Master Thesis, Finance

Option Volume and Evidence of Informed Trading

An empirical study of daily option trading volume from selected S&P500 companies in the period 2009 to 2014.

Fredrik Honningsvåg Stene and Mattias Thuve

Supervisor: Jørgen Haug

NORWEGIAN SCHOOL OF ECONOMICS

Bergen, Spring 2017

Abstract

This thesis seeks to unveil evidence of informed trading in option markets. We use unsigned option volume data to construct a signed modified put call ratio, which is used to analyze unusual trading patterns prior to large spikes in abnormal returns on the underlying equities. The data sample consists of daily option volume of approximately 350 000 options from 232 random companies listed on S&P500 between the 1^{st} of June 2009 to the 6^{th} of August 2014. We conduct statistical tests across time, across firms, and across both simultaneously to identify informed trading under the assumption of a semi efficient market; and investigate any preferences an informed investor might have with regard to selected firm characteristics and timing. We discover evidence of unusual trading patterns one day prior to large spikes in abnormal returns, and find supporting evidence that informed traders prefer out-of-the-money compared to at-the-money and in-the-money options. However, we do not find any significant linkage between a company's market value or price-to-book-value, or that the amount of informed trading in the option market has decreased with time.

Preface

This thesis was written to conclude the degree MSc in Economics and Business Administration at the Norwegian School of Economics during the spring of 2017. The thesis' subject was chosen by the authors, and reflects our interest for the financial markets and background with empirical analysis.

The process has been an educational experience not only from a finance perspective, but also from a programming perspective. The scope of our data set would not have been possible to handle manually, and we are grateful that NHH in the latter years has started offering courses in programming. Our analysis have been conducted in R and the thesis was written in Latex. We believe programming will be a very important tool for economists in the future, and hope that NHH will continue its commitment to a variety of programming courses.

We would like to give a special thanks to our supervisor, Jørgen Haug, for providing excellent guidance, input and constructive criticism throughout the process of writing and finishing our thesis.

Contents

Abstract 1
Preface 2

1 Introduction **7**
1.1 Motivation 7
1.2 Hypotheses 8

2 Literature Review **10**

3 Theory **12**
3.1 Informed Trader 12
3.2 Efficient Market 12
3.3 Option Volume 13
3.4 ATM, ITM and OTM 14

4 Methodology **15**
4.1 Event Studies 15
4.1.1 Step 1 – The Event of Interest 16
4.1.2 Step 2 – Data Selection 17
4.1.3 Step 3 – Modelling Abnormal Returns 22
4.1.4 Step 4 – The Estimation Window 23
4.1.5 Step 5 – Designing the testing framework 24
4.2 Model assumptions 32
4.3 Other issues 33
4.4 Data Presentation 34

5 Empirical Analysis **35**
5.1 Panel Data 35
5.1.1 Abn > 0.5 % 36
5.1.2 Abn > 1.0 % 38
5.2 Time Series 39
5.2.1 Abn > 0.5 % 39
5.2.2 Abn > 1.0 % 47

5.3 Cross Sectional 53
5.3.1 Abn > 0.5 % 53
5.3.2 Abn > 1.0% 59

6 Conclusion **66**

Appendices **72**

A Tickers **72**

B Tables **74**
B.1 Time Series 74
B.2 Cross Sectional 75
B.3 Panel Data 76

C Figures **77**
C.1 Time Series 77
C.2 Cross Sectional 82
C.3 Residual sample plots 89

D LOESS **91**

List of Figures

1 Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs log(MV): Abn > 0.005 42
2 Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.005 42
3 Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs log(MV): Abn > 0.005 . . 43
4 Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.005 44
5 Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.005 44
6 Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.005 45
7 Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs PTBV: Abn > 0.005 . . . 46
8 Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs PTBV: Abn > 0.005 . 46
9 Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs log(MV): Abn > 0.01 49
10 Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.01 50
11 Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs log(MV): Abn > 0.01 . . 50

12 Time-series PCrat$_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.01 51
13 Time-series PCrat$_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.01 52
14 Time-series PCrat$_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.01 52
15 Cross-sectional PCrat$_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.005 56
16 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.005 57
17 Cross-sectional PCrat$_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.005 58
18 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.005 59
19 Cross-sectional PCrat$_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.01 62
20 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.01 63
21 Cross-sectional PCrat$_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.01 64
22 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.01 65
23 Time-series PCrat$_{t-n}^{(*)}$ regression coefficients vs PTBV: Abn > 0.01 . . . 77
24 Time-series PCrat$_{t-n}^{OTM(*)}$ regression coefficients vs PTBV: Abn > 0.01 . 77
25 Time-series PCrat$_{t-n}^{(*)}$ regression coefficients vs log(MV): Abn > 0.015 . . 78
26 Time-series PCrat$_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.015 78
27 Time-series PCrat$_{t-n}^{(*)}$ P-values vs log(MV): Abn > 0.015 79
28 Time-series PCrat$_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.015 79
29 Time-series PCrat$_{t-n}^{(*)}$ regression coefficients vs PTBV: Abn > 0.015 . . . 80
30 Time-series PCrat$_{t-n}^{OTM(*)}$ regression coefficients vs PTBV: Abn > 0.015 . 80
31 Time-series PCrat$_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.015 81
32 Time-series PCrat$_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.015 81
33 Cross-sectional PCrat$_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.015 82
34 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.015 83
35 Cross-sectional PCrat$_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.015 84
36 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.015 85
37 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.005 86
38 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.01 87
39 Cross-sectional PCrat$_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.015 88
40 AAPL residual plot, Abn > 0.05 89

41 01.02.2013 residual plot, Abn > 0.05 90

List of Tables

1 Option data 2009 - 2011 20
2 Option data 2012 - 2014 21
3 Independent variables 2009 - 2011 30
4 Independent variables 2012 - 2014 31
5 Two-way fixed-effect panel regression : Abn > 0.005 36
6 Two-way fixed-effect panel regression : Abn > 0.01 38
7 Time-series regression P-values and coefficients : Abn > 0.005 40
8 Time-series regression explanatory power and observations : Abn > 0.005 41
9 Time-series regression P-values and coefficients : Abn > 0.01 47
10 Time-series regression explanatory power and observations : Abn > 0.01 48
11 Cross-sectional regression P-values and coefficients : Abn > 0.005 54
12 Cross-sectional regression explanatory power and observations : Abn > 0.005 55
13 Cross-sectional regression P-values and coefficients : Abn > 0.01 60
14 Cross-sectional regression explanatory power and observations : Abn > 0.01 61
15 Stock tickers and company names, A - IBM 72
16 Stock tickers and company names, IFF - ZBH 73
17 Time-series regression P-values and coefficients : Abn > 0.015 74
18 Time-series regression explanatory power and observations : Abn > 0.015 74
19 Cross-sectional regression P-values and coefficients : Abn > 0.015 75
20 Cross-sectional regression explanatory power and observations : Abn > 0.015 75
21 Two-way fixed-effect panel regression : Abn > 0.015 76

1 Introduction

1.1 Motivation

Investors with access to private information[1] can choose to trade both in stock and derivative markets, potentially achieving a higher expected profit than under normal circumstances. Informed traders, traders who have better forecasts on stock returns than others, trade frequently on such information(Meulbroek, 1992). Informed trades that are actual corporate insiders[2] undermines the level playing field that is fundamental for a well working, fair and functional capital market by doing so. To balance out such flaws, decades of research have been conducted to detect how and where informed traders trade.

Previous research has mostly been aimed towards trade volume in stock markets, where there has been found a significant relationship between stock volume and abnormal returns (see, Chordia and Swaminathan (2000); Chakravarty et al. (2004); Lamoureux and Lastrapes (1990)), indicating informed trading. The capital market does however consist of many other instruments. Black (1975) already published a paper where he concluded that informed investors would find the option market more favorable, given the higher leverage it offers. Empirical studies such as Easley, O'hara, and Srinivas (1998) have found that option volume contains price discovery, and Ge, Lin, and Pearson (2016) found that certain signed options do predict future abnormal returns. Through research there has been established a relationship between option volume and abnormal returns. Researchers have also investigated further, finding evidence that informed traders prefer out-of-the-money options to in-the-money and at-the-money options, reasoning that the higher leverage in out-of-the-money options is favorable when an investor is informed.

Data used in most of the aforementioned papers is older[3]. This paper aims to see if similar results can be found using newer, unsigned data and proposes the following research question:

[1]Private information is information affecting a company's value but not yet announced to the public

[2]A company employee who can take advantage of their privileged position and knowledge

[3]Easley et al. (1998) use daily data from October and November 1990, Chakravarty et al. (2004) use data for the period 1988 to 1992 and Ge et al. (2016) use data for the period 2005 to 2012.

Do changes in trading volume of options imply evidence of informed trading before large spikes in abnormal returns?

The question will be answered with data gathered from 232 random companies listed on S&P500 between 2009 and 2014.

The thesis is structures as following: The objective and our reasoning for pursuing informed trading is summarized in a hypotheses section. To further support these hypotheses, a literature review of relevant work is then presented. Following this is a presentation of all relevant theories needed to correctly define the analysis and interpret its results, before an extensive and detailed presentation of the methodology framework and associated model assumptions are discussed. Finally, a three-part empirical analysis is presented through figures, tables and text, before a discussion of our main findings are summarized into a final conclusion.

1.2 Hypotheses

By analyzing the relative share of call and put option volume we attempt to discover evidence that informed traders trade on private information. Our analysis will focus on the importance of option volume, with further analysis of the statistical and economic significance of the call to put measure [4] compared to firm size, PTBV and variations over time.

The underlying assumption for the following hypothesis is that informed traders are capital-constrained, and therefore seeking to maximize their profits through leveraging their position, i.e. buying options instead of stocks.

Following are the justifications for all hypotheses[5].

- If there is informed trading in the market, and this information is not further leaked, an investor with capital constraints will benefit from leveraged trading strategies. A leveraged position will maximize an investors profits, and even though the investor might not invest all their capital in option markets, we believe that it is highly likely that they at least partly engage in option trades (Black, 1975). If these assumptions

[4] The variable is formally defined in chapter 4.3.5

[5] The hypotheses presented are formulated as alternative hypotheses.

hold, we would expect to find an abnormal pattern in the relative call to put option volume prior to new information being released to the market, i.e. prior to spikes in abnormal return.

H1: *There is evidence of a relative change in call to put option volume prior to large spikes in abnormal returns on the underlying asset.*

- Since out-of-the-money (OTM) options offer a higher leverage, Chakravarty, Gulen, and Mayhew (2004) suggest that investors prefer OTM over at-the-money (ATM) and in-the-money (ITM) options. If the stock is anticipated to increase, investors could prefer to buy cheap OTM call options which will end up ITM. For stocks anticipated to decrease, investors could prefer to buy cheap OTM put options which will end up ITM when the stock decreases. However, there are several different strategies an investor could pursue to gain a leveraged profit. They could sell ITM put options which falls in value when the underlying stock increased. Following put-call parity, investors could buy ITM puts and the underlying stock, financed by borrowing, mimicking the strategy of buying OTM calls.

  All these strategies are possible, but since we believe that OTM options are most preferred by investors, this will be our area of interest. This is further supported by Chakravarty, Gulen, and Mayhew (2004) who finds that OTM options have a significant higher information share than ATM or ITM options.

  H2: *There is evidence of a relative change in the isolated out-of-the money call to put option volume prior to large spikes in abnormal returns on the underlying asset*

- For a large firm, measured by marked value, new information is less likely to cause a spike in abnormal returns (Chari et al., 1988). The reason for this is that larger firms' profitability is not affected as much by one single piece of information, i.e. a new deal, new products etc., as smaller firms would be. Therefore, it is reasonable to believe that informed traders prefer smaller companies when trading on private information. However, one might also argue that the larger liquidity found in the top firms could better disguise illegal trades.

  H3: *Market value affects an informed trader's decision; informed traders are more likely to trade options in firms with lower market value.*

- This hypothesis suggests that firms with a lower price-to-book-ratio (PTBV) are more likely to show signs of informed trading. A low PTBV (compared to peers) can indicate two things; firstly, that the firm is potentially undervalued (Jensen et al., 1997), and secondly that the firm is close to default. In both cases, the firm's stock will be susceptible to large price changes when new information, good or bad, is released in the market. It is thus reasonable to believe informed trading will happen more often in firms with a low PTBV.

  H4: *Price-to-book-ratio affects an informed trader's decision; informed traders are more likely to trade options in firms with lower price-to-book-ratio.*

- We believe that the significance of the call to put option volume ratio, both statistically and economically, should decrease over time. As earlier research by Roll, Schwartz, and Subrahmanyam (2010) and Johnson and So (2012) has shown, variations of option ratios contain information about future stock returns. Such information should, in line with the efficient market hypothesis, be reflected in stock prices after discovery. As our time frame includes the release of both of these papers, we expect the significance to decline after 2010.

  H5: *There is evidence that the statistical and economic significance of the call to put option volume ratio has decreased over time.*

2 Literature Review

Our work relates to previous research done on the relationship between option volume and future stock movements. This paper is not a direct replica of a specific paper, but rather a summary of previous ideas gathered from various researchers. When we defined the scope of this paper, we took inspiration from literature ranging from establishing a simple relationship between trade volume and changes in price, as done by Morse (1980), to establishing a theory on where informed traders trade, as done by Easley, O'hara, and Srinivas (1998). Several authors have analyzed the relationship between informed trading and option volume. Chakravarty, Gulen, and Mayhew (2004) found through their "information share" approach that option markets contain price discovery, while Ge, Lin, and Pearson (2016) found that it is the embedded leverage in options which contributes most to why option trading predicts stock returns, and Johnson and So (2012) shows that O/S[6] is a signal of private information. It is also suggested that if informed trading takes place in option markets, then option volume is expected to contain information about future stock prices (Pan and Poteshman, 2006).

In accordance with the theoretical prediction that informed traders choose to trade in options markets, several authors have done extensive research on different measures to prove this relationship. First, O/S has been a frequently examined ratio to address what drives volume in options relative to their underlying equities. By using data on option and equity volumes when trade direction is unobserved, Johnson and So (2012) conclude that O/S is a negative cross-sectional signal of private information and future equity return. Roll, Schwartz, and Subrahmanyam (2010) gathers equities and their listed options, and by analyzing the time-series properties of this ratio they conclude that O/S rise sharply prior to an earnings announcement. And Ge, Lin, and Pearson (2016) concludes that O/S is a future stock return predictor. Secondly, option volume has been examined against equity returns, both the absolute value of returns (Blume, Easley, and O'hara, 1994) and abnormal returns (Cao, Chen, and Griffin, 2005), were both papers conclude that option volume and price movements relate. Thirdly, researchers have been examining the moneyness[7] of options and if it affect where informed traders trade. Through an analysis of abnormal option trading volume prior to M&A announcements, Augustin, Brenner, and Subrahmanyam (2015) find that the strongest effects are in the OTM call options. Chakravarty, Gulen, and Mayhew (2004) and Cao et al. (2005) also find supporting results towards the theory that informed traders prefer to trade in OTM options, given the higher leverage achievable. And finally, several authors use daily option data to investigate the informational role of transaction volume in options markets (see, Easley, O'hara, and Srinivas (1998), Augustin, Brenner, and Subrahmanyam (2015)).

While several authors discover option volume to be informative, this conclusion is not unambiguous and other authors find no such evidence (see, Vijh (1990), Stephan and Whaley (1990) and Chan, Chung, and Fong (2002))

[6]The authors define O/S as the option to stock volume ratio. This is the ratio of total option market volume to total equity market volume

[7]Option moneyness is a terminology used to define if an option is in-the-money, at-the-money or out-of-the-money

We apply several of these researcher's techniques and ideas in our thesis when defining our hypotheses, collecting relevant data, deciding the methodology and empirical tests, variable usage and interpretation of results.

3 Theory

3.1 Informed Trader

An informed trader is a trader who has better forecasts of the stock's future returns than others. There are two categories of informed traders. First, the trader could be a corporate insider trading on private information or an investor gaining an advantage by receiving private information externally. Secondly, an informed trader could be a trader that has special talents for correctly interpreting public available information. Since both types of traders have access to information that could lead to higher expected returns than others, they are considered informed traders.

3.2 Efficient Market

Under ideal conditions the capital market allows investors to allocate resources through investments in capital stock, under the assumption that security prices fully reflect all available information. Fama (1970) defines such markets as "efficient", and states that a market is efficient when it is frictionless and consisting of investors all agreeing on the implication new information has on price adjustment.

Fama considered the null hypothesis that security prices at any point "fully reflects" all available information to be too extreme. Instead he categories market efficiency into three sub-categories. Through this he could study at which point the hypothesis breaks down. The three categories are known as weak, semi-strong and strong form of market efficiency and are individually defined as the following:

- The weak form consists of historical information and the subset of interest consist of past returns.

- Semi-strong form consists of all public available information, such as earning announcements, stock splits, mergers etc.
- Strong form consists of all available information, even private information not available for the public. This could be an investor or a group with monopolistic access to any information with relevance for price adjustment.

Extensive empirical work has been performed on all three sub-categories and shows that there are little or no evidence against the weak and semi-strong form of efficiency (see Fama et al. (1969) and Ball and Brown (1968)). It seems that prices efficiently adjust to all public available information. At the strong level of efficiency there are some, but limited, evidence against the hypothesis (Niederhoffer and Osborne, 1966). This could mean that there are individual investors with monopolistic access to information who gain a higher expected trading profit than others. This is however limited to only two documented groups, corporate insiders and specialists. Fama (1970) finds no other evidence that a deviation from the strong form spirals any further down the investment community.

3.3 Option Volume

An investor with access to information prior to its public announcement, may be able to gain a profit by utilizing this private information through informed trading in financial markets. To detect such trades, it is important to first know where informed traders choose to trade. In this section, we will discuss the difference of trading in option and stock markets, and why option volume might reflect information on future stock prices.

Black (1975) describes the differences between trading in option and stock markets as the following; here exemplified by a European call option. If the stock at option maturity equals the exercise price, the option is worthless and the initial investment (price of call option at t = 0, C_0) is lost. The option position is thus worse than the stock position. If the stock goes down, the option will not be exercised and one loses C_0 as in the previously example. The stock position can however go down any amount up to the initial investment (price of stock at t = 0, S_0). If the stock goes up, the option will gain a higher rate of return than the stock, as $C_0 \leq S_0$. Due to higher leverage found in options, realizing a high return has a higher probability when investing in this market. An informed trader, who is be convinced of the coming changes in the underlying stock price, might find the higher leverage in option markets favorable. To secure higher returns, an informed trader will thus be expected to trade more actively in options, rather than in the underlying stock.

If informed traders prefer option markets, then these markets may be venue for information-based trading, suggesting that option trades may reflect information to market participants on future changes in stock and option prices. This is supported by Easley, O'hara, and Srinivas (1998) who provide a theoretical model that shows under which conditions an informed trader would choose to trade in options rather than stocks. Johnson and So (2012) also support this idea by providing theoretical and empirical evidence that informed traders' private information is reflected in the O/S-ratio. In both cases option volume data are used to detect the use of private information.

3.4 ATM, ITM and OTM

The leverage and capital constraints arguments(Chen et al., 2005) imply that the rate of informed trading incident is unevenly distributed over different option moneyness (i.e. different leverage). Option volume is thus categorized into three sub-categories; in-the-money (ITM) for options with strike prices lower than the current stock price (S_t), at-the-money (ATM) for options with strike price equal S_t, and out-of-the-money (OTM) for options with strike price higher than S_t. The limits for ITM and OTM are set to 5 % below or above the current stock price.

For OTM options, theory proposed by Chakravarty, Gulen, and Mayhew (2004) explains that an informed trader will favor the higher leverage and therefore trade more frequently in these options. However, for delta equivalent positions, the option bid-ask spread and trader commissions tend to be the widest/highest for OTM options. ATM options usually have the lowest bid-ask spread, while ITM options usually have the lowest commissions of the three categories. The theory also proposes that it is easier for informed traders to conceal their trades in ATM options, as this category has the highest trading volume due to volatility traders. However, ATM options are more susceptible to risk from changes in the underlying asset's volatility.

It is possible to argue that informed investors would choose to invest in any of the three sub-categories of options. Chakravarty, Gulen, and Mayhew (2004) find, through the empirical implementation of their framework, a relationship between option volume and stock prices. They further extended their research and analyze these effects over moneyness. They concluding that while all three sub-categories contain price discovery, the OTM options have a significant higher amount, indicating that informed traders trade more frequently in OTM options.

This theory section has provided definitions for informed traders, and why option volume and different option types are of interest to such an investor. With the assumption of a semi-efficient market as presented by Fama (1970), our empirical framework presented in the next section allows us to conclude whether or not changes in trading volume of options imply evidence of informed trading before large spikes in abnormal returns.

4 Methodology

This chapter addresses our implementation of the MacKinlay framework (MacKinlay, 1997) for event studies, data collection, parameter definitions and the statistical tests which are to be carried out. We also address some concerns regarding model assumptions.

4.1 Event Studies

Under the assumption that the market is semi-efficient, a release of information to the public should yield an immediate effect in security prices. Financial data can be studied to measure the impact of such an event. The MacKinlay-paper is based on previous work by Ball and Brown (1968) and Fama, Fisher, Jensen, and Roll (1969) and is one of the most widely used frameworks for event studies in empirical finance. The methodology can be summed up in five distinct steps; 1. Defining the event of interest, 2. Data selection, 3. Model for abnormal returns, 4. Estimation window, 5. Designing the testing framework.

The following section will describe the aforementioned methodology behind an event study, define models for measuring and analyzing abnormal performance, and highlight issues arising when conducting these studies.

4.1.1 Step 1 – The Event of Interest

The first step is defining the event of interest. We have defined our events as any day, t, a company, j, has an absolute abnormal stock return higher than x %.

In our case, were we want to examine whether daily option volume data prior to the event influence abnormal returns, we have set the event window to +/- 2 days prior to and past the event. We assume informed traders are interested in trading on private information to gain a competitive market advantage and believe that such trades will occur in the days before a large change in abnormal return. If the market is efficient at any level, such spikes should only occur when new information is released into the market. The fact that the information creates a spike will therefore indicate one of two possibilities: Either *announced* news showed surprising or unexpected results, or the new information was *unannounced*. In both incidences, an investor with private information would be able to profit by trading in the days prior to the information release. While the event itself, by definition, only last for one day, the cause of the event (new information), may have an effect before and after the event[8]. To ensure robustness in the analysis the event window can be set larger than the event of interest so that any estimations does not become biased by the event itself.

In our regression models[9], the events are expressed with the following variable:

- *Abnormal day*, noted $Abn_{j,t}$ (or as an asterisk, *, when modifying other variables), are any day, t, a company, j, has an absolute abnormal stock return higher than x %.

  A binomial variable is created to capture these events, and is defined as the following:

$$Abn_{j,t} = \begin{cases} 1 \; if \; |AR_{j,t}| > x \; \% \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

[8]If the semi-efficient market hypothesis does not hold

[9]Formally defined in 4.1.5

4.1.2 Step 2 – Data Selection

We have aimed to make our analysis as comprehensive as possible, and have ended up with a sample of 232 companies from the S&P500.

The reason for choosing the S&P500 market over others, is based on data availability and market liquidity. There are several databases with high quality data for equities listed on S&P500. More reliable data leads to more robust conclusions; this is thus an important factor to consider when deciding which market to analyze. The S&P500 is also highly liquid compared to smaller markets, i.e. Oslo Børs, which we believe will help us better isolate the effects we are analysing.

The data used in the analysis is gathered from Thomson Reuters Datastream, a database that offers daily updated equity data, and historical equity information as far back as 1973. The database also offers data on bonds, options and other derivatives, which is what we were after. Datastreams option database has data for most of the companies which are or have been listed on the S&P500 dating back to mid-2008.

We have chosen to restrict our data collection to the period from (and including) 01.01.2009, to (and including) 31.12.2014. The lower limit due to 2009 being the first whole year with available data in Datastream, and the upper limit to make sure that most of the options traded in the period are dead. This is a necessity caused by the way Datastream structures its option data, as options are either classified as dead or live. Both classes can be accessed separately with ease, but any combination must be done manually. Our limited time frame for this thesis would not have been enough for such manual work, and thus we set 2014 as our upper limit.

Our 232 companies were selected based on the following criteria. The companies must have been listed on the S&P500 continuously from 2009 to 2014. This were true for a total of 405 companies. The companies must also have available option data on Datastream. This brought our sample down to 398. The order of the 398 companies were randomized to prevent any form of selection bias. We managed to gather data for 232 of these in our limited time frame[10].

Following next is a description of the data downloaded from Datastream and why it has

[10] A complete list of the 232 stock tickers and company names can be found in the appendix

been included.

- *Option Volume* is measured in traded option contracts and is quoted in daily measures. It is our main variable of interest, and therefore crucial to include in the analysis. We have downloaded option volume for both put and call options. The volume is unsigned, i.e. we do not know which way the trades have gone. Each option contract is valid for 100 shares.

- *Strike price* is the specific price that a contract can be exercised at. It is used to measure when an option contract is ITM, ATM or OTM.

- *Adjusted Stock Price* is the adjusted closing price of the underlying equity. As this price is adjusted for dividends and share repurchases it can be used to calculate stock returns.

- *Unadjusted stock price* is the actual closing price of the underlying equity. As this price is unadjusted for dividends and share repurchases it can be used to determine which options are OTM in conjunction with the *Strike price*.

- The *S&P500 index* is necessary to calculate the abnormal returns together with actual stock returns. Under the assumptions of the market model, S&P500 functions as a proxy for the market return[11].

- *Delta*, or hedge ratio, is a measure used to compare changes in the price of the underlying asset to the corresponding change in the derivative. If an option has a 0.5 delta ratio, a $1 change in the underlying asset will generate a $0,5 change in the option. A put option can have a delta value between -1 and 0, and a call options can have a delta value between 0 and 1.

- *Market value* is included to adjust for firm size. Our earlier discussions suggest that larger, more liquid firms, might be better at keeping private information private. Thus, one could expect smaller firms to be more exposed for informed trading. However, we also argued that trading in more liquid firms could better disguise illegal trades by corporate insiders. Market value is meant to control for this, and possibly detect any relationship between firm size and informed trading.

[11]See section 4.1.3

- *PTBV*, Price to tangible book value, is a ratio which expresses the price of a security compared to its tangible book value. It is a theoretical number that represents what an investor would receive per share if a company would liquidate all its assets. PTBV is used as a control variable.

- *EPS*, Earnings per share, is included as a control variable. This controls for whether the level of earnings is related to the abnormal return.

The raw option data was cleaned by removing untraded options and options which contained any kind of errors. Banking holidays were removed for all option and stock data, so that days with zero trade will not interfere and possibly create biases in the regressions. This resulted in 1510 trading days over the period.

The cleaned option data is summarized in table 1 and 2. Note that the numbers for 2009 and 2014 are lower than the other four years, 2010 to 2013, due to the data being cut short by 102 trading days in each end. In terms of different options traded, 2013 was the most active, counting 122 000 different options. The average number of different options traded for the four full years is just above 100 000. 2011 was however the most active year in terms of number of contracts, with just above 840 000 000 contracts traded. The split between call and put contracts is, for all years, between 60 % to 40 % and 65 % to 35 %, and the share of OTM contracts has been somewhere around 40 %. Note that the share of OTM put options is higher than the share of OTM call options for all years. The share of OTM put options is also increasing with time, going from $\sim$ 53 % in 2009 to $\sim$ 65 % in 2014. The share of OTM call options is on the other hand decreasing from $\sim$ 37 % in 2009 to $\sim$ 26 % in 2014. This shift is also illustrated by the change in average values for the independent values in table 3 and 4.

Table 1: Option data 2009 - 2011

	2009				2010				2011			
	Min	Average	Max	Total	Min	Average	Max	Total	Min	Average	Max	Total
Companies				214				232				232
Diff. options	37	204	1 028	47 220	32	370	1 735	85 760	57	476	2 920	110 433
Contracts_{Total}	2.29	687	42 092	159 285	3.75	2 212	83 829	513 200	2.63	3 625	96 409	840 952
Contracts_{Call}	1.25	424	27 992	98 381	1.93	1 470	62 263	341 049	1.66	2 288	57 386	530 803
Contracts_{Put}	0.70	263	14 100	60 904	1.62	742	21 567	172 151	0.77	1 337	40 328	310 149
$\text{Contracts}_{Call}^{(\%)}$	33.85 %	60.85 %	92.26 %		16.45 %	63.91 %	96.30 %		35.76 %	61.48 %	93.91 %	
$\text{Contracts}_{Put}^{(\%)}$	7.74 %	39.15 %	66.15 %		3.70 %	36.09 %	83.55 %		6.09 %	38.52 %	64.24 %	
$\text{Contracts}_{Total}^{OTM}$	0.81	332	25 400	76 923	1.62	920	46 069	213 466	1.20	1 524	53 079	353 454
$\text{Contracts}_{Call}^{OTM}$	0.45	192	17 105	44 457	0.68	515	31 511	119 398	0.56	823	32 888	190 887
$\text{Contracts}_{Put}^{OTM}$	0.30	140	8 295	32 466	0.79	405	14 558	94 068	0.55	701	20 192	162 568
$\text{Contracts}_{Call}^{OTM(\%)}$	8.00 %	37.49 %	69.72 %		1.46 %	31.82 %	99.54 %		4.36 %	34.79 %	99.73 %	
$\text{Contracts}_{Put}^{OTM(\%)}$	23.33 %	53.65 %	90.74 %		24.01 %	55.02 %	96.79 %		26.63 %	53.90 %	99.75 %	
$\text{Contracts}_{Total}^{OTM(\%)}$	17.25 %	43.72 %	72.38 %		3.06 %	39.74 %	98.59 %		10.88 %	41.58 %	99.74 %	
Trading days	98	3 376	24 510	783 129	128	8 475	59 850	1 966 245	292	13 236	153 423	3 070 704
$^{\text{Avg trading days}}/_{\text{option}}$	2.65	13.68	41.72		4.00	19.48	49.98		4.47	23.80	61.42	

The table presents descriptive statistics based on option data gathered from the 1^{st} of June 2009 to the 31^{st} of December 2011. Data for the minimum, average and maximum observed value per company is presented, and the total aggregated value for all companies is found under the column: Total. Contracts_{Total}, Contracts_{Call} , Contracts_{Put} , $\text{Contracts}_{Total}^{(OTM)}$, $\text{Contracts}_{Call}^{(OTM)}$ and $\text{Contracts}_{Put}^{(OTM)}$ are total number of (Total, put or call and OTM or not) contracts observed per company in thousands. $\text{Contracts}_{Call}^{\%}$ and $\text{Contracts}_{Put}^{\%}$ are the percentage of all put or call contracts observed per company, calculated from total contracts. $\text{Contracts}_{Put}^{OTM(\%)}$ and $\text{Contracts}_{Call}^{OTM(\%)}$ are the percentage of put or call options which are OTM, calculated from total put or call contracts observed. Options that are traded over multiple years are listed for each year it is traded. Trading days is the total number of different daily traded options throughout the period and $^{\text{Avg trading days}}/_{\text{option}}$ is the average number of days a specific option is actively traded before it expires. OTM options are defined as an option with $K \geq S_t \cdot 1.05$.

