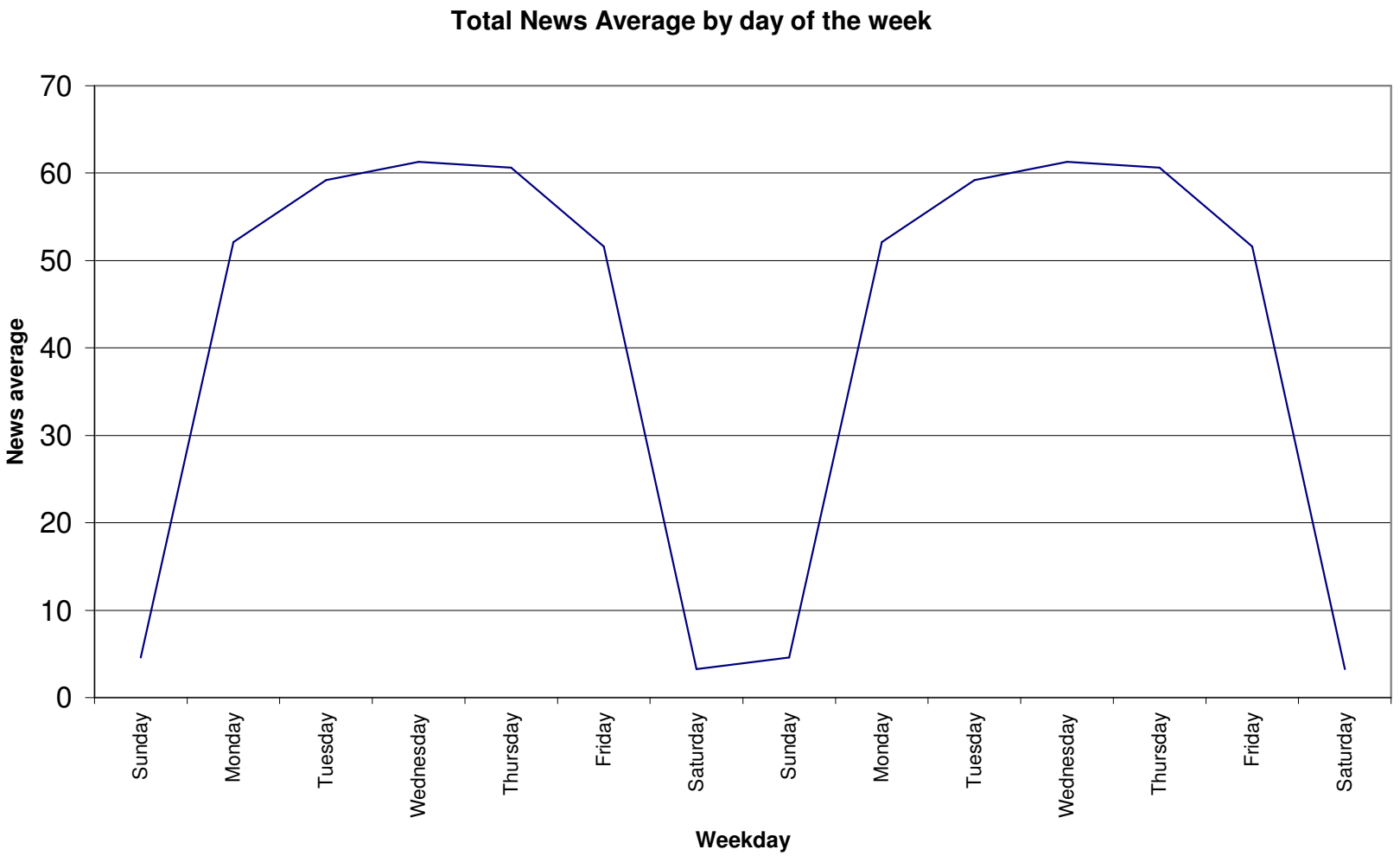


Table 4-1: Factiva classification of companies ("CO" field) on the French market

CO	Company Name
NOVTL	ACCOR (FRA)
NIA	EUROPEAN AERONAUTIC DEFENCE (NETH)
AGF	AGF ASSURANCES (FRA)
LAIRLQ	AIR LIQUIDE,L' (FRA)
CGE	ALCATEL (FRA)
GECALS	ALSTOM (FRA)
ARCELO	ARCELOR SA (LUX)
MIDI	AXA SA (FRA)
BIC	BIC,SOCIETE (FRA)
BANQNP	BNP PARIBAS (FRA)
BOYGUE	BOUYGUES (FRA)
CNPL	CANAL PLUS (FRA)
CGS	CAP GEMINI SA (FRA)
CARFOR	CARREFOUR (FRA)
ECGP	CASINO GUICHARD PERRACHON (FRA)
CCF	CREDIT COMMERCIAL DE FRANCE (FRA)
CRDTAG	CREDIT AGRICOLE SA (FRA)
CRDLYN	CREDIT LYONNAIS SA (FRA)
BSNGD	DANONE,GROUPE (FRA)
DSSYM	DASSAULT SYSTEMES SA (FRA)
GMNKDT	DEXIA (BELG)
ELFAQT	ELF AQUITAINE (FRA)
SITR	EQUANT NV (NETH)
BGHNSY	ERIDANIA BEGHIN SAY SA (FRA)
DGT	FRANCE TELECOM SA (FRA)
LAFAR	LAFARGE (FRA)
MMBFRA	LAGARDERE SCA (FRA)
LEGRND	LEGRAND SA (FRA)
VUIMTH	LVMH MOET HENNESSY LOUIS VUITTON (FRA)
MICHLN	MICHELIN (FRA)
ORNGSA	ORANGE SA (FRA)
ORL	L'OREAL SA (FRA)
BANQDP	PARIBAS (FRA)
PRNOD	PERNOD RICARD (FRA)
PEUGCT	PEUGEOT SA (FRA)
PRNTMP	PINAULT PRINTEMPS REDOUTE (FRA)
PRMODS	PROMODES (FRA)
RNLTL	RENAULT SA (FRA)
RP	AVENTIS SA (FRA)
SGPM	SAINT GOBAIN (FRA)
SANOFI	SANOFI (FRA)
SANSYN	SANOFI SYNTHELABO (FRA)
SCHNDR	SCHNEIDER ELECTRIC SA (FRA)
STMICR	STMICROELECTRONICS (FRA)
SG	SOCIETE GENERALE (FRA)
SODHO	SODEXHO ALLIANCE SA (FRA)
SLEE	SUEZ (FRA)
TFONE	TF1 (FRA)
THMCSF	THALES SA (FRA)
THMMM	THOMSON MULTIMEDIA (FRA)
FDP	TOTAL SA (FRA)
USNR	USINOR (FRA)
FERD	VALEO SA (FRA)
SNRPTB	VINCI SA (FRA)
GNEAU	VIVENDI UNIVERSAL (FRA)
VVENV	VEOLIA ENVIRONNEMENT SA (FRA)

Table 4-2: Factiva classification of industries ("IN" field)

IN	Industry Name
I0	**Agriculture And Forestry**
I01	Agriculture And Horticulture
I02	Forestry/Logging
I03	Fishing
I05	Support Activities For Agriculture
I1	**Energy**
I11	Coal Mining
I12	Coke Products
I13	Crude Oil/Natural Gas
I14	Petroleum/Coal Products
I15	Nuclear Fuel
I16	Electricity And Gas Utilities
I17	Water Utilities
I2	**Metals, Minerals And Chemicals**
I21	Metal Ore Mining
I22	Primary Metal Manufacturing
I23	Non-Metallic Mineral Mining/Quarrying
I24	Pottery/Ceramics/Porcelain
I25	Chemicals
I26	Artificial/Synthetic Fibers
I3	**Metal Goods And Engineering**
I31	Metal Product Manufacturing
I32	Machinery/Industrial Goods
I33	Computers
I34	Electrical And Electronic Engineering
I35	Motor Vehicles And Parts
I36	Vehicules (Other Than Cars) Manufacturing
I37	Instrument Engineering
I4	**Processing Industries**
I41	Food/Beverages/Tobacco
I42	Sugar, Cereals, Drinks, Tobacco
I43	Textiles
I44	Leather Goods
I45	Clothing, Textile Furnishing
I46	Wood Products
I47	Paper Industry (+ Printing, …)
I48	Plastics And Rubber Products
I49	Miscellaneous Manufacturing
I5	**Construction**
I50	Construction
I6	**Distribution, Hotels And Catering**
I61	Wholesalers
I64	Retail
I65	Retail Distribution
I66	Hotels/Restaurants/Casinos
I67	Product Repair Services
I7	**Transport And Communication**
I71	Railroads
I72	Urban Transit/Bus/Taxi/Trucking
I74	Marine Transport
I75	Airlines
I76	Port/Airport Operations
I77	Travel Agencies
I78	Couriers/Messengers
I79	Communications Services
I8	**Financial And Business Services**
I81	Banking And Financial Services
I82	Insurance
I83	Security And Commodity Brokers, Dealers, Exchanges, And Services
I84	Rental/Leasing Services
I85	Real Estate Transactions
I9	**Services And Entertainment**
I92	Waste Mngmt/Building Services
I95	Health Care
I97	Motion Pictures And Entertainment
I98	Personal Services

Table 4-3: Factiva classification of companies ("NS" field)

NS	Name of the subject
CCAT	**Corporate/Industrial News**
C01	Profiles of Companies
C02	Corporate Changes
C11	Plans/Strategy
C12	Legal/Judicial
C13	Regulation/Government Policy
C131	Regulatory Bodies
C132	Standards/Standardization
C133	Patents
C134	Trademarks/Copyrights
C14	Stock Listings
C15	Performance
C151	Earnings
C1511	Annual Meetings
C1512	Dividends
C1513	Sales Figures
C1514	Earnings Surprises
C152	Earnings Projections
C1521	Analyst Comment/Recommendation
C1522	Share Price Movement/Disruptions
C16	Bankruptcy
C17	Funding/Capital
C171	Share Capital
C1711	Initial Public Offerings
C1712	Shareholder-Rights Plans
C172	Corporate Debt Instruments
C173	Financing Agreements
C174	Corporate Credit Ratings
C18	Ownership Changes
C181	Acquisitions/Mergers/Takeovers
C182	Divestitures/Asset Sales
C183	Privatizations
C184	Joint Ventures
C21	Output/Production
C22	New Products/Services
C23	Research/Development
C24	Capacity/Facilities
C25	Information Technology
C26	Product Safety
C31	Marketing
C311	Domestic Markets
C312	External Markets
C313	Market Share
C314	Pricing
C3141	Sales Promotions
C315	Conferences/Exhibitions
C32	Advertising
C33	Contracts/Orders
C331	Defense Contracts
C332	Government Contracts
C333	Contracts, Non-government
C334	Licensing Agreements
C335	Franchises
C34	Monopolies/Antitrust
C341	Government Aid
C41	Management Issues
C411	Management Moves
C42	Labor/Personnel Issues
CINSDL	Insider Dealing
CLABDI	Labor Disputes
CREDUN	Lay-offs/Redundancies
CSHBUY	Share Buybacks
cstksp	Stock Splits

NS	Name of the subject
ECAT	**Economic News**
E11	Economic Performance
E1101	Economic Growth
E1102	Money Supply
E1103	Inflation/Prices
E1104	Personal Income/Average Earnings
E1105	Consumer Sentiment
E1106	Consumer Credit/Expenditure/Savings
E1107	Retail/Wholesale Sales
E1108	Budget Account
E1109	Government Borrowing Requirement
E1111	Industrial Production
E1116	Employment/Unemployment
E1117	Business Sentiment
E1118	Trade Figures
E1123	Car Registrations/Vehicle Sales
E1124	Bankruptcy Figures
E1125	Index of Leading Economic Indicators
E12	Economic/Monetary Policy
E1201	Central Bank Intervention
E1202	Interest Rates
E21	Government Finance
E211	Government Taxation/Revenue
E212	Government Borrowing
E213	Sub-National/Local Taxes
E214	Sub-National/Local Borrowing
E51	Trade/External Payments
E52	Euro Zone/Currency
GCAT	**Political/General News**
MCAT	**Commodity/Financial Market News**
M11	Equity Markets
M12	Debt/Bond Markets
M13	Money/Forex Markets
M131	Money Markets
M132	Foreign-Exchange News
M14	Commodity Markets
M141	Soft Commodity Markets
M142	Metals Markets
M143	Energy Markets
M15	Derivative Securities
M16	Fund Markets
NCAT	**Content Types**
NANL	Analysis
NCRX	Corrected Items
NITV	Interview
NSUM	News Digest
NTOP	Top Wire News

There are 56 different companies in the sample (each of them appears at least once in the CAC40 index over the selected sample period). However, only 40 out of these 56 firms will be analyzed in the subsequent GARCH and MSR econometric models, since some of these companies were created, merged, or acquired during the sample period. Figure 4-4 shows that news releases intensity was the highest for France Telecom SA (16.6%), Vivendi Universal (14.9%), Renault SA (8.0%), Alcatel SA (7.9%) and Total SA (7.6%)[39]. Such results raise the concern that the two major companies ("major" in terms of "news liquidity", i.e. the number of news releases concerning them) could possibly drive the results and lead us to abusive generalizations. We will deal with this problem by providing individual results for each company and showing that those results are consistent and homogeneous across all firms in our sample.

Concerning industry sectors, they are split into ten general categories and 56 sub-categories. Figure 4-5 shows that the three sectors with the highest news frequency from 1999 to 2003 were "Financial and Business Services" (48.2%), "Metal, Goods and Engineering" (44.0%) and "Transport and Communication" (38.2%).

Subjects have also been divided into 5 general categories and 105 more specific ones. From Figure 4-6 one can observe that the 7 most common news subjects among the 105 more specific categories are mergers and acquisitions (M&A) (25.5%), earnings projections (19.4%), earnings (10.8%), funding/capital (5.4%), analyst comment/recommendation (5.2%), contracts/orders (4.1%) and regulation/government policy (3.0%). According to Andersen (1996), we expect the impact of various news releases to be different across categories. One of the purposes of the subsequent econometric analysis will be to give some empirical support to such a theoretical prediction.

Table 4-4 provides an extensive set of summary statistics (mean, min, max, standard deviation, skewness, excess kurtosis and sum), for the whole sample and by company (Panel A), by industry (Panel B), by subject (Panel C) and by timing (Panel D). Note that only the 10 general categories of industries and 55 relevant subcategories of topics are considered in Panels B and C. Concerning Panel D, the "timing" classification of a news release depends whether it is published when the market is open (trading hours) or not (namely overnight and week-end). Daily news frequency ranges from 0 to 142 news

[39] Percentages do not sum up to 100%, because one news release often concerns more than one company, industry or subject. It follows that categories are not exclusive (but they are exhaustive).

releases a day, with an average and a standard deviation of respectively 58.50 and 20.12. The skewness and excess kurtosis (moments of order three and four) are close to zero and indicate that the total news probability distribution is asymptotically roughly Gaussian.

Panels A, B, C and D are quite interesting since they offer valuable information about possible cross-sectional differences in the news arrival rate. One of the main conclusions that can be drawn from the analysis of Table 4-4 is the following: the greater the total number of news releases concerning a company (or an industry, or a subject), the higher the standard deviation and the lower the skewness and kurtosis of the distribution of the daily number of news releases. Indeed, the behavior of the discrete random variable "number of news releases" is a priori best described by a Poisson probability distribution, and this distribution is known to converge to a Gaussian distribution as the number of observations increases. The relatively high level of kurtosis experienced by both Carrefour and Promodes is the direct consequence of the very strong and common outlier on August 30th, 2000 (about 70 press releases on that particular day), caused by the announcement of the merger between Carrefour and Promodes.

Figure 4-4: Total number of news releases by company on the French market

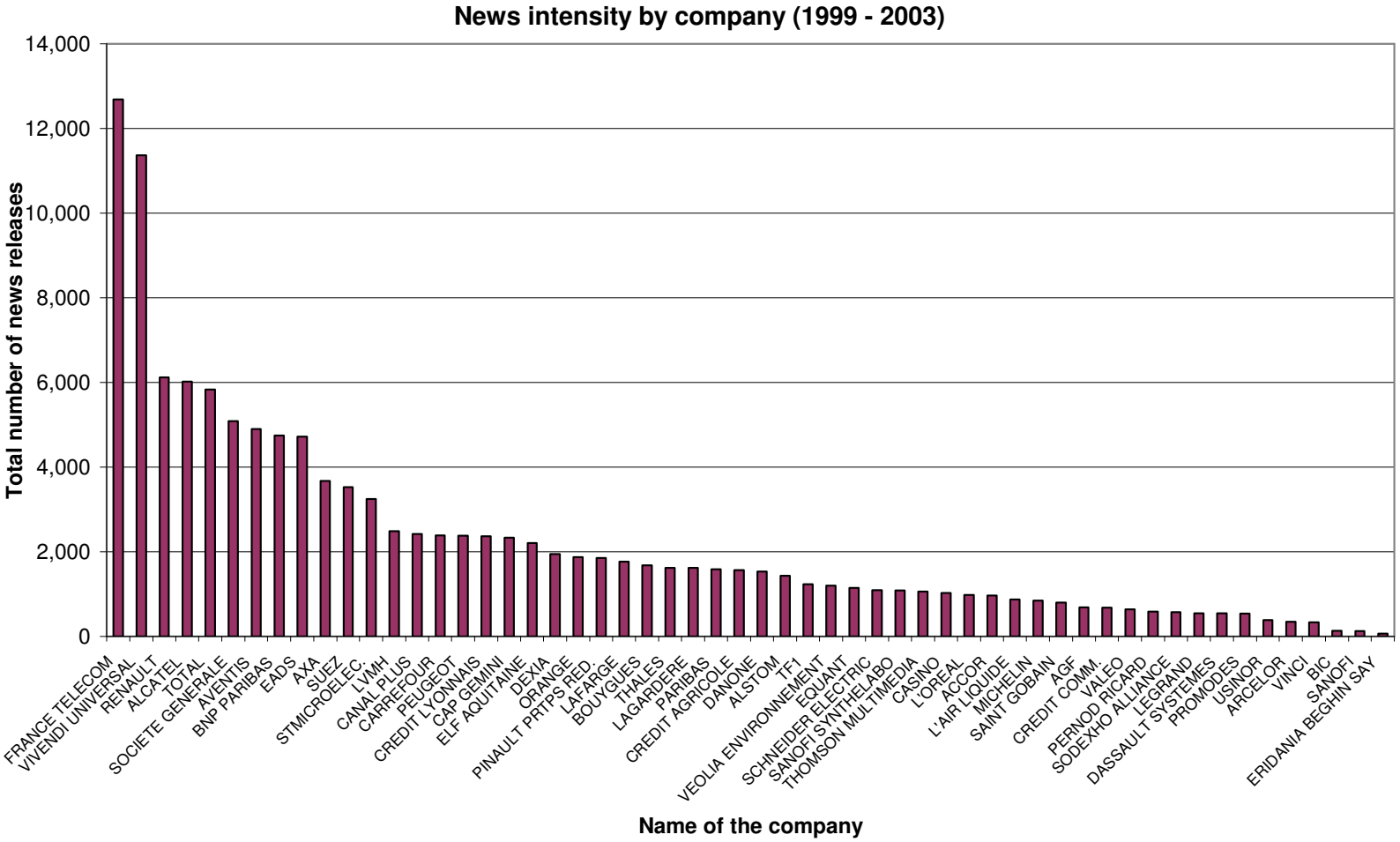


Figure 4-5: Total number of news releases by industry on the French market

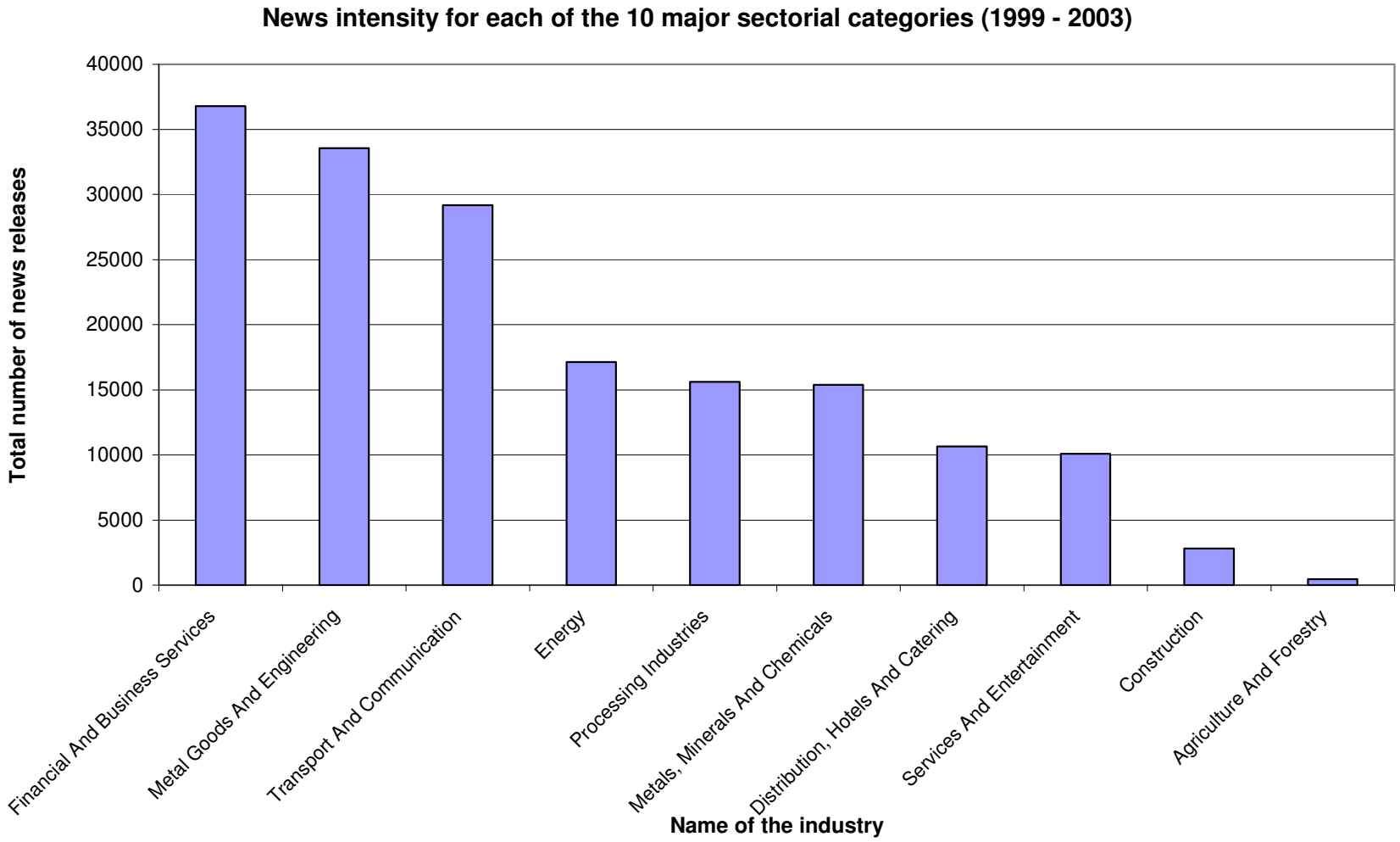


Figure 4-6: Total number of news releases by subject on the French market

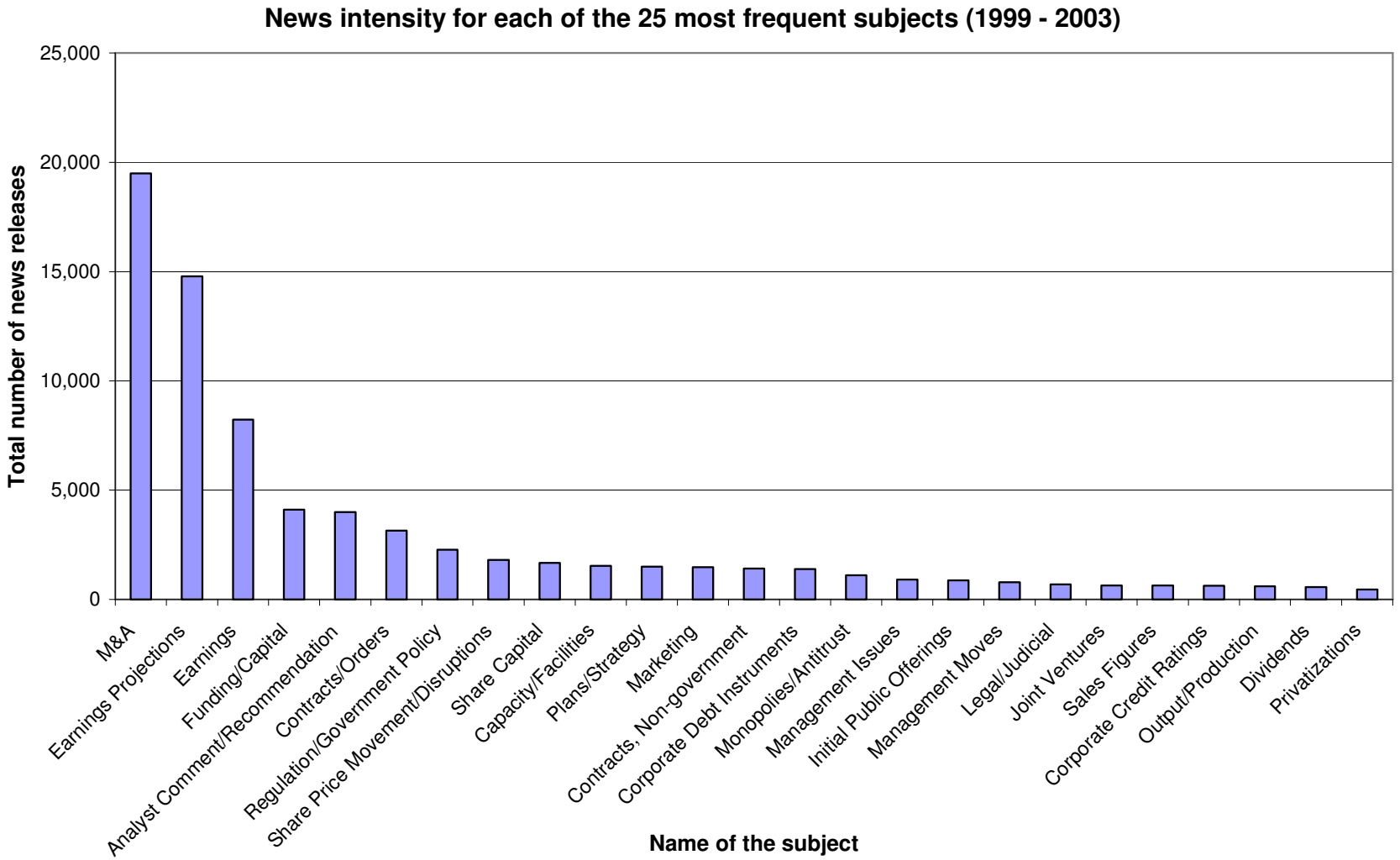


Table 4-4: Summary statistics on the French market

NAME	MEAN	MAX	MIN	STD DEV	SKEWNESS	KURTOSIS	SUM
All	**58.50**	**142**	**0**	**20.12**	**0.18**	**0.66**	**76,341**
PANEL A: Daily number of news releases by company							
FRANCE TELECOM	9.72	106	0	8.61	2.89	18.93	12,690
VIVENDI UNIVERSAL	8.71	93	0	8.74	3.06	16.22	11,370
RENAULT	4.69	43	0	4.89	2.27	8.73	6,121
ALCATEL	4.61	38	0	5.13	2.27	7.60	6,017
TOTAL	4.47	56	0	4.30	3.55	26.80	5,835
SOCIETE GENERALE	3.90	74	0	5.18	5.37	49.85	5,089
AVENTIS	3.75	29	0	3.52	2.09	7.00	4,898
BNP PARIBAS	3.64	62	0	4.92	4.71	37.32	4,745
EADS	3.62	61	0	4.38	4.41	38.01	4,720
AXA	2.81	63	0	3.96	5.61	60.22	3,671
SUEZ	2.70	43	0	3.34	3.42	23.19	3,524
STMICROELEC.	2.49	33	0	3.65	3.31	16.62	3,247
LVMH	1.90	32	0	3.27	3.55	17.73	2,484
CANAL PLUS	1.86	40	0	3.63	3.87	22.06	2,422
CARREFOUR	1.83	72	0	3.67	8.98	135.57	2,384
PEUGEOT	1.82	19	0	2.15	2.26	8.66	2,379
CREDIT LYONNAIS	1.81	42	0	2.94	5.38	53.67	2,363
CAP GEMINI	1.78	22	0	2.86	2.72	9.76	2,329
ELF AQUITAINE	1.69	61	0	4.39	5.82	51.73	2,209
DEXIA	1.49	26	0	2.14	3.59	23.00	1,949
ORANGE	1.44	42	0	2.94	4.75	39.88	1,876
PINAULT PRTPS RED.	1.42	29	0	2.81	4.46	27.98	1,853
LAFARGE	1.35	33	0	2.35	5.91	57.75	1,766
BOUYGUES	1.29	31	0	2.57	4.86	36.74	1,680
THALES	1.24	22	0	2.20	4.12	25.62	1,619
LAGARDERE	1.24	36	0	2.66	4.64	34.41	1,618
PARIBAS	1.21	72	0	4.49	8.06	93.55	1,583
CREDIT AGRICOLE	1.20	37	0	2.41	5.29	50.40	1,566
DANONE	1.18	22	0	2.12	3.78	21.99	1,534
ALSTOM	1.10	23	0	2.03	3.91	22.38	1,433
TF1	0.95	18	0	1.80	3.72	20.57	1,236
VEOLIA ENVIRONNEMENT	0.92	27	0	2.41	4.96	32.67	1,199
EQUANT	0.88	18	0	1.98	3.39	14.09	1,146
SCHNEIDER ELECTRIC	0.84	23	0	1.87	4.98	38.27	1,092
SANOFI SYNTHELABO	0.83	18	0	1.80	3.61	17.53	1,087
THOMSON MULTIMEDIA	0.82	22	0	1.98	4.74	31.60	1,065
CASINO	0.78	14	0	1.71	3.81	17.98	1,024
L'OREAL	0.75	26	0	1.67	6.19	62.66	978
ACCOR	0.74	18	0	1.60	4.90	36.38	967
L'AIR LIQUIDE	0.67	32	0	1.91	7.67	85.79	876
MICHELIN	0.65	17	0	1.61	5.21	36.65	844
SAINT GOBAIN	0.61	28	0	1.71	7.06	76.68	800
AGF	0.52	14	0	1.14	3.80	23.36	684
CREDIT COMM.	0.52	29	0	1.81	7.01	71.59	679
VALEO	0.49	13	0	1.31	4.30	23.79	642
PERNOD RICARD	0.45	27	0	1.54	8.12	97.91	586
SODEXHO ALLIANCE	0.44	23	0	1.48	7.71	85.50	575
LEGRAND	0.42	21	0	1.34	7.12	75.33	547
DASSAULT SYSTEMES	0.42	16	0	1.29	5.47	41.11	546
PROMODES	0.41	69	0	2.64	18.55	423.36	537
USINOR	0.29	8	0	0.79	4.49	27.80	383
ARCELOR	0.26	10	0	0.89	4.92	30.44	345
VINCI	0.26	11	0	0.82	5.73	46.94	334
BIC	0.10	7	0	0.50	7.56	74.64	134
SANOFI	0.10	11	0	0.57	10.20	141.22	126
ERIDANIA BEGHIN SAY	0.05	7	0	0.33	10.81	174.04	69

