On the implications of recent advancements in information technologies and high-dimensional modeling for financial markets and econometric frameworks

Dissertation to obtain the doctoral degree of Economic Sciences (Dr. oec.)

Faculty of Business, Economics and Social Sciences

University of Hohenheim
Institute of Economics (520K)

submitted by
Alexander Schmidt
from Reinbek

Date of submission: August 2019

This thesis has been accepted in 2019 as a dissertation to attain the degree of *Doctor oeconomiae (Dr. oec.)* by the Faculty of Business, Economics and Social Sciences at the University of Hohenheim.

Defense of doctoral thesis: November 26, 2019

Professor Dr. Karsten Hadwich, Dean, University of Hohenheim
Professor Dr. Robert C. Jung, Supervisor, University of Hohenheim
Privatdozent Dr. Thomas Dimpfl, Second assessor, University of Tübingen
Professor Dr. Katja Schimmelpfeng, Chair of the doctoral committee, University of Hohenheim

Contents

List of Figures

List of Tables

Acknowledgments

This thesis is the result of a three-year long journey, during which I've enjoyed continuous support from teachers, colleagues, friends, and family. Thus, before presenting any of the fruits that this journey bore, some thank you's are in order. To begin with, I want to express my sincere gratitude towards Prof. Dr. Robert Jung for welcoming me at his chair, providing me with the opportunity to work and do my doctorate under his guidance, and for his support of my projects. I greatly appreciate the high degree of freedom I was given in both my work and the doctoral thesis and, with this, the trust that he put into me. Largely responsible for this thesis to happen in the first place is PD Dr. Thomas Dimpfl, who has sparked my interest in scientific research and encouraged me to start the PhD. I consider myself very fortunate to have been his student assistant in Tübingen for many years, which – of all university experiences – has probably most profoundly impacted me. I am deeply thankful for all the hours he has invested in me over these past years, for the council he provided me with whenever needed, his honest opinion about my work, and for the faith he puts into me. I also want to thank Prof. Dr. Katja Schimmelpfeng, who generously and on rather short notice agreed to join the committee for the defense of my thesis. Though many of our conversations incidentally took place in hallways, I have always enjoyed them and appreciate her interest in and support of my research.

Besides my supervisory team, I want to thank Dr. Karsten Schweikert for always having an open door for me – figuratively speaking – and my questions, thoughts, and ideas. He always found time whenever I approached him and provided me with guidance in all of my research projects. Likewise did my colleagues Kerstin Bubeck, Konstantin Kuck, Markus Mößler, and Maike Becker provide great support during the time that I have spent at the chair with their helping hands, comments and suggestions on my work, and with outstanding research assistance. A special place in these acknowledgments needs to be reserved for Simon Behrendt. As my co-author, co-founder, most demanding critique, closest friend, and brother, I could list a myriad of situations in which I could count on his unconditional support, his inspiring ideas, and thoughtful advice. Instead, I just say thank you. For everything.

Lastly, I am blessed with a loving family, which always has my back under any circumstances. Though I don't always make it easy for them, they support me at all times and are an essential source of strength for me. For this I cannot express enough gratitude.

Chapter 1

Introduction

Around the turn of the millennium, the Organization for Economic Co-operation and Development (OECD) published an article, which summarizes the organization's expectations towards technological developments of the 21st century (Miller *et al.*, 1998). Of particular interest to the authors are innovations in the area of information technology, highlighting their far-reaching impact on, amongst others, the financial sector. According to the article, the expected increasing interconnectedness of individuals, markets, and economies holds the potential to fundamentally change not only the flow of information in financial markets but also the way in which people interact with each other and with financial institutions. Looking back at the first two decades of the 21st century, these predictions appear to have been quite accurate: The rise of the internet to a "broad platform for a new kind of economy" (Marburger, 2011, p. 209), profoundly impacts how people nowadays receive and process information and subsequently form, share, and discuss their opinions amongst each other. At the financial markets around the globe, trading has become more and more accessible to individuals. Less financial and technical knowledge is required of retail investors to engage in trading, resulting in increased market participation and more heterogeneous trader profiles. This, in turn, influences the dynamics in the financial markets and challenges some of the conventional wisdom concerning market structures. In this context, the interdependencies between the media, retail investors, and the stock market are of particular interest for practitioners (see, for example, Engelberg and Parsons, 2011; Peress, 2014). However, the changed dynamics in the flow and exchange of data and information are also highly interesting from a researcher's perspective, resulting in entire branches of the academic literature devoted to the topic. While these branches have grown in many different directions, this doctoral thesis explores two specific aspects of this field of research: First, it investigates the consequences of the increased interconnectedness of individuals and markets for the dynamics between the new information technologies and the financial markets. This entails both gaining new insights about these dynamics and assessing how investors process certain company-related information for their investment decisions by means of sentiment analysis of large, publicly available data sets. Secondly, it illustrates how an advanced understanding of high-dimensional models, resulting from such analyses of large data sets, can be beneficial in re-thinking and improving existing econometric frameworks.

Before motivating the subsequent chapters, let me first briefly elaborate on the relevance and context of the latter aspect of this thesis. One inevitable consequence of the increasing usage of computers to monitor process flows, execute economic and financial transactions, but also to communicate and exchange opinions is the accumulation of large amounts of data. These data constitute a potentially rich source of information. If analyzed comprehensively, they can provide practitioners and researchers alike with important insights that help to bridge existing knowledge gaps. However, such analyses impose new challenges for econometricians, ranging from plain software restrictions that hinder the analysis of large data sets to analytical challenges – the failure of traditional econometric models to reliably distinguish between signal and noise in huge data sets (Varian, 2014). While the theoretical foundations of statistical learning that provide the tools to address the analytical challenges have already been laid about two decades ago (see, for example, the first edition of Friedman *et al.*, 2001), only the more and more frequent encounter of such large data sets in recent years facilitates rapid methodological progress. From a practitioner's point of view, the most interesting part of the latest developments in statistical learning are the practical implications that follow from the analysis of their data. For researchers, on the other hand, an enhanced understanding of the high-dimensional models themselves used in the analyses has to be achieved first. Thus, in addition to the direct outcome of the analysis of high-dimensional models, new theoretical insights about statistical learning techniques gained from the analysis can be used to approach other types of econometric problems from a new angle. One such technique that is part of this thesis is Tibshirani's (1996) least absolute shrinkage and selection operator (lasso). In contrast to traditional, lower dimensional regression techniques, the lasso enables researchers to investigate the influence and relevance of a broad set of features on some response variable. This is particularly relevant for applications in medical and cancer research on gene expressions (e.g., Simon *et al.*, 2013), yet the lasso's properties are also more and more frequently investigated in a time series context (e.g., Wang *et al.*, 2009). The insights gained from an enhanced understanding of the lasso estimator can then open new paths for improved solutions to different econometric challenges, as will be illustrated below.

Thus, instead of solely focusing on empirical investigations of the changed dynamics in financial markets, this thesis is also concerned with methodological considerations in econometrics. By doing so, the three independent but related research projects of this thesis can give a more holistic picture of the implications that the profoundly changed flow and exchange of data and information of the last decades hold for finance and econometrics. As such, the projects (i) highlight the importance of carefully assessing the dynamics between investor sentiment and stock market volatility in an intraday context, (ii) analyze how investors process newly available, rich sources of information on a firm's Environmental, Social, and Governance (ESG) practices for their investment decisions, and (iii) propose a new approach to detect multiple structural breaks in a cointegrated framework enabled by new insights about high-dimensional models.

The increasingly popular body of the behavioral finance literature that investigates the relationship between investor sentiment and the stock market, which has been briefly touched upon above, constitutes the starting point for the first research project of this thesis. While the question about the nature of this relationship is not new (for early work concerned with this topic see, for example, Barberis *et al.*, 1998; Neal and Wheatley, 1998; Lee *et al.*, 1991), the advent of the internet, resulting in increased interconnectedness among investors, has been hugely influential in altering the dynamics of the relationship. Enabling rapid distribution of financial news among market participants and the subsequent formation and exchange of opinions, analyses of investor sentiment have shifted more and more from relying on financial columns of newspapers (e.g., Tetlock, 2007; Garcia, 2013) to studying the influence of social media (e.g., Chen *et al.*, 2014; Zheludev *et al.*, 2014). Of particular relevance in this context is the concept of noise traders as they are characterized by Kyle (1985) or Black (1986). Such noise traders are assumed to be easily swayed by, among others, opinions published online. In turn, they then form their own opinion based on the content of, for example, social media posts. This often results in impulsive and mostly irrational trading based on their newly formed believes (Barber and Odean, 2007). While large institutional investors in general usually are not assumed to behave similarly to noise traders, they could anticipate and exploit the behavior of these noise traders and gain an advantage by systematically analyzing and subsequently trading on social media information (De Long *et al.*, 1990).

The value of assessing the information contained in social media is not clear a priori, however, it holds the potential to enrich and augment existing forecasting models and to provide new insights into investor behavior as several authors show. In an earlier study, Antweiler and Frank (2004) document the usefulness of stock messages to help to predict market volatility. In their work, the authors analyze daily messages posted on the message boards of Yahoo Finance and Raging Bull in the year 2000. While their sampling period clearly belongs to the early years of online investor sentiment, the authors are among the first to establish a model that shows the predictive power of investor sentiment, extracted from the message boards via a Naive Bayes text classifier, for stock market movements. Since then, the focus of attention has shifted from message boards to social media platforms. Some of these platforms exclusively target the financial community, which is why they appear to be particularly interesting for researchers. For example, Chen *et al.* (2014) use data from Seeking Alpha, which is a crowdsourcing service for financial markets, and also find predictive power of articles and commentaries posted on that platform for future stock returns and earnings surprises. However, not only the analysis of discussion platforms that are particularly targeting participants in the financial markets can help to explore the new dynamics of the relationship between investor sentiment and stock market movements. Financial communities are also forming on more general-interest social media platforms such as Twitter (Yang *et al.*, 2015). Speculating about potential reasons for this, one could argue that noise traders acquire a substantial part of their financial knowledge from such general-interest platforms instead of relying exclusively on finance discussion boards. Recalling the definition of noise traders from above, retail investors' trading on Twitter information matches the concept of uninformed or irrational trading behavior quite well. If such retail investors would consult more specialized online stock market communities as their primary source of financial information, their trading, though still considered as noise trading, might be less irrational and uninformed. This train of thought has led several authors to put Twitter at the center of their attention, exploring the usefulness of a Twitter-based sentiment analysis for financial (prediction) models (e.g., Bollen *et al.*, 2011; Sprenger *et al.*, 2014a,b; Ranco *et al.*, 2015; Oliveira *et al.*, 2017). While Chapter 2 provides more insights about the findings of the related literature, for now it suffices to say that despite their different contexts, all of the studies find significant predictive power of microblogging sentiment for their respective financial variable of interest.

However, the effects that the aforementioned authors find all occur in a daily research framework. Since Twitter could be considered one of the fastest moving online vehicles with most users receiving push notifications on their mobile devices to be informed about new events in real-time, one gap in the literature is the investigation of potential intraday effects of Twitter investor sentiment on the stock market. Such an intraday assessment is also becoming increasingly relevant for the financial market, where the speed with which transactions are being executed has dramatically increased over the past years, as briefly elaborated above. Especially the assessment of intraday volatility has become increasingly important, not only for high frequency traders but also for applications in risk management (Giot, 2005; Engle and Sokalska, 2012). Therefore, the first original work presented in Chapter 2 of this doctoral thesis aims at closing this gap in the literature. More specifically, the paper titled **The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility**, which is joint work with Simon Behrendt, takes a closer look at the dynamics of individual-level stock return volatility, measured by absolute 5-minute returns, and Twitter sentiment and activity in an intraday context.[1] After accounting for the intraday periodicity in absolute returns, we discover some statistically significant co-movements of intraday volatility and information from stock-related Twitter messages (Tweets) for all constituents of the Dow Jones Industrial Average (DJIA). However, economically, the effects are of negligible magnitude, and out-of-sample forecast performance is not improved when including Twitter sentiment and activity as exogenous variables. From a practical point of view, this chapter finds that high-frequency Twitter information is not particularly useful for highly active investors with access to such data for intraday volatility assessment and forecasting when considering individual-level stocks.

Inspired by this first research project, the second original work presented in this thesis keeps its focus on sentiment analysis in the context of the financial markets. At the same time, it decisively diverges from Chapter 2 by exploring another increasingly relevant phenomenon in the financial markets to which investors and the public appear to pay particularly close attention: Companies' management of environmental, social, and governance issues. Since the New York Stock Exchange has started the Principles of Responsible Investment (PRI) in 2006, which was shortly thereafter followed by the Sustainable Stock Exchange Initiative (SSEI) in 2007, the role of socially and environmentally responsible business practices for publicly listed companies has changed dramatically. ESG investing in assets under management amounted to about $20 trillion in the year 2018 – roughly a quarter of all professionally managed assets worldwide (Kell, 2018). ESG-focused assets under management are further expanding at a current growth rate of 20% per year (Reid *et al.*, 2018).

Despite the clear picture that these numbers paint of the relevance of the topic, the extent to which the recently spiking demand for ever increasing ESG-efforts is driven by investors who are willing to pay a premium on ESG-affine companies remains unclear. The academic literature, though increasingly concerned with the issue, has thus far not reached a consensus on the influence of ESG activities on firm value. While some authors find empirical evidence that speaks in favor of value-adding ESG (e.g., Nofsinger and Varma, 2014; Cahan *et al.*, 2015; Lins *et al.*, 2017), there is equally strong support for a more pessimistic view on the effects of socially and environmentally responsible business practices (e.g., Brammer *et al.*, 2006; Krüger, 2015; Capelle-Blancard and Petit, 2019). The increasingly easy access to publicly available information on companies' ESG-activities now provides the chance for new insights about the way in which investors process these information. Instead of relying on annual company reports or single ESG-events, online media offer the latest ESG-related news on a daily basis. By assessing the sentiment conveyed by each piece of news, one can then analyze the correlation between ESG-related news sentiment and stock market movements. Exploiting these newly available tools to monitor and analyze ESG information or the sentiment formed toward them should serve to reflect the clearly intangible nature of ESG efforts in the first place. Since to the best of my knowledge the literature has so far not investigated investors' reaction to newly arriving ESG information by means of sentiment analysis, Chapter 3 addresses this research gap.

[1] This paper has originally been published as *Behrendt, S. and Schmidt, A. (2018) The Twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility, Journal of Banking & Finance, 96, 355-367.* `https://doi.org/10.1016/j.jbankfin.2018.09.016`

Moreover and in contrast to the work presented in Chapter 2, the second original work of this thesis does not base its analysis on a publicly available sentiment index but constructs a domain-specific index from scratch. Domain-specific here refers to the fact that only ESG-related news are considered for the calculation of our sentiment time series. Thus, the second project also elaborates in greater detail on how a sentiment index can be constructed from raw data with an application in behavioral finance that is of utmost importance for today's financial markets. More precisely, Chapter 3, **Sustainable news – A sentiment analysis of the effect of ESG information on stock prices**, investigates the effect of ESG-related news sentiment on the stock market performance of the DJIA constituents. Relying on a large data set of news articles that were published online or in print media between the years of 2010 and 2018, each article's sentiment with respect to ESG-related topics is extracted using a dictionary approach from which a polarity-based sentiment index is calculated. Estimating autoregressive distributed lag models reveals significant effects of both temporary and permanent changes in ESG-related news sentiment on idiosyncratic returns for the vast majority of the DJIA constituents. According to the models' results, one can assign the stocks to different groups depending on their investors' apparent predisposition towards ESG news, which in turn seems to be linked with a stock's financial performance.

This chapter closes the exploration of the first aspect of this thesis – the implications of the changing information technologies for the financial markets. Chapter 4 is then concerned with the question of how an enhanced methodological understanding of high-dimensional datasets can produce new solutions to familiar problems in econometrics. One such a familiar problem is the identification and subsequent consistent estimation of structural breaks in real-world time series. Most economic or financial variables, such as exchange rates, gross domestic product, or spot and futures prices, when observed over multiple years, are likely to experience structural change of one kind or another. These could be related to economic or political events, technological development, or regulatory changes, just to name a few (Perron, 2006; Aue and Horváth, 2013). To complicate matters further, such variables are oftentimes analyzed in multivariate time series models and thus are potentially cointegrated. Cointegration, as described by Engle and Granger (1987), is the condition in which two or more time series are nonstationary in levels, but maintain stable long-run equilibria. Modeling the parameters of this long-run relationship as being constant over time and thus ignoring potential structural breaks renders a cointegration test and subsequent conclusions drawn inaccurate at best, more often outright misleading. Several authors have thus far contributed to this highly relevant problem in econometrics, as Chapter 4 elaborates in greater detail. The proposed testing frameworks, however, can be considered as quite rigid and restrictive, since they often require prior knowledge about the exact amount of break points (Gregory and Hansen, 1996a; Hatemi-J, 2008) or are only applicable for a certain small number of maximum breaks per time series (Kejriwal and Perron, 2010). Therefore, the third research project presented in this doctoral thesis illustrates the benefits of using dimension reduction techniques for high-dimensional models to overcome the limitations of conventional econometric models.

The paper **Multiple structural breaks in cointegrating regressions: A model selection approach**, which is joint work with Karsten Schweikert, proposes the least absolute shrinkage and selection operator, which has been briefly mentioned above, as a tool for consistent breakpoint estimation. The aim of this project is to highlight the usefulness of one of the most commonly applied approaches in statistical learning for the analysis of potentially cointegrated systems of economic time series. In this paper, we propose a new approach to model structural change in cointegrating regressions using penalized regression techniques. First, we consider a setting with fixed breakpoint candidates and show that a modified adaptive lasso estimator can consistently estimate structural breaks in the intercept and slope coefficient of a cointegrating regression. In such a scenario, one could also perceive our method as performing an efficient subsample selection. Second, we extend our approach to a diverging number of breakpoint candidates and provide simulation evidence that timing and magnitude of structural breaks are consistently estimated. Third, we use the adaptive lasso estimation to design new tests for cointegration in the presence of multiple structural breaks, derive the asymptotic distribution of our test statistics and show that the proposed tests have power against the null of no cointegration. Finally, we use our new methodology to study the effects of structural breaks on the long-run PPP relationship.

The subsequent sections present the three research projects of this thesis in greater detail, in the same order that they are mentioned above (*The Twitter myth revisited* in Chapter 2, *Sustainable news* in Chapter 3, and *Multiple structural breaks in cointegrating regressions* in Chapter 4). While each chapter embodies a standalone paper, which introduces and concludes its respective topic, it is worthwhile to reflect upon the findings of the three chapters in a broader context, similar to how they are being motivated in this chapter. The final chapter of this thesis, Chapter 5, is devoted to this purpose.

Chapter 2

The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility

2.1 Introduction

With the ever increasing speed of trading in recent years, intraday volatility assessment and forecasting have gained importance for highly active investors such as derivative traders and hedge funds. Intraday volatility measures are important input factors in high-frequency risk management applications, for the calculation of time-varying liquidity measures, and to concert limit order placement strategies or the optimal scheduling of trades (e.g., Engle and Sokalska, 2012; Giot, 2005). However, not only has the speed of trading increased rapidly but also the way investors can comment or share their opinion about company and stock market performances on social media platforms.

A growing body of behavioral finance literature links investor sentiment, derived from social media, to financial markets (for a recent survey, see Bukovina, 2016). While institutional investors have the means to monitor actively traded stocks constantly, social media represents one channel through which retail investors can easily access stock market relevant information (e.g., Chen *et al.*, 2014). Stock prices, reflecting the trading activities of both institutional and retail investors, might reflect retail investor trading activities that are, at least partially, influenced by sentiment. Let us view the average highly active investor as a professional or institutional investor, close to the definition of an informed investor. By contrast, individual or retail investors are often thought of as having psychological biases and are seen as noise traders in the way portrayed by Kyle (1985) or Black (1986). While professional investors are seen as rational investors, they can still base decisions on less rational factors such as investor sentiment. Early research on investor sentiment has proposed that such rational investors could bet against sentiment driven noise traders to make a profit, albeit with caution to the costs and risks that such strategies would entail (e.g., De Long *et al.*, 1990; Shleifer and Vishny, 1997). Thus, given that retail investors have been shown to trade excessively in attention-grabbing stocks (Barber and Odean, 2007) and to trade in concert (e.g., Kumar and Lee, 2006; Barber *et al.*, 2009), one might think that professional investors could exploit the behavior of retail investors, who use social media platforms as investment forums to obtain information about securities' potential performance.

Recently, the social media platform Twitter has been used to extract a proxy for investor sentiment. For instance, Bollen *et al.* (2011) derive six social mood dimensions from Tweets. Their results indicate that predictions of the DJIA are improved through the inclusion of some of these social mood dimensions. Sprenger *et al.* (2014a) derive good and bad news from more than 400,000 Tweets related to the S&P 500 and find that these news have an impact on the market. In addition, Sprenger *et al.* (2014b) discover a relationship between stock related Twitter sentiment and returns, volume of Tweets and trading volume of the respective stock, as well as disagreement and return volatility. Looking at the transmission and aggregation of information, they also demonstrate that providing above average investment advice is associated with more quotes and an increase in followers. Along this line, Yang *et al.* (2015) unravel the existence of a financial community on Twitter and find that the weighted sentiment of its most influential contributors has significant predictive power for market movement. While Sprenger *et al.* (2014b) focus on some well-known companies from the S&P 100, other studies from the behavioral finance literature look at stock market indices only. Moreover, all studies mentioned above focus on daily stock market and social media data.

Taking an intraday perspective and considering individual-level stocks, this paper has two main objectives: (i) assessing the impact of Twitter investor sentiment and Twitter activity on return volatility and (ii) testing the performance of intraday volatility forecasts augmented with this additional information. Following, among others, Andersen and Bollerslev (1997) and Bollerslev *et al.* (2000), we use high-frequency absolute 5-minute returns as a measure for volatility, since these display greater dynamics, i.e., more persistent autocorrelation patterns and thus conform better to the long-memory property of stock return volatility, than squared returns (for a discussion of this finding see, for example, Ding *et al.*, 1993; Forsberg and Ghysels, 2007). Twitter sentiment and activity, the latter measured as the number of Tweets in a certain time interval, are available to investors through commercial data providers. By assuming the role of a professional investor with access to such data, we obtain intraday prices, Twitter sentiment, and the number of Tweets (henceforth Twitter count) at 1-minute frequency for all constituents of the DJIA and a time period from June 18, 2015 to December 29, 2017 from Bloomberg. We focus on blue-chip stocks such as the DJIA constituents, since other securities are not equally well covered in terms of Twitter sentiment and activity, rendering an intraday analysis infeasible. Before conducting any meaningful time series analysis and intraday volatility forecasting, we address the well-documented intraday periodicity in absolute 5-minute returns within the framework of a two-step estimation procedure involving a Fourier Flexible Form (FFF) estimation (for example, see Andersen and Bollerslev, 1997; Bollerslev *et al.*, 2000). This approach is readily applied to the intraday absolute returns of the DJIA constituents. In order to examine the dynamics between Twitter sentiment, activity, and return volatility, the filtered absolute 5-minute returns obtained from this estimation procedure are then used in a bivariate Vector-AutoRegressive (VAR) model together with average 5-minute Twitter sentiment and 5-minute Twitter count, respectively. Finally, we adapt the Heteroscedastic AutoRegressive (HAR) model of Corsi (2009) to the intraday context and use a panel HAR to forecast filtered absolute 5-minute returns for individual-level stocks. While we observe some statistically significant feedback effects between intraday volatility and Twitter sentiment as well as Twitter count for many stocks, the performance of the panel HAR model, augmented with exogenous information from Twitter, is mixed among the sample of stocks considered in this paper. In general, estimated coefficients are small in magnitude and gains in out-of-sample forecasting performance, if present at all, are negligible in every single case. Thus, professional investors do not benefit from augmenting forecasts with such Twitter data when considering individual-level stocks and an intraday setting. Our results, obtained from the sample of 30 DJIA constituents, clearly differ from research that considers aggregated data in the form of stock market indices and daily observations of index returns and social media sentiment or activity. Gains of using high-frequency financial and social media data are limited, rendering the discussion more interesting for the case of observations at lower frequencies. Moreover, one could say that our results are in line with the notion of professional investors as stated above: the performance of liquid blue-chip stocks should be determined by information that is related to securities' fundamentals and not by investor sentiment obtained from social media platforms. The majority of such stocks are held by institutional investors and thus should be priced more efficiently (e.g., Boehmer and Kelley, 2009).

Our paper is structured as follows: Section 2.2 describes the absolute return time series data and shortly outlines the two-step estimation procedure used to account for the deterministic intraday periodicity that is present in intraday absolute returns. Section 2.3 describes the Twitter data in more detail and assesses the interactions between intraday Twitter sentiment, Twitter count and intraday filtered absolute returns in a bivariate VAR framework. Forecasting of intraday volatility, using exogenous information in the form of Twitter sentiment and Twitter count in a panel HAR setting, is the objective of Section 2.4. Lastly, Section 2.5 concludes.

2.2 Intraday periodicity and long-memory volatility

2.2.1 Data

Intraday prices for all DJIA constituents at 1-minute frequency are obtained from Bloomberg and cover the period from June 18, 2015 to December 29, 2017. For each trading day and stock up to 390 prices are obtained, corresponding to the regular trading hours from 9:30 Eastern Time (ET) to 16:00 ET. Continuously compounded returns are calculated as the log-price changes from one minute to the next. Accordingly, 5-minute returns are calculated as the sum of five 1-minute returns. Excluding overnight returns, this leaves 77 intraday 5-minute returns for each trading day and stock. Occasionally, it is the case that there is no trading recorded by Bloomberg over a given 5-minute time interval, leading to missing values. However, these cases are not frequent and missing values should not distort our empirical results. With a total of 639 trading days, each consisting of 77 intraday 5-minute returns, this leaves us with a total of 49,203 observations. Thus, for all DJIA constituents denote the series of 5-minute returns as $R_{t,n}$, where $t = 1,2,\dots,639$ and $n = 1,2,\dots,77$. Another time series that is used in the two-step estimation procedure to get rid of the pronounced intraday periodicity in absolute returns, explained in more detail below, consists of daily closing prices ranging from January 4, 2010 to December 29, 2017. Analogously, daily returns are calculated as the log-price changes between two consecutive days.

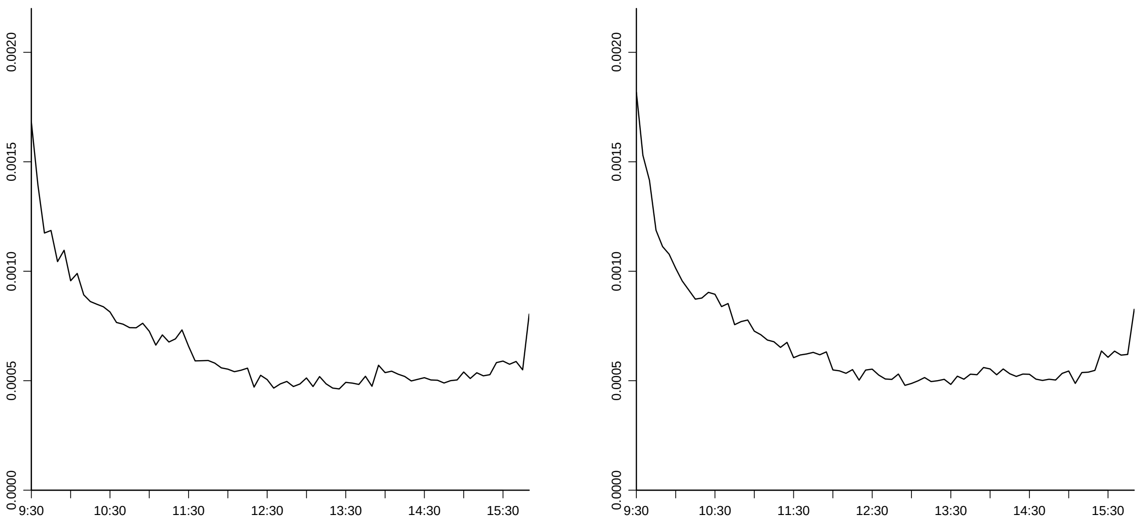
2.2.2 Intraday periodicity in absolute returns

When analyzing 5-minute absolute returns, a clear pattern emerges that has been documented in the literature by, among others, Andersen and Bollerslev (1997), Andersen and Bollerslev (1998), Andersen *et al.*

(2000) and Bollerslev *et al.* (2000): While the volatility process shows clear conditional heteroscedasticity and a pronounced long-memory property on a daily basis, one can observe a strong deterministic intraday periodicity.

The intraday periodicity is illustrated in Figure 2.1, where the average intraday absolute 5-minute return is calculated over the cross-section of each of the n intraday bins. Thus, absolute 5-minute returns are plotted for the "average" trading day. In the following, results are shown for the stocks of two companies, International Business Machines Corporation (*ibm*) and Walmart Inc. (*wmt*), but results are similar for all remaining constituents of the DJIA and are available upon request.[1] Figure 2.1 reveals a distinct difference in the volatility over the trading day. For both stocks one can see that volatility is high at the beginning of the trading day, decreases throughout the day with its minimum around lunch hours, and increases again slightly at the end of the trading day. Several early papers have attributed the pronounced U-shape pattern in intraday stock market volatility to the strategic interaction of traders around market openings and closures (for example, see Admati and Pfleiderer, 1988, 1989). Interestingly, in the case of individual-level stock absolute 5-minute returns, this "U"-shape rather resembles an inverted "J"-shape, since, on average, volatility right after the market opening dwarfs volatility at the market closing.

Figure 2.1: Average of absolute 5-minute returns
The plots show the average of intraday absolute 5-minute returns calculated for two stocks, (a) *ibm* and (b) *wmt*.



(a) *ibm* **(b)** *wmt*

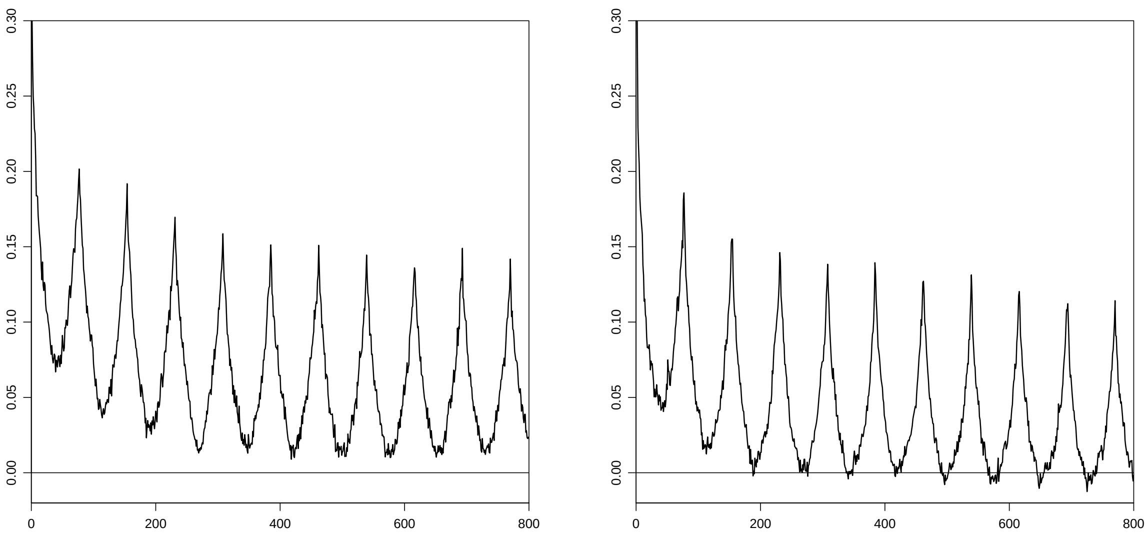
Moreover, this deterministic intraday periodicity induces a certain pattern in the autocorrelation functions (ACFs) of absolute 5-minute returns, which is visible in Figure 2.2. Here, ACFs are plotted over ten trading days. However, the given pattern is consistent over the whole sample period. On the one hand, one can observe the above mentioned long-memory properties of absolute 5-minute returns, since the ACFs decay slowly and are statistically significant over a long time horizon. On the other hand, a clear repetitive pattern can be seen. The deterministic intraday periodicity induces a distorted U-shape in the sample autocorrelations, each of these lasting exactly for one trading day.

Previous research has shown that in order to conduct meaningful time series analysis and to derive

[1]These two stocks are chosen since they display both significant Twitter sentiment and Twitter count terms in all models that we entertain in our empirical analysis. Table 2.5 in the appendix provides an overview over all stocks.

intraday forecasts of 5-minute absolute returns, one has to take both of these patterns into account. One way to purge the intraday absolute 5-minute returns of the periodic component is by applying a two-step procedure based on an FFF estimation (Gallant, 1981). The approach outlined in the following has first been introduced by Andersen and Bollerslev (1997) and is easily applied to individual-level stock absolute return time series. The next section provides a short description of this estimation procedure, while a longer and more technical description can be found in Appendix 2.6.1.

Figure 2.2: ACFs of absolute 5-minute returns
The plots show ACFs of intraday absolute 5-minute returns over 800 5-minute lags, which correspond to approximately ten trading days, for two stocks, (a) *ibm* and (b) *wmt*.



(a) *ibm* **(b)** *wmt*

2.2.3 Two-step estimation procedure

To model the periodic intraday volatility component in high-frequency absolute returns, we follow Andersen and Bollerslev (1997) and decompose 5-minute returns as:

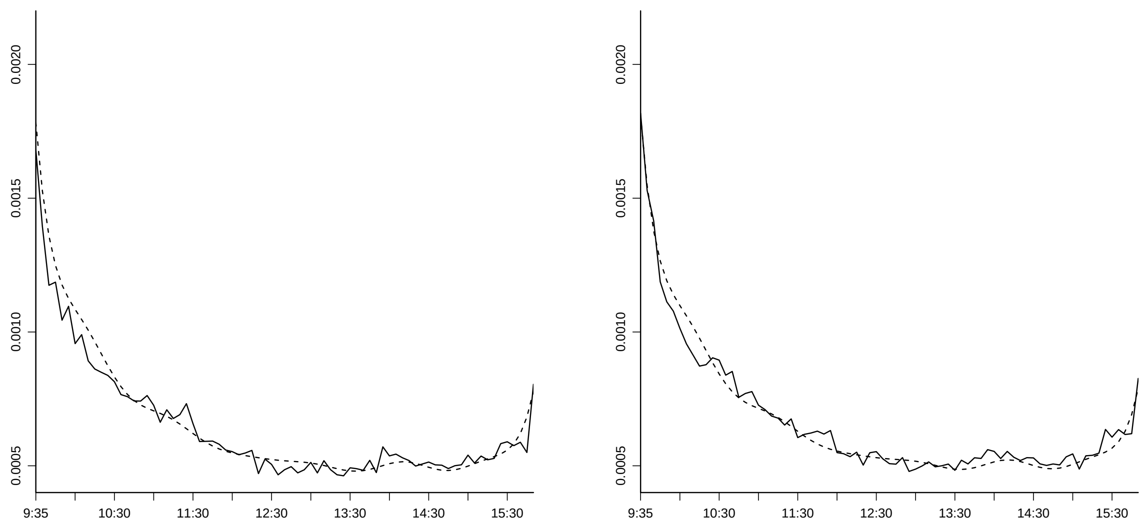
$$R_{t,n} - \mathbb{E}(R_{t,n}) = \varepsilon_{t,n} = s_{t,n}\sigma_{t,n}Z_{t,n} \tag{2.1}$$

where $\sigma_{t,n}$ denotes a 5-minute volatility factor for trading day t and $Z_{t,n}$ is an i.i.d. zero mean and unit variance innovation. The periodic component, $s_{t,n}$, is estimated in a two-step estimation procedure that involves an FFF regression, which is illustrated in more detail in Appendix 2.6.1. In order to obtain $\sigma_{t,n}$, σ_t is estimated using the longer sample of daily returns from January 4, 2010 until December 29, 2017. The sample is chosen such that there is a larger number of observations available for estimation but it excludes the financial crisis. We use an asymmetric power ARCH (A-PARCH) to capture the daily volatility clustering. The A-PARCH of Ding *et al.* (1993) does not only allow for a leverage effect in the volatility equation but also accounts for the empirical finding that the sample autocorrelation of absolute returns is usually higher than that of squared returns. Ding *et al.* (1993) show empirically that the A-PARCH is able to capture the long-memory properties of daily absolute returns. The 5-minute volatility factor for trading day t, $\sigma_{t,n}$, is then simply estimated by $\hat{\sigma}_{t,n} = \hat{\sigma}_t / N^{1/2}$, where N is the number of observations per trading day.

The second step involves estimating the parameters of the FFF specification by Ordinary Least Squared (OLS). Estimation is based on the whole sample of intraday 5-minute returns, instead of simply estimating the average periodic pattern across the trading day. This two-step procedure is not fully efficient. However, Andersen and Bollerslev (1998) show that, in general, the parameter estimates are consistent, given the FFF regression is correctly specified in the second step.

While the actual parameter estimates are difficult to interpret, one can plot the average estimated intraday periodic volatility factor together with the average absolute 5-minute returns in order to see whether or not the estimate provides a sufficient approximation of the intraday shape of average returns. This is depicted in Figure 2.3. The average periodic component seems to approximate the distinct shape of absolute returns quite well and thus the two-step estimation procedure does seem to be a reasonable approach in our case.

Figure 2.3: Average of absolute 5-minute returns and intraday periodic volatility component
The plots show the average of intraday absolute 5-minute returns calculated for two stocks, (a) *ibm* and (b) *wmt*. In addition, the dashed line denotes the superimposed estimated average intraday periodic volatility component, appropriately scaled.



