School of Social Sciences

Master in Business Administration (MBA)

Postgraduate Dissertation

Exploring the Relationship Between Open Source Software Contributions and Firm Performance in Technology Companies

Themistoklis Kyrgos

Supervisor: *Dr. Michail Bekiaris*

Patras, Greece, September 2026

Exploring the Relationship Between Open Source Software Contributions and Firm Performance in Technology Companies

Themistoklis Kyrgos

Supervising Committee

Supervisor:
Michail Bekiaris
Hellenic Open University

Co-Supervisor:
Panagiotis Artikis
Hellenic Open Univercity

Patras, Greece, September 2026

I would like to thank my supervisor Michail Bekiaris for his valuable guidance on this work

I would like to thank Orál for his variform support throughout this process

Abstract

Corporate Open Source Software contribution activity is observable publicly in platforms like GitHub. Prior literature has not evaluated this non-financial signal of OSS engagement in relation to equity return predictability. The current dissertation attempts to examine if corporate OSS activity, measured through the volume of commits in GitHub, can explain stock returns beyond the standard three factor Fama-French model. A sample of 20 firms was examined, of 10 USA and 10 European technology companies, over a period of 10 years from November 2012 to October 2022. Four, augmented with an OSS metric, models were constructed, one with contemporaneous OSS, one with a 1 month OSS lag, one with a 3 months OSS lag window and one with contemporary OSS against 3 months’ future returns. All models were evaluated in pooled OLS and firm fixed effects regressions, for the full, USA and Europe samples. The results show a positive statistically significant association between OSS contribution and excess returns only on firm fixed effects regressions, for the full and USA samples, in the first, third and fourth model. The incremental explanatory power of the introduced factor appears insignificant throughout the cases. This, coupled with the lack of findings across the models, indicate no robust evidence that the OSS commit activity can predict stock returns after adjusted for the Fama-French factors.

Keywords

Open-Source Software, Stock Returns, Fama-French Model

Εξερευνώντας την Σχέση Μεταξύ Συνεισφορών σε Λογισμικό Ανοικτού Κώδικα και Εταιρικής Απόδοσης σε Εταιρίες Τεχνολογίας

Θεμιστοκλής Κύργος

Περίληψη

Η εταιρική συνεισφορά σε Λογισμικό Ανοικτού Κώδικα είναι παρατηρήσιμη δημόσια σε πλατφόρμες όπως το GitHub. Η παρελθοντική βιβλιογραφία δεν έχει αξιολογήσει αυτό το μη-οικονομικό σήμα της ενασχόλησης με το ΛΑΚ σε σχέση με την προβλεψιμότητα της απόδοσης του μετοχικού κεφαλαίου. Η παρούσα διπλωματική επιχειρεί να εξετάσει αν η εταιρική δραστηριότητα στο ΛΑΚ, μετρημένη μέσω του όγκου των commits στο GitHub, μπορεί να εξηγήσει μετοχικές αποδόσεις πέραν του καθιερωμένου μοντέλου Fama-French των τριών παραγόντων. Ένα δείγμα από 20 εταιρίες εξετάστηκε, 10 τεχνολογικές εταιρίες από ΗΠΑ και 10 από Ευρώπη, για μία περίοδο 10 χρόνων από τον Νοέμβριο του 2012 έως τον Οκτώβριο 2022. Τέσσερα μοντέλα, επαυξημένα με μία μετρική ΛΑΚ, κατασκευάστηκαν, ένα με ταυτόχρονη ΛΑΚ, ένα με μετρική ΛΑΚ υστέρησης ενός μηνός, ένα με μετρική ΛΑΚ παραθύρου υστέρησης τριών μηνών και ένα με ταυτόχρονη ΛΑΚ έναντι τριών μηνών μελλοντικών αποδόσεων. Όλα τα μοντέλα αξιολογήθηκαν σε συγκεντρωτικές και σε εταιρικά συγκεκριμένες παλινδρομήσεις, για το συνολικό, το ΗΠΑ και το Ευρωπαϊκό δείγμα. Τα αποτελέσματα δείχνουν μία θετική στατιστικά σημαντική συσχέτιση μεταξύ συνεισφοράς σε ΛΑΚ και υπερβάλλουσες επιστροφές μόνο σε εταιρικά συγκεκριμένες παλινδρομήσεις, για το συνολικό και ΗΠΑ δείγμα, στο πρώτο, τρίτο και τέταρτο μοντέλο. Η αυξητική δύναμη επεξήγησης του εισαχθέντος παράγοντα φαίνεται ασήμαντη για όλες τις περιπτώσεις. Αυτό, μαζί με την έλλειψη ευρημάτων για όλα τα μοντέλα, υποδεικνύουν μη εύρωστα στοιχεία ότι η δραστηριότητα σε commits σε ΛΑΚ μπορεί να προβλέψει μετοχικές αποδόσεις μετά την προσαρμογή των παραγόντων Fama-French.

Λέξεις – Κλειδιά

Λογισμικό Ανοιχτού Κώδικα, Απόδοση Μετοχής, Μοντέλο Fama-French

Table of Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

CAPM	Capital Asset Pricing Model
GSVI	Google Search Volume Index
n.d.	no date
OLS	Ordinary Least Squares
OSS	Open Source Software

1. Introduction

1.1 Background and Context

Open-Source Software (OSS) is a kind of software with open licensing, that allows its users to freely review, execute and alter its source code (Open Source Initiative, 2024). OSS has a long history, resulting from the early programming communities (Open Source Initiative, 2025) and became more formalized in the 1980s by the Free Software Foundation (Free Software Foundation, n.d.). Since then, OSS software has expanded in use and contributions, slowly at first but faster lately, both by individuals as well as by technology companies. Nowadays OSS is no longer a purely community driven phenomenon but also a corporate technology strategy (Fitzgerald, 2006; West & Gallagher, 2006; Dahlander & Magnusson, 2008). Early pioneers, like Red Hat and Google, have been joined by other time opponents, like Microsoft, as companies who both use and contribute to OSS (Red Hat, n.d.; Google Open Source, n.d.; Linux Foundation, 2016).

OSS contributions by firms can take different forms, not just source code but also design, documentation, support, as well as project infrastructure and governance (Finley, 2023; Linux Foundation, n.d.). Technology firms contribute to OSS for different reasons, mostly to improve software ecosystems that they depend on and to make their own products more accessible to larger market shares, but also to reduce development costs through collaboration, build reputation and foster innovation (Nagle, 2018; Henkel, 2006). OSS contributions by firms may not be fully captured in financial statements, like other forms of more traditional R&D, and this also stems from the fact that the value of OSS contributions can not be easily quantified and measured (Association of Chartered Certified Accountants [ACCA], n.d.; Lev, 2003). As a result, OSS contributions can be an under-reported form of innovation investment.

1.2 Research Problem

Financial markets and investors evaluate firms using different signals and metrics, including profitability, revenue growth, innovation and marketing strategies, R&D spending and patent applications (Hall et al., 2005; Chan et al., 2001). While OSS contributions may have economic value for a firm, their relationship with financial outcomes remains under-explored and their value as a market signal less studied (Nagle, 2018; Nagle, 2019).

OSS contributions, through their various forms, can be a sign of an innovation culture, a highly technical leadership and an extroverted strategy, things that usually have a positive impact in a firm's value (Nagle, 2018; Alexy et al., 2013). On the other hand they can also be costly in employee time, organizational resources and infrastructure allocation, presenting as costly activities of uncertain benefits and unrealized cost-cutting opportunities to investors (West & Gallagher, 2006; Linux Foundation, n.d.). All these make it unclear whether OSS contributions are positively, negatively or insignificantly associated with a firm's value and present a research opportunity that this dissertation will attempt to examine.

1.3 Research Scope

OSS contributions can take many forms. This dissertation uses commit data from GitHub as they present the most basic and quantifiable form of OSS contributions. GitHub has been the de facto repository for OSS over the last decade, with more than 2.8 million OSS repositories, small and big, using it for storing their source code and documentation, and handling their everyday development activities (Hoffa, 2016; Kalliamvakou et al., 2014). Commits are, usually small, source code incremental contributions that form the basis of all development activity in modern software projects (GitHub, n.d.-a; GitHub, n.d.-b). Since documentation is usually also stored along the source code, commits are also used to built the documentation of a software project (Finley, 2023). Other forms of contribution on the development of a software project, like pull requests and issues discussion and resolution, stem from or result to commits, so commits can be considered a robust proxy of OSS activity (Kalliamvakou et al., 2014).

Apart from OSS contributions that take many forms, a firm's economic performance can also be evaluated through various measures, like return on equity/assets, valuation multiples, growth ratios and various other accounting-based measures. Given the high pace that technology firms operate in and the spotlight they usually operate under, this dissertation will use the stock returns as the performance measure studied. A firm's stock price is observable in shorter intervals than other economic measures and it reflects the market's expectations of the future firm value taking into account all available information about it, including all other economic measures, under the efficient market perspective (Fama, 1970).

1.4 Aims and Objectives

The aim of this empirical research is to use a structured asset-pricing framework, in order to examine whether OSS contribution activity by publicly traded technology firms can be associated with stock returns, after controlling for market, size and value factors, using the 3 factor Fama-French model (Fama & French, 1993).

In the following chapters, academic literature on OSS, open innovation and firm performance will be reviewed. A dataset of OSS contributions and stock returns will be compiled from openly available sources. A three factor Fama-French model will be used as a base model, augmented with OSS contributions signals. The model results will be studied in order to determine if OSS activity can provide additional explanatory power, beyond the Fama-French factors, across two regions (USA and European). The OSS signal will be used both contemporarily and in lagged form, to study if it holds any predictive qualities. Finally the dissertation will study whether any relationship between OSS activity and stock returns is consistent across firms or driven by specific companies.

2. Literature Review

2.1 Corporate Contributions to OSS

The contributions to OSS from corporate entities seem to be an interesting an initially paradoxical phenomenon. The main goal of a firm is to create value for its shareholders and that usually means it seeks to protect any competitive advantage it might have, with many firms going to great lengths to protect their intellectual property. OSS on the other hand allows public sharing of source code and its use, distribution and modification by any individual or organization. The question that arises here is why firms that seek to make profit contribute their resources into something that would benefit others, including possible competitors. The present dissertation attempts to look at this question by examining whether the OSS contributions of a firm, that are a result of a strategic decision and involve real costs, can contain information about the firm's economic performance and value.

Early work of Lerner and Tirole (2002) proposed theories that explain the OSS participation at an individual level. These include career development, peer recognition, reputation building and continuous learning. These individual level explanations can also be relevant at the firm level, since firms employ developers from the same pool as the OSS projects. Corporate investment in OSS can then be seen as strategy for investing in human resources, reputation and learning. In later work, Lerner and Tirole (2005) widen the discussion by placing the OSS contributions in a technology sharing framework. They argue that OSS strategies can make economic sense when considering the benefits from complementary assets, the weakening of competitors' products, the reliability of a more collaboratively produced software and less barriers for adoption of open software. All these can provide a basis for seeing the OSS activity as something more than a philanthropic activity but also a strategic form of innovation investment.

West and Gallagher 2006, also examine the reasons for which corporations get involved with OSS, under the frame of open innovation. For them, what seems as a paradox can be explained by considering the value that can be captured by the firms. For technology firms this can be done through various ways like using shared software infrastructure and by providing complementary assets and services. Value can also be captured through influencing technical standards which can lead in establishing a leadership position in the ecosystem and gaining reputation. The reputation gains can then be used for attracting developer talent and reducing development costs by pooling resources with the wider community. The majority of these benefits can be considered indirect and presenting more on the long term.

Henkel (2006) examines the concept of selective revealing by studying the contribution of firms to the embedded Linux. Firms appeared to reveal some parts of the internally developed code while keeping others proprietary. This means that openness and secrecy can both be part of a technology strategy. Firms may use open source when they need external contributions and feedback while retaining control of more important and commercially sensitive parts of their technology. This would mean that the GitHub commit activity observed in this dissertation could be only the visible part of a firm's technology achievements.