NAME	MEAN	MAX	MIN	STD DEV	SKEWNESS	KURTOSIS	SUM
PANEL B: Daily number of news releases by industry							
Financial And Business Services	28.18	103	0	12.09	0.63	1.78	36,769
Metal Goods And Engineering	25.71	74	0	12.38	0.46	0.00	33,556
Transport And Communication	22.36	111	0	12.39	1.07	3.39	29,178
Energy	13.14	72	0	9.44	1.51	3.68	17,144
Processing Industries	11.97	74	0	7.49	1.49	5.99	15,615
Metals, Minerals And Chemicals	11.78	50	0	6.90	1.05	1.80	15,379
Distribution, Hotels And Catering	8.16	73	0	5.97	2.38	14.78	10,645
Services And Entertainment	7.73	45	0	5.87	1.78	5.09	10,088
Construction	2.16	33	0	2.89	3.50	21.51	2,815
Agriculture And Forestry	0.35	9	0	0.77	3.39	19.15	463
PANEL C: Daily number of news releases by topic							
M&A	14.94	87	0	8.64	1.56	5.63	19,494
Earnings Projections	11.33	49	0	6.22	0.95	1.79	14,788
Earnings	6.30	41	0	5.60	1.57	3.31	8,223
Funding/Capital	3.14	28	0	2.82	2.36	11.38	4,099
Analyst Comment/Recommendation	3.06	20	0	2.97	1.00	1.18	3,996
Contracts/Orders	2.42	11	0	1.98	1.11	1.55	3,152
Regulation/Government Policy	1.75	15	0	1.96	1.63	3.54	2,278
Share Price Movement/Disruptions	1.38	13	0	1.89	1.98	4.94	1,806
Share Capital	1.29	28	0	1.92	4.54	41.86	1,677
Capacity/Facilities	1.18	8	0	1.31	1.51	2.93	1,541
Plans/Strategy	1.16	16	0	2.13	2.69	8.63	1,508
Marketing	1.13	10	0	1.41	1.79	4.37	1,476
Contracts, Non-government	1.09	9	0	1.41	1.72	3.82	1,419
Corporate Debt Instruments	1.07	13	0	1.44	2.63	12.45	1,393
Monopolies/Antitrust	0.85	8	0	1.25	2.03	5.11	1,113
Management Issues	0.70	8	0	1.01	1.83	4.57	908
Initial Public Offerings	0.67	27	0	1.62	6.63	75.39	877
Management Moves	0.61	8	0	0.95	2.05	5.97	792
Legal/Judicial	0.53	10	0	0.96	2.94	14.07	691
Joint Ventures	0.49	6	0	0.85	2.25	6.32	641
Sales Figures	0.49	5	0	0.95	2.15	4.14	633
Corporate Credit Ratings	0.48	10	0	1.03	3.76	20.14	630
Output/Production	0.47	6	0	0.80	2.18	6.48	608
Dividends	0.44	7	0	0.96	3.07	11.86	570
Privatizations	0.35	8	0	0.81	3.43	16.51	452
New Products/Services	0.35	4	0	0.68	2.33	6.22	452
Labor/Personnel Issues	0.33	11	0	0.73	4.25	39.01	437
Government Contracts	0.33	8	0	0.72	3.46	19.96	428
Stock Listings	0.27	7	0	0.72	3.58	16.42	353
Financing Agreements	0.22	4	0	0.53	2.77	8.96	291
Research/Development	0.21	4	0	0.51	2.77	9.36	280
Defense Contracts	0.21	7	0	0.58	4.18	26.35	269
Divestitures/Asset Sales	0.17	3	0	0.45	2.71	7.44	227
Annual Meetings	0.14	5	0	0.47	4.51	25.50	177
Pricing	0.11	3	0	0.37	3.86	16.93	143
Bankruptcy	0.10	4	0	0.39	5.05	31.83	128
Product Safety	0.09	6	0	0.41	6.48	57.75	122
Market Share	0.09	3	0	0.33	4.23	20.84	115
Licensing Agreements	0.07	2	0	0.28	4.55	21.97	87
External Markets	0.07	3	0	0.26	4.41	23.48	85
Domestic Markets	0.06	2	0	0.25	4.30	19.05	78
Advertising	0.06	2	0	0.25	4.49	21.21	76
Corporate Changes	0.06	2	0	0.25	4.69	23.36	74
Government Aid	0.03	3	0	0.21	7.75	69.83	42
Patents	0.02	2	0	0.16	7.85	68.12	30
Regulatory Bodies	0.02	3	0	0.18	9.73	111.11	29
Conferences/Exhibitions	0.02	3	0	0.17	11.32	156.43	26
Trademarks/Copyrights	0.02	1	0	0.13	7.34	51.96	23
Standards/Standardization	0.01	2	0	0.10	12.31	169.09	12
Earnings Surprises	0.01	1	0	0.10	10.30	104.16	12
Franchises	0.01	2	0	0.10	13.00	189.11	11
Information Technology	0.01	1	0	0.07	13.56	182.14	7
Sales Promotions	0.00	1	0	0.07	14.66	213.33	6
Shareholder-Rights Plans	0.00	1	0	0.06	16.08	256.99	5
Profiles of Companies	0.00	1	0	0.04	25.51	649.99	2
PANEL D: Daily number of news releases - trading hours VS non-trading hours							
Trading hours news	31.99	89	0	12.34	0.21	0.64	41,716
Non-trading hours news	26.55	74	0	10.05	0.53	0.96	34,625

4.1.4 Model specification

One of the core intermediate objectives of our work is to obtain adequate estimates of the daily stock volatility. In this aim, we choose to apply two different frameworks: GARCH(1,1) model, as used in Kalev and al. (2004), and Markov switching model that is thought to be more relevant and appropriate than the former.

4.1.4.1 The GARCH model

First of all, we propose to simply replicate the study of Kalev and al. (2004) on our extensive data set. Their article aimed at testing the relation between firm-specific announcements (as a proxy for information flows) and volatility on the Australian Stock Exchange in a GARCH framework. By testing the same relation in the same framework on a richer and more extensive dataset (namely Euronext news), we want to check that we get results consistent with those of Kalev and al. (2004). However, the study we intend to perform is different from theirs in two respects: while Kalev and al. (2004) investigate the volatility behavior on an intraday and daily scale and were mainly interested in the market volatility, we choose to focus on a daily scale on the individual stock volatility.

GARCH model has now become quite common in the modeling of uncertainty in financial asset returns. We will use it as follows[40]:

$$R_{i,t} = \alpha_i^{MM} + \beta_i^{MM} R_{m,t} + \varepsilon_{i,t} \text{ where } \varepsilon_{i,t} \left| \Omega_{t-1} \sim N(0, \sigma_{i,t}^2) \right. \quad (4\text{-}1)$$

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \quad (4\text{-}2)$$

where: $R_{i,t}$ is the return of stock i at interval t;

$R_{m,t}$ is the return of market index at interval t;

α_i^{MM} and β_i^{MM} are the market model parameters for the stock i;

[40] We use a GARCH model of order 1 since we want to rigorously replicate the study of Kalev and al. (2004). The GARCH(1,1) model is predominant in the volatility literature. Hansen and Lunde (2001) establish a forecast comparison of volatility models and conclude that "the best models do not provide a significantly better forecast than the GARCH(1,1) model". Nevertheless, superior orders and alternative GARCH specifications (like FIGARCH) were tested but they only brought negligible changes to the estimates. For example, the application of the BIC Schwarz's criterium on our data shows that the GARCH(1,1) is better specified than the GARCH(2,2) in 38 cases (= firms) out of 40. Results are available on request.

$\varepsilon_{i,t}$ are the serially uncorrelated errors of stock returns i with mean zero;

$\sigma_{i,t}^2$ is the conditional variance of $\varepsilon_{i,t}$.

The coefficients α_i and β_i reflect the dependence of the current volatility of stock i upon its pasts levels and the sum $\alpha_i + \beta_i$ indicates the degree of volatility persistence.

According to Lamoureux and Lastrapes (1990) and Kalev and al. (2004)'s results, the time-varying pattern of conditional volatility may be generated by serial correlation in the information arrival process. Information clustering is indeed a well-known phenomenon: news releases are not independent and are often concentrated around certain dates. In our data, the daily (resp. weekly) order-one serial correlation of our announcements is 45.9% (resp. 62.7%) and remains high for superior orders. This argument implies that when a proxy for information flows is inserted directly into the conditional variance equation, most of the observed volatility persistence is expected to disappear. Like Kalev and al. (2004), we adopt as a proxy the number of all firm-specific news events announced to the stock market per interval t (N_t), but contrary to them we run the analysis at the individual firm level ($N_{i,t}$) rather than at the market level in order to better and more precisely capture the impact of news announcements on the stock volatility. Importantly, $N_{i,t}$ cannot account for the flow of private information, it does just encompass a major component of the public information set that drives stock prices and volatility. In our first analysis $N_{i,t}$ is set as an exogenous variable in the conditional variance equation of the GARCH (1,1) specification as follows[41]:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \lambda_i N_{i,t} \qquad (4\text{-}3)$$

Two important results follow from the estimation of the above equation. Firstly, the significance of the coefficient λ_i provides evidence about whether the rate of public information arrival influences volatility of the stock i in the presence of conditional heteroscedasticity. Secondly, if the presence of GARCH effect is largely induced by the nature of the information flows, we may legitimately anticipate a significant decrease in the persistence of volatility, $\alpha_i + \beta_i$. This would suggest that the positive serial

[41] Notice that variables $N_{i,t}$, $R_{i,t}$ and $\sigma_{i,t}$ are exactly synchronized. $R_{i,t}$ is the return between the closing price of today and the previous closing price, and $N_{i,t}$ is simply the number of press releases that occurred between these two dates.

correlation in the daily volatilities is partly due to the positive serial correlation in the news arrivals. Since $N_{i,t}$ does not cover all sources of information, we do not expect volatility persistence to disappear entirely after the inclusion of $N_{i,t}$.

After this brief work of "replication", we will continue our quest for an innovative explicative model of the stock volatility and develop one of the basic pieces of our work: the Markov Switching Regression.

4.1.4.2 The two-state market model (TSMM)

We propose a new way to model the stock's variance behavior that is perfectly in line with the MDH and Roll's (1987) results. According to Roll, the true return generating process seems to be better described by a mixture of two distributions: one corresponding to a state of information arrival, and the other to the normal return behavior. In order to study the impact of news on volatility at the firm level, our new approach is based on a combination of the well-established market model (Sharpe, 1963) and the more recent Markov Switching Regression models (MSR), largely introduced and developed by Hamilton (1989 and 1994), and significantly extended in Krolzig (1997). Our initial intuition is simple: if the occurrence of firm-specific events may have a significant impact on the variance of the firm's return generating process, we must capture this perturbation by using a regime switching model.

In this framework, we assume that the error $\varepsilon_{i,t}$ is state dependent. S_t denotes our state variable. We consider the general and conceptual case of a two-state regime model (mixture of two distributions with different variances) with a low-variance regime (S_t = 1) and a high-variance regime ($S_t = 2$):

$$\begin{aligned} Y &= X\delta + \varepsilon_1 \quad if\ S_t = 1 \\ Y &= X\delta + \varepsilon_2 \quad if\ S_t = 2 \end{aligned} \tag{4-4}$$

with Y the vector of dependent variables and X the vector of explanatory variables.

The variance of the residuals for each state is given by:

$$\begin{aligned} E\left[\varepsilon_1 \varepsilon_1' \middle| X\right] &= \sigma_1^2 I \quad if\ S_t = 1 \\ E\left[\varepsilon_2 \varepsilon_2' \middle| X\right] &= \sigma_2^2 I \quad if\ S_t = 2 \end{aligned} \tag{4-5}$$

where $\sigma_2^2 > \sigma_1^2$ and I is the identity matrix.

We assume that the transition between the two regimes is governed by a Markov chain of order 1, for which the transition matrix is given by:

$$P = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix} \tag{4-6}$$

where $p_{ij} = p(S_t = i \mid S_{t-1} = j)$ corresponds to the probability of going from state j to state i. The unconditional probability of the regime is given by (Hamilton, 1994, p. 683):

$$p(S_t = 1) = \frac{1-p_{22}}{2-p_{11}-p_{22}}$$
$$p(S_t = 2) = \frac{1-p_{11}}{2-p_{11}-p_{22}} \tag{4-7}$$

The intuition underpinning our choice is simple: we anticipate that firm-specific events will impact the return variance. We will suppose that the return generating process can be adequately modeled using a two-regime process[42], one regime with normal variance and one regime with high variance (firm-specific event regime). Note that, in both regimes, the MM (market model) parameters are assumed to be the same[43]. Therefore the return generating process is:

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \varepsilon_{i,1,t} \quad \text{if } S_t = 1 \text{ with } \varepsilon_{i,1,t} \sim N(0,\sigma_{i,1})$$
$$R_{i,t} = \alpha_j + \beta_i R_{m,t} + \varepsilon_{i,2,t} \quad \text{if } S_t = 2 \text{ with } \varepsilon_{i,2,t} \sim N(0,\sigma_{i,2})$$
$$\text{and} \quad \sigma_{i,2} > \sigma_{i,1} \tag{4-8}$$

where $R_{i,t}$ is the return of stock i, $R_{m,t}$ is the market return and S_t is a state variable taking the value "1" if we are in the low variance regime and the value "2" if we are in the high variance regime. In the sequel, we will refer to Equation (4-8) as the two-state market model (TSMM). The proposed model is a direct and parsimonious extension of the classical MM. As the regime state variable S_t is not directly observable, we have to

[42] We have realized likelihood ratio test (LR test) to compare the two-state market model (TSMM) and the classical one-state market model (MM). Since we are in the presence of nuisance parameters under the alternative hypothesis, we have implemented the procedure advocated by Davies (1977). Our results (available on request) show that the TSMM dominates the MM in all cases (i.e. for each firm) at a confidence level of 5%. Since a three-regime decomposition appears to be justified only in the presence of strong outliers, we conclude that the two-regime model is an adequate representation of the return generating process. (Note also that a two-regime model is closer to the concept of a mixture of two distributions (MDH)).

[43] This hypothesis is supported by unreported results (available on request). We allow the market model parameters to vary across the two regimes through the estimation of a MSIAH model (Markov Switching Intercept Autoregressive Heteroscedasticity) (see Krolzig (1997)) and we find results very similar to the previous ones, except that the marginal impact of news is slightly stronger by less than one basis point.

specify its statistical properties. The model we propose is therefore based on the estimation of six parameters (α, β, σ_1, σ_2, p_{11} and p_{22}) and, while much more flexible than the classical MM model, remains parsimonious.

The estimation of Markov Switching Regression models is fully presented in Hamilton (1994). It is based on a maximum likelihood approach, for which an efficient estimation algorithm has been developed. The estimated probability of being in a specific state at a specific date is one of the interesting by-products of the advocated approach. It allows one to look for the reasons explaining an increase in the variance of stock returns; in our case, to look for the link between information flow and conditional variance.

The use of this Markov Switching Regression model provides numerous interesting features. Beyond the above-quoted possibility of obtaining the estimated probability of being in a specific regime at a specific date, this specification is in line with Roll's (1987) intuition. In his Presidential Address to the American Finance Association, he clearly highlighted that the true return-generating process seems to be better described by a mixture of two distributions, the first one corresponding to a state of information arrival and the other to the normal return behavior: it's clearly a good candidate for testing the mixture of distribution hypothesis. Once we have estimated the probability for a firm of being in a high variance regime at a specific date, we can test whether it is linked with information flow.

The specification has other attractive features:

- the Gaussian conditional distribution of returns could be misleading. While the model imposes a Gaussian assumption for the return distribution in each state, as shown in Hamilton (1994) and Krolzig (1997), it allows to capture skewness and kurtosis in the unconditional distribution;
- the Markov Switching Regression framework also takes conditional heteroscedasticity into account without imposing a specific form on the conditional dependence of the variance (as in the (G)ARCH framework);
- finally, the estimation process provides us with the estimated variance in each regime, the probability of being in a specific regime at a specific date, and the estimated transition probabilities[44].

[44] All estimations presented in this section have been realized under the Ox econometric software, using the Krolzig MSVAR package. We thank Professor J. Hamilton for advising the use of this package. It is freely downloadbale at: http://www.nuff.ox.ac.uk/Users/Doornik/index.html.

In a certain manner, the TSMM, once we have estimated the probability for a firm of being in a high variance regime at a specific date, could be a better framework than the GARCH model (via the estimated λ_i) in order to estimate the impact of the public information rate on the stock volatility at the individual firm level. This may also give some additional robustness to the news-volatility relation in the pattern of the daily stock return volatility.

4.1.5 Results

4.1.5.1 GARCH framework

The only difference between the basic equation of Kalev and al. (2004)'s regression and ours resides in the introduction of the market return in Equation (4-1), which allows us to control for macroeconomic shocks and to focus on the firm-specific variance of the stock returns. The results reported in Table 5 for each of the forty firms of the sample[45] are largely consistent with Kalev and al. (2004)'s main conclusion: the inclusion of the news variable in the conditional variance equation of the GARCH model reduces volatility persistence (α + β) for most companies of the sample, even though the magnitude of this reduction is highly variable across firms. The volatility persistence without the inclusion of the news proxy in the equation is close to one for all companies, which is symptomatic of the presence of persistence in the stock volatility, while the same volatility persistence once the news variable is introduced in the regression ranges from a very low 0.1555 for CapGemini (-84%) to a hardly unchanged 0.9999 for Alstom. The magnitude of the reduction in volatility persistence is of course closely related to the significance of the coefficient of the news variable: the higher the significance of the impact of news on volatility, the stronger the decrease of the volatility persistence (with some noticeable exceptions like Carrefour and Michelin). Coefficients of the news variable turn out to be positive in all cases and significant for 17 companies out of 40. We will show in a next section of this thesis that the MSR model performs in fact much better. Note also that the coefficient of the market return is positive and highly significant for all companies, which totally justifies its introduction into the GARCH equations.

[45] Note that we encountered convergence problems with 5 companies (Alcatel, Total, STMicroelectronics, Société Générale and Thalès). This kind of problems is typical of GARCH models.

Table 4-5: GARCH model

We apply the Equations (4-1) and (4-2) (without news) and (4-1) and (4-3) (with news) of the GARCH (1,1) model to our data: 1270 observations for each of the 40 securities. We encountered convergence problems for 5 companies.

			VU	*FT*	*Renault*	*Alcatel*	*Aventis*	*Total*	*STMicro*	*SG*	*AXA*	*Suez*
Garch(1,1)	Without NBNews	**C**	-0.0004	-0.0009	0.0009	-0.0003	0.0002	0.0004	0.0003	0.0009	-0.0002	0.0000
		t-stat	*-0.76*	*-1.38*	*1.54*	*-0.47*	*0.43*	*1.26*	*0.48*	*2.02*	*-0.56*	*-0.04*
		Coeff-MR	1.15	1.58	0.85	1.76	0.84	0.79	1.53	1.13	1.31	0.80
		t-stat	*23.47*	*27.60*	*19.56*	*27.36*	*21.71*	*26.46*	*28.58*	*32.22*	*28.09*	*14.20*
		ω	8.13E-06	1.45E-05	8.00E-06	6.32E-06	1.76E-05	2.95E-07	1.91E-05	7.22E-08	3.17E-06	1.13E-05
		t-stat	*0.69*	*1.49*	*0.93*	*1.04*	*1.07*	*0.54*	*1.45*	*0.09*	*1.20*	*1.64*
		α	0.10	0.08	0.07	0.05	0.09	0.04	0.08	0.03	0.08	0.11
		t-stat	*1.63*	*2.54*	*2.16*	*3.86*	*1.51*	*3.15*	*2.68*	*2.75*	*3.29*	*2.74*
		β	0.89	0.90	0.91	0.94	0.86	0.96	0.89	0.97	0.92	0.86
		t-stat	*11.18*	*21.86*	*19.98*	*62.30*	*8.85*	*70.92*	*20.20*	*80.85*	*33.81*	*16.57*
		α+β	**0.9899**	**0.9840**	**0.9866**	**0.9947**	**0.9491**	**0.9987**	**0.9735**	**0.9999**	**0.9929**	**0.9749**
	With NBNews	**C**	-0.0008	-0.0014	0.0009	-	0.0000	-	-	-	-0.0002	0.0000
		t-stat	*-1.60*	*-2.24*	*1.57*	-	*-0.07*	-	-	-	*-0.48*	*0.02*
		Coeff-MR	1.16	1.49	0.81	-	0.83	-	-	-	1.31	0.80
		t-stat	*30.42*	*26.04*	*20.25*	-	*26.64*	-	-	-	*25.04*	*14.78*
		ω	3.09E-05	8.67E-05	8.77E-05	-	6.68E-05	-	-	-	2.30E-05	7.03E-06
		t-stat	*2.15*	*2.13*	*2.34*	-	*2.20*	-	-	-	*0.50*	*0.86*
		α	0.24	0.20	0.18	-	0.19	-	-	-	0.16	0.13
		t-stat	*4.83*	*3.64*	*3.78*	-	*3.71*	-	-	-	*2.41*	*1.93*
		β	0.20	0.14	0.30	-	0.28	-	-	-	0.64	0.83
		t-stat	*1.89*	*0.97*	*1.97*	-	*1.95*	-	-	-	*1.63*	*8.13*
		α+β	**0.4463**	**0.3416**	**0.4834**	**-**	**0.4721**	**-**	**-**	**-**	**0.8014**	**0.9560**
		λ	2.75E-05	3.87E-05	4.10E-05	-	2.92E-05	-	-	-	1.52E-05	4.19E-06
		t-stat	*4.47*	*3.34*	*3.33*	-	*4.72*	-	-	-	*0.66*	*1.13*

			LVMH	BNP	Capgem	Peugeot	Lafarge	Danone	Schneid.	Bouygues	PPR	Carrefour
Garch(1,1)	Without NBNews	C	0.0008	0.0007	-0.0003	0.0007	0.0006	0.0000	0.0006	0.0003	-0.0001	0.0001
		t-stat	*1.74*	*1.80*	*-0.38*	*1.73*	*1.07*	*0.08*	*1.24*	*0.63*	*-0.22*	*0.26*
		Coeff-MR	0.99	0.98	1.40	0.67	0.69	0.44	0.64	0.97	0.90	0.90
		t-stat	*26.30*	*25.71*	*21.79*	*19.76*	*16.30*	*16.78*	*16.18*	*20.63*	*18.51*	*27.03*
		ω	9.69E-06	2.72E-06	2.58E-05	2.03E-05	3.69E-06	1.71E-06	4.27E-06	5.29E-06	9.42E-06	1.66E-06
		t-stat	*1.35*	*0.83*	*1.34*	*1.27*	*1.64*	*1.01*	*0.55*	*1.47*	*1.34*	*0.79*
		α	0.07	0.08	0.04	0.11	0.04	0.06	0.09	0.08	0.04	0.05
		t-stat	*2.44*	*1.70*	*1.91*	*2.17*	*3.90*	*2.32*	*1.66*	*3.88*	*1.83*	*2.22*
		β	0.90	0.91	0.93	0.82	0.95	0.93	0.91	0.92	0.93	0.95
		t-stat	*20.52*	*16.63*	*27.20*	*8.22*	*79.38*	*32.04*	*14.69*	*43.79*	*25.44*	*39.91*
		α+β	**0.9721**	**0.9912**	**0.9724**	**0.9333**	**0.9906**	**0.9944**	**0.9969**	**0.9935**	**0.9752**	**0.9955**
	With NBNews	C	0.0004	0.0007	-0.0008	0.0006	0.0007	-0.0002	0.0001	0.0001	-0.0003	-0.0002
		t-stat	*0.85*	*1.77*	*-1.17*	*1.43*	*1.27*	*-0.51*	*0.25*	*0.22*	*-0.69*	*-0.53*
		Coeff-MR	0.98	0.96	1.31	0.67	0.69	0.44	0.62	0.96	0.92	0.87
		t-stat	*31.27*	*25.55*	*24.82*	*22.19*	*16.04*	*16.87*	*15.62*	*22.98*	*26.25*	*29.73*
		ω	2.70E-05	1.15E-05	3.64E-04	3.89E-05	9.09E-07	3.23E-06	1.79E-04	1.24E-04	1.81E-04	9.52E-06
		t-stat	*1.01*	*1.21*	*6.93*	*1.39*	*0.14*	*0.45*	*2.75*	*3.17*	*2.57*	*1.21*
		α	0.09	0.17	0.12	0.17	0.06	0.12	0.21	0.16	0.08	0.15
		t-stat	*1.87*	*3.35*	*2.90*	*4.33*	*1.49*	*0.79*	*4.91*	*4.09*	*1.94*	*3.78*
		β	0.67	0.73	0.03	0.58	0.92	0.84	0.21	0.35	0.14	0.66
		t-stat	*3.30*	*5.51*	*0.60*	*3.11*	*13.35*	*3.68*	*1.17*	*3.10*	*0.72*	*6.68*
		α+β	**0.7632**	**0.9018**	**0.1555**	**0.7427**	**0.9766**	**0.9587**	**0.4206**	**0.5040**	**0.2222**	**0.8026**
		λ	3.00E-05	3.48E-06	2.16E-04	1.97E-05	6.74E-06	8.03E-06	1.24E-04	1.23E-04	5.74E-05	3.11E-05
		t-stat	*1.67*	*0.95*	*8.16*	*1.78*	*0.95*	*0.45*	*3.44*	*4.06*	*4.66*	*3.15*

		Alstom	*Dassault*	*Sanofi*	*Thales*	*Accor*	*TF1*	*Lagardère*	*Air liq.*	*Michelin*	*L'oréal*
Without NBNews	**C**	0.0001	0.0007	0.0004	0.0002	0.0004	0.0004	0.0006	0.0004	0.0002	0.0000
	t-stat	*0.07*	*0.83*	*0.87*	*0.31*	*0.78*	*0.64*	*1.19*	*0.95*	*0.46*	*-0.05*
	Coeff-MR	0.68	1.26	0.81	0.65	0.81	1.08	0.95	0.77	0.58	0.87
	t-stat	*6.65*	*20.77*	*24.12*	*14.99*	*21.29*	*17.41*	*19.65*	*22.38*	*14.10*	*24.65*
	ω	1.60E-05	1.83E-05	1.44E-06	4.60E-06	3.28E-05	6.26E-06	2.18E-06	1.29E-06	6.39E-05	6.22E-07
	t-stat	*1.16*	*0.38*	*0.44*	*1.42*	*1.58*	*1.03*	*1.18*	*0.95*	*2.82*	*0.34*
	α	0.11	0.06	0.04	0.08	0.11	0.08	0.06	0.04	0.16	0.06
	t-stat	*2.58*	*0.76*	*1.21*	*3.57*	*2.34*	*2.31*	*4.69*	*3.16*	*2.91*	*1.77*
	β	0.90	0.92	0.95	0.92	0.81	0.92	0.93	0.95	0.68	0.94
	t-stat	*29.33*	*7.54*	*22.52*	*43.74*	*8.87*	*27.74*	*69.65*	*62.15*	*8.51*	*26.63*
	α+β	**0.9999**	**0.9821**	**0.9967**	**0.9947**	**0.9206**	**0.9963**	**0.9976**	**0.9952**	**0.8415**	**0.9995**
With NBNews	**C**	-0.0001	0.0001	0.0003	-	0.0003	0.0003	0.0006	0.0004	0.0002	-0.0003
	t-stat	*-0.06*	*0.18*	*0.61*	-	*0.57*	*0.47*	*1.24*	*0.87*	*0.43*	*-0.76*
	Coeff-MR	0.66	1.23	0.81	-	0.81	1.06	0.93	0.76	0.61	0.86
	t-stat	*7.66*	*18.26*	*25.32*	-	*21.85*	*23.20*	*21.36*	*22.16*	*15.91*	*26.42*
	ω	3.63E-07	2.17E-04	3.99E-06	-	9.43E-05	1.24E-06	1.96E-06	1.80E-06	6.99E-05	2.62E-07
	t-stat	*0.02*	*0.83*	*0.65*	-	*2.13*	*0.15*	*0.46*	*0.79*	*2.90*	*0.09*
	α	0.11	0.21	0.09	-	0.18	0.11	0.06	0.04	0.15	0.11
	t-stat	*2.54*	*2.49*	*1.67*	-	*3.99*	*2.21*	*2.88*	*2.58*	*3.80*	*3.02*
	β	0.89	0.50	0.87	-	0.49	0.86	0.92	0.94	0.55	0.86
	t-stat	*20.13*	*1.24*	*10.24*	-	*2.94*	*11.30*	*24.42*	*31.90*	*6.13*	*16.59*
	α+β	**0.9999**	**0.7171**	**0.9602**	**-**	**0.6727**	**0.9655**	**0.9803**	**0.9848**	**0.7030**	**0.9713**
	λ	2.19E-05	1.91E-04	1.45E-05	-	5.44E-05	3.23E-05	7.55E-06	3.62E-06	6.54E-05	1.26E-05
	t-stat	*1.05*	*0.88*	*1.32*	-	*2.07*	*1.15*	*1.26*	*1.00*	*3.11*	*2.10*

			Valeo	*Sodexho*	*Casino*	*Pernod*	*St-Gobain*	*Equant*	*AGF*	*Arcelor*	*Vinci*	*Bic*
Garch(1,1)	**Without NBNews**	**C**	-0.0001	0.0000	0.0002	0.0007	0.0010	0.0004	0.0000	0.0008	0.0006	0.0004
		t-stat	*-0.23*	*-0.06*	*0.46*	*1.34*	*2.19*	*0.43*	*-0.12*	*1.18*	*1.12*	*0.65*
		Coeff-MR	0.62	0.59	0.48	0.23	0.79	1.31	0.55	0.62	0.28	0.20
		t-stat	*14.34*	*7.76*	*12.47*	*4.71*	*10.88*	*18.36*	*11.30*	*11.98*	*9.79*	*4.85*
		ω	2.62E-05	2.97E-06	2.01E-06	2.33E-05	1.23E-05	2.89E-05	2.40E-06	2.09E-05	1.98E-07	2.59E-05
		t-stat	*1.03*	*0.74*	*1.10*	*1.05*	*1.37*	*2.06*	*1.33*	*1.87*	*0.22*	*1.32*
		α	0.10	0.04	0.03	0.14	0.19	0.11	0.06	0.06	0.01	0.11
		t-stat	*1.90*	*3.40*	*3.46*	*1.74*	*1.84*	*3.41*	*3.66*	*3.54*	*2.00*	*2.08*
		β	0.85	0.96	0.97	0.82	0.81	0.87	0.93	0.91	0.99	0.84
		t-stat	*8.96*	*124.38*	*86.98*	*7.18*	*9.51*	*26.38*	*46.67*	*32.90*	*121.78*	*10.21*
		α+β	**0.9522**	**0.9970**	**0.9927**	**0.9580**	**0.9999**	**0.9797**	**0.9927**	**0.9657**	**0.9988**	**0.9506**
	With NBNews	**C**	-0.0003	0.0004	-0.0001	0.0006	0.0008	-0.0001	-0.0001	0.0008	0.0005	0.0003
		t-stat	*-0.62*	*0.75*	*-0.24*	*1.14*	*1.58*	*-0.16*	*-0.33*	*1.18*	*0.63*	*0.50*
		Coeff-MR	0.62	0.65	0.51	0.22	0.78	1.30	0.55	0.62	0.30	0.18
		t-stat	*15.71*	*13.59*	*17.37*	*5.05*	*19.99*	*18.83*	*11.26*	*11.98*	*9.50*	*4.49*
		ω	5.44E-05	1.57E-04	1.39E-04	2.82E-06	9.19E-05	3.88E-05	3.24E-07	2.09E-05	9.89E-05	1.16E-04
		t-stat	*2.01*	*4.52*	*3.21*	*0.22*	*1.38*	*1.60*	*0.18*	*1.87*	*0.86*	*1.06*
		α	0.16	0.21	0.11	0.12	0.18	0.15	0.06	0.06	0.21	0.21
		t-stat	*4.84*	*4.54*	*2.66*	*1.33*	*4.07*	*3.05*	*3.22*	*3.54*	*2.68*	*2.06*
		β	0.66	0.30	0.18	0.84	0.45	0.80	0.93	0.91	0.48	0.51
		t-stat	*6.96*	*3.77*	*1.10*	*6.52*	*1.83*	*10.33*	*36.90*	*32.90*	*1.08*	*1.53*
		α+β	**0.8214**	**0.5153**	**0.2889**	**0.9585**	**0.6269**	**0.9431**	**0.9941**	**0.9657**	**0.6894**	**0.7190**
		λ	8.65E-05	2.26E-04	8.45E-05	1.01E-05	1.26E-04	4.28E-05	3.68E-06	0.00E+00	6.61E-05	2.68E-04
		t-stat	*2.72*	*4.05*	*4.45*	*1.43*	*1.66*	*1.23*	*1.16*	*0.00*	*0.62*	*1.31*

4.1.5.2 MSR framework: the two-state market model (TSMM).

For the many reasons brought up in Section 4.1.2, an MSR framework is expected to be a more appropriate specification and to fit much better than a GARCH model. As it was thoroughly explained in the same section, such an approach leads us to develop a two-state market model that provides us with the (daily) probabilities of being in a high volatility regime on a specific day for each stock of the sample (see Table 4-6 for extensive results). Once these probabilities have been obtained, it is possible to regress them over our explicative and control variables, company by company. Figure 4-7 offers first intuitive evidence of the close relation between the probability of being in a high volatility regime and the daily number of news releases. It displays in parallel the time-series evolution of the probability of being in a high volatility state and the daily news frequency for Vivendi Universal, one of the largest firms of the sample. It clearly results from this figure that peaks of volatility generally occur simultaneously with peaks of public information, which is a clue that stock volatility and public information are quite strongly correlated. The next regressions aim at investigating this expected link more formally and precisely.

