

Working Paper Series — 13/2020

Composite Survey Sentiment as a Predictor of Future Market Returns: Evidence for German Equity Indices

Zuzana Rakovská



Czech National Bank — Working Paper Series — 13/2020

The Working Paper Series of the Czech National Bank (CNB) is intended to disseminate the results of the CNB's research projects as well as the other research activities of both the staff of the CNB and collaborating outside contributors, including invited speakers. The Series aims to present original research contributions relevant to central banks. It is refereed internationally. The referee process is managed by the CNB Economic Research Division. The working papers are circulated to stimulate discussion. The views expressed are those of the authors and do not necessarily reflect the official views of the CNB.

Distributed by the Czech National Bank, available at www.cnb.cz

Reviewed by: František Čech (Charles University, Prague)

Adam Kučera (Czech National Bank)

Project Coordinator: Jan Janků

Issued by: © Czech National Bank, December 2020

Composite Survey Sentiment as a Predictor of Future Market Returns: Evidence for German Equity Indices

Zuzana Rakovská *

Abstract

In this paper, I construct a novel composite sentiment indicator which captures the irrational beliefs of the general population in Germany. This indicator is used to demonstrate that the sentiment of the general public is responsible for temporary overreactions of the aggregate German stock market and also of its narrower segments as embodied in four important equity indices from the DAX family. My results show that population-wide beliefs work as a contrarian predictor of future returns on German equity indices for horizons of six to twelve months. In addition, an out-of-sample framework is developed to underline the degree of improvement achieved by combining several survey-based measures into one composite sentiment indicator. The results reveal that the composite indicator exhibits more accurate forecasting performance than the popular sentiment measure, consumer confidence.

Abstrakt

V tomto článku konstruuji nový kompozitní indikátor sentimentu, který zachycuje iracionální důvěru široké veřejnosti v Německu. Pomocí tohoto indikátoru ukazuji, že sentiment široké veřejnosti má na svědomí dočasné nadměrné reakce německého akciového trhu jako celku i jeho dílčích segmentů, které jsou zahrnuty do čtyř významných akciových indexů z rodiny DAX. Výsledky ukazují, že důvěra sdílená celou populací funguje jako inverzní prediktor budoucích výnosů německých akciových indexů v horizontech od šesti do dvanácti měsíců. Dále vytvářím out-of-sample rámec ke zdůraznění rozsahu vylepšení dosaženého zkombinováním několika ukazatelů z dotazníkových šetření do jednoho kompozitního indikátoru sentimentu. Výsledky ukazují, že kompozitní indikátor vykazuje přesnější predikční schopnost než oblíbený ukazatel sentimentu, kterým je spotřebitelská důvěra.

JEL Codes: G17, G40, G41.

Keywords: Composite indicator, consumer confidence, DAX indices, return predictability, sentiment.

*Zuzana Rakovská, Czech National Bank and Masaryk University, zuzana.rakovska@cnb.cz. The research was supported by Masaryk University [projects MUNI/A/1028/2018 and MUNI/A/1059/2019].

The author notes that the paper represents her own views and not necessarily those of the Czech National Bank.

I would like to thank Adam Kučera, František Čech, Simona Malovaná, Jan Frait and seminar participants at the Czech National Bank for useful comments. All errors and omissions remain the fault of the author.

1. Introduction

A consensus on how to express market sentiment and on the direction of its effect on stock prices (and subsequent returns) in both the short and long term seems to be absent from the existing literature. This might erode the profits of investors who build their strategy on the underlying relation, but it might also have some serious implications for the overall financial system. Specifically, the ambiguity in the sentiment-return relation poses a problem for policy makers concerned with identifying the existence, magnitude, and causes of systemic events such as asset bubbles or quantifying risks to financial stability.

This paper contributes to the literature examining the relationship between sentiment¹ and stock market returns in two main respects. First, I construct a novel survey-based composite sentiment indicator which captures the irrational beliefs of the general public. The importance of the general public in driving market movements was widely discussed in response to the Global Financial Crisis (Akerlof and Shiller, 2010) and most recently, for example, during the technological stock boom in spring 2020 (Bloomberg, 2020). To capture the beliefs of such a wide population, I combine a variety of sentiment measures from the surveys available for the given market while not distinguishing between the respondents' background or the type of questions asked. Each of these measures is cleaned of fundamental factors prior to inclusion in the composite index to ensure that only the irrational part – sentiment in its true sense – is considered.²

There is a vast body of empirical evidence on the relationship between stock market performance and sentiment proxies derived from consumer surveys (Lemmon and Portniaguina, 2006; Schmeling, 2009; Chang et al., 2011) and investment surveys (Fisher and Statman, 2000; Verma and Soydemir, 2006; Schmeling, 2007). All these studies refer to the explanatory variable as “sentiment”, so it is legitimate to hypothesize that these survey-based measures might contain a common component. Moreover, if the expectations of households about their future consumption and future financial position affect stock markets, then the expectations of firms captured by business surveys might do so as well. Hence, I conjecture that a combination of various survey-based proxies, regardless of the respondents' background, will demonstrate population sentiment more clearly than individual measures alone. I test this premise within an out-of-sample framework and show that the survey-based composite sentiment measure provides a better fit than the consumer confidence index, a popular proxy for sentiment adopted in related studies.