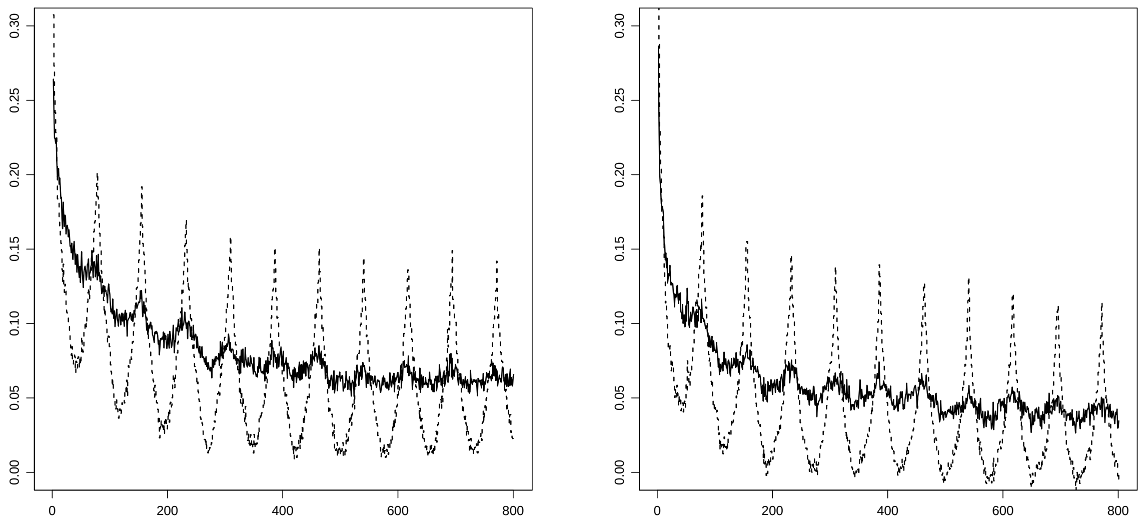
(a) *ibm* **(b)** *wmt*

However, while the A-PARCH estimate, $\hat{\sigma}_t$, may successfully capture the volatility clustering in the daily returns, it might not be a good model for $\hat{\sigma}_{t,n}$. Following, for example, Bollerslev *et al.* (2000) the estimated seasonal component in the 5-minute absolute returns is filtered away to see whether or not the chosen approach is valid empirically. Denote the raw absolute 5-minute returns by $|R_{t,n}|$, the filtered 5-minute absolute returns are then given by:

$$R^*_{t,n} = \frac{|R_{t,n}|}{\hat{s}_{t,n}}, \tag{2.2}$$

where $\hat{s}_{t,n}$ denotes the normalized estimate for the periodic component, as obtained from the two-step estimation procedure. In accordance with Andersen and Bollerslev (1997), the autocorrelogram of the filtered absolute returns should exhibit a strictly positive and slowly declining autocorrelation. This would indicate that the long-memory properties are the characteristic attribute of the return volatility process, after the deterministic intraday component is removed. As can be seen from Figure 2.4, this is exactly the case for the two chosen stocks.

Figure 2.4: ACFs of raw and filtered absolute 5-minute returns
The plots show ACFs of raw (dashed lines) and filtered (solid lines) intraday absolute 5-minute returns over 800 5-minute lags, which correspond to approximately ten trading days, for two stocks, (a) *ibm* and (b) *wmt*.



(a) *ibm* **(b)** *wmt*

The ACFs of the filtered absolute 5-minute returns seem way smoother than the ACFs of the raw absolute 5-minute returns with their distinct U-shaped pattern, exhibiting a strictly positive and slowly declining correlogram. Again, results are similar across all constituents of the DJIA.

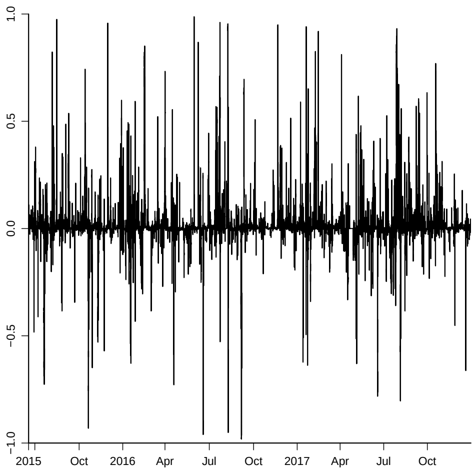
2.3 Twitter sentiment and Twitter count effects

2.3.1 Data

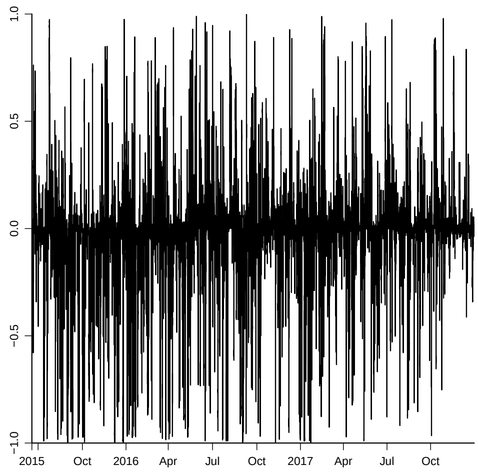
In addition to the return time series, intraday Twitter sentiment and count data for all DJIA constituents at a 1-minute frequency are obtained from Bloomberg for June 18, 2015 through December 29, 2017. While Twitter count measures the overall activity, i.e., the number of Tweets for a given stock and minute, Twitter sentiment ranges continuously from -1 (negative investor sentiment) to 1 (positive investor sentiment). Both measures are based on an undisclosed algorithm used by Bloomberg. Dealing with Twitter count data from Bloomberg is straightforward: In our data set we code Twitter count as zero for a given stock in minutes without any Twitter activity and record the number of Tweets in minutes where Bloomberg registers some Twitter activity. Twitter sentiment is calculated by Bloomberg every minute for all stocks using the last 30 minutes of available data on positively and negatively associated Tweets. However, only if the absolute difference between the newly calculated sentiment value and the previous value is larger than 0.005 does Bloomberg update Twitter sentiment for the respective stock. Accordingly, in our data set we only update the value for Twitter sentiment if for a given stock and minute a change in sentiment is observed in the data obtained from Bloomberg. If there is no observed change in sentiment for a given stock and minute, we instead fill in such missing values with the previously observed change in Twitter sentiment on a given trading day. Let us be more precise and use an example to illustrate how we deal with this issue in our data: If the first observed sentiment value for a given stock is at 10:21 ET and the second one at 10:43 ET on a given trading day, then all values until 10:21 ET are missing values in our data set for this stock and trading day. The first non-missing value is at 10:21 ET and all values between 10:21 ET and 10:43 ET are equal to the sentiment observed at 10:21 ET, only at 10:43 ET do we again record a change in investor sentiment. Thus, we assume that investor sentiment, as obtained from Twitter, remains constant for time periods where Bloomberg does not register a change in Twitter sentiment larger than 0.005 in absolute value. Lastly, in order to match the 5-minute intraday frequency of the return data, time series of 5-minute Twitter sentiment are obtained as the average sentiment over five minutes for each stock, whereas time series of Twitter count constitute the absolute number of counts over each 5-minute time interval for each stock.

Figure 2.5: Twitter sentiment time series and ACFs

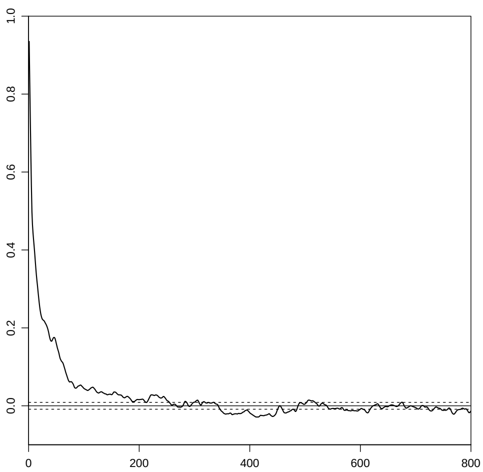
Plots (a) and (b) show the time series of Twitter sentiment (TS) for *ibm* and *wmt*, respectively. Plots (c) and (d) illustrate ACFs of the TS for these stocks over 800 5-minute lags, which correspond to approximately ten trading days. The dashed lines indicate 95% confidence bounds.



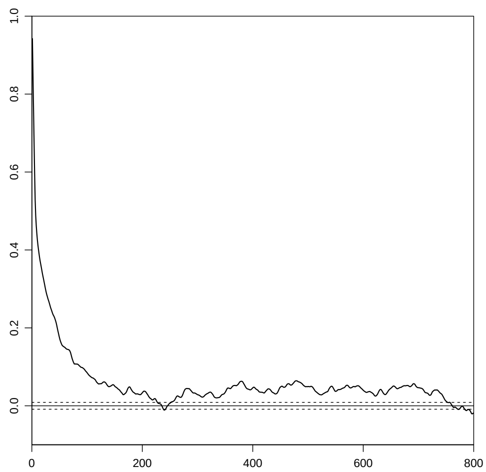
(a) TS: *ibm*



(b) TS: *wmt*



(c) ACF TS: *ibm*



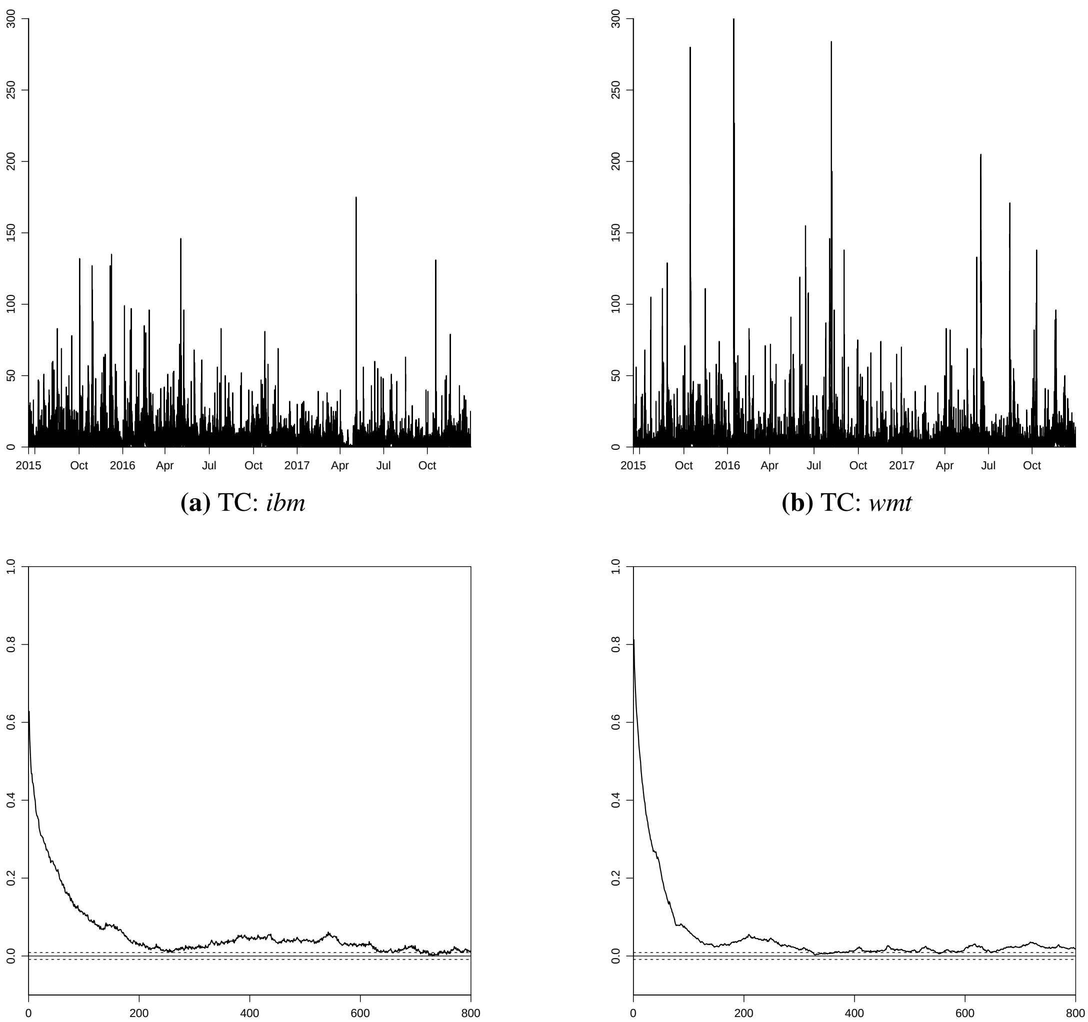
(d) ACF TS: *wmt*

The upper panels of Figure 2.5 illustrate the time series of Twitter sentiment for the stocks of *ibm* and *wmt*, the two lower panels show the respective ACFs over ten trading days. While the sentiment time series for *wmt* has a higher variability compared to *ibm*'s sentiment time series, both display long-memory dependencies in their ACFs. The effect of past sentiment or count values on the present remains significant for more than three trading days for both stocks, with significant lags even after nine trading days for *wmt*.

Similarly, for Twitter count the upper panels of Figure 2.6 illustrate the respective time series for *ibm* and *wmt*, whereas the two lower panels depict their ACFs. One can see that at most times the number of Tweets for *wmt* exceeds the number of Tweets for *ibm*. However, for both stocks the Twitter count time series appears to possess a certain long-memory property. Compared to Twitter sentiment, the memory of the Twitter count time series for *ibm* seems to be longer, as Plot (c) of Figure 2.5 shows significant lags even after seven trading days. For *wmt*, the ACFs of the Twitter count time series behave similarly to its sentiment counterpart with positive and significant lags for up to 10 trading days. Overall, no clear recurring (intraday) pattern can be found in the Twitter time series for our sample of the 30 DJIA constituents.

Figure 2.6: Twitter count time series and ACFs

Plots (a) and (b) show the time series of Twitter count (TC) for *ibm* and *wmt*, respectively. Plots (c) and (d) illustrate ACFs of the TC for these stocks over 800 5-minute lags, which correspond to approximately ten trading days. The dashed lines indicate 95% confidence bounds.



(a) TC: *ibm* **(b)** TC: *wmt*

(c) ACF TC: *ibm* **(d)** ACF TC: *wmt*

2.3.2 Interactions between intraday Twitter sentiment, Twitter count and volatility

In order to allow for a feedback effect of return volatility, given by the filtered absolute 5-minute returns obtained as discussed in Section 2.2, to Twitter sentiment as well as Twitter count and vice versa, a simple bivariate VAR model is entertained. Stated in structural form:

$$\begin{bmatrix} 1 & 0 \\ b_{21}^{0} & 1 \end{bmatrix} \begin{bmatrix} R_{t,n}^{*} \\ twit_{t,n} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} b_{11}^{j} & b_{12}^{j} \\ b_{21}^{j} & b_{22}^{j} \end{bmatrix} \begin{bmatrix} R_{t,n-j}^{*} \\ twit_{t,n-j} \end{bmatrix} + \begin{bmatrix} u_{1t,n} \\ u_{2t,n} \end{bmatrix} \tag{2.3}$$

where $twit_{t,n}$ denotes either Twitter sentiment or count. For the specification with Twitter count we need to relax the assumption of Gaussian white noise innovations $\{u_{it,n}\}_{t=1,n=1}^{TN}$ where $i = 1,2$, $T = 639$, $N = 77$. This assumption is usually made in the VAR context, yet count data are non-negative integers and cannot be normally distributed (Cameron and Trivedi, 1986). The relaxation of the normality assumption of the innovations, however, does not affect our VAR analysis. It has no effect on the estimation of the parameters of the VAR model but is only important for correct inference, which is not a main issue here. We choose the lag order p of the VAR model according to the average lag length suggested by the Schwarz information criterion (SC) across all 30 DJIA constituents, which leads to $p = 17$ for the specification with filtered absolute 5-minute returns and Twitter sentiment and $p = 18$ for the specification with Twitter count data.[2] Even though the ACFs of the VAR residuals still show some significant spikes for *ibm* and *wmt*, in light of the small magnitudes of the coefficient estimates, as reported below, these statistically significant lags do not appear to be of economic relevance.

The results of both VAR specifications are illustrated in Table 2.1. Since we are mostly interested in the question of whether or not lags of Twitter sentiment and Twitter count have a significant impact on filtered absolute 5-minute returns, only the results for Twitter lags significant at a 10% significance level are displayed. Significant effects might indicate that intraday forecast augmentation of return volatility could be possible using exogenous information from Twitter. Panel A shows these significant autoregressive terms for the Twitter variables throughout both models. The Granger causality tests, which can be found in Panel B of Table 2.1, support the finding that Twitter sentiment and count indeed hold statistically significant information about future return volatility. The hypothesis that the Twitter variables do not Granger cause volatility can be rejected for both stocks and specifications, except for *ibm*'s Twitter sentiment.[3] However, the actual estimates of lagged Twitter sentiment and count in the VAR specifications are rather small in magnitude, indicating a statistically significant but economically not relevant influence of the Twitter variables on the filtered absolute 5-minute return series. This impression is reinforced by the contemporaneous correlation matrices of the reduced form VAR residuals, as presented in Table 2.2. The correlations between the filtered absolute return and Twitter variable residuals are smaller than 0.02. Furthermore, decompositions of the filtered absolute returns' forecast error variances show no relevant contribution of either Twitter variable for any of the DJIA constituents (less than 2% of the forecast error variance).

[2] Robustness checks show that adjusting the lag length for each stock individually does not affect our main estimation results.
[3] An overview over the Granger causality tests for all 30 DJIA constituents can be found in Table 2.5 in the appendix.

Table 2.1: VAR model and Granger causality results

The first two columns of Panel A show the results of the VAR model with Twitter sentiment as the second system variable, the last two columns the ones with Twitter count. Only estimates of the respective Twitter variable significant at the 10% level are displayed. Coefficient estimates are multiplied by 10^3. Panel B shows the F-statistics of the Granger causality test. The respective H_0 tested is indicated in the first column. P-values are given in parentheses. Significance of the Granger causality F-statistics at the 10% level is highlighted in bold face.

Panel A: VAR estimation results $\times 10^3$

	Twitter sentiment		Twitter count	
	ibm	*wmt*	*ibm*	*wmt*
$twit_{t,n-1}$			0.0024 (0.0001)	0.0012 (0.0192)
$twit_{t,n-2}$	0.2504 (0.0938)			
$twit_{t,n-4}$				−0.0017 (0.0043)
$twit_{t,n-5}$				0.0019 (0.0013)
$twit_{t,n-9}$			0.0011 (0.0802)	
$twit_{t,n-10}$	0.2807 (0.0696)			
$twit_{t,n-12}$		−0.2104 (0.0041)		
$twit_{t,n-13}$		0.1922 (0.0088)		
$twit_{t,n-14}$				−0.0017 (0.0039)
$twit_{t,n-18}$			0.0019 (0.0021)	0.0018 (0.0006)

Panel B: Granger causality test

	Twitter sentiment		Twitter count	
H_0	*ibm*	*wmt*	*ibm*	*wmt*
$R^* \not\rightarrow twit$	1.2607 (0.2076)	0.8019 (0.6929)	**5.9968** (0.0000)	**40.4163** (0.0000)
$twit \not\rightarrow R^*$	1.0957 (0.3503)	**1.7018** (0.0352)	**3.2663** (0.0000)	**2.8456** (0.0000)

Table 2.2: Contemporaneous residual correlation matrices

The table shows the contemporaneous correlations between the reduced form VAR residuals. For the VAR in Panel A Twitter sentiment is used as the second system variable, in Panel B Twitter count.

Panel A: Twitter sentiment

	ibm R^*	*ibm* $twit$	*wmt* R^*	*wmt* $twit$
R^*	1	-0.0079	1	-0.0032
$twit$	-0.0079	1	-0.0032	1

Panel B: Twitter count

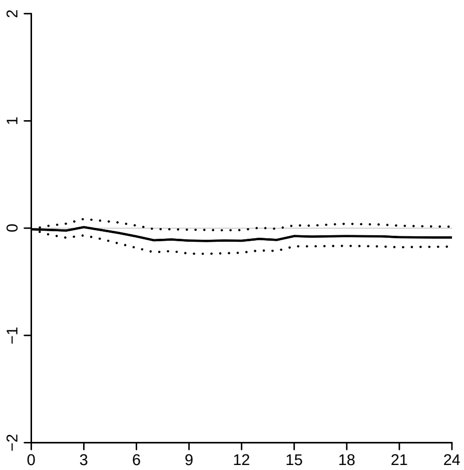
	ibm R^*	*ibm* $twit$	*wmt* R^*	*wmt* $twit$
R^*	1	0.0068	1	0.0191
$twit$	0.0068	1	0.0191	1

Based on the estimation results of the VAR model, we can now investigate the evolution of shocks to either the volatility measure or the respective Twitter variable through the system by means of impulse response analysis. In order to uniquely identify the effect of these shocks, we use a Cholesky decomposition in which return volatility is ordered first, followed by the respective Twitter variable (see Equation

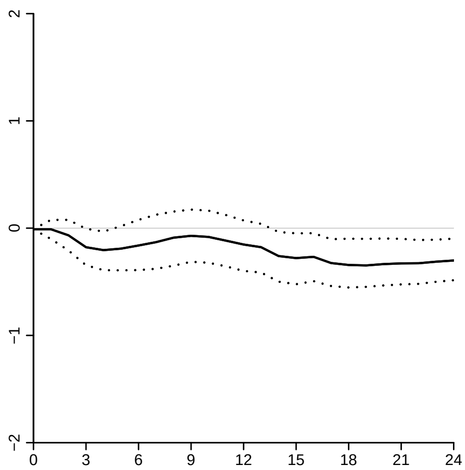
(2.3)). This ordering implies the exclusion restriction that only shocks to volatility can affect the Twitter variables contemporaneously, whereas shocks to the Twitter variables cannot affect absolute returns in the same period. This restriction appears to be sensible, since one would expect a fundamental shock in volatility to appear first, which then, in turn, influences investor sentiment and activity as captured by the Twitter variables (see Dimpfl and Jank, 2016; Lux and Marchesi, 1999). Investor sentiment and activity, on the other hand, can be assumed to be contemporaneously affected by stock performance and changes in return volatility. Figure 2.7 depicts the impulse responses of volatility and Twitter sentiment to shocks in one of the system variables. A 10% shock in absolute returns leads to a negative reaction in Twitter sentiment for both *ibm* and *wmt*. However, this effect is statistically not distinguishable from zero over the first hour (12 lags) after the shock. Only then there appears to be a significant, slightly negative impact of the volatility shock on Twitter sentiment, but solely for *wmt*. A one unit shock in Twitter sentiment leads to minor short-run fluctuations of volatility which are barely distinguishable form zero for both stocks.

Figure 2.7: Impulse response functions Twitter sentiment

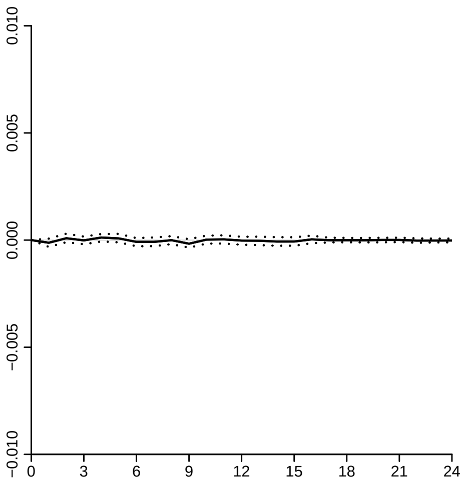
The plots show orthogonal impulse response functions over 24 5-minute lags for the VAR specification with filtered absolute returns (volatility) and Twitter sentiment as system variables. The dashed lines indicate 95% confidence bounds.



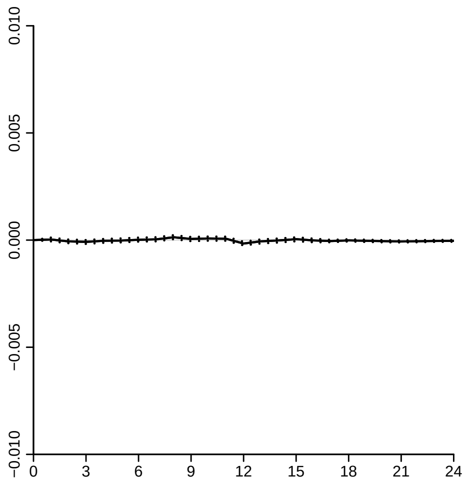
(**a**) *ibm*: volatility → Twitter



(**b**) *wmt*: volatility → Twitter



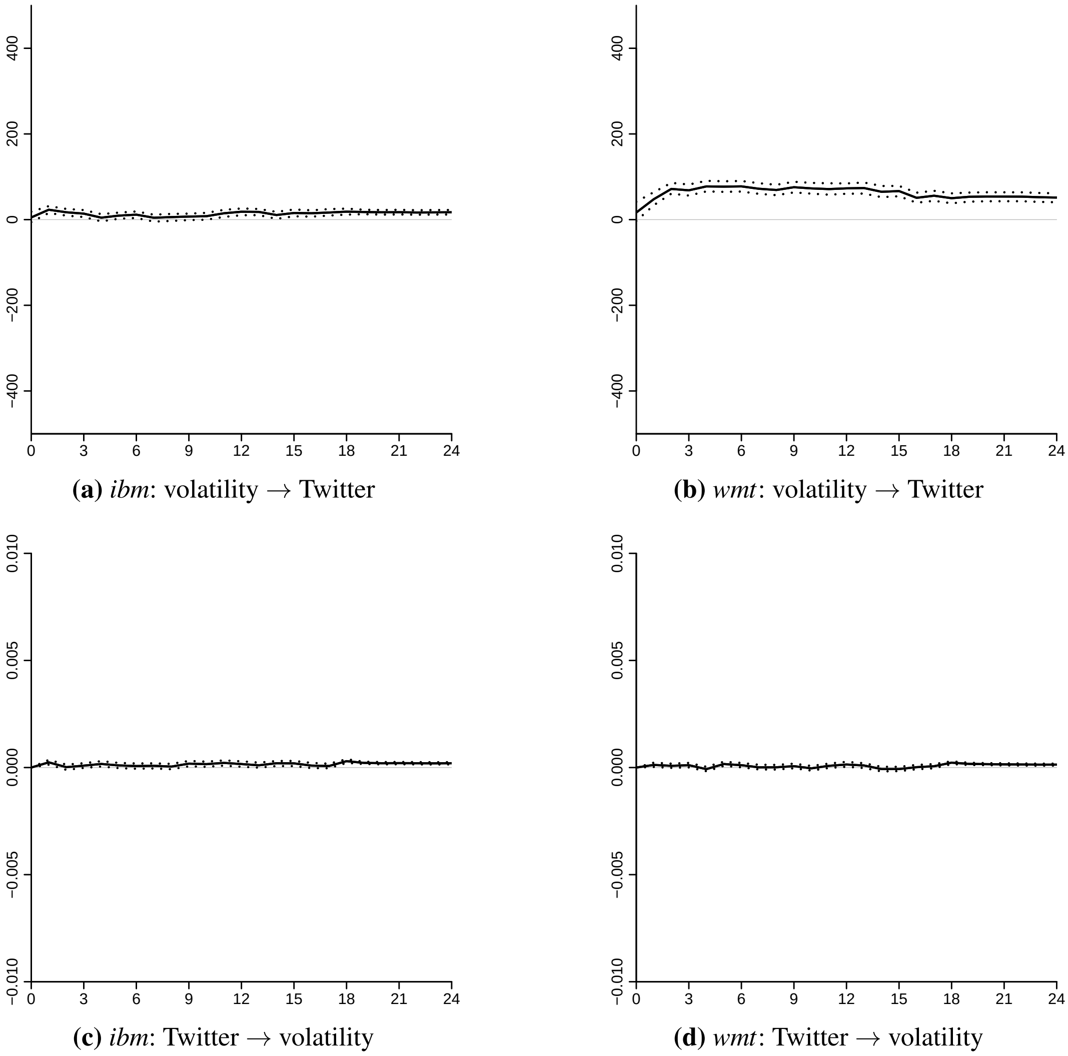
(**c**) *ibm*: Twitter → volatility



(**d**) *wmt*: Twitter → volatility

For the VAR specification with Twitter count as the second system variable, Figure 2.8 shows that a 10% shock in absolute returns leads to an increase in the number of Tweets. While for *ibm* this effect is only slightly significant, the *wmt* stock shows what appears to be a rather persistent, positive reaction from the first period onwards. Turning to the reaction of the filtered returns to a shock in Twitter count, as shown in the two lower graphs, the effect does not appear to be distinguishable from zero.

Figure 2.8: Impulse response functions Twitter count
The plots show orthogonal impulse response functions over 24 5-minute lags for the VAR specification with filtered absolute returns (volatility) and Twitter count as system variables. The dashed lines indicate 95% confidence bounds.



(a) *ibm*: volatility → Twitter **(b)** *wmt*: volatility → Twitter

(c) *ibm*: Twitter → volatility **(d)** *wmt*: Twitter → volatility

The results of the impulse response analysis are robust to a re-ordering of the system variables. All in all, while there are some significant feedback effects between filtered absolute 5-minute returns and the Twitter variables, the prospects for a meaningful forecast augmentation using exogenous information from Twitter seem to be rather poor.

2.4 Forecasting intraday volatility with Twitter information

In light of the previous analysis in Section 2.3, we choose a different approach to forecast intraday volatility and adapt Corsi's (2009) HAR model to the intraday context. The HAR model is chosen since it is a parsimonious model that has been shown to sufficiently capture the long-memory properties of daily realized volatility (e.g., Andersen *et al.*, 2007; Chiriac and Voev, 2011). The HAR model contains aggregates of the absolute filtered returns and Twitter time series as right hand side variables, which should lead to an improvement in terms of predictive power of the HAR model over the previously estimated VAR model. For example, Dimpfl and Jank (2016) use a HAR model that is augmented with lagged Google search queries as an additional exogenous variable and find that their model leads to an improvement in forecast precision out-of-sample in comparison to their VAR specification. For our application, however, the model structure of a conventional HAR model applied to our intraday time series would inevitably produce spillover effects and averages of Twitter and volatility variables that are calculated across trading days. As this is not desirable in the intraday context, we facilitate a panel structure in the HAR model using $t = 1,\dots,639$ as the cross-sectional unit (trading days) and $n = 1,\dots,53$ for intraday periods. Whilst the notation with respect to the indices has not changed in comparison to the VAR model, the HAR model now entails a panel model interpretation. Patton and Sheppard (2015) follow a similar approach in that they set up a panel HAR for volatility modeling based on daily observations to account for firm-specific effects. However, the adaptation of the HAR model to the intraday context has – to the best of our knowledge – not been pursued by other authors so far. The panel HAR model reads as follows:[4]

$$\begin{aligned} R^*_{t,n} &= c + \beta_1 R^*_{t,n-1} + \beta_{12} R^{*12}_{t,n-1} + \beta_{24} R^{*24}_{t,n-1} + \delta_1 twit_{t,n-1} \\ &\quad + \delta_{12} twit^{12}_{t,n-1} + \delta_{24} twit^{24}_{t,n-1} + \gamma_1 sgn(\bar{R}_{t,n-1}) + u_{t,n} \end{aligned} \tag{2.4}$$

where $sgn(\bar{R}_{t,n-1})$ denotes the sign of the average return in the previous 5-minute interval and $R^{*12}_{t,n-1} = \frac{1}{12}\sum_{j=1}^{12} R^*_{t,n-j}$ and $R^{*24}_{t,n-1} = \frac{1}{24}\sum_{j=1}^{24} R^*_{t,n-j}$ are lagged averages for one and two hours of the filtered returns, respectively. $twit^{12}_{t,n-1}$ and $twit^{24}_{t,n-1}$ are calculated analogously for both Twitter variables. Assuming that (professional) investors with access to intraday Twitter data can react swiftly to changes in individual-level stock return volatility and the Twitter variables, further lags are omitted. The sign variable is added to the model to account for the asymmetric effect of returns, i.e., negative returns have a larger effect on volatility than positive returns. The panel HAR of equation (2.4) is subsequently estimated using fixed effects estimation with fixed effects for trading days and adjusted standard errors to account for heteroscedasticity as well as serial correlation.[5]

For forecasting purposes, the overall sample is split into a sample containing 90% (44,283 observations) of the data to which the panel HAR model is fitted. The remaining 10% (4,920 observations) of the data are used to assess the forecasting performance of the model out-of-sample. This is achieved by predicting the absolute 5-minute returns using the coefficient estimates of equation (2.4) together with the exogenously given Twitter sentiment and count data and comparing the predicted values to the actual

[4] We have also estimated a HAR model with more lags of the dependent and exogenous variable, as well as different hourly aggregates as independent variables. Since the results stay robust across all models, we have chosen the most parsimonious one.

[5] The Hausman (1978) test rejects a random effects model in favor of a model with fixed effects. Arellano (1987) standard errors are used, however, the results are robust to other types of robust standard errors.

filtered absolute 5-minute returns. As a measure for forecast performance the root mean squared error (RMSE) is used.

Table 2.3: Panel HAR in-sample results

The table presents the in-sample parameter estimates of the panel HAR model. Panel A shows the results for *ibm* and *wmt* with Twitter sentiment as exogenous variable, Panel B with Twitter count as exogenous variable. P-values are given in parentheses.

	Panel A: Twitter sentiment		**Panel B: Twitter count**	
	ibm	*wmt*	*ibm*	*wmt*
$R^*_{t,n-1}$	0.0441**	0.0540***	0.0440**	0.0536***
	(0.0227)	(0.0000)	(0.0226)	(0.0000)
$R^{*12}_{t,n-1}$	0.1496***	0.0971**	0.1496***	0.1013**
	(0.0027)	(0.0282)	(0.0029)	(0.0203)
$R^{*24}_{t,n-1}$	−0.0805**	−0.1228***	−0.0805**	−0.1230***
	(0.0348)	(0.0000)	(0.0371)	(0.0000)
$sgn(\bar{R}_{t,n-1})$	0.0000	0.0000	0.0000	0.0000
	(0.8952)	(0.3343)	(0.8674)	(0.3169)
$twit_{t,n-1}$	0.0000	0.0000	0.0000***	0.0000*
	(0.7516)	(0.7270)	(0.0065)	(0.0714)
$twit^{12}_{t,n-1}$	−0.0001	0.0001**	0.0000	0.0000
	(0.5852)	(0.0487)	(0.4354)	(0.3725)
$twit^{24}_{t,n-1}$	0.0001**	0.0000	0.0000	0.0000
	(0.0288)	(0.9384)	(0.9543)	(0.1507)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3 summarizes the panel HAR model with Twitter sentiment (Panel A) and Twitter count (Panel B) as additional exogenous variables. For both stocks, lagged values of the respective Twitter variable occasionally show significant coefficient estimates. However, their magnitudes are very small, which does not allow any further meaningful interpretation of the coefficients.[6]

Considering these results, the RMSEs – our forecast accuracy measure – as shown in Table 2.4 should not surprise: Twitter can only slightly lower the forecasting errors across the different HAR specifications and thus does not appear to hold any additional information that is of practical relevance for forecasting volatility.

Table 2.4: Panel HAR forecast evaluation

Overview of the RMSE ($\times 10^4$) for the out-of-sample forecast of the panel HAR model with three specifications: first row HAR model without exogenous variable; second row with Twitter sentiment (TS) as exogenous variable; last row with Twitter count (TC) as exogenous variable.

	ibm	*wmt*
HAR	6.8934	9.0461
HAR + TS	6.8798	9.0309
HAR + TC	6.8706	8.8692

[6] These results are robust to different lag structures as well as to a panel HAR model which includes both Twitter sentiment and count as exogenous variables.

To sum up, the results of the panel HAR models, as shown above, are in line with the results of the VAR models of the previous section: In terms of statistical significance Twitter sentiment and count are indeed relevant for return volatility. Nevertheless, the influence of exogenous Twitter information on the stocks' volatility is, economically speaking, small. With respect to forecasting, Twitter's predictive power can be described as weak at best, just as suggested by the preceding empirical analysis.[7] Including Twitter sentiment and Twitter count does not appear to improve forecast performance significantly.

While we have presented detailed results for two stocks only, Table 2.5 in the appendix sums up the results of our analyses for all constituents of the DJIA. One noticeable difference in our results among the DJIA constituents is a link that exists with the stocks' average trading volume. DJIA constituents that rank low in trading volume more often show a statistically significant influence of Twitter sentiment and count than those with a high average trading volume. While this influence does not appear to be of economic significance for any of the 30 stocks, our results of both the HAR and VAR models are thus not entirely robust across all 30 constituents of the DJIA. Only for the most liquid stocks of the Dow Jones can we rule out a statistically significant effect of Twitter on the respective stock's volatility.

2.5 Concluding remarks

In this paper we use intraday Twitter sentiment and Twitter count data to measure investors' interest in individual-level stocks, in our case the constituents of the DJIA. Measuring intraday volatility with absolute 5-minute returns and after accounting for the pronounced intraday periodicity in absolute returns, we find that there are indeed statistically significant feedback effects of return volatility to Twitter sentiment as well as Twitter count and vice versa in a bivariate VAR framework. However, the estimated coefficients are of small absolute magnitude and the effects do not have a significant economic impact. While Twitter sentiment and count Granger-cause return volatility, the contemporaneous correlations between volatility and both Twitter variables as well as the results from forecast error variance decompositions indicate that incorporating exogenous information from Twitter into intraday prediction models for return volatility is unlikely to have a significant impact on forecast performance. We adapt the HAR model of Corsi (2009) to the intraday context and estimate a panel HAR model, augmented with lagged Twitter sentiment and Twitter count information. As suspected from the preceding analysis, there are no gains in out-of-sample forecast performance, compared to models without exogenous Twitter information. We present our results for stocks of two companies (*ibm* and *wmt*) but results are similar for all constituents of the DJIA and different model specifications.

Thus, it seems that intraday information from Twitter about individual-level stocks, as provided by commercial data vendors, does not constitute a valuable source of information for future volatility and professional, highly active investors with access to such data do not benefit with regards to intraday volatility assessment and forecasting. Our results are in line with the notion of professional investors: The performance of liquid blue-chip stocks such as the DJIA constituents should be linked to information related to fundamentals, indicating that investor sentiment obtained from Twitter should only have a negligible effect on financial volatility. This is even more so, since the intraday frequencies con-

[7] In addition to the panel HAR model, an ARMA model with exogenous variables and up to 19 lags has been estimated too. Though this model is over-parameterized, a necessity to get rid of the autocorrelation in the residuals, the results support the weak predictive power of both Twitter variables that the panel HAR model has found. Results of the ARMA model are available upon request.

sidered here are too high for investors, other than professional investors, to react appropriately to such information. Thus, the empirical analysis of the effects of investor sentiment obtained from social media platforms such as Twitter on stock return properties are most likely rendered more interesting for lower frequencies. This is consistent with previous literature that mainly uses daily observations (e.g., Bollen *et al.*, 2011; Sprenger *et al.*, 2014b). While we rank our overall results by average trading volume, we only consider constituents of the DJIA in this paper. Future research should further investigate the feedback effects between investor sentiment obtained from social media platforms and intraday volatility of less liquid stocks, in order to test the validity and robustness of the findings presented in this paper.

2.6 Appendix

2.6.1 Fourier flexible form estimation procedure

The logarithm of the squared periodic component in Equation (2.1), $ln(s^2_{t,n})$, can be estimated from the following FFF regression:

$$\begin{aligned} 2ln\left(\frac{|R_{t,n}-\bar{R}|}{\hat{\sigma}_t/N^{1/2}}\right) =& c+\delta_{0,1}\frac{n}{N_1}+\delta_{0,2}\frac{n^2}{N_2} \\ &+\sum_{p=1}^{P}\left(\delta_{c,p}cos\frac{2\pi p}{N}n+\delta_{s,p}sin\frac{2\pi p}{N}n\right)+v_{t,n} \end{aligned} \tag{2.5}$$

where $\bar{R}$ denotes the sample mean of the 5-minute returns (might also be set equal to zero, since it is not statistically different from zero for any stock in the sample), $\hat{\sigma}_t$ is a previously obtained estimate of the daily volatility factor, N refers to the number of return intervals per trading day (here $N = 77$), P is a tuning parameter for the number of trigonometric terms, and $N_1 = (N+1)/2$ as well as $N_2 = (N+1)(N+2)/6$ are normalizing constants. In accordance with Bollerslev *et al.* (2000) and other research, the number of polynomial terms is restricted to two.