Table 2: Option data 2012 - 2014

	2012				2013				2014			
	Min	Average	Max	Total	Min	Average	Max	Total	Min	Average	Max	Total
Companies				231				231				231
Diff. options	42	510	5 476	118 429	61	526	4 673	121 968	24	382	2 893	88 548
Contracts_{Total}	2.10	3 224	113 361	748 061	3.34	2 448	63 950	567 830	1.99	1 302	45 377	302 098
Contracts_{Call}	1.71	2 067	69 174	479 501	2.40	1 568	40 402	363 688	1.62	860	31 002	199 413
Contracts_{Put}	0.38	1 158	44 187	268 559	0.72	880	23 548	204 142	0.37	443	14 376	102 685
$\text{Contracts}_{Call}^{(\%)}$	29.85 %	60.63 %	91.67 %		32.86 %	62.27 %	87.30 %		30.94 %	64.25 %	88.87 %	
$\text{Contracts}_{Put}^{(\%)}$	8.33 %	39.37 %	70.15 %		12.70 %	37.73 %	67.14 %		11.13 %	35.75 %	69.06 %	
$\text{Contracts}_{Total}^{OTM}$	0.42	1 256	49 118	291 480	0.74	994	30 566	230 722	1.01	530	14 673	122 932
$\text{Contracts}_{Call}^{OTM}$	0.16	621	29 464	144 103	0.27	471	14 335	109 227	0.25	252	7 983	58 509
$\text{Contracts}_{Put}^{OTM}$	0.26	635	27 432	147 377	0.46	524	16 231	121 495	0.37	278	8 699	64 423
$\text{Contracts}_{Call}^{OTM(\%)}$	3.28 %	29.20 %	69.26 %		5.57 %	26.06 %	77.74 %		0.26 %	26.86 %	94.87 %	
$\text{Contracts}_{Put}^{OTM(\%)}$	30.94 %	57.88 %	90.09 %		33.75 %	60.49 %	96.95 %		23.26 %	64.07 %	98.39 %	
$\text{Contracts}_{Total}^{OTM(\%)}$	9.95 %	39.97 %	68.34 %		19.91 %	38.97 %	83.96 %		11.75 %	40.11 %	95.53 %	
Trading days	137	12 995	281 284	3 014 940	221	13 061	219 141	3 030 080	77	7 424	108 045	1 722 299
$^{\text{Avg trading days}}/_{\text{option}}$	3.26	20.99	51.37		3.62	21.12	46.90		3.21	16.63	37.71	

The table presents descriptive statistics based on option data gathered from the 1^{st} of January 2012 to the 6^{th} of August 2014. Data for the minimum, average and maximum observed value per company is presented, and the total aggregated value for all companies is found under the column: Total. Contracts_{Total}, Contracts_{Call} , Contracts_{Put} , $\text{Contracts}_{Total}^{(OTM)}$, $\text{Contracts}_{Call}^{(OTM)}$ and $\text{Contracts}_{Put}^{(OTM)}$ are total number of (Total, put or call and OTM or not) contracts observed per company in thousands. $\text{Contracts}_{Call}^{\%}$ and $\text{Contracts}_{Put}^{\%}$ are the percentage of all put or call contracts observed per company, calculated from total contracts. $\text{Contracts}_{Put}^{OTM(\%)}$ and $\text{Contracts}_{Call}^{OTM(\%)}$ are the percentage of put or call options which are OTM, calculated from total put or call contracts observed. Options that are traded over multiple years are listed for each year it is traded. Trading days is the total number of different daily traded options throughout the period and $^{\text{Avg trading days}}/_{\text{option}}$ is the average number of days a specific option is actively traded before it expires. OTM options are defined as an option with $K \geq S_t \cdot 1.05$.

4.1.3 Step 3 – Modelling Abnormal Returns

The abnormal returns in our analysis are based on the market model; a statistical model that relates company specific returns to a market portfolio. Even though this is the basis of our analysis, other methods need to be discussed and evaluated as well.

Measuring normal performance can loosely be done two different ways, statistical and economic. The key difference is that the statistical models follow statistical assumptions, while the economic model rely on assumptions concerning investors' behaviour and economic restrictions.

Both the constant mean return model and the market model are statistical approaches that rely on the assumption that all returns are jointly multivariate normal and independently and identically distributed through time. MacKinlay (1997) argues that this assumption, while strong, does not lead to problems in practice because it is empirically reasonable to assume this.

The simplest of the two is the constant mean return model, defined as:

$$R_{j,t} = \mu_j + \zeta_{j,t} \tag{2}$$

$$E[\zeta_{j,t}] = 0, Var[\zeta_{j,t}] = \sigma^2_{\zeta_{j,t}} \tag{3}$$

$R_{j,t}$ is the time t return on asset j, μ_j is the mean return of asset j and $\zeta_{j,t}$ is the disturbance term for asset j at time t. Regardless of its simplicity, Brown and Warner (1980) find that it often yields similar results as the more sophisticated models.

The market model is defined as:

$$R_{j,t} = \alpha_j + \beta_j \cdot R_{m,t} + \varepsilon_{j,t} \tag{4}$$

$$E[\varepsilon_{j,t}] = 0, Var[\varepsilon_{j,t}] = \sigma^2_{\varepsilon_j} \tag{5}$$

Where $R_{j,t}$ and $R_{m,t}$ are the time t return for the individual assets j and the market M. $\varepsilon_{j,t}$ is the zero-mean disturbance term. The market model removes the portion of returns which is related to variance in the market's return. This will reduce the variance of the abnormal returns, which serves as an improvement to the constant mean return model.

The parameters α and β can be estimated using general conditions ordinary least squares (OLS). OLS will estimate the normal returns, while the disturbance term represents the estimated abnormal returns.

$$AR_{j,t} = R_{j,t} - \hat{\alpha}_j - \hat{\beta}_j R_{m,t} \tag{6}$$

Under the null hypothesis, the abnormal returns calculated will be jointly normally distributed with a zero-conditional mean and conditional variance, which means that the abnormal returns and its distributional properties can be used to draw inferences over the events of interest.

Two common economic models are the Capital Asset Pricing Model (CAPM), introduced by Sharpe in 1964, and the Arbitrage Pricing Theory (APT), introduced by Ross in 1976. While intuitive, the CAPM have been shown to be sensitive to its restrictions (Fama and French, 1996) and APT have been shown to have few factors with significant explanatory power (Brown and Weinstein, 1985). To circumvent the restrictions of the economic models, we have taken a statistical approach and will use the market model.

4.1.4 Step 4 – The Estimation Window

The fourth step is defining the estimation window. The most common practice is setting this as a period of a given number of days prior to and past the event window. The length of this period varies, but 30, 100 and 365 days are common values. We have decided upon a window of +/- 100 days.

The estimation window will be the basis for calculating the parameters in equation (4) that are needed to estimate the abnormal returns in equation (6). To exclude biased parametric estimators, the event and event window itself should not be part of this estimation. The parameters in the market model are estimated using OLS.

It is worth noting that the estimation window and event window removes 100 + 2 days on each side of our data, leaving 1306 usable trading days for the regressions.

4.1.5 Step 5 – Designing the testing framework

This subsection presents our testing framework, where we argue that to properly answer all our hypotheses, the data must be analysed three ways: across time, across firms, and across time and firms simultaneously. We will use regressions as our main statistical tool to accomplish this, performing time-series regressions, cross-sectional regressions and panel regressions. In addition to model specification, this subsection also presents variable definitions. Model assumptions will be addressed in the next subsection.

The first two hypotheses, H1 and H2, which cover the correlation between the relative change in call to put option volume and abnormal returns, can be answered with any of the aforementioned regression types. However, regressions across time or firms alone will be limited to any one date or any one firm respectively, and will thus not include all available information. A panel data regression can be performed over the two dimensions simultaneously.

To answer H1 and H2, we have decided upon a two-way fixed effect transformation of the panel data. While a pooled OLS regression assumes that there are no individual or time specific effects[12], the fixed effect model assumes that there are time-invariant individual differences (Wooldridge, 2012). This again implies that every firm will have its own constant term, which is illustrated by the different subscripts on β_0 in equations (7) for pooled OLS and equation (8) for a fixed effect model.

$$Y_{j,t} = \beta_0 + \beta_1 \cdot X_{j,t} + \varepsilon_{j,t} \tag{7}$$

$$Y_{j,t} = \beta_{0,j} + \beta_1 \cdot X_{j,t} + \varepsilon_{j,t} \tag{8}$$

Where $\beta_{0,j} = \beta_0 + D_j$ for firms j in [1, K - 1], and $\beta_{0,K} = \beta_0$ for firm K. D_j is the individual firm constant measured relative to $\beta_{0,K}$ which can be considered the baseline firm[13]. This firm-specific constant also controls for any unobserved time-invariant effects, as they are all conditioned out by this term (Wooldridge, 2002).

[12] which implies a single constant term for the regression

[13] This kind of specification is necessary to control for multicollinearity

Averaging equation (8) over time gives equation (9)

$$\overline{Y}_j = \beta_{0,j} + \beta_1 \cdot \overline{X}_j + \overline{\varepsilon}_j \tag{9}$$

Where, (9) subtracted from (8) gives

$$Y_{j,t} - \overline{Y}_j = \beta_1 \cdot (X_{j,t} - \overline{X}_j) + \varepsilon_{j,t} - \overline{\varepsilon}_j \tag{10}$$

Now, the firm specific constant has disappeared, and the parameters can be estimated with pooled OLS using equation (10)[14] which adheres to the normal OLS assumptions.

$$\hat{Y}_{j,t} = \beta_1 \cdot \hat{X}_{j,t} + \hat{\varepsilon}_{j,t} \tag{11}$$

The two-way fixed effect transformation also includes dummy variables for each period, which conditions out any period effects.

As seen from equation (11), the fixed effect model transforms the data in such a way that one cannot estimate the coefficients for the constant and/or unobserved effect. There is another method for panel data transformation, *the random effect*, which only partially removes the constant and/or unobserved effect in its transformation. This makes the random effect model more efficient than fixed effects, and it can also estimate the parameters of these constant and/or unobserved effects. However, a random effect model has stricter assumptions than a fixed effect model, and assumes that any unobserved effects must be uncorrelated with the explanatory variables (Wooldridge, 2012). Although one can decide the better model with a Hausman-test for endogenity, we argue that this assumption is not plausible[15].

Our transformed two-way fixed effect regression equation is specified as

[14]Written in a collapsed form in equation(11)

[15]E.g. that changes in EPS are not correlated with an unobserved sector for cyclical companies

$$
\begin{aligned}
\widehat{AR}_{j,t} = \beta_1 \cdot \widehat{EPS}_{j,t} + \beta_2 \cdot \widehat{VO}_{j,t} + \beta_3 \cdot \widehat{MV}_{j,t} + \beta_4 \cdot \widehat{PTBV}_{j,t} + \beta_5 \cdot \widehat{Delta}_{j,t} \\
+ \sum_{n=1}^{2} \beta_{5+n} \cdot \widehat{PCrat}_{j,t-n} + \sum_{n=1}^{2} \beta_{7+n} \cdot \widehat{PCrat}^{*}_{j,t-n} \\
+ \sum_{n=1}^{2} \beta_{9+n} \cdot \widehat{PCrat}^{OTM}_{j,t-n} + \sum_{n=1}^{2} \beta_{11+n} \cdot \widehat{PCrat}^{OTM*}_{j,t-n} + \widehat{\varepsilon}_{j,t}
\end{aligned}
\tag{12}
$$

The third and fourth hypotheses, H3 and H4, which cover the informed trader's preferences regarding firm size and firm PTBV, requires time-series regressions to be conducted for each firm individually. Such a regression controls for each firm's average non-observed effect in the constant term, and makes it possible to compare P-values and regression coefficients with each firm's observed market value and PTBV. Our time-series regression is specified as

$$
\begin{aligned}
AR_t = \beta_0 + \beta_1 \cdot EPS_t + \beta_2 \cdot VO_t + \beta_3 \cdot MV_t + \beta_4 \cdot PTBV_t + \beta_5 \cdot Delta_t \\
+ \sum_{n=1}^{2} \beta_{5+n} \cdot PCrat_{t-n} + \sum_{n=1}^{2} \beta_{7+n} \cdot PCrat^{*}_{t-n} \\
+ \sum_{n=1}^{2} \beta_{9+n} \cdot PCrat^{OTM}_{t-n} + \sum_{n=1}^{2} \beta_{11+n} \cdot PCrat^{OTM*}_{t-n} + \varepsilon_t \\
\textit{for all } j
\end{aligned}
\tag{13}
$$

The fifth hypothesis, H5, which covers the magnitude of coefficients and significance over time, requires cross-sectional regressions to be conducted for each date. Such a regression controls for each date's average non-observed effect in the constant term, and makes it possible to see trends in regression coefficients and P-values. Our cross-sectional regression is specified in (14). Note that t in the subscripts behaves as a constant in this equation. Both the time-series and cross-sectional regressions are done with OLS.

$$
\begin{aligned}
AR_j = \beta_0 + \beta_1 \cdot EPS_j + \beta_2 \cdot VO_j + \beta_3 \cdot MV_j + \beta_4 \cdot PTBV_j + \beta_5 \cdot Delta_j \\
+ \sum_{n=1}^{2} \beta_{5+n} \cdot PCrat_{j,t-n} + \sum_{n=1}^{2} \beta_{7+n} \cdot PCrat^{*}_{j,t-n} \\
+ \sum_{n=1}^{2} \beta_{9+n} \cdot PCrat^{OTM}_{j,t-n} + \sum_{n=1}^{2} \beta_{11+n} \cdot PCrat^{OTM*}_{j,t-n} + \varepsilon_j \\
for\ all\ t
\end{aligned}
\tag{14}
$$

The variables used in all three regression specifications are defined on the following pages. While the specification in the definitions include subscripts for both firm, j, and period, t, the type of regression will decide whether one or both subscripts varies. The number of lags in some of the variables, noted n, is set to 2. This is because any effects from further lags did not contribute to the models' explanatory power.

The dependent variable:

- $AR_{j,t}$, Abnormal returns, are calculated as the difference between the stock's expected return given by the market model and the stock's actual return. This is our dependent variable in all our regressions.

$$AR_{j,t} = R^{Actual}_{j,t} - R^{MM}_{t} \tag{15}$$

The control variables are inspired by the work of Ge et al. (2016), Roll et al. (2010) and Johnson and So (2012), but are somewhat limited by data availability of earnings forecasts. As the dependent variable, $AR_{j,t}$, is a signed variable, all variables that are strictly positive must be normalized

- Normalized EPS, $EPS_{j,t}$.

$$EPS_{j,t} = \frac{Actual\ EPS_{j,t}}{\frac{1}{N} \cdot \sum_{i=1}^{N} EPS_{j,t}} - 1 \tag{16}$$

  Where N is the estimation window (i.e 200 observations) at time t.

- Normalized volume, $VO_{j,t}$.

$$VO_{j,t} = \frac{Actual\ VO_{j,t}}{\frac{1}{N} \cdot \sum_{i=1}^{N} VO_{j,t}} - 1 \tag{17}$$

  Where N is the estimation window (i.e 200 observations) at time t.

- Normalized MV, $MV_{j,t}$.

$$MV_{j,t} = \frac{Actual\ MV_{j,t}}{\frac{1}{N} \cdot \sum_{i=1}^{N} MV_{j,t}} - 1 \tag{18}$$

Where N is the estimation window (i.e 200 observations) at time t.

- Normalized PTBV, $PTBV_{j,t}$.

$$PTBV_{j,t} = \frac{Actual\ PTBV_{j,t}}{\frac{1}{N} \cdot \sum_{i=1}^{N} PTBV_{j,t}} - 1 \tag{19}$$

Where N is the estimation window (i.e 200 observations) at time t.

- *Delta* is the average delta for options traded at time t for company j. As the weighted sum of deltas for securities in a portfolio equals the delta of the portfolio, it can be calculated as follows.

$$Delta_{j,t} = \frac{\sum_{i=1}^{M}(Call_{i,j,t}^{Vol} \cdot Call_{i,j,t}^{Delta}) + \sum_{i=1}^{N}(Put_{i,j,t}^{Vol} \cdot Put_{i,j,t}^{Delta})}{\sum_{i=1}^{M} Call_{i,j,t}^{Vol} + \sum_{i=1}^{N} Put_{i,j,t}^{Vol}} \tag{20}$$

M and N is the total number of call and put options respectively, and *Delta* can take values from -1 to 1.

The explanatory variables related to our hypotheses:

- $Abn_{j,t}$, Abnormal day, as defined in equation (1). This variable will only be used in conjunction with others.

- Modified Put-Call ratio, $PCrat_{j,t-n}$. This is the first of our main explanatory variables. The variable is constructed as follows for:

$$PCrat_{j,t-n} = \frac{Call_{j,t}^{Vol} - Put_{j,t}^{Vol}}{Call_{j,t}^{Vol} + Put_{j,t}^{Vol}} \tag{21}$$

Where $\text{Call}_{j,t}^{Vol}$ is the aggregated daily call option volume for the company,

$$Call_{j,t}^{Vol} = \sum_{i=1}^{M} Call_{i,j,t}^{Vol} \text{ for each day, t, and each company, j} \tag{22}$$

and $\text{Put}_{j,t}^{Vol}$ is the aggregated daily put option volume.

$$Put_{j,t}^{Vol} = \sum_{i=1}^{N} Put_{i,j,t}^{Vol} \text{ for each day, t, and each company, j} \tag{23}$$

M and N is the total number of call and put options respectively, varying with combinations of time, t, and company, j. While the volumes are unsigned, the variable itself is signed taking a value somewhere from -1 to 1. This allows us to use signed abnormal returns instead of absolute abnormal returns. The n indicates the ratio for the n-th day prior to the day, t, of the regression.

- Modified Put-Call ratio*, $PCrat^*_{j,t-n}$. This is the second of our main explanatory variables. The variable is constructed by multiplying $PCrat_{j,t-n}$ with the binomial variable *Abnormal day*. If there are any additional information in the $PCrat_{j,t-n}$ prior to an *Abnormal day* event, we would expect this variable to be significant.

$$PCrat^*_{j,t-n} = PCrat_{j,t-n} \cdot Abn_{j,t} \tag{24}$$

  This kind of variable, a continuous one multiplied with a binomial one, affects the slope of the regression.

- Modified OTM Put-Call ratio, $PCrat^{OTM}_{j,t-n}$. The OTM moneyness criteria is defined as any option, put or call, that is *more* than 5 % out of the money.

$$PCrat^{OTM}_{j,t-n} = \frac{Call^{\text{OTM Vol}}_{j,t} - Put^{\text{OTM Vol}}_{j,t}}{Call^{\text{OTM Vol}}_{j,t} + Put^{\text{OTM Vol}}_{j,t}} \tag{25}$$

  Where $\text{Call}^{\text{OTM Vol}}_{j,t}$ is the aggregated daily OTM call option volume for company j at time t,

$$Call^{\text{OTM Vol}}_{j,t} = \sum_{i=1}^{M} Call^{\text{OTM Vol}}_{i,j,t} \text{ for each day, t, and each company, j} \tag{26}$$

  and $\text{Put}^{\text{OTM Vol}}_{j,t}$ is the aggregated daily OTM put option volume.

$$Put^{\text{OTM Vol}}_{j,t} = \sum_{i=1}^{N} Put^{\text{OTM Vol}}_{i,j,t} \text{ for each day, t, and each company, j} \tag{27}$$

- Modified OTM Put-Call ratio*, $PCrat^{OTM*}_{j,t-n}$.

$$PCrat^{OTM*}_{j,t-n} = PCrat^{OTM}_{j,t-n} \cdot Abn_{j,t} \tag{28}$$

Table 3 and table 4 summarizes all the independent variables by year.

Table 3: Independent variables 2009 - 2011

	2009				2010				2011			
	10 %	Average	90 %	No. obs	10 %	Average	90 %	No. obs	10 %	Average	90 %	No. obs
EPS	-1.0000	-0.1320	0.1020	30 813	-0.3586	-0.0574	0.1464	55 256	-0.1242	-0.0195	0.1049	56 701
Vol	-0.4663	-0.0636	0.3785	34 800	-0.4380	0.0028	0.5297	58 464	-0.4264	0.0163	0.5590	58 464
Delta	-0.1771	0.1299	0.4419	31 292	-0.1488	0.1265	0.4027	55 147	-0.1443	0.0977	0.3447	58 134
MV	-0.0857	0.0103	0.1062	34 800	-0.0964	-0.0066	0.0853	58 464	-0.1083	-0.0009	0.0949	58 464
PTBV	-0.0800	0.0304	0.1413	34 800	-0.1115	-0.0102	0.0924	58 464	-0.1168	0.0210	0.1001	58 464
PCrat_{t-1}	-0.5562	0.2085	0.9317	31 296	-0.4565	0.2145	0.8699	55 125	-0.4345	0.1625	0.7615	58 133
PCrat_{t-2}	-0.5556	0.2083	0.9302	31 303	-0.4572	0.2149	0.8713	55 099	-0.4348	0.1620	0.7606	58 132
PCrat^{*}_{t-1}	-0.3795	0.1380	0.8125	31 296	-0.2439	0.1268	0.7115	55 125	-0.2615	0.0974	0.6049	58 133
PCrat^{*}_{t-2}	-0.3857	0.1363	0.8071	31 303	-0.2414	0.1272	0.7098	55 099	-0.2634	0.0973	0.6033	58 132
PCrat^{OTM}_{t-1}	-0.9611	0.0263	1.0000	29 430	-0.9209	-0.0354	0.9022	52 899	-0.8372	-0.0372	0.7959	57 045
PCrat^{OTM}_{t-2}	-0.9604	0.0276	1.0000	29 444	-0.9214	-0.0353	0.9037	52 862	-0.8377	-0.0378	0.7954	57 048
$\text{PCrat}^{OTM*}_{t-1}$	-0.8066	0.0242	0.8926	29 430	-0.7323	-0.0131	0.6935	52 899	-0.6667	-0.0107	0.6232	57 045
$\text{PCrat}^{OTM*}_{t-2}$	-0.8049	0.0264	0.8939	29 444	-0.7345	-0.0120	0.6975	52 862	-0.6667	-0.0099	0.6251	57 048

The table presents descriptive statistics gathered from the 1^{st} of June 2009 to the 31^{st} of December 2011, for all independent variables in the regressions. For each year, data for the 10th percentile (10 %), average value, 90th percentile (90 %) and total number of observations is presented. The data is the accumulated data across all firms and trading days.

Table 4: Independent variables 2012 - 2014

	2012				2013				2014			
	10 %	Average	90 %	No. obs	10 %	Average	90 %	No. obs	10 %	Average	90 %	No. obs
EPS	-0.1243	-0.0218	0.0901	56 933	-0.1954	-0.0457	0.0891	57 230	-0.1316	-0.0164	0.1143	34 598
Vol	-0.4342	-0.0118	0.4726	58 000	-0.4399	-0.0132	0.4785	58 444	-0.4190	0.0303	0.5571	34 650
Delta	-0.1552	0.0970	0.3537	57 272	-0.1273	0.1274	0.3869	57 755	-0.1317	0.1292	0.3925	34 145
MV	-0.0839	-0.0008	0.0811	58 000	-0.0598	0.0021	0.0654	58 464	-0.0579	0.0039	0.0660	34 800
PTBV	-0.0890	0.0025	0.0970	58 000	-0.0847	-0.0151	0.0695	58 464	-0.0703	0.0040	0.0800	34 800
PCrat_{t-1}	-0.4603	0.1524	0.7714	57 273	-0.4118	0.1924	0.7774	57 754	-0.4000	0.2305	0.8208	34 147
PCrat_{t-2}	-0.4595	0.1532	0.7724	57 277	-0.4126	0.1918	0.7767	57 755	-0.3985	0.2310	0.8210	34 146
PCrat^{*}_{t-1}	-0.2575	0.0862	0.5906	57 273	-0.1837	0.1048	0.6016	57 754	-0.1525	0.1203	0.6556	34 147
PCrat^{*}_{t-2}	-0.2575	0.0861	0.5891	57 277	-0.1897	0.1035	0.6012	57 755	-0.1522	0.1196	0.6484	34 146
PCrat^{OTM}_{t-1}	-0.9308	-0.1664	0.7274	55 913	-0.9636	-0.2540	0.6134	56 340	-0.9713	-0.2356	0.6466	33 292
PCrat^{OTM}_{t-2}	-0.9301	-0.1653	0.7286	55 915	-0.9638	-0.2536	0.6137	56 336	-0.9710	-0.2363	0.6457	33 290
$\text{PCrat}^{OTM*}_{t-1}$	-0.7742	-0.0790	0.4902	55 913	-0.8133	-0.1233	0.3501	56 340	-0.8121	-0.1119	0.3792	33 292
$\text{PCrat}^{OTM*}_{t-2}$	-0.7695	-0.0755	0.4985	55 915	-0.8148	-0.1217	0.3539	56 336	-0.8140	-0.1122	0.3745	33 290

The table presents descriptive statistics gathered from the 1^{st} of January 2012 to the 6^{th} of August 2014, for all independent variables in the regressions. For each year, data for the 10th percentile (10 %), average value, 90th percentile (90 %) and total number of observations is presented. The data is the accumulated data across all firms and trading days.

4.2 Model assumptions

If the OLS techniques are going to be applicable, certain assumptions concerning the set of residuals needs to be fulfilled. The Gauss-Markov theorem stats that under certain conditions, OLS estimators are the "best linear unbiased estimators" (BLUE) (Dismuke and Lindrooth, 2006), and the assumptions for this are:

1. $E[\varepsilon_i] = 0$: Zero mean
2. $Var(\varepsilon_i) = \sigma^2 < \infty$: Homoscedastic, the variance for all observations X_i are constant and finite. Since the error variance is a measure of uncertainty, homoscedasticity causes the uncertainty in the model to be identical for all observations. Violating the assumption is a problem often found in cross-sectional data and leads to biased estimates of the variance of the coefficients, which means that standard error and inferences obtained are not valid.
3. $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$: No autocorrelation, the values of ε are not correlated. Violations of this assumptions is often caused by omitted variables, misspecification of the functional form or error of measurements in the dependent variable. In the presence of autocorrelation there will be complications in estimating the true variance which could cause H_0 to be rejected too often.
4. Cov(ε_i,X_i) = 0. I.e. ε_i is uncorrelated with any of the explanatory variables, X_i. If this assumption is violated, then OLS suffers from what is commonly referred to as an endogeneity problem. An endogeneity problem causes the coefficients in OLS to be biased.

When one or more of the OLS assumptions is violated, other models should be considered as better alternatives to OLS.

The first assumption, $E[\varepsilon_i] = 0$, is trivial. The regression constant will absorb any deviation from 0.

The second and third assumption, regarding homoscedasticity and autocorrelation, are not trivial. As our data have both a time and cross-sectional dimension, it is reasonable to assume that for one or more periods or one or more companies, there will be problems with either heteroscedasticity or autocorrelation. Another potential source of heteroscedasticity is the fact that the regression equations (12), (13) and (14) uses $\mathrm{AR}_{j,t}$ as the dependent variable. $\mathrm{AR}_{j,t}$ is calculated from equation (4). Since the dependent variable is calculated by regression parameter estimates, there is a portion of uncertainty in these numbers which introduces errors in the regression. These standard errors are adjusted by using heteroscedasticity consistent standard errors in the linear regression model, which give consistent results even when the presence of heteroscedasticity is of an unknown form (Long and Ervin, 2000). More precisely, we implement tests based on a heteroscedasticity consistent covariance matrix (HCCM) based on White (1980) work, using a version known as HC3 which MacKinnon and White (1985) proves perform better than its alternatives. The correction model is implemented in our regression software package and used directly from there.

For the assumption about endogeneity, we assume it holds.

Although it is not a necessity for OLS to be BLUE according to the Gauss-Markov theorem, residuals should be normally distributed to claim inference to the population. As our data is real life data, the return distributions tend to have thicker tails than a perfect normal distribution. This is also notable in the regression residuals when the explanatory power of the regression equation is low. Such normality problems can be addressed by using the method of Fenn and Liang (2001), which winsorizes the data at the 5th and 95th percentiles, or other similar methods. However, transforming observations this way would be detrimental to our research in this case, as it is exactly these observations that we are trying to explain with option volume. We have thus included all observations in our regressions. Regardless of the normality of ε_i, the normality assumption can by bypassed in large samples according to Greene (2003). When the sample n increases the normal distribution becomes a better approximation of the true regression parameters' distribution. We argue that our samples can be considered large enough for this to hold true.

4.3 Other issues

Previous research has explored option volumes by using data with different time intervals. Therefore, it is important to know the strength of the test given the sample intervals.

Morse (1984) compares daily and monthly data and concludes that decreasing the interval increases the power of the test. MacKinlay (1997) introduces weekly data as well, and draws the same conclusion. It is possible to have even shorter sample intervals then one day, down to single transaction data, but the net benefit of using such data is unclear. We argue that our daily data has enough resolution to capture any effects that the $\text{PCrat}_{j,t-n}$ and any variations thereof may exhibit.

Event dates are not always exact, which leads to uncertainty related to the event window. For example, when collecting news from publications such as the Wall Street Journal, it is possible that the market was informed prior to the news release from the paper. As discussed, this uncertainty can be mitigated by increasing the event window. An event window consisting of two days, day 0 and day +1, is a method often used. MacKinlay (1997) shows that the estimates with a larger event window, set at two days, is still good. If there is uncertainty related to the event of interest, it is therefore better to increase the event window instead of taking the risk of missing the event.

4.4 Data Presentation

For our time-series and cross-sectional regressions we present our regression outputs in tables which summarises the proportion of significance of the variables at 5 % and 10 % together with the regression coefficient medians, 25^{th} and 75^{th} percentiles inspired by the presentation in Roll et al. (2010). We argue that this is a suitable way to present the data in two ways.

The P-value is the probability that, given a true null hypothesis, one observes what one observes (or something more extreme) at random. This implies, for a significance level of e.g. 5 %, that one should expect a proportion of P-values of 0.05 or lower in 5 % of the regressions should the null hypothesis hold true. However, if the reported proportion is significantly higher than 5 %, this would imply that the observed effects are not random (although they might not be significant in every single regression).