Second, the paper extends the work of Schmeling (2009) by employing not only returns on the aggregate stock market, but also returns on several equity indices available for this market as the dependent variable. Schmeling (2009) showed that the importance of survey-based sentiment for future returns builds over time and its effect is negative in both the short and long run. By employing my novel composite sentiment measure, I aim to demonstrate that these findings are applicable to narrower segments of the stock market besides its aggregate form. My results show that the

¹ The definition of sentiment might differ across the relevant literature, but the one offered by Shleifer and Summers (1990) seems to work well when observing the effects of sentiment in the financial markets. According to Shleifer and Summers (1990, p. 19), sentiment can be viewed as “beliefs that are not fully justified by fundamental news.” Even though it is not explicitly stated, it is essential to think about the *aggregate* state of such beliefs, as an individual economic agent hardly affects the financial market with her excessive optimism or pessimism.

² It is reasonable to assume that survey participants form their expectations about the future state of the economy partially based on current or past macroeconomic developments. In other words, good economic prospects might lead to an optimistic assessment of the current situation and the survey responses might contain not only the irrational part, but also the rational part, which needs to be controlled for. As a result, such “orthogonalization” of sentiment to macroeconomic variables has commonly been applied in the existing literature (Baker and Wurgler, 2006; Lemmon and Portniaguina, 2006; Han and Li, 2017).

beliefs of the general population work as a contrarian predictor of future returns on equity indices for horizons of six to twelve months. However, the paper does not find significance of sentiment predictability over the horizon of one to three months, in contradiction with the results in Schmeling (2009). This result is verified by several robustness considerations. In line with the critique formulated in Sibley et al. (2016) and the inconclusive evidence for the short-run effects of sentiment demonstrated in a number of related studies (Brown and Cliff, 2004; Schmeling, 2007; Corredor and Santamaria, 2015), I argue that this discrepancy is caused by the fact that Schmeling (2009) used a single measure of survey-based sentiment and did not clean it of fundamental factors prior to the analysis. The intuition for my results is provided by the overreaction hypothesis (Barberis et al., 1998), based on which uninformed market participants – loosely speaking the general public – overreact to a string of good news, which, in turn, leads to temporary overvaluation in the stock markets. Based on the theory, subsequent mean-reversion in the long run is then ensured by the actions of sophisticated traders, and this is where the negative effect of sentiment in longer periods emerges.

Next, I leverage the fact that the individual equity indices are created based on the size of the companies included in their calculation and that they are mutually exclusive as regards their constituents. The returns on such indices can be then considered returns on portfolios sorted by size – large, medium, and small capitalization – and hence used to confirm the popular finding in Baker and Wurgler (2006). According to the authors, there is cross-sectional variation in the effects of sentiment when one considers stocks with different characteristics. More specifically, stocks that are attractive to speculators, such as small stocks or stocks that are hard to value and hard to arbitrage, are more prone to sentiment than stocks with the opposite characteristics. This premise is further validated in the empirical framework. I am not aware of any existing study exploring the differences in the sentiment-return relation among various equity indices in one country.

Finally, recent decades have seen a process of globalization, characterized by tighter economic and financial integration among countries and regions. This environment is conducive to the spread of sentiment across markets, which can happen in two directions. The first is direct and represents the effect of sentimental foreign investors trading on a given country's market. The second happens via local investors, who simply “catch” the beliefs of investors from abroad (Chang et al., 2011). Baker et al. (2012) were the first to examine how international sentiment affects returns in local stock markets. The authors found that outside-world (“global”) sentiment matters for future portfolio returns, while local sentiment does not. In this paper, I employ an adaptation of the methodology proposed in Baker et al. (2012) to construct a “global” measure of sentiment. I approximate economies whose sentiment might matter for the equity markets studied by economies that are the largest export destinations for goods produced in the country studied. This approach, I believe, provides more reliable information about irrational beliefs shared globally than just considering a few selected countries. The results support the findings in the existing literature, namely, that “global” sentiment drives the contrarian predictability of country-specific sentiment.