Dahlander and Magnusson (2008) examine how corporations interact with open source communities. They identify different ways that firms can utilize these communities for example aligning communities with their corporate objectives, accessing the pooled knowledge and talent or incorporating the community output into their own processes. This can present a limitation on the empirical design of this dissertation, since a firm can extract value from OSS without this being reflected to the number of commits attributed to it.

Nagle (2018) provides some arguments as to why contribution can matter more than just simple OSS usage. The external feedback, that contribution exposes a firm to, can be valuable as a source of learning and improvement. His findings suggest that firms who contribute to OSS can capture more value than firms who only use OSS and are only benefiting from it by using it as a public good. This supports the main idea this dissertation is based upon, that the contribution activity contains information beyond OSS adoption, even though this might be a more long term effect than the timeframe studied in this dissertation.

In a followup study, Nagle (2019), examined the OSS activity and its relationship with firm productivity. The findings suggest that OSS usage can be associated with excess productivity, for firms that possess complementary capabilities. This means that a firm can reap more benefits from OSS if it has the proper organizational, technical and human resources needed to exploit OSS effectively, which is usually the case with technology firms that will be includded in the current study.

A recent work by Conti et al. (2025) has shown that the engagement with OSS can be a visible signal for financing markets. The study found a positive link between GitHub activity of startup firms and the likelihood to attract funding. This supports the idea that OSS activity can serve as a visible signal for technical capability and future potential, at least for the early stages and the small scale of a new company.

Regarding the measurement and quantification of the OSS contribution activity, studies like Calderon et al. (2022) have used GitHub data, including repositories, contributions and commits in order to estimate the cost of OSS innovation at around $36 billions for 2019. Another framework developed by Korkmaz et al. (2024) with the same purpose of estimating the OSS innovation cost through data from GitHub, gave a similar estimate of $38 billions for 2019. These studies support the idea that GitHub data can be used in measuring OSS activity. The present dissertation adapts this idea, from an accounting of innovation context, to a financial market value context.

On the other hand, Young et al (2021), argue about the limitation of such contribution measures. They caution that code changes like those captured by commits and pushes might contain a subset of OSS contribution which can also take the forms of bug reporting, design, infrastructure, community management, public outreach and project management. These forms of contributions can be difficult to quantify or estimate, even for the contributing firms themselves, let alone for external observers. This kind of limitation in qualitative data on the one hand, and the availability of the GitHub data on the other hand, was one of the main reasons that this dissertation used the commit data in estimating the OSS contributions of a firm.

OSS contributions can be viewed as a form of intangible investment since they involve labor and value creation that might not be fully capitalized on financial statements. Chan et al (2001) examined how the stock markets value R&D expenditures and found a somewhat complex relationship. Firms with high R&D do not always perform better than those with less R&D expenditures, but there are instances where the market undervalues firms with high R&D expenditures, especially those having recent bad performance. Eberhart et al (2004) found that an increase in R&D by a company is associated with positive performance and excessive returns in the long term. This can suggest that the markets might not recognize this sort of signal in a timely fashion, which can be a limitation for the current study.

2.2 Fama-French Asset Pricing Framework

In order to examine whether the OSS contribution is related to stock returns, this dissertation uses as its baseline the Fama-French framework for asset pricing. Fama and French (1993), identified three factors that can explain average returns in stocks, the overall market factor, the firm size and the book-to-market factor. Their work extends the Capital Asset Pricing Model, by recognizing that the average returns can not be explained by the market beta only, but also by systemic factors that can affect small firms and high book-to-market firms' returns. The same work also identified two factors that explained bond returns, maturity and default risk, but these are not relevant for the stock market returns examined in the current study. The Fama-French model was selected as the baseline for this dissertation because it can provide a solid benchmark for estimating whether an additional variable can explain returns beyond the established factors. The stricter test this framework provides is important especially for technology firms, which may be exposed to external factors like market cycles, firm size and growth characteristics, that need to be accounted for separately from the OSS contribution activity.

Fama and French (2015) later extended this three factor asset pricing model for stocks into a five factor model, adding a factor for profitability and another one for investment patterns. Fama and French (2017) also tested their model internationally, including markets of Europe, North America and Asia. This is important for this dissertation as it also examines technology stocks outside of the USA and needs regionally relevant model factors. This dissertation uses the three factor Fama-French model, despite the existence of the extended five factor model, because it deemed it enough to test whether there is a correlation between OSS contribution and stock returns, by comparing the baseline and the augmented model. Another reason for the usage of the three factor model is because other relevant works discussed below use it as well, to test the importance of other introduced factors.

There is a body of related finance literature that examines the association or predictability of various public external signals with stock returns. Da et al. (2011) examined the Google Search Volume Index as a proxy signal for investor attention. They found that increased search volume can be associated with price pressure in the short term, suggesting that investor attention can affect prices temporarily. Bank et al. (2011) also studied the GSVI and it's association to stock returns, in German stocks. They found that higher search activity can be associated with increased liquidity and trading activity. Bijl et al. (2016) found that the relationship between search volume and stock returns can be negative, suggesting that high search volume can be a result of negative news for a company.

Nguyen et al. (2019) also examined the use of a GSVI signal as part of an augmented five factor Fama-French model, applied to stocks from Asian markets. They found that there is value in the extra factor but that negative publicity might be more influential to a firm's stock returns than positive publicity. Ekinci and Bulut (2021) did a similar study with a GSVI signal for augmenting a Fama-French model, applied in BIST 100 stocks. They found that the additional factor can be contemporarily associated with excess returns, but it offers little predictive value over future periods. All these works are relevant to this dissertation because they demonstrate that a non-financial signal can contain information about the behavior of a stock. In these cases the signal is external to the firm, whereas in our case the signal originates internally.

Summarizing the literature review, it has been established that OSS contribution is economically meaningful and strategically motivated. Moreover, OSS activity can be measured by publicly available data like GitHub commits. Apart from that, relevant research has already provided a framework for testing whether a public signal has any explanatory power, beyond the established factors. However, previous OSS studies have only examined the OSS contributions more in terms of aggregate economic value and prior asset pricing studies have not used any signals related to OSS contribution activity. The combination of OSS signals and asset pricing has so far been an unexplored area, the exploration of which motivated the work of this dissertation.

3. Research Methodology

3.1 Research Approach and Design

This chapter will explain the research methodology used to examine whether the OSS contributions are associated with stock returns in publicly traded technology companies. The data sources and the data used will be presented, along with the software used. The three factor Fama-French model that was used as the basis for factor adjustment will also be presented. The variables and models' constructions will be explained for the baseline and augmented models, for full, regional and firm-specific analysis. Finally, some limitations of the methodology will be laid out.

The methodology uses a monthly period timeframe even though shorter period data are available for all stock prices, commits data and Fama-French factors. Given the time periods that technology companies operate in and software projects develop and deploy, a monthly timeframe was considered of good enough granularity for examining the contributions to OSS and potential created value reflected to the stock price, as well as having enough data points over the 10 year period examined. Shorter timeframes, like daily or weekly, were deemed too noisy and larger timeframes would leave too few data points to reach any useful conclusions.

The methodology used in this dissertation is quantitative and empirical. It is also an observational study, and not experimental, and its results are interpreted as association evidence rather than evidence of causation.

3.2 The Fama-French Three Factor Model

According to CAPM, the expected excess returns of a stock can be explained by a factor of the market excess return. In a formula form that would mean:

$$R_{i,t} - RF_t = \alpha_i + \beta_i (RM_t - RF_t) + \varepsilon_{i,t}$$

The model does not take into account any other factors that might affect the excess returns of a stock such as the effect of the size of a firm or the effect of the relative value of the firm.

In a paper they published in 1993, Fama and French identified two more factors apart from the excess market returns in equity markets, a size factor and a book to market factor.

Their model also included two more factors for bond markets, which in our case are not relevant since we don’t investigate bond markets.

The three factor Fama-French model is given in the following formula form:

$$R_{i,t} - RF_t = \alpha_i + \beta_{1,i} MktRF_t + \beta_{2,i} SMB_t + \beta_{3,i} HML + \varepsilon_{i,t}$$

for stock i in period t, where R denotes the returns on the asset, RF is the risk free market ratio and their difference is the excess return of the asset which is the dependent variable of the model. MktRF is the market excess return on top of the risk free rate (RM – RF), which, Kenneth French’s data library, defines as the weighted return of eligible firms minus the one month US Treasury bill rate.

The SMB, or the Small-Minus-Big, is a size factor that captures the difference between the small cap and the large cap stocks. Practically, Kenneth French constructs the SMB values as the average returns on three small-stock portfolios minus the average returns on three large-stock portfolios. The HML, or High-Minus-Low, is a factor attempting to capture the difference between high book-to-market stocks (value stocks) and low book-to-market stocks (growth stocks). Kenneth French constructs the HML factor with the average returns of two value portfolios minus the average returns of two growth portfolios.

The β1 coefficient is the sensitivity of the asset’s excess returns to the market behavior. The β2 coefficient is the exposure of the asset’s excess returns to the size factor, with a positive value suggesting that the asset behaves more like a small-cap stock while a negative value suggests that the asset behaves like a large-cap stock. The β3 coefficient is the asset’s excess returns’ exposure to the value factor, with a positive value meaning that the asset behaves more like a value stock while a negative value means that the asset behaves more like a growth stock.

The α in the model is the intercept and can be interpreted as the average excess return that can not be explained by the other three model factors. In a highly interpretive model, this factor should be statistically indistinguishable to zero. The ε is the residual which can be seen as the variation that can not be explained by the rest of the model in a specific period point.

Fama and French later extended their model to include two more factors for a total of five. The first additional factor is the Robust-Minus-Weak (RMW) which is the average returns on firms with robust operating profitability minus the average returns of firms with weak operating profitability. The second additional factor is the Conservative-Minus-Aggressive (CMA) and it is the average returns on firms that invest conservatively minus the average returns on firms that invest aggressively. For the purposes of this dissertation, the three factor Fama-French was deemed sufficient.

3.3 Sample Selection and Data Sources

3.3.1 Firm Sample and Stock Price Data

Trying to cover a larger geographical area of economic activity, technology firms from USA, Europe and Asia were initially considered. The main criterion used in this step was whether a company had at least 10 years of publicly traded stock price data. 10 firms were picked from technology sector of NASDAQ, for the USA, another 10 were picked from Euronext Paris and Euronext Amsterdam, covering the European region, and lastly 10 more from the KORDAQ, in an attempt to cover Asia as well.

The stock data were downloaded from Yahoo Finance, through the yfinance package in the Python language ecosystem. The adjusted monthly closing prices where used, instead of the raw closing price, in order to account for stock events like splits and dividends, given the long period examined. So monthly returns were calculated using the latest adjusted close price for each month starting from November 2012 and ending at October 2022.

The stock prices for the European firms in the Euronext stock exchanges were expressed in EUR. The Fama-French factors on the other hand were given in USD, for both the USA and European regions. In order to be able to use the Fama-French factors and group and compare the companies of the different regions, the EUR prices had to be expressed in USD. This was done using the historical exchange FRED DEXUSEU rates.

3.3.2 OSS Contribution Data

For a metric of the OSS contribution activity, the number of commits on GitHub that could be attributed to a company was used. The attribution of a commit to a firm was done as follows. Each commit to a source code repository contains data like the author’s name, the author’s email address, a timestamp, the changes to the source code itself and more. The standard practice in modern day companies is for each employee to be issued a company email address and be discouraged to use any other email for work activity for reasons of security, compliance and personnel management. In technology companies, more specifically, it is also standard practice for the developer to use their company issued emails when committing source code to repositories both internal and external to the company. Given these common practices, email domains used by each firm were collected from publicly available sources on the internet, in order to search for commits authored by people under these email domains.