Table 4-6: Estimates of the Market Model (MM) and of the Two-State Market Model (TSMM)

The following table gives the estimates of the main parameters of the Market Model (intercept, slope, standard error variance) and of the Two-State Market Model (the same, but with one variance for regime 1 (low regime) and one variance for regime 2 (high regime). The percentage of days in each volatility state is also given. See also Equation (4-8).

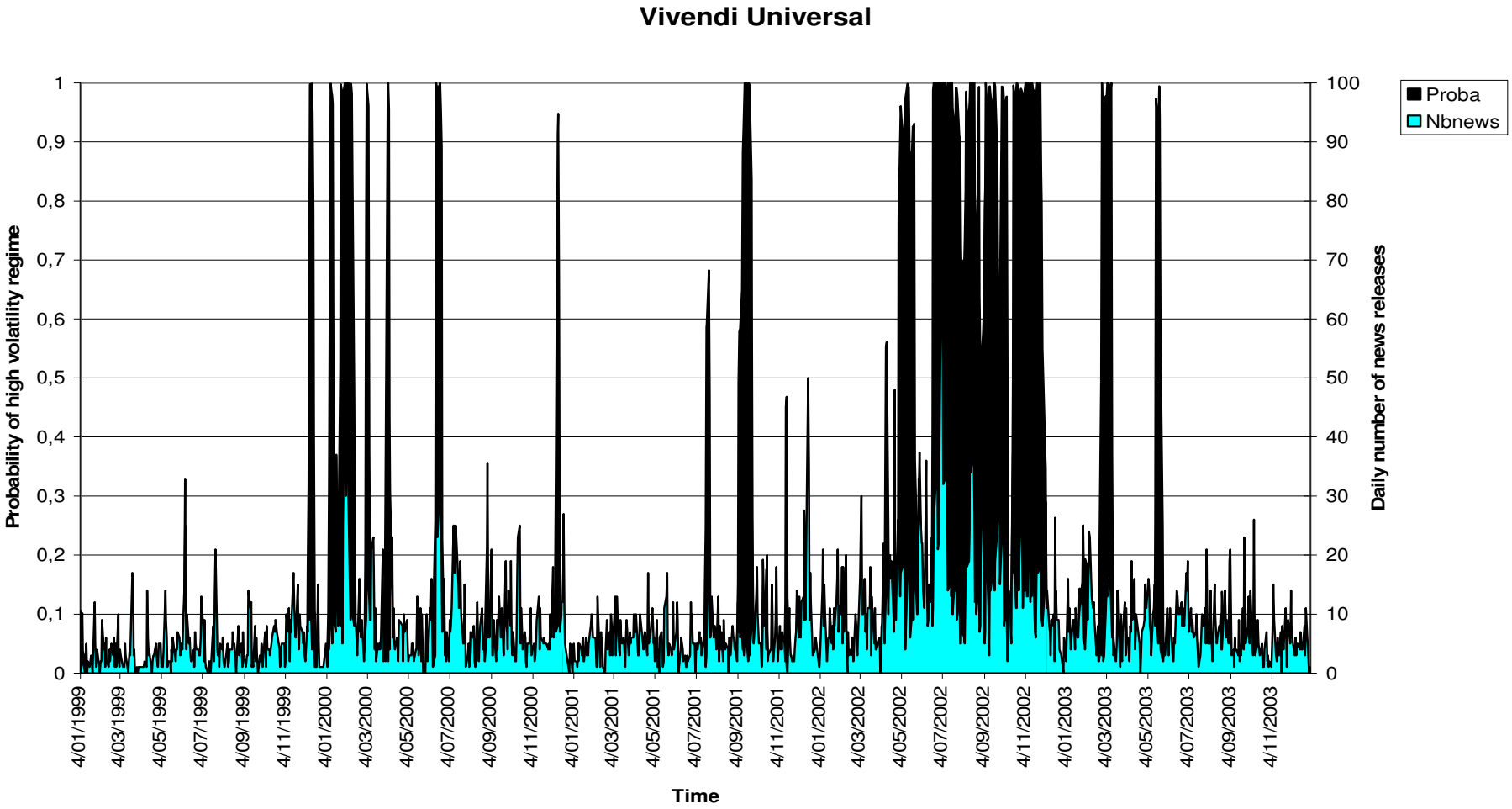
		VU	FT	Renault	Alcatel	Aventis	Total	STMicro	SG	AXA	Suez
MM	α	-0.0005	-0.0001	0.0006	0.0002	0.0004	0.0006	0.0011	0.0008	-0.0001	-0.0003
	β	1.3765	1.6315	0.7940	1.7790	0.8142	0.7459	1.5727	1.0924	1.3868	1.0812
	σ^2	6.65E-04	7.66E-04	5.22E-04	9.58E-04	3.36E-04	2.55E-04	6.80E-04	3.33E-04	3.39E-04	4.13E-04
TSMM	α	-0.0010	-0.0017	0.0007	-0.0001	0.0001	0.0007	0.0004	0.0009	-0.0004	-0.0004
	β	1.2222	1.6387	0.7989	1.7404	0.8605	0.7797	1.5124	1.0563	1.3048	0.7937
	σ^2 (regime 1)	2.13E-04	3.07E-04	1.44E-04	4.80E-04	1.44E-04	8.77E-05	2.98E-04	8.03E-05	1.48E-04	2.07E-04
	σ^2 (regime 2)	2.90E-03	2.83E-03	7.70E-04	3.99E-03	5.80E-04	3.84E-04	1.03E-03	5.45E-04	6.12E-04	1.46E-03
	% State 1	84.90%	83.37%	42.01%	87.85%	56.74%	43.78%	48.71%	46.56%	59.78%	81.54%
	% State 2	16.70%	18.23%	59.59%	13.75%	44.86%	57.82%	52.89%	55.04%	41.82%	20.06%

		LVMH	BNP	Capgem	Peugeot	Lafarge	Danone	Schneid.	Bouygues	PPR	Carrefour
MM	α	0.0008	0.0005	-0.0004	0.0007	0.0002	0.0002	0.0003	0.0007	-0.0003	0.0001
	β	1.0438	0.9907	1.4318	0.6864	0.7063	0.4245	0.6355	1.0446	0.9577	0.8911
	σ^2	3.52E-04	2.65E-04	9.41E-04	2.93E-04	4.26E-04	2.84E-04	5.08E-04	5.52E-04	3.88E-04	3.18E-04
TSMM	α	0.0002	0.0007	-0.0004	0.0003	0.0005	-0.0003	0.0001	0.0003	-0.0005	0.0004
	β	0.9746	0.9605	1.4254	0.6633	0.6530	0.4152	0.6300	0.9577	0.9248	0.8606
	σ^2 (regime 1)	1.21E-04	9.63E-05	4.07E-04	1.28E-04	1.74E-04	7.96E-05	2.30E-04	2.15E-04	1.72E-04	1.39E-04
	σ^2 (regime 2)	7.53E-04	4.51E-04	1.98E-03	5.75E-04	6.39E-04	5.43E-04	1.06E-03	1.03E-03	8.87E-04	6.20E-04
	% State 1	64.34%	55.05%	67.78%	65.16%	47.90%	51.50%	67.49%	58.22%	69.83%	64.69%
	% State 2	37.26%	46.55%	33.82%	36.44%	53.70%	50.10%	34.11%	43.38%	31.77%	36.91%

		Alstom	Dassault	Sanofi	Thales	Accor	TF1	Lagardère	Air liq.	Michelin	L'oréal
MM	α	-0.0008	0.0006	0.0007	0.0001	0.0002	0.0010	0.0005	0.0003	0.0003	0.0002
	β	0.8393	1.2680	0.7818	0.7363	0.8055	1.0999	0.9848	0.7211	0.6191	0.8419
	σ^2	1.78E-03	1.02E-03	4.27E-04	6.16E-04	4.18E-04	8.62E-04	5.54E-04	2.83E-04	3.87E-04	3.12E-04
TSMM	α	-0.0006	0.0000	0.0001	0.0001	0.0001	0.0003	-0.0001	0.0002	-0.0005	-0.0001
	β	0.6789	1.2622	0.8064	0.6651	0.8312	1.0656	0.9201	0.7346	0.5628	0.8275
	σ^2 (regime 1)	6.71E-04	4.85E-04	1.37E-04	2.34E-04	2.55E-04	3.12E-04	2.49E-04	1.45E-04	2.20E-04	1.07E-04
	σ^2 (regime 2)	8.80E-03	2.03E-03	6.08E-04	1.01E-03	8.96E-04	1.56E-03	1.21E-03	6.46E-04	1.26E-03	4.75E-04
	% State 1	87.62%	68.91%	38.89%	54.16%	79.58%	56.66%	68.62%	76.99%	85.02%	39.74%
	% State 2	13.98%	32.69%	62.71%	47.44%	22.02%	44.94%	32.98%	24.61%	16.58%	61.86%

		Valeo	Sodexho	Casino	Pernod	St-Gobain	Equant	AGF	Arcelor	Vinci	Bic
MM	α	-0.0003	-0.0002	0.0000	0.0008	0.0005	-0.0005	0.0001	0.0006	0.0006	0.0000
	β	0.6244	0.6550	0.4966	0.2309	0.8075	1.4390	0.8186	0.6224	0.2688	0.2510
	σ^2	5.16E-04	5.46E-04	2.95E-04	4.95E-04	4.70E-04	1.17E-03	3.03E-04	6.00E-04	3.61E-04	4.87E-04
TSMM	α	-0.0005	-0.0002	-0.0002	0.0002	0.0004	-0.0009	-0.0002	0.0003	-0.0001	-0.0002
	β	0.6130	0.6064	0.4931	0.2206	0.7444	1.3425	0.5733	0.5471	0.2555	0.2335
	σ^2 (regime 1)	2.13E-04	1.93E-04	1.47E-04	1.68E-04	2.12E-04	4.30E-04	1.26E-04	2.11E-04	1.25E-04	1.73E-04
	σ^2 (regime 2)	8.73E-04	1.14E-03	5.78E-04	9.53E-04	1.10E-03	2.36E-03	7.21E-04	9.15E-04	6.74E-04	1.08E-03
	% State 1	53.05%	64.27%	71.54%	61.46%	66.51%	62.54%	70.24%	40.48%	56.52%	67.06%
	% State 2	48.55%	37.33%	30.06%	40.14%	35.09%	39.06%	31.36%	61.12%	45.08%	34.54%

Figure 4-7: Comparison of the volatility with the level of communication for Vivendi



4.1.5.2.1 "Firm by firm" analysis

At this stage, we face two alternatives: either we transform the probability of being in a high volatility regime into a dummy variable and we regress this dummy over our variables of interest through a probit regression, or we directly use the estimated probability of being in a specific volatility state as the dependent variable in a General Method of Moments (GMM) regression[46]. In this latter case, we have opted for a GMM regression rather than a classical OLS regression because we expect the residuals of our regression to present substantial heteroscedasticity and serial correlation problems. The coefficients obtained by this approach are the same as the coefficients delivered by the classical OLS regression, but the GMM procedure provides Student statistics robust to heteroscedasticity and serial correlation of order one. The estimate of the covariance matrix is obtained thanks to the convergent estimator of Newey and West (1987).

4.1.5.2.1.1 Probit regression: the dependent variable is a dummy

In this time-series regression, we transform the state regime into a dummy variable $D_{i,t}$ equal to 1 when the probability of being in the high variance regime at a specific date *t* is greater than 50% and 0 otherwise.

$$D_{i,t} = \beta_{i,0} + \beta_{i,1} N_{i,t} + \beta_{i,2} N_{i,t}^2 + \sum_{j=1}^{J} \delta_{i,j} X_{i,j,t} + \eta_{i,t} \qquad (4\text{-}9)$$

where *i* and *t* represent respectively the security i and the time interval *t*;

$D_{i,t}$ = 1 if $P[S_{i,t} = 2] \geq 50\%$ (« high variance regime »)

0 otherwise (« low variance regime ») ;

$N_{i,t}$ stands for the number of news releases specific to security *i* over time interval *t*;

$X_{i,j,t}$ stands for the j[th] control variable.

[46] Before adopting the GMM framework, we first used an « improved » OLS approach like « weighted least squares »: results are very similar to those obtained by GMM, and even more significant. However, we think that a GMM regression must be preferred to an (even improved) OLS regression since the former is known to be much more robust than the latter to heteroscedasticity and serial correlation problems apparent in our data. Moreover, GMM coefficients are directly interpretable as such, whereas "weighted least squares" coefficients are not, since the variables have been logit-transformed because they are proportions.

For we can a priori suspect the relation between our stock volatility proxy and the daily number of news releases to be non-linear and to exhibit some quadratic pattern, we add the square number of news announcements to the equation. A positive (resp. negative) and significant coefficient of such a variable would reveal the presence of convexity (resp. concavity) in the form of the function relating the news arrival intensity to the stock volatility.

The inclusion of relevant control variables in our regression is necessary in order to test the robustness of the impact of the firm-specific news releases on the conditional variance of the security. The conditional variance of the market portfolio is our main control variable, since all volume proxies (volume levels, differences and log) turned out to be insignificant (at a 5% level) and thus rather irrelevant on a daily scale. This is totally consistent with Kalev and al. (2004)'s results where volume and first of the week variables are shown to be insignificant on a daily scale. Last but not least, the conditional variance of the sector index is not yet introduced here seeing that some stocks of the sample are the major constituents of their respective sector index, which logically induces an annoying strong correlation between the dependent variable and the control variable. Nevertheless, the sector variable will be well taken into account in the later panel study. Both conditional variances of the market portfolio and of the sector index are estimated via a GARCH (1,1) model.

Table 7 reports the results of the probit regression. The coefficient of the news variable is almost always significantly positive, whereas the coefficient of the square news variable is generally negative and significant, which seems to indicate that the marginal impact of news announcements on the stock volatility is positive but its magnitude decreases as the number of news releases increases. Not surprisingly, the market variance coefficient appears to be positive and significant as the stock volatility is known to convey some market component. Such a control variable allows us to account for all macroeconomic shocks that affect the stock volatility and that are not covered by our set of firm-specific press releases. Finally, note the relatively high cross-sectional variability of the R-square of the regressions. The quality of the R-square is of course mainly driven by the significance of the market variance and news variable coefficients that look like quite heterogeneous and variable across companies.

Table 4-7: Probit regressions

We run a probit regression with 1270 observations for each of the 40 securities. The dependent variable is $D_{i,t} = 1$ if $P[S_{i,t} = 2] \geq 50\%$ (« high variance regime »), 0 otherwise (« low variance regime »). C stands for the constant. N(i,t) is the daily number of news releases concerning a given company. N(i,t)² is the square N(i,t) in order to account for a possible quadratic effect. MarketVar is the conditional variance of the market portfolio estimated via a GARCH (1,1). See Equation (4-9).

		VU	FT	Renault	Alcatel	Aventis	Total	STMicro	SG	AXA	Suez
Probit	C	-2.89	-2.64	-0.21	-2.86	-0.96	-0.28	-0.05	-0.11	-2.36	-3.07
	t-stat	*-21.11*	*-17.30*	*-2.64*	*-18.43*	*-10.89*	*-3.06*	*-0.83*	*-1.34*	*-17.79*	*-19.51*
	Coeff-N(i,t)	0.0946	0.0889	0.1032	0.1121	0.1380	0.0983	0.0502	-0.0097	0.0887	0.0106
	t-stat	*7.65*	*5.97*	*6.75*	*4.85*	*6.02*	*3.65*	*2.65*	*-0.35*	*5.01*	*0.30*
	Coeff-N(i,t)²	-0.0007	-0.0011	-0.0025	-0.0014	-0.0042	-0.0003	-0.0018	0.0054	-0.0012	0.0032
	t-stat	*-2.90*	*-3.17*	*-3.91*	*-1.72*	*-3.17*	*-0.15*	*-1.87*	*2.56*	*-2.77*	*1.57*
	Coeff-MarketVar	3048.93	2121.78	536.28	2546.57	1348.79	257.61	-47.35	469.33	8200.15	6155.37
	t-stat	*13.25*	*10.63*	*3.01*	*11.54*	*7.43*	*1.46*	*-0.27*	*2.64*	*15.61*	*16.48*
	R²	0.44	0.16	0.05	0.24	0.09	0.06	0.01	0.04	0.38	0.58

		LVMH	BNP	Capgem	Peugeot	Lafarge	Danone	Schneid.	Bouygues	PPR	Carrefour
Probit	C	-1.10	-1.00	-1.51	-0.98	-0.24	-0.18	-0.94	-0.76	-1.56	-0.69
	t-stat	*-14.96*	*-11.19*	*-18.81*	*-12.87*	*-3.63*	*-2.71*	*-13.92*	*-11.20*	*-20.06*	*-10.23*
	Coeff-N(i,t)	0.1772	0.1241	0.1818	0.0861	0.1034	0.1580	0.2000	0.2782	0.0828	0.1525
	t-stat	*7.85*	*5.24*	*6.31*	*2.40*	*3.58*	*5.01*	*5.78*	*10.40*	*3.43*	*8.44*
	Coeff-N(i,t)²	-0.0053	-0.0003	-0.0072	0.0014	-0.0014	-0.0058	-0.0083	-0.0091	-0.0013	-0.0018
	t-stat	*-4.35*	*-0.22*	*-3.30*	*0.35*	*-0.73*	*-2.22*	*-3.23*	*-6.59*	*-1.04*	*-4.12*
	Coeff-MarketVar	1443.70	1655.81	2698.03	1287.57	866.40	-516.41	838.55	750.30	2351.80	244.55
	t-stat	*8.05*	*8.37*	*12.50*	*7.18*	*4.84*	*-2.88*	*4.69*	*4.29*	*12.32*	*1.38*
	R²	0.11	0.17	0.20	0.07	0.04	0.04	0.05	0.12	0.17	0.08

		Alstom	Dassault	Sanofi	Thales	Accor	TF1	Lagardère	Air liq.	Michelin	L'oréal
Probit	**C**	-2.19	-0.85	-0.05	-0.17	-1.44	-0.36	-0.77	-1.06	-1.92	-0.20
	t-stat	*-22.07*	*-13.32*	*-0.79*	*-2.51*	*-19.12*	*-5.83*	*-11.41*	*-15.13*	*-21.23*	*-3.10*
	Coeff-N(i,t)	0.1555	0.0410	0.1675	0.0662	0.1874	0.1431	0.2146	0.1951	0.2179	0.1552
	t-stat	*3.67*	*0.75*	*4.11*	*2.86*	*4.26*	*3.83*	*8.84*	*5.58*	*4.19*	*3.97*
	Coeff-N(i,t)²	-0.0042	0.0005	-0.0108	-0.0016	-0.0087	-0.0077	-0.0056	-0.0055	-0.0049	-0.0055
	t-stat	*-1.25*	*0.09*	*-2.52*	*-1.09*	*-2.24*	*-2.12*	*-4.35*	*-3.15*	*-1.10*	*-2.00*
	Coeff-MarketVar	2475.55	1345.08	918.05	40.51	1735.52	387.29	99.65	715.30	1871.20	939.36
	t-stat	*12.37*	*7.52*	*4.75*	*0.24*	*9.32*	*2.23*	*0.54*	*3.94*	*9.56*	*5.01*
	R²	0.20	0.05	0.03	0.01	0.10	0.02	0.09	0.04	0.12	0.04

		Valeo	Sodexho	Casino	Pernod	St-Gobain	Equant	AGF	Arcelor	Vinci	Bic
Probit	**C**	-0.44	-1.09	-0.84	-0.59	-1.26	-1.09	-1.35	-0.45	-0.24	-0.95
	t-stat	*-6.96*	*-16.32*	*-12.53*	*-9.02*	*-17.72*	*-15.28*	*-18.93*	*-6.67*	*-3.97*	*-14.74*
	Coeff-N(i,t)	0.3476	0.2341	0.2204	0.0478	0.2560	0.1974	0.1755	0.1929	0.0127	0.5735
	t-stat	*5.86*	*4.43*	*4.62*	*2.15*	*6.49*	*4.68*	*2.87*	*2.88*	*0.16*	*2.52*
	Coeff-N(i,t)²	-0.0283	-0.0064	-0.0114	-0.0016	-0.0091	-0.0074	-0.0068	-0.0163	0.0009	0.0058
	t-stat	*-3.73*	*-1.21*	*-2.16*	*-1.37*	*-3.76*	*-1.63*	*-0.67*	*-1.32*	*0.07*	*0.06*
	Coeff-MarketVar	991.41	2048.54	562.59	1063.78	1702.00	2212.50	2670.75	1774.83	-173.00	1355.79
	t-stat	*5.58*	*10.75*	*3.18*	*5.92*	*9.07*	*11.63*	*13.58*	*8.83*	*-0.98*	*7.64*
	R²	0.05	0.14	0.62	0.03	0.11	0.14	0.20	0.08	0.00	0.08

4.1.5.2.1.2 GMM regression: the dependent variable is the smooth probability itself

Rather than transforming the daily probability of being in the high volatility state into a simple dummy variable and losing some precious and valuable information in the process, it could be more judicious to directly use the above probability as the dependent variable of the time-series regression. Note that the only difference with Equation (4-9) of the probit regression is the nature of the dependent variable. The equation of the GMM regression is then as follows:

$$P\left[S_{i,t} = 2\right] = \beta_{i,0} + \beta_{i,1} N_{i,t} + \beta_{i,2} N_{i,t}^2 + \sum_{j=1}^{J} \delta_{i,j} X_{i,j,t} + \eta_{i,t} \quad (4\text{-}10)$$

Table 4-8 displays the results obtained with this second approach. They are very similar to those provided by the probit regression, although they appear to be still more significant and convincing. The coefficient of the news variable (resp. square news variable) is always positive (resp. generally negative) and significant in 38 (resp. 26) cases out of 40. The positive but marginally decreasing impact of public information on stock volatility seems to be demonstrated for most of the companies over the sample period. The other observations made on the basis of the results of the previous probit regression remain valid.

Table 4-8: GMM regressions

We run a GMM regression with 1270 observations for each of the 40 securities. The covariance matrix is estimated with the convergent estimator of Newey and West (1987). The dependent variable is the daily probability of being in the high-variance regime. C stands for the constant. N(i,t) is the daily number of news releases concerning a given company. $N(i,t)^2$ is the square N(i,t) in order to account for a possible quadratic effect. MarketVar is the conditional variance of the market portfolio estimated via a GARCH (1,1). See Equation (4-10).

		VU	FT	Renault	Alcatel	Aventis	Total	STMicro	SG	AXA	Suez
GMM	**C**	-0.17	-0.09	0.42	-0.07	0.21	0.44	0.47	0.43	0.05	-0.16
	t-stat	*-13.08*	*-5.48*	*12.89*	*-4.73*	*7.81*	*12.48*	*15.73*	*15.44*	*2.32*	*-14.31*
	Coeff-N(i,t)	0.0167	0.0146	0.0364	0.0120	0.0464	0.0394	0.0162	0.0197	0.0158	0.0006
	t-stat	*7.94*	*8.45*	*6.46*	*3.68*	*6.50*	*7.94*	*2.25*	*5.70*	*3.57*	*0.15*
	Coeff-N(i,t)²	-0.0001	-0.0001	-0.0008	0.0001	-0.0015	-0.0006	-0.0006	-0.0002	-0.0001	0.0005
	t-stat	*-2.51*	*-6.35*	*-3.79*	*0.47*	*-3.49*	*-4.98*	*-1.88*	*-2.69*	*-1.06*	*4.41*
	Coeff-MarketVar	742.24	568.67	127.86	534.58	339.08	-111.70	69.99	162.50	1180.85	1271.27
	t-stat	*11.29*	*9.11*	*1.62*	*8.87*	*5.40*	*-1.62*	*0.93*	*2.43*	*23.92*	*32.34*
	R²	0.46	0.27	0.07	0.32	0.09	0.07	0.00	0.04	0.37	0.58

		LVMH	BNP	Capgem	Peugeot	Lafarge	Danone	Schneid.	Bouygues	PPR	Carrefour
GMM	**C**	0.15	0.18	0.06	0.21	0.39	0.46	0.23	0.27	0.11	0.24
	t-stat	*7.38*	*6.74*	*3.28*	*9.30*	*13.27*	*16.37*	*11.48*	*10.62*	*6.88*	*9.74*
	Coeff-N(i,t)	0.0589	0.0338	0.0478	0.0235	0.0367	0.0614	0.0627	0.0921	0.0222	0.0552
	t-stat	*8.27*	*15.25*	*5.95*	*2.82*	*4.10*	*5.49*	*5.92*	*9.94*	*4.19*	*10.91*
	Coeff-N(i,t)²	-0.0016	-0.0003	-0.0017	0.0006	-0.0006	-0.0024	-0.0023	-0.0031	-0.0003	-0.0007
	t-stat	*-3.75*	*-7.52*	*-2.63*	*0.88*	*-2.08*	*-2.94*	*-3.18*	*-4.87*	*-1.46*	*-7.85*
	Coeff-MarketVar	458.35	513.62	772.24	379.63	333.51	-97.68	230.82	222.90	592.20	102.11
	t-stat	*7.27*	*8.48*	*13.59*	*6.18*	*4.55*	*-1.38*	*3.75*	*2.95*	*11.15*	*1.46*
	R²	0.17	0.19	0.27	0.08	0.04	0.04	0.08	0.13	0.21	0.10

		Alstom	Dassault	Sanofi	Thales	Accor	TF1	Lagardère	Air liq.	Michelin	L'oréal
GMM	**C**	-0.06	0.18	0.53	0.39	0.06	0.35	0.22	0.15	0.02	0.50
	t-stat	*-3.81*	*8.27*	*18.36*	*13.19*	*3.78*	*11.92*	*9.76*	*8.04*	*1.17*	*19.81*
	Coeff-N(i,t)	0.0253	0.0068	0.0597	0.0278	0.0508	0.0544	0.0774	0.0670	0.0541	0.0462
	t-stat	*2.73*	*0.42*	*4.33*	*3.37*	*4.48*	*3.68*	*8.73*	*5.70*	*4.28*	*4.44*
	Coeff-N(i,t)²	0.0001	0.0008	-0.0040	-0.0007	-0.0021	-0.0028	-0.0021	-0.0021	-0.0013	-0.0016
	t-stat	*0.10*	*0.44*	*-2.69*	*-1.55*	*-2.09*	*-1.96*	*-3.47*	*-3.87*	*-1.02*	*-2.73*
	Coeff-MarketVar	634.31	513.04	189.65	94.35	446.38	184.28	91.14	191.19	413.80	276.66
	t-stat	*9.68*	*7.72*	*2.88*	*1.11*	*8.17*	*2.18*	*1.44*	*3.21*	*6.25*	*4.74*
	R²	0.24	0.09	0.03	0.02	0.12	0.03	0.12	0.07	0.15	0.04

		Valeo	Sodexho	Casino	Pernod	St-Gobain	Equant	AGF	Arcelor	Vinci	Bic
GMM	**C**	0.35	0.19	0.23	0.27	0.16	0.13	0.04	0.47	0.47	0.22
	t-stat	*13.41*	*9.54*	*12.84*	*10.94*	*8.37*	*5.77*	*1.92*	*24.26*	*20.76*	*10.52*
	Coeff-N(i,t)	0.1221	0.0753	0.0632	0.0143	0.0774	0.0694	0.0445	0.0577	0.0179	0.2244
	t-stat	*6.35*	*6.38*	*5.15*	*1.91*	*6.52*	*5.30*	*3.02*	*3.34*	*0.74*	*5.68*
	Coeff-N(i,t)²	-0.0095	-0.0025	-0.0034	-0.0003	-0.0028	-0.0030	-0.0012	-0.0061	-0.0015	-0.0186
	t-stat	*-3.37*	*-4.18*	*-2.80*	*-0.90*	*-4.14*	*-2.22*	*-0.81*	*-2.03*	*-0.47*	*-2.76*
	Coeff-MarketVar	332.43	566.68	110.11	368.37	522.90	751.74	895.31	400.07	-92.29	387.50
	t-stat	*4.64*	*9.91*	*2.14*	*6.09*	*9.97*	*9.89*	*11.03*	*8.52*	*-1.48*	*5.75*
	R²	0.07	0.17	0.05	0.04	0.14	0.19	0.24	0.08	0.00	0.10

4.1.5.2.2 Panel data analysis

The main advantage of the panel analysis is to work globally and to analyze the combination of time and cross-sectional effects. Modeling in this setting calls for some complex stochastic specification. We therefore use the most common techniques: the fixed effects and random effects approaches.