The focus of this paper is on European financial markets, as analysis of the underlying relation in this region seems to be underrepresented in the existing literature.³ I select Germany as the reference country for this study and perform the analysis on monthly data capturing the dynamics of the German market in the period 2005–2019. This selection was made for three reasons. First, Germany

³ The majority of studies examine the sentiment-returns relation in the US financial markets, and European countries usually enter the sample in studies examining the global/local effects of sentiment (Baker et al., 2012; Corredor and Santamaria, 2015) or studies that leverage a country-specific source of sentiment data (Schmeling, 2007; Rutkowska, 2017).

has a highly developed equity market with a long history of price series for various broad equity indices. It is possible to obtain a nice history of price data for the aggregate stock market captured by the CDAX index and for equity indices that are suitable for verifying the cross-sectional variation of sentiment effects: the DAX (large capitalization index), the MDAX (medium capitalization index), and the SDAX (small capitalization index). Second, there are several well-established providers of survey-based leading indicators for the German economy that are similar to consumer confidence and that are suitable for constructing a composite survey-based sentiment indicator. Last, Germany is the largest European economy based on national GDP and is the most important trading partner for the majority of European countries.⁴ Hence, identification of the effects that the sentiment of the German population has on its equity markets appears to be of great relevance for other European economies. Moreover, given the strong interconnectedness of the euro area and EU financial markets and the current vulnerabilities of national financial systems, potential overreaction (and subsequent repricing) in the German equity markets might generate additional risks to their financial stability.

The rest of this paper is organized as follows. Section 2 summarizes the foundations of the sentiment-return relationship and reviews the empirical results available in the existing literature. It also provides a discourse on the implications of this relationship for financial stability. Section 3 introduces the methodology which underlies the construction of the composite sentiment indicator, and Section 4 offers a description of the sample data. The empirical framework is described in Section 5. Sections 6 and 7 represent the main parts of this paper and list the main results and the results of robustness tests, respectively. Concluding remarks are available in Section 8.

2. Sentiment in Financial Markets

Numerous academics have examined the sentiment-return relation. The significant interest in this topic is due mainly to the vast variety of ways in which sentiment can be measured (which directly implies a vast variety of research designs and subsequent studies), as well as to the fact that the results of the existing studies are still ambiguous, which is driving researchers increasingly to apply alternative approaches. In general, sentiment measures can be grouped into two main categories based on the way sentiment is collected – direct and indirect. Direct sentiment is constructed from some form of survey targeted at either the general population or a subpopulation of individual or institutional investors. The second group of sentiment variables – indirect sentiment – covers a range of measures that are, in a sense, proxy variables or a combination of proxy variables regarded as implicitly indicating the sentiment of investors.⁵ This study only employs direct, survey-based,

⁴ Based on World Integrated Trade Solutions (WITS) statistics (<https://wits.worldbank.org/>) for 2018, Germany ranked as the first trading partner for both exports and imports in 14 out of 26 countries of the EU – Austria, the Czech Republic, Denmark, Finland, France, Hungary, Italy, the Netherlands, Poland, Slovakia, Slovenia, Bulgaria, Romania, and Sweden.

⁵ Indirect sentiment variables can be further divided into three categories according to the data source/technique used to extract the proxy – financial proxies, measures that arise from sentiment analysis, and search-based measures. See, for example, Rakovská and Svoboda (2016) for a discussion of these various approaches.

sentiment measures,⁶ so I will concentrate mainly on the stream of literature that provides empirical results for this setting.

Prior work on the topic suggests two directions of the sentiment-return relation. The first camp is based on positive feedback theory (De Long et al., 1990a), according to which sentiment affects stock returns positively in the short run.⁷ This view is consistent with the underreaction hypothesis postulated in Barberis et al. (1998), who stated that underreaction affects stock returns in the short run of about one to twelve months. This effect is positive and predictable, because underreaction causes current good news to be reflected in stock prices slowly, which in a sense creates positive autocorrelation in the short run of up to one year.

Hengelbrock et al. (2013) examined the effects of survey-based sentiment indicators in the US and Germany and found evidence of a positive relation in the short to medium term but no evidence of a significant relation in the long run. By studying the Chinese stock market, Han and Li (2017) showed that the composite sentiment indicator, which the authors constructed from three indirect sentiment proxies, positively predicted future stock market returns in the short run of up to one year and that this effect reversed and became negative in the long run of the subsequent months.

The reversal of the positive effect and the domination of the negative relationship between sentiment and returns in the long run has its theoretical foundation in the second important hypothesis raised in Barberis et al. (1998) – the overreaction hypothesis. Economic agents seem to overreact to private information signals that exhibit a consistent pattern (a string of positive or negative news)⁸ and by doing so, they erode the stock markets by pushing prices away from their fundamental values. According to Barberis et al. (1998), the correction comes in the long term of about three to five years, because of noise trader risk and limits to arbitrage (De Long et al., 1990b; Shleifer and Summers, 1990), which hinder rational arbitragers from acting in the short run. Based on this theory, which in many studies is also termed the *contrarian view* or the *negative feedback view*, sentiment affects stock market returns negatively in the long term.⁹ The situation of concurrent strategies of two

⁶ There are two reasons for this. First, indirect sentiment measures are formed by underlying proxy variables, which might not always reflect the population sentiment. For example, market turnover, a popular financial proxy (Baker and Stein, 2004), might mimic extreme levels of sentiment only, especially when the market experiences excess optimism or pessimism and the trading volume increases. As a result, the empirical findings in studies applying indirect measures are often inconclusive. Instead, survey sentiment represents the direct responses of the population to questions related to their future expectations, and even the possible risk of untruthful answers should not cause the survey measures not to reflect, at least partially, the participants' sentiment at any point of observation. Second, several well-established indirect sentiment proxies (mainly in US studies), such as closed-end fund discounts and first-day IPO returns, could not be collected for European countries, while data on survey-based sentiment have rather a long history in the European context.