The search for the commits was done using the publicly available GitHub dataset on the Google BigQuery service. Using a SQL type query and the email domains associated with each firm, I was able to extract the count of commits, grouped by firm, for each day of the examined period. The daily data were then aggregated by month at a later step.

Almost all firms from the USA and Europe regions had meaningful, although rather noisy, daily commit data. But none of the firms from KORDAQ had any commits found in the dataset. This can be attributed to the specialization of the Korean technology companies in hardware technology, while the American and European are more focused, or more inclusive of, software technology, so more likely to contribute to OSS. This lack of OSS data from Korean companies led to the eventual exclusion of the KODAQ firms from the research.

3.3.3 Fama-French Factor Data

The Fama-French three factor datasets were downloaded from Kenneth French's faculty page of Dartmouth University (French, n.d.). Separate datasets were used for the monthly USA and European factors. The datasets included the market excess return, the size factor, the value factor and the risk-free rate factors, given in percentages and had to be converted to decimal points before being used in the models.

3.3.4 Data Preparation and Transformation

As mentioned earlier, since the analysis is done at monthly frequency, the adjusted close prices had to be sampled at the end of each month for all tickers examined. The stock returns for each month could then be calculated as

$$R_{i,m}=\frac{P_{i,m}}{P_{i,m-1}}-1$$

where $P_{i,m}$ is the adjusted close value for month m of ticker i, for all months except the first.

Following, the excess returns for each month were calculated as

$$ExcessReturn_{i,m} = R_{i,m} - RF_{r,m}$$

were $RF_{r,m}$ is the risk free rate for region r (USA or Europe) and month m, taken from the Fama-French data.

As mentioned earlier the stock prices for the European firms had to be converted from EUR to USD. This was done simply as

$$Price_{i,m}^{USD} = Price_{i,m}^{EUR} \times DEXUSEU_m$$

where $DEXUSEU_m$ is the most recent exchange rate for the last adjusted close date of month m.

As mentioned earlier, commits where aggregated on monthly basis and, although this aggregation has a smoothing effect on the noisy timeseries, commits were still riddled with extreme values for some months and some firms. This can be seen in Figures 3.1 and 3.2.

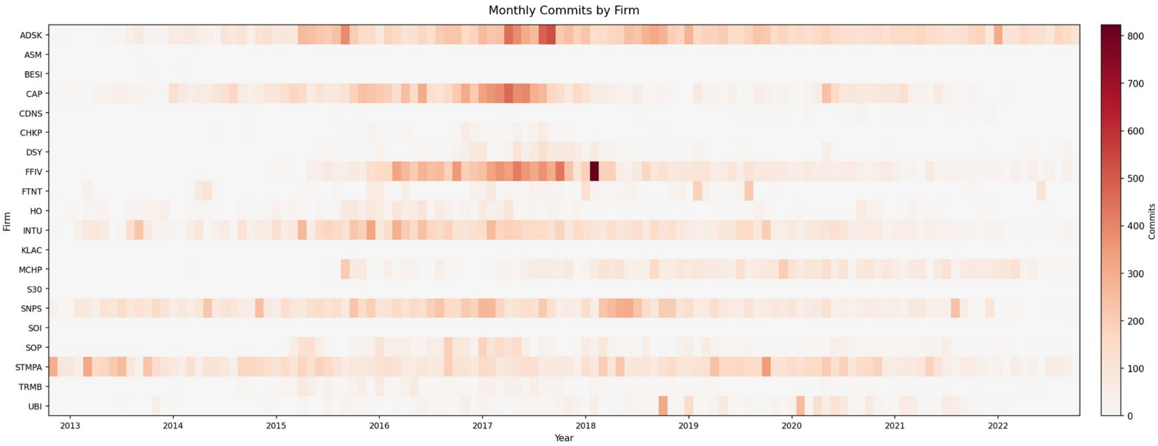

Figure 3.1 Heatmap of Monthly Commits by Firm

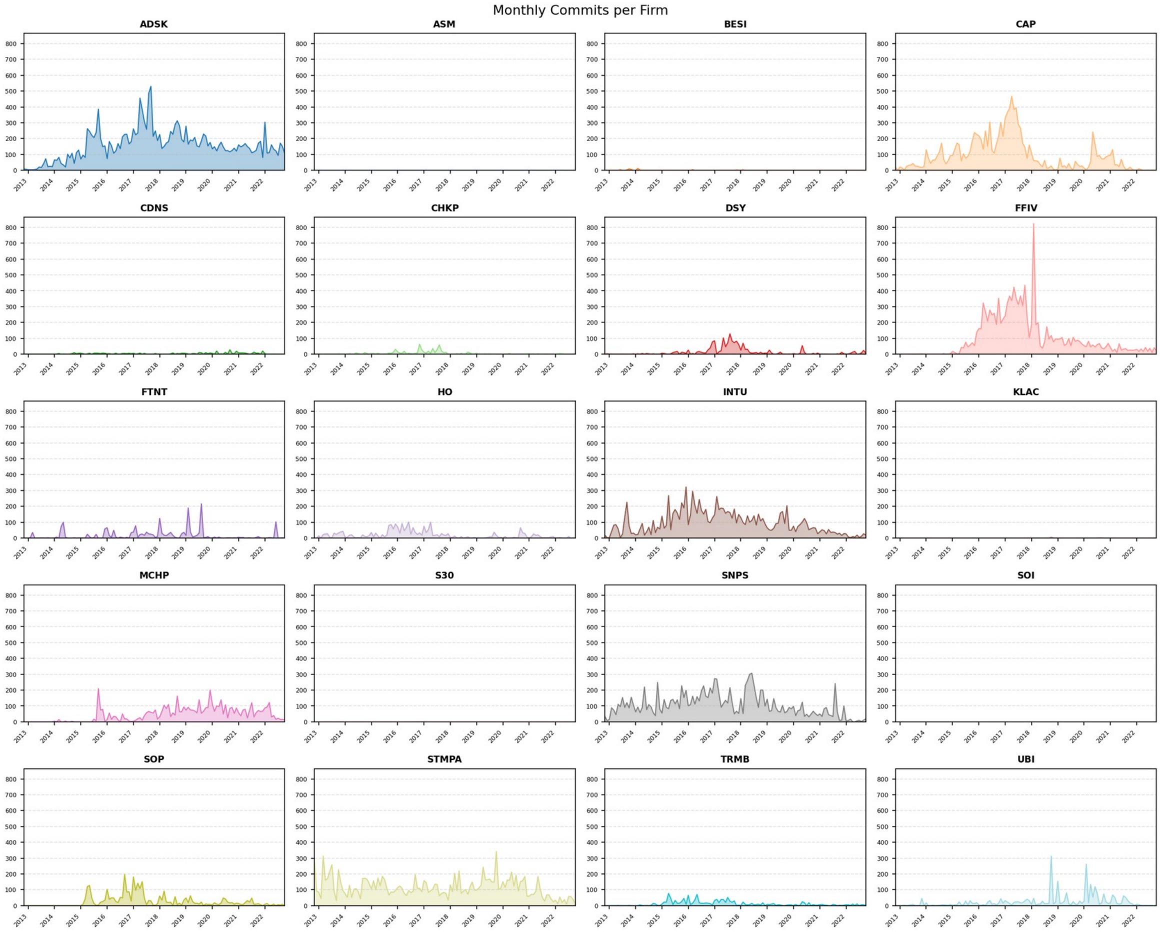

Figure 3.2 Monthly Commits per Individual Firm

To reduce the effect of extreme values, the commits were transformed through a logarithmic transformation:

$$logCommits_{i,m} = \ln(1 + Commits_{i,m})$$

The effect can be seen in Figures 3.3 and 3.4.

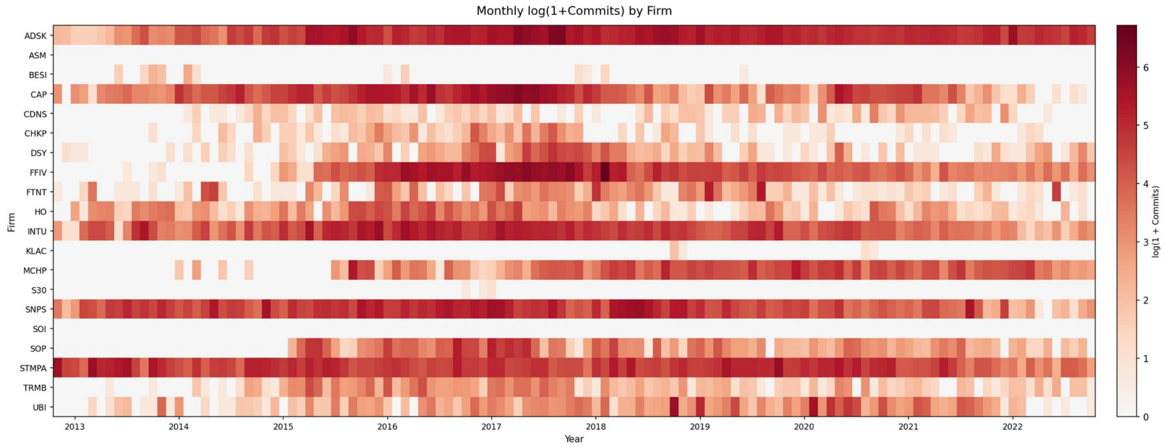

Figure 3.3 Heatmap of Monthly log(1+Commits) by Firm

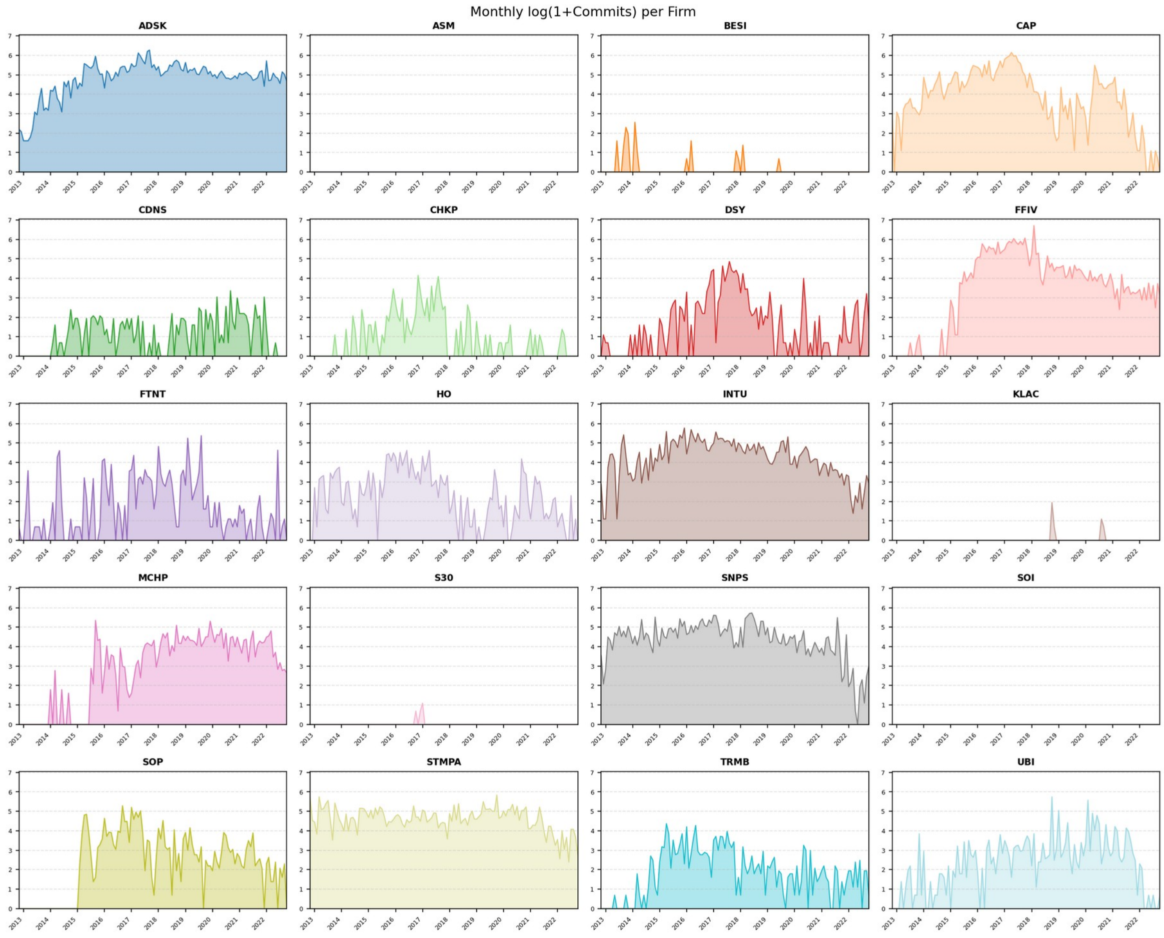

Figure 3.4 Monthly log(1+Commits) per Individual Firm

3.4 Model Specification

3.4.1 Baseline Fama-French Model

As a base line, used to compare the contribution of the OSS data to the excessive returns of the firms studies, a simple 3-factor Fama-French regressive model is used:

$$ExcessReturn_{i,m} = \alpha + \beta_1 MktRF_{r,m} + \beta_2 SMB_{r,m} + \beta_3 HML_{r,m} + \varepsilon_{i,m}$$

where i notes the firm, m the month and r the region. MktRF is the market excess return (Mkt-RF) for the region at each month, SMB the size factor and HML the value factor. This 3-factor Fama-French model is then used as the basis in order to construct models augmented with OSS metrics variables.