We use an A-PARCH specification to estimate σ_t in a first step. In addition, a simple AR(1)-GARCH$(1,1)$ specification serves as a benchmark. Results are not much different compared to the A-PARCH specification. In fact, for many stocks the estimates of σ_t are very similar across these two models. Our model specifies the mean equation in terms of an AR(1) process, since for more than half of the stocks in the sample such an autoregressive structure seems appropriate when considering the statistical significance of the lagged return coefficient. Specifications have also been tested with an MA(1) structure in the mean equation. However, the coefficient of the MA term is statistically significant for a smaller number of stocks. We choose the same model for all stocks, instead of estimating different models for each of the DJIA constituents. More sophisticated ARMA structures are not applied to the mean equation, since the simple autoregressive structure already delivers good empirical results. For the A-PARCH$(1,1)$, the mean equation is given by:

$$R_t = \mu_0 + \mu_1 R_{t-1} + \varepsilon_t \tag{2.6}$$

The variance equation is modeled in the following way:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^{\delta} + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^{\delta} \tag{2.7}$$

where $\delta \in \mathbb{R}^+$ is a Box-Cox transformation of σ_t, and γ the coefficient of the leverage term. Estimates are obtained by assuming conditionally skewed-t distributed standardized innovations, $\varepsilon_t\sigma_t^{-1}$, and the 5-minute volatility estimator is calculated as $\hat{\sigma}_{t,n} = \hat{\sigma}_t/N^{1/2}$.

In the second step, the parameters of the FFF specification are estimated by OLS. The log-transformation is used to draw in outliers and to render the regression more robust. In line with the literature, $P = 6$ is assumed to capture the basic shape of the intraday volatility pattern. Experimenting with different values for P we find that for $P < 6$ the trigonometric terms of different orders were statistically significant for most stocks. However, for $P > 6$ higher order trigonometric terms are only statistically significant for a handful of stocks. Thus, $P = 6$ is chosen as an appropriate order of expansion. Denote the raw absolute

5-minute returns by $|R_{t,n}|$, the filtered 5-minute absolute returns are then given by:

$$R^*_{t,n} = \frac{|R_{t,n}|}{\hat{s}_{t,n}} \tag{2.8}$$

where $\hat{s}_{t,n}$ denotes the normalized estimate for the periodic component, as obtained from the FFF regression. Let $\hat{x}_{t,n}$ denote the estimated value of the right-hand side of the FFF specification. The standardized periodic component is then given by:

$$\hat{s}_{t,n} = TN \frac{exp(\hat{x}_{t,n}/2)}{\sum\limits_{t=1}^{T} \sum\limits_{n=1}^{N} exp(\hat{x}_{t,n}/2)} \tag{2.9}$$

where now $\frac{1}{TN} \sum\limits_{t=1}^{T} \sum\limits_{n=1}^{N} \hat{s}_{t,n} \equiv 1$.

Apart from daily A-PARCH or GARCH models, Engle and Gallo (2006) propose to model the daily volatility component based on the square root of daily realized variance, which is given by $RV_{t,N} = \sum\limits_{j=1}^{1/N} R^2_{t-1+jN,N}$. Similar to the calculation above, the 5-minute volatility estimator is simply given by $\hat{\sigma}_{t,n} = (RV_{t,N}/N)^{1/2}$. In order to reduce the impact of microstructure effects, 10-minute and 15-minute returns are used in the realized variance calculation. Results are robust to calculating the daily volatility component using the square root of daily realized variance over both 10-minute and 15-minute intraday returns.

2.6.2 Overview: Results for all DJIA constituents

In Table 2.5 we present an overview of the statistical significance of Twitter sentiment and count in both VAR and HAR models as well as the Granger causality results for both VAR specifications, considering the whole sample of all 30 DJIA constituents. Additional data on trading volume is taken from Thomson-Reuters Datastream for the period from June 18, 2015 to December 29, 2017.

Table 2.5: Model results for all DJIA constituents

The table summarizes our results for all DJIA constituents in descending order of average trading volume (column 2). Each stock with at least two significant Twitter variables is marked with × in columns 3 to 5. Column 3 summarizes results for both VAR models. Column 4 refers to the panel HAR model in equation (2.4) and column 5 to an alternative model with additional autoregressive terms for 30 minutes and three hours for both R^* and $twit$. F-statistics of the Granger causality test can be found in columns 6 to 7 and columns 8 to 9 for Twitter sentiment and count, respectively. The H_0 tested is indicated above each column. Significance on the 10% level is highlighted.

					Sentiment		Count	
Ticker	Volume	VAR	HAR1	HAR2	$R^* \nrightarrow T$	$T \nrightarrow R^*$	$R^* \nrightarrow T$	$T \nrightarrow R^*$
ge	453.81	×			**2.1684**	0.6976	**5.9543**	**2.8337**
aapl	367.75			×	**1.4721**	0.6154	**5.6040**	**6.0759**
msft	285.29				0.9605	**1.4860**	1.2739	0.8573
pfe	264.22	×			1.0670	0.1066	**3.8680**	1.1818
intc	247.81				1.3820	0.3625	**2.9899**	**1.7498**
csco	233.83			×	1.0993	**1.5003**	0.7855	**2.8876**
jpm	154.75				**1.9473**	0.9993	**1.8870**	1.0505
vz	148.85				0.7375	0.6828	0.4858	1.1721
ko	128.42				1.2546	1.3345	0.9386	**1.7326**
xom	123.74	×	×		1.2731	0.7470	1.1769	**1.7594**
mrk	102.41				0.6854	1.1696	1.2390	0.4046
pg	96.07				0.6206	0.9298	**33.2960**	**1.5941**
nke	94.80	×			0.5325	1.2436	**3.4281**	**2.7298**
wmt	94.38	×	×		0.8019	**1.7018**	**40.4163**	**2.8456**
v	85.88		×	×	0.8170	0.5741	1.3083	1.0021
dis	82.61	×			**1.7140**	0.6003	**25.7273**	**9.8504**
dd	76.55	×			1.2967	1.0419	**9.6421**	**4.7287**
cvx	76.36	×			1.1816	1.1762	**3.0121**	**2.4587**
jnj	69.96	×			**2.6715**	**2.1258**	**3.4540**	**2.6726**
cat	52.74				**2.9053**	1.0739	**10.3373**	**4.5908**
hd	48.94	×			0.9799	**2.1205**	**9.3463**	**3.3585**
axp	48.54	×			0.7538	**2.5288**	**15.9147**	**6.1008**
mcd	48.41	×		×	1.0901	1.0417	**1.9112**	**1.5883**
ibm	42.22	×	×		1.2607	1.0957	**5.9968**	**3.2663**
utx	40.30	×			0.6168	**1.8660**	**20.1767**	**18.2685**
ba	37.95	×			0.8354	0.6984	**8.1744**	**2.1790**
unh	35.05	×			0.8481	1.3064	**4.9244**	**2.4981**
gs	33.80	×			1.0125	**1.6159**	**3.5307**	**2.1091**
mmm	20.98				**1.5369**	0.9916	0.9079	**1.6085**
trv	16.64				0.5003	1.1115	0.8267	0.5104

Chapter 3

Sustainable news – A sentiment analysis of the effect of ESG information on stock prices

3.1 Introduction

From reports about child labor or environmentally harmful supply chains to the distribution of potentially health-threatening food – there is almost no multinational enterprise that has never to some degree been accused of or demonstrably involved in ethically debatable business practices. Such incidents have been observed frequently for many decades now, usually accompanied by a temporary public outrage and, depending on the scale of the scandal, followed by expressions of remorse and promises to improve in the future on the side of the accused business. Recently, however, firms' performance in the area of environmental, social, and governance issues appears to be particularly closely monitored by their investors and the public, manifesting in the emergence of "sustainable investment funds" or demand for "green finance" (Gilbert, 2019). What has also changed is the amount and quality of data available to actually investigate if and if so how investors respond to information about a company's ESG performance. Since empirical evidence that could help to answer this question is still sparse and contradictory in the academic literature, this paper sets out to fill this research gap by extracting ESG information from publicly available news articles and investigating their relationship with the stock market performance of the DJIA constituents.

We are building our research on the work of several authors whose theories, findings, and opinions on this matter are quite controversial. According to an early essay by Friedman (2007), the only social responsibility of a firm is to create legal profits. This view would imply that any kind of ESG activity, which is not part of the core business of an enterprise, should not be undertaken by the company nor should investors incorporate ESG-related information into their investment decisions other than by withdrawing their capital from companies that engage in such activities. Supporting evidence for this effect of ESG activities is provided by Brammer *et al.* (2006), who find that companies with higher social performance scores show smaller returns than those with lower social performance scores. Similarly, Krüger (2015) and Capelle-Blancard and Petit (2019) discover that positive ESG-related information can potentially harm a firm's market value in the short-run. Providing more insights as to the potential mechanism behind such observations, Cheong *et al.* (2017) find that most companies have a reactionary attitude towards ESG matters. These companies excessively engage in ESG activities only after they have experienced negative market and investor sentiment in the previous year, where market and investor sentiment is captured by a modified version of the index of Baker and Wurgler (2006). While this finding on its own would only hold implications with respect to how altruistic a company's motives behind its ESG efforts are, Goss and Roberts (2011) illustrate possible consequences of such behavior. The authors argue that the ESG activities of companies, who react with new ESG efforts in direct response to negative media and investor sentiment, are often perceived as engaging in window-dressing behavior, which will lower the company's perceived creditworthiness and increase their cost of capital.

However, other empirical studies give reason to doubt an entirely pessimistic view on a firm's ESG activities, providing evidence for some (in-)direct financial benefits for companies from being proactive in ESG matters. For example, using yearly sustainability ratings, Lins *et al.* (2017) show that ESG activities can enhance stakeholder trust, which can then be drawn upon in times of economic distress, such as during the financial crisis between 2008 and 2009. Nofsinger and Varma (2014) make a similar argument, elaborating further on the financial performance of socially responsible investment funds under different market conditions. Similarly, Cahan *et al.* (2015) find that companies that show a high level of social responsibility tend to receive a more positive overall news image. The authors argue that such a positive media image helps to build a better reputation, increases investor trust, and might enable the company to reap economic benefits from the increased positive public awareness.

From this brief overview of the ESG literature, it becomes obvious that the link between ESG information and stock market reaction and, thus, investors' reactions to a firm's efforts in the domain of ESG has not been well established yet. Assuming that a firm has no responsibilities other than creating legal profits, ESG activities should be perceived as unnecessary expenses that destroy firm value. Thus, investors holding this view should punish any kind of information on new ESG activities. If investors acknowledge certain positive externalities of ESG-related efforts, they would nevertheless engage in increased selling of their shares when negative ESG news arrive at the market or when newly undertaken ESG efforts are perceived as a mere tool to "greenwash" a firm's public image. Positive ESG-related news, on the other hand, could be seen as an intangible asset that builds investor trust and enhances the reputation or public image of a company. Thus, there appears to be a certain incentive for a company to position itself as a sustainable, socially responsible enterprise, if investors appreciate such ESG efforts and find value in them even though they are not part of the core business. In light of the recent spike of popularity of the topic and financial resources that are devoted to the domain of ESG, with ESG-focused assets under management expanding by almost 20% per year, it appears to be of utmost importance – for both companies and investors alike – to gain a better understanding of the way investors process ESG-related information (Reid *et al.*, 2018).

In contrast to previous ESG-related research, we set out to investigate the relationship between ESG information and investors' reaction by relying on a time series approach: We gather a large data set of news articles with ESG-relevant content from which we extract the articles' sentiment. This sentiment index is then used as an input to AutoRegressive Distributed Lag (ARDL) models to explain stock market returns for the constituents of the DJIA. With this approach we attempt to make several contributions to the ongoing discussion, namely (i) bridging the gap between the findings of several event studies (e.g., Aktas *et al.*, 2011; Naughton *et al.*, 2014; Capelle-Blancard and Petit, 2019) and studies that use low-frequency measures of ESG performance (e.g., Lins *et al.*, 2017; Goss and Roberts, 2011; Cahan *et al.*, 2015) by building an analysis on daily ESG-related information observed over the course of multiple years, (ii) extracting this information on a company's performance in the domains of ESG by official news releases from a third party (newspaper and online media), and (iii) evaluating the ESG information using a domain-specific sentiment approach.

Other studies, for the most part, only briefly touch upon the role of sentiment in identifying the effect of ESG activities on corporate performance (e.g., Cheong *et al.*, 2017; Goss and Roberts, 2011; Naughton *et al.*, 2014). Linking the information contained in a newspaper article on a certain ESG matter, assessed via a sentiment index, with investors' reaction to these information could, however, be a crucial component to further our understanding of the connection between the two. By placing ESG-related sentiment at the center of attention, we build upon the findings of several authors of the behavioral finance literature who identify a significant relationship between sentiment and stock price movements (e.g., Tetlock, 2007; Garcia, 2013; Li *et al.*, 2014). In light of the findings of this strand of the literature, the question arises whether changes in ESG-specific sentiment, i.e., sentiment expressed toward ESG-related activities of companies, also have an impact on a stock's financial performance.

For our approach, we use the idiosyncratic component of daily stock returns as an indicator for the financial performance of a stock. These idiosyncratic returns are estimated as the residual term of an OLS regression of the DJIA constituents' log-returns on the log-returns of the S&P 500 index, resembling the idea of a market model (see, for example, Stapleton and Subrahmanyam, 1983). Using ESG-related news sentiment, constructed by a domain-specific dictionary approach following Loughran and McDonald (2011) and Myšková and Hájek (2018), we find significant effects of both temporary and permanent changes in sentiment on the idiosyncratic returns for the vast majority of the DJIA constituents. We can further identify different groups of stocks according to their ESG - stock-return relationship: The ESG-affine group can be characterized by investors that tend to react with increased buying activity to an increase in positive ESG-related information, whereas the ESG-averse group appears to have investors that rather engage in increased selling of their shares when (positive) ESG-information arrive. Interestingly, the investors of the majority of stocks appear to have no predisposition with respect to ESG activities of the stocks they are invested in. Nevertheless, investors of these stocks tend to be rather pessimistic about positive ESG information, irrespectively of the financial performance of the stocks. Furthermore, our results indicate an inverse relationship between the financial performance and the extent to which investors appreciate ESG activities: Those stocks whose investors predominantly punish increased ESG-related sentiment by withdrawing capital show the highest median log-returns over the sampling period, while those stocks whose investors appear to appreciate ESG-related news financially are among the worst performing stocks of our sample. These findings highlight the ambiguous nature of firms' ESG activities, whose affects appear to be highly context-specific. Especially if companies are aiming at using ESG as a strategic element and signal to investors, our findings plead for a thorough coordination of such activities. In this sense, our results can also mediate between the conflicting findings of the related literature, since they find support for both sides of the argument and offer first insights as to potential mechanisms leading to these stock-specific ESG reactions that may otherwise appear mutually exclusive.

The remainder of this paper is structured as follows. First, Section 3.2 places our work into the context of the relevant literature. Then, Section 3.3 discusses our data set. Section 3.4 continues by presenting our sentiment approach as well as the ARDL models that we use in the analysis of news articles and financial data, before Section 3.5 summarizes our main findings. Lastly, Section 3.6 concludes.

3.2 Related work

3.2.1 A short overview of ESG

The first official mentioning of the term “ESG” is attributed to the study “Who Cares Wins” by the UN Global Compact in 2004 (Compact, 2004). The subsequent “Who Cares Wins Initiative” in the years 2004-2008 was built upon the idea of creating a triple-win situation for the financial industry, society, and the environment.[1] Its main purpose was to sensitize the financial industry for sustainability topics, such that ESG issues will be more and more incorporated into investment decision-making. Ten years later, there are several indications that the initiative is bearing fruits: Estimates on ESG investing in assets under management in 2018 amount to about $20 trillion, which is roughly a quarter of all professionally managed assets worldwide (Kell, 2018). Following the call from the UN Global Compact, the New York Stock Exchange launched the Principles of Responsible Investment (PRI) and the Sustainable Stock Exchange Initiative (SSEI) in 2006 and 2007, respectively. Both of these initiatives provide assistance and guidance but also control in ESG matters to listed companies and professionally managed assets. Since their launch, the financial world has seen a dramatic increase in portfolio managers offering sustainable investment strategies (Lofts, 2018). As an example, the latest PRI annual report[2] shows that many key targets on their agenda have been exceeded in 2018, with 87% (target 80%) and 57% (target 50%) of signatories reporting to consider ESG factors in directly managed assets and portfolio construction, respectively. These developments show that there is a strong concern for the advancement of sustainable business practices from both politics and the financial world. It is plausible to assume a broad set of driving forces behind these concerns – ranging from environmental concerns, political agendas, and the attempt to attract and bind investors long-term on side of the firm to utility gains on side of the investors when investing according to their ethical values and beliefs (Nofsinger and Varma, 2014). Hence, it appears to of interest to all parties involved to unravel the effects on investment decisions emerging from a company’s performance in the domain of ESG.

In face of the recent boost in funding of and attention devoted to ESG-related financial programs, it is surprising that findings about the way investors process ESG activities remain sparse. As mentioned above, the recent literature acknowledges a certain response by investors to companies’ ESG initiatives, which then affects the firm’s stock market performance (Kim *et al.*, 2014; Cahan *et al.*, 2015; Lins *et al.*, 2017; El Ghoul *et al.*, 2011). All of these studies use yearly observations, explaining their findings by concepts evolving around trust-building sustainability efforts, which enable the company to reap financial benefits such as lower costs of capital, favorable media coverage, and an increased overall profitability. These insights prescribe a certain long-term view to investors with respect to the topic of ESG. Such a view seems to be relevant when considering large institutional investors or long-term investment strategies in diversified portfolios. It is plausible to assume that such investment strategies are

[1]Though in the context of this paper we exclusively use the term ESG, it is important to note that this also includes all actions that might in other sources be attributed to Corporate Social Responsibility (CSR). We consider ESG an umbrella term that collects not only socially responsible actions but also covers the aspects of environment and governance – concepts that are not explicitly included in the aforementioned concept of CSR.

[2]The report can be accessed via `https://www.unpri.org/annual-report-2018`, last accessed Aug 19, 2019.

not based on day-to-day news but will rather consider yearly sustainability ratings of companies when incorporating ESG factors into their investment decisions. Nevertheless, this approach ignores both the changing landscape of constantly monitored businesses, in which news spread rapidly over the internet, and the potentially impulsive behavior of the large group of retail investors. These retail investors or noise traders, as described by Kyle (1985) or Black (1986), are assumed to trade irrationally and not based on fundamentals. Thus, it is plausible that this group of investors are more prone to being influenced by sentiment expressed towards a certain company by, for example, public media. The particularly intangible nature of ESG information makes them furthermore more vulnerable to "greenwashing" attempts when giving a considerable weight to company reports, which can be assumed to attempt to paint a more promising picture of the firm's performance in ESG matters than objectively is true. Fortunately, the speed with which more objectively presented information arrive and can potentially be processed by investors has increased dramatically over the past decade, accompanied by increasingly easy access to these information. Exploiting the new possibilities to look beyond year-on-year data could provide some additional insights as to if and if so how investors incorporate ESG-related information into their investment decisions. Not only would this approach account for the increasingly monitored business world and potential noise trading, such a perspective would also allow us to directly relate changes in a firm's ESG performance to stock market movements, instead of relying on a year-on-year average of company performance.

Increasing the frequency of observations can be achieved by relying on public press-releases such as news articles and newswires that report on ESG-issues timely. While also newspapers oftentimes fail to report in a completely objective manner, they should pretty well capture and likewise form public opinion on the reported topic. If we furthermore are willing to assume a certain degree of due diligence present in the research process for an article, we can plausibly argue that news on ESG issues constitute a reasonable compromise between objective ESG reports and fast arriving information. Research conducted by Krüger (2015) and Capelle-Blancard and Petit (2019) resulted in two of a very few studies that do take a similar approach and focus on the effects of ESG news. Both paper conduct an event study and find that investors on average respond negatively to negative news and weakly negatively or insignificantly to positive events. Similar to the argument made by Friedman (2007), Krüger (2015) explains his findings as an agency problem inside the firm: While managers earn a good reputation for socially responsible practices among stakeholders, news on increased spending on such practices are generally bad news for shareholders. However, he also provides another view on this issue, called "doing well by doing good", according to which environmentally and socially sustainable business practices could also create value for shareholders. Even though he does not find support for this latter view in his study, his agency theory and "doing well by doing good" summarize the two potential theories behind the opposing views present in the literature reasonably well. The question we now want to address is: Can we make sense of the ambiguous findings of the literature or mediate between both theories by taking a more holistic time series approach to assess the ESG-investor relationship?

3.2.2 Relevant sentiment literature

Our approach evolves around the increasingly popular body of literature that links sentiment, conveyed by news articles, with the stock market. Research in this area has evolved rapidly over the last two decades. In one of the first comprehensive studies, Tetlock (2007) establishes predictive power of negative words in news articles for downward pressure on stock market prices. While the author uses the General Inquirer's Harvard-IV-4 classification dictionary to identify words with a negative connotation, Loughran and McDonald (2011) advocate using an application-specific dictionary instead of a general English language dictionary when one wants to accurately capture the sentiment conveyed by a text that evolves around a certain topic, for example finance. Following their findings, Garcia (2013) approximates investor sentiment by positive and negative words of two columns of financial news from the *New York Times*, as identified by the Loughran & McDonald dictionary. In his autoregressive distributed lag model of log-returns, he finds that positive words in the financial news help to predict stock returns. Similarly, Li *et al.* (2014) show that sentiment analysis, using both the Harvard and the Loughran & McDonald dictionary, can improve the prediction of price movements at the stock market. We now want to use these promising approaches and transfer them to the ESG domain. Focusing on news articles, published either in print or electronically, that contain information on the areas of environmental, social or governance practices of a firm, we evaluate the sentiment that each of these articles conveys. Thereby, we can differentiate between positive and negative ESG-related information and the respective intensity of sentiment expressed by each piece of news towards an ESG-related subject. In order to assess investors' responses to such information, referring to well-established models from the aforementioned finance literature appears to be well-suited for our endeavor.

3.3 Data

For the subsequent analysis, we collect news articles for the 30 constituents of the DJIA for the time period between January 2010 and December 2018 from the LexisNexis database.[3] Due to limited data availability, the stocks of "The Travelers Companies, Inc." (trv) and "DuPont de Nemours Inc." (dd) are dropped from the sample, such that we are left with data on 28 stocks. A commonly used alternative to news articles in the sentiment literature as basis for the analysis is social media data. As briefly explained above, the reason for consulting news articles instead of social media data is related to the focus of our study on ESG-related information. We try to stay as closely as possible to reports on ESG activities that evaluate a company's performance from a rather objective point of view without adding additional noise to the signal by considering subsequent discussions on these activities in social media. We intentionally do not consider press-releases on ESG activities by the company itself, since there appears to exist an undeniable incentive for a company to exaggerate its sustainability or ESG achievements. Furthermore, such reports are usually only issued annually, which renders an investigation of how investors process the ESG information timely impractical. Relying on publicly available news, we are also able to increase the frequency of our data from annual to daily observations.

For our collection of relevant news articles, we narrow the search by relying on 13 ESG-related search terms in the *SUBJECT*-line of the LexisNexis search for each of the DJIA constituents. We identify these 13 subject terms by comparing the most frequently occurring subjects of articles connected to environmental, social, and governance activities of companies in LexisNexis.[4] Panel A of Table 3.7.1 of the appendix lists those subjects that appear to be included in the vast majority of ESG-related news

[3]Our collection of news articles is based on search results obtained from *Lexis Uni*, accessible via `https://advance.lexis.com/`, last accessed Aug 19, 2019.

[4]Enlarging our list of search terms leads to larger samples of articles collected, yet, after filtering the results as explained in this section, only a few more articles are recorded than for the 13 search terms in Panel A of Table 3.7.1 and the results are not affected.

articles. Furthermore, we exclude non-business news and duplicated, i.e., very similar, articles from our search. In order to avoid that an article is attributed to a certain company simply because it is tagged with that company name, while the article is mainly about another company, we rely on the LexisNexis relevance score (in percent) and attribute an article to a certain company only when it shows a relevance score of more than 50% for that respective company. Analogously, we require a relevance score above 50% for at least one of the ESG subjects.[5]

We obtain daily observations of stock prices for the constituents of the Dow Jones Industrial Average as well as closing prices for the S&P 500 index between January 1, 2010, and December 31, 2018, from Yahoo Finance. Table 3.1 provides an overview of all stocks with their respective ticker symbol that we use in the subsequent analysis, together with a summary of further stock-specific information on the stock's industry sector, market capitalization, ownership structure, CSR rating, and relative frequency of news articles in our sample. More information on the respective data sources are provided in Table 3.1.

Daily closing prices are used to calculate log-returns as $r_t = (log(p_t) - log(p_{t-1})) \cdot 100$ both for each stock i (r_t^i) as well as for the S&P 500 index (r_t^{sp}). Next, we estimate the idiosyncratic component of the individual stock returns by entertaining a simple market model, in the spirit of Stapleton and Subrahmanyam (1983), in which the log-returns of stock i are regressed on a constant and the S&P 500 index log-returns of that respective day by means of OLS, as shown in Equation (3.1).

$$r_t^i = \lambda + \beta \cdot r_t^{sp} + \nu_t, \tag{3.1}$$

where $E(\nu_t) = 0$, $E(\nu_t r_t^{sp}) = 0$. The residual series of this model, ν_t, then represents the part of the log-returns of stock i that cannot be explained by the market return, approximated by the S&P 500 log-returns, and is hence labeled the idiosyncratic return (ir_t) of stock i.

For each stock we observe a total of 2,263 trading days over the nine years that our sampling period encompasses. To each trading day we then match the ESG-sentiment time series, which we extract from our news data sets as illustrated in the following section.

[5] Since we explicitly search for both company and ESG-related subjects, the LexisNexis search already returns relatively high relevance scores on both of these dimensions. Thus, our 50% threshold filter only excludes a few articles.

Table 3.1: Summary DJIA stocks

This table summarizes information on the 28 DJIA constituents that are used in our analysis. The column *GCIS sector* provides information on the industry sector to which a respective stock belongs, according to the MSCI Global Industry Classification Standard, accessible via `https://www.msci.com/gics`, last accessed Aug 19, 2019. Average market capitalization (*M-cap*) of a stock, expressed in Billion USD, and *Ownership*, the average fraction of institutional ownership of a certain stock over the sampling period, are both gathered via Thomson Reuters Eikon. CSR ratings (*CSR*) are provided by CSR Hub's publicly available company overview, available via `https://www.csrhub.com/csrhub/?_ga=2.25363325.1901910551.1561297416-342900800.1561297416`, last accessed Aug 19, 2019. More information on the rating methodology can be accessed via `https://esg.csrhub.com/csrhub-ratings-methodology`, last accessed Aug 19, 2019. Lastly, *Rel news* is a relative measure of how many news articles we observe for a specific stock, where Coca-Cola, as the stock with most ESG-related articles published between 2010 and 2018, receives the value 1.

Name	Ticker	GICS sector	M-cap	Ownership	CSR	Rel news
Apple Inc	aapl	Information Technology	896.336	0.6393	59	0.8071
American Express Co	axp	Financials	102.907	0.8375	57	0.1182
Boeing Co	ba	Industrials	196.544	0.7400	58	0.5918
Caterpillar Inc	cat	Industrials	72.788	0.6774	54	0.2290
Cisco Systems Inc	csco	Information Technology	244.472	0.7419	65	0.5096
Chevron Corp	cvx	Energy	230.818	0.6435	52	0.3908
Walt Disney Co	dis	Communication Services	243.103	0.7140	56	0.4412
General Electrics	ge	Industrials	89.781	0.5514	58	0.6036
Goldman Sachs Group Inc	gs	Financials	71.240	0.7858	54	0.8114
Home Depot Inc	hd	Consumer Discretionary	217.868	0.7267	57	0.0230
International Business Machines Corp	ibm	Information Technology	120.539	0.5964	61	0.0460
Intel Corp	intc	Information Technology	209.747	0.6563	64	0.5681
Johnson & Johnson	jnj	Health Care	371.124	0.6656	64	0.6615
JPMorgan Chase & Co	jpm	Financials	358.991	0.7529	58	0.7598
Coca-Cola Co	ko	Consumer Staples	218.980	0.6650	58	1.0000
McDonald's Corp	mcd	Consumer Discretionary	155.200	0.6876	59	0.6416
3M Co	mmm	Industrials	97.364	0.6904	62	0.3049
Merck & Co Inc	mrk	Health Care	213.644	0.7513	62	0.3715
Microsoft Corp	msft	Information Technology	1,012.000	0.7667	66	0.8849
Nike Inc	nke	Consumer Discretionary	130.864	0.8507	60	0.2427
Pfizer Inc	pfe	Health Care	236.895	0.7233	56	0.4101
Procter & Gamble Co	pg	Consumer Staples	274.361	0.6024	59	0.3360
UnitedHealth Group Inc	unh	Health Care	233.661	0.8931	53	0.1406
United Technologies Corp	utx	Industrials	106.086	0.8305	57	0.3516
Visa Inc	v	Information Technology	372.194	0.8977	59	0.2638
Verizon Communications Inc	vz	Communication Services	235.032	0.6040	56	0.5744
Walmart Inc	wmt	Consumer Staples	308.138	0.8125	57	0.9334
Exxon Mobil Corp	xom	Energy	316.655	0.5111	54	0.0778

3.4 Sentiment and time series approach

3.4.1 Constructing the sentiment index

One crucial element of our approach to assess how investors process ESG information for their investment decisions is the evaluation of the sentiment conveyed by each of the news articles. As elaborated above, sentiment extracted from news articles should provide us with the best proxy of a company's ESG achievements or shortcomings we can obtain that is neither overly biased towards the firm's point of view nor heavily influenced by social media controversies. For computing the sentiment that each individual news article conveys, we rely on a dictionary approach, which utilizes pre-defined lists of words that convey positive and negative meaning, respectively, to identify the sentiment expressed by a text. A commonly used alternative would be to rely on machine learning algorithms. Computing sentiment via a machine learning algorithm has the advantage that one does not need to construct positive and negative word lists but would instead manually classify a subsample of the data according to whether a certain text conveys positive and negative sentiment. The algorithm then "learns" from this training data set the most likely classification of new articles by looking for patterns that are unique to either positive or negative sentiment in the training data. This advantage can, however, also be seen as one drawback of machine learning approaches, since the quality of the sentiment index is heavily dependent on whether or not the training data represent a broad range of features, i.e., words and linguistic structures that imply strong positive or negative sentiment (Siering, 2012). Furthermore, Hutto and Gilbert (2014) point out that many machine learning algorithms, such as neural networks used to construct a sentiment index, derive incomprehensible solutions. To avoid these caveats, we instead choose a parsimonious dictionary approach. More precisely, we use the dictionary developed by Loughran and McDonald (2011), in the following referred to as *LM*, which the authors have specifically tailored to the context of financial texts. Their dictionary is shown to outperform conventionally used English language dictionaries, such as the General Inquirer's Harvard-IV-4 classification dictionary, when applied to finance-related texts.

For our sentiment index, we construct sentiment polarity for each article k from the occurrences of positive and negative words in the LM dictionary as

$$\text{polarity}_k = \frac{\text{positive}_k - \text{negative}_k}{\text{positive}_k + \text{negative}_k}, \tag{3.2}$$

where positive_k and negative_k stand for the sum of positive and negative LM dictionary words, respectively, that appear in article k. Calculating the sentiment index in this way leads to a variable that is defined on the interval $[-1, 1]$. We additionally conduct our analysis using the Valence Aware Dictionary for sEntiment Reasoning (VADER) by Hutto and Gilbert (2014), augmented with the LM dictionary (VADER-LM). The difference to the polarity measure in Equation (3.2) lies in the emphasis on the intensity of the words, i.e., VADER-LM does not only classify words into positive and negative connotations, but it also assesses how strongly positive or negative a certain expression appears to be. Since the results of both approaches are fairly similar with respect to the characteristics observed for the different groups of stocks that we identify below, we focus in the following on the LM approach as the more parsimonious sentiment index. Detailed estimation results for the VADER-LM approach can be found in Tables 3.7.2 and 3.7.3 of the appendix.

While the dictionary by Loughran and McDonald (2011) seems appropriate to capture the sentiment of news content particularly relevant to investors, there is a shortcoming when applying the dictionary to our data set of ESG-related news. If an article combines positive finance-related and negative ESG-specific news, our sentiment index might be biased in favor of the positive financial news at the expense of adequately representing the ESG-related information. For example, one article in 2010 on the company Apple reports on increased revenues of the company while their philanthropic activities have declined. To avoid this pitfall, we follow Nasukawa and Yi (2003) and conduct a domain-oriented sentiment analysis: We rely on the word list by Myšková and Hájek (2018), which comprises words related to sustainability topics, based on sustainable development glossaries provided by the United Nations and the Environmental Protection Agency, amongst others, and evaluate our ESG news articles only around sentences that contain one of the words that are on the list.[6] We then additionally control for the remaining content of an article, which potentially contains other general interest or finance-related information – in the previous example positive news on the company's financial performance – such that these two effects do not become intertwined.

In this way we obtain one ESG-related and one non-ESG-related sentiment index as control variable for each article in our data set. Next, we match each article's sentiment score to a trading day of that respective stock. Since the constituents of the DJIA are traded between 9:30 ET and 16:00 ET, any article published between 0:00 ET and 16:00 ET on a given trading day t is attributed to this trading day.[7] Articles published after 16:00 ET are then attributed to the next trading day. Non-trading days, such as national (United States) holidays and weekends, are treated similarly in that any article published on a non-trading day is attributed to the next trading day. Trading days on which we do not record any news on that specific stock receive a sentiment value of zero. An alternative approach is taken by Behrendt and Schmidt (2018), who use the last observed sentiment value for any missing observation in an intraday context. However, in our application to daily observations this would induce an undesirable persistence of the sentiment time series, since there might be multiple days in a row without new articles released for a company. In such cases, assuming a perpetuating effect of the last observed sentiment value in ESG-related news over multiple days does not appear to be reasonable. Rather, we would expect, and often observe, multiple articles published on the same ESG-related issue over subsequent days as an ESG-story unfolds further, which is then reliably captured by our ESG-sentiment index. In case of multiple news articles attributed to the same trading day, we calculate a weighted average of the individual sentiment scores, loosely following Fang and Peress (2009). More specifically, we give a higher weight to media sources that are entirely focused on finance-related stories, such as *Financial Times* and *Bloomberg*, which we expect to be more influential and relevant to investors than other general interest news sources.[8]

3.4.2 Autoregressive distributed lag models

The sentiment index, constructed as described above, then serves as one input for our ARDL models, with the idiosyncratic returns (ir_t) as the dependent variable. The ARDL model is chosen since it is one of the most widely used, parsimonious models to describe a dependent time series variable as a function

[6]In addition, we test the robustness of our results when assessing only the sentiment of sentences that contained at least one of the words in the multidimensional dictionary for CSR compiled by Pencle and Mălăescu (2016). However, qualitatively all findings remain unchanged.

[7]Assuming that new ESG-related information arriving during the last minutes of a trading day are not reflected in a stock's closing price, we also test how our results change when any article published after 15:30 ET receives the date $t+1$ and is, thus, attributed to the next trading day. Qualitatively, the results are robust to this alternative approach.

[8]For an exhaustive list of all financial media sources we consider, please refer to Panel B of Table 3.7.1 of the appendix.

of its own lagged values as well as contemporaneous and lagged values of several independent variables.

We find support in this approach by, for example, Garcia (2013) who also uses an ARDL model to explain log-returns by, amongst others, a sentiment index. Inspired by the author's approach to distinguish between both positive and negative sentiment as well as the effect of sentiment in the presence or absence of economic recessions, we entertain several versions of the ARDL model. The lag length of both independent and dependent variables is set equal to five. While for some stocks of the sample the BIC information criterion suggests slightly different lag lengths, it appears to be reasonable to focus on the effect of ESG news that occur within one week of that information arriving at the market. Longer horizons would always increase the likelihood of the effect being diluted by other information arriving later. Similarly, weekly or monthly measures of ESG-related news sentiment also appear to weaken the signal contained in the articles, which is why we refrain from entertaining models with such aggregated measures of sentiment. In contrast to Garcia (2013), we do not include lags of the squared dependent variable in the models. While including this measure of the returns' variance as it is proposed in GARCH-in-mean type models (Engle *et al.*, 1987) does not alter our estimation results qualitatively, it would complicate the computation of the long-run multiplier considerably, which is explained in detail below. Thus, despite the potentially time-varying volatility of our idiosyncratic return series we do not find that accounting for this fact adds value in the context of our analysis. Furthermore, all of the following models have two control variables in common, namely day-of-the-week dummy variables and the contemporaneous effect of non-ESG related information contained in an article.[9] OLS with heteroskedasticity- and autocorrelation-consistent standard errors following White (1980) is used for the estimation of the models and the reporting of the results.

The first, "naive" ESG model uses contemporaneous and lagged effects of our ESG-sentiment index to explain daily idiosyncratic returns, and reads

Model (1)

$$ir_t = \alpha + \beta_0' \cdot \mathscr{L}_s(S_t) + \gamma_0' \cdot \mathscr{L}_v(ir_t) + \delta' \cdot \boldsymbol{x}_t + \varepsilon_t, \tag{3.3}$$

where $\mathscr{L}_s$ is a lag operator with $s = 0, 1, \ldots, 5$, while $\mathscr{L}_v$ is a lag operator with $v = 1, 2, \ldots, 5$. S_t is the sentiment index at time t, normalized to have zero mean and unit variance, and $\boldsymbol{x}_t$ are the previously mentioned control variables, namely day-of-the-week dummy variables and non-ESG-related sentiment. ε_t is a zero-mean residual. Thus, β_0' is a vector collecting the six coefficients $\beta_{0,s}$ that belong to the respective lag s of S_t, where the first index of 0 is used to differentiate the sentiment coefficients in Models (1) and (2) from the ones used in Model (3), as will become important in Equation (3.6) below. Analogously, γ_0' is a vector of coefficients for lagged idiosyncratic returns. We call this our "naive" model, since it utilizes the plain ESG news sentiment variable S_t as an input without making any further distinction between positive or negative news or the market conditions during which the news arrive. This model should therefore give us a simple, first impression as to the effect of an increase in ESG-related news sentiment on idiosyncratic returns.