The fixed effects approach takes the constant of the model to be a group-specific term in the regression model. The constant is different for any security. It should be noted that the term "fixed" as used here indicates that the constant term does not vary over time.

The random effects approach assumes that the residual is a group specific random element. In this model, the constant is the same for all securities included in the sample.

The crucial distinction between these two cases is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether this effect is stochastic or not. In our case, the fixed effects model is probably better specified. Indeed, from the relatively high heterogeneity across stocks observed in the individual regressions, we may think that the fixed effects approach fits better, since it assumes the constant to be different for each stock. Moreover, from a statistical perspective, Hausman's test (see the bottom of Table 4-9) indicates that the fixed effects model performs much better than the random effects one.

Table 4-9 presents the results for fixed effect and random effects panel approaches. Columns 1, 2, 3 and 4 indicate successive regressions where explicative variables are added one after the other. Regression one only considers the daily number of announcements; regression two introduces the square number of announcements; and regressions three and four add respectively the conditional market and sector variances. By comparison with the previous firm by firm probit and GMM regressions, the explicative and control variables are unchanged, but the introduction of the conditional variance of the sector index as a new control variable. Both approaches give very similar results: all coefficients are highly significant. Market and especially sector variances seem to strongly affect the stock volatility, and the frequency of news arrivals is shown to have a positive but marginally decreasing impact on the stock volatility. On average, the first news release of the day tends to increase the probability of being in a high volatility regime by roughly 2 basis points, while 50 news releases increase this

same probability by about 43 basis points[47]. Naturally, as it could have already been observed in the individual regressions, the same number of press releases will not impact the same way the volatility of small and big firms. One news wire about Vivendi Universal or France Telecom is of course expected to have much less impact on the volatility of these firms than one news release concerning Valeo for example. Finally, it is worth mentioning that our news variable and two control variables explain altogether not less than 26% of the stock volatility in the fixed effects model.

Quite undoubtedly, the level of firm-specific news intensity clearly influences the conditional variance of returns. Such a conclusion brings strong support in favor of the Mixture of Distribution Hypothesis.

[47] The coefficient of the variable $N_{i,t}$ (resp. $N^2_{i,t}$) in the fixed effects framework when controlling for market and sector variances (column 4) is equal to 0.021872 (resp. -0.000265). That means that the first announcement of the day will **on average** increase the probability of being in the high volatility state by 0.021872 – 0.000265 = 2.1607 basis points. And 50 pieces of news will increase on average this same probability by 50*0.021872 – 50²*0.000265 = 43.11 basis points. Notice that these figures hide a very heterogeneous reality since the news coefficient ranges from a poor 0.0006 for Suez to an impressive 0.2244 for BIC.

Table 4-9: Panel data analysis

We opt for the fixed effects and random effects approaches. The dependent variable is alternatively the volatility regime dummy and the probability of being in the high volatility regime. C stands for the constant. N(i,t) is the daily number of news releases concerning a given company. $N(i,t)^2$ is the square N(i,t) in order to account for a possible quadratic effect. MarketVar and SectorVar are respectively the conditional variance of the market portfolio and the conditional variance of the sector index. Both of them are estimated via a GARCH (1,1). Hausman's test is largely significant and provides evidence in favor of the fixed effects approach. The panel is composed of 1270 observations for each of the 40 securities.

		Y=dummy				*Y=proba*			
		1	***2***	***3***	***4***	***1***	***2***	***3***	***4***
Fixed Effects	**C**	-	-	-	-	-	-	-	-
	t-stat	-	-	-	-	-	-	-	-
	Coeff-N(i,t)	0.020291	0.029837	0.029082	0.022915	0.019403	0.028254	0.027561	0.021872
	t-stat	*36.20*	*35.82*	*35.60*	*28.86*	*42.69*	*41.85*	*41.87*	*34.59*
	Coeff-N(i,t)²	-	-0.000283	-0.000279	-0.000286	-	-0.000262	-0.000258	-0.000265
	t-stat	-	*-15.46*	*-15.56*	*-16.55*	-	*-17.69*	*-17.91*	*-19.27*
	Coeff-MarketVar	-	-	432.799	113.1	-	-	396.695	101.74
	t-stat	-	-	*45.28*	*10.70*	-	-	*51.50*	*12.09*
	Coeff-SectorVar	-	-	-	240.37	-	-	-	221.77
	t-stat	-	-	-	*61.87*	-	-	-	*71.68*
	R²	0.1118	0.1159	0.1502	0.2098	0.1375	0.1428	0.1854	0.2603

		Y=dummy				*Y=proba*			
		1	***2***	***3***	***4***	***1***	***2***	***3***	***4***
Random Effects	**C**	0.319681	0.305827	0.189661	0.144981	0.340982	0.328124	0.221651	0.180393
	t-stat	*13.29*	*12.54*	*7.76*	*5.09*	*42.60*	*14.77*	*9.97*	*6.94*
	Coeff-N(i,t)	0.020217	0.02971	0.028961	0.022828	0.019345	0.028155	0.027468	0.021807
	t-stat	*36.12*	*35.72*	*35.50*	*28.78*	*42.60*	*41.75*	*41.78*	*34.51*
	Coeff-N(i,t)²	-	-0.000281	-0.000278	-0.000285	-	-0.000261	-0.000258	-0.000264
	t-stat	-	*-15.40*	*-15.50*	*-16.49*	-	*-17.62*	*-17.85*	*-19.21*
	Coeff-MarketVar	-	-	432.837	114.189	-	-	396.724	102.505
	t-stat	-	-	*45.28*	*10.81*	-	-	*51.50*	*12.18*
	Coeff-SectorVar	-	-	-	239.58	-	-	-	221.21
	t-stat	-	-	-	*61.73*	-	-	-	*71.56*
	R²	0.0119	0.0133	0.0451	0.0641	0.0154	0.0170	0.0559	0.0803

	Hausman's test:	6.8931***	10.815***	10.521***	26.85***	8.1273***	12.014***	11.704***	27.971***

4.1.5.2.3 Analysis refinement: categorization by topic and by timing

Public information is not homogeneous. We must consider the possibility that a news release about merger and acquisition does not affect the stock volatility the same way as an earnings announcement for example. Similarly, information released during trading hours may not have the same impact on the stock volatility as information released during non-trading hours. From a theoretical perspective (see Andersen (1996)), the topic and the timing of each news release matter. We check this point by running the same regression as above on a limited category of news releases. We define two timing categories: news released during *Trading Hours* and news released during *Non-Trading Hours* (mainly week-end and overnight); and we retain the seven most common subject categories: *Merger and Acquisition (M&A)*, *Earnings*, *Earnings Projections*, *Analyst Comments and Recommendations*[48], *Funding/Capital*, *Regulation and Government Policy*, and, finally, *Contract*. We also choose to add an eighth category named *Earnings Projections clean* which is composed of all *Earnings Projections* except those concerning *Share Price Movement/Disruptions* in order to avoid any endogeneity problem.

Results are synthesized in Table 4-10 and Figure 4-8. If the regressions are run individually for each category, with the conditional market and sector variances as the usual control variables, all coefficients are significant at the 5%-level. With a marginal impact of 9.1 points (0.091) on the stock volatility, the *Regulation and Government Policy* category seems to be the critical news category of our sample. *Analyst Comments and Recommendations* and *Earnings Projections (clean)* come after with coefficients of respectively 6.8 and 4.9. *Contract* (3.4), *M&A* (3.4) and *Earnings* (3.0) follow a bit further and finally comes *Funding and Capital* with a poor 1.2. The high informativeness of the *Regulation and Government Policy* category was not particularly expected. Is it a French specificity due to the important weight of the state in the economy and the society, or does such a type of news impact the same way the stock volatility in other national contexts? This issue could be investigated in further research. Finally, it is interesting to notice that *Earnings Projections* (especially *Analyst Comments and Recommendations*) turn out to be more informative to investors than *Earnings* announcements themselves, since the former induce a greater increase in the

[48] Notice that « analyst comments and recommendations » is actually a subset of « earnings projections ». For obvious reasons of collinearity problems, these two variables will not be introduced together into the multivariate regressions of the next section.

stock volatility than the former. Investors and traders appear to value substantially analyst advice and comments.

From a timing perspective, news announcements released during *Trading Hours* seem to have a stronger effect (3.5) on the stock volatility than news announcements released during the week-end or overnight (2.8). This "timing" difference still becomes more impressive if both variables are introduced together in the same regression. First column of Table 4-11 reports a coefficient of 8.5 for the *Trading Hours* category, weighted by the number of trading hours by day, as opposed to a relatively low coefficient of 1.9 for the *Non-Trading Hours* category, weighted by the number of non trading hours by day. Many reasons can be advanced to explain such a result. It might be that news categories with the strongest impact on volatility (government policy, earnings projections, etc.) occur principally during trading hours; or that investors, consciously or not, don't give the same weight and relevance to information published during trading hours and information published at another time. At this stage, it is too early to definitively answer this question.

We also run a regression with all "category variables" together, except *Analyst Comments and Recommendations* and *Earnings Projections clean* because of obvious problems of collinearity. The last two columns of Table 4-11 generally confirm the results displayed in Figure 4-8, though three types of news, *Earnings*, *Funding and Capital* and *Contract* appear now to be insignificant, while *Regulation and Government Policy*, *M&A* and *Earnings Projections* remain highly significant.

Table 4-10: Panel data analysis for various types of news: categories are not mixed

We opt for the fixed effects approach. The dependent variable is the probability of being in the high volatility regime. C stands for the constant. N(i,t) is the daily number of news releases concerning a given company. $N(i,t)^2$ is the square N(i,t) in order to account for a possible quadratic effect. MarketVar and SectorVar are respectively the conditional variance of the market portfolio and the conditional variance of the sector index. Both of them are estimated via a GARCH (1,1). A first distinction is made between news released during non-trading hours and news released during trading hours. A second distinction is made according to the subject: Mergers and Acquisitions, Earnings, Analyst Comment/Recommendation, Earnings Projections, Funding and Capital, Regulation and Government Policy and, finally, Contracts/Orders. The panel is composed of 1270 observations for each of the 40 securities. The covariance matrix is estimated with the convergent estimator of Newey and West (1987) so that t-stat are robust to serial correlation and heteroscedasticity.

		Dependent Variable : Smooth Probability				
	News category	**All**	**Trading Hours**	**No Trading Hours**	**M&A**	**Earnings**
Fixed effects	**C**	-	-	-	-	-
	t-stat	-	-	-	-	-
	Coeff-N(i,t)	0.021715	0.034466	0.02798	0.033447	0.030733
	t-stat	*15.94*	*14.88*	*11.48*	*19.44*	*6.89*
	Coeff-N(i,t)²	-0.000263	-0.000639	-0.000672	-0.000458	-0.002656
	t-stat	*-4.93*	*-3.99*	*-3.50*	*-9.61*	*-3.02*
	Coeff-MarketVar	89.7641	88.1827	83.0074	87.0495	71.3563
	t-stat	*5.33*	*5.18*	*4.94*	*5.06*	*4.14*
	Coeff-SectorVar	221.797	22.487	229.406	235.186	238.41
	t-stat	*19.09*	*18.93*	*19.86*	*19.82*	*20.00*
	R²	0.2651	0.2653	0.2555	0.2555	0.2479

	News category	**Analysts Comments**	**Eranings Projections (Clean)**	**Funding/Capital**	**Reg/Govn**	**Contract/Orders**
Fixed effects	**C**	-	-	-	-	-
	t-stat	-	-	-	-	-
	Coeff-N(i,t)	0.067531	0.04905	0.012659	0.090504	0.031828
	t-stat	*7.97*	*13.70*	*2.01*	*13.79*	*3.36*
	Coeff-N(i,t)²	-0.005736	-0.001888	-0.000328	-0.003787	-0.006195
	t-stat	*-1.85*	*-3.25*	*-0.41*	*-4.21*	*-2.04*
	Coeff-MarketVar	202.813	75.0816	73.8235	82.3753	73.2795
	t-stat	*11.52*	*4.39*	*4.28*	*4.77*	*4.24*
	Coeff-SectorVar	209.591	234.698	238.445	238.054	239.268
	t-stat	*17.86*	*19.87*	*20.00*	*19.96*	*20.04*
	R²	0.2476	0.2529	0.2468	0.2510	0.2469

Figure 4-8: Coefficients of various regressions with different subsets of news releases

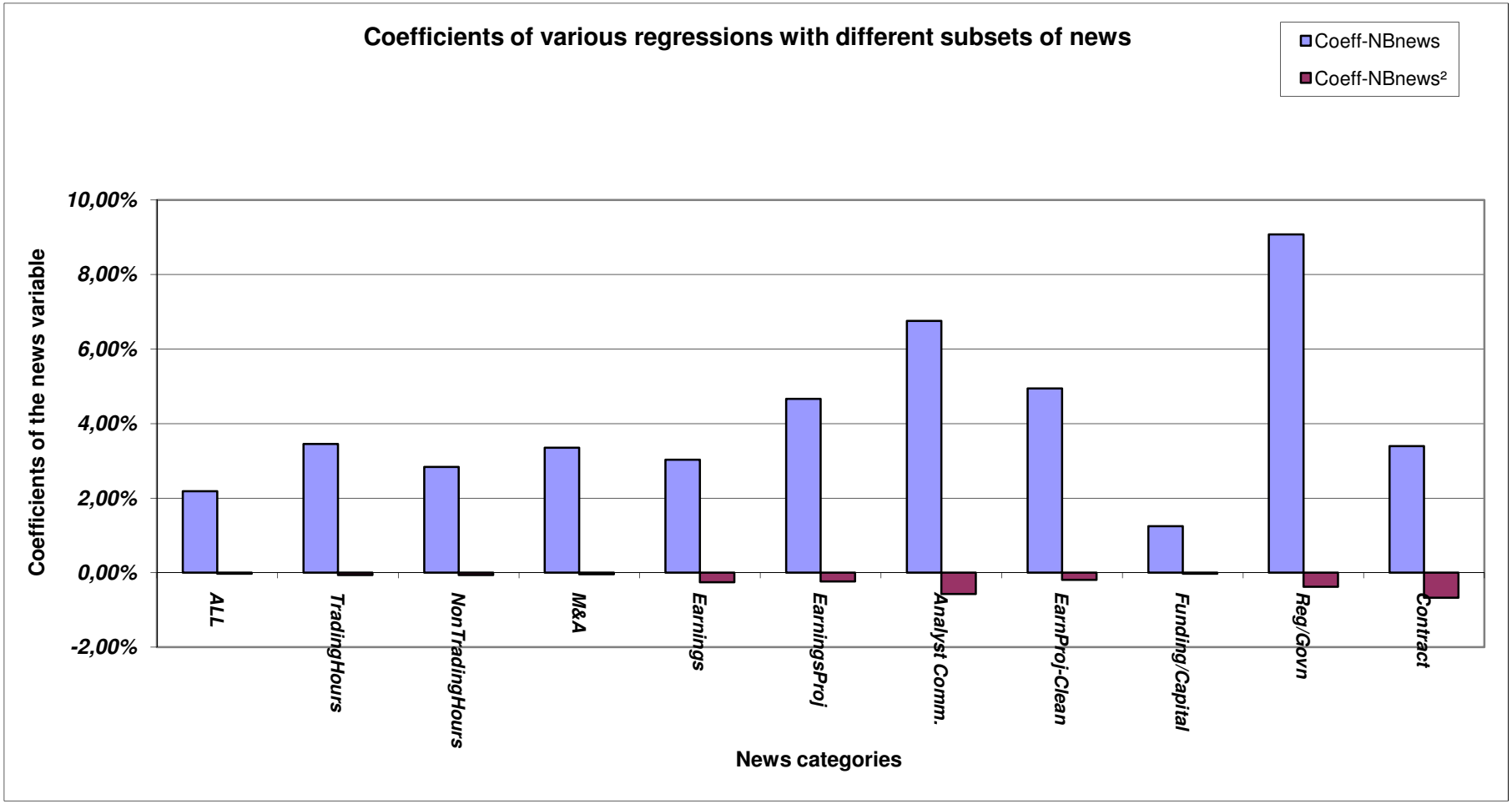


Table 4-11: Panel data analysis for various types of news: categories are mixed together

Fixed effects approach (the dependent variable is the daily probability of being in a high volatility regime). The conditional variance of the market portfolio and the conditional variance of the sector index are estimated via a GARCH (1,1). The panel is composed of 1270 observations for each of the 40 securities. The first regression introduces the two timing categories; the second one introduces six subject categories; and the third one introduces the square variables in addition of the variables of the second regression. The covariance matrix is estimated with the convergent estimator of Newey and West (1987) so that t-stat are robust to serial correlation and heteroscedasticity.

		Y=proba		
		1	***2***	***3***
GMM	**Coeff-N(i,t) (Trading Hours)-weighted**	0.085092	-	-
	t-stat	*14.67*	-	-
	Coeff-N(i,t) (Trading Hours)²	-0.0039444	-	-
	t-stat	*-3.43*	-	-
	Coeff-N(i,t) (Non-Trading Hours)-weighted	0.018452	-	-
	t-stat	*5.18*	-	-
	Coeff-N(i,t) (Non-Trading Hours)²	-0.0011239	-	-
	t-stat	*-2.49*	-	-
	Coeff-N(i,t) (M&A)	-	0.017097	0.023615
	t-stat	-	*9.70*	*12.27*
	Coeff-N(i,t) (M&A)²	-	-	-0.000282
	t-stat	-	-	*-3.61*
	Coeff-N(i,t) (Earnings)	-	-0.001214	0.005683
	t-stat	-	*-0.38*	*1.23*
	Coeff-N(i,t) (Earnings)²	-	-	-0.001475
	t-stat	-	-	*-1.67*
	Coeff-N(i,t) (Earnings Proj)	-	0.025703	0.036752
	t-stat	-	*9.20*	*9.66*
	Coeff-N(i,t) (Earnings Proj)²	-	-	-0.002053
	t-stat	-	-	*-3.19*
	Coeff-N(i,t) (Funding/Kal)	-	0.00024927	-0.001830
	t-stat	-	*0.05*	*-0.41*
	Coeff-N(i,t) (Funding/Kal)²	-	-	-
	t-stat	-	-	-
	Coeff-N(i,t) (Reg/Govn)	-	0.049908	0.063116
	t-stat	-	*6.53*	*9.13*
	Coeff-N(i,t) (Reg/Govn))²	-	-	-0.003961
	t-stat	-	-	*-3.21*
	Coeff-N(i,t) (Contracts)	-	0.00854056	0.017276
	t-stat	-	*1.42*	*1.84*
	Coeff-N(i,t) (Contracts)²	-	-	-0.005311
	t-stat	-	-	*-1.74*
	Coeff-MarketVar	88.632	69.2247	75.8451
	t-stat	*5.38*	*4.05*	*4.58*
	Coeff-SectorVar	222.021	246.928	242.081
	t-stat	*19.69*	*20.71*	*21.25*
	R²	0.2666	0.2576	0.2602

4.1.5.3 Informational content and asymmetry issues

If the daily number of news releases has been proven to be a quite fair proxy of the public information available on the market and to significantly drive the conditional volatility, we still may regret that this variable does not tell anything about the impact of the informational content of news. Indeed, we can legitimately expect that an announcement delivering highly valuable information for investors will induce a stronger increase in the conditional and contemporaneous volatility than any less important piece of news. Our objective is to bring some answer to the following question: "do news releases with different informational contents have different impacts on the conditional volatility?" We opt for the absolute value of the daily abnormal stock return as our proxy of the informational and value content of all the public and idiosyncratic information released each day ($|AR|$), even though we are well aware that this noisy proxy encompasses many more elements and effects than the value effect of information arrivals alone. In practice, we define |AR| as the absolute value of the residual $\varepsilon_{i,t}$ of Equation (4-8).We also compute a measure of informational content per news release by dividing the variable defined above by the number of announcements of the day ("*InfCont*" = "$|AR|/N$").

Another issue is brought up by previous empirical research linked to the vast stream of literature concerned with the development of GARCH models. Works such those of Engle and Ng (1993) have concluded to the asymmetry of the volatility response to news. Bad news appears to induce a larger increase in conditional volatility than good one. Accounting for this differentiated effect seems therefore of prime of importance. We aim at answering the following question: "does good and bad news impact the conditional volatility the same way? Or is there any significant asymmetric effect?" Concretely, we simply use the sign of the daily abnormal return (defined above as the residual of Equation (4-8)) to classify news into good and bad news[49].

In order to gain some insight of the asymmetry and "informational content" patterns across our two volatility regime, Table 4-12 provides some basic and nevertheless interesting descriptive statistics. It shows not surprisingly that the average abnormal return in absolute value ($|AR|$) is greater in the high volatility regime than in the low one, since $|AR|$ is a proxy of the stock volatility as well.

After this first and superficial look at some specificities of our sample, we undertake a more formal analysis and run the following regressions:

$$P[S_{i,t}=2]=\beta_{i,0}+\beta_{i,1}SignAR_{i,t}+\beta_{i,4}|AR_{i,t}|+\sum_{j=1}^{J}\delta_{i,j}X_{i,j,t}+\eta_{i,t} \quad (4\text{-}11)$$

$$P[S_{i,t}=2]=\beta_{i,0}+\beta_{i,1}SignAR_{i,t}+\beta_{i,1}N_{i,t}+\beta_{i,2}N_{i,t}^{2}+\beta_{i,4}|AR_{i,t}|+\sum\delta_{i,j}X_{i,j,t}+\eta_{i,t} \quad (4\text{-}12)$$

$$P[S_{i,t}=2]=\beta_{i,0}+\beta_{i,1}SignAR_{i,t}+\beta_{i,4}InfCont_{i,t}+\sum_{j=1}^{J}\delta_{i,j}X_{i,j,t}+\eta_{i,t} \quad (4\text{-}13)$$

$$P[S_{i,t}=2]=\beta_{i,0}+\beta_{i,1}SignAR_{i,t}+\beta_{i,1}N_{i,t}+\beta_{i,2}N_{i,t}^{2}+\beta_{i,4}InfCont_{i,t}+\sum_{j=1}^{J}\delta_{i,j}X_{i,j,t}+\eta_{i,t} \quad (4\text{-}14)$$

where $SignAR_{i,t} = 1$ if $AR_{i,t}$ is positive and 0 otherwise;

$AR_{i,t} = \varepsilon_{i,t}$ from Equation (4-8);

$InfCont_{i,t} = |AR_{i,t}|/N_{i,t}$

The four columns of Table 4-13 give the results of each of these four regressions.

The first column/regression (4-11) shows that negative abnormal returns have on average a stronger impact on the stock volatility than positive abnormal returns, and this difference of impact is significant at a 5%-level.

Why is the probability of being in a high volatility state higher when the daily AR is negative than when it is positive? A possible explanation could be that days with negative AR experience more news wires than days with positive AR. If it is true, the introduction of the news variable should remove any asymmetric effect between positive and negative AR and make the variable SignAR insignificant. The results of the second column/regression (4-12) lead us to reject this argument: asymmetry persists even after controlling for the impact of the volume of news on the stock volatility.

The absolute value of the daily abnormal return (|AR|) is an interesting proxy of the informational content of all publications released during a day, but it remains a collective measure rather than an individual one, since it does not take into account the information arrivals at all. That is why we choose to also introduce an individual measure of informational content that controls for volume effects, that is, the InfCont defined above. The last two columns/regressions (4-13) and (4-14) show that with such an explicative variable, asymmetry effects completely vanish, since the coefficient of

[49] Notice that with such a classification rule, all news releases of the same day will be classified in the same category.

SignAR turns out to be completely insignificant. We understand this new result as such: one good and one bad announcement with identical informational content affect the conditional volatility the same way. Negative abnormal returns have a stronger impact on the stock volatility than positive AR because they are on average more "informative", or at least they are perceived as such by the investors. Asymmetry between good and bad news disappears once news informational content is accounted for.

Table 4-12: Summary Statistics across volatility regimes on the French market

This table presents some summary statistics for the decomposition regime of the 40 securities studied. The high volatility regime consists of all days where $P[S_{i,t} = 2] \geq 50\%$, while the low volatility regime consists of all remaining days. AR = residual $\varepsilon_{i,t}$ of Equation (4-8). InfCont is defined as ARlevel/N(i,t) when N(i,t) > 0 and is set to 0 if N(i,t) = 0; it is used as a proxy for the average individual informational content of each news release of a given day.

	# Days	Mean-AR	Mean-|AR|	% AR > 0	Mean-News	Mean - InfCont
High Volatility Regime	18642	0.0010	0.0245	50.17%	2.63	0.006855
Low Volatility Regime	32158	-0.0001	0.0114	49.02%	1.81	0.003043
Total	50800	0.0003	0.0162	49.44%	2.11	0.004442

	# Days	Mean-AR	Mean-|AR|	% High regime	Mean-News	Mean - InfCont
Days AR >0	25116	0.0167	0.0167	37.24%	2.16	0.004543
Days AR<0	25684	-0.0157	0.0157	36.17%	2.06	0.004343
Total	50800	0.0003	0.0162	36.70%	2.11	0.004442

Table 4-13: Informational value and asymmetry issues

This table presents various regressions of the smooth probability of being in the high volatility regime on some expected proxies of information content. The purpose of SignAR is to take possible asymmetry effects into account. It is equal to 1 if AR>0 ("good news") and 0 otherwise ("bad news"). N(i,t) is the daily number of news releases. AR = residual $\varepsilon_{i,t}$ of Equation (4-8). InfCont is defined as ARlevel/N(i,t) when N(i,t) > 0 and is set to 0 if N(i,t) = 0; it is used as a proxy for the average individual informational content of each news release of a given day. The covariance matrix is estimated with the convergent estimator of Newey and West (1987) so that the t-stats are robust to serial correlation and heteroscedasticity.

		Dependent Variable : Smooth Probability			
		1	**2**	**3**	**4**
Fixed Effects	**C**	-	-	-	-
	t-stat	-	-	-	-
	Coeff-SignAR	-0.005369	-0.00582657	0.0014457	-0.00013019
	t-stat	*-1.99*	*-2.17*	*0.49*	*-0.05*
	Coeff-N(i,t)	-	0.00828465	-	0.020714
	t-stat	-	*5.93*	-	*17.49*
	Coeff-N(i,t)²	-	-0.00015274	-	-0.00022723
	t-stat	-	*-2.72*	-	*-5.06*
	Coeff-|AR|	9.91529	9.67891	-	-
	t-stat	*47.21*	*46.02*	-	-
	Coeff-InfCont	-	-	9.64871	9.56057
	t-stat	-	-	*46.12*	*46.15*
	Coeff-MarketVar	47.5023	51.8505	63.5813	81.1221
	t-stat	*3.15*	*3.56*	*3.89*	*5.06*
	Coeff-SectorVar	179.498	176.823	228.903	211.283
	t-stat	*16.77*	*17.28*	*20.29*	*19.17*
	R²	0.4059	0.4084	0.2926	0.3100

4.1.6 Conclusion

Without a shadow of doubt, research works aiming at accurately identifying and understanding the link between information diffusion and stock prices are of prime importance and interest, as much for the academic world as for business corporations. In this general context, the main objective of our study is to investigate the impact of the news arrivals intensity on the stock price volatility. We use daily prices and Reuters news releases concerning all stocks pertaining to the CAC40 index over the period ranging from January 1, 1999 to December 31, 2003. We first begin with a partial replication of the Kalev and al. (2004)'s GARCH model and we find, like them, that public information explains a large part of the daily volatility persistence. We apply a Markov switching regression (MSR) model to the data in order to determine the daily probability of the volatility regime to be in a high level state or a low level state. Our developed MSR model presents many interesting features: its structure is more general than the GARCH models' one, they more often converge towards a stable solution and they implicitly rest on a theoretical argument: the Mixture of Distribution Hypothesis. We find for each stock a positive and (5%-)significant relation between the daily number of news and the daily probability of being in a high volatility state. The panel analysis confirms these results since the above detected positive relation remains highly significant, even when controlling for market volume and conditional sector and market variances. The daily number of news alone explains not less than 14% of the probability of being in a high variance regime, and 26% in conjunction with the two main control variables.

A positive and significant relation between news arrivals intensity and stock prices volatility seems thus to be reasonable and consistent. Hence it follows that news contains some relevant and valuable information to the market, since news wires have just been proved to significantly impact the stock return distribution. Such a finding is consistent with Roll's idea (and others before him) that the return distribution is mixed ("with news" and "no news" distributions) and driven by news arrivals frequency. It also brings new support in favor of the Mixture of Distribution Hypothesis.

Moreover, further analysis shows that the impact of news intensity on stock volatility is positive but marginally decreasing. The concavity of the relation between number of news releases and stock volatility could indicate a possible phenomenon of saturation, as if too many news didn't bring anymore additional valuable information to the market.