While the above argument gives an intuitive explanation, the combined p-values have been tested formally with Fisher's combined probability test (Fisher, 1932). The test statistic, is chi-squared distributed with $2 \cdot k$ degrees of freedom, where k is the number of tests.

$$-2\sum_{i=1}^{k} ln(p_i) \sim \chi^2 \tag{29}$$

The results of the test for each regression coefficient is also reported in the tables. As the test assumes independence, a strong assumption, the combined P-values have also been adjusted according to the mean-method presented in Becker (1994). The mean-method has the following test statistic:

$$\sqrt{12 \cdot k} \cdot (0.5 - \sum_{i=1}^{k} p_i) \sim Z \tag{30}$$

The null hypothesis for both of these tests is defined as *the effect of interest is not significant in any of the population studied* by Becker.

5 Empirical Analysis

5.1 Panel Data

The following sections presents a summary of the panel data regression from equation (12). The regression consists of 232 companies observed over 1306 days from the 1^{st} of June 2009 to the 6^{th} of August 2014. The analysis aspires to find evidence of informed trading prior to spikes in abnormal returns, thus answering H1 and H2. The regression outputs are summarized in separate tables for $Abn_{j,t}$ defined with limits of $x > 0.5\%$ and $x > 1.0\%$[16][17]. The table for $Abn_{j,t}$ defined with a limit of $x > 1.5\%$ can be found in the appendix. OTM is, as earlier, defined as all options that are more than 5 % out-of-the-money.

To answer H1 and H2, the important variables are PCrat^*_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$. These variables will indicate whether or not there are any additional explanatory power beyond the day-to-day variations in PCrat_{t-n} and PCrat^{OTM}_{t-n}. If this is the case, the coefficients of PCrat^*_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$ should be statistically significant.

While the hypothesis can be answered on a statistical basis alone, from an empirical finance perspective it is also interesting to answer them in the sense of economic signif-

[16]See equation (1)

[17]The notation $Abn > x$ will from now on be used to describe the limit for which $\text{Abn}_{j,t}$ is defined.

icance. This is done by analyzing the coefficients, which contain information about the impact our variables have regarding both direction and magnitude. This is also somewhat a robustness check, as results with coefficients so high or low that they do not make economic sense can be disregarded.

5.1.1 Abn > 0.5 %

Table 5: Two-way fixed-effect panel regression : Abn > 0.005

		90 % CI		95 % CI			
	Estimate	Lower	Upper	Lower	Upper	T-value	P-value
EPS	-0.0001	-0.0003	0.0001	-0.0004	0.0001	-0.84	0.4028
Vol	-0.0005	-0.0009	-0.0002	-0.0010	-0.0001	-2.34	0.0192
Delta	0.0107	0.0099	0.0115	0.0098	0.0116	23.02	0.0000
MV	0.0184	0.0153	0.0215	0.0147	0.0221	9.81	0.0000
PTBV	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-8.19	0.0000
PCrat_{t-1}	-0.0011	-0.0012	-0.0009	-0.0012	-0.0009	-15.49	0.0000
PCrat_{t-2}	-0.0008	-0.0009	-0.0007	-0.0009	-0.0007	-14.35	0.0000
PCrat^*_{t-1}	0.0002	0.0000	0.0004	-0.0000	0.0004	1.76	0.0782
PCrat^*_{t-2}	0.0000	-0.0002	0.0002	-0.0002	0.0002	0.20	0.8431
PCrat^{OTM}_{t-1}	0.0004	0.0003	0.0004	0.0003	0.0004	8.91	0.0000
PCrat^{OTM}_{t-2}	0.0003	0.0002	0.0004	0.0002	0.0004	7.52	0.0000
$\text{PCrat}^{OTM*}_{t-1}$	0.0002	0.0001	0.0004	0.0001	0.0004	2.62	0.0088
$\text{PCrat}^{OTM*}_{t-2}$	-0.0001	-0.0002	0.0001	-0.0002	0.0001	-0.86	0.3924

Adj. R^2	F-statistic	P-value	DF
0.0305	752.30	0.0000	260431

The table presents an overview of the regression output for Abn > 0.5 %. The values are calculated from a panel data regression consisting of 232 companies and 1306 trading days observed from mid-2009 to mid-2014. The regression coefficients estimates are presented as point estimates, a 90 % confidence interval and a 95 % confidence interval. The table also presents the T-value and P-value for the individual variables, and the *Adjusted* R^2, F-statistic, P-value and the degrees of freedom (DF) for the total regression.

From table 5 we observe that both PCrat_{t-1} and PCrat_{t-2} are highly significant with P-values = 0.00 %. The significance of the PCrat_{t-n} variables are in line with the results from Roll, Schwartz, and Subrahmanyam (2010) that option volume is correlated with future abnormal returns. Our hypothesis, H1, is however more closely related to PCrat^*_{t-1} and PCrat^*_{t-2}. We observe that PCrat^*_{t-1} is significant at a 10 %-level, but not at a 5 %-level with a P-value of 7.82 %. PCrat^*_{t-2} is, on the other hand, not significant at 10

%-level with a P-value of 84.31 %. Thus the null hypothesis is rejected for $\alpha = 10\%$[18], but not at $\alpha = 5\%$, and we claim that there is evidence of a relative change in call to put option volume prior to large spikes in abnormal returns on the underlying asset.

Regarding H2, we see that PCrat_{t-1}^{OTM} and PCrat_{t-2}^{OTM} are significant with P-values = 0.00 %. This is in line with the work of Chakravarty, Gulen, and Mayhew (2004) who finds that OTM options have a higher information share than ATM and ITM options. We also observe supporting results towards H2. $\text{PCrat}_{t-1}^{OTM*}$ is significant at a 5 %-level, with a P-value of 0.88 %. As with PCrat_{t-2}^{*}, the second lag is not significant with a P-value of 39.24 %. Given this result we reject the null hypothesis at $\alpha = 5\%$, and claim that there is evidence of a relative change in the isolated out-of-the money call to put option volume prior to large spikes in abnormal returns on the underlying asset.

The regression coefficients give estimates of the economic impact and magnitude of changes in the dependent variable when the explanatory variables changes. E.g. for a change of 1 in PCrat_{t-1}, the abnormal return is expected to decrease with 0.1 %. The negative sign of the PCrat_{t-n} variables implies that a share of put options > 0.5 in the two previous days is associated with a positive return at $t = 0$. The positive signs of PCrat_{t-n}^{*}, PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-1}^{OTM*}$ mean that a share of call options > 0.5 in the days prior to $t = 0$ is associated with a negative return.

The PCrat_{t-n} variables are the largest with regard to magnitude, with absolute values of 0.0011 for $t-1$ and 0.0008 for $t-2$ respectively. The six other variables have absolute values ≤ 0.0004. The magnitude size of all the regression coefficients of interest can be considered small compared to the estimates for *Delta* and *MV*, being one tenth of the size. Even tough the null hypotheses of H1 and H2 can be rejected statistically, we believe the economic impact of the variables is too small to make profitable trading strategies when considering transaction costs.

As for the total variation in abnormal return captured by the regression, measured by *Adjusted* R^2, the value of 3.02 % is low. This may be due to our choice of model for abnormal returns or unobserved properties or events that are not included as variables.

[18] α defines the significance level

5.1.2 Abn > 1.0 %

Table 6: Two-way fixed-effect panel regression : Abn > 0.01

	Estimate	90 % CI Lower	90 % CI Upper	95 % CI Lower	95 % CI Upper	T-value	P-value
EPS	-0.0001	-0.0003	0.0001	-0.0004	0.0001	-0.81	0.4208
Vol	-0.0006	-0.0009	-0.0002	-0.0010	-0.0001	-2.42	0.0154
Delta	0.0107	0.0099	0.0115	0.0098	0.0116	23.03	0.0000
MV	0.0184	0.0153	0.0215	0.0147	0.0221	9.81	0.0000
PTBV	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-8.32	0.0000
$PCrat_{t-1}$	-0.0011	-0.0012	-0.0010	-0.0013	-0.0010	-16.66	0.0000
$PCrat_{t-2}$	-0.0008	-0.0009	-0.0007	-0.0010	-0.0007	-14.43	0.0000
$PCrat^*_{t-1}$	0.0006	0.0003	0.0010	0.0002	0.0010	3.11	0.0019
$PCrat^*_{t-2}$	0.0002	-0.0002	0.0005	-0.0002	0.0005	0.76	0.4464
$PCrat^{OTM}_{t-1}$	0.0004	0.0003	0.0005	0.0003	0.0005	9.62	0.0000
$PCrat^{OTM}_{t-2}$	0.0003	0.0002	0.0003	0.0002	0.0004	7.20	0.0000
$PCrat^{OTM*}_{t-1}$	0.0003	0.0000	0.0005	-0.0000	0.0006	1.69	0.0902
$PCrat^{OTM*}_{t-2}$	-0.0001	-0.0003	0.0002	-0.0004	0.0002	-0.42	0.6716

Adj. R^2	F-statistic	P-value	DF
0.0306	755.48	0.0000	260431

The table presents an overview of the regression output for Abn > 1.0 %. The values are calculated from a panel data regression consisting of 232 companies and 1306 trading days observed from mid-2009 to mid-2014. The regression coefficients estimates are presented as point estimates, a 90 % confidence interval and a 95 % confidence interval. The table also presents the T-value and P-value for the individual variables, and the *Adjusted* R^2, F-statistic, P-value and the degrees of freedom (DF) for the total regression.

From table 6, when Abn > 1.0 %, we observe many of the same results as in the previous section. Both $PCrat_{t-1}$ and $PCrat_{t-2}$ are highly significant with P-values = 0.00 %, and similar results are seen for $PCrat^{OTM}_{t-1}$ and $PCrat^{OTM}_{t-2}$. Regarding H1, there is however one main difference. We observe that $PCrat^*_{t-1}$ is significant at a 5 %-level, with a P-value = 0.19 %. Compared to the P-value of 7.82 %, when Abn > 0.5 %, this is a notable reduction. $PCrat^*_{t-2}$ is, on the other hand, still not significant at 10 %-level with a P-value of 44.64 %. However, with at least one significant variable we can reject the null hypothesis for $\alpha = 5$ %, and we claim that there is evidence of a relative change in call to put option volume prior to large spikes in abnormal returns on the underlying asset.

We also observe supporting results towards H2, though not as strong as when Abn > 0.5 %. The significance of $PCrat^{OTM*}_{t-1}$ has decreased and it is now only significant at a 10 %-level, with a P-value of 9.02 %. The second lag, $PCrat^{OTM*}_{t-2}$, is not significant with a

P-value of 67.16 %. These results imply a weaker rejection of the null hypothesis, which can now only be rejected at $\alpha = 10\%$. Although a weaker claim, there is still evidence of a relative change in the isolated out-of-the money call to put option volume prior to large spikes in abnormal returns on the underlying asset.

The regression coefficients estimates are almost unchanged compared to Abn > 0.5 %. The most notable change is that the coefficient for $PCrat^{*}_{t-1}$ has tripled from 0.0002 to 0.0006. The magnitude of this coefficient is still lower than those of PCrat_{t-n}, and we still believe the low magnitude limits any profitable trading strategies due to transaction costs.

5.2 Time Series

The following section summarizes the time-series regressions for all 232 companies, with a total of 1306 days observed from the 1^{st} of June 2009 to the 6^{th} of August 2014. The analysis aspires to find supporting evidence towards H1 and H2, but is conducted mainly to answer H3 and H4. I.e. we will not conclude H1 and H2 with these regressions, but rather enhance the finding from the panel data analysis on a firm-by-firm basis. By running a regression for each company j, we are able to map each regression's coefficients and P-values to the company's average market value and PTBV over the period and present the relationships graphically.

The results substantiating H1 and H2 are summarized in tables, while results substantiating H3 and H4 are summarized in figures, both separated in sections for Abn > 0.5 % and Abn > 1.0 %. The tables and figures for Abn > 1.5 % can be found in the appendix.

5.2.1 Abn > 0.5 %

Table 7 shows a summary of the P-values, a combined P-value, and the regression coefficient estimates for the explanatory variables when Abn > 0.5 %. The table is important for two reasons. First, the table presents the proportions of variables that are significant at a 5 % and 10 %-level. Following the P-value argument from section 4.4, a positive deviation between observed proportion of P-values and significance level indicates a higher proportion of significant regression coefficients than what one would expect at

Table 7: Time-series regression P-values and coefficients : Abn > 0.005

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.2251	0.3117	0.0000	0.0000	-0.0011	-0.0006	-0.0001
EPS	0.0866	0.1082	0.2982	0.7702	-0.0019	0.0002	0.0021
Vol	0.0996	0.1818	0.0000	0.0123	-0.0015	0.0000	0.0016
Delta	0.9697	0.9697	0.0000	0.0000	0.0061	0.0099	0.0158
MV	0.1732	0.2597	0.0000	0.0000	0.0004	0.0110	0.0207
PTBV	0.0649	0.1299	0.0403	0.1883	-0.0031	0.0018	0.0124
PCrat_{t-1}	0.4719	0.5844	0.0000	0.0000	-0.0020	-0.0012	-0.0006
PCrat_{t-2}	0.2771	0.3853	0.0000	0.0000	-0.0016	-0.0009	-0.0004
PCrat^*_{t-1}	0.0606	0.1082	0.1432	0.4471	-0.0012	-0.0000	0.0012
PCrat^*_{t-2}	0.0390	0.0823	0.9383	0.9810	-0.0010	0.0000	0.0010
PCrat^{OTM}_{t-1}	0.1861	0.2381	0.0000	0.0000	0.0001	0.0004	0.0009
PCrat^{OTM}_{t-2}	0.1472	0.2900	0.0000	0.0000	0.0001	0.0004	0.0008
$\text{PCrat}^{OTM*}_{t-1}$	0.0433	0.0736	0.3884	0.1939	-0.0006	0.0004	0.0013
$\text{PCrat}^{OTM*}_{t-2}$	0.0519	0.1039	0.6601	0.6923	-0.0009	-0.0001	0.0009

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 0.5 %. The values are calculated from 232 different time-series regressions, with a total of 1306 trading days observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (232). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

random. Second, the coefficients create the baseline for any economic interpretation of the variables.

Regarding H1, we see that PCrat_{t-1} and PCrat_{t-2} have a 47.19 % and 27.71 % portion of significance at a 5 %-level. The lags exhibit a falling trend, indicating that the day closest to a spike in abnormal return is the most significant variable. This is the same as observed in the panel data regression in table 5. The same falling trends are also observed for PCrat^*_{t-1} and PCrat^*_{t-2}. However, the portions 6.06 % and 3.90 % is no larger than what we could expect from a random sample. Our test for combined P-values concludes that PCrat^*_{t-1} and PCrat^*_{t-2} are non-significant at conventional significance levels. This is not in line with the panel data conclusion, which rejected the null hypothesis for H1 at $\alpha = 10\%$.

While the day-to-day OTM option volume, PCrat^{OTM}_{t-1} and PCrat^{OTM}_{t-2}, have a high proportion of significant coefficients, with proportions of 18.61 % and 14.72 % at a 5 %-level

Table 8: Time-series regression explanatory power and observations : Abn > 0.005

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	-0.0011	0.0297	0.0454	0.0622	0.1734
No. obs	135	1024	1246	1303	1306
No. abn. obs	64	580	677	783	999

The table presents summarizing statistical properties when Abn > 0.5 % for a total of 232 different time-series regressions with a total of 1306 trading days observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 0.5 %.

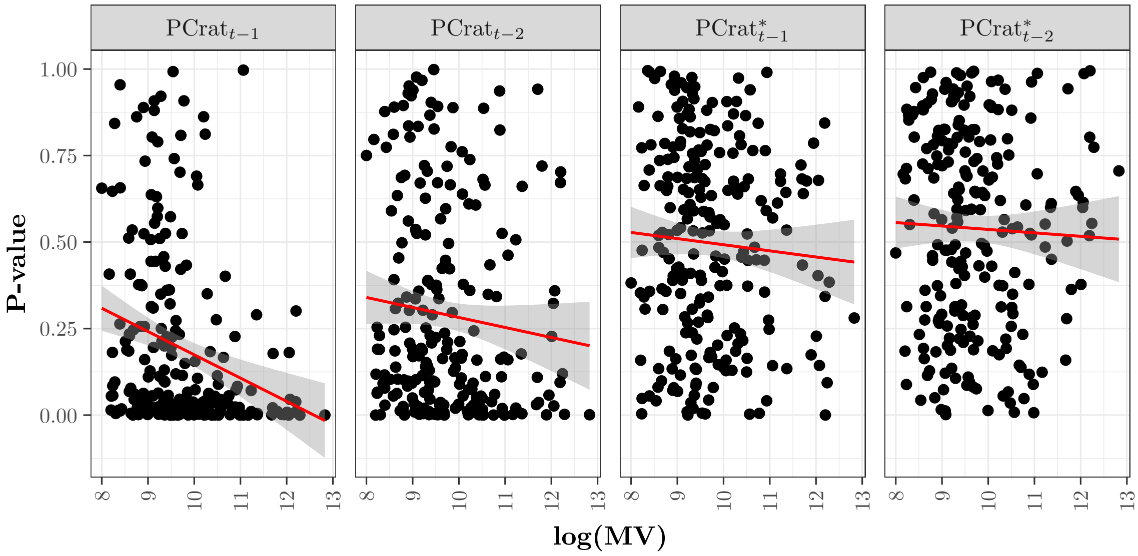
respectfully, the OTM option volume prior to spikes in abnormal return have portions of significance no larger than what we expect from a random sample. The same results are found when comparing at a 10 %-level. The combined P-values substantiates that only the day-to-day volume is significant, both for $\alpha = 5\%$ and $\alpha = 10\%$.

What is noticeable in table 7 is the shifting signs the coefficients have. While PCrat_{t-1}, PCrat_{t-2}, PCrat^{*}_{t-1} and $\text{PCrat}^{OTM*}_{t-2}$ have positive signs, PCrat^{OTM}_{t-1}, PCrat^{OTM}_{t-2}, $\text{PCrat}^{OTM*}_{t-1}$ and PCrat^{*}_{t-2} have negative signs. As the volume used in our analysis is unsigned we cannot draw an unambiguous conclusion of whether call/put options bought/sold causes the abnormal return to increase or decrease, but we can conclude that the majority of PCrat_{t-n} and PCrat^{OTM}_{t-n} pull in different directions. The variable with the largest impact is PCrat_{t-1}, with a median effect of about -0.12 % for an increase in PCrat_{t-1} by 1, while the variable with the lowest impact is PCrat^{*}_{t-2} with a coefficient of zero. It is also worth noting that at least 75 % of the observed coefficients for PCrat_{t-1} and PCrat_{t-2} are negative, and at least 75 % of the coefficients for PCrat^{OTM}_{t-1} and PCrat^{OTM}_{t-2} are positive.

These observations should be viewed alongside the informational content of table 8. The regressions explain at the median 4.54 % of the total variation, and between 2.97 % and 6.22 % in the 25th percentile to 75th percentile interval measured by *Adjusted* R^2. This is in line with the explanatory power of the regressions from Ge, Lin, and Pearson (2016) which also utilizes a form of option ratio as explanatory variables. The table also show that the median number of trading days with abnormal returns > 0.5 % is 677 out of 1246 observations. Since the total number of special observations is 54.33 % of total observations, the special trading days might not be that special. An abnormal return of 0.5 % will occur often, and is most likely caused by random fluctuations, not specific news or information affecting the stock price.

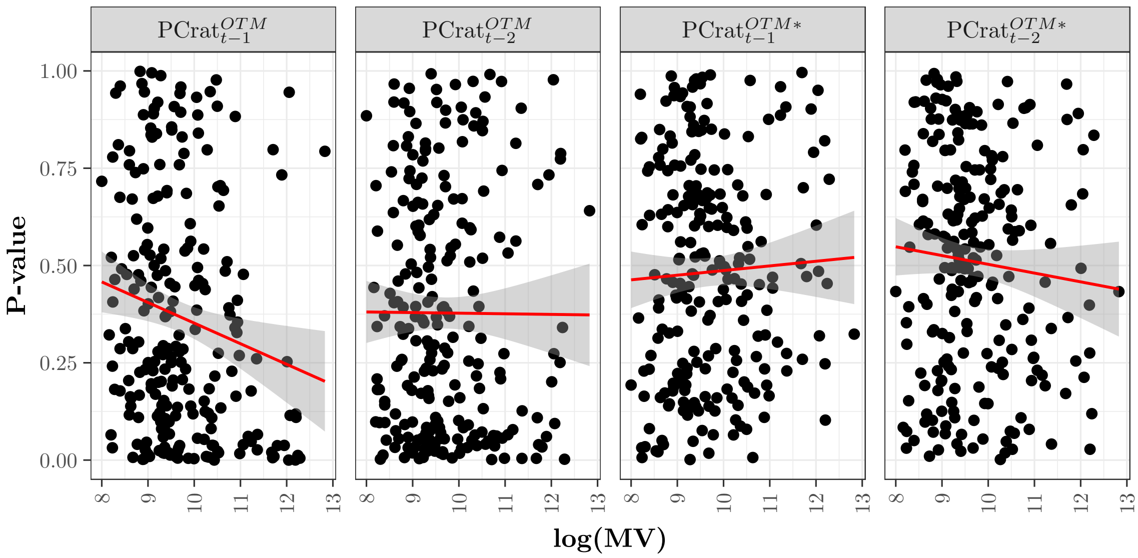
Following are eight different figures showing the variables plotted against market value and PTBV. The interpretation of these figures, i.e. trends and patterns, will conclude our third and fourth hypothesis, H3 and H4.

Figure 1: Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs log(MV): Abn > 0.005



The figure shows a scatter plot of the P-values and the logarithm of market value for the variables PCrat_{t-n} and PCrat^{*}_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 0.5 %.

Figure 2: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.005

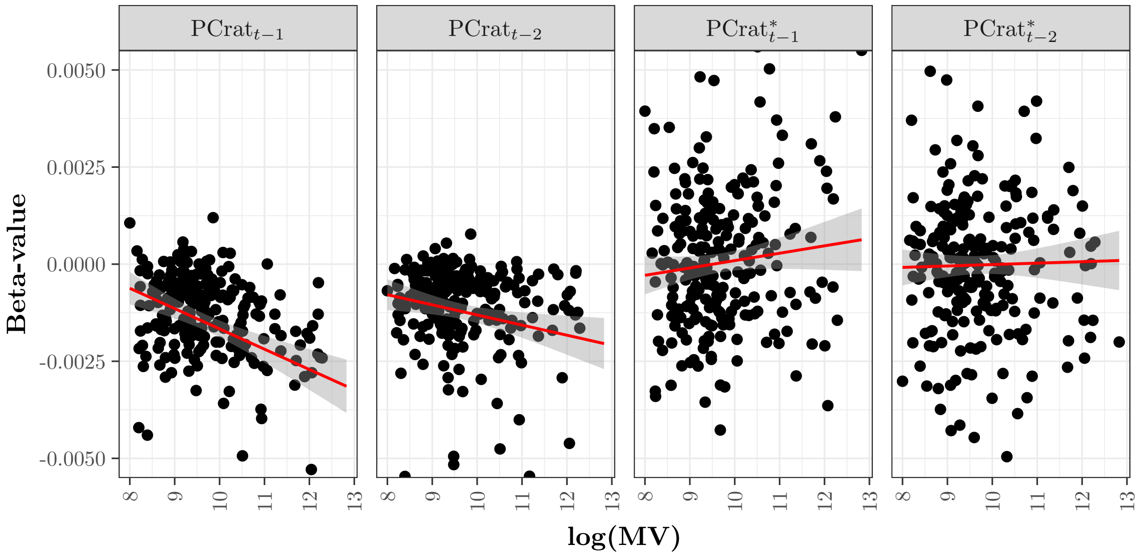


The figure shows a scatter plot of OTM observations for the P-values and the logarithm of market value for the variable PCrat^{OTM}_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 0.5 %. .

In figure 1 and 2 P-values and the logarithm of market value is compared to establish any trends between the two.

Both PCrat_{t-1} and PCrat^{OTM}_{t-1} show significant linear downward trends, indicating that day-to-day option volume for all options and OTM are correlated with firm size for a one day lag. However, the PCrat^{*}_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$ variables do not show any significant trends, as the confidence interval of the slope coefficient of the linearly fitted line contains zero[19]. Therefore, rejection of $H3_0$ cannot be justified for Abn $>$ 0.5 %; *market value does not affect an informed trader's decision.*

Figure 3: Time-series $\text{PCrat}^{(*)}_{t-n}$ regression coefficients vs log(MV): Abn > 0.005



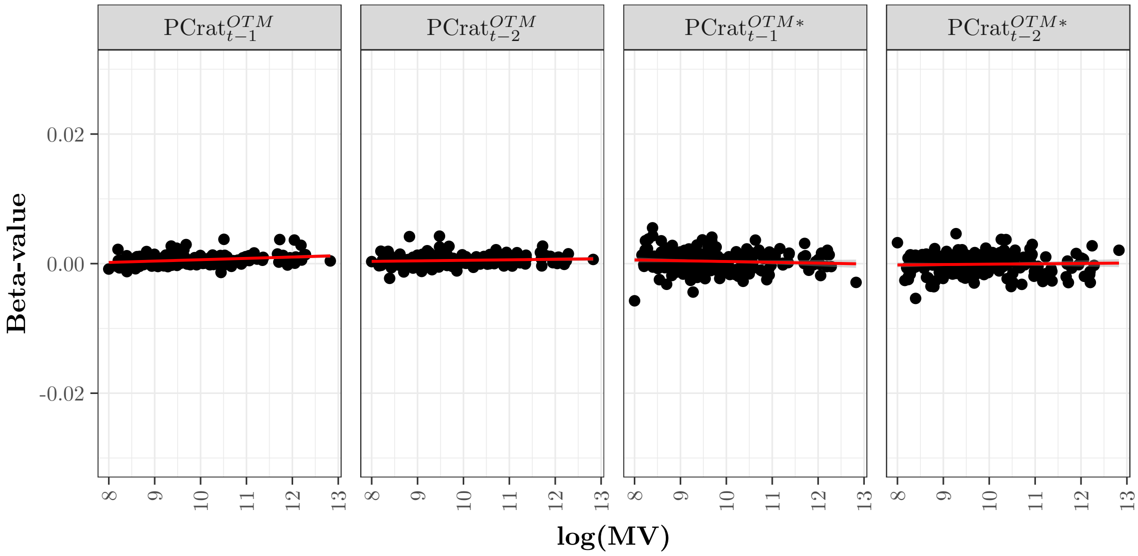
The figure shows a scatter plot of the regression coefficient (β) estimates and the logarithm of market value for the variables PCrat_{t-n} and PCrat^{*}_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn $>$ 0.5 %.

As concluded in the previous paragraph, market value does not affect the portion of informed trading, which is further supported by analyzing the regression coefficients[20] and company market value. From figure 3 we see that for the PCrat_{t-n} variables the trend is significant downward sloping, indicating that the regression coefficients are higher in absolute values for larger firms. For PCrat^{*}_{t-n} however, we cannot conclude that there is a relationship between firm size and the regression coefficient prior to large spikes in abnormal return. The trend line for OTM observations in figure 4 is however

[19]Seen from the grey confidence bands in the figures.

[20]The regression coefficients are referred to as Beta-values in the figures

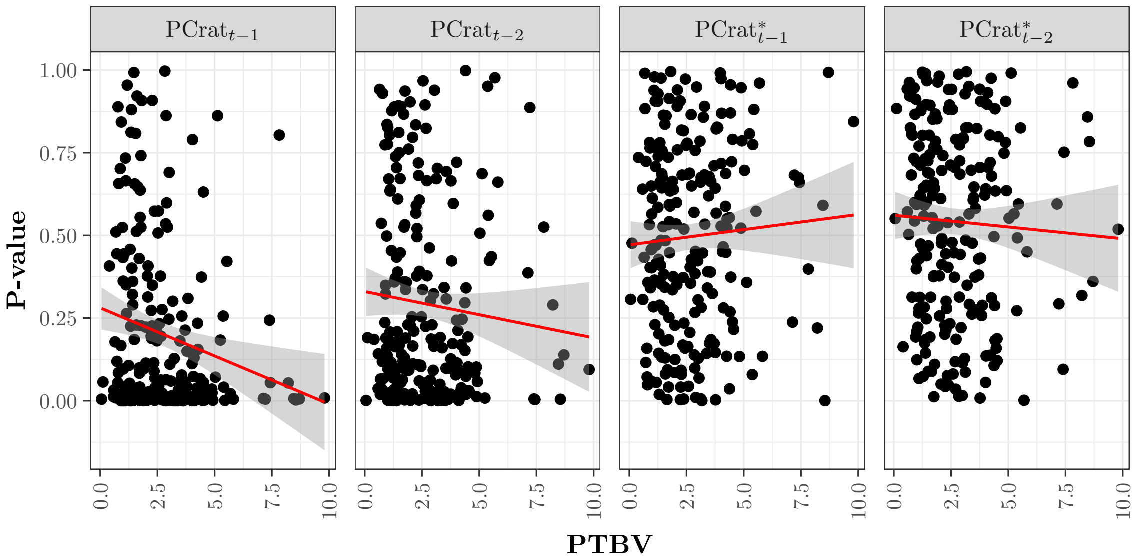
Figure 4: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.005



The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and the logarithm of market value for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 0.5 %.

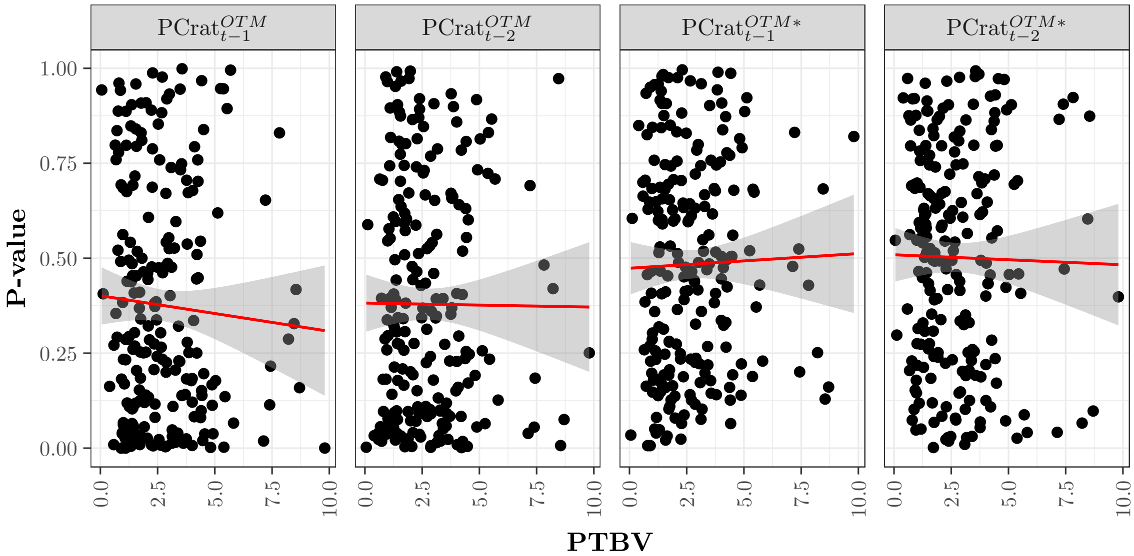
unambiguous, showing a clear flat trend. Market value does not affect the regression coefficients for OTM observations.