⁷ De Long et al. (1990a) argued that rational investors are aware of overreacting noise traders and, because they too have short investment horizons, they prefer to trade in the same direction as noise traders today in anticipation of possible destabilizing effects caused by noise traders tomorrow (De Long et al., 1990a, p. 380). They called this strategy the *positive feedback strategy*, because in certain situations it might be more profitable for rational arbitragers to join the bandwagon with speculators. The expected effect of sentiment caused by overreaction in the financial markets is then positive. This is supported by the *momentum strategy*, which has its source in an anomaly documented in Jagadeh and Titman (2001) and which postulates that buying (selling) stock with a history of high (low) returns over the last three to twelve months yields superior profits.

⁸ One possible explanation is that investors might become overconfident as a result of an illusion that new private information is superior. Moreover, when new signals exhibit a consistent pattern, the overreaction might be produced by representativeness (Kahneman and Tversky, 1974) – a heuristic which postulates that investors tend to overly rely on events that are representative even though they are of low statistical weight.

⁹ I acknowledge that there might be other, non-sentiment-related features that cause overreaction in the financial markets while being independent of fundamental news. These features are linked to market microstructure and

investor groups is also discussed in Kostolany (1990). The author distinguishes between “shaky hands” – the general public, which is sentimental, prone to herding, and acts based on hot tips – and “firm hands”, who represent the professional investor community. Kostolany (1990) states that markets are subject to an everlasting cycle in which securities travel from shaky to firm hands. He also asserts that shaky-hand investors have the ability to affect the financial market in the short run, while the long-run effects are driven by fundamentals. In other words, the switch from a positive to a negative effect in the long run also has its foundation in the investment strategy framework.

An early study by Clarke and Statman (1998) provided support for the *negative feedback hypothesis* by showing that the relation between the direct measure of sentiment and US stock returns for a 4- to 52-week period is negative. Brown and Cliff (2005) employed the same direct measure of sentiment and analyzed the sentiment-return relation over the horizon of six months to three years. They found strong evidence that sentiment is a contrarian predictor of US stock market returns one to three years ahead. In contrast to the above studies, there are also several authors who have found a negative predictive power of sentiment even in the short run, for example, Fisher and Statman (2000) and Huang et al. (2014) for the US market and Schmeling (2009) for a large set of 18 economies using the consumer confidence index as the sentiment variable.

2.1 The Sentiment-Return Relation and Its Implications for Financial Stability

As indicated in the previous section, there is a vast number of empirical papers showing that market sentiment has the capacity to affect equity returns and asset valuation (Dunne et al., 2011). This has some important implications for financial market performance, for example, when investors incorporate the observed dependence into their trading strategies. Nevertheless, there are other important facets of the sentiment-return relation, many of which have already been examined in the preceding literature and which suggest more system-wide implications for the financial markets.

First, studies that analyze the international spillover effects of sentiment have found that returns on local stocks are more prone to foreign (global) sentiment than to national sentiment (Baker et al., 2012; Corredor and Santamaria, 2015; Han and Li, 2017). In other words, enthusiasm and fear in the foreign population might cause excessive shifts in domestic markets, shifts that are more pronounced than those implied by local sentiment. Given the high degree of economic integration among countries, the interconnectedness of financial institutions on the global level, and the vulnerabilities of some financial systems, this might generate additional risks to national financial stability.

Second, there is an influential stream of literature which asserts that market sentiment might explain the existence of asset “bubbles” and relates it to important systemic events such as the Global Financial Crisis (GFC) (Akerlof and Shiller, 2010; Martin and Ventura, 2011; Shefrin, 2016). Martin and Ventura (2011) use a financial accelerator model to show that shocks to investor sentiment cause market bubbles – a crisis represents the end of a bubbly period induced by a shock to investor sentiment. They further argue that when investors are optimistic, stock prices of firms rise, while the investor pessimism accompanied by bubble bursts causes firm valuations to fall. The authors build their reasoning on a premise which is similar to representativeness heuristics, namely, that “during a bubbly episode, investors expect bubbles to survive with a high probability and new bubbles to appear at a rate that is proportional to the stock of existing bubbles” (Martin and Ventura, 2011, p. 22). This might stand as an alternative to the over- and underreaction hypothesis proposed

include price-contingent trading in the form of stop-loss or take-profit trades and automatic portfolio insurance mechanisms that generate price discontinuities (Easley and O’Hara, 1991; Osler and Savaser, 2011; Savaser, 2011).

by Barberis et al. (1998), even though the specific timing of the reversal to fundamentals is not provided.