In order to be able to do a valid comparison with the augmented models, the baseline model is calculated for each of the next four models separately, since each of them has slightly different starting and ending dates because of lagging or forward constructed variables.

3.4.2 Model A: Contemporaneous OSS Activity

The first model uses the OSS metric of the same month in order to investigate the association of the OSS contributions activity with the stock returns of the same month.

$$ExcessReturn_{i,m} = \alpha + \beta_1 MktRF_{r,m} + \beta_2 SMB_{r,m} + \beta_3 HML_{r,m} + \gamma \, logCommits_{i,m} + \varepsilon_{i,m}$$

3.4.3 Model B: Previous Month OSS Activity

The second model uses a lagged OSS metric of the previous month in order to investigate the association of the OSS contributions activity with the stock returns of the month forward.

$$ExcessReturn_{i,m} = \alpha + \beta_1 MktRF_{r,m} + \beta_2 SMB_{r,m} + \beta_3 HML_{r,m} + \gamma \, logCommits_{i,m-1} + \varepsilon_{i,m}$$

3.4.4 Model C: Three Previous Months OSS Activity

The third model uses a lagged OSS metric of the sum of the three previous month in order to investigate the association of the OSS contributions activity with the stock returns of the month forward.

$$ExcessReturn_{i,m} = \alpha + \beta_1 MktRF_{r,m} + \beta_2 SMB_{r,m} + \beta_3 HML_{r,m} + \gamma \, logPrev3Commits_{i,m} + \varepsilon_{i,m}$$

where

$$logPrevious3Commits_{i,m} = \ln\left(1 + Commits_{i,m-1} + Commits_{i,m-2} + Commits_{i,m-3}\right)$$

so the metric of the OSS activity is the sum of the commits of the 3 previous months, passed through a logarithmic transformation.

3.4.5 Model D: Three Future Months Returns

The fourth model attempts to investigate the association of the OSS contribution of a month with the compounded stock returns over the next 3 months.

$$Future3MExcessReturn_{i,m} = \alpha + \beta_1 Future3MMktRF_{r,m} + \beta_2 Future3MSMB_{r,m}$$

$$+ \beta_3 Future3MHML_{r,m} + \gamma\, logCommits_{i,m} + \varepsilon_{i,m}$$

where the stocks' cumulative excess returns over the three next months are calculated as

$$Future3MExcessReturn_{i,m} = Future3MReturn_{i,m} - Future3MRF_{i,m}$$

$$Future3MReturn_{i,m} = (1 + Return_{i,m+1}) * (1 + Return_{i,m+2}) * (1 + Return_{i,m+3}) - 1$$

$$Future3MRF_{r,m} = (1 + RF_{r,m+1}) * (1 + RF_{r,m+2}) * (1 + RF_{r,m+3}) - 1$$

The market excess returns over the future three months are calculated as

$$Future3MMktRF_{r,m} = Future3MMarketReturn_{r,m} - Future3MRF_{r,m}$$

$$Future3MMarketReturn_{r,m} = (1 + MarketReturn_{m+1}) * (1 + MarketReturn_{m+2})$$

$$* (1 + MarketReturn_{m+3}) - 1$$

For SMB and HML the future three month values are the compounded three month factor returns:

$$Future3MSMB_{r,m} = (1 + SMB_{m+1}) * (1 + SMB_{m+2}) * (1 + SMB_{m+3}) - 1$$

$$Future3MHML_{r,m} = (1 + HML_{m+1}) * (1 + HML_{m+2}) * (1 + HML_{m+3}) - 1$$

3.5 Regional, Firm Fixed Effects and Firm Specific Analysis

For all four models, three different types of analysis was done, based on different modes of pulling on the models' coefficients.

The first is a pooled OLS variant that uses both cross sectional and time series variation:

$$Y_{i,m} = \alpha + \beta X_{r,m} + \gamma\, OSS_{i,m} + \varepsilon_{i,m}$$

where Y and X are the returns and Fama-French factors variables for each model and OSS is the different OSS contribution metrics in each model.

Second, a firm fixed effects variation that allows each firm to have its own intercept:

$$Y_{i,m} = \alpha_i + \beta X_{r,m} + \gamma\, OSS_{i,m} + \varepsilon_{i,m}$$

this allows to control for differences between the firms that could affect the association of the variables like the different business models, the long term OSS strategy, the specificity of the technology sector, the inner culture and more.

This essentially examines whether changes in the OSS activity in a specific firm are associated with changes in the factor-adjusted returns. This reduces the risk that specific firms have more influence to the results.

Third, a firm specific variant allows to estimate a different γ factor for each firm, instead of a one common γ estimated in the firm fixed effects variant:

$$Y_{i,m} = \alpha_i + \beta_i X_{r,m} + \gamma_i\, OSS_{i,m} + \varepsilon_{i,m}$$

This mode can be used to examine the heterogeneity of the results and whether they are mainly driven by specific firms, by comparing the signs and significance across them. Since the data for the individual firms are rather noisy, this variant was included mainly for exploratory purposes.

Each of the three different variations for all four models are calculated for the full set of the firms, as well as the USA and European firms separately.

3.6 Baseline and Augmented Model Comparison

As explained earlier, each augmented model has a matching baseline 3 factor Fama-French model, where the Fama-French use only its three factors, while the augmented models use the three factors plus the OSS contributions variable. This can be used to explore the incremental explanatory power of the OSS metrics, by comparing the raw and adjusted R2 values:

$$\Delta R^2 = R^2_{Augmented} - R^2_{Baseline}$$

$$\Delta\, AdjR^2 = AdjR^2_{Augmented} - AdjR^2_{Baseline}$$

The adjusted R2 penalizes the inclusion of the additional explanatory variable, so it is the more important metric of the two. It can be used as a comparison metric along with the sign, magnitude and statistical significance of the OSS coefficient.

3.7 Software and Implementation

All needed data of stock prices and exchange rates were downloaded using Python and the yfinance package that helps with Yahoo Finance data. The commits data that were extracted with a SQL query in Google BigQuery, and the query itself was again constructed with the help of a Python script. The Fama-French factors were downloaded by hand from the official site. All initial, intermediate and result data were save in comma separated format files.

The Python packages of pandas and numpy were used in the cleaning, transformation and calculation of data and variables, and statsmodels package was used for calculating the regressions of all models and variations. Finally, the matplotlib Python package was used for constructing all the diagrams included in the dissertation.

All code, data and results files can be found in the repository mentioned in Appendix A.

3.8 Methodology Limitations

Considering the limitations of the methodology, it has already been stated earlier that, although the commits on the source code of a software project are the base of the effort contributed, there might be other contributions that are not represented directly by them, like the donations in infrastructure or other forms of work around product marketing and project management. Moreover, commits to source code themselves can differ in size and frequency among different software project development styles, that could introduce irregularities between firms, and might not capture the quality of the contributions.

Another issue introducing blind spots could be of firms comfortable enough to let employees contribute to OSS through their personal accounts and emails, even though they are billed directly for the work done. The attribution of commits to a firm can also suffer from other causes like unassociated email domains. Apart from employees contributing on personal accounts, firms can also contribute to OSS project by hiring external contractors, sometimes already established contributors to an OSS project, in order to further the project’s development or develop specific features based on the firm’s needs. One more way a company can contribute OSS, that would not be reflected on this methodology, is by monetary contributions to umbrella OSS organizations like the Linux Foundation and the Apache Foundation, that use the funds to develop OSS projects.

Finally, and looking at the methodology from a more statistical perspective, the stock returns can reflect many different factors internal and external to a firm, that could be unrelated to the OSS contributions. On top of that, OSS contributions could affect a company’s value on timescales longer than the one to three months that were examined in this dissertation. While the methodology is attempting to estimate association and not causation, the small firm sample limits the generalizability of the results.

4. Data analysis and Results

4.1 Sample Description

The final sample used in the dissertation was 20 publicly listed firms, 10 from the US and 10 from Europe, after dropping the 10 Asian firms from the total of 30 initially selected. All the firms samples contain data for the period from November 2012 to October 2022. For models A and B, this is translated to 2,380 observations, 119 per firm, from December 2012 to October 2022. Model C has 2,340 observations, 117 per firm, from February 2012 to October 2022, and model D has also 2,340 observations, 117 per firm, but from November 2012 to July 2022. As for the regional split, each region contains exactly half of the observations, in all models.

4.2 Descriptive Statistics

4.2.1 Dependent Variables

Table 4.1 contains descriptive statistics about the dependent or return variables used in the models, which are the monthly stock returns, the monthly excess returns and, for model D, the future three months' cumulative excess return.

Table 4.1 Descriptive Statistics of Return Variables

	Return	Excess Return	Future M3 Return
count	2380	2380	2340
mean	0.019	0.018	0.056
std	0.096	0.097	0.167
min	-0.547	-0.547	-0.635
25%	-0.035	-0.036	-0.046
50%	0.017	0.017	0.054
75%	0.073	0.073	0.147
max	0.579	0.579	0.992

It can be noted that the mean monthly excess return of 1.8% is consistent with the equity risk premium over a period with low rates. The slightly positive skew can be attributed to the bull market in technology sector from 2013 to 2021.

4.2.2 Fama-French Factors

Table 4.2 contains descriptive statistics for the three Fama-French factors, as well as the risk free rate for each of the two, US and Europe, regions.

Table 4.2 Descriptive Statistics of the Fama-French Factors

Region	USA				Europe			
Factor	Mkt-RF	SMB	HML	RF	Mkt-RF	SMB	HML	RF
N	119	119	119	119	119	119	119	119
Mean	0.0105	-0.0001	-0.0003	0.0005	0.0045	0.0017	-0.0005	0.0005
Std	0.0438	0.0253	0.0354	0.0007	0.0460	0.0170	0.0302	0.0007
Min	-0.1337	-0.0593	-0.1383	0.0000	-0.1544	-0.0422	-0.1130	0.0000
P25	-0.0116	-0.0192	-0.0192	0.0000	-0.0261	-0.0098	-0.0180	0.0000
Median	0.0137	0.0014	-0.0038	0.0001	0.0068	0.0018	-0.0029	0.0001
P75	0.0333	0.0160	0.0162	0.0010	0.0406	0.0128	0.0162	0.0010
Max	0.1360	0.0714	0.1286	0.0023	0.1662	0.0503	0.1209	0.0023

The USA market excess return is notably higher than the European, reflecting the higher growth over this period in the USA over Europe. The mean HML is negative in both regions, and the distribution negatively skewed, suggesting a period of higher returns for growth stocks. The risk free rate is the same in both regions as it is denominated to the US Treasury bill rate for both.