Figure 5: Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.005



The figure shows a scatter plot of the P-values and PTBV for the variables PCrat_{t-n} and PCrat_{t-n}^{*}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 0.5 %.

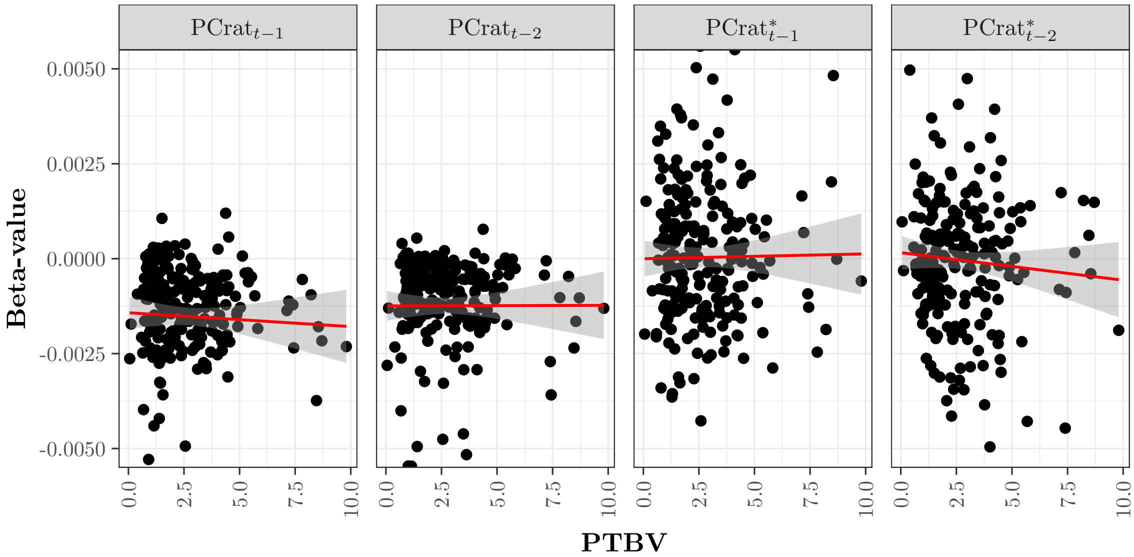
Figure 6: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.005



The figure shows a scatter plot of OTM observations for the P-values and PTBV for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 0.5 %.

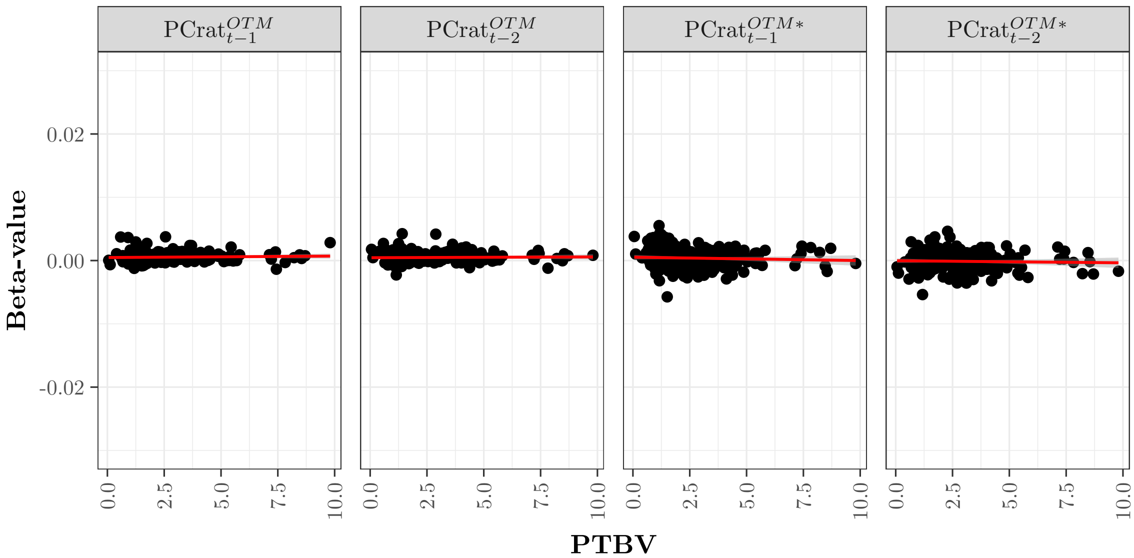
From figure 5 and 6 we find no results that gives reason to reject $H4_0$: *PTBV does not affect an informed traders decision, informed traders are not more likely to trade in firms with lower PTBV*. The only variable that is significant downward sloping is PCrat_{t-1}. With no significant trends in PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$, we fail to reject the null hypothesis. This could be because informed traders do not possess information for multiple companies, and are limited to trade in the ones they are informed about, regardless of PTBV.

Figure 7: Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs PTBV: Abn > 0.005



The figure shows a scatter plot of the regression coefficient (β) estimates and PTBV for the variables PCrat_{t-n} and PCrat_{t-n}^{*}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 0.5 %.

Figure 8: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs PTBV: Abn > 0.005



The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and PTBV for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 0.5 %.

There are no significant trends between regression coefficients and PTBV for neither combination of PCrat, as seen in figure 7 and 8. These results further support our conclusion about H4, and there are still no observed connection between informed trading and PTBV.

5.2.2 Abn > 1.0 %

Table 9: Time-series regression P-values and coefficients : Abn > 0.01

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.2251	0.3030	0.0000	0.0000	-0.0010	-0.0005	-0.0001
EPS	0.0823	0.1126	0.3439	0.8284	-0.0017	0.0001	0.0021
Vol	0.0909	0.1688	0.0000	0.0079	-0.0016	0.0001	0.0016
Delta	0.9654	0.9697	0.0000	0.0000	0.0060	0.0096	0.0159
MV	0.1645	0.2554	0.0000	0.0000	-0.0001	0.0110	0.0201
PTBV	0.0693	0.1212	0.0682	0.2954	-0.0032	0.0018	0.0121
PCrat_{t-1}	0.5022	0.5671	0.0000	0.0000	-0.0021	-0.0013	-0.0007
PCrat_{t-2}	0.2857	0.4329	0.0000	0.0000	-0.0015	-0.0009	-0.0004
PCrat^*_{t-1}	0.0823	0.1255	0.1613	0.3904	-0.0016	0.0001	0.0028
PCrat^*_{t-2}	0.0563	0.1169	0.7974	0.8226	-0.0019	0.0001	0.0020
PCrat^{OTM}_{t-1}	0.2078	0.2944	0.0000	0.0000	0.0001	0.0005	0.0009
PCrat^{OTM}_{t-2}	0.1342	0.2121	0.0000	0.0000	0.0000	0.0003	0.0008
$\text{PCrat}^{OTM*}_{t-1}$	0.0606	0.1082	0.1294	0.0687	-0.0015	0.0005	0.0023
$\text{PCrat}^{OTM*}_{t-2}$	0.0346	0.1082	0.4652	0.4312	-0.0019	-0.0001	0.0015

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 1.0 %. The values are calculated from 232 different time-series regressions, with a total of 1306 trading days observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (232). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

While performing the same analysis and answering the same hypotheses as in the previous section, the threshold for abnormal returns when defining Abn has now increased from 0.5 % to 1.0 %. From table 10, the median number of trading days with an abnormal return > 1.0 % has decreased to 341 out of 1246 trading days (27.37 %).

Comparing table 9 with table 7, PCrat_{t-1}, PCrat_{t-2}, PCrat^*_{t-1} and PCrat^*_{t-2} all increase their proportions of significance at a 5 % level with approximately 2 percentage points each. The diminishing proportions of significance for increased lags are in line with the results from the panel data regression and the time-series regressions for Abn > 5 %. The combined P-values of zero for the first two variables substantiates their significance. The variable PCrat^*_{t-1} has a proportion of significance at a 5 %-level of 8.23 % and at a 10 %-level of 12.55 %. Both these numbers are higher than what we would expect from random observations. PCrat^*_{t-2} exhibit the same trend, with proportions of 5.63 % and

Table 10: Time-series regression explanatory power and observations : Abn > 0.01

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	0.0050	0.0362	0.0500	0.0664	0.1870
No. obs	135	1024	1246	1303	1306
No. abn. obs	27	244	341	440	741

The table presents summarizing statistical properties when Abn > 1.0 % for a total of 232 different time-series regressions with a total of 1306 trading days observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 1.0 %.

11.69 %. The combined P-value do however conclude that both PCrat^*_{t-1} and PCrat^*_{t-2} are non-significant.

For $PCrat^{OTM}_{t-n}$, both lags are significant when testing combined P-values, but with noticeable smaller proportions of significant obervations compared to $PCrat_{t-n}$. For the modified variables, $\text{PCrat}^{OTM*}_{t-1}$ is now significant at a 10 % level according to the adjusted mean procedure, and almost significant at a 10 % level using Fisher's method. This is the same as what was observed using the panel regression. $\text{PCrat}^{OTM*}_{t-2}$ exhibits no systematic changes from Abn > 5 %, and its combination of P-values are non-significant.

The regression coefficients in 9 show many similarities with those in table 7. The variable with the most impact (at the median) is still PCrat_{t-1}. The most notable change in the coefficients is the spread of PCrat^*_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$. The length of the interval from the 25th percentile to the 75th percentile has almost doubled. Some of this increased uncertainty is caused by fewer observations for Abn > 1.0 % than for Abn > 0.5 %, as the second set will, by definition, be at least as large as the first one. On the other hand, it is also economically sensible that the coefficients will have a larger spread and impact when the set of abnormal days have a larger distance between the smallest positive and negative value used in the absolute function which defines the variable.

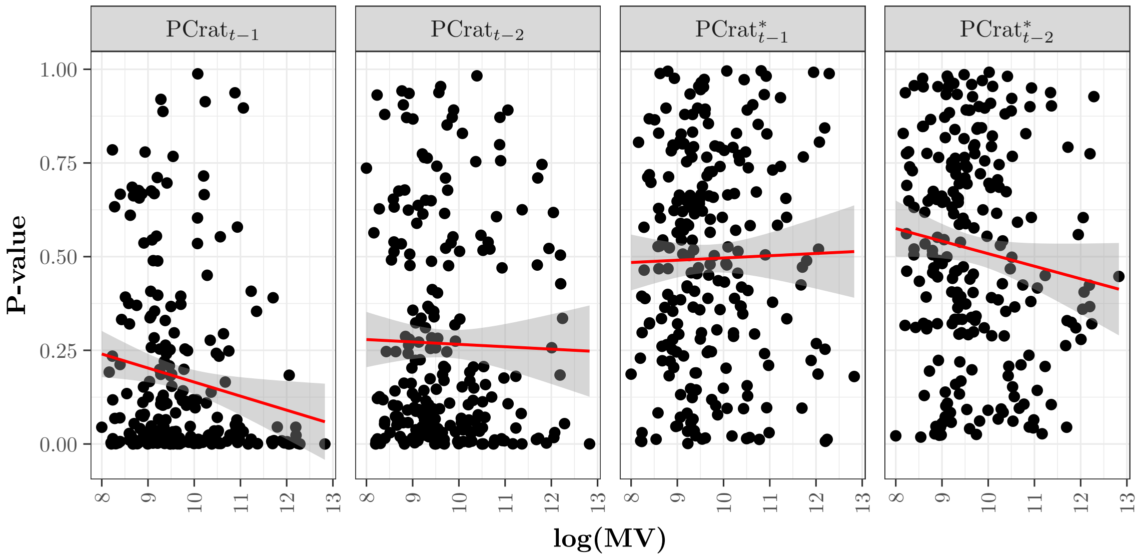
We observe that all PCrat variations show signs of more significance, implying that the model is a better fit for the dependent variable. This is further supported by table 10, which show that our analysis explains at the median 5.00 % of the total variation, and between 3.62% and 6.64 % in the 25th percentile to 75th percentile interval measured by *Adjusted* R^2, an increase from the previous section of about 0.5 percentage points.

The same eighth figures analyzed for Abn > 0.5 %, will now be analyzed for Abn > 1.0

%. The interpretation of these figures, i.e. trends and patterns, will conclude our third and fourth hypothesis, H3 and H4.

In figure 9 and 10 we observe that PCrat$_{t-1}$, PCrat$^{OTM}_{t-1}$ and PCrat$^{OTM}_{t-2}$ show significant downward trends. These trends, except for PCrat$^{OTM}_{t-2}$, are in line with what was observed for Abn > 0.5 %. However, since neither PCrat$^{*}_{t-n}$ or PCrat$^{OTM*}_{t-n}$ show significant trends, rejecting H3$_0$ cannot be justified for Abn > 1.0 %; *market value does not affect an informed trader's decision.*

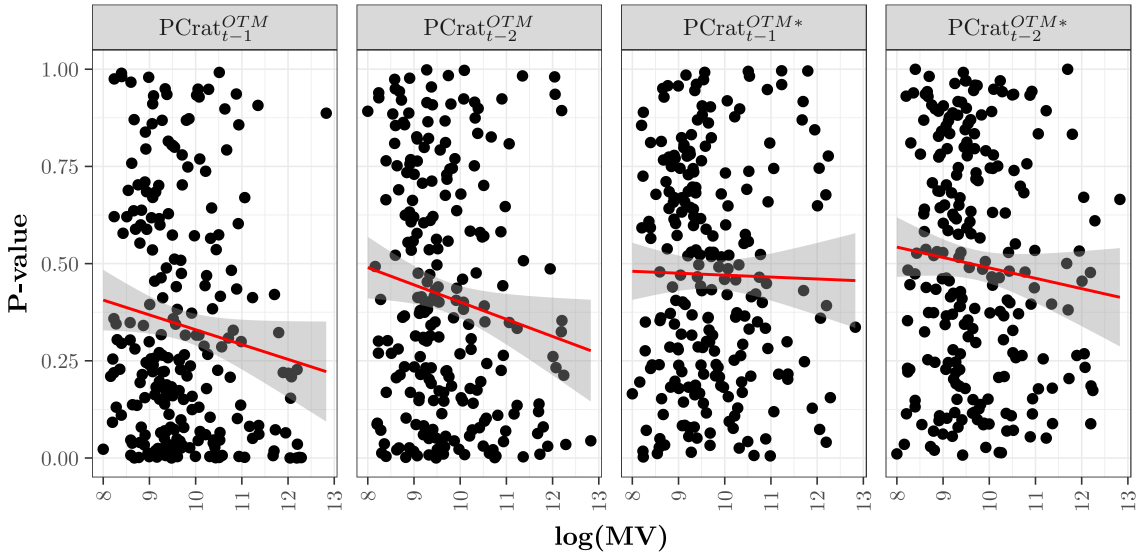
Figure 9: Time-series PCrat$^{(*)}_{t-n}$ P-values vs log(MV): Abn > 0.01



The figure shows a scatter plot of the P-values and the logarithm of market value for the variables PCrat$_{t-n}$ and PCrat$^{*}_{t-n}$. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.0 %.

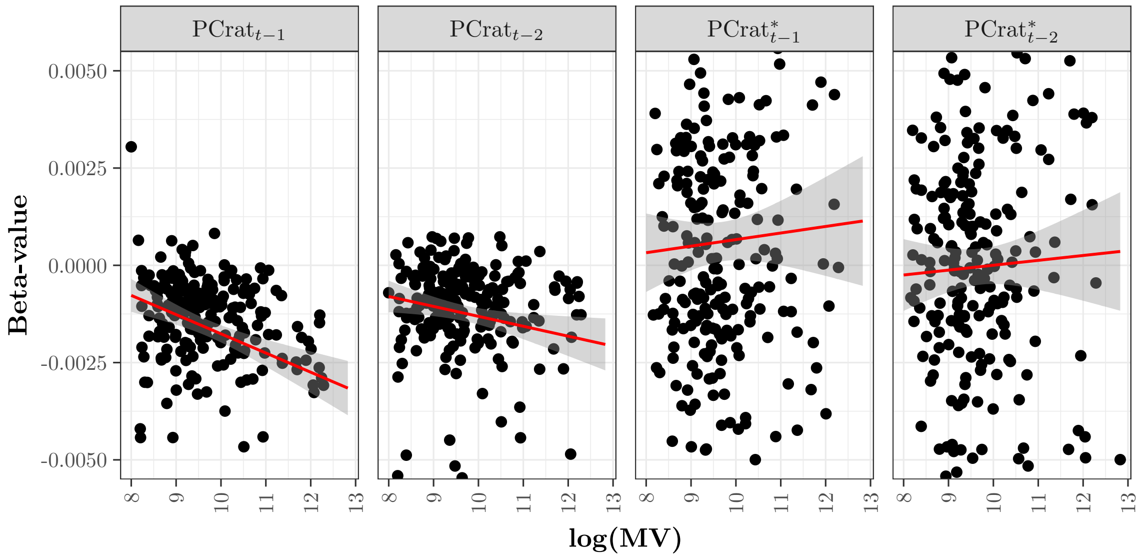
Comparing the regression coefficients and firm value in figure 11 and 12, gives the same concluding remarks as the previous analysis for Abn > 0.5 %. The only variables with a significant trend are PCrat$_{t-1}$ and PCrat$_{t-2}$. Since the confidence interval, for PCrat$^{*}_{t-n}$ and PCrat$^{OTM*}_{t-n}$, of the slope coefficient of the linearly fitted line contains zero, we cannot conclude that there is a relationship between firm size and the regression coefficient prior to large spikes in abnormal return.

Figure 10: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.01



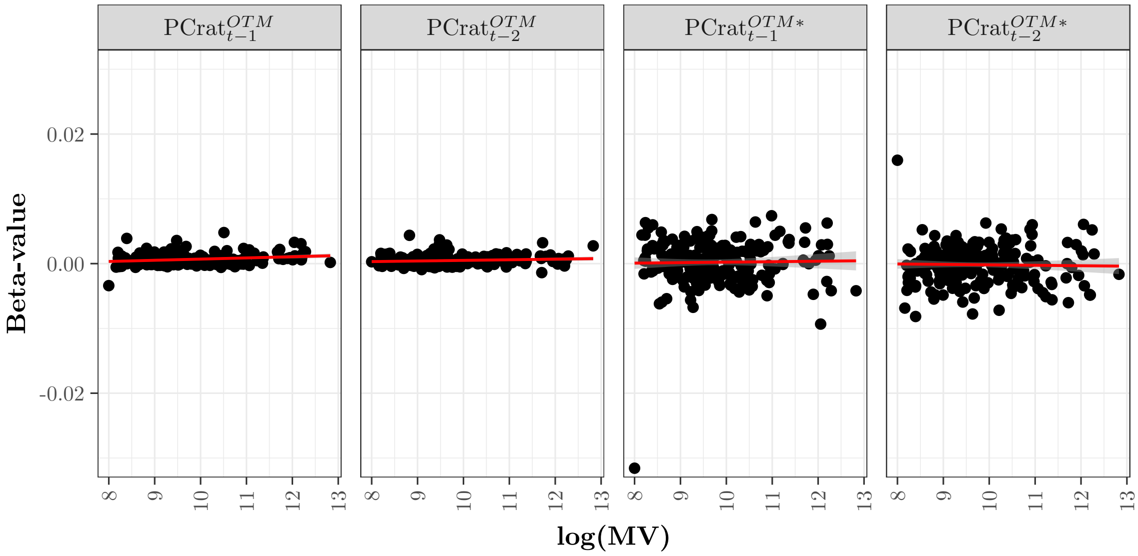
The figure shows a scatter plot of OTM observations for the P-values and the logarithm of market value for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.0 %.

Figure 11: Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs log(MV): Abn > 0.01



The figure shows a scatter plot of the regression coefficient (β) estimates and the logarithm of market value for the variables PCrat_{t-n} and PCrat_{t-n}^{*}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.0 %.

Figure 12: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.01

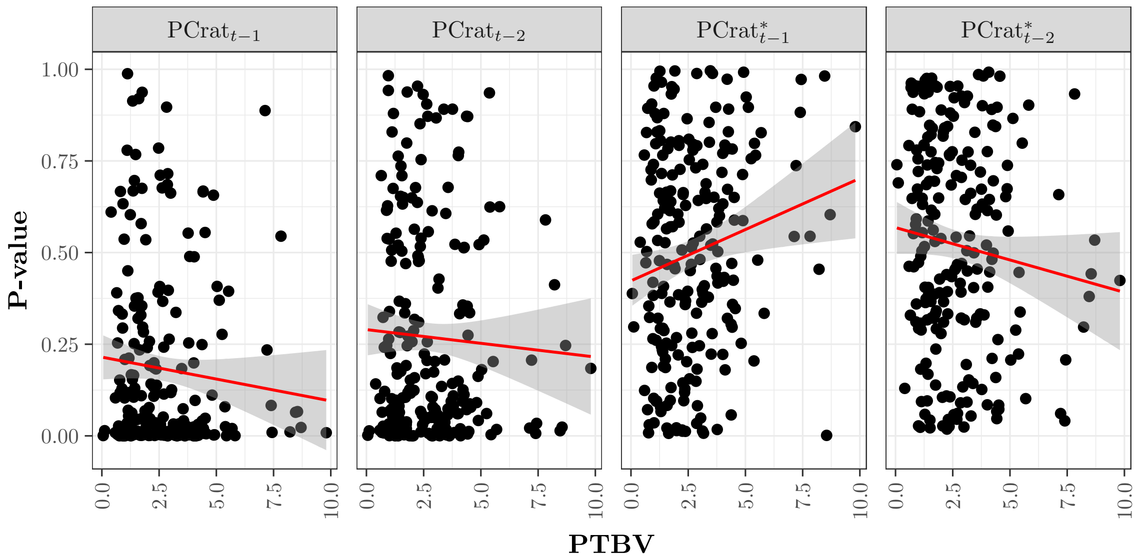


The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and the logarithm of market value for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.0 %.

From figure 13 and 14, we observe that non of the PCrat variations, expect for PCrat_{t-1}^{*}, exhibit a significant trend between P-values and PTBV. However, the significant trend in PCrat_{t-1}^{*} is upward sloping, which is in contrast with what is stated in H4. While we cannot justify rejecting $H4_0$, we can however conclude that there is a significant positive correlation between P-values and PTBV the day prior to spikes in abnormal return > 1.0 %. There were no significant trends between the regression coefficients and PTBV[21].

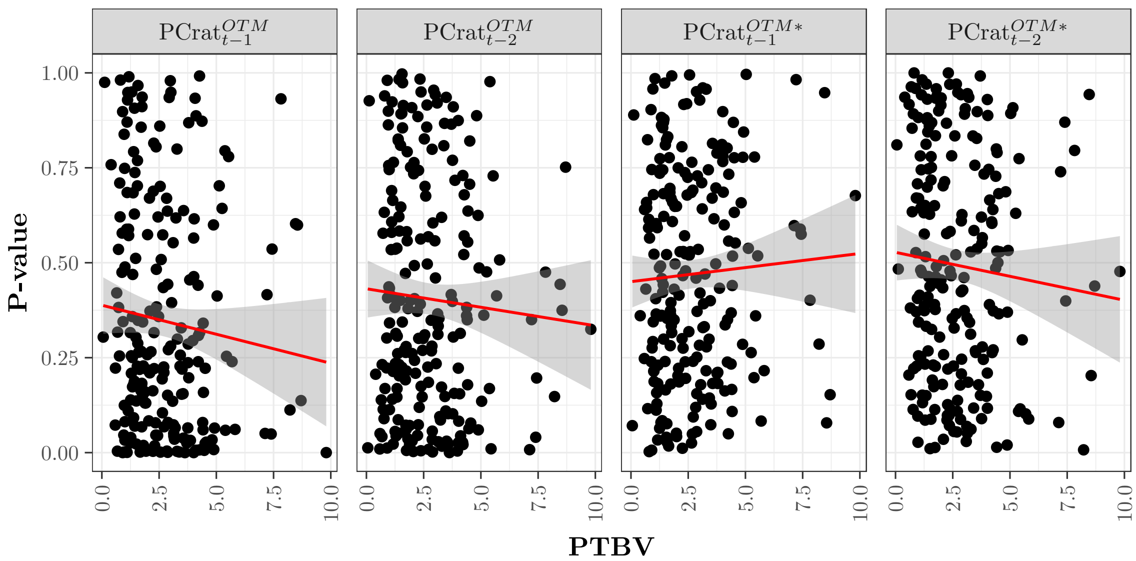
[21] Tables comparing regression coefficients and PTBV for Abn > 1.0 % are found in the appendix

Figure 13: Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.01



The figure shows a scatter plot of the P-values and PTBV for the variables PCrat_{t-n} and PCrat_{t-n}^{*}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.0 %.

Figure 14: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.01



The figure shows a scatter plot of OTM observations for the P-values and PTBV for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.0 %.

5.3 Cross Sectional

The following section summarizes the cross-sectional analysis. For each applicable trading day in the period 1^{st} of June 2009 to the 6^{th} of August 2014 a cross-sectional regression has been carried out. This totals 1306 regression over the period. Unlike the previous approaches, a series of cross sectional regressions will illustrate any changes in statistical and economic significance over time. This technique has been used before by Roll, Schwartz, and Subrahmanyam (2010) and will be utilized to answer H5 about any time-varying differences in the effect of the call to put option volume ratio.

Similar to the time-series sections, we will not conclude H1 and H2, but rather use the regression results to enhance the findings from the panel data analysis on a day-by-day basis. However, the analysis is mainly conducted to find supporting evidence of H5. It is worth noting that our third and fourth hypothesis is not applicable in the cross-sectional setting. The average market value in the cross-sectional regression would somewhat be a proxy for S&P500, and would thus not be an appropriate measure for individual differences.

The results substantiating H1 and H2 are summarized in tables, while results substantiating H5 are summarized in figures, both separated in sections for Abn > 0.5 % and Abn > 1.0 %. The tables and figures for Abn > 1.5 % can be found in the appendix.

5.3.1 Abn > 0.5 %

Table 11 shows a summary of the P-values and the regression coefficient estimates for the explanatory variables when Abn > 0.5 %. There seems to be no evidence of PCrat_{t-1} and PCrat_{t-2} containing any information at a significant level when comparing observed proportions of significance to expected proportions. The proportions of significant observations are 7.35 % and 5.59 % at a 5%-level and 13.94 % and 9.95 % at a 10 %-level for the two, just slightly more than what would be expected from a random sample. The combined test for P-values, which accounts for the *composition* of individual P-values, concludes differently. Both PCrat_{t-1} and PCrat_{t-2} are significant at a 5 %-level according to these tests. With regard to H1, the same proportions are observed for PCrat^*_{t-1} and PCrat^*_{t-2}, with 4.13 % and 5.51 % at a 5 %-level and 9.19 % and 11.49 % at a 10

Table 11: Cross-sectional regression P-values and coefficients : Abn > 0.005

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.1937	0.2695	0.0000	0.0000	-0.0022	-0.0005	0.0012
EPS	0.0689	0.1217	0.0029	0.1189	-0.0025	-0.0001	0.0027
Vol	0.1654	0.2243	0.0000	0.0000	-0.0036	0.0008	0.0052
Delta	0.5674	0.6807	0.0000	0.0000	0.0061	0.0097	0.0139
MV	0.1394	0.2213	0.0000	0.0000	-0.0019	0.0146	0.0338
PTBV	0.0459	0.0789	1.0000	1.0000	-0.0042	0.0007	0.0087
PCrat_{t-1}	0.0735	0.1394	0.0000	0.0001	-0.0023	-0.0009	0.0003
PCrat_{t-2}	0.0559	0.0995	0.0144	0.0015	-0.0019	-0.0008	0.0005
PCrat^{*}_{t-1}	0.0413	0.0919	0.8144	0.6909	-0.0030	-0.0002	0.0028
PCrat^{*}_{t-2}	0.0551	0.1149	0.0913	0.3455	-0.0029	-0.0001	0.0029
PCrat^{OTM}_{t-1}	0.0605	0.1172	0.2789	0.5571	-0.0006	0.0003	0.0012
PCrat^{OTM}_{t-2}	0.0360	0.0819	0.9615	0.8975	-0.0006	0.0002	0.0011
$\text{PCrat}^{OTM*}_{t-1}$	0.0429	0.0995	0.4716	0.5608	-0.0019	0.0003	0.0026
$\text{PCrat}^{OTM*}_{t-2}$	0.0536	0.1103	0.1539	0.2482	-0.0024	-0.0000	0.0022

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 0.5 %. The values are calculated from 1306 different cross-sectional regressions, with a total of 232 companies observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (1306). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

%-level. These modified variables show no significance in their combined P-values, except for PCrat^{*}_{t-2} which is significant at a 10 %-level using Fisher's method.

The variables for everyday OTM option volume, PCrat^{OTM}_{t-1}, PCrat^{OTM}_{t-2}, $\text{PCrat}^{OTM*}_{t-1}$ and $\text{PCrat}^{OTM*}_{t-2}$ all have observed proportions of significance of approximately 5 % at the 5 %-level and 10 % at the 10 %-level. The non-significance of these variables are justified by their combined P-values. Note that the cross-sectional regressions shows less significance for all the PCrat-variations compared to panel data. Fixed effect panel data controls for firm, j, and time, t, specific effects, and time-series control for firm specific. Cross-sectional regression, however, controls time specific effects. As the results observed in the panel data regression and the time series regressions are much closer in terms of significance than the panel data regression and the cross-sectional regressions, this implies that the firm specific effects has a much larger impact on how the regression coefficients of PCrat-variations behaves than the time specific effects.

Table 12: Cross-sectional regression explanatory power and observations : Abn > 0.005

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	-0.0714	0.0401	0.0844	0.1431	0.6463
No. obs	135	204	211	214	225
No. abn. obs	59	105	116	127	185

The table presents summarizing statistical properties when Abn > 0.5 % for a total of 1306 different cross-sectional regressions with a total of 232 companies observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 0.5 %.

As before, one of the most notable observations are the signs for the median observations for the day-to-day options and the OTM options variables. They have, expect for $\text{PCrat}_{t-2}^{OTM*}$, opposite signs. An increase in the $\text{PCrat}_{t-n}^{(*)}$ variables will at the median value predict a decrease in abnormal returns, while an increase in the $\text{PCrat}_{t-n}^{OTM(*)}$ variable predicts an increase in abnormal returns.