Another stream of literature points to a direct relationship between sentiment and systemic risk measures (Tang and Yan, 2010; Smales, 2014; Barone-Adesi et al., 2012; Borovkova et al., 2017).¹⁰ Ergunor (2016) even proposes a novel exuberance index (EI) based on information about industries that exhibit signs of “exuberance” according to the industry-specific sentiment index and suggests that the EI provides a signal of upcoming crisis which comes earlier for regulators to react with preventive measures than other systemic risk indicators.¹¹ Barone-Adesi et al. (2012) conclude that fluctuations in sentiment increase the probability of occurrence of systemic events in financial markets. Therefore, they argue that regulators should closely track not only measures of systemic risk, but also sentiment variables. The reason why systemic risk is affected by sentiment is that increased optimism and overconfidence on the markets usually lead market participants to engage in excessive lending, a practice outlined in Minsky’s financial instability hypothesis (Minsky, 1977).

Lastly, Dow (2011) stresses that effectiveness of macroprudential policy is implied not only by using complex models which describe given parts of the financial system, but also by paying “attention to the indicators of fragility” (p. 246). One such indicator is market sentiment, which plays an essential role in asset valuation and the evolution of systemic risk. Interestingly, she links macroprudential regulation with psychological theory, as she states that macroprudential regulation might be viewed as a “nudging” exercise (Thaler and Sunstein, 2008) designed by central authorities with the aim to induce more rational behavior by financial institutions in the system.

3. A Composite Sentiment Indicator as a Measure of Public Sentiment

In the initial step of my analysis, I considered 13 survey-based sentiment measures. Table 1 provides a short description of them. This initial set contains almost all the survey-based sentiment measures that are available for Germany with a long data history and are conducted on a wide range of respondents. The selection was not affected by the type of questions asked. For example, I considered survey-based measures that reflect not only future beliefs about the dynamics in financial markets, but also expectations about the future development of the economy as a whole. The backgrounds of the respondents also vary across the surveys. For example, all the Sentix sentiment measures are calculated from the responses of either individual or professional investors, while the European Commission Directorate-General for Economic and Financial Affairs (DG ECFIN) uses

¹⁰ Borovkova et al. (2017) constructed a sentiment-based systemic risk indicator, SenSR, using the tone of news (its sentiment) about systemically important financial institutions (SIFIs). The authors found that this novel measure Granger-causes three of the most common measures of systemic risk at different time lags – up to 12 weeks. These measures are SRISK (Brownlees and Engle, 2017), the VIX (Brenner and Galai, 1989), and the LIBOR-OS spread (which indicates systemic stress in money markets). In other words, they found that the sentiment measure provides an earlier warning signal about the stability of the financial system than other common systemic risk measures. In a similar vein, Smales (2014) studied how sentiment gathered from news about major banks relates to market credit risk as captured by changes in credit default swap (CDS) spreads and bank credit risk as captured by the LIBOR-OS spread. The author found that the relation for CDS spreads is negative but asymmetric in news polarity, i.e., that positivism imprinted in news decreases ex-ante credit risk expectations, but the relation is stronger for negative news. No significant relation was observed for the bank credit risk measure. Similarly, Tang and Yan (2010) found that survey-based sentiment (as measured by the consumer confidence indicator) is the most prevalent determinant of market-wide credit spreads.

¹¹ Ergunor (2016) reviewed two groups of systemic risk indicators applied by central banks – threshold and systematic – and argue that both types of measure provide signals of upcoming crisis but these signals are not early enough. The sentiment index applied in the EI is based on the effects with which stock market returns of industries respond to both momentum and liquidity.

the responses of households to calculate the Consumer Confidence Indicator (CCI). Even though these measures track various states/factors in the economy (depending on the target of the questions employed), the overall optimism of the German population might be a common component among them. As a result, I decided to combine all these 13 measures into one composite sentiment indicator which captures the sentiment of the general public. The underlying process is described below.

Table 1: Description of the Survey-Based Sentiment Measures

Provider	Mnemonics	Description	Freq.	Start	Transf.
DG ECFIN	bci	Business Confidence Indicator	M	1985	-
DG ECFIN	cci	Consumer Confidence Indicator	M	1985	-
DG ECFIN	esi	Economic Sentiment Indicator	M	1985	<i>s100</i>
Ifo	grifpbus	ifo Business Climate Index for Germany	M	2005 ^b	<i>s100</i>
Ifo	grifpex	ifo Business Expectations for Germany	M	2005 ^b	<i>s100</i>
Ifo	geifombc	ifo Current Assessment for Germany	M	2005 ^b	-
Sentix	sntmdxh1	Sentix Sentiment German Equities (Large Caps) Headline Index	W ^a	2001	<i>m100</i>
Sentix	sntmdxi1	Sentix Sentiment German Equities (Large Caps) Institutional Investors	W ^a	2001	<i>m100</i>
Sentix	sntmdxp1	Sentix Sentiment German Equities (Large Caps) Individual Investors	W ^a	2001	<i>m100</i>
ZEW	grzecurr	ZEW Economic Situation Germany	M	1991	-
ZEW	grzewi	ZEW Indicator of Economic Sentiment Germany	M	1991	-
ZEW	grzedeme	ZEW Germany DAX Stock Market Expectations	M	1991	-
ZEW	grzedemt	ZEW Germany TecDAX Stock Market Expectations	M	1991	-