4.2.3 OSS Activity Variable

Table 4.3 summarizes the descriptive statistics for each of the OSS contribution metrics used in the four models. Log commits used by models A and D, one month lagged log commits used by model B and a rolling window of three months lagged log commits used by model C.

Table 4.3 Descriptive Statistics for OSS Contribution Metrics

	Log Commits	Log Commits 1M lag	Log Commits 3M Window
count	2380	2380	2340
mean	2.107	2.105	2.959
std	1.990	1.991	2.356
min	0.000	0.000	0.000
25%	0.000	0.000	0.000
50%	1.792	1.792	3.113

75%	4.043	4.043	5.205
max	6.714	6.714	7.153

It is very clear that the activity for a lot of the firm-months is at zero. This happens because many of the European firms have a near zero activity, but also because there is large dispersion between the number of commits both between firms and within the same firm, as was seen in Chapter 3.

Table 4.4 and Figure 4.1 show the statistical correlation between the three OSS metrics, where we see a high pairwise correlation across all of them.

Table 4.4 Autocorrelation of OSS Variables

	Log Commits	logCommits_lag1	logCommits_last3
logCommits	1.000	0.901	0.907
logCommits_lag1	0.901	1.000	0.949
logCommits_last3	0.907	0.949	1.000

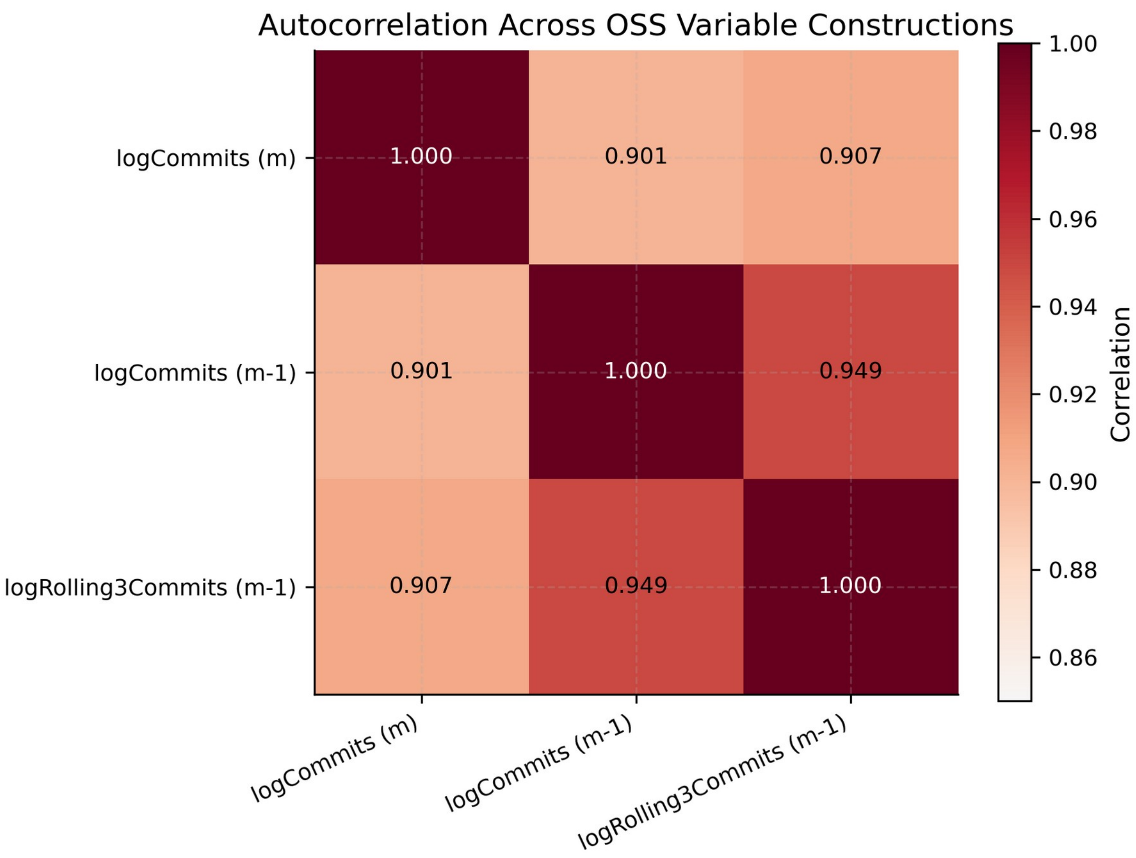

Figure 4.1 Correlation Heatmap of the OSS Variables

The unconditional correlation of the excess returns with the OSS metrics for each of the models, as seen in the correlation files in the repository, is between -0.003 and 0.006. These close to zero values are expected since the unconditional correlation draws most of the differences from the cross sectional and time series variation rather than the OSS activity.

4.3 Baseline Fama-French Three Factor Model Results

As explained in Chapter 3 when discussing about he methodology, the baseline Fama-French model was estimated for each of the augmented models to serve as benchmarks for the adjusted R^2 changes. Tables 4.5, 4.6 and 4.7 contain the coefficients for each model, along with the R^2 and adjusted R^2 values, for each of the modes examined and for both regions and full sample.

Table 4.5 Coefficients for Full Sample

Full	Model A		Model B		Model C		Model D	
Metric	FFE	OLS	FFE	OLS	FFE	OLS	FFE	OLS
Intercept	0.007	0.009	0.007	0.009	0.006	0.009	0.019	0.029
Mkt_RF	1.239	1.231	1.239	1.231	1.238	1.231	1.147	1.115
SMB	0.129	0.141	0.129	0.141	0.123	0.135	0.380	0.426
HML	-0.241	-0.240	-0.241	-0.240	-0.248	-0.247	-0.251	-0.247
R^2	0.336	0.330	0.336	0.330	0.336	0.330	0.302	0.284
Adj. R^2	0.330	0.329	0.330	0.329	0.329	0.329	0.296	0.283

Table 4.6 Coefficients for USA Sample

USA	Model A		Model B		Model C		Model D	
Metric	FFE	OLS	FFE	OLS	FFE	OLS	FFE	OLS
Intercept	0.008	0.005	0.008	0.005	0.007	0.005	0.022	0.019
Mkt_RF	1.199	1.199	1.199	1.199	1.200	1.200	1.052	1.052
SMB	0.082	0.082	0.082	0.082	0.076	0.076	0.317	0.317
HML	-0.234	-0.234	-0.234	-0.234	-0.241	-0.241	-0.257	-0.257
R^2	0.419	0.413	0.419	0.413	0.419	0.413	0.367	0.344
Adj. R^2	0.413	0.411	0.413	0.411	0.412	0.411	0.360	0.343

Table 4.7 Coefficients for Europe Sample

Europe	Model A		Model B		Model C		Model D	
Metric	FFE	OLS	FFE	OLS	FFE	OLS	FFE	OLS
Intercept	0.018	0.012	0.018	0.012	0.018	0.009	0.057	0.039
Mkt_RF	1.273	1.273	1.273	1.273	1.272	1.231	1.187	1.187
SMB	0.256	0.256	0.256	0.256	0.250	0.135	0.626	0.626
HML	-0.249	-0.249	-0.249	-0.249	-0.253	-0.247	-0.248	-0.248
R^2	0.289	0.286	0.289	0.286	0.288	0.330	0.275	0.265
Adj. R^2	0.282	0.284	0.282	0.284	0.281	0.329	0.267	0.263

The intercept for all cases is very close to zero, meaning that in all cases the three factors are enough to explain the excess returns. The MktRF beta is consistently between 1.05 and 1.27, reflecting the broad market exposure of the firms. The HML coefficient is negative for all cases, noting that all firms are growth oriented. The SMB coefficient is positive for all cases which could mean that the firms selected are relatively on the smaller capitalization scale.

The R^2 values range from 0.28 to 0.42, indicating that the three Fama-French factors explain a moderate but not dominant share of the variance in the firms' excess returns. In most models the firm fixed effect mode raises the R^2 by a few percentage points compared to the Pooled OLS mode, which is expected since firm fixed effect takes into account the differences between the firms due to the firm-individual intercepts.

4.4 Model A Results

Model A investigates the relationship between the contemporaneous OSS activity and the excess returns of the firms. Since the same months' contributions are examined along the same months' excess returns, no causality direction can be established by this model.

4.4.1 Pooled OLS Results

Table 4.8 contains the coefficients values, the standard errors values and the p-values for the full sample, the USA region and the European region.

Table 4.8 Model A Pooled OLS Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	0.0097	0.0030	0.0000	0.0044	0.0040	0.2220	0.0133	0.0030	0.0000

Mkt_RF	1.2315	0.0760	0.0000	1.1992	0.0900	0.0000	1.2739	0.1290	0.0000
SMB	0.1405	0.0980	0.1530	0.0824	0.0870	0.3410	0.2579	0.2480	0.2980
HML	-0.2411	0.0790	0.0020	-0.2334	0.1040	0.0240	-0.2497	0.1370	0.0680
logCommits	-0.0005	0.0010	0.5250	0.0003	0.0010	0.7340	-0.0005	0.0010	0.5510

In all three modes, the logCommits coefficient is not statistically significant. The baseline factor structure is preserved with a significant positive MktRF beta and a negative HML coefficient.

4.4.2 Firm Fixed Effects Results

Table 4.9 contains the coefficients, standard errors and p-values for the full, USA and European region samples.

Table 4.9 Model A FFE Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	-0.0023	0.0020	0.2300	0.0013	0.0030	0.6890	0.0175	0.0010	0.0000
T.ASM	0.0202	0.0020	0.0000						
T.BESI	0.0297	0.0020	0.0000				0.0094	0.0001	0.0000
T.CAP	0.0045	0.0010	0.0000				-0.0180	0.0010	0.0000
T.CDNS	0.0101	0.0010	0.0000	0.0075	0.0020	0.0000			
T.CHKP	-0.0027	0.0010	0.0410	-0.0053	0.0020	0.0100			
T.DSY	0.0048	0.0010	0.0000				-0.0164	0.0010	0.0000
T.FFIV	-0.0109	0.0000	0.0000	-0.0118	0.0010	0.0000			
T.FTNT	0.0131	0.0010	0.0000	0.0110	0.0020	0.0000			
T.HO.PA	0.0067	0.0010	0.0000				-0.0149	0.0010	0.0000
T.INTU	0.0002	0.0000	0.2480	-0.0002	0.0000	0.5970			
T.KLAC	0.0127	0.0020	0.0000	0.0094	0.0030	0.0000			
T.MCHP	0.0007	0.0010	0.3250	-0.0007	0.0010	0.5230			
T.S30	0.0151	0.0020	0.0000				-0.0051	0.0000	0.0000
T.SNPS	0.0011	0.0000	0.0000	0.0007	0.0000	0.0040			
T.SOI	0.0181	0.0020	0.0000				-0.0021	0.0000	0.0000
T.SOP	0.0062	0.0010	0.0000				-0.0155	0.0010	0.0000
T.STMPA.PA	0.0077	0.0010	0.0000				-0.0153	0.0020	0.0000
T.TRMB	-0.0039	0.0010	0.0000	-0.0059	0.0020	0.0000			
T.UBI	0.0051	0.0010	0.0000				-0.0165	0.0010	0.0000

Mkt_RF	1.2373	0.0770	0.0000	1.1997	0.0910	0.0000	1.2697	0.1290	0.0000
SMB	0.1298	0.0990	0.1880	0.0843	0.0870	0.3350	0.2450	0.2490	0.3240
HML	-0.2373	0.0790	0.0030	-0.2308	0.1040	0.0270	-0.2458	0.1380	0.0740
logCommits	0.0020	0.0000	0.0000	0.0013	0.0010	0.0160	0.0026	0.0000	0.0000

A statistically significant and positive commits coefficient is emerging, for all three regions, now that the heterogeneity between firms is absorbed by the individual intercepts.