The PCrat_{t-n} variables have the largest impact in regard to the magnitude of the coefficients, with the first and second lag having a median effect of just under -0.1 % for an increase in the ratio of 1[22]. Note that the coefficients spreads for the modified variables, PCrat_{t-n}^{*} and Crat_{t-n}^{OTM*}, are wider comparing the cross-sectional regressions with the time-series regressions. This is sensible when taking the total number of observations and number of *Abnormal day* observations into account. For each cross-sectional regression there is a median of 211 observations while the time-series regressions have a median of 1246 observations.

As seen from the summary of regression meta data in table 12, our analysis explains at the median 8.4 % of the total variation, and between 4.0 % and 14.3 % in the 25^{th} percentile to 75^{th} percentile interval measured by *Adjusted* R^2. With a median of 116 observations of *Abnormal Days* and a minimum of 59, there are no reason to be concerned that the findings are purely a mechanical consequence of the statistical methods. This is also shown in figure 37 in the appendix.

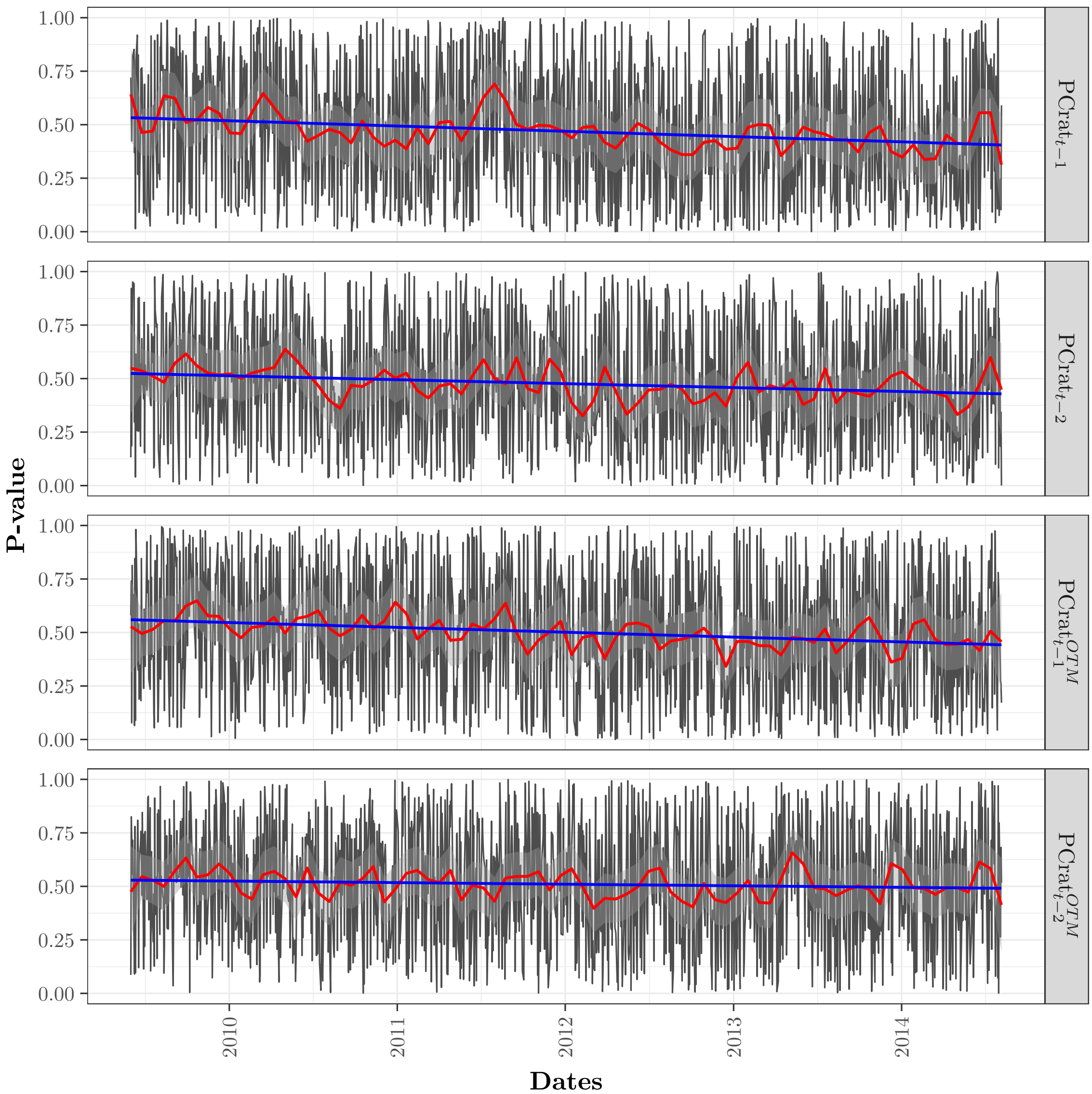
Our fifth hypothesis, H5, cannot be answered by the summary tables. The changes in P-values and regression coefficients over time is presented graphically in the figures on the following pages. This is inspired by the work of Roll, Schwartz, and Subrahmanyam

[22]Equation (21)

(2010). The trends have been summarized with two statistical techniques. The blue line is a simple linear regression on time, while the red line is a series of localized regressions (LOESS) paired with a 95 % confidence band. LOESS is explained in the appendix.

In our hypothesis we argued that the information rate in option ratios should have decreased over the time period, as research by Roll, Schwartz, and Subrahmanyam (2010) and Johnson and So (2012) have shown that option ratios contain information about future stock returns.

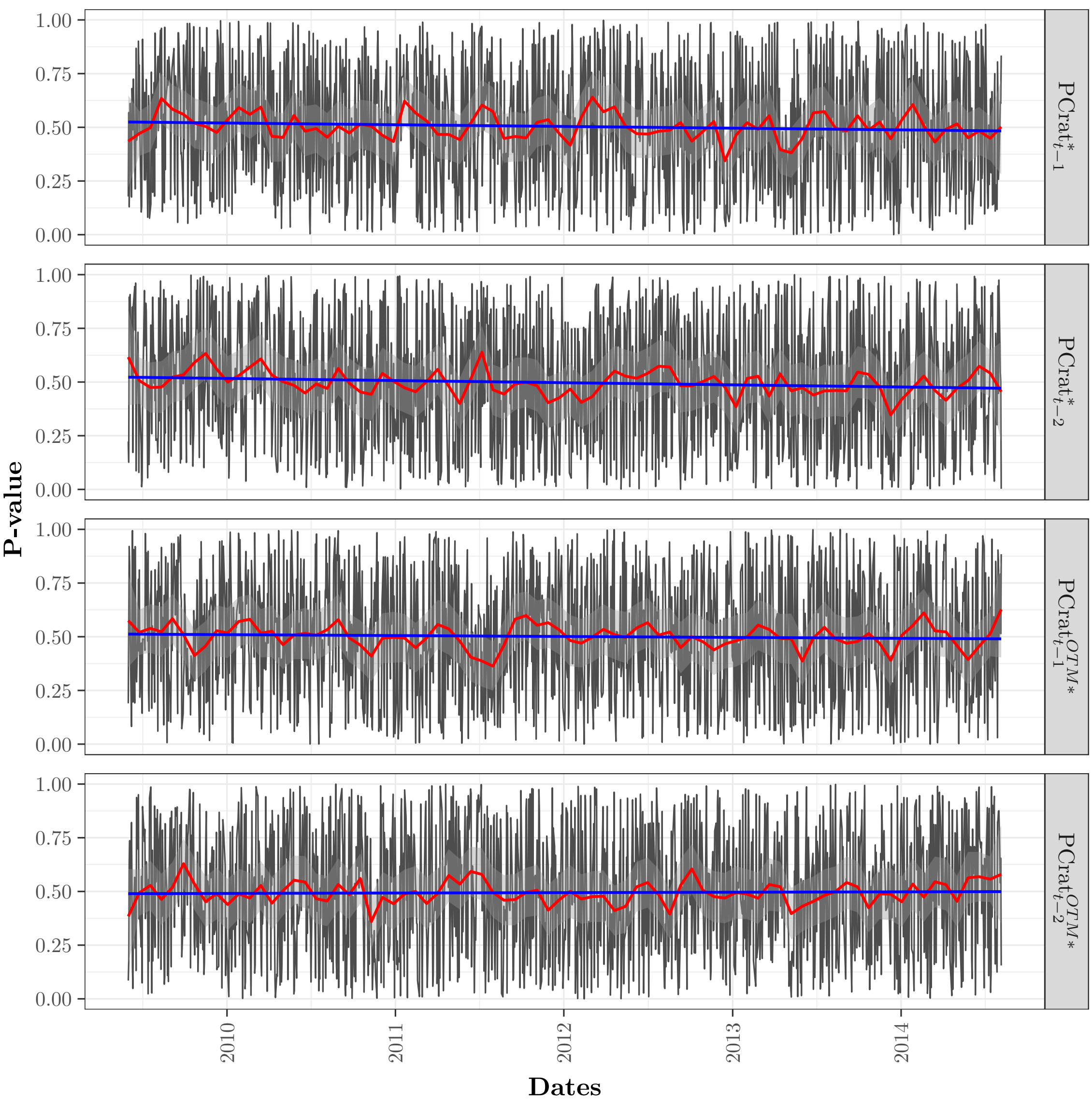
Figure 15: Cross-sectional $\mathrm{PCrat}_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.005



The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n} and $\mathrm{PCrat}_{t-n}^{OTM}$ when Abn > 0.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

As seen in figure 15, all variations of the PCrat_{t-n} and PCrat_{t-n}^{OTM} variables have a downwards trend in their P-values over time. This is not in line with our hypothesis, which would suggest and upward trend, i.e. less significance with time. All fitted linear trends have a negative slope coefficient, but time explains very little of the actual variation. Flat trends are observed for PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ in figure 16

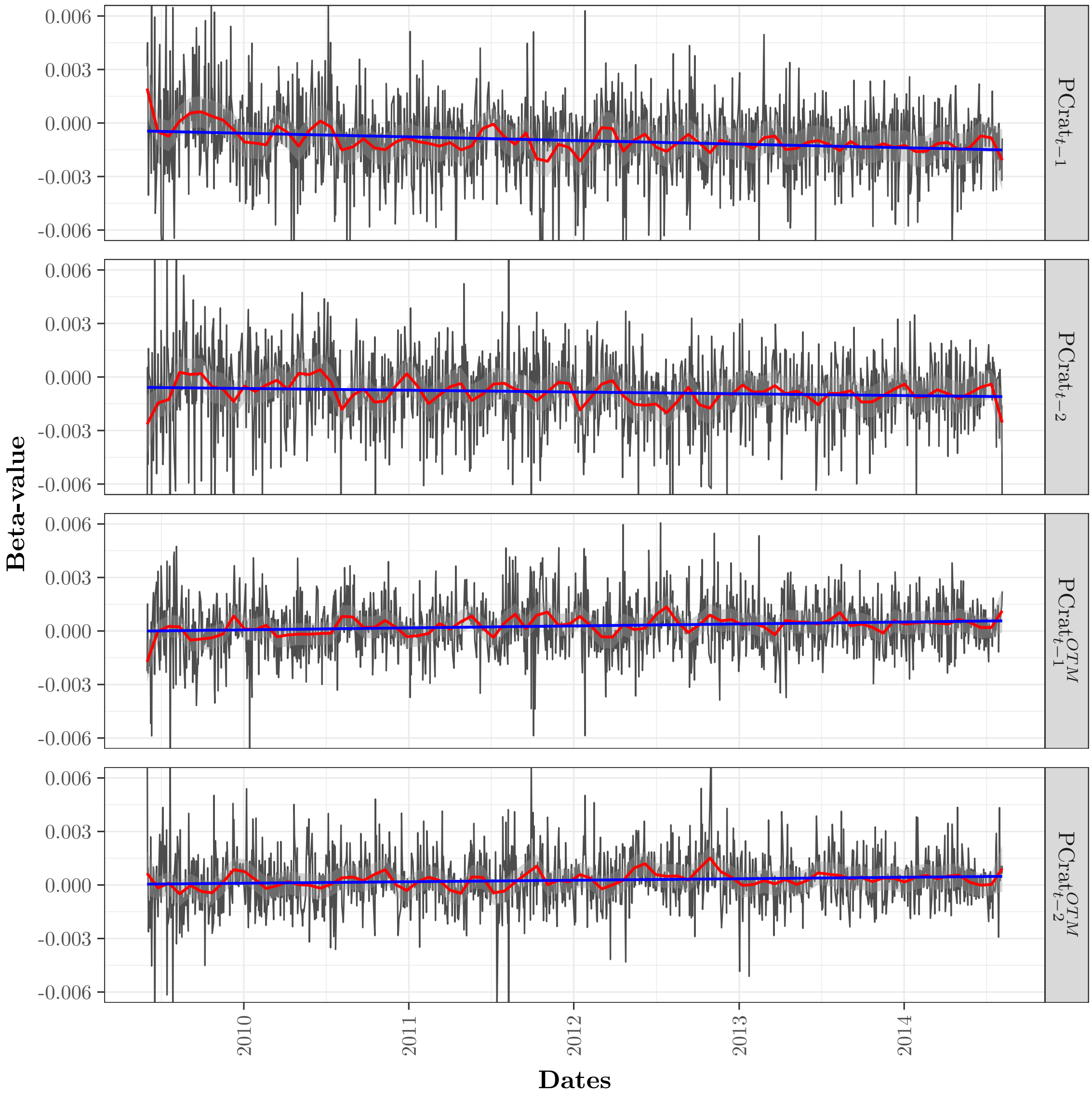
Figure 16: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.005



The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ when Abn > 0.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

The regression coefficients and thus the economic impact of the ratios have increased in magnitude over time. The coefficients for PCrat_{t-n} have decreased from approximately 0 to -0.0015, which is an increase in absolute value and its effect on abnormal returns. The betas for PCrat_{t-n}^{OTM} on the other hand has increased from approximately 0 to 0.0015, which is also an increase in its absolute value and effect on abnormal returns.

Figure 17: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.005

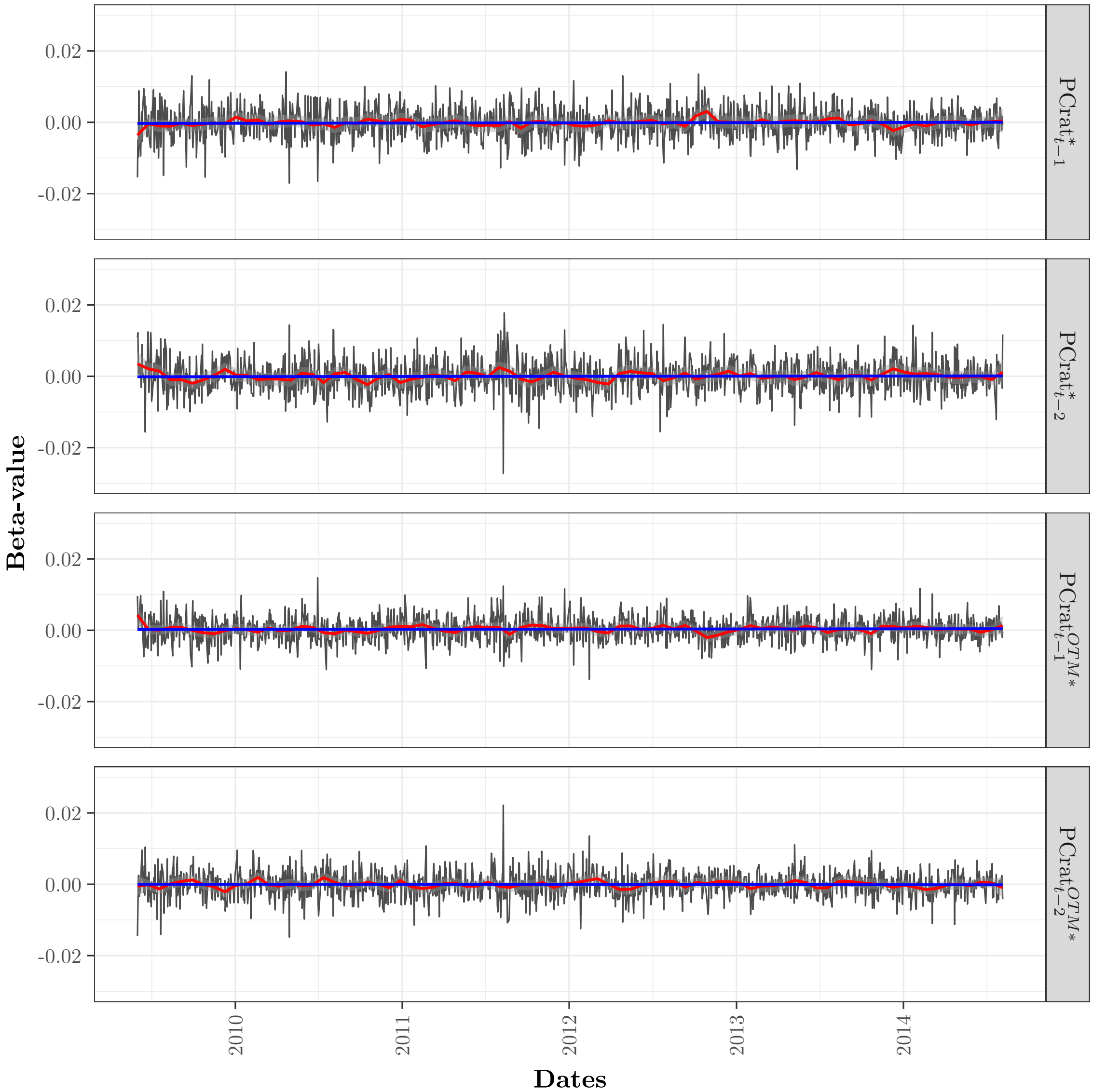


The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n} and PCrat_{t-n}^{OTM} when Abn > 0.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

While the absolute values of the coefficients in figure 17 have increased, the variance shows a decreasing trend over time. The coefficients for PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ in figure 18 does not have the same interesting findings as the previous figures. The coefficients show no trends or variations in variance over time, but have been included for completeness.

All in all, the null hypothesis of H5 cannot be rejected. Any trends in the P-values have indicated increased significance, and the trends in regression coefficients have shown increasing absolute values.

Figure 18: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.005



The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ when Abn > 0.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

5.3.2 Abn > 1.0%

The threshold for *Abnormal Days* has now increased to 1.0%. While our cross-sectional regressions for Abn > 0.5 % had few significant coefficients, regressing with the new limit for *Abn* in table 13 increases the significance for multiple variables. PCrat_{t-1} and

Table 13: Cross-sectional regression P-values and coefficients : Abn > 0.01

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.1784	0.2519	0.0000	0.0000	-0.0021	-0.0004	0.0012
EPS	0.0636	0.1080	0.0596	0.1972	-0.0025	-0.0000	0.0028
Vol	0.1547	0.2144	0.0000	0.0000	-0.0035	0.0007	0.0052
Delta	0.5299	0.6531	0.0000	0.0000	0.0058	0.0093	0.0136
MV	0.1363	0.2052	0.0000	0.0000	-0.0022	0.0147	0.0330
PTBV	0.0406	0.0697	1.0000	1.0000	-0.0042	0.0006	0.0086
PCrat_{t-1}	0.0819	0.1539	0.0000	0.0000	-0.0023	-0.0010	0.0003
PCrat_{t-2}	0.0689	0.1248	0.0000	0.0000	-0.0020	-0.0008	0.0004
PCrat^*_{t-1}	0.0521	0.0965	0.4872	0.6629	-0.0052	-0.0001	0.0050
PCrat^*_{t-2}	0.0651	0.1187	0.0839	0.7073	-0.0053	0.0001	0.0050
PCrat^{OTM}_{t-1}	0.0658	0.1340	0.0000	0.0002	-0.0006	0.0003	0.0013
PCrat^{OTM}_{t-2}	0.0620	0.1118	0.1383	0.5700	-0.0007	0.0002	0.0012
$\text{PCrat}^{OTM*}_{t-1}$	0.0536	0.1064	0.2464	0.7190	-0.0036	0.0006	0.0047
$\text{PCrat}^{OTM*}_{t-2}$	0.0605	0.1034	0.0915	0.5541	-0.0043	-0.0001	0.0041

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 0.5 %. The values are calculated from 1306 different cross-sectional regressions, with a total of 232 companies observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (1306). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

PCrat_{t-2} have had their proportions of significant results increase slightly compared to Abn > 0.5 %, from 7.35 % to 8.19 % and from 5.59 % to 6.89 % at the 5% level, and from 13.94 % to 15.39 % and from 9.95 % to 12.48 % at the 10% level. Their combined P-values have also decreased to zero for both, making them more significant than before. I.e. the put call ratio is more evident as an explanatory variable in days prior to a larger stock price change.

The modified ratios, PCrat^*_{t-n}, are not significant, with no evident changes in the proportions of significance compared to Abn > 0.5 %. PCrat^*_{t-2} is the exception when using the Fisher method, still being significant at the 10 %-level.

The most notable change in significance after increasing the limit for *Abnormal Days* concern PCrat^{OTM}_{t-1}. PCrat^{OTM}_{t-1} is now significant according to both combined P-values at all conventional significance levels. As argued in the previous section, controlling for firm specific differences isolates the effect of the PCrat-variables more than controlling

Table 14: Cross-sectional regression explanatory power and observations : Abn > 0.01

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	-0.0661	0.0647	0.1132	0.1775	0.6636
No. obs	135	204	211	214	225
No. abn. obs	11	50	60	71	152

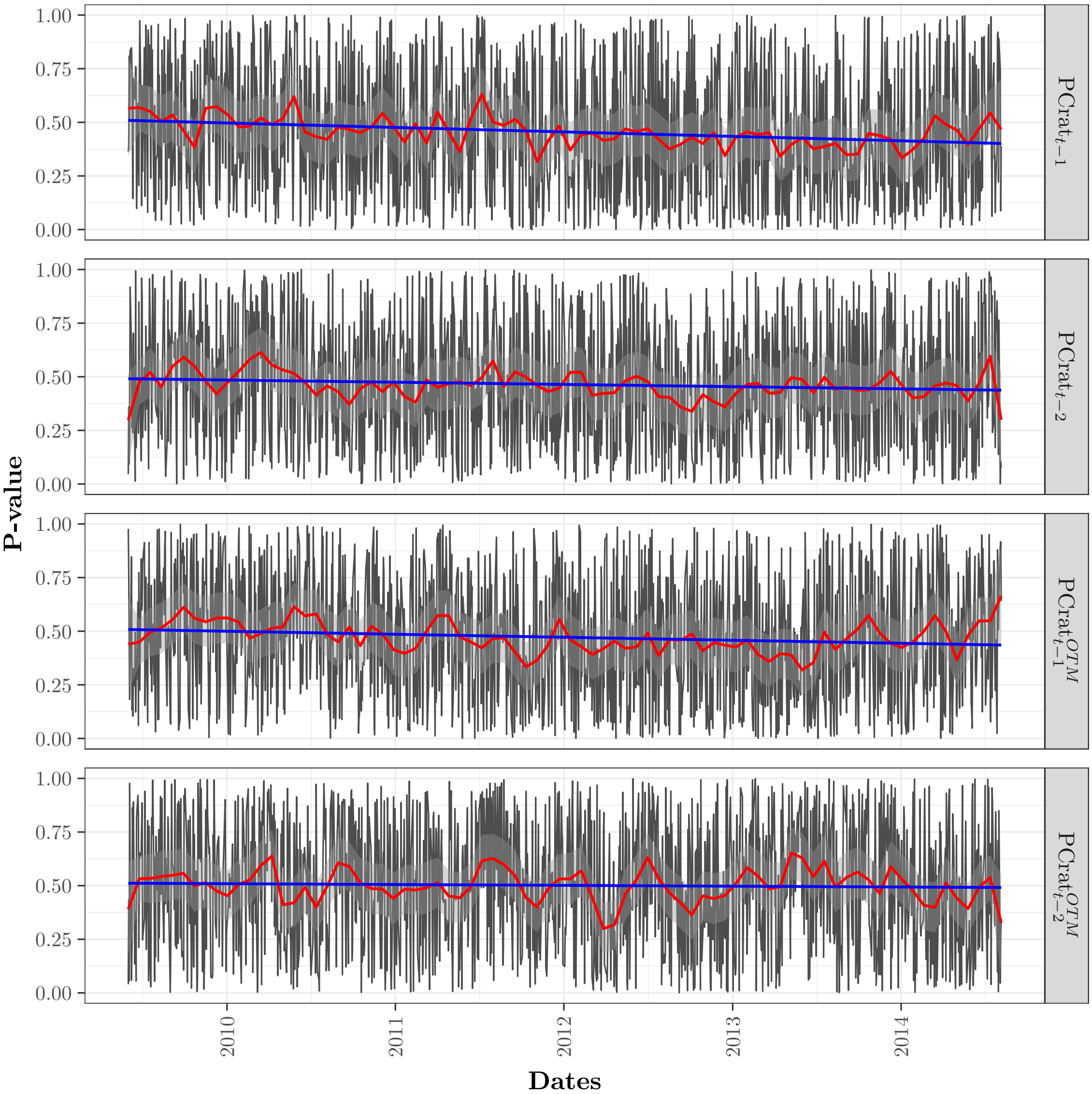
The table presents summarizing statistical properties when Abn > 1.0 % for a total of 1306 different cross-sectional regressions with a total of 232 companies observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 1.0 %.

for time specific effects. From both the panel regression and the time-series section, the effect for all option volume and OTM option volume is stronger at a one-day lag than at a two-day lag. The observation that PCrat_{t-1}^{OTM} becoming significant while PCrat_{t-2}^{OTM} is not, is thus in line with the previous findings. The increased significance for PCrat_{t-1}^{OTM} from Abn > 0.5 % to Abn > 1.0 % implies that OTM option volume becomes increasingly important for larger values of abnormal returns.

When assessing the economic significance of the regression results in table 13 we note the many similarities with the coefficients from table 11. In order of magnitude, the coefficients (at the median) with most impact are still PCrat_{t-1} and PCrat_{t-2}. The observed median in table 13 for all coefficients are almost identical as in table 11 with the exception of PCrat_{t-2}^{*} which now is negative.

From table 14 we observe that the total explained variation, expressed with Adj.R^2, has increased compared to the 0.5 % limit for abnormal days, from 8.4 % to 11.3 %. The median number of observations of abnormal days has decreased to 60, with 50 % of the regressions having several such observations in the interval [50, 71]. There is, however, still no relation between this number and the observed P-values for the regressions which makes our conclusions robust. This can be seen in figure 38 in the appendix.

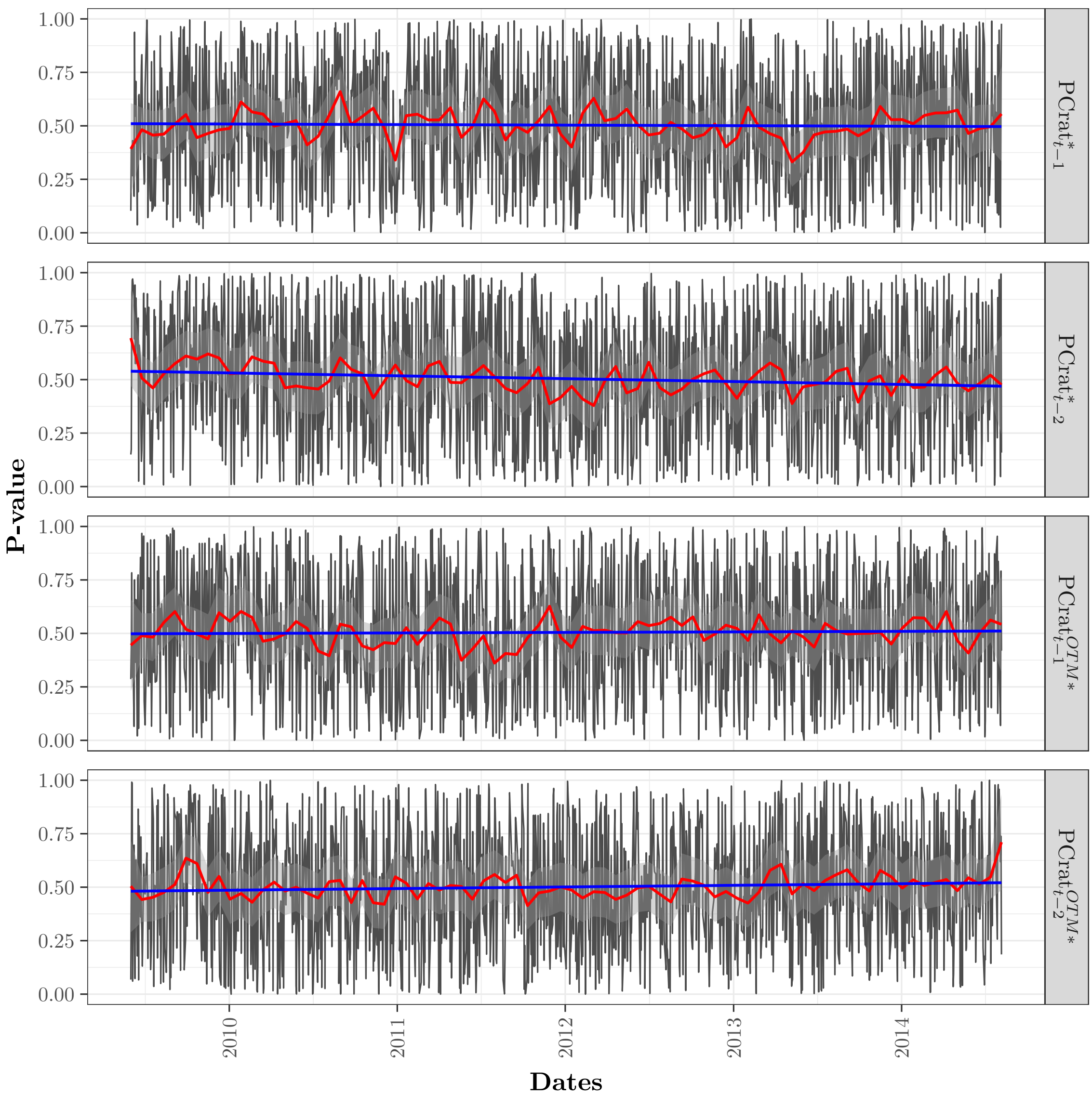
Figure 19: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.01



The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n} and PCrat_{t-n}^{OTM} when Abn > 1.0 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

While there are some differences regarding H1 and H2 when adjusting Abn from > 0.5 % to > 1.0 %, there are no big differences concerning H5. The same downward sloping trend for all $\text{PCrat}_{t-n}^{(OTM)}$-variations is observed in figure 19 and a flat trend for all $\text{PCrat}_{t-n}^{((OTM)*)}$-variations in 20. As with Abn > 0.5 %, this is not in line with H5, which would suggest and upward trend and less significance with time.