Note: The table provides a short description of the initial set of survey-based sentiment indicators considered in this study. The first column lists the official provider of the given series, the second column lists the mnemonics used throughout this study, the fourth column identifies the available frequency of the data (W – weekly, M – monthly), and the fifth column indicates the official start of calculation of the series. The original scales of the variables differed, so I decided to transform all the series so that they are centered around 0 and range between -100 and 100. The type of transformation used is reported in the last column of the table (*d100* – 100 was subtracted from the variable, *m100* – the variable was multiplied by 100, (-) – no transformation). *bci*, *cci*, and *esi* were downloaded from the DG ECFIN website. The data source for all the other indicators is Bloomberg.

^a Sentix sentiment variables were transformed to monthly frequency by taking the simple average of the weekly observations in the respective month.

^b The ifo Business Climate Index for industry and trade, which had a more than 40-year history, was replaced by an updated ifo Business Climate Index as of April 2018. Due to the change in its composition, its historical values have been restored only until 2005 (Sauer and Wohlrabe, 2018).

Each sentiment variable in my sample expresses, in a numerical way, how optimistic the survey respondents were at the time the survey was conducted. It is reasonable (and also common in the existing literature) to assume that the current or past economic situation influences respondents' perceptions about the future, and if the macroeconomic prospects are good today and were good in the past, respondents tend to extrapolate these optimistic prospects to their responses in the survey. Therefore, one might object that survey-based sentiment contains not only the irrational part (sentiment in its true sense), but also the rational part, which is closely linked to macroeconomic fundamentals. Nowzohour and Stracca (2020) states that a central issue with using confidence and uncertainty measures of sentiment is the risk that these measures reflect actual news rather than "animal spirits", in which case it is only possible to observe shifts in fundamental factors.

The accepted method in the influential stream of literature for "cleaning" raw sentiment measures of rational (fundamental) factors was proposed in Baker and Wurgler (2006) and further in Lemmon and Portniaguina (2006). I decided to follow the authors and decompose all the initial 13 sentiment measures into rational and irrational parts by regressing each of them on the set of macroeconomic variables. The residuals of those regressions then represent the irrational part of sentiment, while the fitted values stand for rational beliefs that are affected by market fundamentals.

The same set of macroeconomic controls was used to decompose the initial battery of survey sentiment variables. Specifically, I employed the yearly change in industrial production (ip_t) as a proxy for economic growth (Baker and Wurgler, 2006; Schmeling, 2009), the yearly change in the final consumption expenditure of households and non-profit institutions serving households (NPISH)¹² as a proxy for consumption growth (Baker and Wurgler, 2006), the term spread, defined as the difference between ten-year and two-year German government bond yields smoothed by its 12-month moving average (ts_t), as a proxy for future economic expectations (Baker et al., 2012), and the yearly change in the nominal effective exchange rate ($neer_t$) (Han and Li, 2017). Summary statistics of these macroeconomic controls can be found in Table A1.

I acknowledge that I might be missing some important fundamental variables in the analysis; however, I believe that the proposed set is reasonable and captures the macroeconomic conditions of the German economy well.¹³ In addition, I acknowledge that orthogonalization is a statistical procedure which is based on a simple regression equation and, as such, the resulting two elements which I refer to as *rational* and *irrational* might differ from rational and irrational expectations as understood in the theory. In fact, I model the rationality of sentiment as a reflection of historical time series of fundamental factors, so I admittedly assume it to be backward-looking only. As a result, the irrational part of sentiment might contain that part of the excessive optimism or pessimism of the population which reflects forward-looking announcements made by national authorities, such as projections of future economic prospects. In other words, in the event of a crisis outbreak, the approach assumes that any currently expected further deterioration of the economic conditions on top of the current macroeconomic state is considered "irrational". To facilitate discussion of whether the selection of macroeconomic variables used in the cleaning procedure affects the results, I perform the robustness analysis using the "uncleaned" variant of sentiment (see Section 6.3).

Figure A1 plots the dynamics of each initial sentiment variable together with its irrational and rational parts. Clearly, the selected controls nicely capture the dynamics of the first six individual

¹² Data on final consumption expenditure are available at quarterly frequency only, so I used cubic spline interpolation in order to obtain monthly observations for this variable.

¹³ I experimented with several other macroeconomic controls, such as the yearly changes in the house price index (calculated by the ECB), the consumer price index, and disposable income, but the results either showed no significant change or were biased due to multicollinearity in the individual regressions.

sentiment indicators during and between the recession years. As a result, the irrational sentiment measures do not exhibit such extreme peaks during this period compared to the original series. All the sentiment indicators provided by ZEW except for the ZEW Indicator of Economic Sentiment Germany (*grzewi*) are captured well by the control variables, which might suggest that these sentiment measures do not contain much information about irrational sentiment. Lastly, the Sentix sentiment variables and *grzewi* seem largely not to be fitted by the selected controls; therefore, irrational sentiment mimics the tendencies of the original versions of the respective survey-based variables.