However, since several authors reveal striking differences between the effect of positive and negative sentiment in similar frameworks (e.g., Garcia, 2013; Capelle-Blancard and Petit, 2019; Krüger, 2015; Balahur *et al.*, 2010), our second model makes such a distinction and reads

[9]Entertaining the models with month-of-the-year and year dummy variables does not change the results substantially. Thus, we only report the results of the most parsimonious model, while the results of all robustness checks are available upon request.

Model (2)

$$ir_t = \alpha + \beta_0'^{+} \cdot \mathscr{L}_s(S_t^+) + \beta_0'^{-} \cdot \mathscr{L}_s(S_t^-) + \gamma_0' \cdot \mathscr{L}_v(ir_t) + \delta' \cdot \boldsymbol{x}_t + \varepsilon_t, \tag{3.4}$$

where S_t^+ and S_t^- stand for positive and negative ESG news sentiment, respectively. For the construction of S_t^+ (S_t^-), we interact the standardized S_t variable with an indicator variable that equals one if $S_t > 0$ ($S_t < 0$), and is zero otherwise. For better interpretability of the estimates in $\beta_0'^{-}$, we then take the absolute value of S_t^-, such that an increase in S_t^- indicates an increase in negative sentiment. In order to facilitate the distinction between the β-coefficients of positive and negative sentiment, we annotate the β-coefficients of Equation (3.4) with the superscripts $+$ and $-$ for the effects of positive and negative ESG-related sentiment, respectively.

Next, we further modify the approach taken by Garcia (2013), who investigates differences in the effects of sentiment on financial data between economic recessions and the absence of recessions. While investigating the influence of overall economic recessions would require many more years of news articles, which are more and more scarcely available the further one goes back in time, we do find it plausible that the current financial performance of a stock influences the degree to which investors pay attention to and process ESG-related information. Furthermore, we observe a boom in "sustainable finance" only in recent years, which also speaks against including more years of observations. In this context, one hypothesis would be that if a firm's stock is currently yielding negative idiosyncratic returns, investors might perceive reports on ESG efforts undertaken by this firm as window-dressing – an attempt to distract investors from or attract investors despite of a current lack of performance of the company's core business. On the other hand, positive ESG-related news could also keep investors from selling their shares during falling stock prices, following the argument that proactive engagement in ESG activities enhances investor trust. In order to investigate potentially varying effects of ESG information on stock prices given different states of a stock's current financial performance, we specify a third ARDL model that accounts both for negative idiosyncratic returns as well as for positive and negative ESG-related news sentiment and reads

Model (3)

$$\begin{aligned} ir_t \quad = \quad & \alpha + D_t \left[\beta_1'^{+} \cdot \mathscr{L}_s(S_t^+) + \beta_1'^{-} \cdot \mathscr{L}_s(S_t^-) + \gamma_1' \cdot \mathscr{L}_v(ir_t)\right] \\ & + (1 - D_t) \left[\beta_2'^{+} \cdot \mathscr{L}_s(S_t^+) + \beta_2'^{-} \cdot \mathscr{L}_s(S_t^-) + \gamma_2' \cdot \mathscr{L}_v(ir_t)\right] \\ & + \delta' \cdot \boldsymbol{x}_t + \varepsilon_t, \end{aligned} \tag{3.5}$$

where D_t is a dummy variable that equals 1 if a stock is currently underperforming, i.e., if its idiosyncratic returns in period t are negative. $\beta_1'^{+}$ ($\beta_1'^{-}$) then collects the coefficients for the interaction term between an underperforming stock and positive (negative) sentiment, while $\beta_2'^{+}$ ($\beta_2'^{-}$) are the coefficients for the interaction term between a stock that yields zero or positive idiosyncratic returns and positive (negative) ESG-related news sentiment. In addition to the previously used control variables, the $\boldsymbol{x}_t$ of Equation (3.5) also includes the dummy variable D_t in order to control for the direct effect of negative idiosyncratic returns on the dependent variable at time t.[10]

While potentially each of the lagged β-coefficients of a model holds valuable information with respect to how a stock reacts to a change in ESG-related news sentiment, we only want to distinguish between the effect of a temporary change in ESG-related sentiment on idiosyncratic returns, also called the impact multiplier, and the effect of a permanent shock to ESG-related sentiment, the long-run multiplier (LRM). The impact multiplier as we use this term in the context of our research is given by the contemporaneous effect at time t of the sentiment variable of interest of the respective model. The LRM, as derived in Brissimis (1976) and Bewley (1978), is calculated as

$$\theta_l^{(*)} = \frac{\sum_{s=0}^{5} \beta_{l,s}^{(*)}}{1 - \sum_{v=1}^{5} \gamma_{l,v}}, \quad \text{for } |\sum_{v=1}^{5} \gamma_{l,v}| < 1, \quad l = 0,1,2 \quad * = +,- \tag{3.6}$$

where $\beta_{l,s}^{(*)}$ and $\gamma_{l,v}$ represent the coefficient estimates of the independent variable of interest and the dependent variable, respectively. The superscript of β is noted in brackets, since only Models (2) and (3) distinguish between positive and negative sentiment. For more information on the LRM and its derivation, please refer to Section 3.7.1 of the appendix. Standard errors of each $\theta_l^{(*)}$ are calculated using the delta method, relying on the heteroskedasticity- and autocorrelation-consistent variance-covariance matrix of the ARDL coefficient estimates.

3.5 Results of the sentiment analysis

3.5.1 Sentiment and estimation results

Table 3.2 provides descriptive statistics for the sentiment index that we construct for each of the 28 stocks over the sampling period prior to standardization. We observe throughout positive mean values for the sentiment index. Considering that this variable is bounded, all but three stocks' sentiment appears to cover the entire range with minimum values of -1 and maximum values of 1. For illustrative purposes, we show plots of both the time series and ACFs of the sentiment series for two of the 28 stocks in the sample in Figure 3.1, namely Apple (*aapl*) and Visa (*v*).[11] Beginning with the time series plots, we observe that sentiment extracted from ESG-related news articles on the company Apple does not appear to be predominantly positive nor negative, while for Visa the plot shows fewer negative sentiment scores than positive ones. The ACFs of both stocks do not show clear patterns or significant lags that would suggest serially correlated series.

[10] While the interaction terms of the sentiment variables with D_t are of main interest in Model (3), additionally including D_t as control variable appears to be very important in order to account for the current performance of the stock. ARDL models that omit D_t as additional control show coefficient estimates of the sentiment parameters that simply reflect the current underperforming state (throughout significant, negative coefficient estimates) or not underperforming state (throughout significant, positive coefficient estimates) of the respective stock but are not informative about the effect of a change in sentiment under either of these conditions.

[11] The stocks of Apple and Visa are chosen since they represent all major features that the sentiment time series plots and ACFs of the entire sample of 28 stocks show, namely differing ratios of positive to negative ESG-related sentiment and non-significant ACFs. Plots for the remaining stocks of the sample are available upon request.

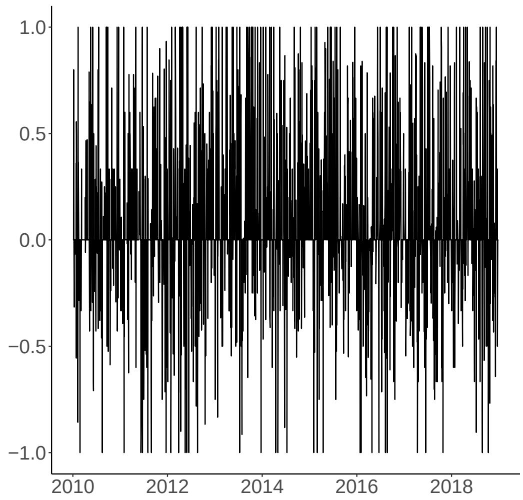
Table 3.2: Summary statistics ESG sentiment

This table provides an overview of the descriptive statistics for the LM sentiment index prior to standardization. Company references to the ticker symbols used in the column *Ticker* can be found in Table 3.1. *N* stands for the number of observations in our sample, while *Mean* and *Sd* provide values for the mean and standard deviation of each sentiment time series, respectively. *Min* and *Max* are minimum and maximum values of each series, respectively, and the 25%- and 75%-quantiles are given in the accordingly labeled remaining two columns.

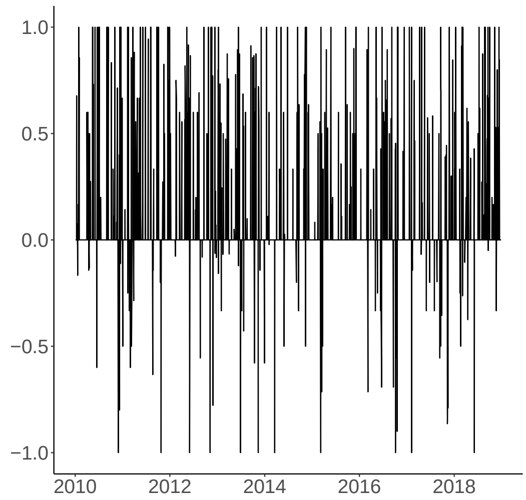
Ticker	N	Mean	Sd	Min	Q25	Q75	Max
aapl	2,263	0.0734	0.3695	-1.0000	0.0000	0.1735	1.0000
axp	2,263	0.0374	0.1891	-1.0000	0.0000	0.0000	1.0000
ba	2,263	0.1936	0.3695	-1.0000	0.0000	0.4543	1.0000
cat	2,263	0.0608	0.2303	-1.0000	0.0000	0.0000	1.0000
csco	2,263	0.2121	0.3517	-0.6667	0.0000	0.5000	1.0000
cvx	2,263	0.0544	0.3071	-1.0000	0.0000	0.0000	1.0000
dis	2,263	0.0866	0.3206	-1.0000	0.0000	0.0000	1.0000
ge	2,263	0.1908	0.3505	-1.0000	0.0000	0.4167	1.0000
gs	2,263	0.1415	0.4030	-1.0000	0.0000	0.3394	1.0000
hd	2,263	0.0054	0.0713	-0.6667	0.0000	0.0000	1.0000
ibm	2,263	0.0155	0.1212	-1.0000	0.0000	0.0000	1.0000
intc	2,263	0.0577	0.2082	-0.8333	0.0000	0.0000	1.0000
jnj	2,263	0.1877	0.3683	-1.0000	0.0000	0.4444	1.0000
jpm	2,263	0.1644	0.3861	-1.0000	0.0000	0.4023	1.0000
ko	2,263	0.3007	0.4076	-1.0000	0.0000	0.6364	1.0000
mcd	2,263	0.1095	0.4080	-1.0000	0.0000	0.1753	1.0000
mmm	2,263	0.1011	0.2895	-1.0000	0.0000	0.0000	1.0000
mrk	2,263	0.0969	0.2945	-1.0000	0.0000	0.0000	1.0000
msft	2,263	0.2236	0.4036	-1.0000	0.0000	0.5327	1.0000
nke	2,263	0.0468	0.2493	-1.0000	0.0000	0.0000	1.0000
pfe	2,263	0.0659	0.3041	-1.0000	0.0000	0.0000	1.0000
pg	2,263	0.1010	0.2959	-1.0000	0.0000	0.0000	1.0000
unh	2,263	0.0421	0.1835	-1.0000	0.0000	0.0000	1.0000
utx	2,263	0.1542	0.3267	-1.0000	0.0000	0.0000	1.0000
v	2,263	0.0662	0.2644	-1.0000	0.0000	0.0000	1.0000
vz	2,263	0.0964	0.3445	-1.0000	0.0000	0.1658	1.0000
wmt	2,263	0.1247	0.4186	-1.0000	0.0000	0.3835	1.0000
xom	2,263	0.0140	0.1392	-1.0000	0.0000	0.0000	1.0000

Figure 3.1: ESG sentiment time series and ACFs

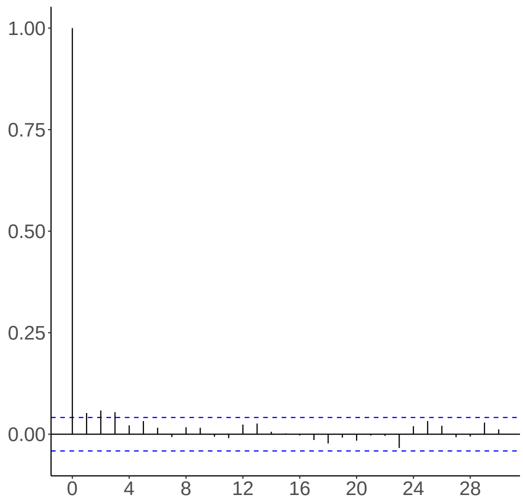
The four panels of this figure show the time series (TS) of ESG-related sentiment for Apple (aapl) and Visa (v) in plots (a) and (b), respectively. Plots (c) and (d) then illustrate the ACFs of the time series over 30 trading days. The dashed lines indicate 95% confidence bounds.



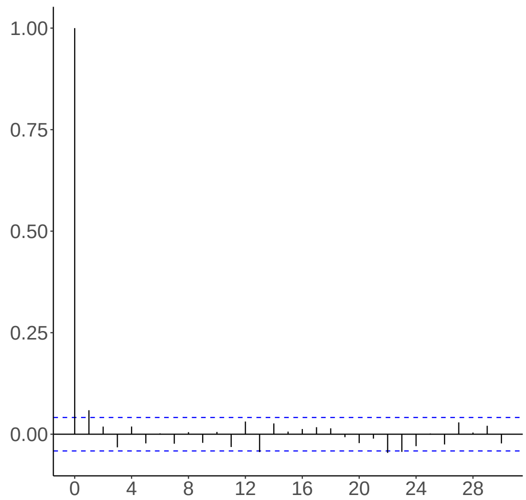
(a) TS: aapl



(b) TS: v



(c) ACF: aapl



(d) ACF: v

Analogously, Table 3.3 summarizes the characteristics of the idiosyncratic return series that serve as dependent variable in all of our ARDL models. It is important to note here as well as for subsequent comments on the size of the returns that the log-returns that were used for the estimation of the idiosyncratic returns are multiplied by the factor 100. By construction, we observe that the idiosyncratic return series for all 28 stocks are centered around zero. This feature is also clearly visible in the time series plots in Figure 3.2. Furthermore, one sees a few major peaks and valleys in both plots, i.e., spikes of high idiosyncratic returns in absolute value. When investigating the corresponding ESG-related sentiment, these spikes correlate with matching preceding or contemporaneous positive or negative sentiment values. To give one example, the highest idiosyncratic return for Apple, observed on April 24, 2014, coincides with a series of ESG-related news for Apple mainly evolving around the company's announcement to offer free recycling of all its devices, published between Monday, April 21, and Thursday, April 24.[12] Lastly, the ACF plots of both stocks do not reveal any kind of persistent patterns, which is supported by Augmented Dickey-Fuller (ADF) tests that show no reason to doubt the stationarity of all idiosyncratic returns time series.[13]

Table 3.3: Summary statistics idiosyncratic returns

This table provides an overview of the descriptive statistics for the idiosyncratic returns of each stock. Company references to the ticker symbols used in the column *Ticker* can be found in Table 3.1. *N* stands for the number of observations in our sample, while *Mean* and *Sd* provide values for the mean and standard deviation of the idiosyncratic returns, respectively. *Min* and *Max* are minimum and maximum values of each series, respectively, and the 25%- and 75%-quantiles are given in the accordingly labeled remaining two columns.

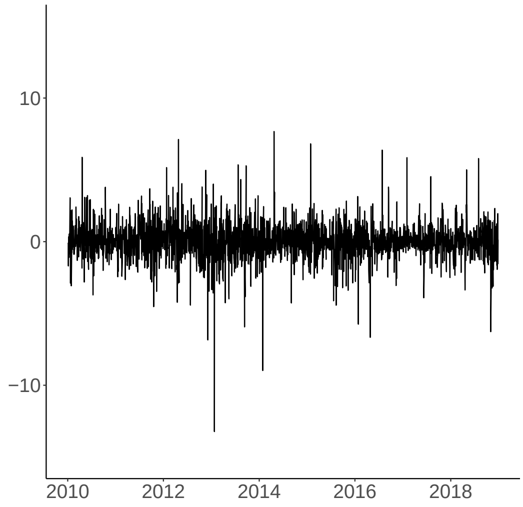
Ticker	N	Mean	Sd	Min	Q25	Q75	Max
aapl	2,263	0.0000	1.3195	-13.2267	-0.6501	0.6947	7.6713
axp	2,263	0.0000	1.0263	-15.1245	-0.4495	0.5096	8.7987
ba	2,263	0.0000	1.0981	-8.1444	-0.5929	0.5488	9.3516
cat	2,263	0.0000	1.1603	-7.1079	-0.6356	0.6116	6.7664
csco	2,263	0.0000	1.2332	-17.2174	-0.4643	0.4923	12.4507
cvx	2,263	0.0000	0.9429	-5.2926	-0.5111	0.4917	5.1598
dis	2,263	0.0000	0.9225	-9.9465	-0.4878	0.4708	7.7500
ge	2,263	0.0000	1.1645	-10.7603	-0.4693	0.4843	9.7814
gs	2,263	0.0000	1.1570	-11.5849	-0.6108	0.5831	5.2170
hd	2,263	0.0000	0.9245	-4.4652	-0.5089	0.4987	5.9795
ibm	2,263	0.0000	0.9472	-9.3446	-0.3870	0.4367	8.4673
intc	2,263	0.0000	1.1581	-8.2858	-0.5811	0.5746	8.7915
jnj	2,263	0.0000	0.7253	-9.3626	-0.3535	0.3656	3.9404
jpm	2,263	0.0000	1.0283	-9.2806	-0.5443	0.5245	5.9994
ko	2,263	0.0000	0.7471	-7.3232	-0.3895	0.4136	4.7213
mcd	2,263	0.0000	0.8331	-5.2004	-0.3989	0.4159	6.8556
mmm	2,263	0.0000	0.7411	-7.2405	-0.3513	0.3751	5.5859
mrk	2,263	0.0000	0.9829	-6.8579	-0.5173	0.4947	9.2433
msft	2,263	0.0000	1.0583	-12.2876	-0.5490	0.5099	9.6897
nke	2,263	0.0000	1.2244	-12.1845	-0.5636	0.5671	10.7111
pfe	2,263	0.0000	0.8804	-4.7786	-0.4724	0.4626	5.9461
pg	2,263	0.0000	0.7873	-6.7920	-0.3965	0.4006	8.4520
unh	2,263	0.0000	1.1295	-5.8859	-0.5502	0.5688	7.2215
utx	2,263	0.0000	0.7908	-6.8507	-0.3999	0.4145	3.9634
v	2,263	0.0000	1.1401	-14.2343	-0.5387	0.4895	13.0729
vz	2,263	0.0000	0.9248	-4.8525	-0.5033	0.5189	7.4511
wmt	2,263	0.0000	0.9900	-10.4371	-0.4660	0.4918	9.9096
xom	2,263	0.0000	0.8040	-4.4463	-0.4327	0.4359	4.6622

[12]There were a total of 14 articles published between April 21 and April 24, featuring headlines such as “Apple Cleans Up Its Act” (April 21, USNews.com), “Apple unveils free recycling of all its devices, vows to increase reliance on renewable energy” (TheRecord.com), “SOLID WASTE: Apple will offer free recycling at stores” (April 22, Greenwire), or “BUSINESS: Apple, Google race to reduce greenhouse gas footprints” (April 24, ClimateWire).

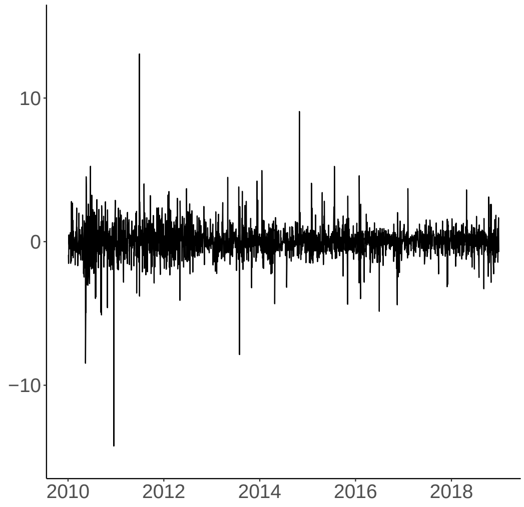
[13]Results of the ADF tests are available upon request.

Figure 3.2: Idiosyncratic returns time series and ACFs

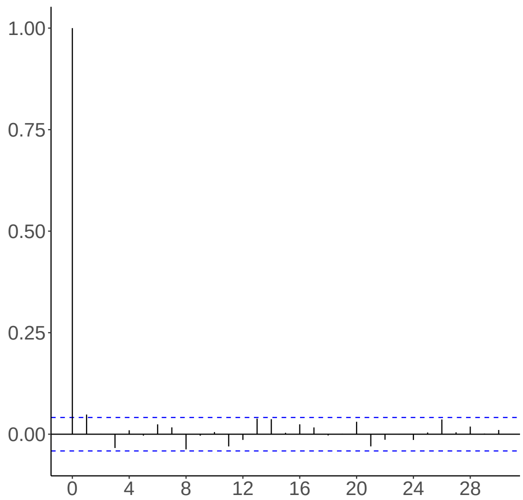
The four panels of this figure show the time series (TS) of idiosyncratic returns for Apple (aapl) and Visa (v) in plots (a) and (b), respectively. Plots (c) and (d) then illustrate the ACFs of the time series over 30 trading days. The dashed lines indicate 95% confidence bounds.



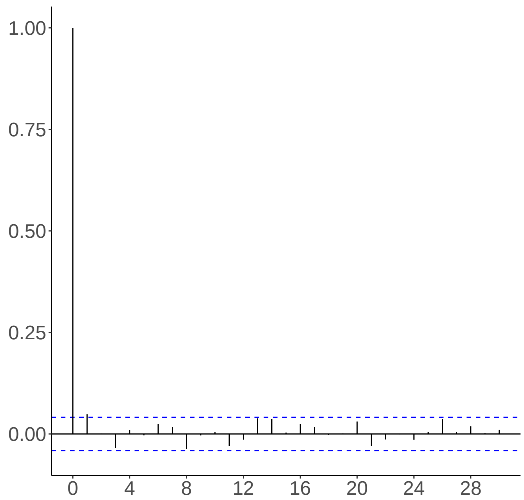
(a) TS: aapl



(b) TS: v



(c) ACF: aapl



(d) ACF: v

Turning to the estimation results of our models, Tables 3.4 and 3.5 summarize the impact and long-run multiplier, respectively, for the 28 DJIA constituents that remained in our sample.[14] It is important to note that only those multiplier are reported that are significant on an $\alpha \leq 0.1$. In order to facilitate interpretability of the relatively large tables, we divide each table into four panels according to the estimated multiplier of the respective stocks in the naive model: Panel A presents those stocks that show a negative sign of the respective multiplier in Model (1), while Panel B presents the stocks with a positive multiplier effect in Model (1). Panel C then collects those stocks that do not show a significant impact or long-run multiplier in Model (1) and Panel D, finally, those stocks that throughout all three ARDL

[14] We additional test the effect of a sample split that divides our time series roughly in two equal shorter samples. For the results of the ARDL models using these two subsamples, ranging from 2010 to 2014 and 2015 to 2018, respectively, please refer to Tables 3.7.4 to 3.7.7 of the appendix. The last paragraph of this sections briefly summarizes these results.

specifications do not show a significant impact or long-run multiplier. The reasoning behind the construction of these panels follows the idea that Model (1) represents a first-best guess with respect to the correlation between ESG-related sentiment and the respective stock's performance on the market. The remaining two models should then help to gain a better understanding of the effects that ESG news have on the returns of a stock and to identify further patterns or the absence of such. Since Models (2) and (3) contain multiple sentiment variables, β's with super- and subscripts at the head of each column of Table 3.4 indicate the respective coefficients whose impact multiplier effect, i.e., the contemporaneous coefficient estimate, a column summarizes. Analogously, Table 3.5 uses θ's with super- and subscripts at the head of each column to distinguish the LRM effects from the impact multipliers of Table 3.4.

Starting with the estimates of the impact multiplier, Panel A of Table 3.4 reveals that only for one of the DJIA constituents idiosyncratic returns tend to decrease on average when ESG-related news sentiment temporarily increases. Since an increase in ESG-related sentiment would indicate more positive ESG-related information arriving at the market, we could label such a stock an "ESG-averse" stock: Investors tend to punish ESG activities by withdrawing capital, even if those activities or new information on such were increasingly positive in nature. This hypothesis finds support in the estimate of the contemporaneous β_0^+-coefficient of Model (2), which shows a negative sign, indicating investor resentment towards positive ESG news. Though for the impact effects we only find one ESG-averse stock, these considerations will become more important when turning to the LRM effects below. The larger group of "ESG-affine" stocks, which show a positive contemporaneous reaction in idiosyncratic returns to an increase in ESG-sentiment, are collected in Panel B of Table 3.4. While, again, by construction, these five stocks show a throughout positive reaction in idiosyncratic returns to an increase in ESG-sentiment, we find that this response is sensitive to the nature of ESG news (positive vs. negative) but to weaken when further interacted with different states of performance of the stock (negative vs. zero or positive idiosyncratic returns): Idiosyncratic returns on average increase when positive ESG-related information arrive at the market, whereas they decrease or are not effected when negative ESG news arrive, as indicated by the positive and negative coefficient estimates for the β_0^+- and β_0^--column, respectively. These effects are only faintly visible in Model (3), in which the coefficient estimates of the interaction terms are mostly insignificant. Panel C of Table 3.4 then collects stocks for which our naive model does not show a tendency with respect to how investors value ESG activities, such that we could label them the "ESG-neutral" stocks. For 17 of these 21 stocks, which are more than half of all DJIA constituents, investors do not appear to react at all to positive ESG-related information. Interestingly, three out of four stocks that do react statistically significantly tend to show a decrease in idiosyncratic returns when positive ESG-information arrives at the market. This tendency is supported by Model (3), in which we find that irrespectively of the financial performance investors either do not react at all to a temporary increase in positive sentiment or they engage in increased selling of their shares. These findings suggest that investors of stocks that don't have a specific pre-defined opinion towards ESG-activities are particularly sceptical towards positive ESG-headlines contemporaneously.

Table 3.4: Impact multiplier of ESG sentiment on idiosyncratic returns

This table summarizes the estimated impact multiplier, i.e., the effects of temporary changes in ESG-related sentiment on idiosyncratic returns, in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative impact multiplier in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to temporary changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant impact multiplier in Model (1), while Panel D collects stocks that show no significant estimates for the impact multiplier throughout all models. The respective model is indicated by its equation number and respective sentiment coefficient. β_l^+ and β_l^- are the coefficients of positive $(+)$ and negative $(-)$ sentiment variables, respectively, where $l = 0$ represents the sentiment coefficient of Model (1) and $l = 1, 2$ represent the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	β_0	β_0^+	β_0^-	β_1^+	β_2^+	β_1^-	β_2^-
Panel A: negative impact multiplier Model (1)							
unh	-0.0403	-0.0351					
Panel B: positive impact multiplier Model (1)							
axp	0.0414	0.0640					
csco	0.0642		-0.2300				-0.1753
gs	0.0416	0.0633					
ko	0.0576	0.1034			0.0485		
xom	0.0430	0.0554			0.0302	0.0531	
Panel C: insignificant impact multiplier Model (1)							
aapl				-0.1113			0.0935
cat						0.0558	-0.0629
cvx			0.0636			0.0815	
dis			-0.0893	-0.0555		-0.0664	
ge		-0.0794		-0.1008			-0.0960
hd							-0.1330
ibm			-0.0299			0.0379	-0.0892
intc				-0.0518			
jnj			-0.0644				-0.0688
jpm			0.0890			0.0852	0.0783
mcd							0.0912
mmm						0.0491	
mrk			0.0704			0.0628	
msft					-0.0818	0.0689	
nke		0.0568	0.1112			0.0866	
pfe			-0.0657			-0.1046	
pg				-0.0424	-0.0436		
utx			0.1786			0.0922	0.1117
v		-0.0628		-0.0881	-0.0536		
vz						0.0524	-0.0434
wmt		-0.0959	-0.0678				
Panel D: insignificant impact multiplier in all models							
ba							

Table 3.5: Long-run multiplier of ESG sentiment on idiosyncratic returns

This table summarizes the long-run multiplier of ESG-related sentiment on idiosyncratic returns in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative LRM estimate in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to permanent changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant LRM in Model (1), while Panel D collects stocks that show no significant LRM effects throughout all models. The respective model is indicated by its equation number and respective LRM coefficient. θ_l^+ and θ_l^- are the LRM estimates for the positive $(+)$ and negative $(-)$ sentiment variables, respectively, where $l = 0$ represents the LRM of Model (1) and $l = 1, 2$ represent the LRM belonging to the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	θ_0	θ_0^+	θ_0^-	θ_1^+	θ_2^+	θ_1^-	θ_2^-
Panel A: negative LRM Model (1)							
aapl	-0.1411	-0.2665		-0.1907	-0.1475		
hd	-0.0571	-0.0714			-0.0704	-0.0652	
unh	-0.0678		0.2313				
utx	-0.0640		0.1805				
Panel B: positive LRM Model (1)							
ko	0.0446	0.1826	0.1028				
mrk	0.0818			0.1330			
pfe	0.0820		-0.1057			-0.1061	
Panel C: insignificant LRM Model (1)							
axp			0.1909			0.4955	
ba			0.1697				
cat				0.1112		0.2994	
cvx						0.0996	
dis		-0.0754	-0.1882	-0.0808		-0.1081	-0.1528
ge							-0.3906
ibm			-0.0525			0.0874	-0.2411
intc					0.1044		0.6508
jnj					-0.1466		-0.2536
jpm					-0.1312		
mmm					-0.1034		
msft							-0.1398
nke					-0.0848		
pg			-0.1816	-0.0723		-0.1617	
v			-0.2523				
vz					-0.0994	0.1295	
wmt		-0.1785	-0.1112		-0.2351		-0.1709
Panel D: insignificant LRM in all models							
csco, gs, mcd, xom							

Table 3.5 reveals that only for a few stocks significant effects in reaction to temporary changes in ESG-related sentiment manifest into significant long-run multiplier. However, those LRM effects we do observe appear to broadly support the findings summarized for the impact multiplier estimates: Throughout all models, the four ESG-averse stocks in Panel A of Table 3.5 show either significant, negative LRM, i.e., a decrease in idiosyncratic returns, in reaction to an increase in positive ESG-related sentiment or do not show significant LRM at all. While we find more ESG-averse stocks for the LRM effects than in case

of the impact multiplier, the group of stocks that show an overall ESG-affine attitude now only consists of three stocks. These three, however, show similar patterns as we observe for their impact multiplier counterparts. Panel C, again, constitutes the largest group of stocks. For 15 out of the 17 ESG-neutral stocks we do not observe any significant reaction to an increase in positive ESG-related sentiment, similar to above. When interacted with the financial performance of a stock, we see that the skeptical attitude of investors towards positive ESG news that we observe for temporary changes in sentiment is still clearly visible with predominantly negative LRM effects in the first two columns of Model (3). In contrast to the impact multiplier effects, investors appear to punish a permanent increase in negative ESG-related news more severely if paired with a stock that is financially performing well. If, however, a stock yields negative idiosyncratic returns, investors of the large group of ESG-neutral stocks do not systematically withdraw their capital.

With respect to the magnitude of the effects, by construction, the long-run multiplier are on average larger than their impact counterparts. However, all coefficients estimates are relatively small, indicating that such ESG-based information is unlikely to provide a valuable input to ESG-related trading strategies. With the smallest, absolute LRM of 0.05 and the largest of 0.49, a permanent change in ESG-related sentiment by one standard deviation is estimated to lead to a shift in conditional average idiosyncratic returns of between 0.5 to 4.9 basis points over one day. A one standard deviation temporary shock in sentiment is estimated to lead to a change in conditional average idiosyncratic returns of between 0.3 to 2.3 basis points.

Summarizing the results of our ARDL models, we find that for the majority of the DJIA constituents investors do not appear to be inherently inclined to favor or reject ESG activities, but to perceive ESG news differently depending on the content of the news, the firm's current stock market performance and depending on the nature of the ESG-related shock (temporary or permanent) under consideration. For example, for the majority of the stocks we find that the investors react cautiously both to a temporary and to a permanent increase in positive ESG-related sentiment by withdrawing capital, similar to the findings in Krüger (2015). While no clear patterns in investors' reaction to a temporary increase in negative ESG-related sentiment can be observed, they do engage in increased selling of their shares if such a negative shock is of permanent nature and the stock is currently not underperforming. Moreover, there are some stocks whose idiosyncratic returns react inherently negatively to positive ESG-related news and another group of stocks whose investors appear to appreciate positive ESG news. The latter finding would speak in favor of the beneficial aspect of ESG reported by Lins *et al.* (2017), who attribute this to the trust-building nature of ESG activities. The group of inherently ESG-affine stocks, however, decreases in number when considering a permanent shock to ESG-related sentiment, whereas more stocks appear to be negatively predisposed towards ESG activities in comparison to the scenario of a temporary shock to ESG-related sentiment. Most of these findings, with the exception of the specific stocks belonging to each group, appear to be robust both to the alternative sentiment dictionary approach (VADER-LM), briefly introduced in Section 3.4.1, and to a sample split in 2015, as the Tables 3.7.2 to 3.7.7 of the appendix show.

3.5.2 Economic implications

In search for more lessons that can be learned from our findings, Table 3.6 summarizes the characteristics of the different groups of stocks – ESG-averse, ESG-affine, ESG-neutral, and stocks with throughout insignificant impact and long-run multiplier – that we have identified via our ARDL estimation results. More specifically, by investigating averages of some economic indicators for each group we are looking for feature patterns that are shared among the stocks in each group and differ in comparison across groups. Since we often have only a few stocks in each group, median instead of mean values over all stocks in a respective group are reported to prevent large outlier values from overly influencing the group values. Especially for the ESG-averse and insignificant groups in Panel A of Table 3.6 one needs to keep in mind that there is only one stock in each of them.

Table 3.6: Group summary descriptives for estimation results
This table provides an overview of some summary statistics for the groups of stocks in Tables 3.4 (Panel A) and 3.5 (Panel B). The entries in the column *Subsample* refer to the respective panels of Tables 3.4 and 3.5, i.e., "ESG-averse" refers to Panel A, "ESG-affine" to Panel B, "ESG-neutral" to Panel C, and "insignificant" to Panel D of each respective table. Each column shows the group's median value for the respective category, most of which are explained in Table 3.1. The last two columns, *Log-returns* and *Sentiment*, provide the groups' median log-returns, multiplied by 100, and sentiment values over the entire sampling period, respectively. Here, log-returns instead of idiosyncratic returns are chosen since the latter, by construction, show a mean and median close to zero.

Subsample	M-cap	Ownership	CSR	Rel news	Log-returns	Sentiment
Panel A: impact multiplier						
ESG-averse	233.6610	0.8931	53.0	0.1406	0.0923	0.0421
ESG-affine	218.9800	0.7419	57.0	0.5096	0.0223	0.1415
ESG-neutral	230.8180	0.6904	59.0	0.4101	0.0366	0.0964
insignificant	196.5440	0.7400	58.0	0.5918	0.0780	0.1936
Panel B: long-run multiplier						
ESG-averse	225.7645	0.7786	57.0	0.2461	0.0755	0.0577
ESG-affine	218.9800	0.7233	58.0	0.4101	0.0320	0.0969
ESG-neutral	230.8180	0.6904	58.0	0.4412	0.0366	0.0964
insignificant	199.8360	0.7148	56.5	0.5756	0.0128	0.1255

Starting from left to right, the stocks' average market capitalization, expressed in billion US Dollars, does not appear to be indicative of the group's reaction towards a change in ESG-related sentiment. By construction, all constituents of the DJIA show a rather high market capitalization, which is why we might not see any further patterns here. A high average percentage of institutional ownership, on the other hand, appears to be a feature of ESG-averse stocks. Intuitively, it is plausible to argue that institutional shareholders care less about ESG-activities of a company but instead are focused on economic performance in the short run – a hypothesis that supports the results of Kim *et al.* (2014). This finding would also speak in favor of the agency-problem point of view mentioned in Krüger (2015) and the profit argument made by Friedman (2007). We furthermore look into publicly available corporate sustainability ratings published by *CSR Hub* in order to see, for example, whether ESG-averse stocks might show differences in their CSR rating from ESG-averse stocks. Indeed, investors of stocks with the lowest CSR ratings react more aversely to temporary ESG news on average, as Panel A reveals, while companies with the highest CSR ratings do not necessarily have the most ESG-affine investors. However, for the LRM results summarized in Panel B the differences in CSR ratings across the groups are only marginal. Moreover, ESG-averse stocks show a rather low median news coverage, which is a relative measure of how many ESG-related articles were published over the sampling period for a specific company, where Coca-Cola (ko) as the company with most news story counts receives the value 1. The clearest pattern in economic indicators across both our impact and long-run multiplier analysis can be found in the median log-returns of the groups. Log-returns, a measure for how well the respective group of stocks performed over the entire sampling period, are to some extent inversely related to their investors' attitude towards ESG activities: Those stocks that show a throughout negative reaction towards increased ESG efforts have performed best between 2010 and 2018, while the ESG-affine stocks are amongst the worst performing groups of stocks. Similar to the findings in Brammer *et al.* (2006) and contrary to Lins *et al.* (2017), this would imply that stocks whose investors react positively to more positive ESG-related news do not appear to be able to reap economic benefits from their ESG efforts in comparison with other DJIA constituents. While yielding the largest returns, the ESG-averse stocks show the lowest sentiment scores both for a temporary sentiment shock and for the LRM.

3.6 Conclusion

With steadily increasing public and regulatory demand for publicly listed companies to adhere to socially responsible business practices and environmentally sustainable modes of operation, furthering our understanding of how companies' efforts in the domain of environmental, social, and governance are perceived by investors and affect corporate performance appears to be of utmost importance. Using nine years worth of ESG-related news articles, we entertain several autoregressive distributed lag models to analyze how idiosyncratic returns of the Dow Jones Industrial Average constituents react towards stock-specific ESG information. In order to evaluate the content of each ESG-related news article, we extract its sentiment using the dictionary approach advocated by Loughran and McDonald (2011) and subsequently calculate a polarity ESG-sentiment index for each trading day of each stock of the sample.

The vast majority of DJIA constituents' idiosyncratic returns react significantly to a temporary as well as to a permanent change in ESG-related news sentiment. While we find supporting evidence both for a beneficial and for a detrimental effect on firm value, for the majority of DJIA constituents the type of the ESG-information (positive or negative), the duration of the shock (temporary or permanent), and the stock's current financial performance are critical for the response in idiosyncratic returns. Investors of most DJIA constituents are on average either sceptical or indifferent towards positive ESG-related sentiment. On the other hand, they are more attentive to a negative, permanent shift in ESG-related sentiment when, financially, a stock is performing well than when it yields negative idiosyncratic returns. Furthermore, companies with high institutional ownership appear to have predominantly investors that are particularly ESG-averse, yet these stocks' returns perform best over the entire sampling period. Stocks that show the tendency to react with an increase in idiosyncratic returns to positive ESG news, on the other hand, show one of the weakest financial performances of all stocks of the sample.