4.4.3 Incremental Explanatory Power

The following table contains the adjusted R^2 values and their difference from the baseline model, for the two modes and three regions examined.

Table 4.10 Model A Adjusted R^2

	Full		USA		Europe	
	OLS	FE	OLS	FE	OLS	FE
Model A	0.32917	0.33030	0.41088	0.41265	0.28378	0.28217
Baseline	0.32937	0.33002	0.41134	0.41277	0.28431	0.28214
Diff	-0.00019	0.00028	-0.00046	-0.00012	-0.00053	0.00003

It can be seen that all OLS augmented models don’t add any explanatory power. The only positive difference in adjusted R2 is in the firm fixed effects for the full sample and marginally for the Europe sample.

4.4.4 Regressions on Specific Firms

As explained in Chapter 3, the specific firm regressions are the implementation of OLS in the individual time series. For model A, none of the firms produce any statistically significant OSS contributions coefficient, although the majority of the coefficient signs are positive. This means that the OSS signal does not appear concentrated on a small number of firms.

4.5 Model B Results

Slightly different from model A, model B examines the relationship between the OSS activity of the previous month and the excess returns of the firms.

4.5.1 Pooled OLS Results

Table 4.11 has the coefficients values, the standard errors and the p-values for the full sample, the USA region and the European region.

Table 4.11 Model B Pooled OLS Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	0.0109	0.0025	0.0000	0.0052	0.0035	0.1401	0.0147	0.0027	0.0000
Mkt_RF	1.2323	0.0761	0.0000	1.1991	0.0904	0.0000	1.2752	0.1281	0.0000
SMB	0.1408	0.0985	0.1526	0.0820	0.0869	0.3456	0.2600	0.2488	0.2960
HML	-0.2417	0.0791	0.0022	-0.2342	0.1037	0.0240	-0.2513	0.1373	0.0673
LogCommits M1 lag	-0.0010	0.0007	0.1618	-0.0001	0.0008	0.9386	-0.0013	0.0009	0.1422

It can be seen that in all three modes the LogCommits M1 lag coefficient is not statistically significant, negative and of small relative value. The baseline factor structure is once again preserved with a significant positive MktRF beta and a negative HML coefficient.

4.5.2 Firm Fixed Effects Results

Table 4.12 shows the coefficients, standard errors and p-values for the full, USA and European region samples.

Table 4.12 Model B FFE Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	0.0056	0.0040	0.1340	0.0053	0.0040	0.2110	0.0175	0.0010	0.0000
T.ASM	0.0122	0.0040	0.0010						
T.BESI	0.0220	0.0030	0.0000				0.0098	0.0000	0.0000
T.CAP	0.0028	0.0010	0.0040				-0.0076	0.0050	0.1230
T.CDNS	0.0039	0.0030	0.1450	0.0044	0.0030	0.1250			
T.CHKP	-0.0091	0.0030	0.0010	-0.0086	0.0030	0.0040			
T.DSY	-0.0005	0.0020	0.8290				-0.0120	0.0020	0.0000
T.FFIV	-0.0131	0.0010	0.0000	-0.0129	0.0010	0.0000			
T.FTNT	0.0079	0.0020	0.0000	0.0084	0.0020	0.0000			
T.HO.PA	0.0025	0.0020	0.2180				-0.0087	0.0030	0.0030
T.INTU	-0.0008	0.0000	0.0640	-0.0007	0.0000	0.1320			
T.KLAC	0.0047	0.0030	0.1590	0.0055	0.0040	0.1360			
T.MCHP	-0.0025	0.0010	0.0630	-0.0022	0.0010	0.1310			

T.S30	0.0072	0.0040	0.0440				-0.0050	0.0000	0.0000
T.SNPS	0.0003	0.0000	0.4100	0.0003	0.0000	0.3410			
T.SOI	0.0101	0.0040	0.0050				-0.0021	0.0000	0.0000
T.SOP	0.0020	0.0020	0.3080				-0.0091	0.0030	0.0020
T.STMPA.PA	0.0074	0.0010	0.0000				-0.0026	0.0060	0.6610
T.TRMB	-0.0088	0.0020	0.0000	-0.0084	0.0020	0.0000			
T.UBI	0.0010	0.0020	0.6120				-0.0101	0.0030	0.0010
Mkt_RF	1.2383	0.0770	0.0000	1.1992	0.0910	0.0000	1.2735	0.1290	0.0000
SMB	0.1290	0.0990	0.1910	0.0819	0.0870	0.3460	0.2564	0.2490	0.3030
HML	-0.2409	0.0790	0.0020	-0.2334	0.1040	0.0250	-0.2493	0.1380	0.0710
logCommits	0.0003	0.0010	0.6630	0.0005	0.0010	0.5510	-0.0002	0.0010	0.8950

Contrary to model A, there is no statistically significant lagged commits coefficient for all three regions.

4.5.3 Incremental Explanatory Power

Table 4.13 contains the adjusted R^2 values and their difference from the baseline model, for the two modes and three regions examined.

Table 4.13 Model B Adjusted R^2

	Full		USA		Europe	
	OLS	FE	OLS	FE	OLS	FE
Model B	0.32952	0.32975	0.41085	0.41232	0.28425	0.28153
Baseline	0.32937	0.33002	0.41134	0.41277	0.28431	0.28214
Diff	0.00015	-0.00027	-0.00049	-0.00045	-0.00006	-0.00061

We can see that almost none of the OLS or fixed effects augmented models add any explanatory power. The only positive difference in adjusted R^2 is in the pooled OLS for the full sample but it it still near zero.

4.5.4 Regressions on Specific Firms

For model B, none of the firms produce any statistically significant OSS contributions coefficient in the specific firm regressions on the individual time series, although most of the coefficient signs are also positive like in model A. This means that the 1 month lagging OSS signal does not appear concentrated on any small number of firms.

4.6 Model C Results

Extending model B, model C examines the relationship between the sum of OSS activity over a window of the previous three months and the excess returns of the firms' stocks.

4.6.1 Pooled OLS Results

Table 4.14 presents the coefficients, the standard errors and the p-values for the full, the USA region and the European region samples.

Table 4.14 Model C Pooled OLS Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	0.0103	0.0028	0.0003	0.0028	0.0038	0.4607	0.0148	0.0031	0.0000
Mkt_RF	1.2316	0.0758	0.0000	1.1999	0.0913	0.0000	1.2739	0.1272	0.0000
SMB	0.1342	0.0975	0.1689	0.0765	0.0861	0.3739	0.2559	0.2445	0.2953
HML	-0.2478	0.0790	0.0017	-0.2398	0.1029	0.0198	-0.2534	0.1381	0.0665
LogCommits M3 lag	-0.0006	0.0006	0.3286	0.0006	0.0006	0.3314	-0.0010	0.0007	0.1200

In all of the three modes the LogCommits M3 lag coefficient does not appear to be statistically significant. The baseline factor structure is again the same as the previous models, with a significant positive MktRF beta and a negative HML coefficient.

4.6.2 Firm Fixed Effects Results

Table 4.15 shows the coefficients values, standard errors values and p-values for the full, USA and European region samples.

Table 4.15 Model C FFE Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	-0.0046	0.0030	0.1230	-0.0049	0.0050	0.3470	0.0176	0.0010	0.0000
T.ASM	0.0225	0.0030	0.0000						
T.BESI	0.0318	0.0030	0.0000				0.0095	0.0000	0.0000
T.CAP	0.0047	0.0010	0.0000				-0.0153	0.0040	0.0000
T.CDNS	0.0107	0.0020	0.0000	0.0112	0.0030	0.0000			
T.CHKP	-0.0020	0.0020	0.2980	-0.0016	0.0030	0.6200			
T.DSY	0.0064	0.0020	0.0000				-0.0148	0.0020	0.0000

T.FFIV	-0.0106	0.0010	0.0000	-0.0105	0.0010	0.0000			
T.FTNT	0.0125	0.0010	0.0000	0.0128	0.0020	0.0000			
T.HO.PA	0.0078	0.0010	0.0000				-0.0129	0.0030	0.0000
T.INTU	0.0010	0.0000	0.0000	0.0011	0.0000	0.0050			
T.KLAC	0.0139	0.0030	0.0000	0.0145	0.0040	0.0010			
T.MCHP	0.0011	0.0010	0.2130	0.0014	0.0020	0.3640			
T.S30	0.0167	0.0030	0.0000				-0.0057	0.0000	0.0000
T.SNPS	0.0021	0.0000	0.0000	0.0021	0.0000	0.0000			
T.SOI	0.0196	0.0030	0.0000				-0.0029	0.0000	0.0000
T.SOP	0.0048	0.0010	0.0010				-0.0161	0.0020	0.0000
T.STMPA.PA	0.0066	0.0010	0.0000				-0.0130	0.0040	0.0020
T.TRMB	-0.0039	0.0010	0.0050	-0.0036	0.0020	0.1110			
T.UBI	0.0065	0.0010	0.0000				-0.0142	0.0030	0.0000
Mkt_RF	1.2366	0.0770	0.0000	1.2000	0.0920	0.0000	1.2702	0.1290	0.0000
SMB	0.1211	0.0970	0.2140	0.0777	0.0860	0.3650	0.2418	0.2460	0.3250
HML	-0.2455	0.0780	0.0020	-0.2371	0.1030	0.0210	-0.2533	0.1370	0.0650
logCommits3	0.0019	0.0000	0.0000	0.0020	0.0010	0.0080	0.0014	0.0010	0.0560

Contrary to model B, there is statistically significant three months lagged commits coefficient for the full and USA regions. In all three regions the OSS coefficient is positive for model C.

4.6.3 Incremental Explanatory Power

Table 4.16 contains the adjusted R^2 values and their difference from the baseline model, for the two modes and three regions examined.

Table 4.16 Model C Adjusted R^2

	Full		USA		Europe	
	OLS	FE	OLS	FE	OLS	FE
Model C	0.32889	0.32975	0.41094	0.41304	0.28335	0.28068
Baseline	0.32897	0.32947	0.41121	0.41250	0.28349	0.28111
Diff	-0.00008	0.00028	-0.00027	0.00054	-0.00014	-0.00043

Here it can be observed that the statistically significant OSS coefficient in the firm fixed effects in the full and USA regions also adds a small amount to the explanatory power.

4.6.4 Regressions on Specific Firms

In model C, only one of the firms produced a statistically significant OSS coefficient but that specific firm lacks a significant number of non-zero commit values, so it can not be interpreted any further.

4.7 Model D Results

Model D differs from the three previous models because it examines how the current contributions affect the excess returns on the future three months period.

4.7.1 Pooled OLS Results

Table 4.17 shows the coefficient values, the standard errors and the p-values for the full, the USA region and the European region samples.

Table 4.17 Model D Pooled OLS Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	0.0333	0.0074	0.0000	0.0157	0.0106	0.1367	0.0455	0.0084	0.0000
Mkt_RF	1.1192	0.0957	0.0000	1.0519	0.0891	0.0000	1.1917	0.1645	0.0000
SMB	0.4239	0.1222	0.0005	0.3184	0.1236	0.0100	0.6439	0.2598	0.0132
HML	-0.2498	0.0965	0.0097	-0.2555	0.1008	0.0113	-0.2512	0.1987	0.2063
logCommits	-0.0022	0.0021	0.3129	0.0013	0.0022	0.5587	-0.0041	0.0027	0.1320

For all three modes the logCommits OSS coefficient appears to be statistically insignificant. The baseline factor structure is similar as all the previous models, with a significant positive MktRF beta and a negative HML coefficient.

4.7.2 Firm Fixed Effects Results

Table 4.18 contains the coefficients, standard errors and p-values for the full, USA and European region samples.