As the next step, I used these newly created measures to construct a single sentiment indicator. Following Baker and Wurgler (2006), I used principal components analysis (PCA) to deliver the result.¹⁴ All 13 variables were standardized to have zero mean and unit variance prior to PCA. The resulting composite survey-based sentiment indicator represents the first principal component of the underlying analysis. It explains 46.7% of the sample variance and can be characterized by the following equation:

$$\begin{aligned}
 sent_t = & 0.937bci_t^{ir} + 0.920cci_t^{ir} + 0.915esi_t^{ir} + 0.890grifpbus_t^{ir} + \\
 & +0.563grifpex_t^{ir} + 0.750geifombc_t^{ir} - 0.239sntmdxh_t^{ir} - \\
 & -0.210sntmdxi_t^{ir} - 0.327sntmdxp_t^{ir} + 0.906grzcurr_t^{ir} - \\
 & -0.357grzewi_t^{ir} - 0.633grzedeme_t^{ir} - 0.525grzedemt_t^{ir}
 \end{aligned} \tag{1}$$

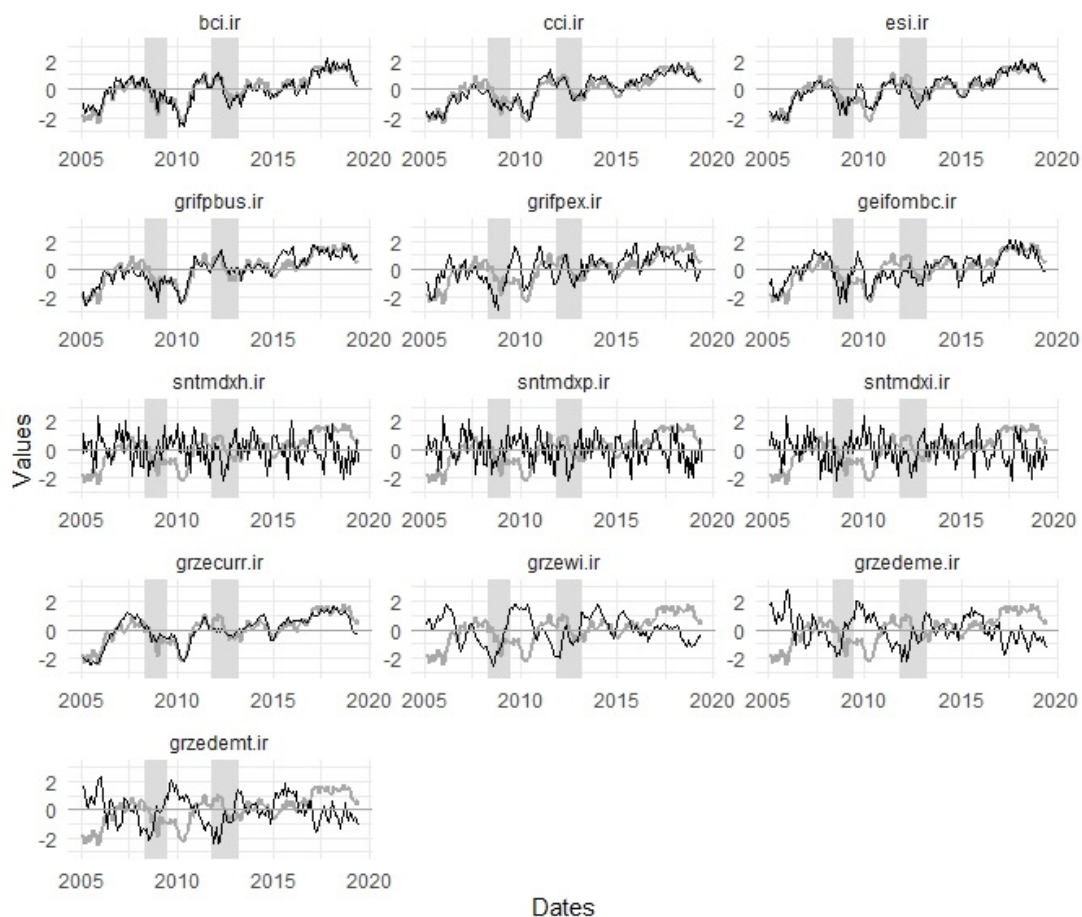
where $sent_t$ is the composite sentiment indicator and the superscript ir identifies irrational sentiment.