Our findings offer several take-aways for companies and investors alike but also show new avenues for further research. First of all, companies should pay close attention their activities in the domain of ESG. While our results indicate that ESG activities should not be thoughtlessly used as a means to greenwash a firm's public image and oftentimes fail to distract from financial underperformance, they do appear to hold the potential to mitigate financial losses, if investors' attitude towards them is generally positive. Thus, our results mediate the existing opposing views about the value of ESG activities in both academia and industry in that we show that ESG-efforts are neither unambiguously detrimental nor valuable, but context dependent. Therefore, it would be worthwhile to further explore the relationship between ESG-activities and stock market returns in different scenarios. For example, the striking imbalance in financial performance that we find between the ESG-averse and ESG-affine groups of stocks is an aspect that would be of utmost interest to both practitioners and researchers. While limitations in data availability did not allow for a larger sample of stocks for our analysis, access to larger data sets of news articles could provide additional insights as to whether the patterns we found also hold for, for example, the S&P 500 stocks. Larger samples could then also allow to compare or identify industry clusters or further correlations between market capitalization and ESG-effects. While we intentionally excluded social media data, such data could proof valuable, if one would want to change the focus to the influence of ESG-related investor sentiment on returns or volatility. Even though the results of the VADER-LM approach to calculating the sentiment index broadly supports the findings of the LM approach, it would be interesting to further investigate the differences between the two, and potentially also several other, sentiment measures. Lastly, our findings should by no means discourage companies from engaging in socially responsible behavior but rather serve to illustrate the high degree of complexity of the topic, which requires a coordinated approach by both firms and investors to assess the true value of ESG activities.

3.7 Appendix

3.7.1 Long-run multiplier derivation

This section provides a brief, closer investigation of the long-run multiplier, which is mainly based on Brissimis (1976) and Bewley (1978), in order to help to facilitate a better understanding of the concept. Taking the simplest case of Equation (3.3), in period $t+5$ it would read

$$ir_{t+5} = \alpha + \sum_{s=0}^{5} \beta_{0,s} \cdot S_{t+5-s} + \sum_{v=1}^{5} \gamma_{0,v} \cdot ir_{t+5-v} + \delta' \cdot \boldsymbol{x}_{t+5} + \varepsilon_{t+5}. \tag{3.7.1}$$

Next, we define $S_{t+5}, S_{t+4}, \ldots, S_t$ as a function of the continuous variable h. Permanent changes to h, starting in period t, would then lead to $\frac{\partial S_{t+5-s}}{\partial h} > 0$ for $s = 0, 1, \ldots, 5$. Since the idiosyncratic returns are a function of S_t, changes in h would also effect ir_t. Consequentially, taking the first partial derivative $\frac{\partial ir_{t+5}}{\partial h}$ leads to

$$\frac{\partial ir_{t+5}}{\partial h} = \sum_{s=0}^{5} \beta_{0,s} \cdot \frac{\partial S_{t+5-s}}{\partial h} + \sum_{v=1}^{5} \gamma_{0,v} \cdot \frac{\partial ir_{t+5-v}}{\partial h}. \tag{3.7.2}$$

Assuming unit changes, i.e., $\frac{\partial S_{t+5-s}}{\partial h} = 1$, and substituting the partial derivatives of $\frac{\partial ir_{t+5-v}}{\partial h}$ by θ_{j-v}, we get

$$\theta_j = \sum_{s=0}^{5} \beta_{0,s} + \sum_{v=1}^{5} \gamma_{0,v} \cdot \theta_{j-v}. \tag{3.7.3}$$

While not explicitly shown here, the individual multipliers θ_{j-v} are monotonously increasing, however at a decreasing rate such that the sequence of multipliers converges towards what we consider the long-run multiplier θ. This LRM, the change in idiosyncratic returns as a reaction to a permanent change in the sentiment variable S_t as given in Equation (3.6), can thus be understood as the converging sequence of the individual multipliers θ_{j-v}. Therefore, we can derive the LRM by setting $\theta_j = \theta_{j-1} = \cdots = \theta_{j-5} = \theta$ and solving for θ.

Analogously, we can also express the LRM for models (2) and (3) by Equation (3.6) by substituting the nominator by the respective β-coefficient of the sentiment variable or interaction term between sentiment and financial performance indicator D_t of each model.

3.7.2 Supplementary tables

Table 3.7.1: ESG-related subject terms and financial media sources
Panel A lists the ESG-related search terms that are used to filter the LexisNexis data base for news articles, whose subject belongs to the domain of environmental, social, and governance issues. Panel B provides an overview of the finance-related publishing sources that we consider particularly relevant to investors, such that their news articles receive a higher weight in our daily news average sentiment calculation.

Panel A: ESG subjects	
Charities	Ethics & corporate citizenship
Corporate environmental responsibility	Gender equality
Corporate responsibility	Societal issues
Corporate social responsibility	Sustainable development
Corporate sustainability	Sustainable investing
ESG	Wages & salaries
Ethical investing	
Panel B: Financial media sources	
Bloomberg	MainStreet
CNBC	Money Management
CNN	Nordic Region Pensions & Investments News
European Pensions Investments News	OptionsProfits
Financial Adviser	Pensions Expert
Financial Times	Pensions Management
Follow The Money	Professional Wealth Management
Foreign Direct Investment	RealMoney
Fox Business	SEC Filings
Fox News Network	Stocks Under $10
FT Energy Newsletters	The Daily Swing Trade
FT.com	TheStreet
FTAdviser	Thomson Reuters
Growth Seeker	Tribune Content Agency
High Net Worth	USA Today
Information Bank Abstracts	USNEWS
Investment Adviser	Wall Street
Investorschronicle	WebNews English
Jim Cramer's Action Alerts PLUS	WebNews Japanese
Kiplinger	Zacks

Table 3.7.2: Impact multiplier of ESG sentiment on idiosyncratic returns (VADER-LM)

This table summarizes the estimated impact multiplier, i.e., the effects of temporary changes in ESG-related sentiment on idiosyncratic returns, for the VADER-LM sentiment approach in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative impact multiplier in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to temporary changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant impact multiplier in Model (1), while Panel D collects stocks that show no significant estimates for the impact multiplier throughout all models. The respective model is indicated by its equation number and respective sentiment coefficient. β_l^+ and β_l^- are the coefficients of positive (+) and negative (−) sentiment variables, respectively, where $l = 0$ represents the sentiment coefficient of Model (1) and $l = 1, 2$ represent the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	β_0	β_0^+	β_0^-	β_1^+	β_2^+	β_1^-	β_2^-
Panel A: negative impact multiplier Model (1)							
ibm	-0.0312	-0.0399	-0.0124			0.0065	-41.8248
intc	-0.0435	-0.0534		-0.0557			-0.0176
utx	-0.0319		0.0247		-0.0332		0.0170
Panel B: positive impact multiplier Model (1)							
ba	0.0494	0.0626			0.0478	0.0211	
cat	0.0648		-0.0626		0.0665		-0.0364
pg	0.0413	0.0368	-0.0187				-0.0164
xom	0.0323	0.0443			0.0375		
Panel C: insignificant impact multiplier Model (1)							
aapl				-0.0423			0.0487
cvx					-0.0287	0.0294	
dis			-0.0265	-0.0576			
ge			-0.0325				-0.0345
gs					-0.0326		
hd			-0.0141				-0.0865
jnj		-0.0272	-0.0268				
jpm						0.0322	
ko			-0.0187	-0.0347		0.0164	-0.0284
mcd		0.0373					0.0469
msft						0.0258	
nke			0.0667				
pfe		-0.0341	-0.0493	-0.0273	-0.0450	-0.0347	
unh							-0.0176
v		-0.0878			-0.0759		
Panel D: insignificant impact multiplier in all models							
axp, csco, mmm, mrk, vz, wmt							

Table 3.7.3: Long-run multiplier of ESG sentiment on idiosyncratic returns (VADER-LM)

This table summarizes the long-run multiplier of ESG-related sentiment on idiosyncratic returns for the VADER-LM sentiment approach in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative LRM estimate in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to permanent changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant LRM in Model (1), while Panel D collects stocks that show no significant LRM effects throughout all models. The respective model is indicated by its equation number and respective LRM coefficient. θ_l^+ and θ_l^- are the LRM estimates for the positive $(+)$ and negative $(-)$ sentiment variables, respectively, where $l = 0$ represents the LRM of Model (1) and $l = 1, 2$ represent the LRM belonging to the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	θ_0	θ_0^+	θ_0^-	θ_1^+	θ_2^+	θ_1^-	θ_2^-
Panel A: negative LRM Model (1)							
aapl	-0.1518	-0.2251	-0.0921	-0.0908	-0.1220		
hd	-0.0590	-0.0622			-0.0776	-0.0175	0.0734
utx	-0.0543	-0.0484					
Panel B: positive LRM Model (1)							
cat	0.1158	0.0898					
cvx	0.0609	0.0554					
mrk	0.0787	0.0783		0.1077			
Panel C: insignificant LRM Model (1)							
axp						-0.0613	0.0630
ba							0.0895
dis						-0.0773	
gs			-0.0951				
ibm						0.0391	-38.9691
intc				-0.0809			0.1377
jnj			-0.0945		-0.0833		-0.0964
jpm							-0.0732
nke				0.0886		0.0838	
pg						-0.0460	
unh			0.0496				
wmt			-0.0739			0.0486	-0.0964
Panel D: insignificant LRM in all models							
csco, ge, ko, mcd, mmm, msft, pfe, v, vz, xom							

Table 3.7.4: Impact multiplier of ESG sentiment on idiosyncratic returns (2010-2014)

This table summarizes the estimated impact multiplier, i.e., the effects of temporary changes in ESG-related sentiment on idiosyncratic returns, for the subsample encompassing the years 2010 to 2014 in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative impact multiplier in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to temporary changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant impact multiplier in Model (1), while Panel D collects stocks that show no significant estimates for the impact multiplier throughout all models. The respective model is indicated by its equation number and respective sentiment coefficient. β_l^+ and β_l^- are the coefficients of positive (+) and negative (−) sentiment variables, respectively, where $l = 0$ represents the sentiment coefficient of Model (1) and $l = 1, 2$ represent the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	β_0	β_0^+	β_0^-	β_1^+	β_2^+	β_1^-	β_2^-
Panel A: negative impact multiplier Model (1)							
aapl	-0.0648	-0.1092		-0.1794			
cvx	-0.0314		0.0748			0.1046	
v	-0.0950	-0.1208		-0.1473			
Panel B: positive impact multiplier Model (1)							
axp	0.0425	0.0670					
csco	0.0952						
gs	0.0509	0.1125					
ko	0.0493	0.0840					
xom	0.0273	0.0459	0.0574	0.0359		0.0569	
Panel C: insignificant impact multiplier Model (1)							
ba					-0.0707		-0.0803
cat							-0.1284
ge							-0.1455
hd			-0.1427	-0.0664			-0.3194
ibm			-0.0796				
intc				-0.0549			-0.1023
jnj							-0.0456
jpm			0.1444			0.1149	
mcd							0.1390
mrk							0.1351
msft				0.0699	-0.1039		
nke					-0.0701		-0.0740
pfe						-0.1487	
pg					-0.0502		
utx			0.1108		-0.0343	0.0972	
wmt		-0.0821	-0.0723				
Panel D: insignificant impact multiplier in all models							
dis, mmm, unh, vz							

Table 3.7.5: Impact multiplier of ESG sentiment on idiosyncratic returns (2015-2018)

This table summarizes the estimated impact multiplier, i.e., the effects of temporary changes in ESG-related sentiment on idiosyncratic returns, for the subsample encompassing the years 2015 to 2018 in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative impact multiplier in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to temporary changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant impact multiplier in Model (1), while Panel D collects stocks that show no significant estimates for the impact multiplier throughout all models. The respective model is indicated by its equation number and respective sentiment coefficient. β_l^+ and β_l^- are the coefficients of positive (+) and negative (−) sentiment variables, respectively, where $l = 0$ represents the sentiment coefficient of Model (1) and $l = 1, 2$ represent the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	β_0	β_0^+	β_0^-	β_1^+	β_2^+	β_1^-	β_2^-
Panel A: negative impact multiplier Model (1)							
ibm	-0.0383	-0.0704				0.0361	-0.0779
unh	-0.0532	-0.0680		-0.0462			0.2419
Panel B: positive impact multiplier Model (1)							
ba	0.0972	0.1245			0.0955		
ko	0.0697	0.1498			0.0878		
Panel C: insignificant impact multiplier Model (1)							
aapl							0.1657
cat						0.1124	0.2066
csco			-0.2668			-0.2597	
cvx		0.0825					
dis		-0.1030	-0.1259	-0.0667			
gs					-0.0723		
hd		0.0397	0.0790				
jnj			-0.1158				-0.0883
jpm		0.0580		0.0789			
mmm						0.1005	
mrk		0.0945				0.1295	
nke		0.1071	0.1725			0.1016	
pfe		-0.0855	-0.0894		-0.0675	-0.0884	
pg			-0.1112	-0.0738			
utx			0.3648				0.2401
v					-0.0409		-0.0470
wmt		-0.1125					
xom					0.0776		-0.1342
Panel D: insignificant impact multiplier in all models							
ge, mcd, msft, vz							

Table 3.7.6: Long-run multiplier of ESG sentiment on idiosyncratic returns (2010-2014)
This table summarizes the long-run multiplier of ESG-related sentiment on idiosyncratic returns for the subsample encompassing the years 2010 to 2014 in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative LRM estimate in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to permanent changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant LRM in Model (1), while Panel D collects stocks that show no significant LRM effects throughout all models. The respective model is indicated by its equation number and respective LRM coefficient. θ_l^+ and θ_l^- are the LRM estimates for the positive $(+)$ and negative $(-)$ sentiment variables, respectively, where $l = 0$ represents the LRM of Model (1) and $l = 1, 2$ represent the LRM belonging to the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	θ_0	θ_0^+	θ_0^-	θ_1^+	θ_2^+	θ_1^-	θ_2^-
Panel A: negative LRM Model (1)							
aapl	-0.2240	-0.3159		-0.2405			
msft	-0.1044	-0.1486			-0.3109		
unh	-0.1386	-0.1156	0.3082			0.4966	
utx	-0.0701		0.2707			0.1462	
Panel B: positive LRM Model (1)							
Panel C: insignificant LRM Model (1)							
axp			0.1992			0.4019	0.1581
ba			0.2500	-0.1016		-0.1896	
csco							-0.6819
cvx						0.1118	
dis		-0.0947	-0.2078				
ge							-0.4639
hd		-0.0873			-0.0693	-0.1274	0.1699
ibm						0.1809	-9.7532
intc					0.1230	0.4165	0.6023
jnj							-0.2509
jpm						0.2739	
ko		0.1784					
mcd						-0.1256	0.1701
mrk					-0.0800		
nke			-0.2816	-0.1010			
wmt					-0.1652		
xom					-0.0787	-0.1190	0.0762
Panel D: insignificant LRM in all models							
cat, gs, mmm, pfe, pg, v, vz							

Table 3.7.7: Long-run multiplier of ESG sentiment on idiosyncratic returns (2015-2018)

This table summarizes the long-run multiplier of ESG-related sentiment on idiosyncratic returns for the subsample encompassing the years 2015 to 2018 in all ARDL models that are significant on an $\alpha \leq 10\%$. Panel A presents those stocks that show a predominantly negative LRM estimate in Model (1), while Panel B focuses on those stocks that tend to show a positive reaction in idiosyncratic returns to permanent changes in ESG-related sentiment. Panel C then collects all stocks that do not show a significant LRM in Model (1), while Panel D collects stocks that show no significant LRM effects throughout all models. The respective model is indicated by its equation number and respective LRM coefficient. θ_l^+ and θ_l^- are the LRM estimates for the positive $(+)$ and negative $(-)$ sentiment variables, respectively, where $l = 0$ represents the LRM of Model (1) and $l = 1, 2$ represent the LRM belonging to the interaction terms of positive (1) and negative (2) idiosyncratic returns with ESG-related sentiment. The idiosyncratic returns used as the dependent variable are computed using log-returns scaled by the factor 100.

	Model (1)	**Model (2)**		**Model (3)**			
Ticker	θ_0	θ_0^+	θ_0^-	θ_1^+	θ_2^+	θ_1^-	θ_2^-
Panel A: negative LRM Model (1)							
jpm	-0.0789				-0.2416		
wmt	-0.1173	-0.3266	-0.2062		-0.3079		-0.2778
Panel B: positive LRM Model (1)							
cat	0.1373	0.2110		0.2159		1.1952	
cvx	0.1150						
mrk	0.1709	0.1790		0.4020			-0.2378
pfe	0.1168		-0.2066			-0.1282	
v	0.0548		-0.2573		-0.1039		
Panel C: insignificant LRM Model (1)							
aapl		-0.2210					
ba					0.2305	0.2308	
gs			-0.2748				
hd					-0.1232		
ibm					0.1611		-0.1277
jnj					-0.2420		-0.2822
ko		0.1848					
nke			0.1670				
pg				-0.1912			
unh				-0.0990			
xom					0.1037		
Panel D: insignificant LRM in all models							
csco, dis, ge, mcd, mmm, msft, utx, vz							

Chapter 4

Multiple structural breaks in cointegrating regressions: A model selection approach

4.1 Introduction

Reliably detecting structural change in multivariate time series models has increasingly gained importance over the last decade. A diverse literature has emerged which is concerned with estimation and testing of unknown structural breaks (see, for example, Perron, 2006; Qu and Perron, 2007; Kejriwal and Perron, 2010; Aue and Horváth, 2013; Perron and Yamamoto, 2016; Qian and Su, 2016; Kurozumi and Skrobotov, 2017). It is well-known that coefficients of time series regressions are potentially inconsistently estimated if structural breaks are not accounted for. Further, statistical inference in these situations is unreliable as the size and power properties of statistical tests are distorted. This holds particularly for cointegration models in the spirit of Engle and Granger (1987), for which a long-run equilibrium relationship is estimated under the assumption of parameter constancy. Ignoring break dates at which the cointegrated system attains a new equilibrium might severely confound the cointegration analysis. Thus, accounting for unknown structural breaks in cointegration models and consistently estimating them if they occur during the sample period is of utmost importance for applied economic research.

For this matter, we propose a new approach to detect structural breaks in a potentially cointegrated regression using penalized regression techniques. Based on this approach, we develop residual-based tests for cointegration which are valid in the presence of multiple structural breaks.

A structural break in a cointegrating regression, as it will be dealt with in this paper, occurs when either the intercept or the slope coefficient (or both) change substantially at one point in time. Early on, Gregory and Hansen (1996a) have recognized the need for cointegration tests which account for the presence of structural breaks. They allow the cointegrating vector to change at an unknown point in time during the sample period and use a grid search approach to determine the break date. Their test performs well in situations where commonly applied residual-based tests fail to detect a cointegration relationship (Gregory *et al.*, 1996). However, their test is limited to only one such change in the long-run equilibrium and performs poorly in the presence of more than one break. Moreover, they do not consider whether the breakpoint's timing and the magnitude of parameter changes are consistently estimated. Hatemi-J (2008) developed a test for similar models with two structural breaks. The problem with grid search algorithms for structural breaks, as they are being used for the aforementioned tests, lies in the exponential increase in computing time with an increasing number of breaks and the crucial assumption that the exact number of breaks has to be known a priori.

Bai and Perron (1998) proposed a procedure which allows to detect multiple structural breaks sequentially. This approach was further developed in Kejriwal and Perron (2010) and Maki (2012), adapting it to cointegrating regressions. In contrast to grid search algorithms, their sequential approach does not require the user to know the exact number of breaks. Instead, only a maximum number of breaks needs to be specified. After estimating a baseline model without structural breaks, each added breakpoint is tested as to whether it improves the fit of the linear regression model. Westerlund and Edgerton (2006) design Lagrange Multiplier (LM)-based test statistics invariant to structural breaks to test the null of no cointegration and Davidson and Monticini (2010) use subsample procedures to account for structural breaks in their cointegration tests. Carrion-i Silvestre and Sanso (2006) and Arai and Kurozumi (2007) propose a CUSUM-based approach to test the null hypothesis of cointegration with a structural break against the alternative hypothesis of no cointegration.

One major disadvantage of most approaches to model structural change in the cointegration literature is that they mainly focus on improving the cointegration test. However, if we plan to further analyze cointegrated data after having tested for cointegration, e.g., specifying error correction models, we are interested in obtaining consistently estimated cointegration residuals. This necessitates to find the exact break dates and to estimate the magnitude of the breaks consistently. Hence, with our new approach we pursue three objectives, namely (i) detecting multiple structural breaks in cointegrating regressions, (ii) consistently estimating the magnitude of those breaks, and (iii) testing for cointegration in the presence of multiple structural breaks.

We achieve these three objectives by perceiving the task of detecting and estimating structural breaks as a model selection problem. We could potentially shift and turn the regression hyperplane at every point in time using the appropriate indicator functions. Finding the true breakpoints corresponds to selecting the right indicators and eliminating irrelevant indicators. This leads to a high-dimensional setting with the total amount of parameters of the model close to the number of observations. The lasso estimator, introduced by Tibshirani (1996), in principle has attractive properties in these settings. However, quite restrictive regularity conditions about the design matrix are needed for simultaneous variable selection and consistent parameter estimation. Knight and Fu (2000) discuss the asymptotic behavior of the lasso estimator under different regularity conditions. They show that the lasso is an inconsistent estimator if the regressors are highly correlated. Subsequent studies built on their results and propose slightly different extensions such as the fused lasso (Tibshirani *et al.*, 2005), the adaptive lasso (Zou, 2006) and grouped lasso (Yuan and Lin, 2006), among others. Particularly, the adaptive lasso is shown to have the oracle property under a broad set of assumptions which means it performs consistent variable selection and parameter estimation.

The use of lasso estimators in cointegrating regressions has been discussed in a few recent studies. To begin with, Mendes (2011) investigates the asymptotic properties of the adaptive lasso estimator in cointegrating regressions with additional stationary components. He shows that also in this context the adaptive lasso estimator has the oracle property under some regularity conditions. The adaptive lasso estimator has further been used by Liao and Phillips (2015) to simultaneously estimate the cointegrating rank and autoregressive order of a Vector Error Correction Model (VECM). An extension to the conventional Johansen model in high-dimensional settings with a short sampling period can be found in Wilms and Croux (2016). Their sparse cointegration model using a lasso approach is shown to outperform the Johansen method in terms of forecasting precision if some elements of the cointegrating vectors are exactly zero. Koo *et al.* (2017) apply the lasso to predictive regressions involving highly persistent and potentially cointegrated time series. Their proposed lasso approach leads to superior forecasting performance relative to the OLS estimator based on the full model.

For our purposes, we rely on a modified version of the adaptive lasso procedure, similar to the one presented by Kock (2016), and proceed as follows. First, we operate in a setting with a fixed set of breakpoint candidates and provide a technical proof that our estimator has the oracle property in cointegrating regressions. While being quite restrictive, such a setting, nevertheless, appears to be of practical relevance. One could, for example, be confronted with a situation in which a fixed number of crises occurred at well-known points during the sample period, all of which could potentially have influenced the cointegration relationship. An important question then would be which of these crises actually changed the long-run equilibrium, i.e., which of these crises led to breaks in either slope or intercept (or both) and which did not. In this context, we allow the breakpoints of intercept and slope coefficient to occur at different points in time (i.e., at different crises). Thus, one could also perceive our method as performing an efficient subsample selection.

Second, we build on these results and extend the procedure to more general situations where we do not have any prior information about potential breakpoint candidates but the breaks can occur at any point in time. This corresponds to a diverging number of parameters in the full model. Despite the increased complexity of the setting, we can provide simulation evidence that our procedure still estimates the breakpoints consistently. Lastly, we discuss how our modified adaptive lasso procedure can be used for residual-based tests for cointegration in the presence of multiple structural breaks.

The remainder of this paper is organized as follows. In Section 4.2, we describe the models considered, discuss the asymptotic properties of the adaptive lasso estimator and propose suitable cointegration tests under the presence of multiple structural breaks. Section 4.3 is devoted to the Monte Carlo simulation study and Section 4.4 presents an empirical application in the context of Purchasing Power Parity (PPP). Section 4.5 summarizes our results and states objectives for future research.

4.2 Methodology

To present the main idea, we restrict our analysis to a bivariate cointegration system with structural breaks in the intercept and slope coefficient. While the focus on bivariate systems might appear restrictive at first glance, there is abundant, highly relevant empirical research dealing with such systems, ranging from studies on the PPP and the spot-futures relationship to discussions on asymmetric price transmission along supply chains (see, for example, Taylor and Taylor, 2004; Bekiros and Diks, 2008; Schweikert, 2019). Possible extensions to multivariate cointegration systems are discussed in Section 4.5.

Let $\{y_t\}_1^\infty$ denote a scalar process generated by

$$y_t = \mu_t + \beta_t x_t + u_t, \qquad t = 1, 2, \ldots, \tag{4.2.1}$$

where μ_t and β_t are time-varying coefficients and $\{x_t\}_1^\infty$ follows a random walk process

$$x_t = x_{t-1} + v_t, \qquad t = 1, 2, \ldots, \tag{4.2.2}$$

where $x_0 = 0$. $\{u_t\}_1^\infty$ and $\{v_t\}_1^\infty$ are mean-zero weakly stationary error processes. We make the following assumptions about the vector process $w_t = (u_t, v_t)'$:

Assumption 1. *The vector process $\{w_t\}_1^\infty$ satisfies the following conditions*

1. $Ew_t = 0$ *for* $t = 1, 2, \ldots$.
2. $\{w_t\}_1^\infty$ *is weakly stationary.*
3. $\{w_t\}_1^\infty$ *is strong mixing with mixing coefficients of size* $-\delta\beta/(\delta-\beta)$ *and* $E|w_t|^\delta < \infty$ *for some* $\delta > \beta > 5/2$.

We further make some assumptions about the coefficients μ_t and β_t concerning the number of total changes in a given sample. We treat structural breaks as rare events and assume that parameter changes persist for some time. This assumption is easily justified by the intended application on economic long-run equilibrium relationships which, in order to be meaningful, have to hold over long periods of time. For true random coefficient models without such strict sparsity assumptions, we refer to Quintos and Phillips (1993), Kuo (1998), Park and Hahn (1999), Xiao (2009a), Xiao (2009b) and Bierens and Martins (2010), among others.

Assumption 2. *The total number of distinct values in any set* $\{\mu_1, \ldots, \mu_T\}$ *is* $p+1$*, where* $0 \le p \le p^* \ll T$ *and the total number of distinct values in any set* $\{\beta_1, \ldots, \beta_T\}$ *is* $m+1$*, where* $0 \le m \le m^* \ll T$*. Further, we assume that* $p^* + m^* \ll T$.

We assume that the maximum number of breaks in the intercept, p^*, and slope, m^*, is known beforehand and thereby follow Bai and Perron (1998). The true number of p breaks in the intercept and m breaks in the cointegrating vector is unknown and can be determined from the data. For models with fixed breakpoint candidates and in contrast to Bai and Perron (1998), we allow that the intercept and the slope coefficient have a different number of breaks at different points in time. We denote the distinct coefficients in samples of length T as

$$\mu_t = \tilde{\mu}_i, \quad \text{for } t = T_{1,i-1}, T_{1,i-1}+1, \ldots, T_{1,i}-1, \quad i = 1, \ldots, p+1, \tag{4.2.3}$$

and

$$\beta_t = \tilde{\beta}_j, \quad \text{for } t = T_{2,j-1}, T_{2,j-1}+1, \ldots, T_{2,j}-1, \quad j = 1, \ldots, m+1, \tag{4.2.4}$$

where $T_{1,0} = T_{2,0} = 1$ and $T_{1,p+1} = T_{2,m+1} = T+1$. The relative timing of breakpoints is denoted by $\tau_{1,i} = T_{1,i}/T, i \in \{1, \ldots p\}$ and $\tau_{2,j} = T_{2,j}/T, j \in \{1, \ldots, m\}$, respectively. To study the consistency of our estimator, we need some additional assumptions about the magnitude of the breaks and the distance between them.

Assumption 3. *(i) The minimum break intervals are defined as* $I_{1,\min} = \min_{1 \le i \le p^*+1} |T_{1,i} - T_{1,i-1}| \ge T\varepsilon_1$ *and* $I_{2,\min} = \min_{1 \le j \le m^*+1} |T_{2,j} - T_{2,j-1}| > T\varepsilon_2$ *for some constants* $\varepsilon_1, \varepsilon_2 > 0$.
(ii) The break magnitudes are bounded by $J_{1,\min} < |\tilde{\mu}_i - \tilde{\mu}_{i-1}| < J_{1,\max}$ *for* $2 \le i \le p+1$ *and* $J_{2,\min} < |\tilde{\beta}_j - \tilde{\beta}_{j-1}| < J_{2,\max}$ *for* $2 \le j \le m+1$*, where* $J_{1,\min}$*,* $J_{1,\max}$*,* $J_{2,\min}$ *and* $J_{2,\max}$ *are positive constants.*

Assumption 3(i) requires that the length of the regimes between breaks increases with the sample size and in the same proportions to each other. This allows us to consistently detect and estimate the true break fractions as it makes the break dates asymptotically distinct (Perron, 2006). Assumption 3(ii) excludes the possibility of infinite shifts in the parameters and requires parameter changes to be of a substantial magnitude to distinguish active breaks from inactive breaks.

4.2.1 Fixed breakpoint candidates

First, we consider a special setting where we have prior information about the timing of potential breakpoint candidates in our sample and are interested in efficient subsample selection. Hence, the values of p^*, m^* and $(\tau_{1,1},\ldots,\tau_{1,p^*},\tau_{2,1},\ldots,\tau_{2,m^*})$ are known. The total amount of coefficients in the full model, i.e., of baseline regressors and all breakpoint candidates, is then given by the fixed scalar $d^* = p^* + m^* + 2$. In this case, we can express the long-run equilibrium equation in a regime-specific form such that

$$y_t = \sum_{i=1}^{p^*+1} \mu_i^* \varphi_{t,\tau_{1,i-1}} + \sum_{j=1}^{m^*+1} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} + u_t, \tag{4.2.5}$$

where the indicator variable $\varphi_{t,\tau_{k,l}}$ is defined as

$$\varphi_{t,\tau_{k,l}} = \begin{cases} 0 & \text{if } t < [T\tau_{k,l}] \\ 1 & \text{if } t \geq [T\tau_{k,l}] \end{cases}, \qquad k \in \{1,2\}, \quad t = 1,2,\ldots, \tag{4.2.6}$$

and $\tau_{k,0} = 0$. The coefficients in regime-specific form are $\mu_1^* = \tilde{\mu}_1$, $\mu_i^* = \tilde{\mu}_i - \tilde{\mu}_{i-1}$ for $i = 2,\ldots,p^*+1$ and $\beta_1^* = \tilde{\beta}_1$, $\beta_i^* = \tilde{\beta}_j - \tilde{\beta}_{j-1}$ for $j = 2,\ldots,m^*+1$, i.e., μ_1^* and β_1^* denote the parameter values until the first breakpoint (baseline model), while μ_i^*, $i = 2,\ldots,p^*+1$ and β_j^*, $j = 2,\ldots m^*+1$ denote the parameter changes at all breakpoint candidates. We are primarily interested in a procedure to detect the true number of breakpoints and to consistently estimate the magnitude of the parameter change. Relevant breakpoints should be indicated by nonzero coefficients while irrelevant breakpoints should be eliminated from the model. For that matter, we estimate the cointegrating regression with potentially multiple breaks using an objective function which shrinks irrelevant breakpoints to zero.

A natural choice for such an objective function would be the lasso of Tibshirani (1996). It allows to select relevant coefficients, i.e., those corresponding to active breakpoints, and shrinks the coefficients of irrelevant coefficients, i.e., those corresponding to the other potential but non-active breakpoints, to zero. However, it is well-known that the least squares estimator of μ has convergence rate $\sqrt{T}$ while the least squares estimator of β is superconsistent at rate T. Since we want to recover breaks in both μ and β, we should not shrink both types of coefficients with the same tuning parameter. Another well-known fact is that the plain lasso estimator is not oracle efficient. One way to deal with these issues is to assign individually chosen weights to each coefficient, as in the adaptive lasso of Zou (2006). With these weights the coefficients will experience different degrees of shrinkage even though there is still only one global tuning parameter in the model. The objective function that we will use in the following is a variant of the adaptive lasso objective function and similar to the objective function used in Kock (2016) who investigates model selection in stationary and nonstationary autoregressions. He modifies the objective function such that a different exponent of the weights is added (either γ_1 or γ_2) which depends on the convergence rate of the least squares estimator. It allows us to shrink all elements of the sets $\{\mu_1,\ldots,\mu_T\}$ and $\{\beta_1,\ldots,\beta_T\}$ to zero where no structural change occurs. Subsequently, we detect structural breaks using the index of all nonzero coefficients left after optimization. The objective function can be written as

$$
\begin{aligned}
V_T(\{\mu_t,\beta_t\}) \quad &= \quad \sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{p^*+1}\mu_i^*\varphi_{t,\tau_{1,i-1}} - \sum_{j=1}^{m^*+1}\beta_j^* x_t \varphi_{t,\tau_{2,j-1}}\right)^2 \\
&+ \quad \lambda_T \sum_{i=2}^{p^*+1} w_{1i}^{\gamma_1}|\mu_i^*| + \lambda_T \sum_{j=2}^{m^*+1} w_{2j}^{\gamma_2}|\beta_j^*|, \qquad \gamma_1,\gamma_2 > 0, \qquad (4.2.7)
\end{aligned}
$$

where $w_{1i} = 1/|\hat{\mu}_{I,i}^*|$ for $i = 2,\dots,p^*+1$ and $w_{2j} = 1/|\hat{\beta}_{I,j}^*|$ for $j = 2,\dots,m^*+1$ are coefficient-specific weights based on initial estimates of the coefficients. Note that we do not apply any shrinkage to the baseline model.

The value of the global tuning parameter, λ_T, is generally unknown and has to be estimated from the data. Cross-validation approaches are commonly used for this matter. However, since we later also consider cointegration tests where the null hypothesis is no cointegration, we cannot meaningfully apply these approaches. Further, it is unclear how the training sample should be selected if the number of breakpoints and their timing are unknown. Instead, we follow Kock (2016) and select the tuning parameter by minimizing the BIC,

$$
BIC(\lambda_T) = \log(\hat{u}_\lambda'\hat{u}_\lambda/T) + \log(T)/T\cdot \mathrm{df}, \qquad (4.2.8)
$$

where *df* are the respective degrees of freedom of the model, i.e., amount of nonzero coefficient estimates, and $\hat{u}_\lambda$ are the residuals resulting from the adaptive lasso estimation of Equation 4.2.7.

In the case of fixed breakpoint candidates, the initial estimates $\hat{\mu}_I^*$ and $\hat{\beta}_I^*$ can be obtained from least squares estimation of the long-run equilibrium equation with multiple structural break indicators. The least squares estimator is consistent and yields appropriate weights. However, if the total number of coefficients d^* exceeds the number of observations T, ordinary least squares estimation is not an option and alternatives, e.g., ridge regression or the dimension-reduction procedure outlined in Subsection 4.2.2, have to be considered.

In the following, we establish the first of our main results. We prove that the adaptive lasso estimator tuned to perform consistent model selection has the oracle property in bivariate cointegrating regressions with multiple structural breaks. We show that the adaptive lasso performs correct model selection which requires that the probability of including all truly nonzero coefficients in the model tends to one while the probability of keeping irrelevant variables tends to zero. This satisfies the first property an oracle procedure should possess. Further, the estimator should have an asymptotic normal distribution (Fan and Li, 2001). We show that our estimator has the same asymptotic distribution as the least squares estimator. However, since our regression involves nonstationary components, the asymptotic distribution of the least squares estimator is naturally given as a functional of Brownian motions. Hence, we say that our estimator satisfies a nonstandard oracle property to distinguish it from its stationary counterpart.

We use the following notation to present our main results: A vector of T observations for the variable y_t is denoted by $y = (y_1,\dots,y_T)'$. Similarly, we denote $x = (x_1,\dots,x_T)'$ and $u = (u_1,\dots,u_T)'$ and $\mathbf{1}$ as a T-dimensional vector of ones. Further, we define

$$
\varphi_{\tau_{k,l}} = (\underbrace{0\dots0}_{T_{k,l}-1} \quad \underbrace{1\dots1}_{T-T_{k,l}+1}), \qquad k\in\{1,2\},
$$

to denote break indicators in vector form. We define the identity matrix $\boldsymbol{I}$ and the matrix

$\boldsymbol{X} = (\mathbf{1}, \varphi_{\tau_{1,1}}, \ldots, \varphi_{\tau_{1,p^*}}, x, x\varphi_{\tau_{2,1}}, \ldots, x\varphi_{\tau_{2,m^*}})$ with column rank d^* containing the baseline regressors and all potential break indicator variables. We decompose the matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ into one part $\boldsymbol{X}_1 = (\mathbf{1}, \varphi_{\tau_{1,1}}, \ldots, \varphi_{\tau_{1,p^*}})$ containing the constant and the intercept break indicators and one part $\boldsymbol{X}_2 = (x, x\varphi_{\tau_{2,1}}, \ldots, x\varphi_{\tau_{2,m^*}})$ containing the regressor and the slope break indicators. $\Sigma = E(\boldsymbol{X}_1'\boldsymbol{X}_1)$ is the covariance matrix of $\boldsymbol{X}_1$. We define the index set $\mathscr{A} = \mathscr{A}_1 \cup \mathscr{A}_2$, where $\mathscr{A}_1 = \{1 \leq i \leq p^* + 1 : \mu_i^* \neq 0\}$ denotes the set of active intercept break indicators and $\mathscr{A}_2 = \{1 \leq j \leq m^* + 1 : \beta_j^* \neq 0\}$ denotes the set of active slope break indicators. $|\mathscr{A}|$ denotes the cardinality of the set $\mathscr{A}$ and $\mathscr{A}^c$ denotes the complementary set, i.e., the index set of truly zero coefficients. In that sense, $\hat{\mu}_{T,\mathscr{A}_1}$ ($\hat{\mu}_{T,\mathscr{A}_1^c}$) represents the vector of estimated intercept coefficients belonging to the set of active (inactive) breaks, and $\hat{\beta}_{T,\mathscr{A}_2}$ ($\hat{\beta}_{T,\mathscr{A}_2^c}$) are active (inactive) slope coefficients. $\Sigma_{\mathscr{A}_1}$ indexes the rows and columns of the covariance matrix belonging to the active variables. $B(s)$ denotes Brownian motion with variance σ^2 and $U(s)$ denotes the weak limit of u_t. For notational convenience we use '$\Rightarrow$' to signify weak convergence of the associated probability measures and $\xrightarrow{p}$ to denote convergence in probability. Continuous stochastic processes such as $B(s)$ on [0,1] are simply written as B if no confusion is caused. We also write integrals with respect to the Lebesgue measure such as $\int_0^1 B(s)ds$ simply as $\int_0^1 B$. We use $x\varphi_{\tau_{2,j}} = x \odot \varphi_{\tau_{2,j}}$ as short-hand notation for the Hadamard product involving indicator terms where no confusion arises.