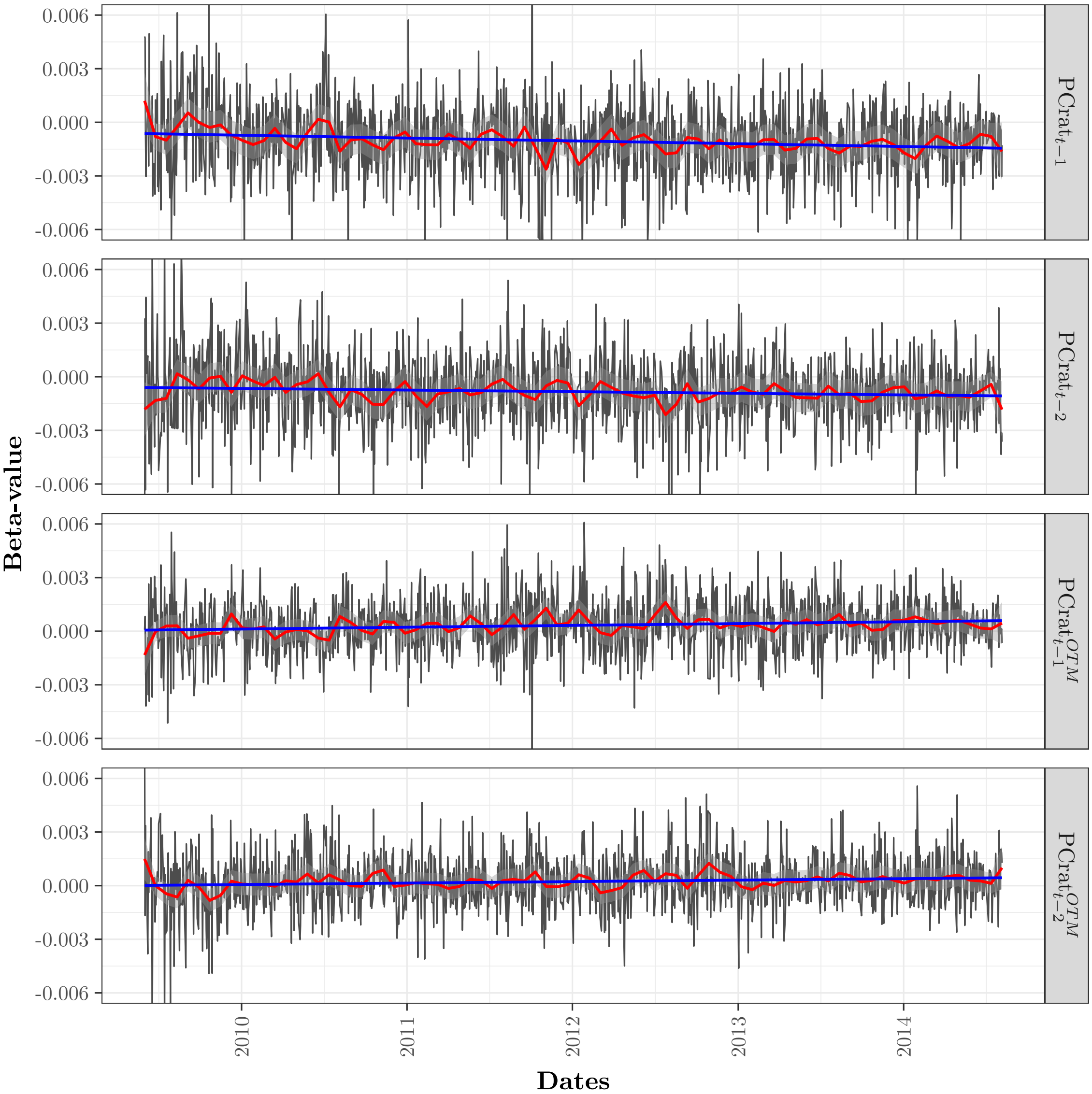
Figure 20: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.01



The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ when Abn > 1.0 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

The fitted linear trends for the coefficients for Abn > 1.0 % are similar to those of Abn $>$ 0.5 %. Comparing figure 21 and 17, we observe that they are almost identical. Although, coefficients for Abn > 0.5 % have an increased number of extreme observations and some increased variance, the fitted linear trend lines do not deviate by much. This is contrary to what we would expect, i.e. Abn > 0.5 % having a lower variance than Abn > 1.0 %.

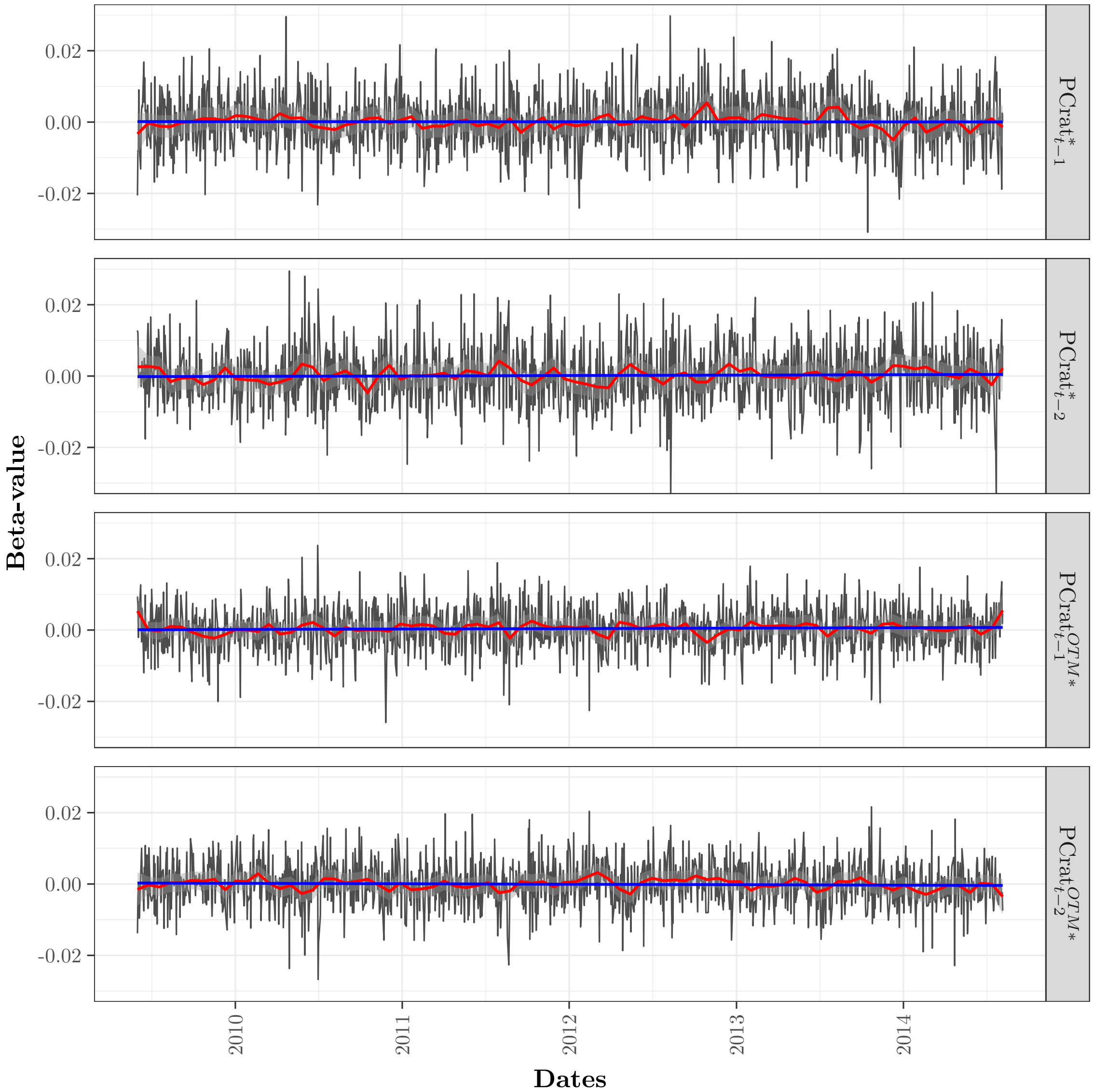
Figure 21: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.01



The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n} and PCrat_{t-n}^{OTM} when Abn > 1.0 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

The last figure, 22, for the coefficients of the modified variables, PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$, have also not changed much from Abn > 0.5 % to Abn > 1.0 %. The fitted linear regression lines shows no trend upwards or downwards. We note that the variance has increased with the new parameter value, but this is to be expected, as we have fewer observations for Abn > 1.0 % than for Abn > 0.5 %.

Figure 22: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.01



The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ when Abn > 1.0 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

As with Abn > 0.5 % we fail to reject the null hypothesis for H5 for Abn > 1.0 %. For the first four variables, PCrat_{t-n} and PCrat_{t-n}^{OTM}, the variables become both more statistically and economically significant with time. The P-values are decreasing on average, and the betas are increasing in absolute value. The modified variables, PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$, have no trends for either P-value or coefficients.

6 Conclusion

Volume is an extensive part of the financial markets. While previous research of volume primarily has focused on stock volume and its role as a stock price predictor, multiple studies of option volume has been conducted the last decade. Published research includes option volume as a stock price predictor and the relative informational content of options with different moneyness. For investors with private information, trading in option markets can be a more lucrative opportunity than stock markets due to e.g. capital restrictions. This thesis takes the research of option volume as a price predictor one step further. While previous research show that option volume leads stock price changes on a day-to-day basis, we examine whether option volume has additional predictive power in days prior to large changes in abnormal returns of the underlying asset. Under the assumption that the market is semi-efficient there should be no extra information in option volume prior to these changes, and any significant change in trading patterns will be a result of informed trading. Through our analyses we find significant evidence of unusual trading patterns prior to these spikes, indicating that informed trading is present in option markets.

This paper specifically investigates a modified put call ratio, $PCrat$, that takes the difference between unsigned call and put option volume divided by the total option volume. Such a ratio was also constructed for the out-of-the money options with a moneyness criteria of 5 %, $PCrat^{OTM}$. The data set is comprehensive and covers 232 companies over 1306 trading days for a total of approximately 350 000 different options. The effect of these put to call ratios was analysed at a 0.5 % and a 1.0 % threshold for abnormal returns.

A fixed effect panel data regression concluded that the modified put to call ratio display a different pattern the day prior to large stock price changes. This is true for both the total option volume ratio, $PCrat$, and the out-of-the money ratio, $PCrat^{OTM}$. At the 0.5 % threshold for abnormal returns, the out-of-the-money option ratio is more significant than the total option ratio. At the 1.0 % threshold, the significance is reversed with $PCrat^{OTM}$ being the most significant. The fixed effect panel data regression concludes our first and second sub-hypothesis, and we claim that there is evidence of a relative change in both the total and the isolated out-of-the-money call to put option volume prior to large spikes in abnormal return on the underlying asset. The effect is, however, only significant for a one-day lag.

The collection of time-series regressions, one for each company in the sample, concludes our sub-hypotheses H3 and H4. There is a significant correlation between market value and the day-to-day put to call ratios, but there is no significant change in pattern prior to days with a high abnormal return. The same is true for price-to-book-value, where no significant change in patterns were observed. There is however a significant correlation between price-to-book-value and the total put to call ratio, *PCrat* for a one-day lag at both thresholds for abnormal returns. For out-of-the-money options, there is no significant correlation with price-to-book-value for any lag. As some informed traders cannot choose which companies they obtain information of, such as corporate insiders, no systematic correlation between company characteristics and changes in day-to-day option volume pattern prior to large changes in abnormal returns is sensible. We fail to reject our null hypotheses for both H3 and H4, and claim that informed traders have no preference with regard to market value and price-to-book-ratio.

The collection of cross-sectional regressions, one for each day in the sample, concludes our last sub-hypothesis, H5. The significance of the modified put to call ratio should decrease over time under the assumption of a semi-efficient market. As previous research has concluded option volume as a significant predictor of future stock returns, the predictive power of option volume should diminish over time as the market is now informed of the predictive power of option volume. Our analysis concludes with the contrary, as both the average statistical significance and the absolute values of the coefficients of the variables have increased over time. We fail to reject the null hypothesis, as the effects are increasing, not decreasing.

We note that the economic impact of changes in any of the modified put to call ratios have been of small magnitude in all our analyses; but the results from the empirical results, combined with the assumption that the market is semi-efficient, sustain a claim that changes in trading volume of options imply evidence of informed trading before large changes in abnormal returns.

References

P. Augustin, M. Brenner, and M. G. Subrahmanyam. Informed options trading prior to m&a announcements: Insider trading? 2015.

R. Ball and P. Brown. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, pages 159–178, 1968.

B. J. Becker. Combining significance levels. *The Handbook of Research Synthesis*, pages 215–230, 1994.

F. Black. Fact and fantasy in the use of options. *Financial Analysts Journal*, pages 36–72, 1975.

L. Blume, D. Easley, and M. O'hara. Market statistics and technical analysis: The role of volume. *The Journal of Finance*, 49(1):153–181, 1994.

S. J. Brown and J. B. Warner. Measuring security price performance. *Journal of Financial Economics*, 8(3):205–258, 1980.

S. J. Brown and M. I. Weinstein. Derived factors in event studies. *Journal of Financial Economics*, 14(3):491–495, 1985.

C. Cao, Z. Chen, and J. M. Griffin. Informational content of option volume prior to takeovers. *The Journal of Business*, 78(3):1073–1109, 2005.

S. Chakravarty, H. Gulen, and S. Mayhew. Informed trading in stock and option markets. *The Journal of Finance*, 59(3):1235–1257, 2004.

K. Chan, Y. P. Chung, and W.-M. Fong. The informational role of stock and option volume. *Review of Financial Studies*, 15(4):1049–1075, 2002.

V. Chari, R. Jagannathan, and A. R. Ofer. Seasonalities in security returns. *Journal of Financial Economics*, 21(1):101 – 121, 1988.

C. R. Chen, P. P. Lung, and N. S. Tay. Information flow between the stock and option markets: Where do informed traders trade? *Review of Financial Economics*, 14(1): 1–23, 2005.

T. Chordia and B. Swaminathan. Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance*, 55(2):913–935, 2000.

W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.

C. Dismuke and R. Lindrooth. Ordinary least squares. *Methods and Designs for Outcomes Research*, 93:93–104, 2006.

D. Easley, M. O'hara, and P. S. Srinivas. Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2):431–465, 1998.

E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

E. F. Fama and K. R. French. The capm is wanted, dead or alive. *The Journal of Finance*, 51(5):1947–1958, 1996.

E. F. Fama, L. Fisher, M. C. Jensen, and R. Roll. The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21, 1969.

G. W. Fenn and N. Liang. Corporate payout policy and managerial stock incentives. *Journal of Financial Economics*, 60(1):45–72, 2001.

R. A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, 4 edition, 1932.

L. Ge, T.-C. Lin, and N. D. Pearson. Why does the option to stock volume ratio predict stock returns? *Journal of Financial Economics*, 120(3):601–622, 2016.

W. H. Greene. *Econometric analysis*. Pearson Education, 2003.

G. R. Jensen, R. R. Johnson, and J. M. Mercer. New evidence on size and price-to-book effects in stock returns. *Financial Analysts Journal*, pages 34–42, 1997.

T. L. Johnson and E. C. So. The option to stock volume ratio and future returns. *Journal of Financial Economics*, 106(2):262–286, 2012.

C. G. Lamoureux and W. D. Lastrapes. Heteroskedasticity in stock return data: volume versus garch effects. *The Journal of Finance*, 45(1):221–229, 1990.

J. S. Long and L. H. Ervin. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224, 2000.

A. C. MacKinlay. Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39, 1997.

J. G. MacKinnon and H. White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3): 305–325, 1985.

L. K. Meulbroek. An empirical analysis of illegal insider trading. *The Journal of Finance*, 47(5):1661–1699, 1992.

D. Morse. Asymmetrical information in securities markets and trading volume. *Journal of Financial and Quantitative Analysis*, 15(05):1129–1148, 1980.

D. Morse. An econometric analysis of the choice of daily versus monthly returns in tests of information content. *Journal of Accounting Research*, pages 605–623, 1984.

V. Niederhoffer and M. F. Osborne. Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, 61(316):897–916, 1966.

J. Pan and A. M. Poteshman. The information in option volume for future stock prices. *Review of Financial studies*, 19(3):871–908, 2006.

R. Roll, E. Schwartz, and A. Subrahmanyam. O/S: The relative trading activity in options and stock. *Journal of Financial Economics*, 96(1):1–17, 2010.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

D. Ruan, J. A. Fessler, and J. Balter. Real-time prediction of respiratory motion based on local regression methods. *Physics in Medicine and Biology*, 52(23):7137, 2007.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

J. A. Stephan and R. E. Whaley. Intraday price change and trading volume relations in the stock and stock option markets. *The Journal of Finance*, 45(1):191–220, 1990.

A. M. Vijh. Liquidity of the cboe equity options. *The Journal of Finance*, 45(4):1157–1179, 1990.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.

J. M. Wooldridge. *Econometric analysis of cross section and panel data.* MIT Press, 1 edition, 2002.

J. M. Wooldridge. *Introductory econometrics: A modern approach.* Cengage Learning, 5 edition, 2012.

Appendices

A Tickers

Table 15: Stock tickers and company names, A - IBM

Ticker	Company name	Ticker	Company name
A	Agilent Technologies Inc.	DE	Deere & Co.
AAPL	Apple Inc.	DGX	Quest Diagnostics
ABC	AmerisourceBergen Corp.	DHI	D. R. Horton
ABT	Abbott Laboratories	DHR	Danaher Corp.
ADI	Analog Devices Inc.	DIS	The Walt Disney Co.
ADP	ADP, LLC.	DOW	Dow Chemical
ADSK	Autodesk Inc.	DPS	Dr Pepper Snapple Group
AEE	Ameren Corp.	DRI	Darden Restaurants
AEP	American Electric Power	DUK	Duke Energy
AES	AES Corp.	DVN	Devon Energy Corp.
AFL	AFLAC Inc.	EA	Electronic Arts
AIG	AIG, Inc.	ECL	Ecolab Inc.
AIZ	Assurant Inc.	ED	Consolidated Edison
AKAM	Akamai Technologies Inc.	EIX	Edison Int'l
ALL	Allstate Corp.	EMC	EMC Corp.
ALTR	Altera Corp.	EMN	Eastman Chemical
AMAT	Applied Materials Inc.	EMR	Emerson Electric Co
AMGN	Amgen Inc.	EOG	EOG Resources
AMP	Ameriprise Financial	EQR	Equity Residential
AN	AutoNation Inc.	EQT	EQT Corporation
APH	Amphenol Corp.	ETFC	E*Trade Financial Corp.
ATI	Allegheny Technologies Inc.	EXC	Exelon Corp.
AVB	AvalonBay Communities	EXPD	Expeditors Int'l
AXP	American Express Co	EXPE	Expedia Inc.
BAC	Bank of America Corp.	FAST	Fastenal Co.
BAX	Baxter International Inc.	FCX	Freeport-McMoRan Inc.
BEAM	Beam Inc.	FDX	FedEx Corporation
BEN	Franklin Resources	FE	FirstEnergy Corp
BF.B	Brown-Forman Corp.	FIS	Fidelity Nat. Info. Services
BK	BNY Mellon	FISV	Fiserv Inc
BLL	Ball Corp	FITB	Fifth Third Bancorp
BMS	Bemis	FLIR	FLIR Systems
BSX	Boston Scientific	FLR	Fluor Corp.
BTUUQ	Peabody Energy	FLS	Flowserve Corporation
C	Citigroup Inc.	FTR	Frontier Communications
CBG	CBRE Group	GAS	AGL Resources Inc.
CCL	Carnival Corp.	GE	General Electric
CHK	Chesapeake Energy	GME	GameStop Corp.
CHRW	C. H. Robinson Worldwide	GNW	Genworth Financial Inc.
CI	Cigna Corp.	GOOGL	Alphabet Inc. Class A
CL	Colgate-Palmolive	GPS	Gap Inc.
CMA	Comerica Inc.	GS	Goldman Sachs Group
CME	CME Group Inc.	GT	Goodyear Tire & Rubber
CNP	CenterPoint Energy	GWW	W. W. Grainger, Inc.
CNX	Consol Energy Inc.	HAL	Halliburton Co.
COG	Cabot Oil & Gas	HAS	Hasbro Inc.
COH	Coach Inc.	HBAN	Huntington Bancshares
COL	Rockwell Collins	HCBK	Hudson City Bancorp
COP	ConocoPhillips	HCP	HCP Inc.
COST	Costco Co.	HD	Home Depot
CPB	Campbell Soup	HIG	Hartford Financial Services
CSX	CSX Corp.	HOG	Harley-Davidson
CTAS	Cintas Corp.	HRB	H&R Block, Inc.
CTSH	Cognizant Technology Sol.	HRS	Harris Corp.
CTXS	Citrix Systems	HST	Host Hotels & Resorts
CVS	CVS Health Corp.	HSY	The Hershey Company
CVX	Chevron Corp.	HUM	Humana Inc.
D	Dominion Resources	IBM	IBM Corp.

Table 16: Stock tickers and company names, IFF - ZBH

Ticker	Company name	Ticker	Company name
IFF	Intl Flavors & Fragrances	PFG	Principal Financial Group
INTU	Intuit Inc.	PH	Parker-Hannifin
IP	International Paper Co.	PHM	PulteGroup Inc
IPG	Interpublic Group	PM	Philip Morris Int'l
ISRG	Intuitive Surgical Inc.	PNW	Pinnacle West Capital
IVZ	Invesco Ltd.	PPG	PPG Industries
JBL	Jabil Circuit	PRU	Prudential Financial
JNJ	Johnson & Johnson	PSA	Public Storage
JPM	JPMorgan Chase & Co.	PX	Praxair Inc.
JWN	Nordstrom	PXD	Pioneer Natural Resources
KLAC	KLA-Tencor Corp.	RAI	Reynolds American Inc.
KO	The Coca Cola Co.	RDC	Rowan Companies plc
KR	Kroger Co.	RF	Regions Financial Corp.
KSS	Kohl's Corp.	RL	Polo Ralph Lauren Corp.
L	Loews Corp.	RRC	Range Resources Corp.
LB	L Brands Inc.	RSG	Republic Services Inc.
LH	LabCorp	RTN	Raytheon Co.
LLL	L-3 Comm's Holdings	SBUX	Starbucks Corp.
LLTC	Linear Technology Corp.	SCHW	Charles Schwab Corp.
LLY	Eli Lilly and Co.	SHW	Sherwin-Williams Co.
LM	Legg Mason	SJM	J.M. Smucker Co.
LMT	Lockheed Martin Corp.	SNDK	SanDisk Corporation
LNC	Lincoln National	SPG	Simon Property Group Inc.
LO	Lorillard Inc.	SRCL	Stericycle Inc.
LOW	Lowe's Cos.	STI	SunTrust Banks
LUK	Leucadia National Corp.	STJ	St. Jude Medical, Inc.
LUV	Southwest Airlines	STT	State Street Corp.
MA	Mastercard Inc.	SWK	Stanley Black & Decker
MAS	Masco Corp.	SWN	Southwestern Energy
MAT	Mattel Inc.	SYY	Sysco Corp.
MCD	McDonald's Corp.	TAP	Molson Coors Brewing Co.
MCHP	Microchip Technology	TE	TECO Energy
MCK	McKesson Corp.	TIF	Tiffany & Co.
MET	MetLife Inc.	TMO	Thermo Fisher Scientific
MKC	McCormick & Co.	TSN	Tyson Foods
MMC	Marsh & McLennan	TWX	Time Warner Inc.
MO	Altria Group Inc.	TXT	Textron Inc.
MON	Monsanto Co.	UNH	United Health Group Inc.
MRK	Merck & Co.	UNP	Union Pacific
MRO	Marathon Oil Corp.	UPS	United Parcel Service
MSI	Motorola Solutions Inc.	USB	U.S. Bancorp
MTB	M&T Bank Corp.	VAR	Varian Medical Systems
MU	Micron Technology	VZ	Verizon Communications
MUR	Murphy Oil	WAT	Waters Corp.
MWV	MeadWestvaco Corp.	WFC	Wells Fargo
NBL	Noble Energy Inc.	WFM	Whole Foods Market
NEE	NextEra Energy	WIN	Windstream Comm's
NKE	Nike, Inc.	WMB	Williams Companies
NOV	National Oilwell Varco Inc.	WMT	Wal-Mart Stores
NWL	Newell Rubbermaid Co.	WU	Western Union Co.
OI	Owens-Illinois Inc.	WY	Weyerhaeuser Corp.
ORCL	Oracle Corp.	WYNN	Wynn Resorts Ltd.
PBCT	People's United Bank	X	U.S. Steel Corp.
PBI	Pitney-Bowes	XL	XL Capital
PCP	Precision Castparts	XLNX	Xilinx Inc.
PDCO	Patterson Companies	XRX	Xerox Corp.
PEG	Public Serv. Enterprise Inc.	YHOO	Yahoo Inc.
PFE	Pfizer Inc.	ZBH	Zimmer Biomet Holdings

B Tables

B.1 Time Series

Table 17: Time-series regression P-values and coefficients : Abn > 0.015

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.2251	0.3160	0.0000	0.0000	-0.0010	-0.0005	-0.0001
EPS	0.0823	0.1039	0.4134	0.8223	-0.0018	0.0002	0.0021
Vol	0.0952	0.1602	0.0002	0.0092	-0.0016	0.0000	0.0016
Delta	0.9567	0.9740	0.0000	0.0000	0.0059	0.0096	0.0155
MV	0.1602	0.2511	0.0000	0.0000	-0.0003	0.0107	0.0201
PTBV	0.0563	0.1212	0.0855	0.3355	-0.0030	0.0016	0.0126
PCrat_{t-1}	0.3983	0.5368	0.0000	0.0000	-0.0023	-0.0012	-0.0006
PCrat_{t-2}	0.3030	0.3983	0.0000	0.0000	-0.0016	-0.0010	-0.0005
PCrat^*_{t-1}	0.0649	0.1082	0.2067	0.4448	-0.0031	0.0007	0.0047
PCrat^*_{t-2}	0.0519	0.1082	0.8012	0.7262	-0.0032	0.0007	0.0045
PCrat^{OTM}_{t-1}	0.2078	0.3117	0.0000	0.0000	0.0001	0.0006	0.0010
PCrat^{OTM}_{t-2}	0.1385	0.2121	0.0000	0.0000	-0.0000	0.0004	0.0009
$\text{PCrat}^{OTM*}_{t-1}$	0.0433	0.0866	0.6031	0.5466	-0.0030	0.0004	0.0033
$\text{PCrat}^{OTM*}_{t-2}$	0.0736	0.1126	0.3471	0.6974	-0.0034	-0.0001	0.0026

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 1.5 %. The values are calculated from 232 different time-series regressions, with a total of 1306 trading days observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (232). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

Table 18: Time-series regression explanatory power and observations : Abn > 0.015

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	0.0024	0.0424	0.0601	0.0794	0.2383
No. obs	135	1024	1246	1303	1306
No. abn. obs	12	103	167	247	516

The table presents summarizing statistical properties when Abn > 1.5 % for a total of 232 different time-series regressions with a total of 1306 trading days observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 1.5 %.

B.2 Cross Sectional

Table 19: Cross-sectional regression P-values and coefficients : Abn > 0.015

	P-values		Combined P-values		Regression coefficients		
	Proportion significant at 5%	Proportion significant at 10%	Fisher's method	Adj. mean	25^{th} percentile	Median	75^{th} percentile
Intercept	0.1696	0.2387	0.0000	0.0000	-0.0019	-0.0004	0.0011
EPS	0.0591	0.1090	0.2923	0.6675	-0.0024	-0.0000	0.0027
Vol	0.1343	0.1972	0.0000	0.0000	-0.0034	0.0006	0.0047
Delta	0.4942	0.6309	0.0000	0.0000	0.0054	0.0090	0.0132
MV	0.1236	0.2034	0.0000	0.0000	-0.0025	0.0142	0.0321
PTBV	0.0307	0.0737	1.0000	1.0000	-0.0042	0.0006	0.0082
PCrat_{t-1}	0.0890	0.1581	0.0000	0.0000	-0.0023	-0.0009	0.0003
PCrat_{t-2}	0.0721	0.1305	0.0000	0.0000	-0.0021	-0.0008	0.0005
PCrat^*_{t-1}	0.0514	0.0967	0.2702	0.9952	-0.0094	0.0003	0.0100
PCrat^*_{t-2}	0.0714	0.1228	0.0014	0.7945	-0.0092	0.0003	0.0112
PCrat^{OTM}_{t-1}	0.0675	0.1213	0.0002	0.0038	-0.0007	0.0003	0.0013
PCrat^{OTM}_{t-2}	0.0652	0.1120	0.0472	0.3951	-0.0008	0.0002	0.0012
$\text{PCrat}^{OTM*}_{t-1}$	0.0553	0.0990	0.9487	0.9991	-0.0070	0.0011	0.0082
$\text{PCrat}^{OTM*}_{t-2}$	0.0675	0.1097	0.2594	0.9921	-0.0077	-0.0002	0.0074

The table summarizes the P-value distributions and the coefficient estimates for the explanatory variables when Abn > 1.5 %. The values are calculated from 1306 different cross-sectional regressions, with a total of 232 companies observed from mid-2009 to mid-2014. Proportion significant at x % is the number of regressions with a P-value < x %, divided by total regressions (1306). The dependent variable is daily abnormal returns. EPS, Vol, Delta, MV and PTBV are control variables, and the variables of interest are the different PCrat variations.

Table 20: Cross-sectional regression explanatory power and observations : Abn > 0.015

	Min	25^{th} percentile	Median	75^{th} percentile	Max
Adj.R^2	-0.0334	0.0898	0.1534	0.2227	0.6978
No. obs	135	204	211	214	225
No. abn. obs	4	24	30	39	117

The table presents summarizing statistical properties when Abn > 1.5 % for a total of 1306 different cross-sectional regressions with a total of 232 companies observed from mid-2009 to mid-2014. Adj.R^2 is the "goodness-of-fit" measures, No. obs is the total number of days with registered option trades for all firms, while No. abn. obs is the number of observations with an abnormal return > 1.5 %.