By splitting our news sample into different categories according to the timing and the subject, we also find that announcements released during trading hours have a greater impact than news published during non-trading hours. Regarding the content of the news releases, we show that news concerning *Regulation and Government Policy* has the greatest impact on the stock volatility. *Earnings Projections* (especially the subset of *Analyst Comments and Recommendations*) and *Mergers and Acquisitions* come just after and seem to significantly affect the stock volatility too. The impact of *Earnings* is significant too, but to a limited extent only.

Asymmetry effects reported by the GARCH literature are examined too. Results are globally consistent with previous works, since bad news releases seem to affect more the conditional volatility than good ones, even though their significance is relatively modest. Interestingly, this well-known asymmetric effect vanishes once we account for the informational content of the news releases: good and bad announcements with identical informational content affect the conditional volatility the same way.

Now that the link between news frequency and stock price volatility is quite ascertained, future research could be extended in order to account for private information through the introduction of an appropriate proxy of this phenomenon. From a more dynamic point of view, switching from a daily scale to an intraday scale in order to capture very short-term volatility effects is another potential promising trail of research.

4.2 The US case

4.2.1 Sample Selection

Our final sample covers a period ranging from May 13, 1998 to December 31, 2003 (i.e. 1418 trading days). Our firms' universe is composed of the 458 US stocks that were part of the S&P 500 index in May 2006 and which survived the filtering process. We have indeed deleted all stocks (i.e. 42) that were acquired, merged or that have purely vanished over the sample period, as well as stocks with insufficient news or financial data. We do not consider news prior to May 13, 1998, because the time of arrival of news is a relatively new feature of Factiva and was not systematically provided before this date. As this item appears to be useful in the aggregation process of announcements and the subsequent regressions, and since we wish to work on a dataset as homogenous as possible, we decide to restrict our attention to this specific time range, as well as to 458 companies out of the 500 initially considered, for some necessary data lack for 42 of them.

Eventually, we are left with a broad and representative sample of the US market since it accounts for not less than 75% of the US market value. The average total market value of our sample is $25,000.489 billion. The mean (median) company market value is $55.653 billion ($2.175 billion). For all firms included in this sample, we use daily closing prices from May 13, 1998 to December 31, 2003. The three largest sectors in the sample are the commercial banking (5,8%), Electricity/Gas Utilities (3,8%) and Semiconductors (3,4%). We use as a market portfolio the CRSP EW. The data are obtained from the CRSP-tapes.

All press releases were collected through the Factiva data base web site[50]. Factiva is a Dow Jones and Reuters company that provides professionals with business news and information sources like The Wall Street Journal or The Financial Times for example. Our sample is composed of all corporate and industrial news wires concerning firms included in the S&P 500 index and recorded by Dow Jones and Reuters from May 13, 1998 to December 31, 2003. Dow Jones and Reuters news wires are the two major sources of news available in Factiva and are of course redundant. We have performed the whole analysis with each dataset separately, and we have found very close results.

[50] http://www.factiva.com/

Therefore, we have chosen to present only the statistics and results concerning Reuters news in this work.

For each news wire, we have got the following fields: the accession number (AN), headline (HD), word count (WC), publication date (PD) and time (ET), source name (SN) and code descriptor (SC), lead paragraph (LP), company (concerned by the news release) code (CO), industry code (IN), subject code (NS), region code (RE), Dow Jones codes (DJIC) and descriptors (DJID), information provider codes (IPC) and Reuters codes (RBBCM). Company, industry and subject codes are the main fields of interest, since they allow us to classify each news release in sector and in topic. One piece of news can of course concern several companies, industries and subjects.

To avoid any redundancy and duplicate announcements that do not bring any additional information value, we restrict the sample to news released by Reuters only (or Dow Jones only when we construct the Dow Jones sample). For the same reason, we eliminate all news releases with the same headlines and lead paragraphs. Eventually, after having excluded news items with one missing field or more, we are left with 335,202 Reuters financial communications over the whole sample period Nov 1998 – Dec 2003.

Figure 4-9 (resp. Figure 4-10) displays the evolution over time of the total daily (resp. weekly) number of news wires. We can distinguish three different “regimes”. The first one ranges from November 1995 to May 1998 and presents a relatively low news arrival rate. Indeed, at that time, the time of the news release was not yet systematically provided. Since May 1998 this feature has been generalized and the frequency of announcements becomes much more significant. That is why we begin our sample in May 1998 in order to have a homogenous sample. Latter years exhibit a decline in the number of announcements, because the news releases are not integrated immediately in Factiva but progressively, and it may occur that some pieces of news dating back to 2003 are only incorporated in the Factiva database in 2006 only. That is why we stop the sample on December 30, 2003, in order to ensure a minimal homogeneity. From now, all subsequent figures and tables will be given for this specific sample that ranges from May 13, 1998, to December 31, 2003.

Figure 4-9: Daily number of news releases on the US market from Nov. 1995 to Dec. 2005

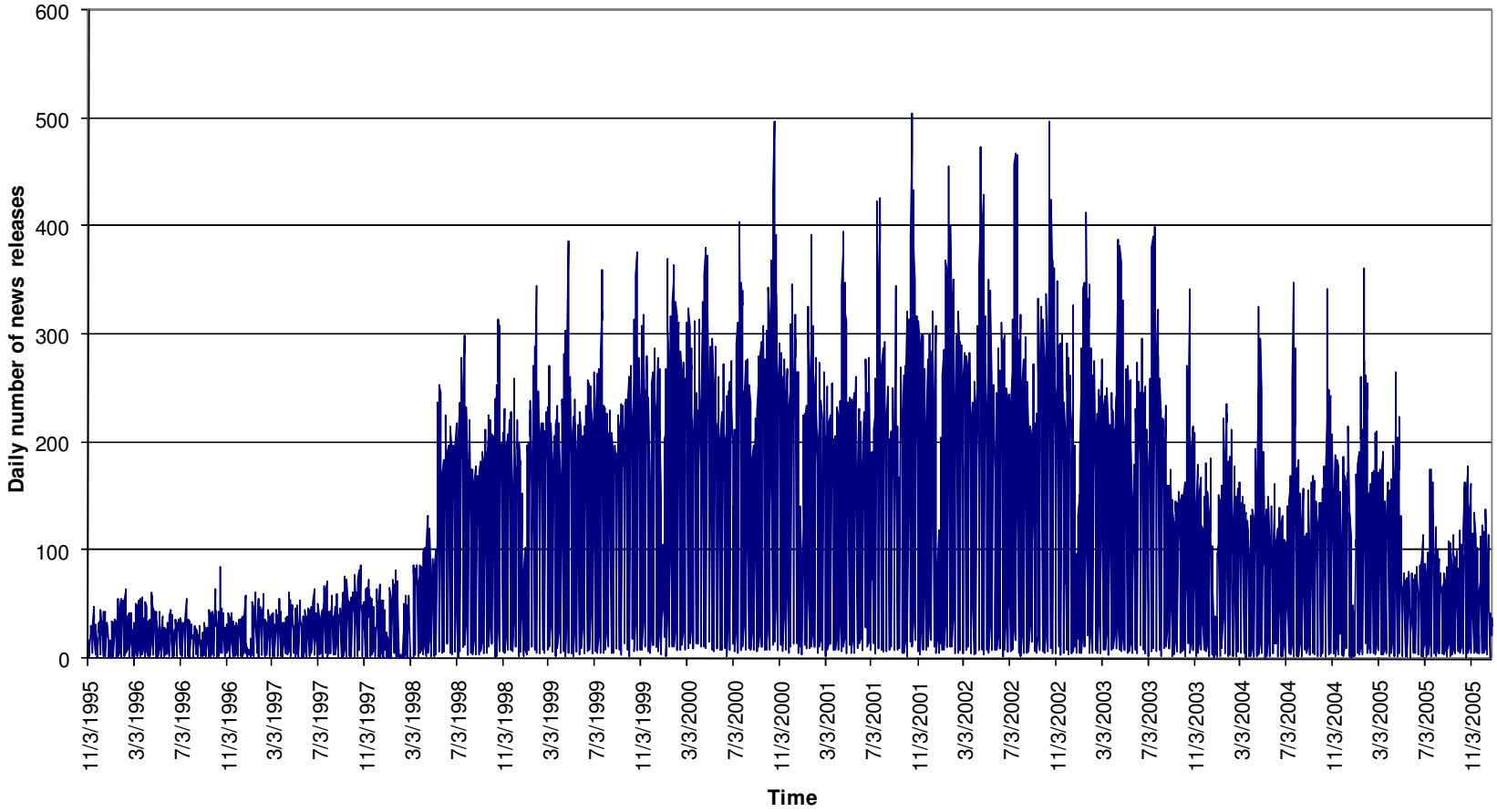


Figure 4-10: Weekly number of news releases on the US market from Nov. 1995 to Dec. 2005

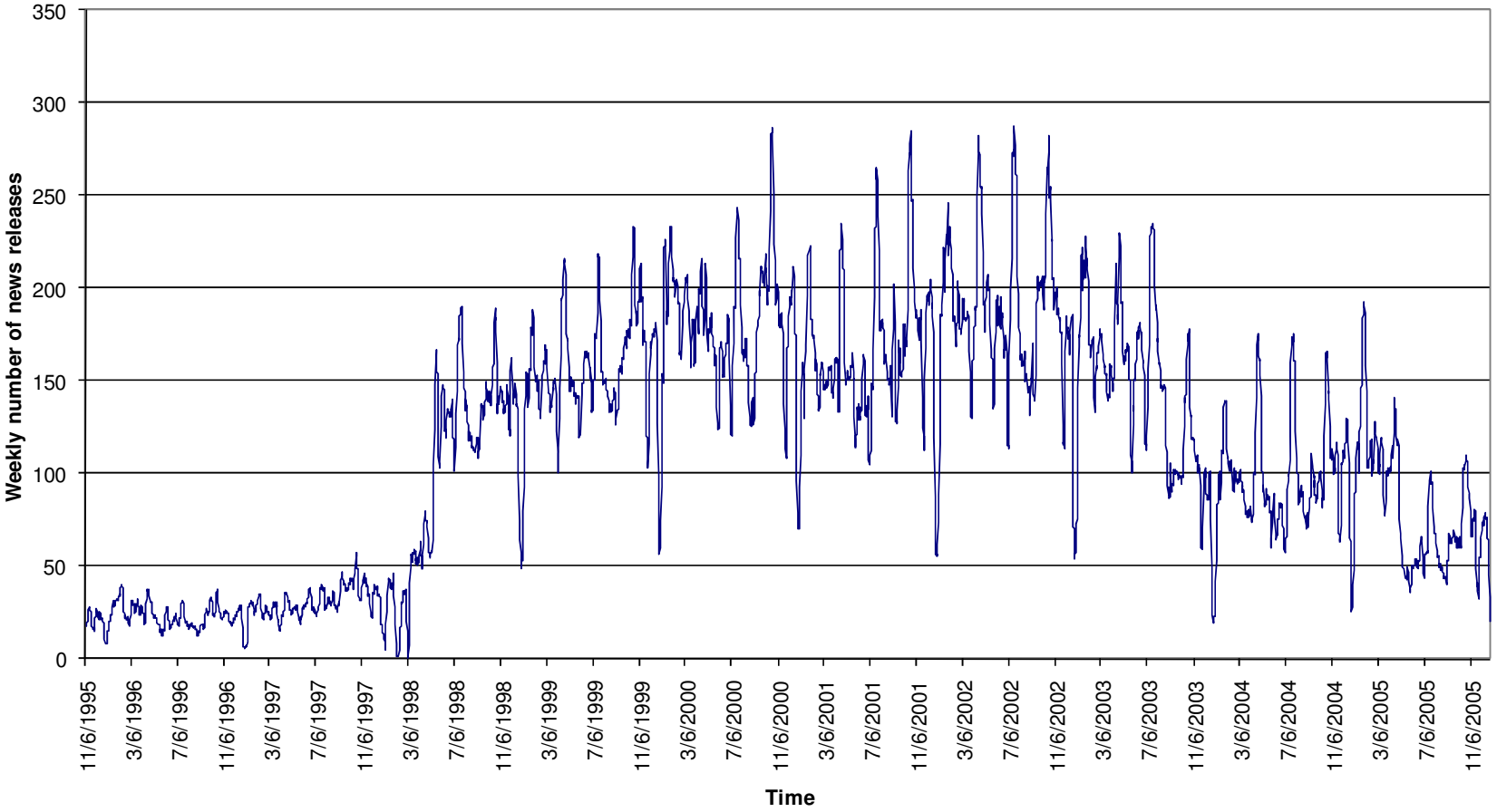


Figure 4-11: Total news releases average by day of the week on the US market

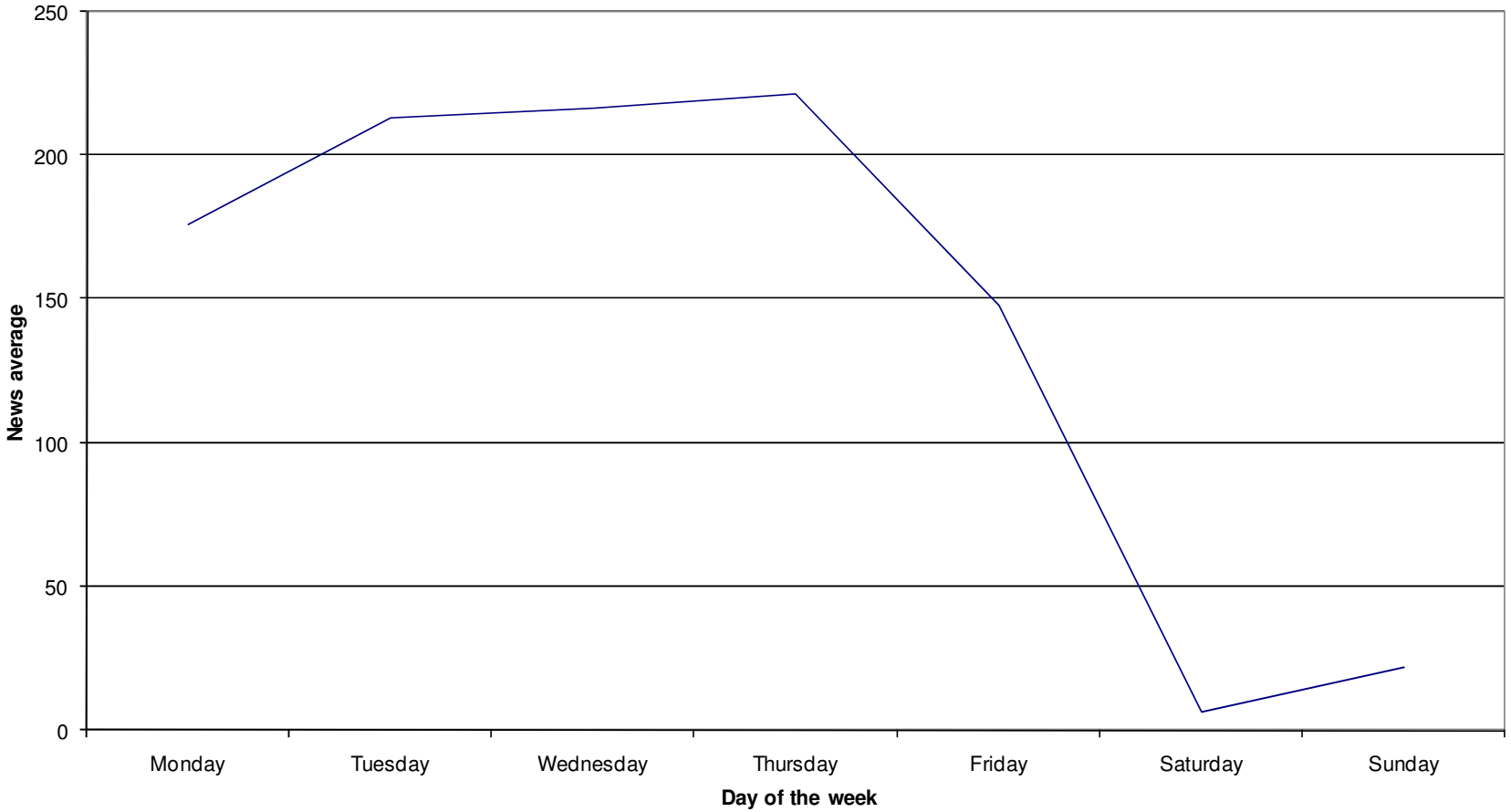
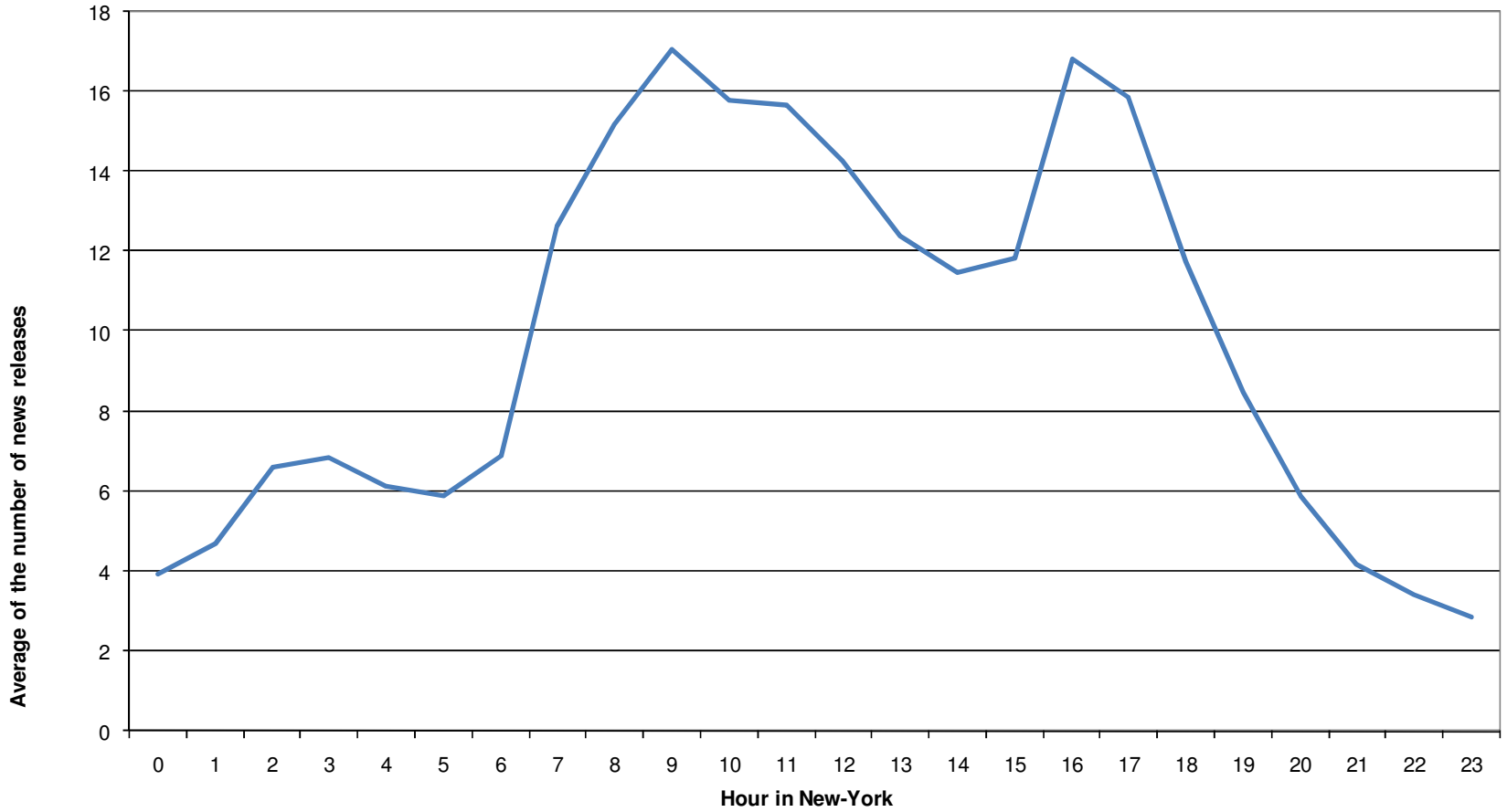


Figure 4-12: Total news releases average by hour of the day in New-York on the US market



At first sight, the absence of any clear trend in Figure 4-9 and Figure 4-10 indicates that the news time-series is rather stationary[51]. This characteristic of our data presents the advantage to eliminate the risk of spurious results due to a possible simultaneous increase over time of the stock volatility. Not surprisingly at all, the Christmas-New Year period experiences a news arrivals rate much below the average. Finally, there are recognizable peaks every quarter (around each January, April, July and October 20), that probably correspond to the wave of the quarterly releases of the intermediate figures and earnings of companies.

Figure 4-11 focuses on a possible and expected weekly seasonality in the data and shows indeed that the average number of news announcements released during the week-end is much lower than the one of the other weekdays. Since the stock exchange is closed during the week-end, this seasonality is not problematic: news released during the week-end will simply be aggregated to news published on the next Monday. It can also be noticed that the number of announcements released on Tuesday, Wednesday and Thursday is roughly the same, with Monday and especially Friday lower than the former.

Figure 4-12 depicts the evolution of the frequency of announcements arrivals throughout the day, at New-York's time. There are peaks at 9 am and 4 pm, and the activity seems to be globally more sustained in the morning than in the afternoon. The "lunch drop" is easily recognizable.

Table 4-14 gives the Factiva classification of companies. However, because of the high number of companies in the sample (458), Table 4-14 is restricted to the 50 most important companies ranked by their sales in May 2006. The Factiva classification of industries and subjects is the same as the one of the French market. For a reminder, the reader is referred to Table 4-2 and Table 4-3.

[51] We have checked this stationarity hypothesis more formally through the unit root test of Dickey-Fuller that rejects the hypothesis of unit root at a 0.1% level. So we can conclude that the series is stationary.

Table 4-14: Factiva classification of 50 companies ("CO" field) on the US market

CO	Company Name
EXXN	ExxonMobil Corporation
WLMRT	Wal-Mart Stores, Inc.
SOCAL	Chevron Corporation
GNMOC	General Motors Corporation
PHILP	ConocoPhillips
FRDMO	Ford Motor Company
GNELC	General Electric Company
AIU	American International Group, Inc.
PHILM	Altria Group, Inc.
IBM	International Business Machines Corp.
HEWLP	Hewlett-Packard Company
VLRNGY	Valero Energy Corp.
HOMD	Home Depot, Inc.
FORMK	McKesson Corporation
CCRED	Citigroup Inc.
BELATT	Verizon Communications
CDIST	Cardinal Health, Inc.
USSTEL	Marathon Oil Corporation
KROGR	The Kroger Co.
NCNBCO	Bank of America Corporation
PRGML	The Procter & Gamble Company
PCLTD	Dell Inc.
BOEING	Boeing Company, The
ALCHS	AmerisourceBergen Corp.
FEDNAT	Fannie Mae
PCLB	Costco Wholesale Corp.
DAYHUD	Target Corporation
DWITD	Morgan Stanley
PFIZ	Pfizer Inc.
JONJON	Johnson & Johnson
KMRT	Sears Holdings Corporation
MERLY	Merrill Lynch & Co., Inc.
DOWCH	The Dow Chemical Company
UHELC	UnitedHealth Group Inc.
CNYC	JPMorgan Chase & Co.
ANMINC	WellPoint, Inc.
METLFE	MetLife, Inc.
SBCATT	AT&T Inc.
AMRONL	Time Warner Inc.
GLDMNS	Goldman Sachs Group, Inc.
LOWCO	Lowes Companies, Inc.
UTC	United Technologies Corporation
UPSER	United Parcel Service
WLGRN	Walgreen Company
ALBERS	Albertsons, Inc.
MCROST	Microsoft Corporation
HWLY	Tyco International Ltd.
INTL	Intel Corporation
SFWFS	Safeway Inc.

There were initially 500 different companies in the sample (all constituents of the S&P 500 in May 2006). However, only 458 out of these 500 firms will be analyzed in the subsequent GARCH and MSR econometric models, since some of these companies were created, merged, or acquired during the sample period, or present insufficient financial or news data. Figure 4-13 shows that news releases intensity was the highest for Microsoft (5.19% of all news releases concerned Microsoft in some way), General Motors (4.52%), Intel (3.59%), Ford (3.53%) and Time Warner (3.21%)[52]. Contrary to the French market for example, the sample is quite homogeneous: we do not observe one or two firms presenting news intensity levels much above the rest of the market.

Concerning industry sectors, they are split into ten general categories and 56 sub-categories. Figure 4-14 shows that the three sectors with the highest news frequency from May 1998 to December 2003 were *Metal, Goods and Engineering* (44.98%), *Financial and Business Services* (36.01%) and *Transport and Communication* (22.96%).

Subjects have also been divided into 5 general categories and 105 more specific ones, but only the general category "corporate and industrial news" is retained in our sample. From Figure 4-15 one can observe that the 9 most common news subjects are *earnings projections* (22.44%), *earnings* (17.99%), *ownership changes* (17.51%), *mergers and acquisitions* (hereafter *M&A*, which is in fact a sub-category of *ownership changes*) (16.07%), *funding/capital* (6.65%), *capacity/facilities* (4.80%), *analyst comment/recommendation* (a subcategory of *earnings projections*) (4.33%), *contracts/orders* (4.26%) and *regulation/government policy* (3.67%). According to Andersen (1996), we expect the impact of various news releases to be different across categories. One of the purposes of the subsequent econometric analysis will be to give some empirical support to such a theoretical prediction.

Table 4-12 provides an extensive set of summary statistics (mean, min, max, standard deviation, skewness, excess kurtosis and sum), for the whole sample and by company (Panel A), by industry (Panel B), by subject (Panel C) and by timing (Panel D). Note that only the 10 general categories of industries and all subcategories of topics of *corporate/industrial news* (CCAT) are considered in Panels B and C. Concerning Panel D, the "timing" classification of a news release depends whether it is published when

[52] Percentages do not sum up to 100%, because one news release often concerns more than one company, industry or subject. It follows that categories are not exclusive (but they are exhaustive).

the market is open (trading hours) or not (namely overnight and week-end). Daily news frequency ranges from 25 to 943 news releases a day[53], with an average and a standard deviation of respectively 236.41 and 66.64. The skewness and excess kurtosis (moments of order three and four) are moderate and indicate that the total news probability distribution is asymptotically roughly Gaussian, even if it presents some substantial tails.

Panels A, B, C and D are quite interesting since they offer valuable information about possible cross-sectional differences in the news arrival rate. One of the main conclusions that can be drawn from the analysis of Table 4-12 is the following: the greater the total number of news releases concerning a company (or an industry, or a subject), the higher the standard deviation and the lower the skewness and kurtosis of the distribution of the daily number of news releases. Indeed, the behavior of the discrete random variable "number of news releases" is a priori best described by a Poisson probability distribution, and this distribution is known to converge to a Gaussian distribution as the number of observations increases.

[53] We only consider **trading** days, i.e. days when the market is open. If a piece of news is released on Saturday, it will be aggregated to the news releases of the next trading day (here, if there is no holiday, Monday).