Theorem 1. *Suppose that the scalar processes $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ are cointegrated as described by Equation* (4.2.1)*, Assumptions 1- 3 hold, $\frac{\lambda_T}{\sqrt{T}} \to 0$, $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and $\frac{\lambda_T}{T^{1-\gamma_2}} \to \infty$. Then, the adaptive lasso estimator has nonstandard oracle properties:*

(a)
$$P(\hat{\mu}_{T,\mathscr{A}_1^c} = 0) \to 1$$
$$P(\hat{\beta}_{T,\mathscr{A}_2^c} = 0) \to 1$$

(b)
$$\begin{bmatrix} \sqrt{T}(\hat{\mu}_{T,\mathscr{A}_1} - \mu_{\mathscr{A}_1}) \\ T(\hat{\beta}_{T,\mathscr{A}_2} - \beta_{\mathscr{A}_2}) \end{bmatrix} \Rightarrow \begin{bmatrix} \Sigma_{\mathscr{A}_1} & 0 \\ 0 & \int_0^1 B_{\tau,\mathscr{A}_2} B'_{\tau,\mathscr{A}_2} \end{bmatrix}^{-1} \times \begin{bmatrix} N(0, \Upsilon_{\mathscr{A}_1}\sigma^2) \\ \int_0^1 B_{\tau,\mathscr{A}_2} dU + C^*_{\mathscr{A}_2} \end{bmatrix}$$

$$C^*_{\mathscr{A}_2} = [\Lambda, (1-\tau_{2,1})\Lambda, \ldots, (1-\tau_{2,m^*})\Lambda]', \qquad \Lambda = \sum_{t=0}^{\infty} E(v_t u_0),$$

$$\Upsilon = \begin{bmatrix} 1 & \tau_{1,1} & \ldots & \tau_{1,p^*} \\ \tau_{1,1} & \tau_{1,1} & & \\ \vdots & & & \vdots \\ \tau_{1,p^*} & \ldots & & \tau_{1,p^*} \end{bmatrix}.$$

Statement (a) of the theorem establishes the convergence to zero of truly zero coefficients with probability approaching one. In statement (b), we derive the limit distribution for truly nonzero coefficients. It follows from statement (b) that truly nonzero coefficients are consistently estimated and converge with the same rate as the least squares estimators. As in Theorem 1 of Kock (2016), the exponents of the weights have to satisfy the restrictions $\gamma_1 > 0, \gamma_2 > 1/2$ to guarantee the nonstandard oracle property. However, note that for consistency of the estimates, the exponents γ_1 and γ_2 do not need to deviate from unity, since the weights w_{1i} and w_{2j} already account for the different rates of convergence of μ and β.

The proof of Theorem 1, which can be found in Section 4.6 of the appendix, is given under the assumption that initial least squares estimators are available. The idea of the proof is similar to the proof of Theorem 2 in Zou (2006) and Theorem 1 in Kock (2016).

Remark 1. It is shown in Theorem 2 of Kock (2016) that selecting the tuning parameter via the BIC results in consistent variable selection. His results for nonstationary autoregressions extend straightforwardly to cointegrating regressions under our assumptions.

Remark 2. The results presented in Theorem 1 describe the pointwise asymptotic distribution of our estimator. As Pötscher and Schneider (2009) show, using more general asymptotic theory, the oracle property does not hold uniformly over the parameter space and the rate of convergence can be substantially slower than $(\sqrt{T},T)'$.

4.2.2 Diverging number of breakpoint candidates

If we begin our analysis without any prior information about the timing of the structural breaks, we could consider each $0 < t < T$ to be a potential breakpoint for both μ and β. This would result in a high-dimensional estimation problem with a diverging number of parameters of the full model as $T \to \infty$. Zhang and Huang (2008) investigate the properties of the lasso estimator in a similar, high-dimensional setting under a general sparsity condition. They find that the lasso tends to include variables with large coefficients but also selects some irrelevant variables. Still, the lasso substantially reduces the dimensionality of the estimation problem and provides coefficient estimates which can be used to construct weights for the adaptive lasso method.

We follow Horowitz and Huang (2013) and use the adaptive lasso again to distinguish between active and non-active structural breaks when the number of potential breaks diverges with the sample size, $d_T^* = O(T)$.[1] Since we want to reduce the dimensionality of our estimation problem a priori as much as possible, we impose two restrictions: First, we apply some lateral trimming to exclude the possibility of selecting structural breaks at the beginning or the end of the sample. The degree of trimming is denoted with the parameter ξ. Again, this is motivated by our intended empirical applications in which each regime relates to a newly attained long-run equilibrium and should persist for some time. Second, we do not consider breaks in the intercept for our initial estimation. In a later estimation step, we relax this restriction and instead require the intercept breaks to have the same timing as the slope breaks which is a common restriction in most studies on structural breaks in multiple regression models (Bai and Perron, 1998). A lateral trimming of 5% leaves us with $d_T^* = 0.9 \cdot T + 2$ parameters of the full model and we satisfy the condition $d_T^* < T$ when $T > 20$.

To further reduce the dimensionality of the problem and to obtain useful weights, we apply the plain lasso estimator to our full model,

$$y_t = \mu + \beta_1^* x_t + \sum_{j=2}^{T_\xi} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}} + u_t, \tag{4.2.9}$$

where $T_\xi = [(1-2\xi)T]$ and $\tau_{2,j} \in (\xi, 1-\xi)$ on an equidistant grid. While the lasso estimator does not shrink all irrelevant breaks exactly to zero, we nevertheless obtain consistent estimates of all coefficients.

[1] The subscript T for the scalar d_T^* is added to emphasize the dependence of model complexity on the sample size.

Particularly, we do not eliminate any relevant variables asymptotically if the coefficients exceed a certain threshold.

When a variable is not selected by the first stage lasso estimation, we do not include the variable in the subsequent adaptive lasso stage. If the coefficient is nonzero, we take the reciprocal absolute value as its weight (w_j) and optimize the objective function

$$V_T(\{\mu,\beta_t\}) = \sum_{t=1}^{T}\left(y_t - \mu - \beta_1^* x_t - \sum_{j=2}^{T_\xi} \beta_j^* x_t \varphi_{t,\tau_{2,j-1}}\right)^2 + \lambda_T \sum_{j=2}^{T_\xi} w_j |\beta_j^*|,$$

where the exponent of the weights is unity (see Equation (4.2.7)). The tuning parameter is selected by minimizing the modified BIC proposed in Wang *et al.* (2009),

$$BIC^*(\lambda_T) = \log(\hat{u}_\lambda' \hat{u}_\lambda / T) + \log(T)/T \cdot \mathrm{df} \cdot \log\log d_T^*. \tag{4.2.10}$$

This generalization of the BIC can still identify the true model consistently with a diverging number of parameters as long as $d_T^* < T$ holds. The second stage estimation might still indicate more structural breaks in the slope coefficient than assumed a priori. This could be caused by noisy parameter estimates which are close to zero but not exactly zero. We eliminate remaining irrelevant breaks by using a post-lasso OLS estimation proposed by Belloni and Chernozhukov (2013) as our third and final estimation stage. We select only m^* break indicators corresponding to the m^*-th largest parameter changes for the OLS model. This also allows us to relax the assumption about a constant intercept which might be unrealistic in practice. At the third stage, we can easily add intercept break indicators with the same timing as the slope breaks obtained from the adaptive lasso estimation, i.e., $m^* = p^*$ and $\tau_{1,j} = \tau_{2,j}$ for $j = 1, 2, \ldots, m^*$.[2] However, a crucial aspect for our procedure is the performance of the adaptive lasso estimator if the model is misspecified with respect to the intercept. Taking into consideration the different convergence rates of the least squares estimators and the much higher variation of the slope break indicators, we should be able to detect the slope and indicator breaks sequentially. We analyze this aspect again in our simulation experiments in Section 4.3.

4.2.3 Testing for cointegration

The previous sections have revealed that the (modified) adaptive lasso estimator can be used to detect multiple structural breaks in cointegrating regressions. These results hinge on Assumption 1, which specifies u_t as a stationary error term. In most practical applications, we do not know with certainty whether a particular set of nonstationary variables hold a long-run equilibrium relationship. Therefore, we consider residual-based tests for cointegration which allow for the possibility of multiple structural breaks. The regression in Equation (4.2.1) becomes spurious under the null hypothesis of no cointegration. In this case, the error term u_t is a cumulative sum of innovations and hence integrated of order one. If we apply our adaptive lasso estimator in such situations, we not only obtain spurious least squares coefficients but also have to deal with penalization terms applied to non-existing structural breaks. However, the degree of shrinkage naturally depends on the value of the tuning parameter λ. Further, we still assume that we

[2] As Belloni and Chernozhukov (2013) show, the post-lasso estimation only performs well if all components of the true model are included as a subset of the selected model and the selected model is sufficiently sparse. To avoid multicollinearity in our design matrix, we treat estimated breakpoints with adjacent break dates as a single breakpoint in the post-lasso estimation.

know the maximum number of breaks if the variables were indeed cointegrated and thereby limit the number of location shifts of the error term in situations when the variables are not cointegrated.

Asymptotic distributions of cointegration test statistics in the presence of structural breaks naturally depend on nuisance parameters. More specifically, we encounter the well-known problem that the breakpoints' locations are only identified under the alternative (Andrews and Ploberger, 1994; Hansen, 1996). Additionally, we consider cases in which the number of breakpoints is equally unknown and only limited by a maximum number of breaks. Using a grid of values for the tuning parameter, $\lambda \in \mathscr{L} \subset (0,\infty)$, allows us to construct infimum statistics similar to those used in Gregory and Hansen (1996b) and Hatemi-J (2008). The grid of length L should encompass all model selection choices between the maximum number of breaks and no structural break. We select λ such that the cointegration test statistic is minimized and evaluate the null hypothesis only at this point which provides the most evidence for the alternative.

We conduct our cointegration test in three steps. First, we apply the adaptive lasso estimator to a potentially cointegrated regression with pre-specified maximum number of slope breaks m^* and constant intercept. Without prior knowledge of any break date, we begin to shrink the number of breakpoints from T_ξ and select the m^*-th largest breaks. Second, we re-estimate the long-run equilibrium equation with the selected slope breaks and corresponding intercept breaks ($m^* = p^*$ and $\tau_{1,j} = \tau_{2,j}$ for $j = 1,2,\ldots,m^*$) using post-lasso OLS. Finally, we test the residuals for stationarity using ADF-type and bias-corrected test statistics (Phillips, 1987). The infimum statistics in case of the bias-corrected test statistics Z_λ is given by,

$$Z^* = \inf_{\lambda \in \mathscr{L}} Z_\lambda, \tag{4.2.11}$$

and the ADF-type statistic is constructed analogously.

We make the following assumptions to present the asymptotic distributions of our test statistics. The observed data $z_t = (y_t, x_t)$ is generated as a random walk under the null hypothesis. We define the innovation vector $\Delta z_t = \xi_t$ and its cumulative sum $S_t = \sum_{j=1}^{t} \xi_j$ so that $z_t = z_0 + S_t$. We assume that z_t conforms to the following regularity conditions:

Assumption 4.

1. *z_0 is a random vector with $E|z_0| < \infty$.*
2. *$E\xi_t = 0$ for $t = 1,2,\ldots$.*
3. *$\{\xi_t\}_1^\infty$ is weakly stationary.*
4. *$\{\xi_t\}_1^\infty$ is strong mixing with mixing coefficients of size $-\delta\beta/(\delta-\beta)$ and $E|\xi_t|^\delta < \infty$ for some $\delta > \beta > 5/2$.*
5. *The long-run variance of S_t,*
$$\Omega = \lim_{T\to\infty} T^{-1} E S_t S_t',$$
is positive definite.

We denote the post-lasso cointegrating residuals by $\hat{e}_{t\lambda}$, where the subscript λ indicates that the residual vector depends on the selected number of breaks and their timing, i.e., on the value of λ. We consider the following auxiliary regression,

$$\hat{e}_{t\lambda} = \rho_\lambda \hat{e}_{t-1\lambda} + v_{t\lambda}, \tag{4.2.12}$$

and estimate the bias-corrected first-order serial correlation coefficient as suggested in Phillips (1987),

$$\hat{\rho}_\lambda^* = \sum_{t=2}^{T}(\hat{e}_{t\lambda}\hat{e}_{t-1\lambda} - \hat{\psi}_\lambda) \Big/ \sum_{t=2}^{T}\hat{e}_{t-1\lambda}^2, \tag{4.2.13}$$

where the bias-correction term, $\hat{\psi}_\lambda$, is an estimate of the weighted sum of autocovariances,

$$\hat{\psi}_\lambda = \sum_{j=1}^{M} w(j/M)\hat{\gamma}_\lambda(j), \qquad \hat{\gamma}_\lambda(j) = \frac{1}{T}\sum_{t=j+1}^{T}\hat{v}_{t-j\lambda}\hat{v}_{t\lambda}. \tag{4.2.14}$$

The kernel weights $w(\cdot)$ satisfy standard conditions and the bandwidth is a function of the sample size so that $M \to \infty$ and $M/T^5 = O(1)$. The estimate of the long-run variance of $\hat{v}_{t\lambda}$ is then given by

$$\hat{\sigma}_\lambda^2 = \hat{\gamma}_\lambda(0) + 2\hat{\psi}_\lambda. \tag{4.2.15}$$

We employ the standardized bias-corrected test statistic to evaluate the null hypothesis of no cointegration. The test statistic is given by

$$Z_\lambda = (\hat{\rho}_\lambda^* - 1)/\hat{s}_\lambda, \qquad \hat{s}_\lambda^2 = \hat{\sigma}_\lambda^2 / \sum_{t=2}^{T}\hat{e}_{t-1\lambda}^2 \tag{4.2.16}$$

for each λ. Alternatively, we regress $\Delta\hat{e}_{t\lambda}$ upon $\hat{e}_{t-1\lambda}$ and K lagged differences $\Delta\hat{e}_{t-1\lambda}, \ldots, \Delta\hat{e}_{t-K\lambda}$. In practice, we use order selection rules such as AIC, BIC or a general-to-specific pretesting procedure to determine the lag truncation parameter. We follow Chang and Park (2002) and require that K increases with the sample size.

Assumption 5. *K increases with T in such a way that $K = o(T^{1/2})$.*

The ADF test statistic is the t-ratio for the regressor $\hat{e}_{t-1\lambda}$. We express the asymptotic distribution of our primary test statistic, Z^*, as functionals of Brownian motion.

Theorem 2. *If the scalar processes $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ are generated under the null hypothesis of no cointegration and Assumption 4 holds, then*

$$Z^* \Rightarrow \inf_{\lambda \in \mathscr{L}} \int_0^1 W_\lambda dW_\lambda \Big/ \left(\int_0^1 W_\lambda^2\right)^{1/2} \left(1 + \kappa_\lambda' D_\lambda \kappa_\lambda\right)^{1/2},$$

where

$$W_\lambda = W_1(s) - \int_0^1 W_1 W_{2\lambda}\left[\int_0^1 W_{2\lambda}W_{2\lambda}'\right]^{-1} W_{2\lambda}(s),$$

$$\kappa_\lambda = \left[\int_0^1 W_{2\lambda}W_{2\lambda}'\right]^{-1}\int_0^1 W_{2\lambda}W_1.$$

The dimensionality of $W_{2\lambda}$ depends on the number of selected breakpoint for each value of λ. D_λ is a quadratic matrix with rank equal to the column rank of $W_{2\lambda}$.

Remark 3. The distribution of our test statistic is an infimum statistic over all possible model selection outcomes in the post-lasso estimation. Hence, it depends on the choice of the pre-specified grid $\mathscr{L}$. For a maximum number of m^* breaks in the slope coefficient, we have m^*+1 choices for the number of breaks which can be selected by the lasso estimation procedure depending on the value of the tuning parameter. Suppose that $\lambda \to \infty$, our procedure does not select any (false) structural breaks under the null hypothesis of no cointegration and no structural breaks. On the other hand, if we tune the adaptive lasso to perform conservative model selection ($\lambda \to 0$), we increase the probability of choosing the maximum number of breaks.

Following Phillips and Ouliaris (1990) and Gregory and Hansen (1996a), we expect that the asymptotic distributions of the ADF-type and bias-corrected test statistics are the same if Assumption 5 holds.[3]

4.3 Simulation results

Monte Carlo experiments are utilized to evaluate the finite sample performance of the (modified) adaptive lasso estimator. First, we consider cointegrated systems with several structural break specifications to investigate the theoretical claims developed in Subsection 4.2.1. Particularly, we want to find out whether the timing of the breaks is accurately indicated and whether the estimated change in the parameters is becoming more accurate if we increase the sample size successively. Second, we evaluate the performance of our estimator in models with a diverging number of breakpoint candidates described in Subsection 4.2.2. Here, we are primarily interested in the precision of our estimator when breaks occur in the intercept and slope coefficients simultaneously but only the slope coefficients are specified correctly. Third, we study the size and power of our residual-based tests proposed in Subsection 4.2.3 for different configurations of the Data Generating Process (DGP) and finite sample sizes. Approximate critical values of the bias-corrected (Z_τ) and ADF test statistics are reported in Table 4.6.5.

The following DGP is employed to model a bivariate cointegrated system with multiple structural breaks,

$$\begin{aligned} y_t &= \mu_t + \beta_t x_t + e_t & & \\ x_t &= x_{t-1} + \omega_t & \omega_t &\sim N(0, \sigma_\omega^2) \\ \Delta e_t &= \rho\, e_{t-1} + \vartheta_t & \vartheta_t &\sim N(0, \sigma_\vartheta^2). \end{aligned} \tag{4.3.17}$$

Using this framework, we study the performance of our estimation procedure and residual-based tests under different breakpoint specifications. In order to evaluate the performance of the modified adaptive lasso estimator in a fixed breakpoint setting, we consider seven potential breakpoint candidates located at the break fractions $\tau = (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875)$. The first specification has one active break at 0.5 where the baseline coefficients are $\mu_1^* = 2$ and $\beta_1^* = 2$ before the break and $\mu_2^* = 4$ and $\beta_2^* = 4$ after the break. The second specification has a first break of both coefficients at 0.25 and a second break at 0.75. The final specification involves three breaks at 0.25, 0.5, and 0.75. The parameter values are increased by 2 at each breakpoint. If the adaptive lasso performs as described by Theorem 1, we should see that the estimates of truly zero coefficients tend to zero very quickly. In our setting, all indicator terms except for the true breakpoints should have zero coefficients. Further, we should observe different convergence rates for the truly nonzero intercept and slope coefficients, where the convergence rate for the intercept coefficients should be slower.

[3]Since in the proof of Theorem 2, which can be found in Section 4.6 of the appendix, we only refer to results for a fixed number of break indicator regressors, we follow Gregory and Hansen (1996a) and expect Theorem 4.2 of Phillips and Ouliaris (1990) to hold. Our simulation results in Section 4.3 seem to support this claim.

The results for increasing sample sizes are summarized in Table 4.3.1. To begin with, we observe quite accurate estimates of the breakpoints for the slope coefficients. The coefficient estimates approach the true parameter values with increasing sample size, while the variance for truly zero coefficients becomes very small at sample sizes of $T = 800$. It appears more difficult to obtain precise estimates of breaks in the intercept than for the slope coefficients. The estimates for nonzero intercept changes on average underestimate the true change which can be attributed to a non-detection of true breakpoints in some replications. Moreover, they have variances which are some magnitudes larger than the variances of slope changes. At this point, it should be emphasized that we present the results for the fixed breakpoint setting without imposing a maximum number of breakpoints. Hence, large variances are mostly driven by adjacent breaks where an initial positive (negative) change is immediately offset by a negative (positive) change to optimize the fit. This behaviour appears less frequently with increasing sample sizes. Since parameter changes are estimated from one regime to the next, estimation errors made in previous regimes accumulate and influence the values of parameter changes at later break dates. Consequently, the variance of parameter changes at later break dates is generally larger than the variance at earlier break dates. Thus, these results mostly fulfill our expectations with respect to the different convergence rates of intercept and slope coefficients.

The results for a diverging number of breakpoint candidates are reported in Table 4.3.2. We consider one break located at $\tau = 0.5$, two breaks at $\tau = (0.33, 0.67)$, and four breaks at $\tau = (0.2, 0.4, 0.6, 0.8)$ to have an equidistant spacing on the unit interval. We first compute the percentages of correct estimation (pce) of m and measure the accuracy of the break date estimation conditional on the correct estimation of m. For this matter, we compute the average Hausdorff distance[4] and divide it by T (hd/T) to compare the values across different sample sizes. We find that the number of breaks is detected with increasing precision and the distance between estimated breakpoints and true breakpoints is becoming smaller for increasing sample sizes. The estimated break fractions are obtained from the second estimation stage involving a misspecified intercept which is falsely assumed to be constant over the sample period. However, the estimates are already very accurate at small sample sizes. Using these second stage results, we are able to specify the post-lasso estimation and obtain consistent estimates for the intercept and slope changes. As expected, the parameter changes of models with fewer breakpoints can be estimated more precisely than those of models with a higher number of breakpoints, as indicated by larger variances of the latter models at all sample sizes.

[4]We define $\Delta(A,B) = \sup_{b \in B} \inf_{a \in A} |a-b|$ for any two sets A and B, then $\max\{\Delta(A,B), \Delta(B,A)\}$ is the Hausdorff distance between A and B.

Table 4.3.1: Estimation results (fixed breakpoint candidates)

This table summarizes the parameter estimates of the model with multiple structural breaks in the intercept and slope coefficient for the case of fixed breakpoint candidates. We use 10,000 replications of the data-generating process given in Equation (4.3.17) with seven breakpoint candidates using equidistant spacing $\tau = (0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875)$. The adjustment coefficient is $\rho = -1$ (i.i.d. error terms), the variance of the error terms is $\sigma_{\omega}^2 = 1$ and $\sigma_{\vartheta}^2 = 1$, respectively. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.25$, $\tau_2 = 0.75$ and the third panel has three active breakpoints at $\tau_1 = 0.25$, $\tau_2 = 0.5$, $\tau_3 = 0.75$. The baseline coefficients and parameter changes at all breakpoints take the value 2. We use initial least squares estimates to compute the adaptive lasso weights. Standard deviations are given in parentheses.

SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2\}$, $(\tau = 0.5)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	2.143 (7.383)	0.004 (0.734)	0.017 (0.763)	0.032 (24.372)	1.744 (27.585)	0.020 (0.733)	0.003 (0.569)	0.007 (0.548)
200	2.136 (2.530)	0.007 (0.469)	0.002 (0.486)	−0.127 (12.809)	1.928 (15.603)	0.019 (0.376)	0.001 (0.390)	−0.005 (0.389)
400	2.101 (1.260)	0.004 (0.391)	0.003 (0.389)	−0.161 (8.530)	1.963 (10.476)	0.008 (0.293)	−0.001 (0.287)	−0.002 (0.297)
800	2.063 (0.904)	0.003 (0.349)	0.000 (0.336)	0.064 (4.385)	1.795 (5.870)	0.006 (0.273)	−0.002 (0.255)	−0.004 (0.255)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.198 (0.393)	0.000 (0.023)	0.003 (0.054)	0.285 (1.255)	1.364 (1.376)	0.005 (0.073)	0.001 (0.027)	0.001 (0.037)
200	2.078 (0.137)	0.000 (0.002)	0.001 (0.015)	0.211 (0.619)	1.645 (0.688)	0.000 (0.010)	0.000 (0.009)	0.000 (0.011)
400	2.050 (0.076)	0.000 (0.001)	0.000 (0.004)	0.075 (0.307)	1.843 (0.345)	0.000 (0.005)	0.000 (0.001)	0.000 (0.003)
800	2.031 (0.046)	0.000 (0.000)	0.000 (0.001)	0.012 (0.101)	1.938 (0.131)	0.000 (0.002)	0.000 (0.000)	0.000 (0.000)

SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2,3\}$, $(\tau_1 = 0.25,\ \tau_2 = 0.75)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	2.049 (16.263)	0.313 (13.717)	1.616 (20.403)	0.010 (0.613)	0.009 (0.622)	−0.009 (21.770)	2.000 (26.376)	0.009 (1.225)
200	1.997 (8.298)	−0.069 (8.413)	2.070 (13.372)	0.011 (0.440)	0.004 (0.455)	−0.295 (13.982)	2.335 (19.690)	−0.010 (1.025)
400	2.126 (3.550)	0.001 (5.305)	1.869 (7.313)	0.007 (0.428)	0.007 (0.441)	−0.151 (9.491)	1.959 (16.354)	0.002 (0.867)
800	2.091 (2.273)	0.032 (1.834)	1.868 (3.702)	−0.001 (0.413)	0.003 (0.438)	0.026 (4.531)	1.877 (12.696)	−0.003 (0.871)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.641 (0.938)	0.111 (0.779)	1.273 (1.193)	0.005 (0.069)	0.004 (0.061)	0.333 (1.132)	1.243 (1.255)	0.005 (0.071)
200	2.281 (0.453)	0.115 (0.447)	1.604 (0.685)	0.000 (0.009)	0.001 (0.013)	0.225 (0.629)	1.589 (0.764)	0.001 (0.027)
400	2.153 (0.211)	0.034 (0.212)	1.830 (0.325)	0.000 (0.001)	0.000 (0.002)	0.081 (0.301)	1.799 (0.430)	0.000 (0.004)
800	2.100 (0.131)	0.004 (0.048)	1.911 (0.154)	0.000 (0.000)	0.000 (0.001)	0.015 (0.096)	1.902 (0.219)	0.000 (0.001)

SB3: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,\dots,4\}$, $(\tau_1 = 0.25,\ \tau_2 = 0.5,\ \tau_3 = 0.75)$

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*	μ_6^*	μ_7^*	μ_8^*
100	1.979 (20.169)	0.286 (11.902)	1.661 (18.975)	0.092 (15.905)	2.029 (21.521)	0.227 (17.692)	2.009 (22.915)	0.024 (1.388)
200	1.956 (9.361)	−0.104 (7.719)	2.181 (14.867)	−0.097 (13.264)	2.142 (18.240)	−0.381 (14.153)	2.550 (20.722)	−0.013 (1.190)
400	2.098 (4.266)	−0.015 (5.490)	1.907 (9.705)	−0.107 (7.940)	2.199 (15.172)	−0.085 (9.055)	2.005 (19.680)	0.006 (1.190)
800	2.080 (2.804)	0.021 (2.011)	1.898 (5.948)	0.041 (3.465)	1.863 (11.065)	0.019 (3.208)	2.159 (18.747)	0.002 (1.337)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*	β_6^*	β_7^*	β_8^*
100	2.710 (1.111)	0.138 (0.704)	1.293 (1.232)	0.349 (1.004)	1.408 (1.279)	0.386 (1.020)	1.289 (1.181)	0.005 (0.068)
200	2.325 (0.503)	0.101 (0.415)	1.620 (0.798)	0.195 (0.615)	1.721 (0.877)	0.215 (0.614)	1.603 (0.832)	0.001 (0.020)
400	2.180 (0.248)	0.032 (0.212)	1.851 (0.429)	0.047 (0.263)	1.912 (0.524)	0.054 (0.266)	1.784 (0.529)	0.000 (0.004)
800	2.123 (0.162)	0.004 (0.053)	1.912 (0.224)	0.005 (0.074)	1.980 (0.262)	0.006 (0.064)	1.870 (0.337)	0.000 (0.001)

Table 4.3.2: Estimation results (diverging number of breakpoint candidates)

This table summarizes the parameter estimates of the model with multiple structural breaks in the intercept and slope coefficient for the case of a diverging number of breakpoint candidates. We use 5,000 replications of the data-generating process given in Equation (4.3.17). The adjustment coefficient is $\rho = -1$ (i.i.d. error terms), the variance of the error terms is $\sigma_{\omega}^2 = 1$ and $\sigma_{\vartheta}^2 = 2$, respectively. Models with a better signal-to-noise ratio yield more precise estimates for all sample sizes. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third panel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. The baseline coefficients and parameter changes at all breakpoints take the value 2. We use the procedure detailed in Subsection 4.2.2 to compute the adaptive lasso weights and apply post-lasso estimation to obtain the estimates for the intercept breaks. Standard deviations are given in parentheses.

SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2\}$, $(\tau = 0.5)$

T	pce	hd/T	τ	μ_1^*	μ_2^*	β_1^*	β_2^*
100	99.8	0.54	0.502 (0.041)	2.04 (1.510)	1.98 (2.270)	2.01 (0.156)	1.99 (0.282)
200	100	0.32	0.500 (0.019)	2.02 (0.774)	1.96 (1.247)	2.00 (0.074)	2.00 (0.106)
400	100	0.19	0.499 (0.008)	2.00 (0.405)	1.98 (0.739)	2.00 (0.035)	2.00 (0.048)
800	100	0.15	0.499 (0.006)	2.00 (0.244)	1.99 (0.503)	2.00 (0.017)	2.00 (0.025)

SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2,3\}$, $(\tau_1 = 0.33,\ \tau_2 = 0.67)$

T	pce	hd/T	τ_1	τ_2	μ_1^*	μ_2^*	μ_3^*	β_1^*	β_2^*	β_3^*
100	98.7	1.26	0.335 (0.054)	0.677 (0.044)	2.07 (2.230)	1.96 (3.470)	1.93 (3.820)	2.03 (0.292)	2.01 (0.500)	1.96 (0.569)
200	99.8	0.80	0.331 (0.038)	0.673 (0.033)	2.03 (1.130)	2.00 (2.250)	1.95 (2.340)	2.01 (0.171)	2.00 (0.201)	1.98 (0.247)
400	99.9	0.48	0.329 (0.020)	0.671 (0.020)	2.02 (0.602)	2.00 (1.009)	2.02 (1.296)	2.00 (0.084)	2.00 (0.121)	1.99 (0.147)
800	99.9	0.37	0.328 (0.009)	0.670 (0.007)	2.01 (0.362)	1.98 (0.733)	2.01 (0.961)	2.00 (0.030)	2.00 (0.044)	2.00 (0.043)

SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,\ldots,5\}$, $(\tau_1 = 0.2,\ \tau_2 = 0.4,\ \tau_3 = 0.6,\ \tau_4 = 0.8)$

T	pce	hd/T	τ_1	τ_2	τ_3	τ_4
100	94.4	2.61	0.208 (0.050)	0.410 (0.060)	0.612 (0.057)	0.809 (0.039)
200	97.2	2.49	0.204 (0.044)	0.407 (0.053)	0.610 (0.053)	0.808 (0.035)
400	98.7	1.67	0.202 (0.040)	0.405 (0.045)	0.609 (0.049)	0.806 (0.033)
800	99.2	1.16	0.202 (0.037)	0.403 (0.036)	0.604 (0.034)	0.803 (0.024)

T	μ_1^*	μ_2^*	μ_3^*	μ_4^*	μ_5^*
100	2.11 (3.72)	2.19 (8.47)	1.81 (8.92)	1.85 (6.40)	2.03 (6.44)
200	2.09 (1.92)	2.03 (2.95)	1.92 (3.51)	1.92 (3.81)	1.96 (3.86)
400	2.09 (1.10)	1.94 (2.80)	1.99 (3.40)	1.95 (3.08)	1.97 (2.66)
800	2.07 (0.91)	2.00 (2.41)	1.92 (2.60)	1.99 (1.74)	1.99 (1.68)

T	β_1^*	β_2^*	β_3^*	β_4^*	β_5^*
100	2.09 (0.507)	2.01 (1.436)	2.03 (1.428)	1.98 (0.733)	1.89 (0.831)
200	2.06 (0.344)	2.02 (0.427)	2.01 (0.581)	1.99 (0.575)	1.90 (0.564)
400	2.04 (0.266)	2.01 (0.312)	2.02 (0.298)	2.01 (0.362)	1.90 (0.526)
800	2.04 (0.233)	2.00 (0.321)	2.00 (0.246)	2.00 (0.184)	1.95 (0.308)

Turning to the cointegration test next, we compare the performance of our test against the performance of well-known cointegration tests with structural breaks described in Gregory and Hansen (1996a), Hatemi-J (2008), and Maki (2012) – henceforth GH-test, HJ-test, and Maki-test.[5] Further, we use the benchmark cointegration test for constant coefficients by Engle and Granger (1987), henceforth EG-test, to evaluate the performance for a DGP without any structural breaks. The results for $T = 100$ and $T = 200$ are reported in Table 4.3.3 and Table 4.3.4, respectively. For the power simulations, we do not set a fixed number of breakpoint candidates in our adaptive lasso framework, but let every observation along T, excluding the lateral trimming, be a potential breakpoint.

Our proposed cointegration test based on the adaptive lasso and the biased-corrected test statistic appears to be slightly oversized for an increasing maximum number of breaks. However, this can also be observed for the Maki-test and the HJ-test. A reason for this pattern might be the small number of observations per regime invalidating asymptotic approximations. The size-adjusted power of our test increases with the sample size and faster speed of adjustment. It is generally low for small adjustment coefficients and always in favor of the ADF test statistic over the bias-corrected test statistic for larger values of the adjustment coefficients. Low power against the cointegration alternative with slow adjustment can be attributed to the inaccuracy of detecting the correct breakpoints if the cointegration residuals are near unit root processes. Choosing the correct maximum number of breaks exerts substantial influence on the power curves, as the power of the proposed test is always some magnitudes higher for the correct model choice. The results thereby show that, in terms of power, it is generally better to select irrelevant breaks than to restrict the number of breaks and to miss a crucial parameter change, which should also be a guiding principle in applied work with our proposed general-to-specific modeling approach.

Our test outperforms the EG-test for specifications with structural breaks. Further, it also performs better than the GH-test for more than one break and better than the HJ-test for more than two breaks if the adjustment is moderate. The Maki-test appears to perform better for a high number of breakpoints and slow adjustment. However, it does not provide a comprehensive modeling framework as it does not attempt to estimate all breakpoints consistently.

[5] The Maki-test with one break at most (M1) is conceptually identical to the GH-test. The only reason to explain the small differences observed in the results is the trimming parameter which is $\xi = 0.05$ for the Maki-test and $\xi = 0.15$ for the GH-test.

Table 4.3.3: Size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 100$)

This table shows the size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 100$). $ADF_{\tau,1}$, $ADF_{\tau,2}$, and $ADF_{\tau,4}$ denote the cointegration test based on the ADF regression with one, two, and four possible breakpoints, respectively. Correspondingly, $Z_{\tau,1}$, $Z_{\tau,2}$, and $Z_{\tau,4}$ denote the cointegration test based on the bias-corrected test statistic. GH denote the cointegration test with one structural break (model C/S) by Gregory and Hansen (1996a) and HJ denotes the corresponding test with two structural breaks by Hatemi-J (2008). $M1$, $M2$, and $M4$ denote the cointegration test by Maki (2012) allowing for one, two, and four structural breaks, respectively. EG denotes the Engle-Granger cointegration test with constant coefficients. The results are obtained from 2,500 replications. All tests are conducted at the 5% significance level.

ρ	$ADF_{\tau,1}$	$ADF_{\tau,2}$	$ADF_{\tau,4}$	$Z_{\tau,1}$	$Z_{\tau,2}$	$Z_{\tau,4}$	GH	HJ	$M1$	$M2$	$M4$	EG
	SB0: $\mu = 2$, $\beta = 2$											
Size:	0.061	0.059	0.053	0.076	0.074	0.083	0.076	0.099	0.054	0.068	0.097	0.041
−0.05	0.102	0.115	0.120	0.080	0.083	0.060	0.069	0.056	0.068	0.072	0.062	0.099
−0.10	0.174	0.165	0.134	0.136	0.114	0.066	0.106	0.078	0.106	0.106	0.080	0.226
−0.25	0.604	0.388	0.173	0.533	0.304	0.101	0.496	0.274	0.492	0.370	0.218	0.885
−0.50	1.000	0.983	0.740	0.998	0.943	0.496	0.997	0.933	0.998	0.968	0.760	1.000
	SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2\}$, $(\tau = 0.5)$											
−0.05	0.095	0.102	0.133	0.073	0.066	0.071	0.663	0.695	0.683	0.571	0.376	0.042
−0.10	0.157	0.176	0.160	0.123	0.106	0.083	0.753	0.746	0.766	0.644	0.450	0.048
−0.25	0.642	0.545	0.353	0.564	0.427	0.197	0.934	0.879	0.935	0.829	0.638	0.066
−0.50	0.995	0.950	0.829	0.992	0.941	0.707	1.000	0.994	1.000	0.995	0.924	0.086
	SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2,3\}$, $(\tau_1 = 0.33,\ \tau_2 = 0.67)$											
−0.05	0.056	0.107	0.143	0.044	0.065	0.061	0.225	0.815	0.256	0.606	0.370	0.043
−0.10	0.057	0.154	0.181	0.044	0.097	0.083	0.237	0.852	0.272	0.668	0.420	0.045
−0.25	0.080	0.505	0.424	0.064	0.396	0.239	0.283	0.950	0.320	0.827	0.580	0.047
−0.50	0.122	0.966	0.884	0.097	0.955	0.784	0.334	0.999	0.372	0.976	0.888	0.054
	SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,\ldots,5\}$, $(\tau_1 = 0.2,\ \tau_2 = 0.4,\ \tau_3 = 0.6\ \tau_4 = 0.8)$											
−0.05	0.028	0.044	0.154	0.021	0.027	0.056	0.089	0.257	0.100	0.108	0.403	0.039
−0.10	0.033	0.047	0.190	0.022	0.032	0.087	0.085	0.263	0.102	0.118	0.432	0.038
−0.25	0.031	0.054	0.431	0.019	0.037	0.249	0.091	0.284	0.107	0.122	0.542	0.038
−0.50	0.032	0.064	0.902	0.022	0.046	0.814	0.100	0.317	0.113	0.148	0.773	0.039

Table 4.3.4: Size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 200$)

This table shows the size and size-adjusted power of cointegration tests in the presence of multiple regime shifts ($T = 200$). $ADF_{\tau,1}$, $ADF_{\tau,2}$, and $ADF_{\tau,4}$ denote the cointegration test based on the ADF regression with one, two, and four possible breakpoints, respectively. Correspondingly, $Z_{\tau,1}$, $Z_{\tau,2}$, and $Z_{\tau,4}$ denote the cointegration test based on the bias-corrected test statistic. *GH* denote the cointegration test with one structural break (model C/S) by Gregory and Hansen (1996a) and *HJ* denotes the corresponding test with two structural breaks by Hatemi-J (2008). $M1$, $M2$, and $M4$ denote the cointegration test by Maki (2012) allowing for one, two, and four structural breaks, respectively. *EG* denotes the Engle-Granger cointegration test with constant coefficients. The results are obtained from 2,500 replications. All tests are conducted at the 5% significance level.