B.3 Panel Data

Table 21: Two-way fixed-effect panel regression : Abn > 0.015

	Estimate	90 % CI Lower	90 % CI Upper	95 % CI Lower	95 % CI Upper	T-value	P-value
EPS	-0.0001	-0.0003	0.0001	-0.0003	0.0001	-0.78	0.4362
Vol	-0.0006	-0.0010	-0.0002	-0.0011	-0.0002	-2.64	0.0083
Delta	0.0107	0.0099	0.0115	0.0098	0.0116	23.04	0.0000
MV	0.0184	0.0153	0.0215	0.0147	0.0221	9.81	0.0000
PTBV	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-8.45	0.0000
PCrat_{t-1}	-0.0011	-0.0012	-0.0010	-0.0013	-0.0010	-15.96	0.0000
PCrat_{t-2}	-0.0009	-0.0010	-0.0008	-0.0010	-0.0008	-15.19	0.0000
PCrat^{*}_{t-1}	0.0012	0.0006	0.0018	0.0005	0.0020	3.44	0.0006
PCrat^{*}_{t-2}	0.0008	0.0002	0.0013	0.0001	0.0015	2.16	0.0307
PCrat^{OTM}_{t-1}	0.0004	0.0003	0.0005	0.0003	0.0005	9.60	0.0000
PCrat^{OTM}_{t-2}	0.0003	0.0002	0.0004	0.0002	0.0004	7.15	0.0000
$\text{PCrat}^{OTM*}_{t-1}$	0.0004	-0.0001	0.0008	-0.0001	0.0009	1.44	0.1503
$\text{PCrat}^{OTM*}_{t-2}$	-0.0003	-0.0008	0.0002	-0.0008	0.0003	-1.03	0.3042

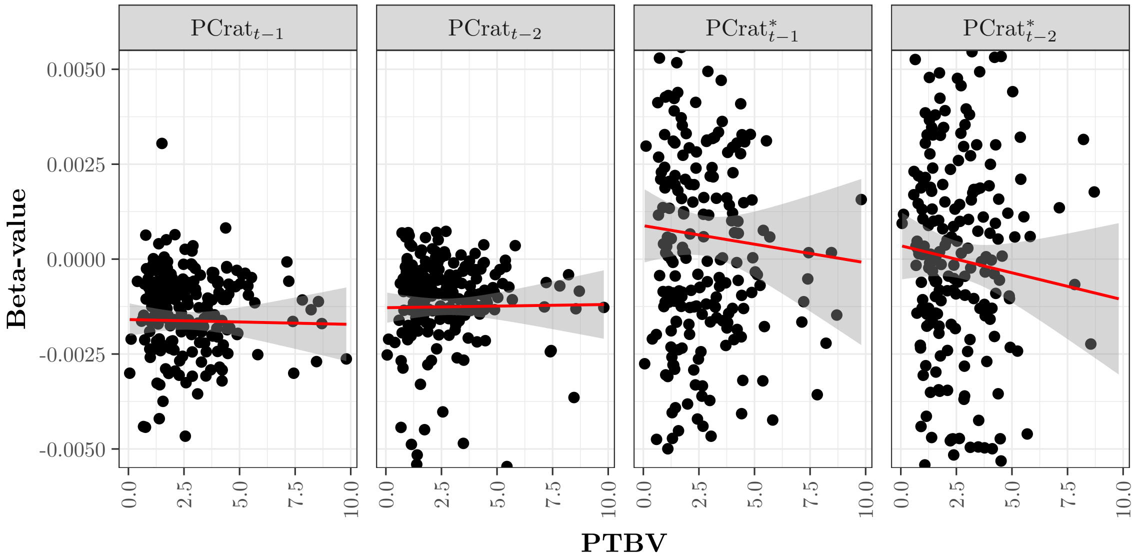
Adj. R^2	F-statistic	P-value	DF
0.0309	762.62	0.0000	260431

The table presents an overview of the regression output for Abn > 1.5 %. The values are calculated from a panel data regression consisting of 232 companies and 1306 trading days observed from mid-2009 to mid-2014. The regression coefficients estimates are presented as point estimates, a 90 % confidence interval and a 95 % confidence interval. The table also presents the T-value and P-value for the individual variables, and the *Adjusted* R^2, F-statistic, P-value and the degrees of freedom (DF) for the total regression.

C Figures

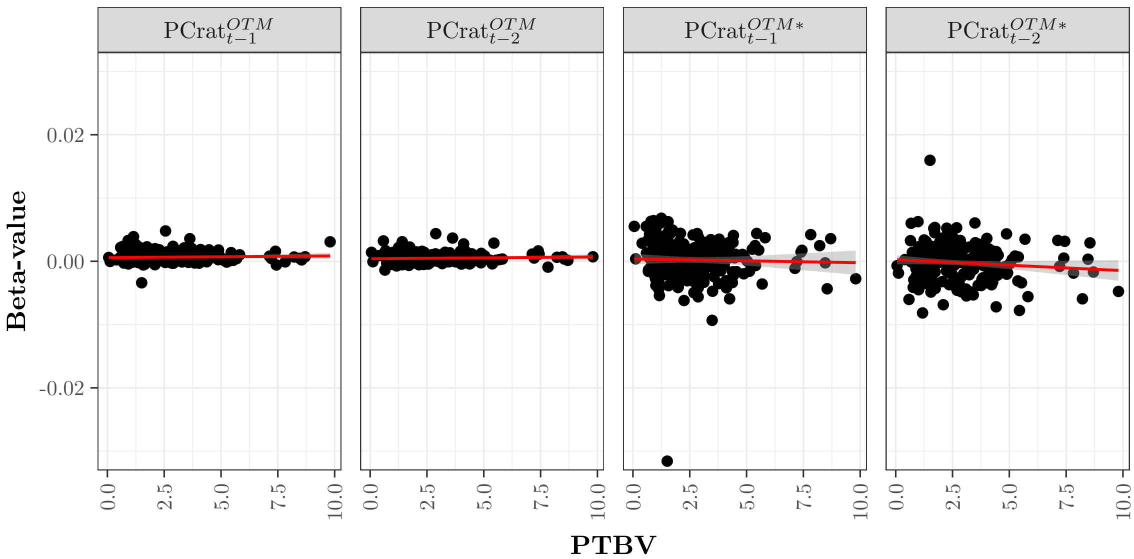
C.1 Time Series

Figure 23: Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs PTBV: Abn > 0.01



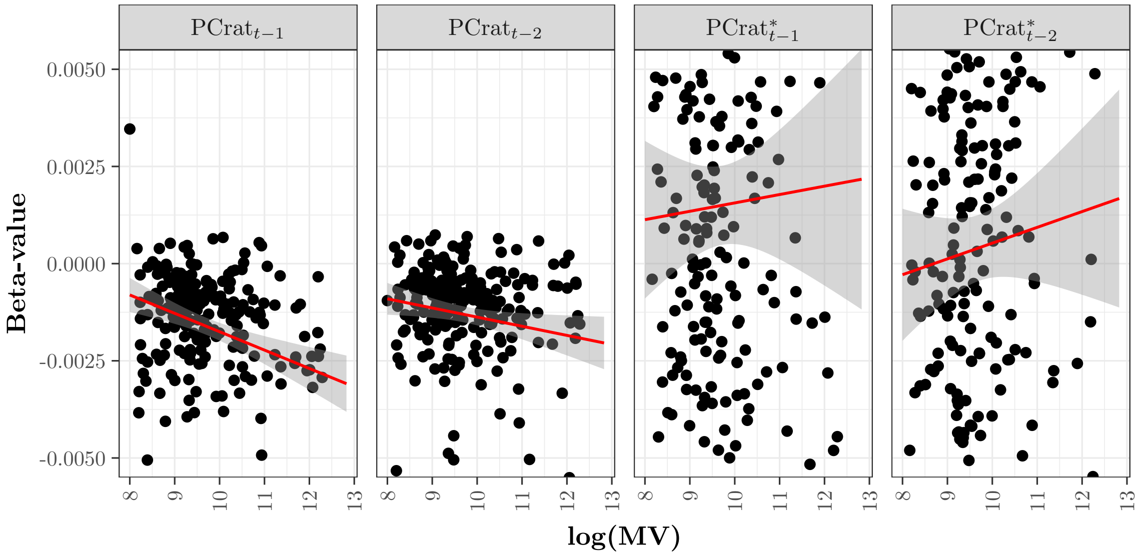
The figure shows a scatter plot of the regression coefficient (β) estimates and PTBV for the variables PCrat_{t-n} and PCrat^*_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.0 %.

Figure 24: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs PTBV: Abn > 0.01



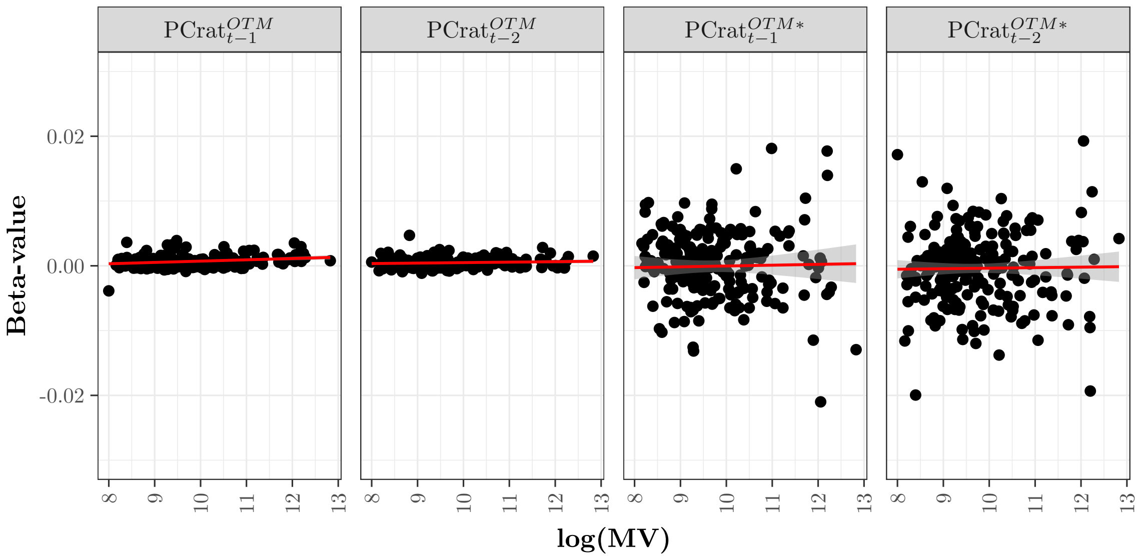
The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and PTBV for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.0 %.

Figure 25: Time-series $\text{PCrat}_{t-n}^{(*)}$ regression coefficients vs log(MV): Abn > 0.015



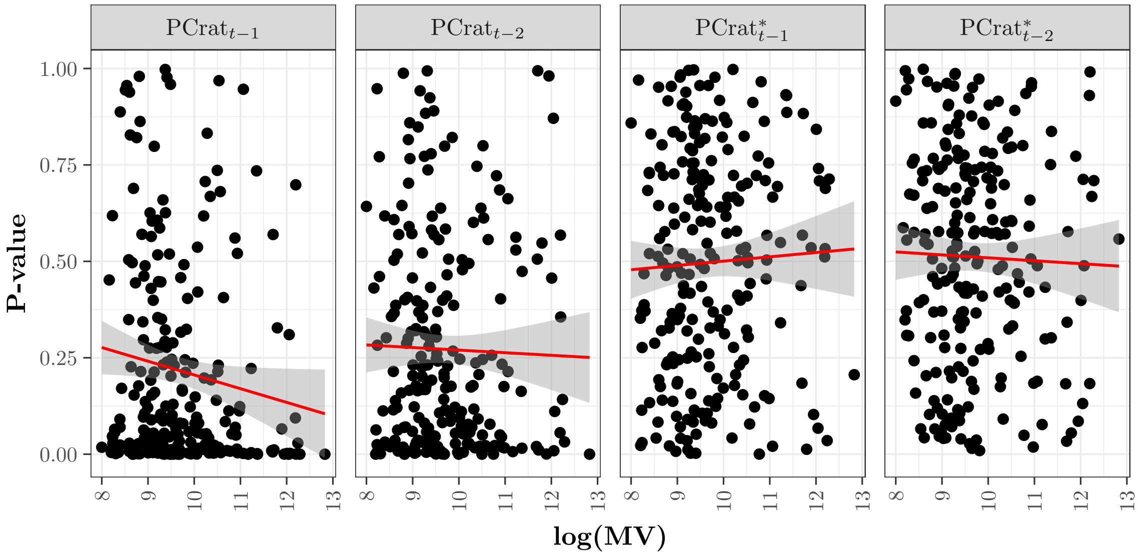
The figure shows a scatter plot of the regression coefficient (β) estimates and the logarithm of market value for the variables PCrat_{t-n} and PCrat^{*}_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.5 %.

Figure 26: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ regression coefficients vs log(MV): Abn > 0.015



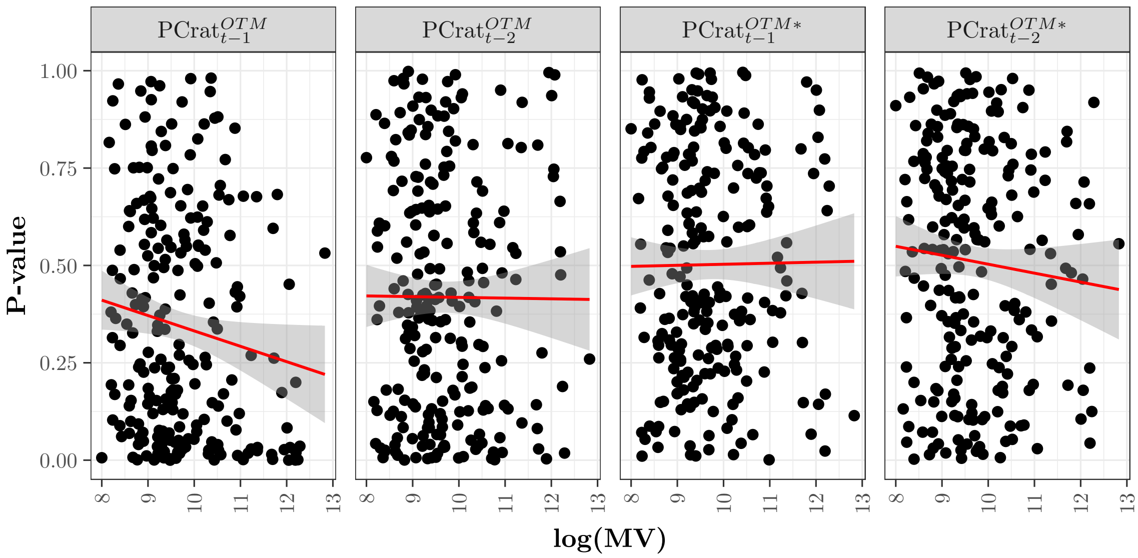
The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and the logarithm of market value for the variable PCrat^{OTM}_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.5 %.

Figure 27: Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs log(MV): Abn > 0.015



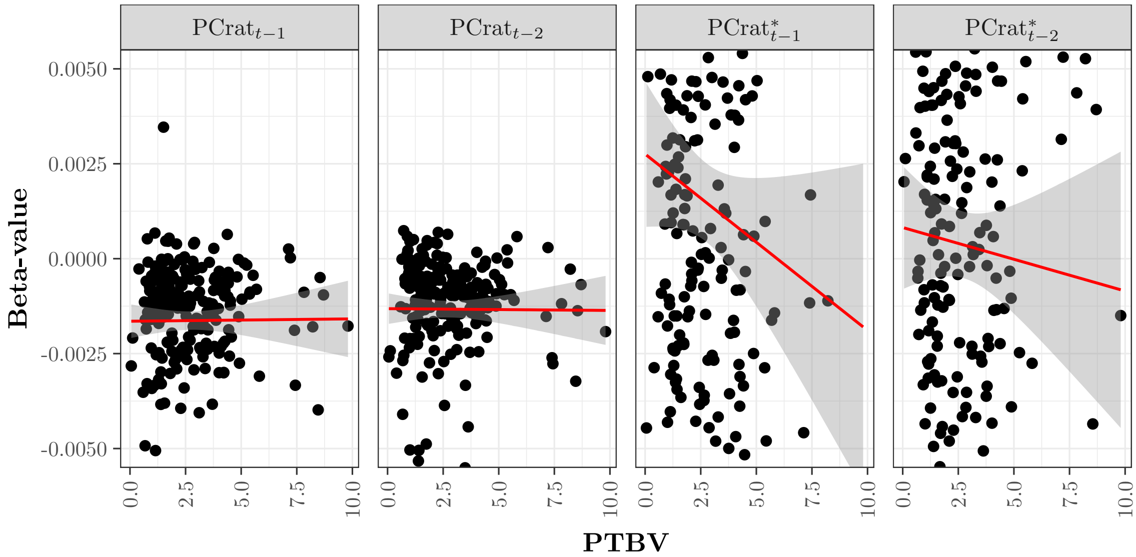
The figure shows a scatter plot of the P-values and the logarithm of market value for the variables PCrat_{t-n} and PCrat^*_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.5 %.

Figure 28: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs log(MV): Abn > 0.015



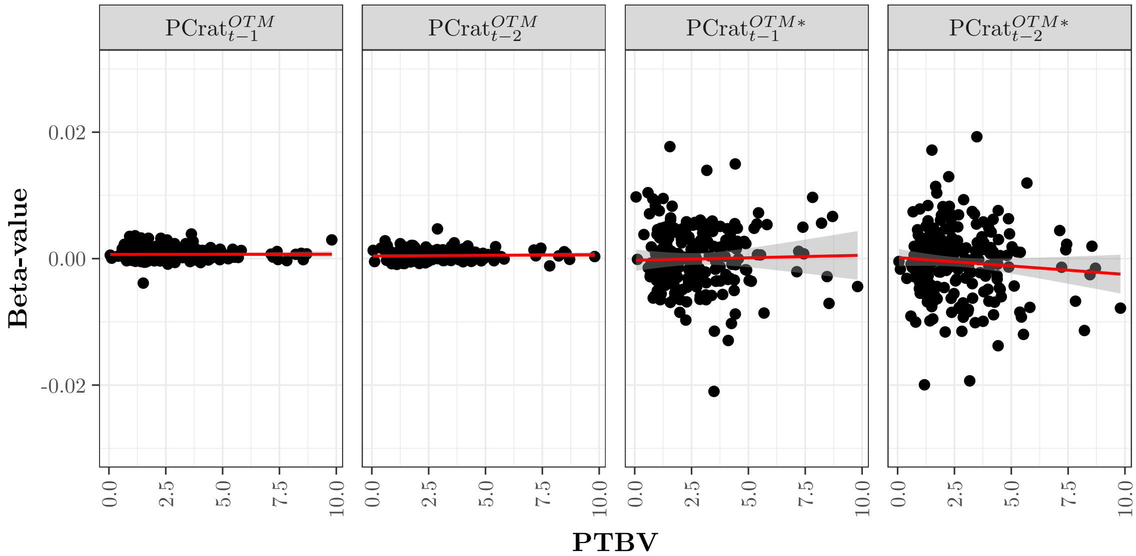
The figure shows a scatter plot of OTM observations for the P-values and the logarithm of market value for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.5 %.

Figure 29: Time-series $\text{PCrat}^{(*)}_{t-n}$ regression coefficients vs PTBV: Abn > 0.015



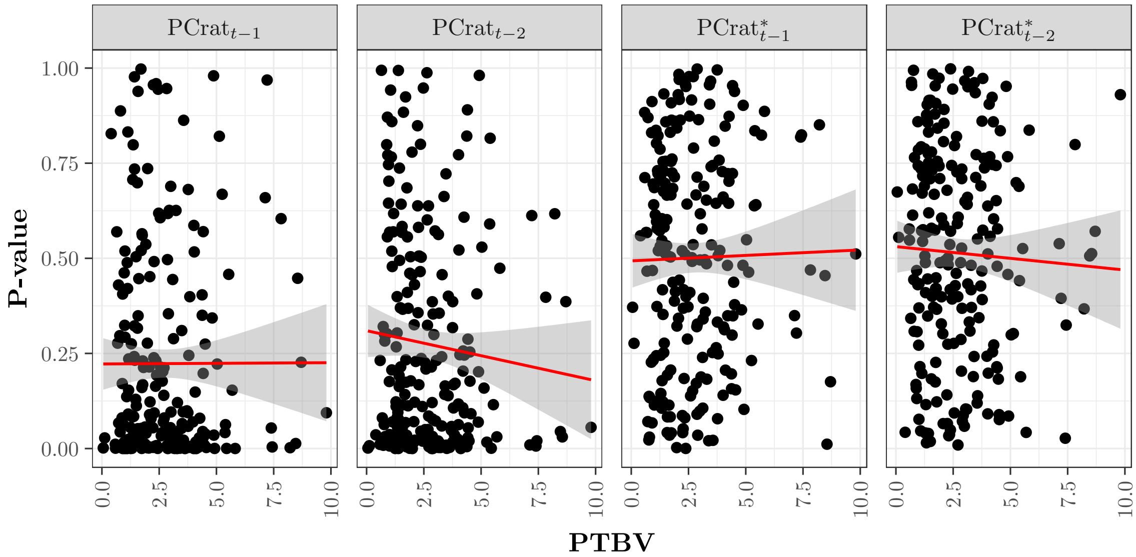
The figure shows a scatter plot of the regression coefficient (β) estimates and PTBV for the variables PCrat_{t-n} and PCrat^{*}_{t-n}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.5 %.

Figure 30: Time-series $\text{PCrat}^{OTM(*)}_{t-n}$ regression coefficients vs PTBV: Abn > 0.015



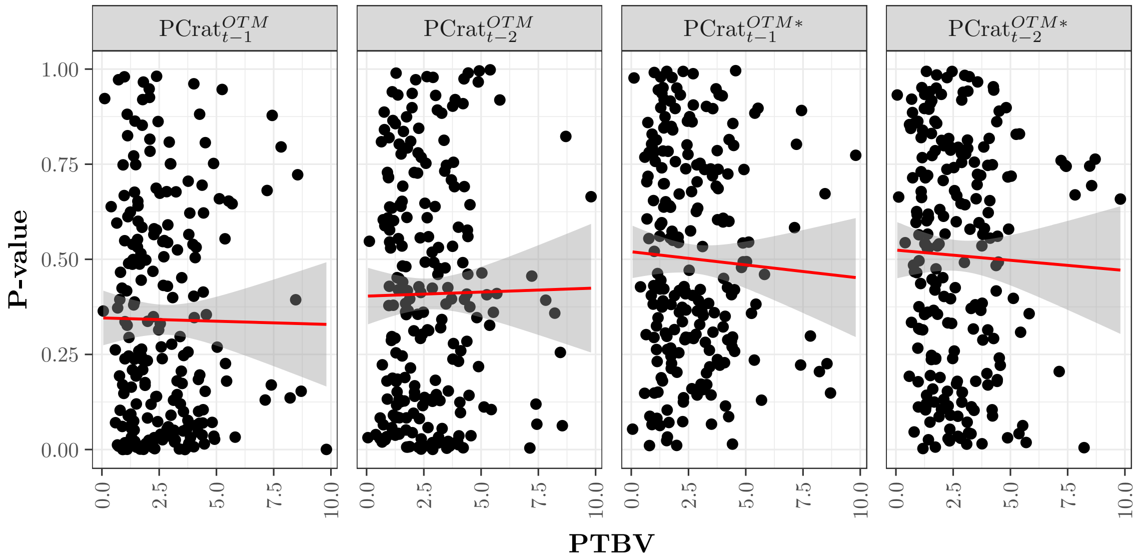
The figure shows a scatter plot of OTM observations for the regression coefficient (β) estimates and PTBV for the variable PCrat^{OTM}_{t-n} and $\text{PCrat}^{OTM*}_{t-n}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.5 %.

Figure 31: Time-series $\text{PCrat}_{t-n}^{(*)}$ P-values vs PTBV: Abn > 0.015



The figure shows a scatter plot of the P-values and PTBV for the variables PCrat_{t-n} and PCrat_{t-n}^{*}. The trend line is computed by linear regression and includes a 95 % confidence interval. Abn > 1.5 %.

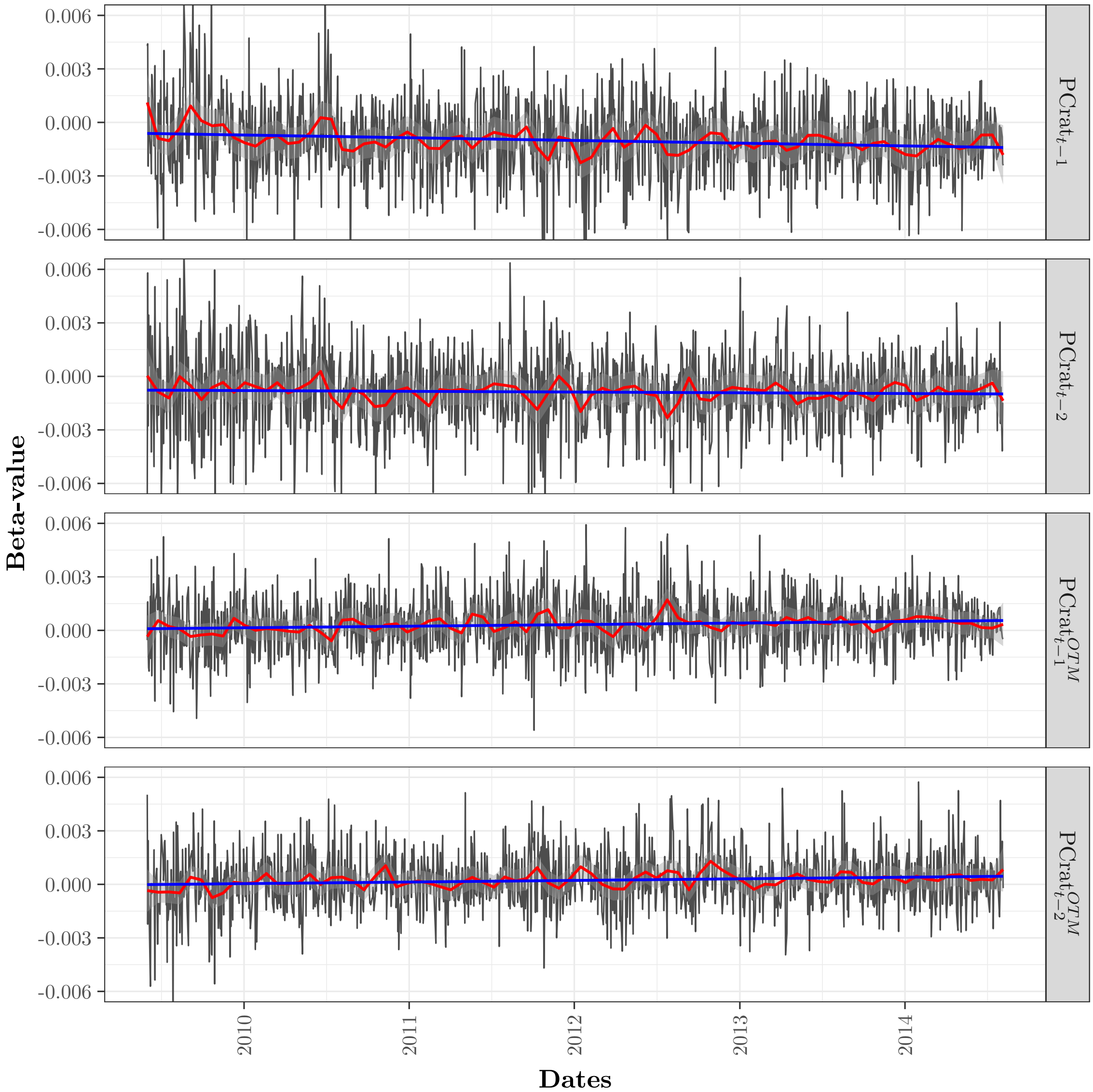
Figure 32: Time-series $\text{PCrat}_{t-n}^{OTM(*)}$ P-values vs PTBV: Abn > 0.015



The figure shows a scatter plot of OTM observations for the P-values and PTBV for the variable PCrat_{t-n}^{OTM} and $\text{PCrat}_{t-n}^{OTM*}$. The trend line is computed by linear regression and includes a 95 % confidence interval, visualized as the grey area. Abn > 1.5 %.