Table 4.18 Model D FFE Results

	Full			USA			Europe		
	coef	std_err	pvalue	coef	std_err	pvalue	coef	std_err	pvalue
Intercept	-0.0016	0.0100	0.8720	-0.0031	0.0120	0.8000	0.0574	0.0020	0.0000
T.ASM	0.0608	0.0090	0.0000						
T.BESI	0.0848	0.0090	0.0000				0.0244	0.0010	0.0000

T.CAP	0.0137	0.0030	0.0000				-0.0357	0.0150	0.0160
T.CDNS	0.0330	0.0070	0.0000	0.0363	0.0080	0.0000			
T.CHKP	-0.0112	0.0070	0.1120	-0.0078	0.0080	0.3520			
T.DSY	0.0150	0.0070	0.0250				-0.0413	0.0060	0.0000
T.FFIV	-0.0297	0.0020	0.0000	-0.0285	0.0030	0.0000			
T.FTNT	0.0389	0.0060	0.0000	0.0417	0.0070	0.0000			
T.HO.PA	0.0185	0.0050	0.0010				-0.0355	0.0090	0.0000
T.INTU	0.0056	0.0010	0.0000	0.0061	0.0010	0.0000			
T.KLAC	0.0351	0.0090	0.0000	0.0394	0.0100	0.0000			
T.MCHP	0.0012	0.0040	0.7460	0.0029	0.0040	0.4920			
T.S30	0.0447	0.0090	0.0000				-0.0160	0.0001	0.0000
T.SNPS	0.0112	0.0010	0.0000	0.0116	0.0010	0.0000			
T.SOI	0.0511	0.0090	0.0000				-0.0097	0.0000	0.0000
T.SOP	0.0138	0.0050	0.0100				-0.0402	0.0090	0.0000
T.STMPA.PA	0.0249	0.0020	0.0000				-0.0223	0.0180	0.2090
T.TRMB	-0.0148	0.0050	0.0070	-0.0121	0.0060	0.0610			
T.UBI	0.0201	0.0050	0.0000				-0.0337	0.0090	0.0000
Mkt_RF	-0.2467	0.0950	0.0100	-0.2505	0.1000	0.0130	-0.2474	0.1970	0.2100
SMB	1.1425	0.0990	0.0000	1.0515	0.0900	0.0000	1.1860	0.1650	0.0000
HML	0.3785	0.1200	0.0020	0.3215	0.1220	0.0080	0.6205	0.2640	0.0190
logCommits	0.0043	0.0020	0.0210	0.0052	0.0020	0.0180	0.0013	0.0040	0.7290

Similar to model C, the OSS coefficient is statistically significant for the full and USA regions but not the Europe region. In all three regions the OSS coefficient is positive for model C.

4.7.3 Incremental Explanatory Power

Table 4.19 shows the adjusted R^2 values and their difference from the baseline model, for the two modes and three regions examined.

Table 4.19 Model D Adjusted R^2

	Full		USA		Europe	
	OLS	FE	OLS	FE	OLS	FE
Model D	0.28298	0.29628	0.34255	0.36181	0.26434	0.26663
Baseline	0.28262	0.29569	0.34277	0.35999	0.26337	0.26721

Diff	0.00036	0.00059	-0.00021	0.00182	0.00098	-0.00058

A small increase in the adjusted R^2 values appears in both pooled OLS and fixed firm effects for the full sample. The fixed firm effects for the USA sample also appears to increase in explanatory power, the biggest increase in a model so far, as well as the pooled OLS model for Europe.

4.7.4 Regressions on Specific Firms

The results in the specific firm regressions in model D appear the same as in model C, with one of the firms producing a statistically significant OSS coefficient but that specific firm lacks a significant number of non-zero commit values.

5. Discussion and Conclusions

5.1 Interpretation of Core Findings

This dissertation attempted to examine whether the contribution in Open Source Software by publicly traded technology firms carries any information that can be associated with their excess stock returns, beyond what the three factor Fama-French model can already explain.

The pooled OLS regressions for all models examined, across all three regional samples (Full, USA, Europe), do not produce any significant OSS contribution coefficients, as can be seen in Figure 5.1. In economic terms this can be interpreted as that firms that commit to source code more on average, do to have higher or lower returns on their stocks, compared to firms that commit less. In other words a portfolio long on high OSS committing firms and short on low OSS committing firms, is not statistically expected to have higher returns. The Fama-French model indicates that the dominant driver of the excess returns would be the overall market excess returns.

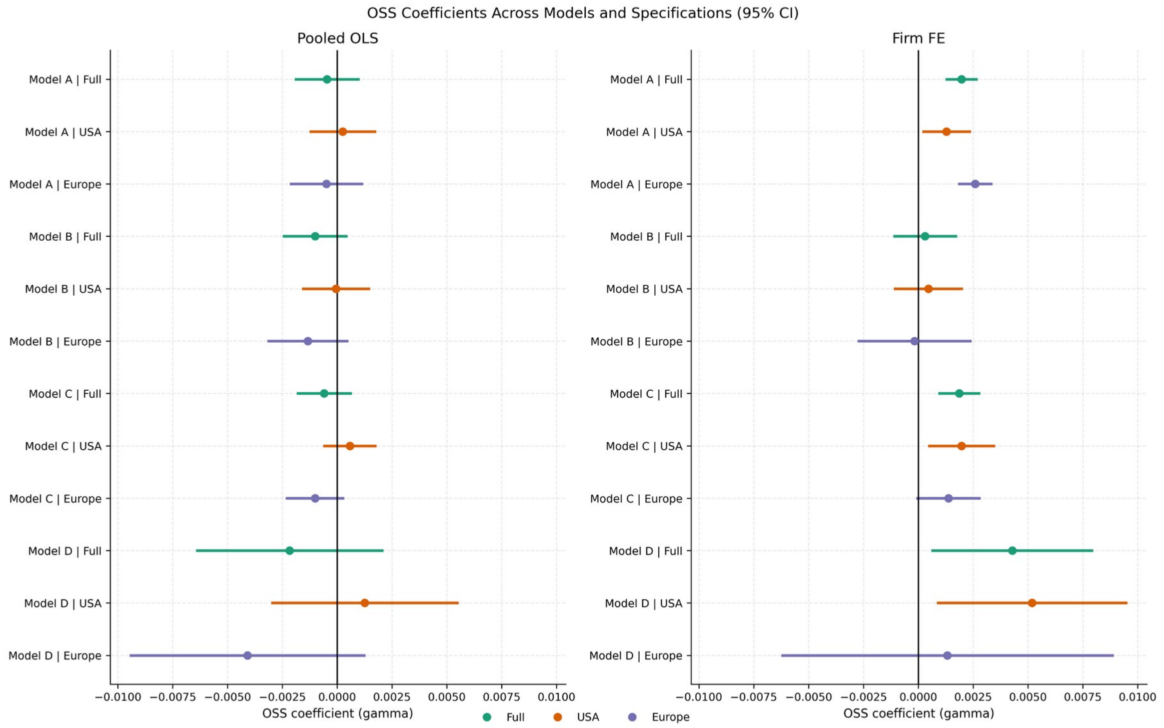

Figure 5.1 Summary diagram of all OSS coefficients, for Pooled OLS and Firm FE modes, for all regions

Once the heterogeneity between the firm is removed, in the firm fixed effects mode, a positive and statistically significant association emerges for some models and some areas, as can be also seen in Figure 5.1. More specifically this is seen in the contemporaneous association model (model A) for all regions, in the 3 month lagging window model (model C) for the full and USA regions and in the future 3 months model (model D) for the full and USA regions again. This means that months with high OSS contribution activity can coincide with months of higher excess stock returns. On the other hand, model B which is the one month lagging model, does not produce a significant OSS contribution coefficients for any of the regions examined. This could mean that, if the preceding intra firm variation in OSS activity could be clearly associated with excess returns this would also work on the one month lag, so the firm fixed effects are not predictive.

This heterogeneous pattern across the models, which is significance in A, insignificance in B and then significance in C and D, appears paradoxical if the hypothesis of OSS contribution to the excess stock returns was holding. The non-monotonic pattern can be explained by the high degree of autocorrelation between the contemporaneous and 3 month lagging window metric. This is more likely a data artifact rather than something to do with the timescale on which the work on OSS is transferred to value creation for the firm.

Even in the cases where the OSS coefficient appears statistically significant, the incremental explanatory power appears insignificant. In other words, adding a commit metric to the three factor Fama-French model only explains less than 0.2% of the variation in returns (Figure 5.2), which is not meaningful from an economic perspective or portfolio selection perspective. So even if we were to treat the OSS signal as genuine, it would still not be exploitable realistically.

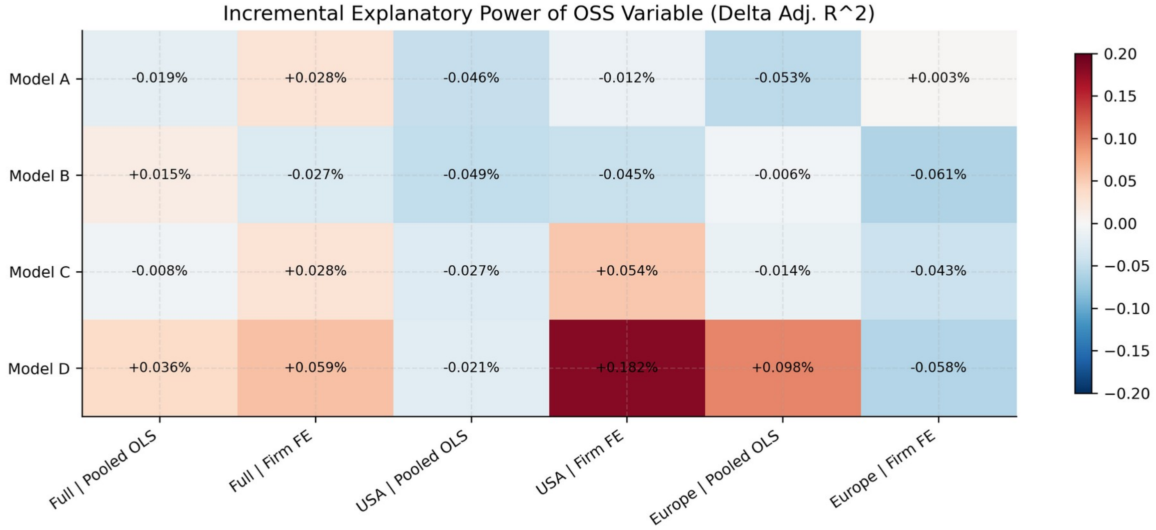

Figure 5.2 Incremental Exploratory Power of OSS Variable

When examining the model on an individual firm level, no firm shows a reliable signal apart from an outlier firm driven by a very small amount of non-zero commit months. This means that the positive associations in the intra firm cases for models A, C and D are not concentrated nor driven by a small group of firms, but there is a rather diffused signal manifesting as significant in the general estimator but too weak to be seen in any individual firm. The distribution of the OSS coefficients for specific firms across each model can be seen in Figure 5.3. This strengthens the idea that if the OSS signal holds no real power, because if it did it would manifest itself in some of the firms.