Figure 4-13: Total number of news releases by company on the US market

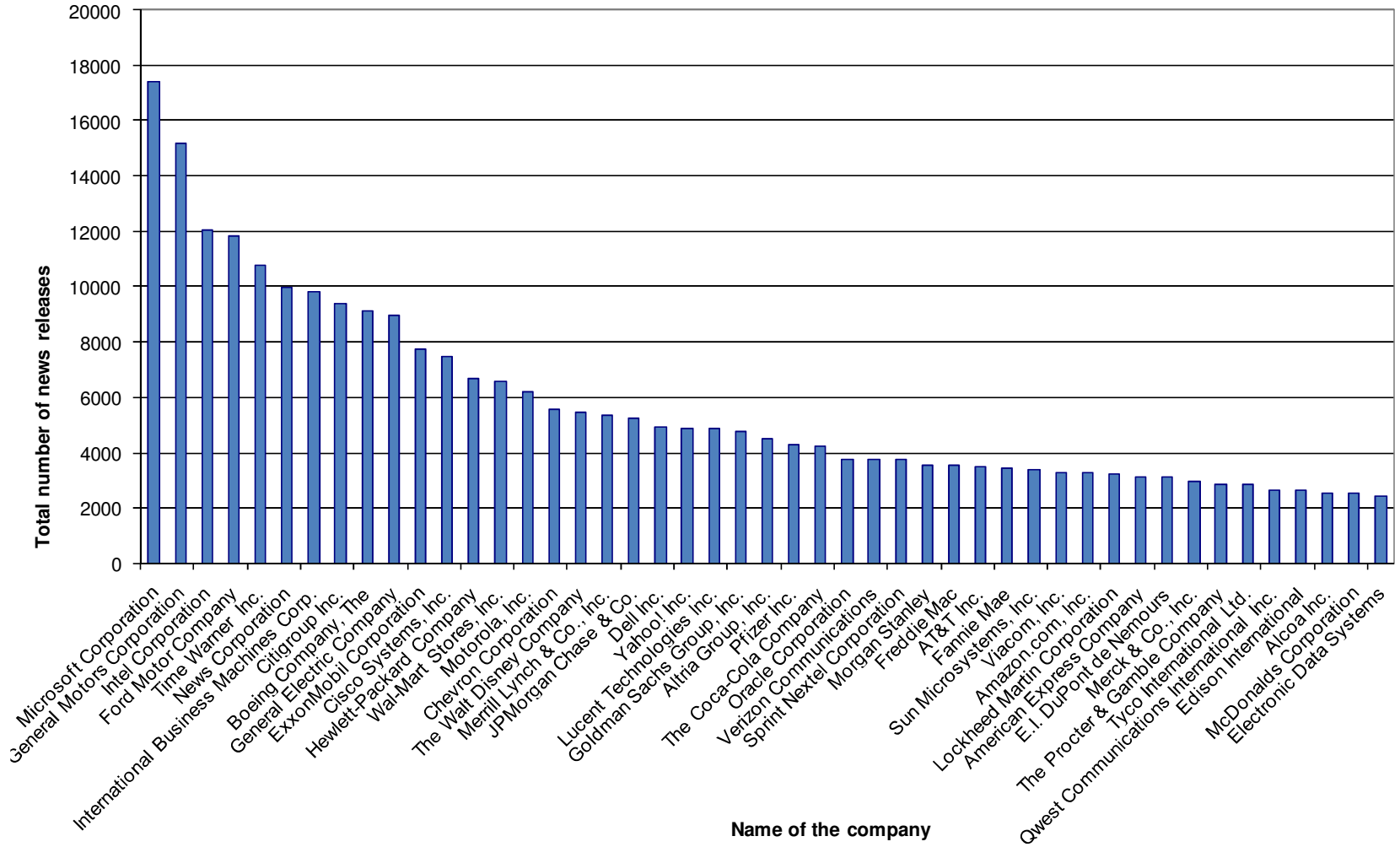


Figure 4-14: Total number of news releases by industry on the US market

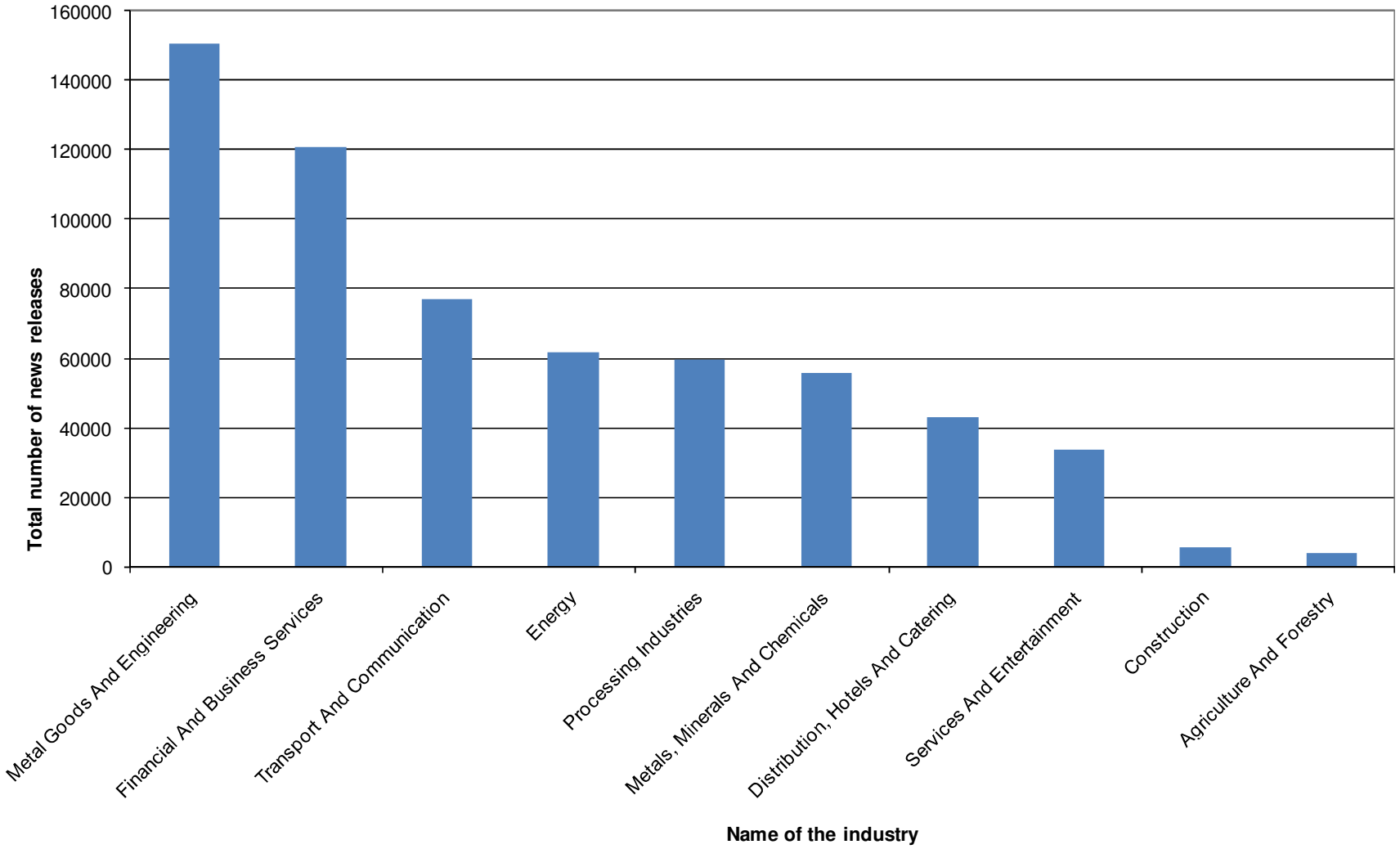


Figure 4-15: Total number of news releases by subject on the US market

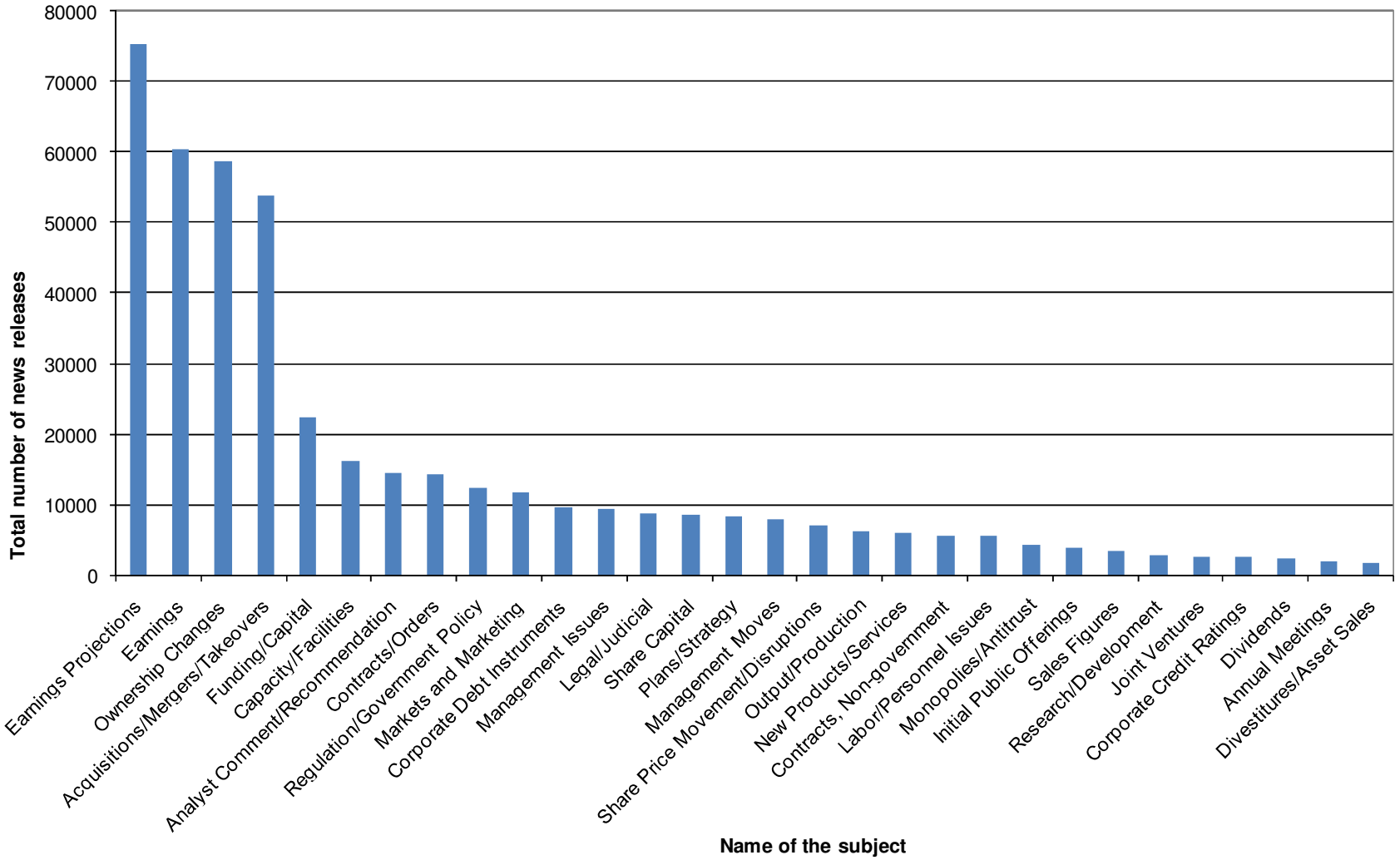


Table 4-15: Summary statistics on the US market

NAME	MEAN	MEDIAN	MAX	MIN	STD	SKEWNESS	KURTOSIS	SUM
All	236.41	229	943	25	66.64	1.31	9.72	335,232
Panel A: Daily number of news releases by company								
Microsoft Corporation	12.28	9	180	0	12.52	4.93	41.14	17,410
General Motors Corporation	10.68	9	61	0	7.49	2.03	6.89	15,148
Intel Corporation	8.48	6	177	0	11.28	5.57	52.19	12,023
Ford Motor Company	8.34	7	77	0	6.79	2.47	12.19	11,824
Time Warner Inc.	7.60	6	183	0	8.25	8.86	156.57	10,771
News Corporation	7.05	6	42	0	4.92	1.95	6.78	9,992
International Business Machines Corp.	6.91	5	152	0	9.52	6.12	61.25	9,796
Citigroup Inc.	6.62	5	53	0	5.59	2.77	12.89	9,390
Boeing Company, The	6.45	5	52	0	5.11	2.61	13.14	9,142
General Electric Company	6.31	5	48	0	5.37	2.54	10.22	8,953
ExxonMobil Corporation	5.46	5	105	0	5.24	7.81	118.59	7,736
Cisco Systems, Inc.	5.28	3	117	0	8.88	5.67	45.12	7,493
Hewlett-Packard Company	4.70	3	103	0	7.24	5.60	47.85	6,663
Wal-Mart Stores, Inc.	4.65	3	39	0	5.04	2.43	7.99	6,592
Motorola, Inc.	4.36	3	65	0	6.59	4.79	30.77	6,186
Chevron Corporation	3.91	3	57	0	3.72	3.99	38.37	5,545
The Walt Disney Company	3.85	3	35	0	3.85	3.08	15.14	5,454
Merrill Lynch & Co., Inc.	3.78	3	26	0	3.39	1.99	6.04	5,353
JPMorgan Chase & Co.	3.69	2	66	0	4.81	5.39	49.12	5,238
Dell Inc.	3.48	2	88	0	6.51	5.85	46.48	4,936
Yahoo! Inc.	3.44	2	77	0	5.70	4.79	35.28	4,884
Lucent Technologies Inc.	3.44	2	87	0	6.08	6.71	66.89	4,874
Goldman Sachs Group, Inc.	3.36	2	26	0	3.45	2.23	7.65	4,768
Altria Group, Inc.	3.18	2	49	0	4.30	3.93	23.46	4,516
Pfizer Inc.	3.02	2	90	0	4.62	7.71	111.13	4,289
The Coca-Cola Company	3.00	2	58	0	4.11	4.55	37.76	4,249
Oracle Corporation	2.67	1	61	0	5.02	5.02	33.76	3,788
Verizon Communications	2.66	2	41	0	3.44	3.97	26.36	3,775
Sprint Nextel Corporation	2.65	1	73	0	4.46	6.00	60.58	3,761
Morgan Stanley	2.50	2	45	0	2.76	4.47	47.23	3,540
Freddie Mac	2.50	2	48	0	3.22	5.56	56.89	3,540
AT&T Inc.	2.47	2	38	0	2.93	3.51	23.73	3,497
Fannie Mae	2.44	2	26	0	2.67	2.69	12.68	3,463
Sun Microsystems, Inc.	2.39	1	47	0	4.27	4.71	29.87	3,387
Viacom, Inc.	2.34	2	31	0	2.58	3.36	20.95	3,314
Amazon.com, Inc.	2.32	1	54	0	4.09	4.66	34.12	3,293
Lockheed Martin Corporation	2.28	2	43	0	2.74	4.62	45.61	3,226
American Express Company	2.20	1	22	0	2.75	2.19	7.14	3,116
E.I. DuPont de Nemours	2.20	1	38	0	3.18	3.92	25.34	3,113
Merck & Co., Inc.	2.10	1	76	0	3.94	8.69	126.32	2,977
The Procter & Gamble Company	2.03	1	74	0	3.96	7.56	97.17	2,885
Tyco International Ltd.	2.03	1	54	0	4.59	5.31	37.74	2,880
Qwest Communications International Inc.	1.89	1	68	0	4.12	7.23	84.64	2,679
Edison International	1.87	1	32	0	2.84	3.98	26.21	2,646
Alcoa Inc.	1.81	1	43	0	3.37	5.28	41.14	2,566
McDonalds Corporation	1.79	1	31	0	2.95	4.06	23.67	2,543
Electronic Data Systems	1.74	1	66	0	3.22	9.04	144.04	2,464

NAME	MEAN	MEDIAN	MAX	MIN	STD	SKEWNESS	KURTOSIS	SUM
Panel B: Daily number of news releases by industry								
Metal Goods And Engineering	106.33	101	356	5	37.60	1.02	3.19	150,771
Financial And Business Services	85.14	82	403	7	28.67	1.66	11.69	120,723
Transport And Communication	54.29	51	286	0	24.23	1.04	5.75	76,985
Energy	43.59	42	156	3	15.14	1.06	4.45	61,813
Processing Industries	42.02	40	155	2	16.01	0.73	1.98	59,585
Metals, Minerals And Chemicals	39.26	36	117	1	16.64	0.96	1.43	55,664
Distribution, Hotels And Catering	30.22	28	104	4	13.51	1.07	1.95	42,845
Services And Entertainment	23.77	23	92	1	10.09	0.87	1.81	33,712
Construction	3.77	3	20	0	2.66	1.27	2.97	5,347
Agriculture And Forestry	2.77	2	31	0	2.43	2.57	16.71	3,921
Panel C: Daily number of news releases by topic								
Earnings Projections	53.05	49	177	7	21.91	1.09	1.95	75,222
Earnings	42.54	29	238	0	40.79	2.11	4.93	60,324
Ownership Changes	41.39	40	122	4	14.20	0.94	2.13	58,693
Acquisitions/Mergers/Takeovers	38.00	36	105	2	13.45	0.94	1.95	53,886
Funding/Capital	15.73	15	76	0	6.85	0.90	4.07	22,305
Capacity/Facilities	11.35	11	60	0	5.30	1.11	5.17	16,092
Analyst Comment/Recommendation	10.24	11	40	0	8.66	0.25	-0.85	14,524
Contracts/Orders	10.08	9	39	0	4.88	0.92	1.61	14,297
Regulation/Government Policy	8.67	8	39	0	5.04	1.13	2.12	12,287
Markets and Marketing	8.27	7	45	0	6.13	1.73	4.38	11,725
Corporate Debt Instruments	6.82	6	33	0	4.73	0.95	1.05	9,676
Management Issues	6.58	6	23	0	3.43	0.70	0.85	9,337
Legal/Judicial	6.17	6	51	0	3.89	2.03	13.97	8,749
Share Capital	6.03	5	47	0	4.08	1.88	9.16	8,546
Plans/Strategy	5.88	2	42	0	7.17	1.42	1.39	8,339
Management Moves	5.57	5	22	0	3.09	0.80	1.22	7,902
Share Price Movement/Disruptions	4.94	5	21	0	4.37	0.48	-0.44	6,998
Output/Production	4.43	3	47	0	4.08	2.11	10.69	6,276
New Products/Services	4.17	4	20	0	3.02	1.25	2.20	5,913
Contracts, Non-government	3.92	4	18	0	3.69	0.62	-0.33	5,553
Labor/Personnel Issues	3.90	3	24	0	2.93	1.30	3.04	5,525
Monopolies/Antitrust	2.96	2	48	0	3.37	3.34	28.20	4,193
Initial Public Offerings	2.73	2	23	0	3.34	1.70	3.86	3,877
Sales Figures	2.38	1	30	0	4.46	3.11	10.66	3,379
Research/Development	1.93	2	11	0	1.78	1.31	2.38	2,743
Joint Ventures	1.81	1	21	0	2.36	1.90	5.81	2,566
Corporate Credit Ratings	1.80	1	17	0	2.25	2.22	7.09	2,555
Dividends	1.58	1	13	0	2.11	1.75	3.47	2,244
Annual Meetings	1.34	0	53	0	3.59	6.92	69.87	1,905
Divestitures/Asset Sales	1.25	1	10	0	1.51	1.76	3.87	1,770
Bankruptcy	1.16	1	38	0	1.97	6.76	96.74	1,640
Pricing	1.03	0	12	0	1.44	1.82	4.67	1,463
Defense Contracts	1.01	1	10	0	1.25	1.76	4.52	1,438
Advertising	0.98	0	11	0	1.43	1.97	5.09	1,387
Lay-offs/Redundancies	0.83	0	12	0	1.56	2.60	8.20	1,177
Financing Agreements	0.82	1	7	0	1.05	1.68	3.76	1,161
Product Safety	0.78	0	11	0	1.35	2.77	10.71	1,104
Government Contracts	0.78	0	7	0	1.17	1.74	3.12	1,099
Privatizations	0.67	0	18	0	1.34	4.07	29.79	947
Stock Listings	0.62	0	8	0	1.01	2.19	6.58	874
Market Share	0.49	0	5	0	0.79	1.99	4.79	690
Domestic Markets	0.46	0	6	0	0.76	1.99	4.93	646
External Markets	0.38	0	4	0	0.65	1.89	4.21	542
Corporate Changes	0.34	0	5	0	0.70	2.57	8.06	479
Patents	0.34	0	6	0	0.73	2.91	11.17	476
Labor Disputes	0.30	0	8	0	0.76	3.68	18.71	428
Licensing Agreements	0.25	0	4	0	0.57	2.44	6.11	360
Regulatory Bodies	0.19	0	4	0	0.51	3.41	14.00	265
Earnings Surprises	0.14	0	7	0	0.53	5.78	44.90	196
Conferences/Exhibitions	0.14	0	3	0	0.40	3.10	9.79	196
Share Buybacks	0.10	0	11	0	0.44	12.59	274.82	138
Trademarks/Copyrights	0.10	0	4	0	0.35	4.64	27.71	135
Sales Promotions	0.09	0	4	0	0.33	4.65	27.22	121
Information Technology	0.07	0	2	0	0.28	3.73	13.94	106
Government Aid	0.07	0	3	0	0.28	4.45	23.18	98
Standards/Standardization	0.07	0	2	0	0.26	4.03	16.63	95
Shareholder-Rights Plans	0.06	0	2	0	0.25	4.92	25.98	78
Insider Dealing	0.04	0	2	0	0.21	5.07	26.43	59
Stock Splits	0.02	0	1	0	0.13	7.34	51.93	25
Franchises	0.02	0	2	0	0.13	8.81	84.19	22
Profiles of Companies	0.00	0	1	0	0.07	14.14	198.28	7
Panel D: Daily number of news releases - trading hours VS non-trading hours								
Trading hours news	149.22	144	830	15	46.87	2.78	31.98	211,592
Non-trading hours news	87.19	87	196	5	26.85	0.27	0.94	123,640

4.2.2 Model Specification

The main model econometric specifications are the same in the US case and in the French case. We refer the reader to Section 4.1.4 for a reminder.

4.2.3 Results

4.2.3.1 GARCH framework

We begin with the GARCH specification already presented in the previous French application. In a first step, we merely reproduce the methodology commonly used in the literature (see Kalev et al. (2004) and Janssen (2004)). It sums up to insert the number of daily announcements concerning a stock into the equation of its variance in a GARCH (1,1) model. The estimation of the latter after inclusion of the variable *NBN* (daily number of news releases) converges for 445 stocks. Out of those 445 estimations, the coefficient associated to the NBN variable is significant at the 5% (resp. 10%) level in 161 (resp. 212) cases and its inclusion decreases the persistence (alpha + beta) in 403 cases. It is important to notice that the estimation of a GARCH on such a long period may not be without problems. Therefore, results are probably not totally exempt of methodological biases. Beyond this reserve, it gives a first insight of how the information stream affects the conditional volatility and to which extent it can explain the well-known persistence of the latter. It is unfeasible to give extensively here the results for each of the 458 companies of the sample, but they are available upon request.

4.2.3.2 MSR framework

4.2.3.2.1 Panel data analysis with the smooth probability

4.2.3.2.1.1 Global analysis

Like in the French case developed in Section 4.1, the first step is to estimate the "smoothed" probabilities to be in each of the two regimes of volatility, low and high. It is then possible to test different hypotheses on the expected relation between those probabilities and the information stream. Our main hypothesis is the MDH, the mixture of distributions hypothesis, already exposed above, which supposes that the distribution of stock returns is linked to the current informational state of the stock. When a large

amount of information occurs on a stock, the contemporaneous return of this stock is expected to be in the high volatility regime.

Our first specification aims at measuring the impact of the announcements arrivals on the probability to be in the high volatility regime, i.e. the regime where the estimated variance is the largest. In order to assess this impact we conduct a fixed-effect panel study estimation[54]:

$$Prob_{i,t} = C_i + C_t + \boldsymbol{\beta}'\mathbf{x}_{i,t} + \varepsilon_{i,t} \quad (4\text{-}15)$$

where $Prob_{i,t}$ stands for the probability of stock *i* at time *t* to be in the high volatility regime and $\mathbf{x}_{i,t}$ the variables for which we anticipate some kind of association with $Prob_{i,t}$. The heterogeneity, i.e. the cross-sectional effect between stocks, is modeled by C_i where C_i is a constant term specific to individuals. We also add a time effect, C_t, in order to take into account the possible evolution of the intercept through time. Specifically, we code this variable C_t as five annual dummies D_k (*k*= 1999, 2000, 2001, 2002 and 2003) equal to 1 if the year of the observation = *k*.

The use of a probability as a dependent variable mainly raises an issue of heteroscedasticity (see Greene (2002)). We solve it by using the Newey-West robust estimator of the covariance matrix.

In this first regression, our proxy for the information flow is a dummy equal to one if there is at least one news release concerning stock *i* in *t* and 0 otherwise. The control variables which are not accounted for by the fixed effect, because they are not constant on the selected period, are the conditional volatility of the market and of the sector index (the latter is formed on the basis of the two first digits of the SIC code).

$$Prob_{i,t} = C_i + C_t + \beta_1 D_{i,t} + \beta_2 varmrkt_t + \beta_3 varsect_{i,t} + \varepsilon_{i,t} \quad (4\text{-}16)$$

[54] In our setting, the Haussman's test (available on request) leads us to prefer the fixed effect specification to the random effect one, like in the French case.

Table 4-16: Panel fixed-effect regression of *Prob* on *D*

Results from the estimation of the fixed-effect panel regression (4-16) for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable is a dummy *D* equal to one if at least one news release specific to stock *i* has been observed in *t*. The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i* *varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2		3	
D	0.0707	***	0.0678	***	0.0675	***
t-stat	*43.9141*		*43.3044*		*43.2025*	
varmrkt			798.3310	***	759.7610	***
t-stat					*46.9459*	
varsect					41.4987	***
t-stat					*4.0297*	
R^2	0.1957		0.2428		0.2442	

It follows from Table 4-16 that after controlling for annual, fixed, market and sectorial effects, the fact to be on a date where at least one Reuters news release has been recorded for a stock increases the probability to be in the high variance regime of about 6.75 points for this stock. Not surprisingly, the coefficients of the market and sector variances appear to be positive and significant, since the stock volatility is known to convey some market component. Such control variables allows us to account (at least partially) for all macroeconomic shocks that affect the stock volatility and that are not covered by our set of firm-specific press releases.

It also appears (from results available on request) that the annual dummies denoted by C_t are systematically significant when they are introduced, which seems to suggest that the probability to be in the high variance regime exhibits some pattern through time and cannot be taken as independent of time. Therefore, we will perform our regressions of Section 4.2.3.2.1.3 (except in Table 4-21 in order to be able to compare US results to the French ones) by systematically including this variable C_t in order to control for the time pattern of the probability of being in the high variance regime.

As a next step we wish to measure the impact of the information volume on this latter probability. Our idea is to introduce four dummies $D_{k,i,t}$ (k=1, …, 4) in order to split the trading days in different categories according to the level of communication observed everyday: $D_{1,i,t}$ (resp. $D_{2,i,t}$, $D_{3,i,t}$, $D_{4,i,t}$) = 1 if the trading day t is in the top 20% (resp.

10%, 5%, 1%) of the days presenting the highest number of news releases concerning stock *i*, and 0 otherwise:

$$Prob_{i,t} = C_i + \beta_1 D_{k,i,t} + \beta_2 varmrkt_t + \beta_3 varsect_{i,t} + \varepsilon_{i,t} \quad (4\text{-}17)$$

Table 4-17: Panel fixed-effect regression of *Prob* on four dummies *D(k)*

Results from the estimation of the fixed-effect panel regression (4-17) for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable is a dummy $D_{1,i,t}$ (resp. $D_{2,i,t}$, $D_{3,i,t}$, $D_{4,i,t}$) = 1 if the trading day *t* is in the top 20% (resp. 10%, 5%, 1%) of the days presenting the highest number of news releases concerning stock *i*. The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2		3		4		5	
D1	0.0959	***							0.0490	***
t-stat	*50.5733*								*22.4654*	
D2			0.1295	***					0.0299	***
t-stat			*57.7950*						*9.1186*	
D3					0.1725	***			0.0693	***
t-stat					*52.8485*				*17.7371*	
D4							0.2657	***	0.1268	***
t-stat							*41.5187*		*19.3808*	
varmrkt	754.4650	***	754.3420	***	754.7200	***	757.8390	***	753.6100	***
t-stat	*43.3078*		*43.2204*		*43.2293*		*43.2443*		*43.2599*	
varsect	45.8325	***	45.8505	***	46.0006	***	46.1884	***	45.5920	***
t-stat	*3.6766*		*3.6663*		*3.6756*		*3.66413**		*3.6567*	
R^2	0.2467		0.2473		0.2471		0.2439		0.2502	

Results of Table 4-17 show clearly that the probability to be in the high variance regime increases with the volume of announcements: a trading day situated in the top 1% of the most "news-intensive" trading days brings 26.6 extra points to the variance of the concerned stock, which cannot be considered as negligible, both from a statistical and economical points of view.

This new evidence leads us to propose a specification including directly the number of announcements in the regression equation:

$$Prob_{i,t} = C_i + \beta_1 NBN_{i,t} + \beta_2 NBN_{i,t}^2 + \beta_3 varmrkt_t + \beta_4 varsect_{i,t} + \varepsilon_{i,t} \quad (4\text{-}18)$$

$NBN_{i,t}$ is the total number of announcements observed for stock *i* on day *t* while $NBN_{i,t}^2$ is simply the square of $NBN_{i,t}$. This last variable is introduced in order to take into account the possibility of a marginally decreasing (or increasing) effect of the number of news releases on the probability of high variance regime.

Table 4-18: Panel fixed-effect regression of *Prob* on the number of news releases *NBN*

Results from the estimation of the fixed-effect panel regression (4-18) for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t*. The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i* *varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2		3		4	
NBN	0.0137	***	0.0205	***	0.0200	***	0.0200	***
t-stat	*26.8405*		*28.9200*		*28.9036*		*28.8303*	
NBN²			-0.0002	***	-0.0002	***	-0.0002	***
t-stat			*-9.0210*		*-8.9011*		*-8.9014*	
varmrkt					798.4440	***	760.1080	***
t-stat					*60.7061*		*47.0891*	
varsect							41.2439	***
t-stat							*4.0235*	
R²	0.1972		0.1990		0.2461		0.2475	

Results of Table 4-18 demonstrate the positive and significant impact of the volume of specific announcements on the conditional volatility of each stock. The significant negative sign associated to *NBN²* suggests the existence of a marginally decreasing effect, with a maximum situated around 64.5 news releases.

In line with Kalev et al. (2002), p. 1444, we introduce the trading volume in value ($) into the regression equation, "as this variable does, to a limited degree, contain private information not yet reflected in public news. It could also influence volatility in ways that are unrelated to the information arrival". Indeed, our variable NBN, because it is of public knowledge, does not take into account the private dimension of information. In order to avoid any endogeneity problem, we lag this volume variable by one period:

$$Prob_{i,t} = C_i + \beta_1 NBN_{i,t} + \beta_2 NBN_{i,t}^2 + \beta_3 Vol_{i,t-1} + \beta_4 varmrkt_t + \beta_5 varsect_{i,t} + \varepsilon_{i,t} \quad (4\text{-}19)$$

According to Table 4-19, the impact of trading volume on the probability of high variance regime turns out to be positive and significant, which is a well-known result of the literature. We refer the interested reader to Karpoff (1987) for a detailed survey of earlier empirical work, most of which shows evidence that higher volatility is accompanied by higher trading volume. More recent studies such as Jain and Joh (1988), Schwert (1989), Gallant et al. (1992), Richardson and Smith (1994), Foster and Viswanathan (1995), and Andersen (1996) also document a similar relation. Notice that this impact of the volume trading remains significant after the inclusion of our public information proxy (*NBN*).

Table 4-19: Panel fixed-effect regression of *Prob* on *NBN* and *Vol*

Results from the estimation of the fixed-effect panel regression (4-19) for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t*. Vol(-1) is the trading volume in value ($) for stock *i* in *t-1*.The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2		3		4	
Vol(-1)	1.33E-10	***	1.17E-10	***	1.14E-10	***	8.40E-11	***
t-stat	*10.8287*		*10.5669*		*10.3908*		*8.3637*	
NBN							0.0184	***
t-stat							*26.8625*	
NBN²							-0.0002	***
t-stat							*-9.0817*	
varmrkt			793.8910	***	757.2090	***	756.2640	***
t-stat			*60.1675*		*47.6380*		*47.6190*	
varsect					39.6547	***	39.5121	***
t-stat					*4.0476*		*4.0216*	
R²	0.1948		0.2413		0.2426		0.2492	

4.2.3.2.1.2 Introduction of a time effect

Up to now, we have not considered the possibility that the impact of the announcements frequency on the smooth probability could vary through time. We therefore propose to perform the following panel regression:

$$Prob_{i,t}$$
$$= C_i + C_t + \beta_1 NBN_{i,t} + \beta_2 NBN_{i,t}^2 + \beta_3 \cdot D_{1999} \cdot NBN_{i,t} + \beta_4 \cdot D_{2000} \cdot NBN_{i,t} + \beta_5 \cdot D_{2001} \cdot NBN_{i,t}$$
$$+ \beta_6 \cdot D_{2002} \cdot NBN_{i,t} + \beta_7 \cdot D_{2003} \cdot NBN_{i,t} + \beta_8 varmrkt_t + \beta_9 varsect_{i,t} + \varepsilon_{i,t} \quad (4\text{-}20)$$

C_t takes into account as before the possible evolution of the intercept through time, while the coefficients β_3 to β_7 control for the potential evolution through time of the coefficient of our main variable of interest.