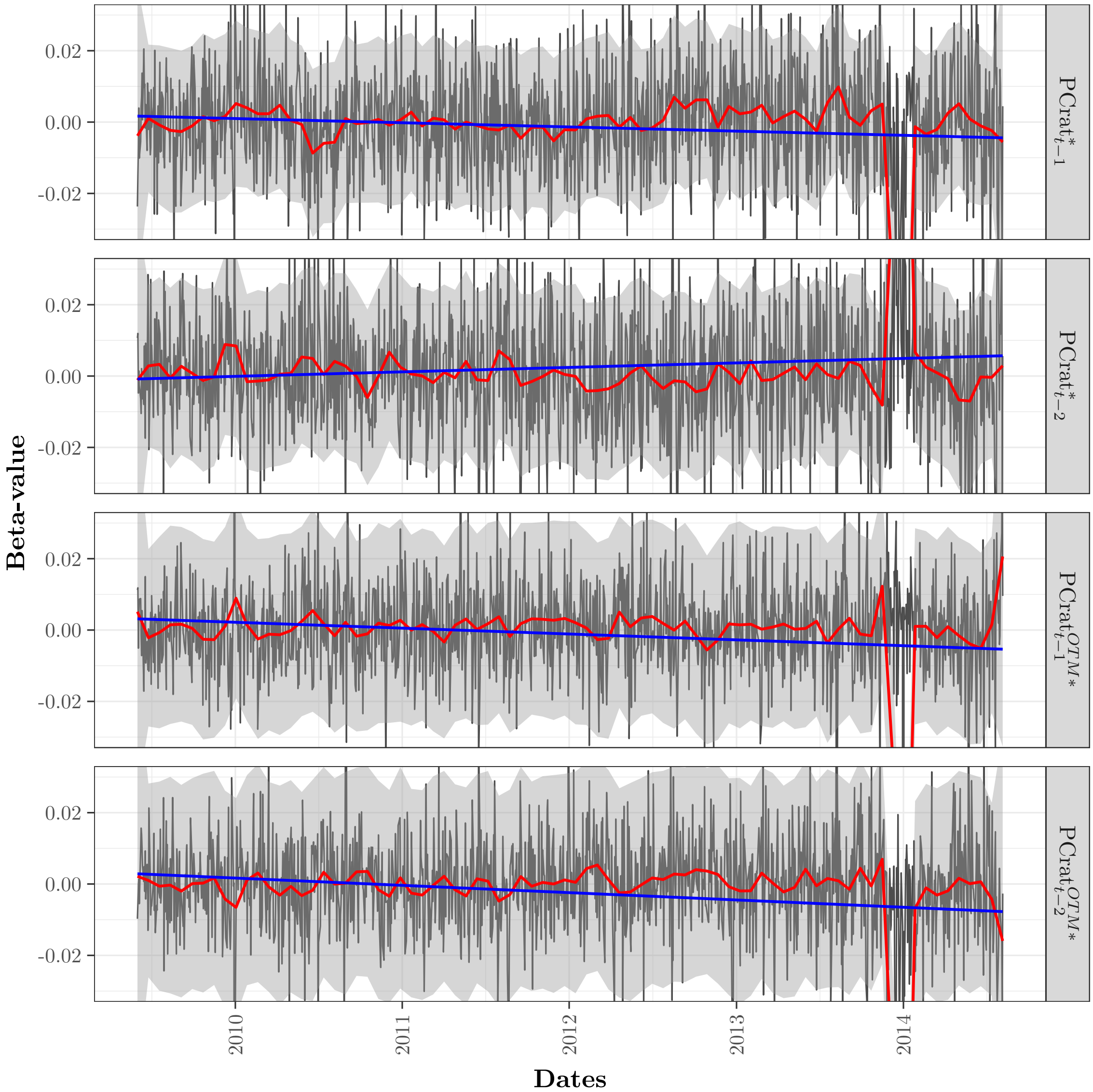
C.2 Cross Sectional

Figure 33: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)}$ regression coefficients vs Dates: Abn > 0.015



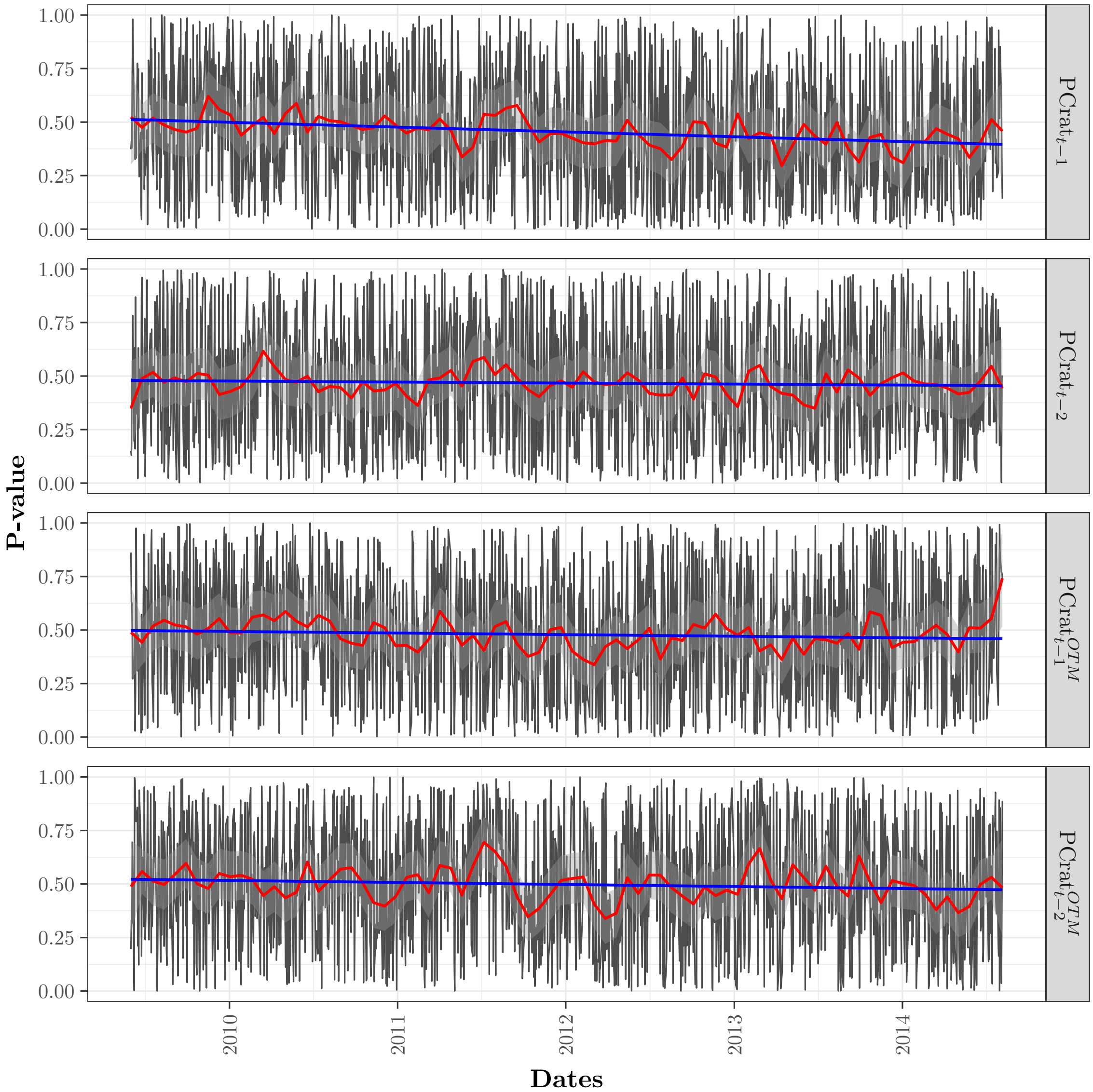
The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n} and PCrat_{t-n}^{OTM} when Abn > 1.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

Figure 34: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ regression coefficients vs Dates: Abn > 0.015



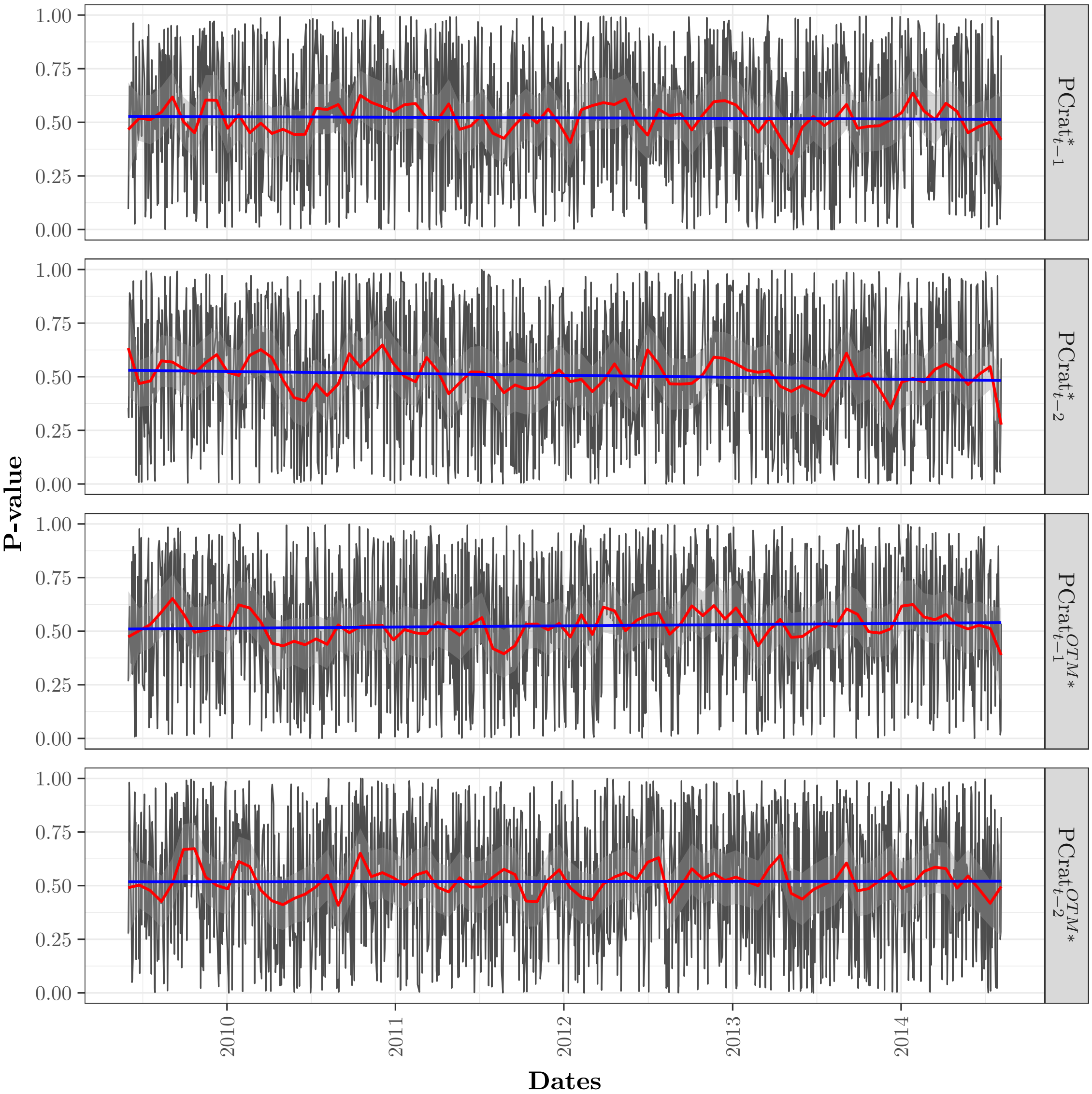
The figure shows a scatter plot of the development of the regression coefficient (β) estimates over time for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$ when Abn > 1.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

Figure 35: Cross-sectional $\mathrm{PCrat}_{t-n}^{(OTM)}$ P-values vs Dates: Abn > 0.015



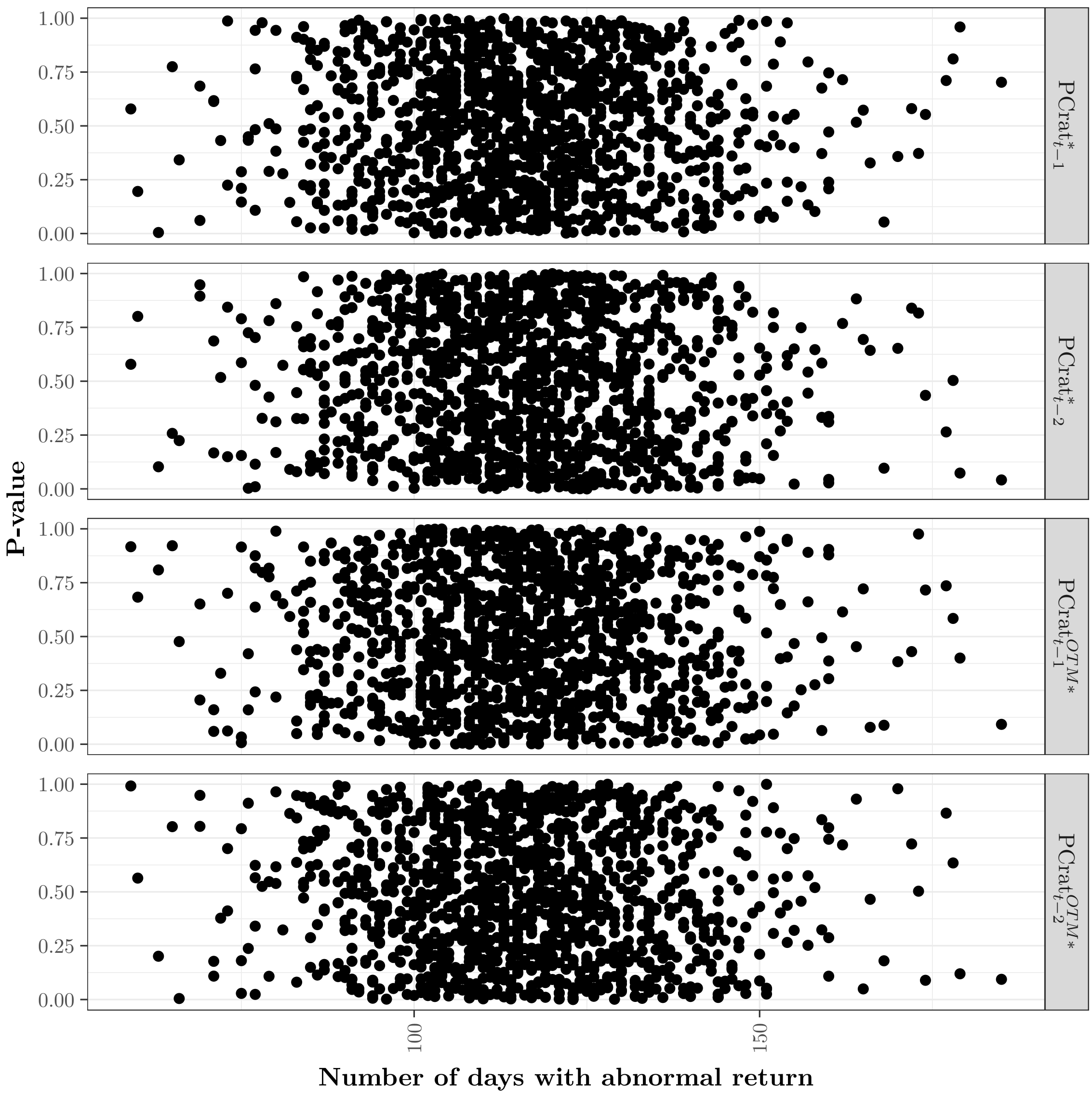
The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n} and $\mathrm{PCrat}_{t-n}^{OTM}$ when Abn > 1.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

Figure 36: Cross-sectional $\mathrm{PCrat}_{t-n}^{(OTM)*}$ P-values vs Dates: Abn > 0.015



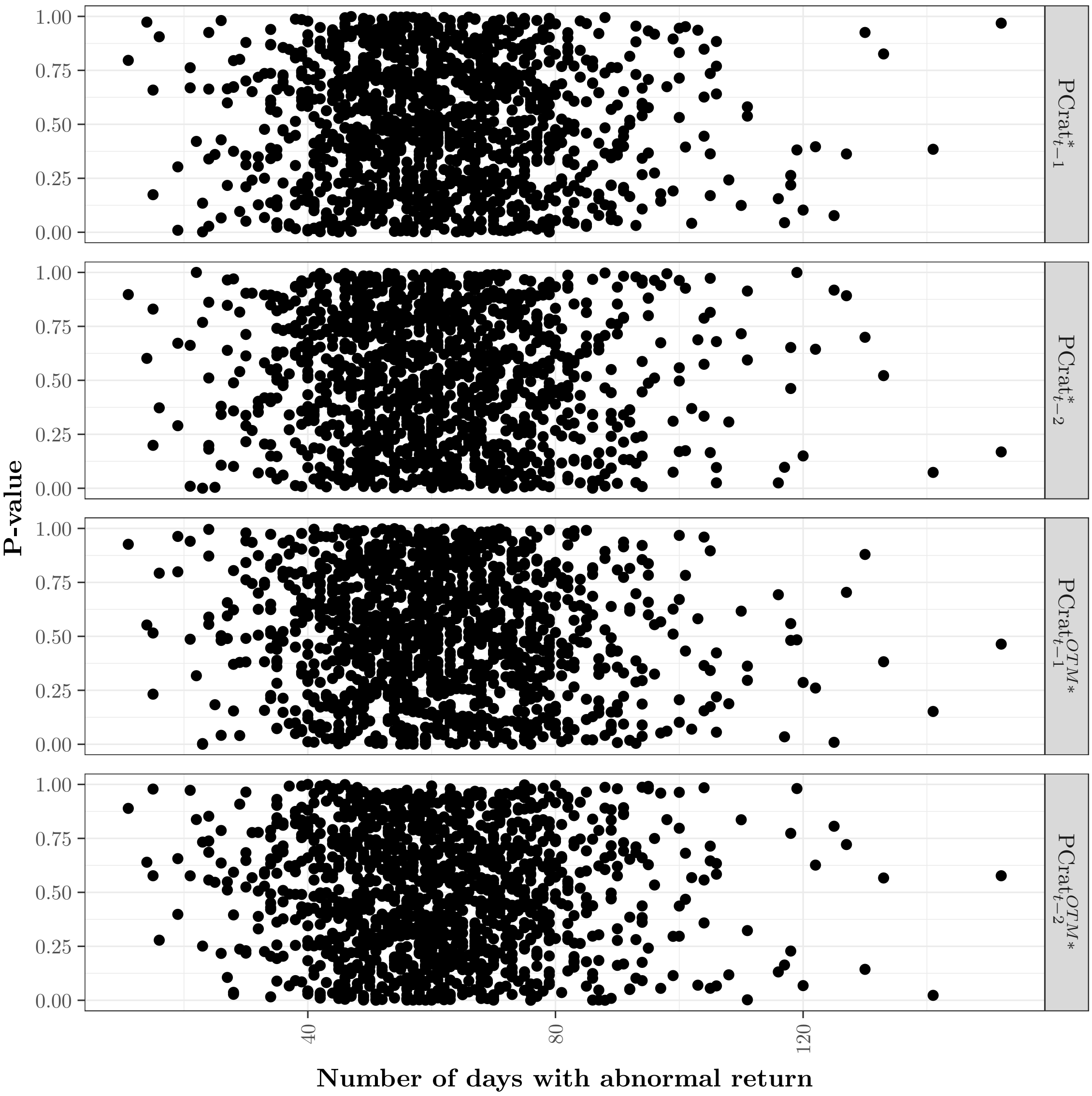
The figure shows a scatter plot of the development of P-values over time for the variables PCrat_{t-n}^{*} and $\mathrm{PCrat}_{t-n}^{OTM*}$ when Abn > 1.5 %. The blue trend line is computed by a linear regression and the red trend line with a 95 % confidence interval is computed by a LOESS regression with the parameter $\alpha = 0.05$.

Figure 37: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.005



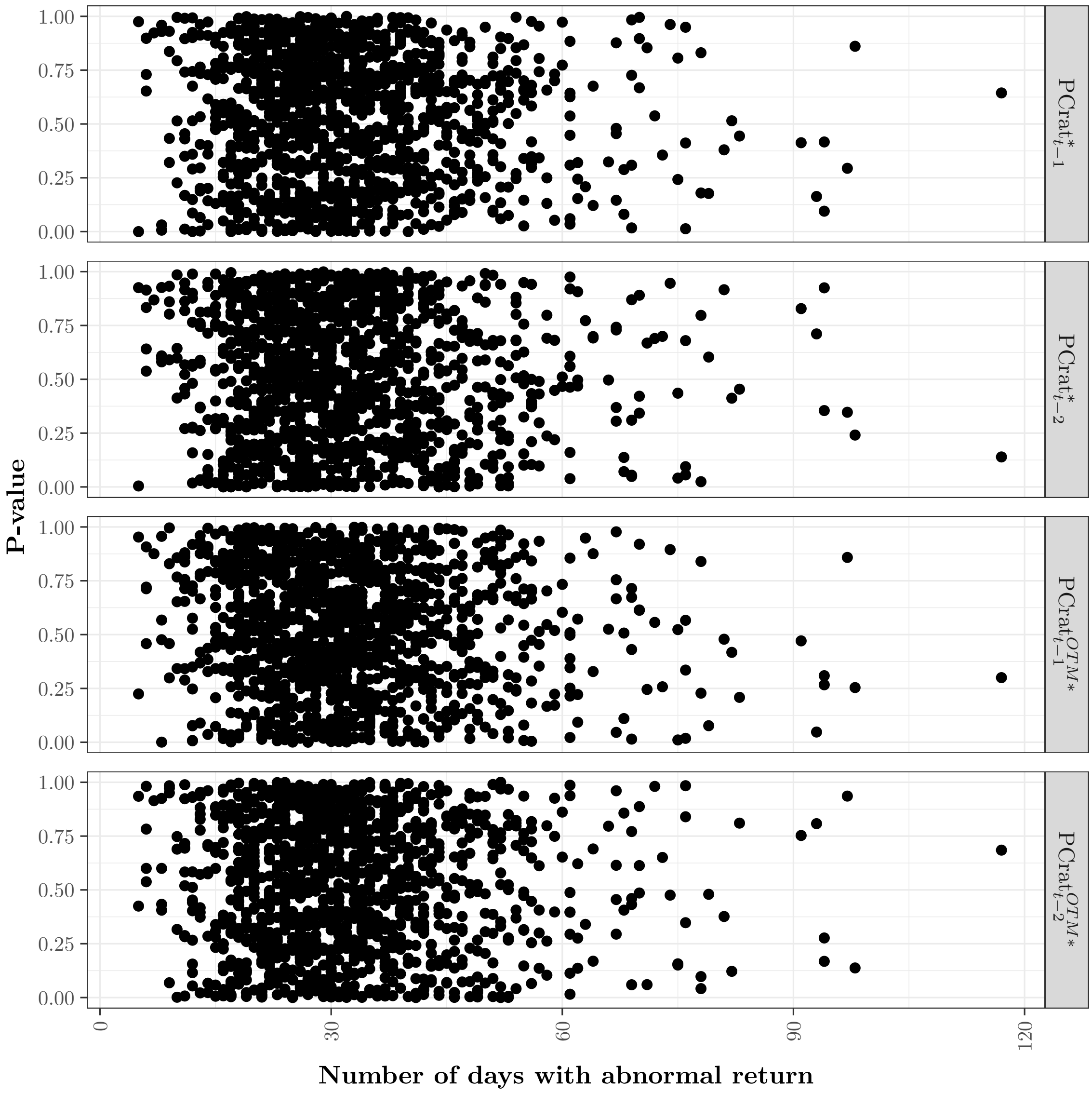
The figure shows a scatter plot of P-values and the number of days with Abn > 0.5 % for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$.

Figure 38: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.01



The figure shows a scatter plot of P-values and the number of days with Abn > 1.0 % for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$.

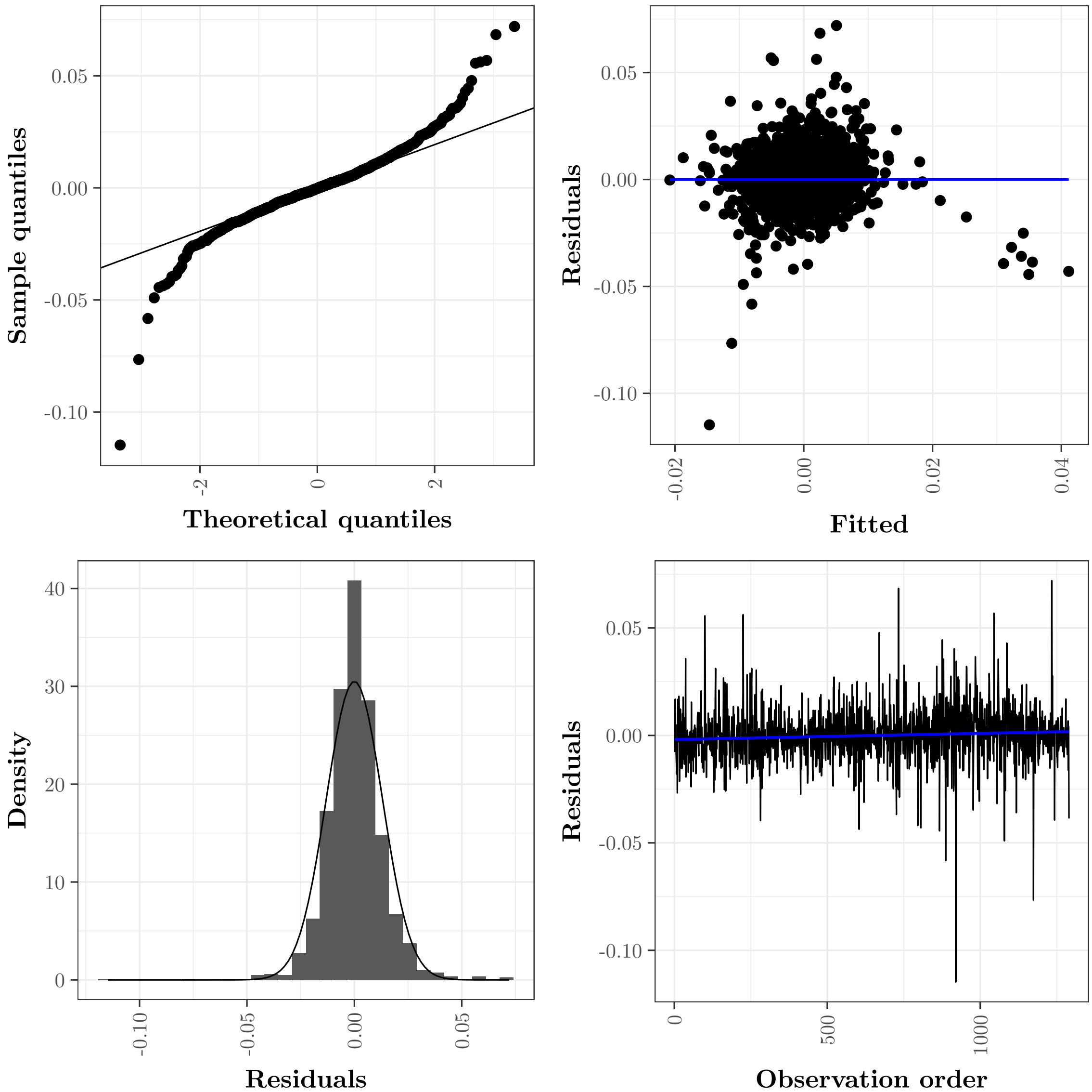
Figure 39: Cross-sectional $\text{PCrat}_{t-n}^{(OTM)*}$ P-values vs Number of abnormal observations: Abn > 0.015



The figure shows a scatter plot of P-values and the number of days with Abn > 1.5 % for the variables PCrat_{t-n}^{*} and $\text{PCrat}_{t-n}^{OTM*}$.

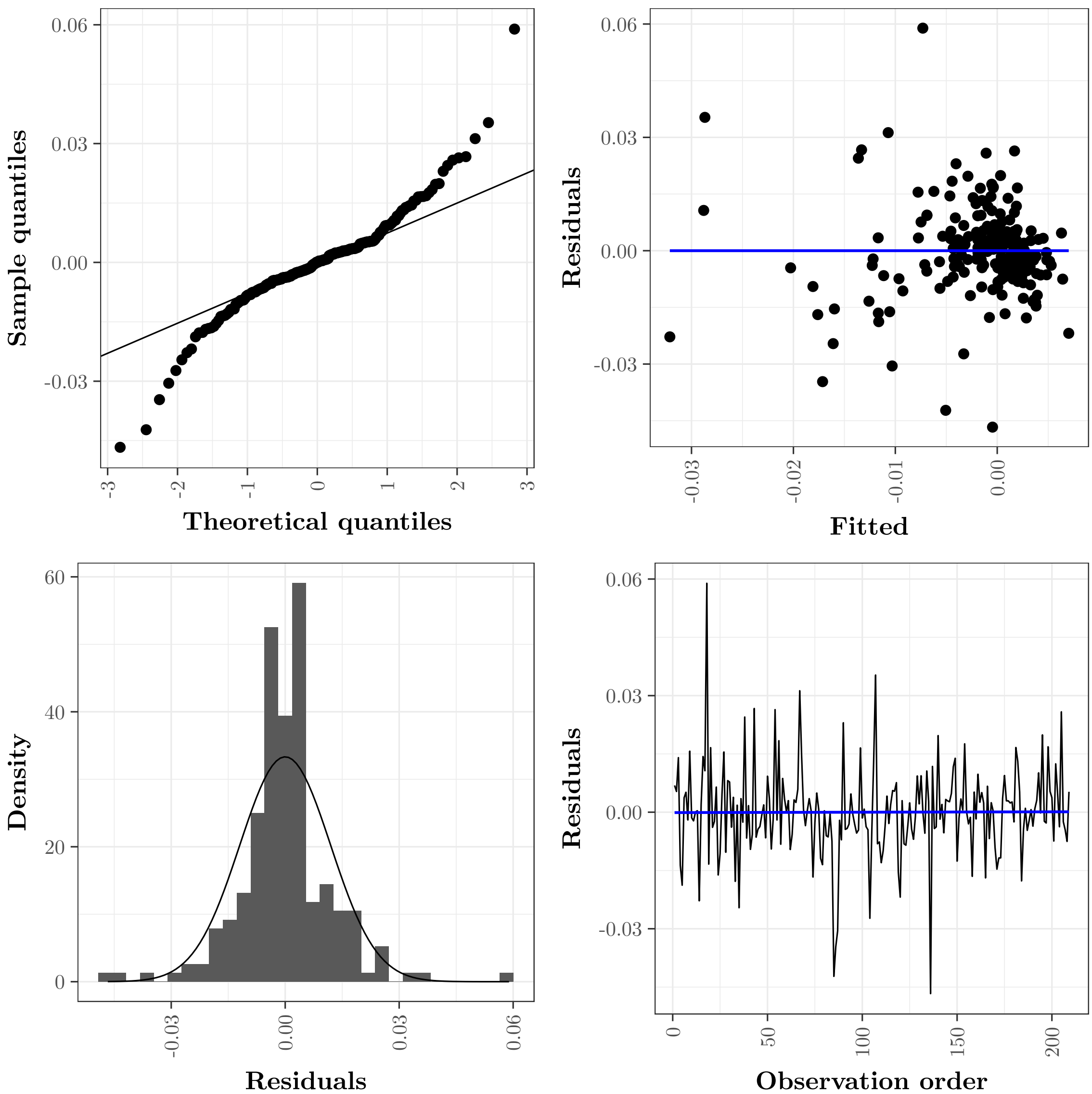
C.3 Residual sample plots

Figure 40: AAPL residual plot, Abn > 0.05



The figure shows four different residual plots for the company AAPL when Abn > 0.05. The four figures are a normal probability plot, a residual histogram, residuals vs. fitted values, and residuals vs. observation order.

Figure 41: 01.02.2013 residual plot, Abn > 0.05



The figure shows four different residual plots for the date 01.02.2013 for Abn > 0.05. The four figures are the normal probability plot, a residual histogram, residuals vs. fitted values, and residuals vs. observation order.

D LOESS

Local regression (LOESS) is most commonly used to enhance the visual information on a scatter plot, by computing and plotting smoothed points. The method was originally proposed by Cleveland (1979), and represents a non-parametric regression that avoids assuming a fixed model structure and solving it. This differs from a linear regression which makes assumptions of the underlying model structure and tries to explicitly solve for the model parameters (Ruan et al., 2007).

The regression is local because the fitting at any point x is determined by a nearest neighbors' algorithm, using subsets of data for each weighted least square fit, where the weights for (x_i and x_k) is large if x_i is close to x_k, and small otherwise. It focuses on estimating the fitted model, and not estimating parameters like α and β found in linear regressions.

When constructing LOESS, the parameter α needs to be set. α is a smoothing parameter and controls for flexibility in the LOESS regression function. Large values of α generates the smoothest output, while smaller α will generate output closer to the data. α is defined on the interval [0,1].

The advantage of LOESS is most importantly that it does not require the specification of a function to fit a model, but only the decision of the size of the smoothing parameter used. However, LOESS requires larger and denser sets of data to produce a good model. LOESS is almost explicitly used for visual purposes, and therefore does not produce a regression function easily interpreted through a mathematical formula.