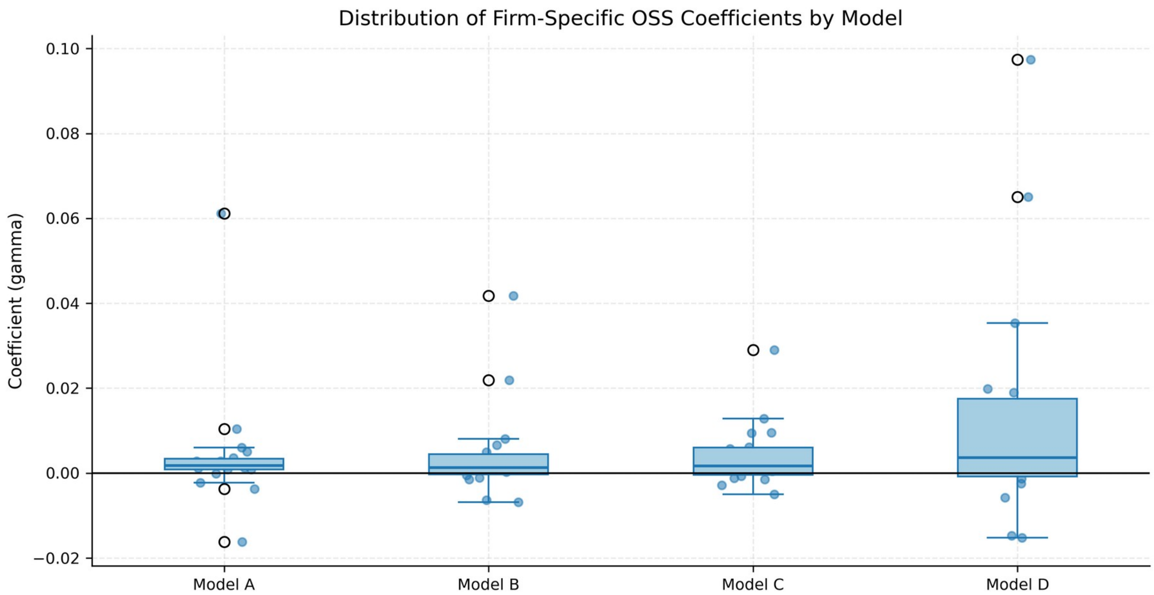

Figure 5.3 Distribution of Firm-Specific OSS Coefficients by Model

When considering the USA versus Europe regions, the heterogeneity between them is obvious, Europe's results are often null relative to the USA sample. Possible explanations for this can be the relatively small OSS footprint among the European firms, the differences between the fragmented and less liquid European market versus the unified USA market, and the sectoral composition of the two regions, with Europe being more heavy in semiconductor and aerospace adjacent industries while the USA being more heavy on pure software enterprises.

From a market perspective, we can say that the OSS signal appears to be either already incorporated in the stock price of the firm or not economically significant in order to drive the stock price of a firm. On the other hand one could argue that the OSS signal might be too noisy or indirect for the markets to be able to interpret it and use it.

5.2 Limitations

On the limitations of this study I have to mention firstly the relatively small size of the firm sample examined. Most of the firms examined are relatively small capitalization and publicly listed, so the results can not be generalized to higher capitalization or privately owned firms.

Another limitation has to do with the metric of the OSS contributions. The commits count might not be the best proxy for OSS contribution as it is a measure of quantity rather than quality. Moreover, the attribution of the commits to specific firms based on email domains may result in false negative or false positive data points.

The characteristics of the sample period might also introduce artifacts in the data and thus be considered as a limitation. Most of the period is covered by a bull USA market with a sharp correction at the end, which means that most of the period was affected by very low risk free rates that might not be a neutral base for testing asset pricing signals. On top of that, after 2022 we have the introduction of AI systems capable to meaningfully contribute to source code and their gradual adoption by technology companies. This could introduce different kinds of artifacts in the OSS data from that moment onward, since firms can now use AI systems to contribute more to OSS projects, less costly and more frequently.

Another effect of the length of the period examined is that it could introduce a survivorship bias. Firms that contributed to OSS but failed to operate over all this time period, either because they were delisted or because they were listed later, will be missing from the sample set.

5.3 Future Research

The current research has the potential for further expansion into different directions. One way is by incorporating a richer OSS contribution activity measure, like adjusting the value of commits by the repository’s popularity or significance, or using some code quality measure. Other sources of OSS can also be used, apart from GitHub, like GitLab, or DockerHub.

Another way for furthering the research would be to use a broader and more diverse sample. Adding more firms from USA and Europe, but also from other markets, could solidify or challenge the results of this dissertation. The period horizon could also be extended to include years further that 2022.

Different models could also be studied, were the dependent variables could be the excess returns over longer horizons, and the OSS metrics could also be over longer rolling windows of the past. Different frequencies could also be studied in order to reveal if major or concentrated OSS contributions can have a more imminent effect on stock prices. Lastly, instead of the three factor Fama-French model, a future study could use the extended five factor Fama-French model or even an entirely different asset pricing model.

5.4 Conclusions

This study examined whether publicly traded technology firms’ OSS contribution activity, measured in monthly commit counts, can be associated with excess stock returns, after adjusting for the three Fama-French factors. The augmented Fama-French model was applied to 20 firms (10 from the USA and 10 from Europe) over a period from November 2012 to October 2022. Four different models and two different pool types (pool OLS and firm fixed effect) were examined. A positive and statistically significant association was identified only in three cases for the full and USA sample, while all other cases revealed no statistically significant associations. Even the statistically significant associations had a very small explanatory increment of less that 0.2%. Examined individually, no firm showed any reliable relationship between OSS contributions and excess returns.

OSS contribution activity as measured in this dissertation, does not constitute a predictive signal for excess returns on stock of the publicly traded firms examined over this specific period. This could mean that the market prices the pubic OSS activity information efficiently or assigns no significant value to it. This finding does not imply that the OSS contributions from technology firms do not have strategic value or competitive effects, influencing on different timescales or through different channels, the value of the firm.

References

Alexy, O., George, G., & Salter, A. J. (2013). Cui bono? The selective revealing of knowledge and its implications for innovative activity. *Academy of Management Review*, *38*(2), 270–291. https://doi.org/10.5465/amr.2011.0193

Association of Chartered Certified Accountants. (n.d.). *Research and development*. https://www.accaglobal.com/gb/en/student/exam-support-resources/fundamentals-exams-study-resources/f7/technical-articles/rd.html

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, *25*(3), 239-264. https://doi.org/10.1007/s11408-011-0165-y

Bijl, L., Kringhaug, G., Molnar, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, *45*, 150-156. https://doi.org/10.1016/j.irfa.2016.03.015

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, *56*(6), 2431-2456. https://doi.org/10.1111/0022-1082.00411

Conti, A., Guzman, J., & Roche, M. P. (2025). Beefing IT up for your investor? Engagement with open source communities, innovation and startup funding: Evidence from GitHub. *Organization Science*, *36*(4), 1551-1573. https://doi.org/10.1287/orsc.2023.18348

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461-1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dahlander, L., & Magnusson, M. G. (2008). How do firms make use of open source communities? *Long Range Planning*, *41*(6), 629-649. https://doi.org/10.1016/j.lrp.2008.09.003

Eberhart, A. C., Maxwell, W. F., & Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, *59*(2), 623-650. https://doi.org/10.1111/j.1540-6261.2004.00644.x

Ekinci, C., & Bulut, A. E. (2021). Google search and stock returns: A study on BIST 100 stocks. *Global Finance Journal*, *47*, 100518. https://doi.org/10.1016/j.gfj.2020.100518

French, K. R. (n.d.). *Data library*. Tuck School of Business, Dartmouth College. Retrieved April 28, 2026, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics, 123*(3), 441-463. https://doi.org/10.1016/j.jfineco.2016.11.004

Finley, K. (2023, June 13). *Non-code contributions are the secret to open source success*. GitHub. https://github.com/readme/featured/open-source-non-code-contributions

Fitzgerald, B. (2006). The transformation of open source software. *MIS Quarterly, 30*(3), 587–598. https://doi.org/10.2307/25148740

Free Software Foundation. (n.d.). *FSF history*. Retrieved May 22, 2026, from https://www.fsf.org/history/

GitHub. (n.d.-a). *About commits*. GitHub Docs. Retrieved May 22, 2026, from https://docs.github.com/en/pull-requests/committing-changes-to-your-project/creating-and-editing-commits/about-commits

GitHub. (n.d.-b). *About pull requests*. GitHub Docs. Retrieved May 22, 2026, from https://docs.github.com/pull-requests/collaborating-with-pull-requests/proposing-changes-to-your-work-with-pull-requests/about-pull-requests

Google Open Source. (n.d.). *Google Open Source*. Retrieved May 22, 2026, from https://opensource.google/

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *The RAND Journal of Economics, 36*(1), 16–38. https://doi.org/10.2307/1593752

Henkel, J. (2006). Selective revealing in open innovation processes: The case of embedded Linux. *Research Policy*, *35*(7), 953-969. https://doi.org/10.1016/j.respol.2006.04.010

Hoffa, F. (2016, June 30). *GitHub on BigQuery: Analyze all the open source code*. Google Cloud Blog. https://cloud.google.com/blog/topics/public-datasets/github-on-bigquery-analyze-all-the-open-source-code

Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D. M., & Damian, D. (2014). The promises and perils of mining GitHub. In *Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 92–101). Association for Computing Machinery. https://doi.org/10.1145/2597073.2597074

Korkmaz, G., Santiago Calderon, J. B., Kramer, B. L., Guci, L., & Robbins, C. A. (2024). From GitHub to GDP: A framework for measuring open source software innovation. *Research Policy*, *53*(3), Article 104954. https://doi.org/10.1016/j.respol.2024.104954

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The Journal of Industrial Economics*, *50*(2), 197-234. https://doi.org/10.1111/1467-6451.00174

Lerner, J., & Tirole, J. (2005). The economics of technology sharing: Open source and beyond. *Journal of Economic Perspectives*, *19*(2), 99-120. https://doi.org/10.1257/0895330054048678

Lev, B. (2003). Remarks on the measurement, valuation, and reporting of intangible assets. *Federal Reserve Bank of New York Economic Policy Review*, *9*(3), 17–22. https://www.newyorkfed.org/medialibrary/media/research/epr/03v09n3/0309levy.pdf

Linux Foundation. (2016, November 16). *Microsoft fortifies commitment to open source, becomes Linux Foundation Platinum member*. https://www.linuxfoundation.org/press/press-release/microsoft-fortifies-commitment-to-open-source-becomes-linux-foundation-platinum-member

Linux Foundation. (n.d.). *Participating in open source communities*. Retrieved May 22, 2026, from https://www.linuxfoundation.org/resources/open-source-guides/participating-in-open-source-communities

Nagle, F. (2018). Learning by contributing: Gaining competitive advantage through contribution to crowdsourced public goods. *Organization Science, 29*(4), 569-587. https://doi.org/10.1287/orsc.2018.1202

Nagle, F. (2019). Open source software and firm productivity. *Management Science, 65*(3), 1191-1215. https://doi.org/10.1287/mnsc.2017.2977

Nguyen, C. P., Schinckus, C., & Nguyen, T. V. H. (2019). Google search and stock returns in emerging markets. *Borsa Istanbul Review, 19*(4), 288-296. https://doi.org/10.1016/j.bir.2019.07.001

Open Source Initiative. (2024, February 16). *The Open Source Definition (Annotated)*. https://opensource.org/definition-annotated

Open Source Initiative. (2025, June 3). *History of the Open Source Initiative*. Retrieved May 22, 2026, from https://opensource.org/about/history-of-the-open-source-initiative

Red Hat. (n.d.). *Red Hat: We make open source technologies for the enterprise*. Retrieved May 22, 2026, from https://www.redhat.com/en

Santiago Calderon, J. B., Robbins, C. A., Guci, L., Korkmaz, G., & Kramer, B. L. (2022). Measuring the cost of open source software innovation on GitHub (BEA Working Paper No. WP2022-10). *U.S. Bureau of Economic Analysis*. https://www.bea.gov/research/papers/2022/measuring-cost-open-source-software-innovation-github

West, J., & Gallagher, S. (2006). Challenges of open innovation: The paradox of firm investment in open-source software. *R&D Management, 36*(3), 319-331. https://doi.org/10.1111/j.1467-9310.2006.00436.x

Young, J.-G., Casari, A., McLaughlin, K., Trujillo, M. Z., Hebert-Dufresne, L., & Bagrow, J. P. (2021). *Which contributions count? Analysis of attribution in open source.* arXiv. https://doi.org/10.48550/arXiv.2103.11007

Appendix A: "Source Code and Result Files"

The source code, along with the data files and the result files, can be found in the following repository:

https://github.com/thmsk/oss-contributions-and-firm-performance

Author's Statement:

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.