Table 4-20: Panel fixed-effect regression of *Prob* on *NBN* with time effect

Results from the estimation of the fixed-effect panel regression (4-20) for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t*. Dummies Dy2 to Dy6 are yearly dummies, with y2 = 1999 and y6 = 2003. The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2	
NBN	0.0163	***	0.0206	***
t-stat	*9.3472*		*13.7261*	
NBN*Dy2	-0.0028		-0.0011	
t-stat	*-1.1416*		*-0.6274*	
NBN*Dy3	-0.0060	***	0.0002	
t-stat	*-3.0588*		*0.1116*	
NBN*Dy4	-0.0037	*	-0.0027	
t-stat	*-1.7396*		*-1.5471*	
NBN*Dy5	-0.0018		0.0006	
t-stat	*-0.9371*		*0.3746*	
NBN*Dy6	-0.0017		-0.0018	
t-stat	*-0.8288*		*-1.0857*	
NBN^2	-		-0.0002	***
t-stat	-		*-9.2207*	
varmrkt	774.4910	***	773.7530	***
t-stat	*48.0008*		*48.0333*	
varsect	41.2058	***	40.9447	***
t-stat	*4.0302*		*4.0240*	
R^2	0.2450		0.2466	

Results of Table 4-20 show that the link between this probability and our explicative variable does not seem to depend on time. It provides thus some evidence in favor of the stability of the relation between the news flow and the volatility.

4.2.3.2.1.3 Analysis by category

4.2.3.2.1.3.1 Categorization by topic

Like in the French application, we split our announcements into different categories according to the topic.

Table 4-21: Panel fixed-effect regression on various topics of news

Results from the estimation of the fixed-effect panel regression for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t* about a given topic. The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	Dependent Variable: Smooth Probability									
	C13		C17		C18		C181		C33	
News Category	Regulation/Government Policy		Funding/Capital		Ownership Changes		M&A		Contracts/Orders	
NBN	0.0592	***	0.0333	***	0.0309	***	0.0320	***	0.0226	***
t-stat	*13.1882*		*8.7407*		*16.9998*		*16.8694*		*6.7067*	
NBN²	-0.0037	***	-0.0014	*	-0.0005	***	-0.0005	***	-0.0008	
t-stat	*-8.3001*		*-1.6629*		*-5.1766*		*-5.2768*		*-1.5052*	
varmrkt	924.2530	***	923.2670	***	923.9000	***	924.1180	***	923.6340	***
t-stat	*56.5965*		*56.5991*		*56.6024*		*56.6022*		*56.5421*	
varsect	48.4219	***	48.3434	***	48.3117	***	48.3312	***	48.4323	***
t-stat	*4.2109*		*4.2176*		*4.2107*		*4.2101*		*4.2167*	
R²	0.0814		0.0811		0.0824		0.0824		0.0808	

	C151		C1514		C152		C1521	
News Category	Earnings		Earnings surprises		Earnings Projections		Analyst Comment/Recommendation	
NBN	0.0377	***	0.3224	***	0.0503	***	0.0553	***
t-stat	*24.8088*		*6.5913*		*25.0825*		*12.1319*	
NBN²	-0.0019	***	-0.0809	***	-0.0019	***	-0.0060	***
t-stat	*-14.2971*		*-3.4858*		*-9.2970*		*-4.8391*	
varmrkt	919.4520	***	922.7440	***	918.3020	***	921.0510	***
t-stat	*56.5606*		*56.4987*		*56.8263*		*56.6012*	
varsect	48.0572	***	48.4495	***	47.4693	***	48.2015	***
t-stat	*4.2245*		*4.2198*		*4.2077*		*4.2218*	
R²	0.0831		0.0808		0.0868		0.0813	

Results are very similar to the ones obtained on the French market. If we forget about the "Earning surprises" categories for a while, the *Regulation/Government Policy* news is still the kind of announcement with the highest impact on the stock conditional volatility, followed by the *Analyst Comment/Recommendation* category. Other categories present coefficients similar on the French and US market, with some only slight variations.

A particular emphasis must put on a newly introduced topic: "*Earnings surprises*". Because the number of news releases on this topic was too low on the French market, we could only use it on the US market. The coefficient is impressive: one earning surprise increases (on average) the probability of being in a high volatility regime by about 24 points (=coefficient of *NB* + coefficient of NB^2). It seems that this kind of news release increases significantly the uncertainty of investors about future prospects of the firm.

We give now the results for three subjects in particular: *Regulation / Government Policy* (*C13*), *Ownership Change* (*C18*) and *Contracts* (*C33*), and we subdivide these latter general categories into more accurate ones to investigate what kind of news brings more or less volatility to the stock return. Notice also that we now introduce temporal dummies to take into account of a possible time pattern of the conditional volatility.

Table 4-22: Panel fixed-effect regression of *Prob* on category *Reg/Gov Policy*

Results from the estimation of the fixed-effect panel regression for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t* about "*Regulation / Government Policy*". The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	Dependent variable: Smooth Probability			
	C13		C131	
News Category	Regulation/Government Policy		Regulatory Bodies	
NBN	0.0275	***	0.1498	***
t-stat	*6.2578*		*4.2512*	
NBN^2	-0.0017	**	-0.0360	**
t-stat	*-1.9841*		*-2.1003*	
varmrkt	763.0210	***	762.7590	***
t-stat	*16.2586*		*46.9135*	
varsect	42.0959	***	42.1055	***
t-stat	*4.0498*		*4.0499*	
R^2	0.2395		0.2394	

	Dependent variable: Smooth Probability					
	C132		C133		C134	
News Category	Standards/Standardization		Patents		Trademarks/Copyrights	
NBN	0.0181		0.0248	**	0.0172	
t-stat	*0.6405*		*2.2178*		*0.7887*	
NBN^2	-		-		-	
t-stat	-		-		-	
varmrkt	762.7480	***	762.7510	***	762.7370	***
t-stat	*46.9070*		*46.9074*		*46.9056*	
varsect	42.1139	***	42.1122	***	42.1149	***
t-stat	*4.0499*		*4.0499*		*4.0499*	
R^2	0.2394		0.2394		0.2394	

Table 4-23: Panel fixed-effect regression of *Prob* on category *Ownership Change*

Results from the estimation of the fixed-effect panel regression for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t* about "*Ownership Change*". The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	Dependent variable: Smooth Probability									
	C18		C181		C182		C183		C184	
News Category	Ownership Changes		M&A		Divestitures/Asset Sales		Privatizations		Joint Ventures	
NBN	0.0164	***	0.0174	***	0.0057		-0.0622	***	0.0093	
t-stat	*10.1143*		*10.3427*		*0.7370*		*-4.9282*		*1.5233*	
NBN²	-0.0001	*	-0.0002	**	-		0.0034	***	-	
t-stat	*-1.7567*		*-2.1550*		-		*3.5650*		-	
varmrkt	763.9160	***	763.9810	***	762.7540	***	762.7050	***	762.7050	***
t-stat	*46.9785*		*46.9787*		*46.9074*		*46.9122*		*46.9062*	
varsect	42.0733	***	42.0812	***	42.1135	***	42.1021	***	42.1080	***
t-stat	*4.0445*		*4.0442*		*4.0500*		*4.0497*		*4.0498*	
R^2	0.2399		0.2399		0.2394		0.2394		0.2394	

The more detailed subdivision that we apply here delivers a very interesting fact: the coefficient of the category *C183 Privatization* is significantly negative (-0.0622). If we compare it to the coefficient of *C131 Regulatory Bodies*, which is in contrast the highest one, with 0.1498, we have two extreme types of topics and of effects. While privatization news seems to calm down the investors and to decrease the uncertainty around the stock value, news from the regulator affects the volatility in a completely different way and tends to increase the uncertainty around the prospects of the company. The fact that investors resent the intervention of the State as more unsafe than the intervention of private investors or companies is maybe not so unexpected, but tells a lot about the current prejudices and judgments of the market…

Table 4-24: Panel fixed-effect regression of *Prob* on category *Contracts*

Results from the estimation of the fixed-effect panel regression for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBN* is the number of news releases specific to stock *i* at time *t* about "*Contracts*". The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	Dependent variable: Smooth Probability					
	C33		C331		C332	
News Category	Contracts/Orders		Defense Contracts		Government Contracts	
NBN	0.0039		0.0185	***	0.0181	*
t-stat	*1.4908*		*2.5592*		*1.8881*	
NBN²	-		-		-	
t-stat	-		-		-	
varmrkt	762.7610	***	762.7610	***	762.7350	***
t-stat	*46.9067*		*46.9070*		*46.9066*	
varsect	42.1116	***	42.1139	***	42.1137	***
t-stat	*4.0495*		*4.0500*		*4.0499*	
R²	0.2394		*0.2394*		*0.2394*	

	Dependent variable: Smooth Probability					
	C333		C334		C335	
News Category	Non-government Contracts		Licensing Agreements		Franchises	
NBN	0.0102	**	0.0245	*	0.0135	
t-stat	*2.0499*		*1.6732*		*0.2338*	
NBN²	-		-		-	
t-stat	-		-		-	
varmrkt	762.7140	***	762.7340	***	762.7370	***
t-stat	*46.9093*		*46.9067*		*46.9061*	
varsect	42.1047	***	42.1112	***	42.1142	***
t-stat	*4.0498*		*4.0499*		*4.0499*	
R²	0.2394		*0.2394*		*0.2394*	

4.2.3.2.1.3.2 Categorization by timing

Table 4-25: Panel fixed-effect regression of *Prob* on *NBN* split by timing

Results from the estimation of the fixed-effect panel regression for the 458 stocks of the sample. The dependent variable is the daily smooth probability to be in the high variance regime and the explicative variable *NBNTR* (resp. *NBNNOTR*) is the number of news releases specific to stock *i* at time *t* occurring when the market is open (resp. closed). The two control variables are the conditional variance of the market return *varmrkt* and the conditional variance of the sectorial index of stock *i varsect*. The period of study covers 1418 trading days. The estimation is done by GMM and t-stats are robust to heteroscedasticity and autocorrelation of order 20 (Newey-West).

	1		2	
NBNTR	0.0224	***	-	
t-stat	*31.6948*		-	
NBNNOTR	-		0.0008	***
t-stat	-		*7.9596*	
NBNTR²	0.0080	***	-	
t-stat	*13.8077*		-	
NBNNOTR²	-		*0.0000*	
t-stat	-		*-1.1865*	
varmrkt	774.2720	***	775.2980	***
t-stat	*48.0220*		*47.8471*	
varsect	41.0912	***	41.6894	***
t-stat	*4.0273*		*4.0464*	
R²	0.2463		0.2399	

As in the French case, we can make a distinction between the announcements arriving when the market is open or closed. Results are qualitatively similar to the ones obtained on the French market. An announcement occurring during a trading day seems to bring more uncertainty than another arriving during the WE or overnight for example.

4.2.3.2.2 Cross-sectional data analysis with the transition probability

A nice characteristic of the MS-var estimation is to offer us the possibility to obtain the transition probabilities from one state (here, one variance regime) to another. Notice that those transition probabilities are not available on a daily basis as the previous smooth probabilities, but only in the cross-sectional dimension. It follows that we have got two transition probabilities (from the high regime to the low one, and vice-versa) for each firm. As a next step in our analysis, we can try to explain these probabilities by some characteristics specific to the firm. Our first attempt is to investigate the link between those transition probabilities and the level of communication prevailing around a given firm. We choose as a first proxy of the latter the log of the daily average number of announcements $\ln(\overline{NBN_i})$ (which is supposed to measure the **intensity** of the flow of information for a given stock):

$$ProbTrans_i = C + \beta \ln(\overline{NBN_i}) + \delta \ln(\overline{MV_i}) + \varepsilon_i \quad (4\text{-}21)$$

Where $\ln(\overline{MV_i})$ is the average market value of the firm on the sample period. We use it as a control variable, because the number of news releases is logically higher for large firms than for smaller ones.

As a second proxy, we choose *NBDays-NBN*: the percentage of trading days where at least one announcement was observed (which is supposed to measure the **regularity** of the information flow).

In order to estimate the persistence of the communication of each company, we specify and perform an AR(1) individually for each firm and regress the number of news releases in *t* on the number of news releases on the previous day *t-1*:

$$NBN_t = C + \ln(Rh\hat{o}) \cdot NBN_{t-1} + \varepsilon_t \quad (4\text{-}22)$$

ln*(Rhô),* our third proxy, measures the **persistence** of the communication: if ln(*Rhô*) is high, it implies that a day with a high number of news releases is likely to be followed by another day with a high number of announcements as well.

Table 4-26: Cross-sectional regression of the transition probabilities

Results from the estimation of the cross-sectional regression (4-21) for the 458 stocks of the sample. The dependent variable is the transition probability from one variance regime to another. Explicative variables measure some characteristics (intensity, regularity, persistence values) of the information flow of the firm. The estimation is done by GMM and t-stats are robust to heteroscedasticity (White).

Transition probability from high to low variance regime								
	1		2		3		4	
C	0.2068	***	0.1237	***	0.1064	***	0.2115	***
t-stat	*6.4728*		*6.8726*		*8.1611*		*6.6138*	
ln(Mean-NBN)	0.0130	***					0.0185	***
t-stat	*3.6807*						*3.9880*	
NBDays-NBN			0.0236	*			-0.0290	*
t-stat			*1.9072*				*-1.9433*	
ln(Rhô)					0.0021		-0.0013	
t-stat					*1.2455*		*-0.5030*	
ln(MV)	-0.0099	***	-0.0056	***	-0.0038	***	-0.0094	***
t-stat	*-5.2787*		*-4.0820*		*-5.4472*		*-4.9977*	
R^2	0.1088		0.0763		0.0714		0.1144	

Transition probability from low to high variance regime								
	1		2		3		4	
C	0.8137	***	0.4223	***	0.4125	***	0.8446	***
t-stat	*6.5351*		*6.1252*		*7.2495*		*6.7699*	
ln(Mean-NBN)	0.0581	***					0.0796	***
t-stat	*4.1339*						*3.9849*	
NBDays-NBN			0.0879	*			-0.1412	*
t-stat			*1.7209*				*-1.8817*	
ln(Rhô)					0.0180	**	0.0034	
t-stat					*2.3049*		*0.4969*	
ln(MV)	-0.0413	***	-0.0208	***	-0.0164	***	-0.0390	***
t-stat	*-5.7004*		*-3.8703*		*-5.4290*		*-5.4435*	
R^2	0.0771		0.0462		0.0502		0.0840	

Table 4-26 presents results of univariate and multivariate regressions. Notice that the coefficient of the log proxies must be interpreted as elasticities, or semi-elasticities. For example, we deduce from the coefficients of $\ln(\overline{NBN_i})$ that if a stock has a daily average number of announcements NBN_i 10% higher than another stock, its probability of transition from the high to the low (resp. from the low to the high) variance regime is about 13 (resp. 58) basis points higher. Our proxies of communication characteristics

seem to have a higher impact on the probability of transition from the low variance regime to the high variance regime, and thus a stronger effect on the probability to present a more volatile behaviour of stock returns. This seems consistent with our previous results: if the probability of transition L-H increases more than the probability of transition H-L when there are a lot of announcements, it is logical that the probability of being in the high variance regime is higher as well. The higher and the more regular the communication intensity of a firm, and the faster the volatility of its stock will arrive in the high regime. However, once the volatility is in the high regime, a firm communicating a lot can expect to come back quicker to the low volatility regime than a firm which does not.

Our interpretation must nevertheless remain cautious. We cannot exclude the presence of an endogeneity problem, since the sense of the causation is not straightforward: does a higher communication trigger a change of regime, or, the other way around, is it a change of regime that triggers a higher communication about the company? The current results leave the issue open, and we consider the latter as a potentially nice track for future research.

The interested reader will find descriptive statistics about the main variables used in this section in Appendix B.

4.2.4 Conclusion

This US application has confirmed on a wider sample the previous evidence brought by the French application, and delivered additional insights about the link existing between information flow and conditional volatility.

Similarly to the French case, we use daily prices and Reuters news releases in this US application, but the scope has increased to 458 stocks (instead of 40 for the French study) and the covered period ranges from May 13, 1998 to December 31, 2003 (instead of January 1, 1999 to December 31, 2003 for the French study). We first begin with a partial replication of the Kalev and al. (2004)'s GARCH model and we find, like them, that public information explains a large part of the daily volatility persistence. We apply a Markov switching regression (MSR) model to our panel data in order to determine the daily probability of the volatility regime to be in a high level state or a low level state.

This panel analysis confirms the existence of a positive and highly significant relation between information flow and conditional volatility, even when controlling for market volume (a proxy of private information, which turns out to have a significant impact on conditional volatility as well) and conditional sector and market variances. We interpret it as strong support in favor of the Mixture of Distribution Hypothesis (MDH).

Moreover, further analysis shows (like in the French case) that the impact of news intensity on stock volatility is positive but marginally decreasing. After we split the announcements into different categories according to their respective subject, we find that strong evidence that *Earnings Surprises* are typically the kind of news that brings the highest uncertainty on the market. After this particular category, we also find that news concerning *Regulation and Government Policy* has the greatest impact on the stock volatility, as it was the case in the French application. Very interestingly, a more precise categorization allows us to observe that news concerning *Privatization* leads to a significant decrease of the conditional volatility. This could indicate that information issued by the public authorities tends to increase the uncertainty among investors, while news announcing private shareholders instead of public ones rather leads to more confident investors and more stable distribution of stock returns.

A subdivision of news releases according to their timing give results similar to the French application: an announcement occurring when the market is active has a stronger impact on the conditional volatility than if it occurs overnight or during the WE.

The US study offered the possibility to conduct some tests on the transition probabilities that were no conducted in the previous French study. All results lead to identical conclusions. Companies with superior intensity, regularity and persistence of communication also present higher transition probabilities from one state to another. Furthermore, the increase is larger for the probability of transition from the low volatility regime to the high one than the other probability (from high to low regime). In summary, a more active long-term communication policy increases the probability of transition from one state to another: in a regime of low volatility, it is more likely to switch soon to a higher volatility state, and the other way around. However, an issue of endogeneity may arise here. It will be the purpose of future research to deal with it.

Conclusion

The objective of this dissertation and of the whole research process that finally lead to it was to investigate the multiple relations and mechanisms governing the triad "public information – conditional volatility – asset price". Traders, shareholders, companies, market authorities, are directly concerned by the impact that a change in the communication policy of a company or a specific piece of news may have on the asset prices. History is full of examples of disastrous profit warnings or awkward disclosures that provoked unexpected and extremely negative market reactions. The better understanding of the mechanisms at stake should provide guidance for the different actors evolving on the market to better interpret some difficult situations they could meet.

The first step of our study was to set up a simple theoretical model of "asset pricing" in a universe of constrained, or not optimally diversified, investors. It builds on the classical CAPM model and generalizes it to the case of not optimally diversified investors. The main result of such a model is to make appear an additional idiosyncratic variance term in the final asset pricing equation, besides the usual systematic risk (the covariance with the market) of the CAPM. This additional variance term is in interaction with our concept of deviation, positive if the investors overweight the stock in the portfolio, negative if they underweight it. We draw several interesting conclusions, and especially the following: the higher the deviation in a stock, the higher its price. We also show that the relation between the variance and the price can be potentially significantly affected by the magnitude of the deviation. So much affected, in fact, that under certain conditions, the price can become increasing in the variance when the deviation reaches a certain positive threshold. However, this phenomenon is likely to be purely theoretical, and we do not have practical evidence of it on the current financial markets.

The second step was to test empirically our "deviation effect". Does a higher deviation actually lead to a higher valorization of the firm? A natural experiment for it is the issue of the index effect. A vast literature has examined the effect of additions to and deletions from heavily traded index, typically the S&P 500. Many authors have provided evidence that the inclusion of a firm in the S&P 500 has a positive effect on its value. There are many possible explanations for such an effect, and we ourselves bring our own contribution and perspective on the topic. Our deviation concept may indeed

help to explain, at least partially, this famous index effect. We therefore set up econometric specifications able to disentangle the different (and sometimes opposite) effects announced by the competing hypotheses. Our results allow us to then confirm or infirm those various explanations of the index. We tend to reject Downward-Sloping Demand Curve Hypothesis, the Certification of an Index Member Hypothesis and the classical version of the Investor Recognition Hypothesis, while we confirm (at least partly) the Price Pressure Hypothesis and a short-term version of the Investor Recognition Hypothesis. The Information Content Hypothesis (in its form: inclusion causes better future performance) and the Liquidity Cost Hypothesis are not directly tested in our specification.

Until now, the information component of our triad has not been exploited yet. This is the goal of the second part of this document to remedy it. We begin with a theoretical model which directly continues the model developed in the first step. We enrich it with the concept of signals, which are continuously released and reveal public information to the market. Our point is to investigate how the quality and quantity of micro and macro public signals impact the main determinants of our pricing equation: the conditional expectation, variance and covariance. The main (and rather intuitive) conclusion is that more precise and more frequent information decreases both the conditional covariance and the variance of a stock price. The resulting effect on the price is ambiguous: it depends (among other things) on the deviation. In a classical CAPM framework, when all investors are perfectly diversified, a better signaling decreases the systematic risk and consequently increases the stock price as well. However, when a positive deviation appears and gets larger and larger, the decrease of conditional variance induced by a better signaling does not lead to an increase of price as large as before. The higher the deviation, the lower the sensibility of price to the conditional variance. That is why firms benefiting from a high deviation have less incentive to communicate than others, because for a same increase of their communication quality, the price of the firm with a low deviation will increase more than the price of the firm with a high deviation, ceteris paribus.

The last section of this dissertation is purely empirical and aims at investigating the link between public information and conditional volatility. It uses an original sample of several tens of thousands of news releases from Reuters and Dow Jones. The empirical estimation is conducted both on the French and the US markets. Thanks to various

econometric specifications like GARCH models and Markov Switching Regressions, we show that a larger daily number of news releases increases the probability to be in the high probability regime. More information is on average synonymous of more volatility, at least in the short run. But information is not uniform. We show that the impact of information is strongly dependent on the topic and the timing of the release of this information. Interestingly, news carrying information about government and regulation policy increases the most the volatility, while news concerning privatization is the only kind of announcements to significantly decrease the volatility. Should we conclude that investors consider private interventions as less uncertain and worrying than public ones? Results seem to provide some evidence in favor of this conclusion. Another interesting fact is the similarity of our results of both French and US markets. The link between information and volatility seems to be quite stable across those two markets.

Tracks for future research are numerous. Firstly, our current theoretical model is discrete. A possibility could be to make it continuous. Secondly, it models the relationship between public information (this kind of information related to the semi-strong-form efficiency hypothesis) and asset prices. If we wanted to enrich it, we could also introduce private information (this kind of information related to the strong-form efficiency hypothesis) information coming from historical prices (this kind of information related to the weak-form efficiency hypothesis). Thirdly, our model allows to make some predictions about the relationships between the micro and macro public information, the conditional expectation, variance and covariance, and the price. Our last empirical study focused "only" on the relation between "micro public information" and "conditional volatility". We could extend micro information to macro information, and conditional volatility to conditional covariance for example. Those suggestions are all promising venues of future research.

References

Amihud, Y. and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics* 17, 223-249.

Andersen, T. G., 1996, Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* 51, 169-204.

Anderson, T. W., 2003, An Introduction to Multivariate Statistical Analysis, 3rd ed., Hoboken, NJ: John Wiley & Sons.

Arnott, R. D. and S. J. Vincent, 1986, S&P additions and deletions: a market anomaly, *Journal of Portfolio Management*, Fall, 26-33.

Ausubel, L.M., 1990, Insider trading in a rational expectations economy, *American Economic Review* 80(5), 1022-1041.

Bai, Y., E. C. Chang and J. Wang, Asset Prices under Short-Sales Constraints, Working paper under review (revised and re-submission), Journal of Finance.

Baiman, S. and R. Verrecchia, 1996, The relation among capital markets, financial disclosure, production efficiency, and insider trading, *Journal of Accounting Research* 34, 1-22.

Banz, R. W., 1981, The Relationship between Return and Market Value of Common Stocks, *Journal of Financial Economics* 9, 3-18.

Barberis, N., A. Shleifer and J. Wurgler, 2005, Comovement, *Journal of Financial Economics*, 75(2), 283-317.

Barry, C. and S. Brown, 1985, Differential information ad security market equilibrium, *Journal of Financial and Quantitative Analysis* 20, 407-422.

Beaver, William H., 1968, The information content of annual earnings announcements, *Empirical Research in Accounting* 6, 67-92.

Belsley, D. A., 1991, Conditioning Diagnostics: collinearity and weak data in regression, New York : Wiley.

Beneish, M. and J. Gardner, 1995, Information Costs and Liquidity Effects from Changes in the Dow Jones Industrial Average List, *The Journal of Financial and Quantitative Analysis*, 30(1), 135-157.

Beneish, M. D. and R. E. Whaley, 1996, An Anatomy of the "S&P Game": The Effects of Changing the Rules, *Journal of Finance* 51(5), 1909-1930.

Beneish, M. D. and R. E. Whaley, 1997, A scorecard from the S&P game, *The Journal of Portfolio Management* 23(2), 16-23.

Berry, T. D., and K. M. Howe, 1993, Public information arrival, *Journal of Finance* 49, 1331–1346.

Bessembinder, H., 1992, "Systematic Risk, Hedging Pressure, and Risk Premiums in Future Markets", *Review of Financial Studies* 5, 637-67.

Blitzer, D. and S. Dash, 2004, Index Effect Revisited: What Happened to the Post-Addition Price Pop?, Standard and Poor's Publication Series, September 2004.

Blouin, J., J. Raedy and D. Shackelford, 2003, Capital Gains Taxes and Equity Trading: Empirical Evidence, *Journal of Accounting Research* 41(4) , 611–651.

Blume, M. and R. Edelen, 2001, On S&P 500 index replication strategies, Working paper, University of Pennsylvania.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 31, 307-327.

Braun, M. and B. Larrain, 2008, Do IPO's Affect the Prices of Other Stocks? Evidence from Emerging Markets, forthcoming *Review of Financial Studies*.

Brennan, M., 1975, The Optimal Number of Securities in a Risky Asset Portfolio when There are Fixed Costs of Transacting: Theory and Some Empirical Results, *Journal of Financial and Quantitative Analysis* 10, 483-496.

Brooks, C., K. Kappou and C. Ward, 2004, Gambling on the S&P 500's gold seal: new evidence on the index effect, ISMA Center Discussion Papers in Finance, University of Reading.

Campa, J. M. and S. Kedia, 2002, Explaining the Diversification Discount, *Journal of Finance* 57(4), 1731-1762.

Campbell, J. Y., M. Lettau, B. Malkiel and Y. Xu, 2001, Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk, *Journal of Finance* 56, 1-43.

Chae, J., 2005, Trading volume, information asymmetry, and timing information, *Journal of Finance* 60(1), 413-442.

Chakrabarti R., W. Huang, N. Jayaraman and J. Lee, 2002, "The index effect" on stock prices and trading volumes: international evidence, 2002-2003 Working Paper Series, Georgia Tech Center for International Education and Research.

Chamberlain, G. and M. Rothschild, 1983, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, 51(5), 1281-1304.

Chan, L. K. C., N. Jegadeesh and J. Lakonishok, 1995, Evaluating the performance of value versus glamour stocks: The impact of selection bias, *Journal of Financial Economics* 38, 269–296.

Chauvet, M., and J. D. Hamilton, 2005, Dating business cycle turning points, National bureau of economic research, Working paper 11422.

Chen, H., G. Noronha and V. Singal, 2004, The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation, *Journal of Finance* 59(4), 1901-1929.

Chordia, T., 2001, Liquidity and returns: The impact of inclusion into the S&P 500 index, Working paper, Emory University.

Chordia, T., R. Roll, and A. Subrahmanyam, 2005, Evidence on the Speed of Convergence to Market Efficiency, *Journal of Financial Economics* 76(2), 271-292.

Clark, P. K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135-155.

Coles, J., U. Loewenstein and J. Suay, 1995, On equilibrium pricing under parameter uncertainty, *The Journal of Financial and Quantitative Analysis* 30, 347-374.

Cousin, J.-G. and T. de Launois, 2006, "News Intensity and Conditional Volatility on the French Stock Market", *Finance*, 27, 7-60.

Cusick, P., 2002, Price effects of addition or deletion from the Standard and Poor's 500 index: evidence of increasing market efficiency, *Financial Markets, Institutions and Instruments* 11, 349-383.

Damodaran, A. and C. H. Liu, 1993, Insider trading as a signal of private information, *Review of Financial Studies* 6(1), 79-119.

Davies, R.B., 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, 247-254.

Degroot, M, 1970, Optimal Statistical Decisions, Wiley and Sons, Hoboken NJ.

Deininger, C., C. Kaserer and S. Roos, 2000, Stock price effects associated with index replacements in Germany, Working Paper N°7, Free University of Bozen-Bolzano.

De Long, J. B., A. Shleifer, L. Summers and R. Waldmann, 1990, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance* 45, 379-395.

Denis, D., J. McConnell, A. Ovtchinnikov and Y. Yu, 2003, S&P 500 index additions and earnings expectations, *Journal of Finance* 58(5), 1821–1840.

Dhillon, U. and H. Johnson, 1991, Changes in the Standard and Poor's 500 list, *Journal of Business* 64, 75–86.

Diamond, D. W. and R. E. Verrecchia, 1987, Constraints on Short-Selling and Asset Price Adjustment to Private Information, *Journal of Financial Economics* 18, 277-311.

Diamond, D. and R. Verrecchia, 1991, Disclosure, liquidity and the cost of capital, *Journal of Finance* 46, 1325-1359.

Douglas, G. W., 1969, "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency", *Yale Economic Essays*, Vol. 9, pp. 3-45.

Easley, D. and M. O'Hara, 2004, Information and the Cost of Capital, *Journal of Finance* 59(4), 1553-1583.

Edelen, R. and J. Warner, 2001, Aggregate price effects of institutional trading: a study of mutual fund flow and market returns, *Journal of Financial Economics* 59(2).

Ederington, L.H. and J. H. Lee, 1993, How markets process information: News releases and volatility, *Journal of Finance* 48, 1161–1191.

Edmister, R., S. Graham and W. Pirie, 1994, Excess returns of index replacement stocks: evidence of liquidity and substitutability, *Journal of Financial Research* 17(3), 333-346.

Edmister, R., S. Graham and W. Pirie, 1996, Trading cost expectations: evidence from S&P 500 index replacement stock announcements, *Journal of Economics and Finance* 20(2), 75-85.

Elayan F., W. Li and J. Pinfold, 2000, Price effects of changes to the composition of New Zealand share indices, Working Paper Series, N° 00.24, Massey University, Auckland.

Engle, R., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, *Econometrica* 50, 987-1008.

Engle, R. and V. Ng, 1993, Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 48, 1749-1777.

Erwin, G. R. and J. M. Miller, 1998, The liquidity effects associated with addition of a stock to the S&P 500 index: evidence from bid/ask spreads, *The Financial Review* 33(1), 131-146.

Falkenstein, E. G., 1996, “Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings”, *Journal of Finance* 51(1), 111-135

Fama, E. F., 1970, Efficient Capital Markets : a review of theory and empirical work, *Journal of Finance* 25, 383-417.

Fama, E. F., 1976, Foundations of Finance, Basic Books, New York.