ρ	$ADF_{\tau,1}$	$ADF_{\tau,2}$	$ADF_{\tau,4}$	$Z_{\tau,1}$	$Z_{\tau,2}$	$Z_{\tau,4}$	GH	HJ	$M1$	$M2$	$M4$	EG
	SB0: $\mu = 2$, $\beta = 2$											
Size:	0.062	0.054	0.051	0.066	0.067	0.074	0.062	0.080	0.045	0.056	0.060	0.040
−0.05	0.162	0.152	0.131	0.140	0.104	0.079	0.102	0.064	0.090	0.092	0.083	0.235
−0.10	0.406	0.296	0.180	0.365	0.231	0.124	0.312	0.154	0.285	0.222	0.167	0.705
−0.25	1.000	0.952	0.557	0.998	0.909	0.386	0.993	0.872	0.990	0.930	0.730	1.000
−0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SB1: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2\}$, $(\tau = 0.5)$											
−0.05	0.148	0.157	0.158	0.128	0.100	0.090	0.707	0.736	0.722	0.598	0.450	0.069
−0.10	0.416	0.358	0.247	0.374	0.278	0.159	0.874	0.848	0.870	0.738	0.588	0.087
−0.25	0.996	0.972	0.784	0.995	0.956	0.682	1.000	0.997	1.000	0.990	0.914	0.121
−0.50	0.998	0.974	0.930	0.998	0.973	0.930	1.000	1.000	1.000	1.000	1.000	0.138
	SB2: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,2,3\}$, $(\tau_1 = 0.33,\ \tau_2 = 0.67)$											
−0.05	0.070	0.138	0.156	0.060	0.092	0.084	0.296	0.830	0.339	0.643	0.436	0.055
−0.10	0.090	0.320	0.291	0.077	0.240	0.187	0.352	0.916	0.388	0.761	0.547	0.058
−0.25	0.134	0.966	0.865	0.114	0.946	0.780	0.418	0.998	0.460	0.973	0.873	0.066
−0.50	0.168	0.987	0.937	0.138	0.986	0.936	0.456	1.000	0.505	0.992	1.000	0.069
	SB4: $\mu_k^* = 2$, $\beta_k^* = 2$, $k = \{1,\ldots,5\}$, $(\tau_1 = 0.2,\ \tau_2 = 0.4,\ \tau_3 = 0.6\ \tau_4 = 0.8)$											
−0.05	0.038	0.042	0.166	0.030	0.028	0.098	0.104	0.284	0.112	0.143	0.462	0.048
−0.10	0.038	0.049	0.297	0.030	0.035	0.189	0.109	0.302	0.119	0.152	0.545	0.050
−0.25	0.043	0.061	0.866	0.030	0.043	0.769	0.119	0.346	0.131	0.173	0.796	0.052
−0.50	0.044	0.075	0.960	0.033	0.049	0.948	0.132	0.382	0.149	0.214	0.948	0.055

4.4 Empirical application

This section considers an application of our proposed framework in the context of the long-run PPP between the US and the UK with more than a century of data and several potential regime shifts. PPP is a well-known theory in macroeconomics which postulates that the nominal exchange rate between two currencies should be equal to the ratio of the domestic to the foreign price level. Extensive studies have been conducted to investigate whether the proposition of PPP holds over the long-run (for a comprehensive review, see Taylor and Taylor, 2004). While unit root tests provide one means to test for the existence of strong-form PPP, one strand of the literature investigates the potentially cointegrated relationship between the nominal exchange rate and the price ratio, in order to test for weak-form PPP (see, for example, Corbae and Ouliaris, 1988; Taylor, 1988; Kim, 1990). However, empirical evidence on the subject still remains contradictory. Particularly the examined data span seems to have a substantial influence on the outcomes of empirical studies (Taylor, 2006; Karoglou and Morley, 2012). The diverging findings could on one hand be rooted in the generally slow adjustment of the real exchange rate which necessitates long samples for robust results. On the other hand, these long samples might be composed of regimes shaped by very different macroeconomic environments which have to be accounted for. For example, the US and the UK moved between several fixed and floating exchange rate regimes in the last century.

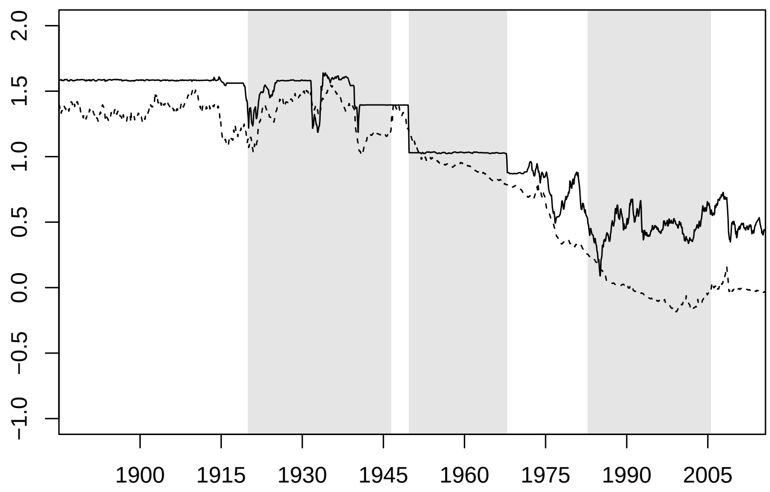
We thus consider the potentially cointegrated relationship between the nominal exchange rate and price ratio, both of which are observed monthly between 1885 and 2015, a suitable application to illustrate our methodology with an empirical example. The dataset is an extended version of the data collected in Grilli and Kaminsky (1991) and Engel and Kim (1999). It is unique in the sense that such a comparatively high sampling frequency is not available for other country pairs over such a long sampling period. The logarithm of the nominal USD/GBP exchange rate is denoted by ex_t, while p_t is the log US price level and p_t^* is the log UK price level, respectively. We estimate the regression,

$$ex_t = \alpha + \beta(p_t - p_t^*) + u_t, \tag{4.4.18}$$

where ex_t, p_t, and p_t^* are generated by integrated processes of order one. Under PPP, we would expect to observe a mean-zero stationary error term u_t so that nominal exchange rate and relative prices are cointegrated. Strong-from PPP, assuming strict proportionality, would be given by the restriction $\alpha = 0$ and $\beta = 1$. Since we can only speculate which macroeconomic events might have changed the long-run equilibrium relationship, we apply our adaptive lasso procedure without the pre-specification of breakpoint candidates. The maximum number of breaks is chosen to be six, which is a reasonable compromise between flexibility and average regime length. It also corresponds to the number of changes from fixed to floating exchange rates and vice versa in our sample. A detailed description of US/UK exchange rate regimes is provided in Craighead (2010). The minimum length of a regime was chosen to be one year.

First, we assume constancy of the parameters and ignore potential structural breaks. Estimation of the long-run equilibrium coefficients yields OLS estimates $\hat{\alpha} = 0.46$ and $\hat{\beta} = 0.77$. The Engle-Granger test based on an ADF regression rejects the null hypothesis at the 1% level. Similar results can be obtained for the Phillips-Ouliaris test. The adjustment is slow ($\hat{\rho} = -0.016$) with a half life period of disequilibrium states of more than 3.5 years. Next, we compare several previously mentioned structural break models with our model selection approach. The GH-test finds evidence for cointegration with a breakpoint at 1949 m02, while the HJ-test indicates breakpoints at 1949 m02 and 1982 m01. The Maki-test selects the breakpoints 1919 m07, 1949 m08, 1967 m11, 1978 m07, 1987 m03 and rejects the null hypothesis of no cointegration as well. Our new general-to-specific procedure yields break dates 1919 m12, 1946 m07, 1949 m09, 1967 m12, 1982 m11, 2005 m09 and rejects the null hypothesis of no cointegration at the 1% (5%) level for the ADF (bias-corrected) test statistic.

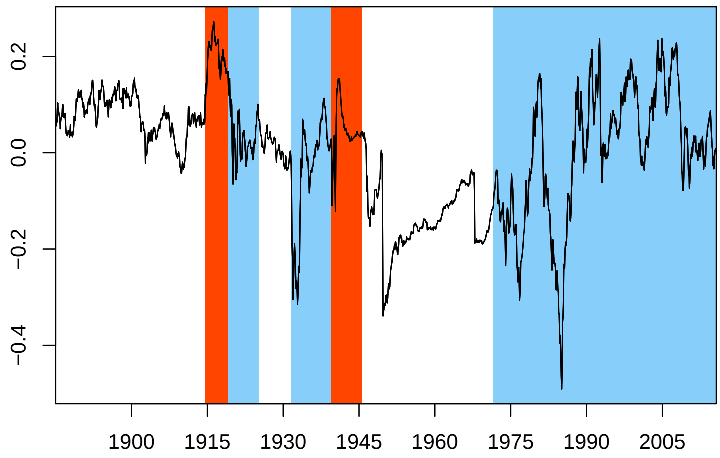
Figure 4.4.1: Nominal exchange rate and relative price levels time series
This plot shows the nominal exchange rate (solid) and relative price levels (dashed). The shaded areas correspond to the regimes identified by the adaptive lasso procedure.



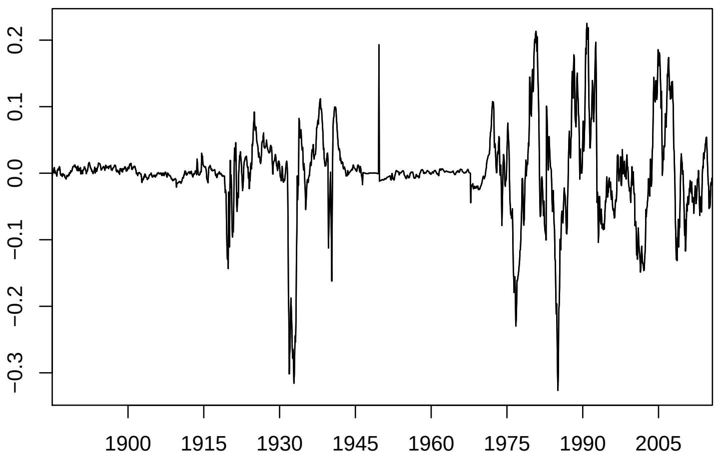
The indicated break dates (see Figure 4.4.1) can all be related to economic events affecting the real exchange rate. The first regime extends to the end of the First World War and spans the classical gold standard period with a fixed price of a pound sterling at USD 4.87. The next regime extends to the end of the Second World War and comprises fixed and floating exchange rate regimes in the inter-war period. The third breakpoint is found in September 1949 which coincides with a devaluation of pound sterling by roughly 30% and is followed by another breakpoint after Britain devalued the pound in November 1967. After the Bretton Woods system ended, we find two more breakpoints where the first one can be associated with a deep recession in the UK and the second one slightly pre-dates the financial crisis. We have to emphasize that the break dates might be affected by the usual lead and lag effects, since the parameter changes are representative for the following regime.

Figure 4.4.2 displays the residual plots for (a) the OLS regression and (b) the post-lasso regression. We can clearly see from the lower panel of Figure 4.4.2 that the post-lasso residuals are characterized by high and low volatility periods. The high volatility periods match with the floating exchange rate periods, which supports the hypothesis formulated by Mussa (1986), who argues that the nominal exchange rate system is a major determinant of the real exchange rate volatility. By contrast, for the OLS residuals, we do not find a clear pattern and in general observe trending behavior. Accounting for the different exchange rate regimes yields a much faster adjustment ($\hat{\rho} = -0.062$) and a half life of disequilibrium states of less than a year. Our findings reveal that long-run PPP can only be properly assessed if we model structural instabilities induced by policy decisions, which lead to substantial moves in the exchange rate.

Figure 4.4.2: OLS residuals and post-lasso residuals
These plots depict (a) the OLS residual series and (b) post-lasso residual series. Fixed exchange rate regimes are marked by white areas, floating exchange rate regimes are marked by light-blue areas and both World Wars are marked by orange-red areas.



(a) OLS residual series



(b) Post-lasso residual series

4.5 Conclusion

In this article, we propose a new general-to-specific method for the accurate detection of an unknown number of structural breaks in cointegrating regressions. Furthermore, we design a new test for cointegration under the presence of multiple structural breaks based on our proposed adaptive lasso estimator. Our main goal is to build a comprehensive modeling framework which does not only focus on improving the performance of a specific test for cointegration under parameter instability but also provides consistent estimates of the corresponding parameter changes. Our procedure for a diverging number of breakpoint candidates imposes very few a priori restrictions on the timing of the breakpoints but only requires to specify a maximum number of breaks. This promises a high degree of flexibility for applied economists in finding the right model specification.

Although some authors have already used lasso estimators in a cointegration framework, our specific application of the adaptive lasso to cointegrating regressions with multiple structural breaks has not been discussed in the literature before. Thus, we first prove the suitability of the adaptive lasso estimator in our context by showing its nonstandard oracle property for a fixed number of breakpoint candidates. Subsequently, we provide extensive Monte Carlo evidence that our framework can be extended to a diverging number of breakpoint candidates. Our results show an accurate and consistent estimation of the parameter changes for different choices for the maximum number of possible breakpoints.

The analysis of this paper is confined to a bivariate cointegration model. A generalization to multivariate cointegration models is straightforward in situations where fixed breakpoint candidates are available and $d^* \ll T$. However, the same does not hold for a diverging number of breakpoint candidates which would require a different penalty for nonzero parameters reducing the overall number of penalized parameters. A suitable solution could be adapting the group-fused lasso estimator applied in Chan *et al.* (2014) and Qian and Su (2016) to nonstationary regressions. Their estimator forces all slope coefficients to have breaks at a common break date.

Our residual-based cointegration tests appear to perform well in terms of power, especially for moderate adjustment. They outperform standard tests and are only inferior to the Maki-test if the adjustment is very slow. It seems that the adaptive lasso estimator needs a strong signal to find true breakpoints which directly determines the power of the cointegration tests. It would be interesting to develop an analogous testing framework which reverses the null hypothesis and alternative similar to the tests described in Carrion-i Silvestre and Sanso (2006) and Arai and Kurozumi (2007).

These findings show a promising new direction for cointegration model specification and cointegration testing in the presence of multiple structural breaks. While the proposed estimation and testing procedures provide the possibility of several extensions, they also constitute a solid benchmark for other general-to-specific approaches dealing with structural change in cointegration models.

4.6 Appendix

Proof of Theorem 1

According to Assumption 1, the scalar partial sum process in Equation (4.2.2) satisfies the functional central limit theorem (FCLT). For $s \in [0,1]$ and as $T \to \infty$, it holds that

$$x_{[Ts]} = T^{-1/2} \sum_{t=1}^{[Ts]} v_t \Rightarrow B(s), \tag{4.6.19}$$

where $B(s)$ is a Brownian motion process with variance σ^2. This is shown by Herrndorf (1984) and extended to the vector case by Phillips and Durlauf (1986).

Next, we define the objective function $V_T(\boldsymbol{b})$ by

$$\begin{aligned} V_T(\boldsymbol{b}) &= \sum_{t=1}^{T} \left[(u_t - \boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1})^2 - u_t^2 \right] \\ &+ \lambda_T \sum_{i=2}^{p^*+1} w_{1i}^{\gamma_1} |\mu_i^* + b_{1i}/\sqrt{T}| + \lambda_T \sum_{j=2}^{m^*+1} w_{2j}^{\gamma_2} |\beta_j^* + b_{2j}/T|, \end{aligned} \tag{4.6.20}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1', \boldsymbol{b}_2')'$, $\delta_T = diag(T^{1/2}\boldsymbol{I}_{p^*+1}, T\boldsymbol{I}_{m^*+1})$ and

$$\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_1', \hat{\boldsymbol{b}}_2')' = \arg\min V_T(\boldsymbol{b}) \tag{4.6.21}$$

is the minimizer of V_T with $\hat{b}_{1i} = \sqrt{T}(\hat{\mu}_{T,i} - \mu_i^*)$ and $\hat{b}_{2j} = T(\hat{\beta}_{T,j} - \beta_j^*)$.

First, we consider the asymptotic counterparts to the least squares terms

$$-2\sum_{t=1}^{T} u_t \boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1} + \sum_{t=1}^{T} \boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1}\delta_T^{-1}\boldsymbol{X}_t'\boldsymbol{b}.$$

We use the decomposition $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ to express the weak convergence result

$$T^{-1/2}\boldsymbol{I}_{m^*+1}\boldsymbol{X}_{2,[Ts]} \Rightarrow (B(s), B(s)\varphi_{\tau_{2,1}}(s), \ldots, B(s)\varphi_{\tau_{2,m^*}}(s)) = B_\tau(s), \tag{4.6.22}$$

where

$$\varphi_{\tau_{k,l}}(s) \begin{cases} 0 & \text{if } s < \tau_{k,l} \\ 1 & \text{if } s \geq \tau_{k,l} \end{cases}, \qquad k \in \{1,2\}, \quad s \in [0,1]. \tag{4.6.23}$$

Using (A.4) in Gregory and Hansen (1996a) and the continuous mapping theorem (CMT, see Billingsley, 1999), Theorem 2.7), we observe that

$$\sum_{t=1}^{T} \boldsymbol{b}'\boldsymbol{X}_t\delta_T^{-1}\delta_T^{-1}\boldsymbol{X}_t'\boldsymbol{b} \Rightarrow \boldsymbol{b}' \begin{bmatrix} \Upsilon & 0 \\ 0 & \int_0^1 B_\tau(s)B_\tau(s)'ds \end{bmatrix} \boldsymbol{b},$$

where the weak convergence is uniform over the vector $(\tau_{1,1}, \ldots, \tau_{1,p^*}, \tau_{2,1}, \ldots, \tau_{1,m^*}) \in \mathscr{T}$. Further, using (A.3) in Gregory and Hansen (1996a) and Theorem 3.1 in Hansen (1992), we have the weak convergence to a stochastic integral

$$\sum_{t=1}^{T} u_t \boldsymbol{b}' \boldsymbol{X}_t \delta_T^{-1} \Rightarrow \boldsymbol{b}' \begin{bmatrix} U \\ \int_0^1 B_\tau(s) dU(s) \end{bmatrix} + \boldsymbol{b}' \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Lambda \\ (1-\tau_{2,1})\Lambda \\ \vdots \\ (1-\tau_{2,m^*})\Lambda \end{bmatrix},$$

where $U \sim N(0, \Upsilon\sigma^2)$, $\Lambda = \sum_{t=0}^{\infty} E(v_t u_0)$ and $U(s)$ is the weak limit of u_t.

Under Assumption 2 and Assumption 3, the maximum number of breakpoints is limited and initial least squares estimates are available for the weights of the adaptive lasso estimator. We investigate the consistency of the individual coefficients and distinguish between the true coefficients being zero or nonzero:

(a) If $\mu_i^* \neq 0$, we have

$$\begin{aligned} & \lambda_T w_{1i}^{\gamma_1} \left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*| \right] \\ = \quad & \frac{\lambda_T}{\sqrt{T}} \left| \frac{1}{\hat{\mu}_{I,i}} \right|^{\gamma_1} \sqrt{T} \left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*| \right] \\ \xrightarrow{p} \quad & 0, \end{aligned} \tag{4.6.24}$$

since (i) $\frac{\lambda_T}{\sqrt{T}} \to 0$, (ii) $\left|\frac{1}{\hat{\mu}_{I,i}}\right|^{\gamma_1} \xrightarrow{p} \left|\frac{1}{\mu_i^*}\right|^{\gamma_1}$, if the initial estimator is consistent, and (iii) $\sqrt{T}\left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*|\right] \xrightarrow{p} b_{1i}\operatorname{sgn}(\mu_i^*)$ as in Zou (2006).

(b) If $\mu_i^* = 0$, we have

$$\begin{aligned} \lambda_T w_{1i}^{\gamma_1} \left[|\mu_i^* + b_{1i}/\sqrt{T}| - |\mu_i^*| \right] \quad & = \quad \frac{\lambda_T}{\sqrt{T}} \left| \frac{1}{\hat{\mu}_{I,i}} \right|^{\gamma_1} |b_{1i}| \\ & = \quad \frac{\lambda_T}{T^{1/2-\gamma_1/2}} \left| \frac{1}{\sqrt{T}\hat{\mu}_{I,i}} \right|^{\gamma_1} |b_{1i}| \\ & \Rightarrow \quad \begin{cases} \infty & \text{if} \quad b_{1i} \neq 0 \\ 0 & \text{if} \quad b_{1i} = 0 \end{cases}, \end{aligned} \tag{4.6.25}$$

since (i) $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and (ii) the initial least squares estimator is tight and converges to a normal distribution, $\sqrt{T}\hat{\mu}_{I,i} \Rightarrow W_{1i} \sim N(0, \frac{\sigma^2}{\tau_{1,i}(1-\tau_{1,i})})$.

(c) If $\beta_j^* \neq 0$, we have

$$\begin{aligned} & \lambda_T w_{2j}^{\gamma_2} \left[|\beta_j^* + b_{2j}/T| - |\beta_j^*| \right] \\ = \quad & \frac{\lambda_T}{T} \left| \frac{1}{\hat{\beta}_{I,j}} \right|^{\gamma_2} T \left[|\beta_j^* + b_{2j}/T| - |\beta_j^*| \right] \\ \xrightarrow{p} \quad & 0, \end{aligned} \tag{4.6.26}$$

since (i) $\frac{\lambda_T}{T} \to 0$, (ii) $\left|\frac{1}{\hat{\beta}_{I,j}}\right|^{\gamma_2} \xrightarrow{p} \left|\frac{1}{\beta_j^*}\right|^{\gamma_2}$, if the initial estimator is consistent, and (iii) $T\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] \xrightarrow{p} b_{2j}\,\mathrm{sgn}(\beta_j^*)$.

(d) If $\beta_j^* = 0$, we have

$$\begin{aligned}
\lambda_T w_{2j}^{\gamma_2}\left[|\beta_j^* + b_{2j}/T| - |\beta_j^*|\right] &= \frac{\lambda_T}{T}\left|\frac{1}{\hat{\beta}_{I,j}}\right|^{\gamma_2} |b_{2j}| \\
&= \frac{\lambda_T}{T^{1-\gamma_2}}\left|\frac{1}{T\hat{\beta}_{I,j}}\right|^{\gamma_2} |b_{2j}| \\
&\Rightarrow \begin{cases} \infty & \text{if} \quad b_{2j} \neq 0 \\ 0 & \text{if} \quad b_{2j} = 0 \end{cases},
\end{aligned} \tag{4.6.27}$$

since (i) $\frac{\lambda_T}{T^{1-\gamma_2}} \to \infty$, (ii) the least squares estimator is tight and has the following nonstandard distribution

$$T\hat{\beta}_{I,j} \Rightarrow W_{2j} = \frac{\left(\int_0^1 B_{\tau_{2,j}}(s)dU(s) + (1-\tau_{2,j})\Lambda\right)}{\int_0^1 B_{\tau_{2,j}}^2(s)ds}, \tag{4.6.28}$$

and (iii) $P(W_{2j} = 0) \overset{a.s.}{\to} 0$.

Thus, $V_T(\boldsymbol{b}) \Rightarrow V(\boldsymbol{b})$, where

$$V(\boldsymbol{b}) = \begin{cases} \boldsymbol{b}'A\boldsymbol{b} - 2\boldsymbol{b}'B - 2\boldsymbol{b}'C & \text{if} \quad b_k = 0 \text{ for all } k \in \mathscr{A}^c \\ \infty & \text{if} \quad b_k \neq 0 \text{ for some } k \in \mathscr{A}^c \end{cases} \tag{4.6.29}$$

with

$$A = \begin{bmatrix} \Upsilon & 0 \\ 0 & \int_0^1 B_\tau(s)B_\tau(s)'ds \end{bmatrix},$$

$$B = \begin{bmatrix} U \\ \int_0^1 B_\tau(s)dU(s) \end{bmatrix}, \qquad U \sim N(0, \Upsilon\sigma^2),$$

$$C = [0, \ldots, 0, \Lambda, (1-\tau_{2,1})\Lambda, \ldots, (1-\tau_{2,m^*})\Lambda]'.$$

Since V_T is a convex function and V has a unique minimum, it follows from Knight and Fu (2000) that

$$\arg\min V_T(\boldsymbol{b}) = \hat{\boldsymbol{b}} = \begin{bmatrix} \sqrt{T}(\hat{\mu}_T - \mu^*) \\ T(\hat{\beta}_T - \beta^*) \end{bmatrix} \Rightarrow \arg\min V(\boldsymbol{b}). \tag{4.6.30}$$

From these results, we can deduce that

$$\begin{aligned}
\sqrt{T}(\hat{\mu}_{T,\mathscr{A}_1^c} - \mu^*_{\mathscr{A}_1^c}) &\Rightarrow \delta_0^{|\mathscr{A}_1^c|} \\
\sqrt{T}(\hat{\mu}_{T,\mathscr{A}_1} - \mu^*_{\mathscr{A}_1}) &\Rightarrow N(0, [\Sigma_{\mathscr{A}_1}]^{-1}\Upsilon_{\mathscr{A}_1}\sigma^2),
\end{aligned} \tag{4.6.31}$$

where δ_0 denotes the Dirac measure at 0. Correspondingly, we have

$$\begin{aligned} T(\hat{\beta}_{T,\mathscr{A}_2^c} - \beta^*_{\mathscr{A}_2^c}) &\Rightarrow \delta_0^{|\mathscr{A}_2^c|} \\ T(\hat{\beta}_{T,\mathscr{A}_2} - \beta^*_{\mathscr{A}_2}) &\Rightarrow \left[\int_0^1 B_{\tau,\mathscr{A}_2} B'_{\tau,\mathscr{A}_2}\right]^{-1} \left[\int_0^1 B_{\tau,\mathscr{A}_2} dU + C^*_{\mathscr{A}_2}\right]. \end{aligned} \tag{4.6.32}$$

It remains to show that coefficients of inactive variables are set to zero with probability approaching one. We begin with a proof of $P(\hat{\mu}_{T,\mathscr{A}_1^c} = 0) \to 1$. Consider the event that $\hat{\mu}_{T,i} \neq 0$ although $i \in \mathscr{A}_1^c$. We know from the Karush-Kuhn-Tucker (KKT) optimality conditions that the first order condition for a minimum is given by

$$\frac{2\varphi'_{\tau_{1,i}}(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{\sqrt{T}} = \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}}. \tag{4.6.33}$$

Note that

$$\left| \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}} \right| = \frac{\lambda_T}{T^{1/2-\gamma_1/2}} \left| \frac{1}{\sqrt{T}\hat{\mu}_{I,i}} \right|^{\gamma_1} \to \infty, \tag{4.6.34}$$

since (i) $\frac{\lambda_T}{T^{1/2-\gamma_1/2}} \to \infty$ and (ii) $\sqrt{T}\hat{\mu}_{I,i}$ is tight. The left hand side of the equation is equivalent to

$$\frac{2\varphi'_{\tau_{1,i}}(u - x\delta_T^{-1}\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})')}{\sqrt{T}} = \frac{2\varphi'_{\tau_{1,i}} u}{\sqrt{T}} - \frac{2\varphi'_{\tau_{1,i}} x\delta_T^{-1}\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})'}{\sqrt{T}}. \tag{4.6.35}$$

For the first term, we have the weak convergence

$$\frac{\varphi'_{\tau_{1,i}} u}{\sqrt{T}} \Rightarrow N(0, \sigma^2 \tau_{1,i}) \tag{4.6.36}$$

and for the second term, we have the weak convergence of $\frac{\varphi'_{\tau_{1,i}} x\delta_T^{-1}}{\sqrt{T}}$, which depends on the timing of the break fraction $\tau_{1,i}$ relative to all other possible break fractions. Say $\tau_{1,i} = \tau_{1,p^*} > \tau_{2,m^*}$ holds, then

$$\frac{\varphi'_{\tau_{1,i}} x\delta_T^{-1}}{\sqrt{T}} \Rightarrow \left(0, \ldots, 0, \int_0^1 B_{\tau_{2,1}}(s)ds, \ldots, \int_0^1 B_{\tau_{2,m^*}}(s)ds\right). \tag{4.6.37}$$

Further, we have already shown the weak convergence of $\delta_T(\hat{\mu}'_T - \mu^{*\prime}, \hat{\beta}'_T - \beta^{*\prime})'$. Hence, the distribution of the first term is tight and

$$P(\hat{\mu}_{T,i} \neq 0) \leq P\left(\frac{2\varphi'_{\tau_{1,i}}(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{\sqrt{T}} - \frac{\lambda_T w_{1i}^{\gamma_1} \operatorname{sgn}(\hat{\mu}_{T,i})}{\sqrt{T}} = 0\right) \to 0. \tag{4.6.38}$$

Next, we show that $P(\hat{\beta}_{T,\mathscr{A}_2^c} = 0) \to 1$. Again, we consider the event that $\hat{\beta}_{T,j} \neq 0$ although $j \in \mathscr{A}_2^c$. The KKT optimality condition in this case is given by

$$\frac{2(x\varphi_{\tau_{2,j}})'(y - x(\hat{\mu}'_T, \hat{\beta}'_T)')}{T} = \frac{\lambda_T w_{2j}^{\gamma_2} \operatorname{sgn}(\hat{\beta}_{T,j})}{T}, \tag{4.6.39}$$

where the factor T substitutes the factor $\sqrt{T}$ in Equation (4.6.33). For the right hand side of the equation, we observe that

$$\left|\frac{\lambda_T w_{2j}^{\gamma_2}\,\mathrm{sgn}(\hat{\beta}_{T,j})}{T}\right| = \frac{\lambda_T}{T^{1-\gamma_2}}\left|\frac{1}{T\hat{\beta}_{I,j}}\right|^{\gamma_2} \to \infty, \tag{4.6.40}$$

since $T\hat{\beta}_{I,j}$ is tight. For the left hand side,

$$\frac{2(x\varphi_{\tau_{2,j}})'u}{T} - \frac{2(x\varphi_{\tau_{2,j}})'x\delta_T^{-1}\delta_T(\hat{\mu}_T' - \mu^{*\prime}, \hat{\beta}_T' - \beta^{*\prime})'}{T}, \tag{4.6.41}$$

we have the weak convergence of the first term using

$$\frac{(x\varphi_{\tau_{2,j}})'u}{\sqrt{T}} \Rightarrow \int_0^1 B_{\tau_{2,j}}(s)dU(s) + (1-\tau_{2,j})\Lambda. \tag{4.6.42}$$

The expression of the weak convergence result for the second term depends on the timing of the break fraction $\tau_{2,j}$. Say $\tau_{2,j} = \tau_{1,m^*} > \tau_{1,p^*}$ holds, we have

$$\begin{aligned}\frac{(x\varphi_{\tau_{2,j}})'x\delta_T^{-1}}{T} \Rightarrow \quad & \Bigg(\int_0^1 B_{\tau_{2,j}}(s)ds, \int_0^1 B_{\tau_{1,1}}(s)ds, \ldots, \int_0^1 B_{\tau_{2,p^*}}(s)ds, \\ & \int_0^1 B_{\tau_{2,j}}^2(s)ds, \int_0^1 B_{\tau_{1,1}}^2(s)ds, \ldots, \int_0^1 B_{\tau_{2,m^*}}^2(s)ds\Bigg),\end{aligned} \tag{4.6.43}$$

and as before $\delta_T(\hat{\mu}_T' - \mu^{*\prime}, \hat{\beta}_T' - \beta^{*\prime})'$ is tight. Finally, we have shown that

$$P(\hat{\beta}_{T,j} \neq 0) \leq P\left(\frac{2(x\varphi_{\tau_{2,j}})'(y - x(\hat{\mu}_T', \hat{\beta}_T')')}{T} - \frac{\lambda_T w_{2j}^{\gamma_2}\,\mathrm{sgn}(\hat{\beta}_{T,j})}{T} = 0\right) \to 0 \tag{4.6.44}$$

and this completes the proof. □

Proof of Theorem 2

For ease of exposition, we assume that the maximum number of breaks is $m^* = 2$ and the true intercept is known to be $\mu_t = 0$ for all t. In this case, we obtain three possible model selection outcomes regarding the number of breaks for the post-lasso regressions:

1. All coefficients of break indicator regressors are shrunk to zero

$$y_t = \beta_1 x_t + e_{t\tau_0}. \tag{4.6.45}$$

2. One structural break is (falsely) detected

$$y_t = \beta_1 x_t + \beta_2 x_t \varphi_{t,\tau_{2,1}} + e_{t\tau_1}. \tag{4.6.46}$$

3. Two structural breaks are (falsely) detected

$$y_t = \beta_1 x_t + \beta_2 x_t \varphi_{t,\tau_{2,1}} + \beta_3 x_t \varphi_{t,\tau_{2,2}} + e_{t\tau_2}. \tag{4.6.47}$$

Note that the specification of indicator terms in Equation (4.6.46) and Equation (4.6.47), i.e., the timing of (falsely) detected breaks, depends on the tuning parameter λ. We continue the proof for the bias-corrected test statistic, Z_2, corresponding to the case of two (falsely) detected structural breaks with unknown timing. The asymptotic distribution of the test statistics for the two remaining cases can be easily deduced from our derivations. We decompose the cumulative sum into $S_t = (S_{1t}, S_{2t})'$. Further, we define the break fraction vector $\tau_2 = (\tau_{2,1}, \tau_{2,2})'$ as a compact set on $(0,1) \times (0,1)$ and define the matrix $X_{t\tau_2} = (S_t', S_{2t}\varphi_{t,\tau_{2,1}}, S_{2t}\varphi_{t,\tau_{2,2}})' = (S_{1t}, X_{2t\tau_2}')'$.

Using the result

$$T^{-1/2} S_{[\tau_{2,i}T]} \Rightarrow B(\tau_{2,i}) \tag{4.6.48}$$

and the CMT yields the weak convergence of

$$\frac{1}{T^2} \sum_{t=[\tau_{2,i}T]}^{T} S_t S_t' = \int_{\tau_{2,i}}^{1} BB'. \tag{4.6.49}$$

Further, (4.6.48) and Theorem 4.1 of Hansen (1992) yield the weak convergence of

$$\frac{1}{T} \sum_{t=[\tau_{2,i}T]}^{T} S_{t-1} u_t' = \int_{\tau_{2,i}}^{1} BdB' + (1 - \tau_{2,i})\Lambda. \tag{4.6.50}$$

The result in (4.6.49) can straightforwardly be extended to

$$\frac{1}{T^2} \sum_{t=1}^{T} X_{t\tau_2} X_{t\tau_2}' = \int_{0}^{1} X_{\tau_2} X_{\tau_2}', \tag{4.6.51}$$

where $X_{\tau_2} = (B_1, B_2, B_2\varphi_{\tau_{2,1}}, B_2\varphi_{\tau_{2,2}})' = (B_1, X_{2\tau_2}')'$ and

$$\varphi_{\tau_{2,i}}(s) = \begin{cases} 0 & \text{if } s < \tau_{2,i} \\ 1 & \text{if } s \geq \tau_{2,i} \end{cases}, \qquad s \in [0,1], i \in \{1,2\}. \tag{4.6.52}$$

Define $\hat{\beta}_{\tau_2} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ as the post-lasso least squares estimator and set $\hat{\eta}_{\tau_2} = (1, -\hat{\beta}_{\tau_2}')'$ so that

$$\hat{\eta}_{\tau_2} \Rightarrow \begin{bmatrix} 1 \\ -\left(\int_0^1 X_{2\tau_2} X_{2\tau_2}'\right)^{-1} \int_0^1 X_{2\tau_2} B_1 \end{bmatrix} = \eta_{\tau_2}. \tag{4.6.53}$$

We partition

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \qquad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \tag{4.6.54}$$

in conformity with S_t and define $\Lambda_{2\cdot} = (\Lambda_{21}, \Lambda_{22})$ and $\Lambda_{\cdot 2} = (\Lambda_{12}, \Lambda_{22})'$.