Fama, 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575-1617.

Fama, E., Fisher, L., Jensen, M., Roll, R., 1969, The adjustment of stock prices to new information, *International Economic Review* 10, 1-21.

Fama, E. and J. MacBeth 1973, “Risk, Return, and Equilibrium: Empirical Tests”, *Journal of Political Economy*, Vol. 81, pp. 607-636.

Foster, F. D. and S. Viswanathan, 1995, Can speculative trading explain the volume-volatility relation?, *Journal of Business and Economic Statistics*, 13, 379–396.

Gallant, A. R., P. E. Rossi and G. Tauchen, 1992, Stock prices and volume, *Review of Financial Studies*, 5, 199–242.

Goetzmann, W. N. and M. Garry, 1986, Does delisting from the S&P 500 affect stock price? *Financial Analyst Journal* 42, 64–69.

Goetzmann, W.N. and M. Massa, 2003, Index Funds and Stock Market Growth, *The Journal of Business*, 76(1), 1-28.

Goldsmith, D., 1976, Transaction Costs and the Theory of Portfolio Selection, *Journal of Finance* 31, 1127-1139.

Goyal, A., and P. Santa-Clara, 2003, Idiosyncratic Risk Matters!, *Journal of Finance* 58, 975-1007.

Graham, S. and W. Pirie, 1994, Index fund rebalancing and market efficiency, *Journal of Economics and Finance* 18(2), 219-229.

Green, Richard and K. Rydqvist, 1997, “The Valuation of Nonsystematic Risks and the Pricing of Swedish Lottery Bonds”, *Review of Financial Studies* (10), 447-480.

Greene, W. H., 1997, Econometric Analysis, Third edition, Prentice Hall.

Greene, W. H., 2002, Econometric Analysis, Fourth edition, Prentice Hall.

Greenwood, R., 2005, Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* 75, 607-649.

Grossman, S. J. and J. E. Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, 393-408.

Hamilton, J. D., 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and The Business Cycle, *Econometrica* 57, 357-84.

Hamilton, J. D., 1994, Time series analysis, Princeton University Press.

Harris, L. and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–830.

Harrison, J. M. and D. M. Kreps, 1978, Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations, *Quarterly Journal of Economics* 92, 323-336.

Hasbrouck, J., 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

Hegde, S. and J. McDermott, 2003, The liquidity effects of revisions to the S&P 500 index: An empirical analysis, *Journal of Financial Markets* 6, 413–459.

Hellwig, M. F., 1980, On the aggregation of information in competitive markets, *Journal of Economic Theory*, 22(3), 477–498.

Hrazdil, K. and T. Scott, 2007, S&P 500 index revisited: Do index inclusion announcements convey information about firms' future performance?, SSRN working paper series.

Jacques, W., 1988, The S&P 500 membership anomaly, or would you join this club?, *Financiel Analyst Journal* 44(6), 73-75.

Jain, P., 1987, The effect on stock price of inclusion in or exclusion from the S&P 500, *Financial Analysts Journal* 43, 58-65.

Jain, P. and G. Joh, 1988, The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, 23, 269–283.

Kalev, P. S., W.-M. Liu, P. K. Pham and E. Jarnecic, 2004, Public Information Arrival and Volatility of Intraday Stock Returns, *Journal of Banking and Finance* 28(6), 1447-1467.

Karpoff, J.M., 1987, The relation between price changes and trading volume: A survey, *Journal of Financial and Quantitative Analysis*, 22, 109–126.

Kaul, A., V. Mehrotra, and R. Morck, 2000, Demand curves for stocks do slope down: New evidence from an index weights adjustment, *Journal of Finance* 55, 893–912.

King, M., E. Sentana and S. Wadhwani, 1994, "Volatility and Links Between National Stock Markets", *Econometrica* 62(4), 901-933.

Klasa, S., 2007, Why do controlling families of public firms sell their remaining ownership their remaining ownership stake?, *Journal of Financial and Quantitative Analysis* 42(2), 339-368.

Krolzig, H.M., 1997, Markov-Switching Vector Autoregressions, University of Oxford, UK Lecture Notes in Economics and Mathematical Systems, Vol. 454, Springer, XIV.

Lakonishok, J., A. Shleifer and Robert Vishny, 1991, Do institutional investors destabilize stock prices? Evidence on herding and feedback trading, NBER working paper 3846.

Lakonishok, J., A. Shleifer and R. Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32, 23-43.

Lambert, R. A., C. Leuz and R. E. Verrecchia, 2007, Accounting Information, Disclosure, and the Cost of Capital, *Journal of Accounting Research* 45(2), 385-420.

Lamoureux, C. and J. Wansley, 1987, Market effects of changes in the Standard and Poor's 500 index, *Financial Review* 22(1), 53-69.

Lardic, S. and V Mignon, Analyse intraquotidienne de l'impact des news sur le marché boursier français, *Economie Appliquée* 56(2), 205-237.

Leape, J., 1987, Taxes and Transaction Costs in Asset Market Equilibrium, *Journal of Public Economics* 33, 1-20.

Lehmann, B. N., 1990, Residual Risk Revisited, *Journal of Econometrics*, 45, 71-97.

Lespagnol, C. and J. Teïletche, 2005, La dynamique de la volatilité à très haute fréquence des taux longs euro, *Finance* 26(2), 87-131.

Levy, H., 1978, Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in a Portfolio, *American Economic Review* 68, 643-658.

Li, L. and B. M. Fleisher, 2004, Heterogeneous expectations and stock prices in segmented Markets: application to Chinese firms, *Quarterly Review of Economics and Finance* 44, 521-538

Li, K. and N. R. Prabhala, Self-Selection Models in Corporate Finance, Chapter 2 in: Handbook of Corporate Finance: Empirical Corporate Finance, Handbooks in Finance Series, Elsevier/North Holland, 2007.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47(1), 13-37.

Lynch, A. and R. Mendenhall, 1997, New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index, *Journal of Business* 70(3), 351-383.

Malkiel, B. and A. Radisich, 2001, The growth of index funds and the pricing of equity securities, *Journal of Portfolio Management* 27(2), 9-21.

Malkiel, B. G. and Y. Xu, 2004, Idiosyncratic Risk and Security Returns, SSRN working paper series.

Massa, M. and W. Goetzman, 2003, Index Funds and Stock Market Growth, *The Journal of Business* 76(1), 1-28.

Markowitz, H. M., 1999, The early history of portfolio theory: 1600-1960, *Financial Analysts Journal* 55(4).

Masse I., R. Hanrahan, J. Kushner and F. Martinello, 2000, The Effect of Additions to or Deletions from the TSE 300 Index on Canadian Share Prices, *The Canadian Journal of Economics* 33(2).

Mayshar, J., 1979, Transaction Costs in a Model of Capital Market Equilibrium, *Journal of Political Economy* 87, 673-700.

Mayshar, J., 1981, Transaction Costs and the Pricing of Assets, *Journal of Finance*, 36, 583-597.

Merton, R. C., 1987, Presidential address: A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42(2), 483–510.

Mikkelson, H. and M. Partch, 1985, Stock price effects and costs of secondary distributions, *Journal of Financial Economics* 14, 165-194.

Miller, E., 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151-1168.

Mitchell, M. L. and J. H. Mulherin, 1994, The Impact of Public Information on the Stock Market, *Journal of Finance* 49, 923-950.

Morck, R. and F. Yang, 2002, The mysterious growing value of S&P 500 membership, NBER Working Paper, N° 8654.

Mossin, J., 1966, Equilibrium in a Capital Asset Market, *Econometrica* 34, 768-783.

Newey, W. K. and K. D. West, 1987, A Simple Positive Semi-definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Petajisto, A., 2006, The Index Premium and Its Implications for Index Funds, Yale School of Management, Working Paper, February 15, 2006.

Petersen, M. A., 2008, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies*, Forthcoming.

Piazzesi, M., M. Schneider and S. Tuzel, 2007, Housing, Consumption, and Asset Pricing, Journal of Financial Economics 83, 531-569.

Pruitt, S. W. and K. C. J. Wei, 1989, Institutional ownership and changes in the S&P 500, *Journal of Finance* 44, 509–514.

Richardson, M. and T. Smith, 1994, A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information, *Journal of Financial and Quantitative Analysis*, 29, 101–116.

Roll, R., 1984, Orange Juice and weather, *American Economic Review* 74(5), 861-880.

Roll, R., 1988, R-Squared, *Journal of Finance* 43, 541-566.

Ross, S. A., 1989, Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy, *Journal of Finance* 44, 1-17.

Scholes, M., 1972, The market for securities: substitution versus price pressure and the effects of information on share prices, *Journal of Business* 45(2), 179-211.

Schwarz, G., 1978, Estimating the Dimension of a Model, *The Annals of Statistics* 6(2), 461-464.

Schwert, G.W., 1989, Why does market volatility change over time?, *Journal of Finance*, 44, 1115–1153.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19 (3), 425-442

Shleifer, A., 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.

Subrahmanyam, A. and S. Titman, 2001, Feedback from Stock Prices to Cash Flows, *The Journal of Finance* 56(6), 2389–2413.

Treynor, J. L., 1961, Market Value, Time, and Risk, Unpublished manuscript.

Van Oppens, H., 2007, Insider trading, private information and market efficiency: An empirical analysis through complementary studies, Université Catholique de Louvain – Louvain School of Management, Ph.D. Dissertation.

Verrecchia, R., 2001, Essays on disclosure, *Journal of Accounting and Economics* 32, 97-180.

Vespro, C., 2006, Stock Price and Volume Effects Associated with Compositional Changes in European Stock Indices, *European Financial Management* 12(1).

Viard, A. D., 1995, The Asset Pricing Effects of Fixed Holding Costs: An Upper Bound, *Journal of Financial and Quantitative Analysis* 30(1), 43-59.

Villalonga, B. and R. Amit, 2006, How Do Family Ownership, Control, and Management Affect Firm Value?, *Journal of Financial Economics* 80, 385–417.

Warther, V., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209-235.

Wei, S. and C. Zhang, 2005, Idiosyncratic risk does not matter: A re-examination of the relationship between average returns and average volatilities, *Journal of Banking and Finance,* 29(3), 603-621.

Woolridge, R. and C. Ghosh, 1986, Institutional trading and security prices: the case of changes in the composition of the S&P 500 index, *Journal of Financial Research* 9(1), 13-24.

Wurgler, J. and K. Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks? *Journal of Business* 75(4), 583–608.

Appendix A

We prove that the formula for one signal can be extended to take into account Q_j and Q_M signals concerning the stock and the market.

We begin with the conditional expectation. We have to prove that:

$$E\left(\tilde{p}_{j,t+1} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right)$$
$$= \mu_j + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\left(z_j - \mu_j\right)\right) Var\left(\tilde{p}_{j,t+1} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right)$$
$$+ \left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\left(z_M - \mu_M\right)\right) Cov\left(\tilde{p}_{j,t+1}, \tilde{p}_{M,t+1} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}\right)$$

We first prove the result for two signals on stock *j* and two signals on the market. The above more general result follows by recursivity.

We have to prove that:

$$E\left(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{j,2}, z_M\right)$$
$$= \mu_j + \left(\sum_{q=1}^{2} v_{j,q}^{-1}\left(z_{j,q} - \mu_j\right)\right) Var\left(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{j,2}, z_M\right)$$
$$+ v_M^{-1}\left(z_M - \mu_M\right) Cov\left(\tilde{p}_{j,t+1}, \tilde{p}_{M,t+1} \mid z_{j,1}, z_{j,2}, z_M\right)$$

The development is the following:

$$\mathbf{m}_{1|2} = \mathbf{m}_1 + \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\left(\begin{pmatrix} z_{j,1} \\ z_{j,2} \\ z_M \end{pmatrix} - \mathbf{m}_2\right) = \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \begin{pmatrix} \sigma_j^2 & \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \begin{pmatrix} \sigma_j^2 + \nu_{j,1} & \sigma_j^2 & \sigma_{j,M} \\ \sigma_j^2 & \sigma_j^2 + \nu_{j,2} & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_{j,M} & \sigma_M^2 + \nu_M \end{pmatrix}^{-1} \begin{pmatrix} z_{j,1} - \mu_j \\ z_{j,2} - \mu_j \\ z_M - \mu_M \end{pmatrix}$$

$$= \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \frac{\begin{pmatrix} \sigma_j^2 & \sigma_j^2 & \sigma_{j,M} \\ \sigma_{j,M} & \sigma_{j,M} & \sigma_M^2 \end{pmatrix} \begin{pmatrix} \left(\sigma_j^2 + \nu_{j,2}\right)\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2 & \sigma_{j,M}^2 - \sigma_j^2\left(\sigma_M^2 + \nu_M\right) & -\sigma_{j,M}\nu_{j,2} \\ \sigma_{j,M}^2 - \sigma_j^2\left(\sigma_M^2 + \nu_M\right) & \left(\sigma_j^2 + \nu_{j,1}\right)\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2 & -\sigma_{j,M}\nu_{j,1} \\ -\sigma_{j,M}\nu_{j,2} & -\sigma_{j,M}\nu_{j,1} & \left(\sigma_j^2 + \nu_{j,1}\right)\left(\sigma_j^2 + \nu_{j,2}\right) - \sigma_j^4 \end{pmatrix} \begin{pmatrix} z_{j,1} - \mu_j \\ z_{j,2} - \mu_j \\ z_M - \mu_M \end{pmatrix}}{\sigma_j^2\left(\nu_{j,1} + \nu_{j,2}\right)\left(\sigma_M^2 + \nu_M\right) + \nu_{j,1}\nu_{j,2}\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2\left(\nu_{j,1} + \nu_{j,2}\right)}$$

$$= \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \frac{\begin{pmatrix} \nu_{j,2}\left[\sigma_j^2\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2\right] & \nu_{j,1}\left[\sigma_j^2\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2\right] & \sigma_{j,M}\nu_{j,1}\nu_{j,2} \\ \sigma_{j,M}\nu_{j,2}\nu_M & \sigma_{j,M}\nu_{j,1}\nu_M & \sigma_M^2\left[\sigma_j^2\left(\nu_{j,1} + \nu_{j,2}\right) + \nu_{j,1}\nu_{j,2}\right] - \sigma_{j,M}^2\left(\nu_{j,1} + \nu_{j,2}\right) \end{pmatrix} \begin{pmatrix} z_{j,1} - \mu_j \\ z_{j,2} - \mu_j \\ z_M - \mu_M \end{pmatrix}}{\sigma_j^2\left(\nu_{j,1} + \nu_{j,2}\right)\left(\sigma_M^2 + \nu_M\right) + \nu_{j,1}\nu_{j,2}\left(\sigma_M^2 + \nu_M\right) - \sigma_{j,M}^2\left(\nu_{j,1} + \nu_{j,2}\right)}$$

If we divide the numerator and the denominator of the second term by $\nu_{j,1}\nu_{j,2}\nu_M$, we get:

$$= \begin{pmatrix} \mu_j \\ \mu_M \end{pmatrix} + \frac{\begin{pmatrix} \nu_{j,1}^{-1}\left[\sigma_j^2\left(\sigma_M^2\nu_M^{-1} + 1\right) - \sigma_{j,M}^2\nu_M^{-1}\right] & \nu_{j,2}^{-1}\left[\sigma_j^2\left(\sigma_M^2\nu_M^{-1} + 1\right) - \sigma_{j,M}^2\nu_M^{-1}\right] & \sigma_{j,M}\nu_M^{-1} \\ \sigma_{j,M}\nu_{j,1}^{-1} & \sigma_{j,M}\nu_{j,2}^{-1} & \nu_M^{-1}\left\{\sigma_M^2\left[\sigma_j^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right) + 1\right] - \sigma_{j,M}^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)\right\} \end{pmatrix} \begin{pmatrix} z_{j,1} - \mu_j \\ z_{j,2} - \mu_j \\ z_M - \mu_M \end{pmatrix}}{\left(\sigma_j^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right) + 1\right)\left(\sigma_M^2\nu_M^{-1} + 1\right) - \sigma_{j,M}^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)\nu_M^{-1}}$$

And finally:

$$E\left(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{j,2}, z_M\right)$$

$$= \mu_j + \frac{\left(\sum_{q=1}^{2} v_{j,q}^{-1}\left(z_{j,q} - \mu_j\right)\right)\left(\sigma_j^2 + \left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)\nu_M^{-1}\right)}{\left(\sigma_j^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)+1\right)\left(\sigma_M^2\nu_M^{-1}+1\right) - \sigma_{j,M}^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)\nu_M^{-1}} + \frac{\nu_M^{-1}\left(z_M - \mu_M\right)\sigma_{j,M}}{\left(\sigma_j^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)+1\right)\left(\sigma_M^2\nu_M^{-1}+1\right) - \sigma_{j,M}^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)\nu_M^{-1}}$$

And by symmetry, we also show that:

$$E\left(\tilde{p}_{j,t+1} \mid z_j, z_{M,1}, z_{M,2}\right)$$

$$= \mu_j + \frac{\left(\nu_j^{-1}\left(z_j - \mu_j\right)\right)\left(\sigma_j^2 + \left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)\right)}{\left(\sigma_j^2\nu_j^{-1}+1\right)\left(\sigma_M^2\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)+1\right) - \sigma_{j,M}^2\nu_j^{-1}\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)} + \frac{\left(\sum_{q=1}^{2} v_{M,q}^{-1}\left(z_{M,q} - \mu_M\right)\right)\sigma_{j,M}}{\left(\sigma_j^2\nu_j^{-1}+1\right)\left(\sigma_M^2\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)+1\right) - \sigma_{j,M}^2\nu_j^{-1}\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)}$$

So that finally we can conclude that:

$$E\left(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{j,2}, z_{M,1}, z_{M,2}\right)$$

$$= \mu_j + \frac{\left(\sum_{q=1}^{2} v_{j,q}^{-1}\left(z_{j,q} - \mu_j\right)\right)\left(\sigma_j^2 + \left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)\nu_M^{-1}\right) + \left(\sum_{q=1}^{2} v_{M,q}^{-1}\left(z_{M,q} - \mu_M\right)\right)\sigma_{j,M}}{\left(\sigma_j^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)+1\right)\left(\sigma_M^2\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)+1\right) - \sigma_{j,M}^2\left(\nu_{j,1}^{-1} + \nu_{j,2}^{-1}\right)\left(\nu_{M,1}^{-1} + \nu_{M,2}^{-1}\right)}$$

$$= \mu_j + \left(\sum_{q=1}^{2} v_{j,q}^{-1}\left(z_{j,q} - \mu_j\right)\right) Var\left(\tilde{p}_{j,t+1} \mid \Im_t, \tilde{z}_{j,1}, \tilde{z}_{j,2}, \tilde{z}_{M,1}, \tilde{z}_{M,2}\right) + \left(\sum_{q=1}^{2} v_{M,q}^{-1}\left(z_{M,q} - \mu_M\right)\right) Cov\left(\tilde{p}_{j,t+1}, \tilde{p}_{M,t+1} \mid \Im_t, \tilde{z}_{j,1}, \tilde{z}_{j,2}, \tilde{z}_{M,1}, \tilde{z}_{M,2}\right)$$

By recursivity, the general formula with Q signals follows. This concludes the proof.

Now that we have proven the general form of the conditional expectation, we continue with the variance and we propose to prove that:

$$Var(\tilde{p}_{j,t+1} \mid z_{j,1},..., z_{j,q},..., z_{j,Q_j}, z_{M,1},..., z_{M,q},..., z_{M,Q_M}) = \frac{1}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\sigma_M^2 + \frac{1}{\sum_{q=1}^{Q_M} v_{M,q}^{-1}}}}}$$

We first consider the case of two signals on stock j and two signals on the market:

$$\begin{aligned}
&Var(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{j,2}, z_{M,1}, z_{M,2}) \\
&= \frac{1}{v_{j,2}^{-1} + \frac{1}{Var(\tilde{p}_{j,t+1} \mid z_{j,1}, z_{M,1}) - \frac{Cov^2(\tilde{p}_{j,t+1}, \tilde{p}_{M,t+1} \mid z_{j,1}, z_{M,1})}{Var(\tilde{p}_{M,t+1} \mid z_{j,1}, z_{M,1}) + \frac{1}{v_{M,2}^{-1}}}}} \\
&= \frac{1}{v_{j,2}^{-1} + \frac{1}{\frac{\sigma_j^2 \sigma_M^2 v_{M,1}^{-1} + \sigma_j^2 - \sigma_{j,M}^2 v_{M,1}^{-1}}{v_{j,1}^{-1}\left(\sigma_j^2 \sigma_M^2 v_{M,1}^{-1} + \sigma_j^2 - \sigma_{j,M}^2 v_{M,1}^{-1}\right) + \sigma_M^2 v_{M,1}^{-1} + 1} - \frac{\left(\frac{\sigma_{j,M}}{v_{j,1}^{-1}\left(\sigma_j^2 \sigma_M^2 v_{M,1}^{-1} + \sigma_j^2 - \sigma_{j,M}^2 v_{M,1}^{-1}\right) + \sigma_M^2 v_{M,1}^{-1} + 1}\right)^2}{\frac{v_{j,1}^{-1}\left(\sigma_j^2 \sigma_M^2 - \sigma_{j,M}^2\right) + \sigma_M^2}{v_{j,1}^{-1}\left(\sigma_j^2 \sigma_M^2 v_{M,1}^{-1} + \sigma_j^2 - \sigma_{j,M}^2 v_{M,1}^{-1}\right) + \sigma_M^2 v_{M,1}^{-1} + 1} + \frac{1}{v_{M,2}^{-1}}}}}
\end{aligned}$$

$$= \frac{1}{v_{j,2}^{-1} + \dfrac{1}{\dfrac{\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}}{v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1} - \dfrac{\left(\dfrac{\sigma_{j,\underline{M}}}{v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1}\right)^2}{\left(\begin{array}{l} v_{j,1}^{-1} v_{\underline{M},2}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_{\underline{M}}^2 v_{\underline{M},2}^{-1} \\ + v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1 \end{array}\right)}{\left(v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1\right) v_{\underline{M},2}^{-1}}}}$$

$$= \frac{1}{v_{j,2}^{-1} + \dfrac{1}{\dfrac{\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}}{v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1} - \dfrac{\dfrac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1}}{v_{j,1}^{-1}\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) + \sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1}}}$$

$$= \frac{1}{v_{j,2}^{-1} + \dfrac{\left[v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right) + \sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1\right]\left[v_{j,1}^{-1}\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) + \sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1\right]}{\left\{\begin{array}{l} v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \end{array}\right\}}}$$

$$= \frac{1}{v_{j,2}^{-1} + \dfrac{\left\{\begin{array}{l} v_{j,1}^{-2}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + v_{j,1}^{-1}\left[\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1\right) + \left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right)\right] \\ + \left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + 1\right) \end{array}\right\}}{\left\{\begin{array}{l} v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right) + 1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \end{array}\right\}}}$$

$$= \frac{1}{v_{j,2}^{-1} + \frac{\left\{\begin{array}{l} v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \end{array}\right\}\left\{ v_{j,1}^{-1} + \frac{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}} \right\}}{\left\{\begin{array}{l} v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \end{array}\right\}}}$$

The factorization of the numerator of the large fraction in the equation above is easy to verify for the term in $v_{j,1}^{-2}$ and for the independent term. However, it is much more tricky for the term $v_{j,1}^{-1}$. We propose to check now that the product of the two factors resulting from the factorization, $\left\{\begin{array}{l} v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right) \\ + \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \end{array}\right\}$ and $\left\{ v_{j,1}^{-1} + \frac{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}} \right\}$, is indeed identical to

$v_{j,1}^{-2}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right)$

$+ v_{j,1}^{-1}\left[\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) + \left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right)\right]$ in the $v_{j,1}^{-1}$ term:

$+ \left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)$

Therefore, we must verify the following equality:

$$\begin{aligned} &\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1} \\ &+ \frac{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) - \sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}} \\ &= \left[\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1} + \sigma_j^2 - \sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right) + \left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2 - \sigma_{j,\underline{M}}^2\right) + \sigma_j^2\right)\right] \end{aligned}$$

Notice that it comes back to check the following subsequent equalities:

$$\Rightarrow \frac{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}$$

$$=\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)$$

$$\Rightarrow\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)$$

$$\left\{\frac{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)}\right\}$$

$$=\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)$$

$$\Rightarrow\left\{\frac{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)}{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)}}-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)}\right\}=1$$

$$\Rightarrow\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)-\left[\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)}\right]\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}$$

$$=\left[\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)}\right]\left[\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)\right]$$

$$\Rightarrow \left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)}=\frac{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)}$$

$$\Rightarrow \left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}=\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)$$

$$\Rightarrow \sigma_j^2\sigma_{\underline{M}}^4 v_{\underline{M},1}^{-2}+\sigma_j^2\sigma_{\underline{M}}^4 v_{\underline{M},1}^{-1}v_{\underline{M},2}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},2}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-2}-\sigma_{j,\underline{M}}^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}v_{\underline{M},2}^{-1}-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}$$
$$=\sigma_j^2\sigma_{\underline{M}}^4 v_{\underline{M},1}^{-2}-\sigma_{\underline{M}}^2\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-2}+\sigma_j^2\sigma_{\underline{M}}^4 v_{\underline{M},1}^{-1}v_{\underline{M},2}^{-1}-\sigma_{\underline{M}}^2\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}v_{\underline{M},2}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},2}^{-1}-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}+\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2$$

This last equality is verified. The factorization is correct.

We can continue the initial development:

$$=\frac{1}{v_{j,2}^{-1}+\frac{\left\{\begin{array}{l}v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)\\+\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}\end{array}\right\}\left\{v_{j,1}^{-1}+\frac{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}\right\}}{\left\{\begin{array}{l}v_{j,1}^{-1}\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)\left(\sigma_j^2\sigma_{\underline{M}}^2-\sigma_{j,\underline{M}}^2\right)+\sigma_j^2\right)\\+\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}\end{array}\right\}}}$$

$$=\frac{1}{v_{j,2}^{-1}+v_{j,1}^{-1}+\frac{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}{\left(\sigma_j^2\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+\sigma_j^2-\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}\right)\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)-\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}}=\frac{1}{v_{j,1}^{-1}+v_{j,2}^{-1}+\frac{1}{\sigma_j^2-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},1}^{-1}}{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}-\frac{\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1}+v_{\underline{M},2}^{-1}\right)+1\right)\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}}}$$

$$= \frac{1}{v_{j,1}^{-1} + v_{j,2}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1} \frac{v_{\underline{M},1}^{-1}\left(\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1\right)+\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}}}$$

$$= \frac{1}{v_{j,1}^{-1} + v_{j,2}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1} \frac{v_{\underline{M},1}^{-1}\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)+v_{\underline{M},1}^{-1}\sigma_{\underline{M}}^2 v_{\underline{M},2}^{-1}+\sigma_{j,\underline{M}}^2 v_{\underline{M},2}^{-1}}{\left(\sigma_{\underline{M}}^2 v_{\underline{M},1}^{-1}+1\right)}}}$$

$$= \frac{1}{v_{j,1}^{-1} + v_{j,2}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)+1}\left(v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}\right)}} = \frac{1}{v_{j,1}^{-1} + v_{j,2}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,\underline{M}}^2}{\sigma_{\underline{M}}^2 + \frac{1}{v_{\underline{M},1}^{-1} + v_{\underline{M},2}^{-1}}}}}$$

By a recursive reasoning, we finally have that:

$$Var(\tilde{p}_{j,t+1} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}) = \frac{1}{\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \frac{1}{\sigma_j^2 - \frac{\sigma_{j,M}^2}{\sigma_M^2 + \frac{1}{\sum_{q=1}^{Q_M} v_{M,q}^{-1}}}}}$$

What has been done for the variance can also be done for the covariance in order to show that:

$$Cov(\tilde{p}_{j,t+1}, \tilde{p}_{M,t+1} \mid z_{j,1},...,z_{j,q},...,z_{j,Q_j}, z_{M,1},...,z_{M,q},...,z_{M,Q_M}) = \frac{\sigma_{j,M}}{1+\sigma_j^2\sum_{q=1}^{Q_j} v_{j,q}^{-1} + \sigma_M^2\sum_{q=1}^{Q_M} v_{M,q}^{-1} + \left(\sum_{q=1}^{Q_j} v_{j,q}^{-1}\right)\left(\sum_{q=1}^{Q_M} v_{M,q}^{-1}\right)\left(\sigma_j^2\sigma_M^2 - \sigma_{j,M}^2\right)}$$

Appendix B

	Moyenne	Std-dev	Min	Max	Skewness	Kurtosis
Prob	0.33	0.38	0.00	1.00	0.15	0.80
D	0.26	0.44	0.00	1.00	0.19	1.11
D1	0.12	0.32	0.00	1.00	0.10	2.41
D2	0.06	0.25	0.00	1.00	0.06	3.55
D3	0.03	0.18	0.00	1.00	0.03	5.10
D4	0.01	0.09	0.00	1.00	0.01	11.03
NBN	0.76	2.61	0.00	185.00	6.80	12.84
varmrkt	1.02E-04	1.04E-04	2.20E-05	1.10E-03	1.07E-08	3.88
varsect	1.75E-04	3.36E-04	5.76E-06	7.51E-02	1.13E-07	141.20
vol(-1)	1.06E+08	2.74E+08	0.00	3.17E+10	7.48E+16	12.76

	Prob	D	D1	D2	D3	D4	NBN	varmrkt	varsect	vol(-1)
Prob	1.00									
D	0.09	1.00								
D1	0.09	0.61	1.00							
D2	0.09	0.45	0.73	1.00						
D3	0.08	0.32	0.52	0.72	1.00					
D4	0.06	0.15	0.25	0.34	0.48	1.00				
NBN	0.09	0.50	0.45	0.45	0.45	0.39	1.00			
varmrkt	0.26	0.02	0.02	0.02	0.02	0.00	0.01	1.00		
varsect	0.12	0.02	0.02	0.01	0.01	0.01	0.03	0.30	1.00	
vol(-1)	0.09	0.27	0.11	0.08	0.07	0.04	0.39	0.03	0.10	1.00

Autocorrelations					
	lag 1	lag 2	lag 3	lag 4	lag 5
Prob	0.96	0.91	0.86	0.81	0.77
D	0.40	0.35	0.34	0.33	0.33
D1	0.19	0.09	0.08	0.07	0.07
D2	0.16	0.06	0.05	0.04	0.04
D3	0.14	0.04	0.03	0.03	0.02
D4	0.09	0.02	0.01	0.01	0.01
NBN	0.59	0.44	0.40	0.39	0.38
varmrkt	0.91	0.85	0.76	0.68	0.62
varsect	0.49	0.34	0.29	0.26	0.25
vol(-1)	0.88	0.84	0.821	0.813	0.805