For each element of $S_{2t\tau_2} = (S_{2t}\varphi_{t,\tau_{2,1}}, S_{2t}\varphi_{t,\tau_{2,2}})$ it holds that

$$\Delta S_{2t}\varphi_{t,\tau_{2,i}} = \Delta S_{2t}\varphi_{t,\tau_{2,i}} + S_{2t-1}\Delta\varphi_{t,\tau_{2,i}} \tag{4.6.55}$$

and $\Delta\varphi_{t,\tau_{2,i}} = \varphi_{t,\tau_{2,i}} - \varphi_{t-1,\tau_{2,i}} = \mathbb{1}\{t = [\tau_{2,i}T]\}$. Since $\varphi_{t-1,\tau_{2,i}}\Delta\varphi_{t,\tau_{2,i}} = 0$ and

$$d(B_2(s)\varphi_{\tau_{2,i}}(s)) = dB_2(s)\varphi_{\tau_{2,i}}(s) + d\varphi_{\tau_{2,i}}(s)B_2(s) \tag{4.6.56}$$

for the asymptotic counterpart, we have the identities

$$\int_0^1 BdB_2(s)\varphi_{\tau_{2,i}}(s) = \int_{\tau_{2,i}}^1 BdB_2(s) + B(\tau_{2,i})B_2(\tau_{2,i}) \tag{4.6.57}$$

and

$$\int_0^1 B_2\varphi_{\tau_{2,i}}dB_2(s)\varphi_{\tau_{2,i}}(s) = \int_{\tau_{2,i}}^1 B_2\varphi_{\tau_{2,i}}dB_2(s) = \int_{\tau_{2,i}}^1 B_2dB_2(s). \tag{4.6.58}$$

Consequently, we can state the following important weak convergence results

$$\frac{1}{T}\sum_{t=2}^{T} X_{t-1\tau_2}\Delta S'_{2t\tau_2} \quad \Rightarrow \quad \int_0^1 X_{\tau_2}dB_{2\tau_2} + \begin{bmatrix} (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{12} \\ (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,2})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22}, \end{bmatrix} \tag{4.6.59}$$

where $dB_{2\tau_2} = (d(B_2(s)\varphi_{\tau_{2,1}}(s)), d(B_2(s)\varphi_{\tau_{2,2}}(s)))$ and

$$\frac{1}{T}\sum_{t=2}^{T} X_{t-1\tau_2}\Delta X'_{t\tau_2} \Rightarrow \int_0^1 X_{\tau_2}dX'_{\tau_2} + \Lambda_{\tau_2}, \tag{4.6.60}$$

where

$$\Lambda_{\tau_2} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & (1-\tau_{2,1})\Lambda_{12} & (1-\tau_{2,2})\Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} & (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,1})\Lambda_{12} & (1-\tau_{2,1})\Lambda_{22} & (1-\tau_{2,1})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \\ (1-\tau_{2,2})\Lambda_{12} & (1-\tau_{2,2})\Lambda_{22} & (1-\tau_{2,2})\Lambda_{21} & (1-\tau_{2,2})\Lambda_{22} \end{bmatrix}. \tag{4.6.61}$$

Under the null hypothesis, the cointegration residuals can be written as $\hat{e}_{t\tau_2} = \hat{\eta}'_{\tau_2}X_{t\tau_2}$ and we can show weak convergence of the sample moments. It holds that

$$\begin{aligned} \frac{1}{T^2}\sum_{t=1}^{T}\hat{e}^2_{t\tau_2} &= \hat{\eta}'_{\tau_2}\frac{1}{T^2}\sum_{t=1}^{T} X_{t\tau_2}X'_{t\tau_2}\hat{\eta}_{\tau_2} \\ &\Rightarrow \eta'_{\tau_2}\int_0^1 X_{\tau_2}X'_{\tau_2}\eta_{\tau_2} = \sigma^2\int_0^1 W^2_{\tau_2}, \end{aligned} \tag{4.6.62}$$

where $W_{\tau_2}(s) = W_1(s) - \left(\int_0^1 W_1 W_{2\tau_2}'\right)\left(\int_0^1 W_{2\tau_2} W_{2\tau_2}'\right)^{-1} W_{2\tau_2}$ and

$$\begin{aligned} \frac{1}{T}\sum_{t=2}^{T} \hat{e}_{t-1\tau_2}\Delta\hat{e}_{t\tau_2} &= \hat{\eta}_{\tau_2}' \frac{1}{T}\sum_{t=2}^{T} X_{t-1\tau_2}\Delta X_{t\tau_2}' \hat{\eta}_\tau \\ &\Rightarrow \eta_{\tau_2}' \left(\int_0^1 X_{\tau_2} dX_{\tau_2}' + \Lambda_{\tau_2}\right) \eta_{\tau_2} = \sigma^2 \int_0^1 W_{\tau_2} W_{\tau_2}' + \eta_\tau' \Lambda_\tau \eta_\tau. \end{aligned} \tag{4.6.63}$$

Next, we consider the bias-correction term for the first-order serial correlation coefficient. We denote the kernel weights as $w(j/M) = w_j$ and can show that

$$\hat{\psi}_{\tau_2} = \sum_{j=1}^{M} w_j \frac{1}{T}\sum_t \Delta\hat{e}_{t-j\tau_2}\Delta\hat{e}_{t\tau_2} + o_p(1). \tag{4.6.64}$$

Hence, we have the weak convergence result

$$\hat{\psi}_{\tau_2} = \hat{\eta}_{\tau_2}' \sum_{j=1}^{M} w_j \frac{1}{T}\sum_t \Delta X_{t-j\tau_2}\Delta X_{t\tau_2}\hat{\eta}_{\tau_2} + o_p(1) \Rightarrow \eta_{\tau_2}' \Lambda_{\tau_2} \eta_{\tau_2}. \tag{4.6.65}$$

For the long-run variance, we obtain the result

$$\hat{\sigma}_{\tau_2}^2 \Rightarrow \eta_{\tau_2}' \Omega_{\tau_2} \eta_{\tau_2}, \tag{4.6.66}$$

where

$$\begin{aligned} \Omega_{\tau_2} &= \begin{bmatrix} \sigma^2 & \Omega_{12} & (1-\tau_{2,1})\Omega_{12} & (1-\tau_{2,2})\Omega_{12} \\ \Omega_{21} & \Omega_{22} & (1-\tau_{2,1})\Omega_{22} & (1-\tau_{2,2})\Omega_{22} \\ (1-\tau_{2,1})\Omega_{12} & (1-\tau_{2,1})\Omega_{22} & (1-\tau_{2,1})\Omega_{21} & (1-\tau_{2,2})\Omega_{22} \\ (1-\tau_{2,2})\Omega_{12} & (1-\tau_{2,2})\Omega_{22} & (1-\tau_{2,2})\Omega_{21} & (1-\tau_{2,2})\Omega_{22} \end{bmatrix} \\ &= \begin{bmatrix} 1, -\kappa_{\tau_2}' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D_{\tau_2} \end{bmatrix} \begin{bmatrix} 1 \\ -\kappa_{\tau_2} \end{bmatrix} = \sigma^2(1 + \kappa_{\tau_2}' D_{\tau_2} \kappa_{\tau_2}) \end{aligned} \tag{4.6.67}$$

and

$$D_{\tau_2} = \begin{bmatrix} 1 & (1-\tau_{2,1}) & (1-\tau_{2,2}) \\ (1-\tau_{2,1}) & (1-\tau_{2,1}) & (1-\tau_{2,2}) \\ (1-\tau_{2,2}) & (1-\tau_{2,2}) & (1-\tau_{2,2}) \end{bmatrix}. \tag{4.6.68}$$

Now, we use the CMT to show that

$$
\begin{aligned}
Z_2 &= \frac{\frac{1}{T^2}\sum\limits_{t=2}^{T}\hat{e}_{t-1\tau_2}\Delta\hat{e}_{t\tau_2} - \hat{\psi}_{\tau_2}}{\frac{1}{T^2}\sum\limits_{t=2}^{T}\hat{e}^2_{t-1\tau_2}} \times \left(\frac{1}{\hat{\sigma}^2_\tau T^2}\sum_{t=2}^{T}\hat{e}^2_{t-1\tau_2}\right)^{1/2} \\
&\Rightarrow \frac{\sigma^2\int\limits_0^1 W_{\tau_2}dW_{\tau_2} + \eta'_{\tau_2}\Lambda_{\tau_2}\eta_{\tau_2} - \eta'_{\tau_2}\Lambda_{\tau_2}\eta_{\tau_2}}{\sigma^2\int\limits_0^1 W^2_{\tau_2}} \times \left(\frac{1}{\sigma^2(1+\kappa'_{\tau_2}D_{\tau_2}\kappa_{\tau_2})}\sigma^2\int_0^1 W^2_{\tau_2}\right)^{1/2} \\
&= \frac{\int\limits_0^1 W_{\tau_2}dW_{\tau_2}}{\left(\int\limits_0^1 W^2_{\tau_2}\right)^{1/2}\left(1+\kappa'_{\tau_2}D_{\tau_2}\kappa_{\tau_2}\right)^{1/2}}
\end{aligned} \tag{4.6.69}
$$

for each configuration of τ_2. Correspondingly, the test statistics for the remaining model selection outcomes have the asymptotic distributions

$$
Z_1 \sim \int_0^1 W_{\tau_1}dW_{\tau_1} \Bigg/ \left(\int_0^1 W^2_{\tau_1}\right)^{1/2} \left(1+\kappa'_{\tau_1}D_{\tau_1}\kappa_{\tau_1}\right)^{1/2}, \tag{4.6.70}
$$

$$
W_{\tau_1} = W_1(s) - \left[\int_0^1 W_1 W_{2\tau_1}\right]\left[\int_0^1 W_{2\tau_1}W'_{2\tau_1}\right]^{-1} W_{2\tau_1}(s),
$$

$$
\kappa_{\tau_1} = \left[\int_0^1 W_{2\tau_1}W'_{2\tau_1}\right]^{-1}\left[\int_0^1 W_{2\tau_1}W_1\right],
$$

$$
W_{2\tau_1} = \left[W_2(s), W_2(s)\varphi_{\tau_{2,1}}(s)\right],
$$

and

$$
Z_0 \sim \int_0^1 W_{\tau_0}dW_{\tau_0} \Bigg/ \left(\int_0^1 W^2_{\tau_0}\right)^{1/2}, \tag{4.6.71}
$$

$$
W_{\tau_0} = W_1(s) - \left[\int_0^1 W_1 W_2\right]\left[\int_0^1 W_2^2\right]^{-1} W_2(s),
$$

respectively. Naturally, the distributions of Z_2 and Z_1 depend on the timing of the breakpoint. Finally, selecting the infimum statistic over all potential model selection outcomes is a continuous transformation so that we can use the CMT to complete the proof. □

Table 4.6.5: Approximate critical values

The table summarizes the approximate critical values of our proposed cointegration test, where *ADF* marks the columns reporting critical values of the ADF test statistic while Z_τ reports those of the bias-corrected test statistic. *Note*: The lag truncation parameter in the ADF regression is determined using the AIC. Critical values for other order selection rules are not reported but can be obtained from the authors upon request. We use 25,000 replications to compute the finite sample critical values.

	ADF			Z_t		
$m^* = p^* = 1$						
T	10%	5%	1%	10%	5%	1%
100	−3.91	−4.28	−4.93	−4.09	−4.45	−5.13
200	−3.89	−4.24	−4.88	−4.03	−4.37	−5.04
400	−3.88	−4.23	−4.86	−3.96	−4.31	−4.92
∞	−3.86	−4.19	−4.83	−3.92	−4.26	−4.85
$m^* = p^* = 2$						
T	10%	5%	1%	10%	5%	1%
100	−4.51	−4.90	−5.59	−4.79	−5.18	−5.87
200	−4.51	−4.88	−5.53	−4.70	−5.07	−5.73
400	−4.48	−4.85	−5.47	−4.63	−4.99	−5.63
∞	−4.48	−4.84	−5.47	−4.58	−4.94	−5.57
$m^* = p^* = 3$						
T	10%	5%	1%	10%	5%	1%
100	−4.96	−5.35	−6.07	−5.34	−5.74	−6.45
200	−4.98	−5.36	−6.01	−5.25	−5.62	−6.30
400	−4.99	−5.34	−6.00	−5.19	−5.54	−6.20
∞	−4.98	−5.34	−6.02	−5.10	−5.48	−6.15
$m^* = p^* = 4$						
T	10%	5%	1%	10%	5%	1%
100	−5.29	−5.70	−6.43	−5.79	−6.21	−6.94
200	−5.38	−5.78	−6.46	−5.71	−6.11	−6.84
400	−5.42	−5.80	−6.44	−5.66	−6.03	−6.70
∞	−5.40	−5.78	−6.45	−5.57	−5.95	−6.60
$m^* = p^* = 5$						
T	10%	5%	1%	10%	5%	1%
100	−5.58	−6.00	−6.70	−6.18	−6.64	−7.46
200	−5.70	−6.10	−6.79	−6.11	−6.52	−7.29
400	−5.74	−6.16	−6.82	−6.03	−6.43	−7.15
∞	−5.77	−6.16	−6.86	−5.97	−6.35	−7.06
$m^* = p^* = 6$						
T	10%	5%	1%	10%	5%	1%
100	−5.77	−6.21	−7.02	−6.46	−6.98	−7.84
200	−5.97	−6.39	−7.08	−6.45	−6.88	−7.67
400	−6.05	−6.47	−7.22	−6.39	−6.80	−7.56
∞	−6.10	−6.49	−7.17	−6.33	−6.72	−7.41

Chapter 5

Conclusion and final remarks

After having presented the original work of this thesis in the preceding chapters, let me conclude by briefly reflecting upon the main findings of the individual studies in a broader context. This should serve to summarize the insights that the studies provide about the two major aspects of this thesis, namely the implications of the changing information technologies for the financial markets and the opportunities that an advanced understanding of high-dimensional models holds for re-thinking and improving existing econometric frameworks. Furthermore, this chapter critically discusses the limitations of each study, indicates potentially fruitful avenues for future related research and closes with some final remarks.

Chapter 2 starts by addressing one highly relevant aspect of the question about the dynamics between social media, investors, and the stock market. It analyzes the relationship of Twitter sentiment and activity with individual-level stock return volatility in an intraday context. Although the study finds a statistically significant influence of the respective Twitter variable on stock return volatility, these effects are of negligible size for practitioners who cannot expect to achieve an informational advantage by analyzing Twitter data for their intraday trading. This finding might appear to contradict one's expectations, assuming that information provided by a fast moving microblogging platform such as Twitter fits the increasing speed of trading in and relatively easy access to the financial markets. Instead, it indicates that retail investors do not process these information instantaneously or at least not in such a way that anticipating their behavior would give institutional investors any kind of advantage. Thus, while the relevance of recent innovations in information technology, resulting in platforms such as Twitter, for financial markets are documented in the literature quite well for daily observations (e.g., Bollen *et al.*, 2011; Sprenger *et al.*, 2014b), their influence does not appear to extent to an intradaily response of the stock market. One apparent shortcoming is the study's focus on blue-chip companies. Since the constituents of the DJIA that we use for the analysis in Chapter 2 all show a fairly high market capitalization and high percentage of institutional ownership, it would be interesting to rerun the models using stocks in which the amount of shares held by retail investors is higher. It appears to be reasonable to assume that such stocks are more easily influenced by investor sentiment due to their differing shareholder structure. Thus, in order to check the robustness of the key takeaway of our study, testing the assumption of no informational gain for intraday trading through systematic analysis of Twitter sentiment for this second class of stocks constitutes a valuable contribution. While we did not find these stocks being sufficiently covered by the Twitter community to allow the construction of a sentiment and Twitter activity time series, alternative online platforms such as Seeking Alpha or StockTwits might provide adequate data for such an analysis.

Shifting the focus from social media to publicly available news in general, Chapter 3 exploits the availability of daily news on firms' performance in the domain of environment, social, and governance, to investigate how investors incorporate these information into their investment decisions. The findings reveal that investors' predisposition towards ESG activities appears to differ quite distinctively between the firms of the sample: While for the majority of stocks in the sample investors, on average, react to ESG-information differently depending on the type of information and the stock's current financial performance, one group of companies that shows the highest median log-returns over the sampling period has rather ESG-averse investors, while the investors of another small group of stocks that shows the lowest median returns are rather ESG-affine. In light of the recent spike in popularity of ESG-investing touched upon in Chapter 1 the highly context-sensitive findings of this study clearly highlight the complexity of ESG-investing itself. Nevertheless, there are two key takeaways from Chapter 3: First, it is important for companies to realize that their ESG-activities are being closely monitored by investors and the public, as almost all stocks of the sample show a significant reaction in idiosyncratic returns to changes in ESG-related sentiment. This is likely due to the increasingly easy access to company-specific information that investors enjoy nowadays as well as the aforementioned lower entry barriers to trading for retail investors. It could also be indicative of the more and more pronounced necessity that companies are facing to include sustainability dimensions into their business models to secure their success, as Marburger (2011) argues. Secondly, due to the complexity of the topic, the work presented in Chapter 3 still has some shortcomings, which I want to address in the following. These, in turn, present potential ways in which future research can further contribute to unraveling the value of ESG information to investors and companies.

First of all, while the research presented in Chapter 3 already uses a fairly large data set of news articles, for some of the stocks data are still only quite sparsely available, which limits the general validity and explanatory power of the results. Even larger data sets of news articles could be obtained from various subscription-based resources.[1] Tapping into such data bases could also be useful to control for further news stories that are not included in the analysis. So far, each article that is used in the analysis has some ESG-related content. The main model used for the estimation of the effect of changes in ESG-related sentiment then only controls for non-ESG related news that are part of these ESG articles. Constructing an investor sentiment time series from general interest news, which can be entirely unrelated to sustainability and environmental topics, could serve as an useful, additional control variable to further isolate investors' reaction to ESG news. Another valuable contribution that future work could be devoted to is to further conceptualize the financial theory of ESG-investing. While the presented research assumes retail investors that directly engage in trading as a reaction to ESG-related news, a more advanced understanding of the role of ESG-concerned portfolio management in financial markets appears to be highly relevant. With respect to the ESG news themselves, it would also be interesting to explore to which kind of ESG-related content retail investors are most prone to react to. This could then be used to refine the lexicon approach used in Chapter 3 by giving a higher weight to more important ESG-related words than to less important ones. In this context, Chapter 3 often mentions potential "green-washing" attempts of companies. While in the context of the presented work it appears to be reasonable to assume that official news sources have fewer incentives to green-wash a company's ESG-image, the way in which media and companies themselves engage in such attempts is highly interesting from both the researcher's as well as the practitioner's point of view.

[1]The following page provides several such sources of news databases, the majority of which are fee-based `https://guides.library.cornell.edu/news_online`, last accessed July 31, 2019.

By proposing a new approach of detecting and consistently estimating multiple structural breaks in bivariate, potentially cointegrated systems, the last chapter of this doctoral thesis gives one example of how an advanced understanding of high-dimensional models can provide opportunities to improve existing econometric frameworks. Though often neglected, utilizing new insights gained in one domain of research, such as statistical learning, to approach well-known but not yet ideally solved problems of another domain from a different angle can often prove to be a worthwhile endeavor. In this spirit, the study presented in Chapter 4 builds upon recent advances in the understanding of the properties of the lasso to reframe the problem of identifying structural breaks in cointegrated regressions as a model selection problem. Despite its focus on bivariate systems and thus the method's limited applicability to more general cases, the flexibility and accuracy of the approach to detect and estimate multiple structural breaks provides a good alternative to conventional methods, which are often shown to be less flexible and accurate. Our findings should furthermore encourage future research on the topic of detecting and estimating structural breaks to further explore the benefits of reframing this as a model selection problem. Possible extensions, besides the obvious generalization to more than two system variables, could be using a group-lasso approach to model structural breaks in multiple-equation systems or in VECMs. From a more general perspective, there appear to be more domains of research that could benefit from re-thinking their respective existing econometric framework by considering alternative, potentially high-dimensional approaches. The literature on modeling expected surgery procedure times in hospitals, for example, has only recently started to explore the implementation of dimension reduction techniques to produce more flexible and accurate forecasts (see, for example, Jovanovic *et al.*, 2016). Another domain of research that is both related to the other two main chapters of this thesis and is shown to benefit from dimension-reduction techniques is sentiment analysis. While sentiment is traditionally modeled via lexicon approaches or machine learning algorithms, Hu *et al.* (2013) demonstrate that the lasso could also be a very useful tool to assess the sentiment of microblogging data.

Let me conclude the last chapter of this doctoral thesis by providing some closing remarks on the presented research. Given the considerably vast frame of the implications of a changing landscape in information technology and recent advances in high-dimensional modeling for financial markets and econometric modeling that this thesis is embedded in, one inevitably has to be selective in the specific aspects that are being covered. Thus, through a combination of subjectively perceived relevance of the issues, personal interest, and opportunity the three main chapters of this thesis are formed. They should first of all serve to illustrate the importance of exploring recent dynamics between (online) media, investors, and financial markets, for which the first two research projects appear to be well suited. Moreover, all three main research projects of this thesis show the benefits that can arise by combining econometric theory and real-world applications. Lastly, the original work presented above has the potential to stimulate the academic discourse and should encourage further research to critically discuss the studies and their approaches to advance our understanding of the changing dynamics in financial markets and new opportunities for improving existing econometric frameworks.

Bibliography

Admati, A. R. and Pfleiderer, P. (1988) A theory of intraday patterns: Volume and price variability, *Review of Financial Studies*, **1**, 3–40.

Admati, A. R. and Pfleiderer, P. (1989) Divide and conquer: A theory of intraday and day-of-the-week mean effects, *Review of Financial Studies*, **2**, 189–224.

Aktas, N., De Bodt, E. and Cousin, J.-G. (2011) Do financial markets care about SRI? Evidence from mergers and acquisitions, *Journal of Banking & Finance*, **35**, 1753–1761.

Andersen, T. G. and Bollerslev, T. (1997) Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, **4**, 115–158.

Andersen, T. G. and Bollerslev, T. (1998) Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance*, **53**, 219–265.

Andersen, T. G., Bollerslev, T. and Cai, J. (2000) Intraday and interday volatility in the Japanese stock market, *Journal of International Financial Markets, Institutions and Money*, **10**, 107–130.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007) Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics*, **89**, 701–720.

Andrews, D. W. K. and Ploberger, W. (1994) Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, **62**, 1383–1414.

Antweiler, W. and Frank, M. Z. (2004) Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance*, **59**, 1259–1294.

Arai, Y. and Kurozumi, E. (2007) Testing for the null hypothesis of cointegration with a structural break, *Econometric Reviews*, **26**, 705–739.

Arellano, M. (1987) Computing robust standard errors for within group estimators, *Oxford Bulletin of Economics and Statistics*, **49**, 431–434.

Aue, A. and Horváth, L. (2013) Structural breaks in time series, *Journal of Time Series Analysis*, **34**, 1–16.

Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes, *Econometrica*, **66**, 47–78.

Baker, M. and Wurgler, J. (2006) Investor sentiment and the cross-section of stock returns, *Journal of Finance*, **61**, 1645–1680.

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., Pouliquen, B. and Belyaeva, J. (2010) Sentiment analysis in the news, in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta.

Barber, B. M. and Odean, T. (2007) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies*, **21**, 785–818.

Barber, B. M., Odean, T. and Zhu, N. (2009) Do retail trades move markets?, *Review of Financial Studies*, **22**, 151–186.

Barberis, N., Shleifer, A. and Vishny, R. (1998) A model of investor sentiment, *Journal of Financial Econometrics*, **49**, 307–343.

Behrendt, S. and Schmidt, A. (2018) The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility, *Journal of Banking & Finance*, **96**, 355–367.

Bekiros, S. D. and Diks, C. G. (2008) The relationship between crude oil spot and futures prices: Cointegration, linear and nonlinear causality, *Energy Economics*, **30**, 2673–2685.

Belloni, A. and Chernozhukov, V. (2013) Least squares after model selection in high-dimensional sparse models, *Bernoulli*, **19**, 521–547.

Bewley, R. A. (1978) A further note on dynamic multipliers, *Bulletin of Economic Research*, **30**, 54–55.

Bierens, H. J. and Martins, L. F. (2010) Time-varying cointegration, *Econometric Theory*, **26**, 1453–1490.

Billingsley, P. (1999) *Convergence of Probability Measures*, Wiley, New York, 2 edn.

Black, F. (1986) Noise, *Journal of Finance*, **41**, 529–543.

Boehmer, E. and Kelley, E. K. (2009) Institutional investors and the informational efficiency of prices, *Review of Financial Studies*, **22**, 3564–3594.

Bollen, J., Mao, H. and Zeng, X.-J. (2011) Twitter mood predicts the stock market, *Journal of Computational Science*, **2**, 1–8.

Bollerslev, T., Cai, J. and Song, F. M. (2000) Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US Treasury bond market, *Journal of Empirical Finance*, **7**, 37–55.

Brammer, S., Brooks, C. and Pavelin, S. (2006) Corporate social performance and stock returns: UK evidence from disaggregate measures, *Financial Management*, **35**, 97–116.

Brissimis, S. N. (1976) Multiplier effects for higher than first order linear dynamic econometric models, *Econometrica*, **44**, 593–595.

Bukovina, J. (2016) Social media big data and capital markets - An overview, *Journal of Behavioral and Experimental Finance*, **11**, 18–26.

Cahan, S. F., Chen, C., Chen, L. and Nguyen, N. H. (2015) Corporate social responsibility and media coverage, *Journal of Banking & Finance*, **59**, 409–422.

Cameron, A. C. and Trivedi, P. K. (1986) Econometric models based on count data. Comparisons and applications of some estimators and tests, *Journal of Applied Econometrics*, **1**, 29–53.

Capelle-Blancard, G. and Petit, A. (2019) Every little helps? ESG news and stock market reaction, *Journal of Business Ethics*, **157**, 543–565.

Carrion-i Silvestre, J. L. and Sanso, A. (2006) Testing for the null hypothesis of cointegration with structural breaks, *Oxford Bulletin of Economics and Statistics*, **68**, 623–646.

Chan, N. H., Yau, C. Y. and Zhang, R.-M. (2014) Group lasso for structural break time series, *Journal of the American Statistical Association*, **109**, 590–599.

Chang, Y. and Park, J. Y. (2002) On the asymptotics of ADF tests for unit roots, *Econometric Reviews*, **21**, 431–447.

Chen, H., De, P., Hu, Y. and Hwang, B.-H. (2014) Wisdom of crowds: The value of stock opinions transmitted through social media, *Review of Financial Studies*, **27**, 1367–1403.

Cheong, C. W., Sinnakkannu, J. and Ramasamy, S. (2017) Reactive or proactive? Investor sentiment as a driver of corporate social responsibility, *Research in International Business and Finance*, **42**, 572–582.

Chiriac, R. and Voev, V. (2011) Modelling and forecasting multivariate realized volatility, *Journal of Applied Econometrics*, **26**, 922–947.

Compact, U. G. (2004) Who cares wins: Connecting financial markets to a changing world, *New York*.

Corbae, D. and Ouliaris, S. (1988) Cointegration and tests of purchasing power parity, *Review of Economics and Statistics*, **70**, 508–511.

Corsi, F. (2009) A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.

Craighead, W. D. (2010) Across time and regimes: 212 years of the US-UK real exchange rate, *Economic Inquiry*, **48**, 951–964.

Davidson, J. and Monticini, A. (2010) Tests for cointegration with structural breaks based on subsamples, *Computational Statistics and Data Analysis*, **54**, 2498–2511.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990) Noise trader risk in financial markets, *Journal of Political Economy*, **98**, 703–738.

Dimpfl, T. and Jank, S. (2016) Can internet search queries help to predict stock market volatility?, *European Financial Management*, **22**, 171–192.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993) A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, **1**, 83–106.

El Ghoul, S., Guedhami, O., Kwok, C. C. and Mishra, D. R. (2011) Does corporate social responsibility affect the cost of capital?, *Journal of Banking & Finance*, **35**, 2388–2406.

Engel, C. and Kim, C.-J. (1999) The long-run U.S./U.K. real exchange rate, *Journal of Money, Credit and Banking*, **31**, 335–356.

Engelberg, J. E. and Parsons, C. A. (2011) The causal impact of media in financial markets, *Journal of Finance*, **66**, 67–97.

Engle, R. F. and Gallo, G. P. (2006) A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, **131**, 3–27.

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and error correction: Representation, estimation and testing, *Econometrica*, **55**, 251–276.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987) Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica*, **55**, 391–407.

Engle, R. F. and Sokalska, M. E. (2012) Forecasting intraday volatility in the US equity market. Multiplicative component GARCH, *Journal of Financial Econometrics*, **10**, 54–83.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.

Fang, L. and Peress, J. (2009) Media coverage and the cross-section of stock returns, *Journal of Finance*, **64**, 2023–2052.

Forsberg, L. and Ghysels, E. (2007) Why do absolute returns predict volatility so well?, *Journal of Financial Econometrics*, **5**, 31–67.

Friedman, J., Hastie, T. and Tibshirani, R. (2001) *The elements of statistical learning*, vol. 1, Springer series in statistics New York.

Friedman, M. (2007) The social responsibility of business is to increase its profits, in *Corporate Ethics and Corporate Governance*, Springer, pp. 173–178.

Gallant, A. R. (1981) On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form, *Journal of Econometrics*, **15**, 211–245.

Garcia, D. (2013) Sentiment during recessions, *Journal of Finance*, **68**, 1267–1300.

Gilbert, M. (2019) The rising cost of investing responsibly, Bloomberg Opinion, `https://www.bloomberg.com/opinion/articles/2019-01-11/the-rising-cost-of-esg-and-socially-responsible-investing`, last accessed Aug 19, 2019.

Giot, P. (2005) Market risk models for intraday data, *European Journal of Finance*, **11**, 309–324.

Goss, A. and Roberts, G. S. (2011) The impact of corporate social responsibility on the cost of bank loans, *Journal of Banking & Finance*, **35**, 1794–1810.

Gregory, A. W. and Hansen, B. E. (1996a) Residual-based tests for cointegration in models with regime shifts, *Journal of Econometrics*, **70**, 99–126.

Gregory, A. W. and Hansen, B. E. (1996b) Tests for cointegration in models with regime and trend shifts, *Oxford Bulletin of Economics and Statistics*, **58**, 555–560.

Gregory, A. W., Nason, J. M. and Watt, D. G. (1996) Testing for structural breaks in cointegrated relationships, *Journal of Econometrics*, **71**, 321–341.

Grilli, V. and Kaminsky, G. (1991) Nominal exchange rate regimes and the real exchange rate - Evidence from the United States and Great Britain, 1885-1986, *Journal of Monetary Economics*, **27**, 191–212.

Hansen, B. E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory*, **8**, 489–500.

Hansen, B. E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.

Hatemi-J, A. (2008) Tests for cointegration with two unknown regime shifts with an application to financial market integration, *Empirical Economics*, **35**, 497–505.

Hausman, J. A. (1978) Specification tests in econometrics, *Econometrica*, **46**, 1251–1271.

Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables, *Annals of Probability*, **12**, 141–153.

Horowitz, J. and Huang, J. (2013) Penalized estimation of high-dimensional models under a generalized sparsity condition, *Statistica Sinica*, **23**, 725–748.

Hu, X., Tang, L., Tang, J. and Liu, H. (2013) Exploiting social relations for sentiment analysis in microblogging, in *Proceedings of the sixth ACM international conference on Web search and data mining*, ACM, pp. 537–546.

Hutto, C. J. and Gilbert, E. (2014) VADER: A parsimonious rule-based model for sentiment analysis of social media text, in *Eighth international AAAI conference on weblogs and social media*.

Jovanovic, M., Radovanovic, S., Vukicevic, M., Van Poucke, S. and Delibasic, B. (2016) Building interpretable predictive models for pediatric hospital readmission using tree-lasso logistic regression, *Artificial Intelligence in Medicine*, **72**, 12–21.

Karoglou, M. and Morley, B. (2012) Purchasing power parity and structural instability in the US/UK exchange rate, *Journal of International Financial Markets, Institutions and Money*, **22**, 958–972.

Kejriwal, M. and Perron, P. (2010) Testing for multiple structural changes in cointegrated regression models, *Journal of Business & Economic Statistics*, **28**, 503–522.

Kell, G. (2018) The remarkable rise of ESG, Online, `https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/#7cea69891695`, last accessed Aug 19, 2019.

Kim, Y. (1990) Purchasing power parity in the long run: A cointegration approach, *Journal of Money, Credit and Banking*, **22**, 491–503.

Kim, Y., Li, H. and Li, S. (2014) Corporate social responsibility and stock price crash risk, *Journal of Banking & Finance*, **43**, 1–13.

Knight, K. and Fu, W. (2000) Asymptotics for lasso-type estimators, *Annals of Statistics*, **28**, 1356–1378.

Kock, A. B. (2016) Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions, *Econometric Theory*, **32**, 243–259.

Koo, B., Anderson, H., Seo, M. H. and Yao, W. (2017) High-dimensional predictive regression in the presence of cointegration, Tech. rep., SSRN Working Paper.

Krüger, P. (2015) Corporate goodness and shareholder wealth, *Journal of Financial Econometrics*, **115**, 304–329.

Kumar, A. and Lee, C. M. C. (2006) Retail investor sentiment and return comovements, *Journal of Finance*, **61**, 2451–2486.

Kuo, B.-S. (1998) Test for partial parameter instability in regressions with I(1) processes, *Journal of Economics*, **86**, 337–368.

Kurozumi, E. and Skrobotov, A. (2017) Confidence sets for the break date in cointegrating regressions, *Oxford Bulletin of Economics and Statistics*, **80**, 514–535.

Kyle, A. S. (1985) Continuous auctions and insider trading, *Econometrica*, **53**, 1315–1335.

Lee, C. M., Shleifer, A. and Thaler, R. H. (1991) Investor sentiment and the closed-end fund puzzle, *Journal of Finance*, **46**, 75–109.

Li, X., Xie, H., Chen, L., Wang, J. and Deng, X. (2014) News impact on stock price return via sentiment analysis, *Knowledge-Based Systems*, **69**, 14–23.

Liao, Z. and Phillips, P. C. B. (2015) Automated estimation of vector error correction models, *Econometric Theory*, **31**, 581–646.

Lins, K. V., Servaes, H. and Tamayo, A. (2017) Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis, *Journal of Finance*, **72**, 1785–1824.

Lofts, G. (2018) Who cares wins in the brave world of ESG, Online, `https://www.ft.com/content/64ce9225-0c55-3256-967c-5fae6622c637`, last accessed Aug 19, 2019.

Loughran, T. and McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance*, **66**, 35–65.

Lux, T. and Marchesi, M. (1999) Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature*, **397**, 498–500.

Maki, D. (2012) Tests for cointegration allowing for an unknown number of breaks, *Economic Modelling*, **29**, 2011–2015.

Marburger, J. H. (2011) Science, technology and innovation in a 21st century context, *Policy Sciences*, **44**, 209.

Mendes, E. F. (2011) Model selection consistency for cointegrating regressions, Tech. rep., Northwestern University.

Miller, R., Michalski, W., Stevens, B. and Secretariat, O. (1998) The promises and perils of 21st century technology: An overview of the issues, *21st*, p. 7.

Mussa, M. (1986) Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications, *Carnegie-Rochester Conference on Public Policy*, **25**, 117–214.

Myšková, R. and Hájek, P. (2018) Sustainability and corporate social responsibility in the text of annual reports - the case of the IT services industry, *Sustainability*, **10**, 4119.

Nasukawa, T. and Yi, J. (2003) Sentiment analysis: Capturing favorability using natural language processing, in *Proceedings of the 2nd international conference on Knowledge capture*, ACM, pp. 70–77.

Naughton, J. P., Wang, C. and Yeung, I. (2014) Are CSR expenditures affected by investor sentiment?, Tech. rep., Citeseer.

Neal, R. and Wheatley, S. M. (1998) Do measures of investor sentiment predict returns?, *Journal of Financial and Quantitative Analysis*, **33**, 523–547.

Nofsinger, J. and Varma, A. (2014) Socially responsible funds and market crises, *Journal of Banking & Finance*, **48**, 180–193.

Oliveira, N., Cortez, P. and Areal, N. (2017) The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices, *Expert Systems with Applications*, **73**, 125–144.

Park, J. Y. and Hahn, S. B. (1999) Cointegrating regressions with time varying coefficients, *Econometric Theory*, **15**, 664–703.

Patton, A. J. and Sheppard, K. (2015) Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics*, **97**, 683–697.

Pencle, N. and Mălăescu, I. (2016) What's in the words? Development and validation of a multidimensional dictionary for CSR and application using prospectuses, *Journal of Emerging Technologies in Accounting*, **13**, 109–127.

Peress, J. (2014) The media and the diffusion of information in financial markets: Evidence from newspaper strikes, *Journal of Finance*, **69**, 2007–2043.

Perron, P. (2006) Dealing with structural breaks, in *Palgrave Handbook of Econometrics - Volume 1: Econometric Theory* (Eds.) H. Hassani, T. Mills and K. Patterson, Palgrave Macmillan UK, pp. 278–352.

Perron, P. and Yamamoto, Y. (2016) On the usefulness or lack thereof of optimality criteria for structural change tests, *Econometric Reviews*, **35**, 782–844.

Phillips, P. C. B. (1987) Time series regression with a unit root, *Econometrica*, **55**, 277–301.

Phillips, P. C. B. and Durlauf, S. N. (1986) Multiple time series regression with integrated processes, *Review of Economic Studies*, **53**, 473–495.

Phillips, P. C. B. and Ouliaris, S. (1990) Asymptotic properties of residual based tests for cointegration, *Econometrica*, **58**, 165–193.

Pötscher, B. M. and Schneider, U. (2009) On the distribution of the adaptive lasso estimator, *Journal of Statistical Planning and Inference*, **139**, 2775–2790.

Qian, J. and Su, L. (2016) Shrinkage estimation of regression models with multiple structural changes, *Econometric Theory*, **32**, 1376–1433.

Qu, Z. and Perron, P. (2007) Estimating and testing structural change in multivariate regressions, *Econometrica*, **75**, 459–502.

Quintos, C. E. and Phillips, P. C. B. (1993) Parameter constancy in cointegrating regressions, *Empirical Economics*, **18**, 675–706.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M. and Mozetič, I. (2015) The effects of Twitter sentiment on stock price returns, *PLoS ONE*, **10**, e0138441.

Reid, J., Templeman, L. and Mahtani, S. (2018) Big data shakes up esg investing, *Deutsche Bank Konzept*, `https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000478852/Big_data_shakes_up_ESG_investing.PDF`, last accessed Aug 19, 2019.

Schweikert, K. (2019) Asymmetric price transmission in the US and German fuel markets: a quantile autoregression approach, *Empirical Economics*, **56**, 1071–1095.

Shleifer, A. and Vishny, R. W. (1997) The limits of arbitrage, *Journal of Finance*, **52**, 35–55.

Siering, M. (2012) “boom” or “ruin”– Does it make a difference? Using text mining and sentiment analysis to support intraday investment decisions, in *2012 45th Hawaii International Conference on System Sciences*, IEEE, pp. 1050–1059.

Simon, N., Friedman, J., Hastie, T. and Tibshirani, R. (2013) A sparse-group lasso, *Journal of Computational and Graphical Statistics*, **22**, 231–245.

Sprenger, T. O., Sandner, P. G., Tumasjan, A. and Welpe, I. M. (2014a) News or noise? Using Twitter to identify and understand company-specific news flows, *Journal of Business Finance & Accounting*, **41**, 791–830.

Sprenger, T. O., Sandner, P. G., Tumasjan, A. and Welpe, I. M. (2014b) Tweets and trades: The information content of stock microblogs, *European Financial Management*, **20**, 926–957.

Stapleton, R. C. and Subrahmanyam, M. G. (1983) The market model and capital asset pricing theory: a note, *Journal of Finance*, **38**, 1637–1642.

Taylor, A. M. and Taylor, M. P. (2004) The purchasing power parity debate, *Journal of Economic Perspectives*, **18**, 135–158.

Taylor, M. P. (1988) An empirical examination of long-run purchasing power parity using cointegration techniques, *Applied Economics*, **20**, 1369–1381.

Taylor, M. P. (2006) Real exchange rates and purchasing power parity: Mean-reversion in economic thought, *Applied Financial Economics*, **16**, 1–17.

Tetlock, P. C. (2007) Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance*, **62**, 1139–1168.

Tibshirani, R. (1996) Regression selection and shrinkage via the lasso, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **58**, 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005) Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **67**, 91–108.

Varian, H. R. (2014) Big data: New tricks for econometrics, *Journal of Economic Perspectives*, **28**, 3–28.

Wang, H., Li, B. and Leng, C. (2009) Shrinkage tuning parameter selection with a diverging number of parameters, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **71**, 671–683.

Westerlund, J. and Edgerton, D. L. (2006) New improved tests for cointegration with structural breaks, *Journal of Time Series Analysis*, **28**, 188–224.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–838.

Wilms, I. and Croux, C. (2016) Forecasting using sparse cointegration, *International Journal of Forecasting*, **32**, 1256–1267.

Xiao, Z. (2009a) Functional-coefficient cointegration models, *Journal of Econometrics*, **152**, 81–92.

Xiao, Z. (2009b) Quantile cointegrating regression, *Journal of Econometrics*, **150**, 248–260.

Yang, S., Mo, S. and Liu, A. (2015) Twitter financial community sentiment and its predictive relationship to stock market movement, *Quantitative Finance*, **15**, 1637–1656.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **68**, 49–67.

Zhang, C. H. and Huang, J. (2008) The sparsity and bias of the lasso selection in high-dimensional linear regression, *Annals of Statistics*, **36**, 1567–1594.

Zheludev, I., Smith, R. and Aste, T. (2014) When can social media lead financial markets?, *Scientific reports*, **4**, 4213.

Zou, H. (2006) The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429.