As outlined in the Introduction, the aim of this paper is to examine how the sentiment of the general population affects future stock returns. To capture the beliefs of such a wide population, I combined the irrational parts of 13 sentiment proxies that represent almost all the survey-based sentiment variables available for Germany. I did not filter the series based on the respondents' background or the type of questions asked; hence, the resulting composite sentiment indicator $sent$ exhibits the sentiment which is the most common among the surveyed population. The sign and magnitude of the loadings listed in Equation (1), however, reveal some interesting aspects. First, the signs of six out of the 13 ingredients of the composite index are negative, while the other sentiment variables enter $sent_t$ with a positive sign. A possible explanation for this might lie in the different characteristics of the survey respondents. The Sentix surveys are distributed to either individual or institutional investors, who might be better informed about what is going on in the financial markets than the general public. A similar characteristic can be drawn for the ZEW indicators, which result from surveying experts from banks, insurance companies, and the financial departments of selected corporations. In contrast, the first six indicators from Equation (1) emerge from surveys whose participants are households on the one hand, or companies from various sectors on the other. This population might thus act as a noise trader or "dumb money", while investors who respond to the Sentix or ZEW surveys behave in line with the theory as rational arbitragers or "smart money" (Shleifer and Summers, 1990). In other words, investors might expect noise traders (the general population) to overreact in the short run; therefore, they form expectations in the opposite direction in order to drive prices back to their intrinsic values. The explanation above, however, does not apply to the opposite sign of $grzcurr^{ir}$ and $grzewi^{ir}$, which both emerge from the same respondents'

¹⁴ There are other techniques available to obtain the composite index, such as the Kalman filter (Brown and Cliff, 2004; Oliveira et al., 2017) and partial least squares (PLS) (Huang et al., 2014). However, the PCA technique is the most common.

answers. Nevertheless, the former sentiment proxy represents experts' assessment of the current economic situation, while the latter reflects their expectations. The pairwise correlation between these two measures is moderately negative and significant (-0.42), which suggests that the two views are contrasting in nature.

Figure 1: Composite Sentiment Indicator and Individual Sentiment Variables



Note: This figure depicts the dynamics of the composite sentiment indicator (the grey line) and the respective individual sentiment indicators (the black line). Shaded areas represent CEPR recession months.

Second, regarding the magnitude of the loadings, it is not surprising that the three DG ECFIN sentiment indicators share the largest effects. The Economic Sentiment Indicator (*esi*) represents a weighted combination of confidence indicators calculated for five different economic sectors: industry, services, consumers, construction, and retail trade (European Commission, 2007), while the underlying information from the industrial and consumer sectors is applied in the construction of *bci* and *cci*, respectively. Further, the share of the variance explained by the Sentix measures clearly deviates from the rest of the indicators. Again, this might be caused by the fact that the Sentix surveys are fed by answers from the population of investors, while the rest of the surveys address a general audience. Moreover, the Sentix measures display a larger degree of fluctuation than the other variables (Figure 1). This is because they originate from weekly instead of monthly survey rounds. Considering these facets, *sent* can also be thought of as the irrational overreaction by the public compared to expert views. I elaborate more on this topic in Section 6.3.

4. Data

This paper examines how public sentiment affects the return dynamics in the German equity market. To this end, I employed a sample of monthly data with a span of almost 15 years – from January 2005 to May 2019. The sample size was influenced by the availability of two sets of sentiment variables. In April 2018, the ifo Institute changed the methodology for constructing the ifo Business Climate Index for Germany so that it also contains the service sector (Sauer and Wohlrabe, 2018). The old version, the ifo Business Climate Index for industry and trade, is available for the period before 2005; however, since the sentiment indicators provided by Sentix GmbH have a history only from 2001, I decided not to include the old version of the ifo index in the sample and to align the start point of all the data to January 2005. The period of observation is well comparable to other studies that have examined the sentiment-return relationship in European markets.¹⁵

4.1 Market Returns

Three general methodological approaches to selecting market returns are available in the existing literature. In the first, authors examine the effects of sentiment on the returns on individual stocks either separately for each firm (Kurov, 2010) or in a panel data framework (Aissia, 2016). In the second and most widely used approach, authors use returns on portfolios sorted by a characteristic such as size or value as the dependent variable (Brown and Cliff, 2005; Baker and Wurgler, 2006, 2007; Chang et al., 2011) and show empirically that the effect of sentiment is more a story of the given characteristic. The third avenue of research employs market returns, which are generally expressed as changes in the value of a stock market index such as the S&P 500 (Clarke and Statman, 1998; Fisher and Statman, 2000; Huang et al., 2014; Chu et al., 2016) or the DAX30, the German blue-chip equity index (Schmelting, 2007; Schmitz et al., 2007; Hengelbrock et al., 2013).

In this paper, I decided to combine the last two approaches. Specifically, I examine the returns on five German equity market indices (the third approach): a large-capitalization (large-cap) index – the DAX, a medium-capitalization (mid-cap) index – the MDAX, a small-capitalization (small-cap) index – the SDAX, a technology sector index – the TecDAX, and the overall German equity market index – the CDAX.¹⁶ Because three of them (the DAX, the MDAX, and the SDAX) are mutually exclusive as regards their constituents, I treat these three indices as portfolios which are sorted by size (market capitalization). This allows me to examine the cross-sectional effects of sentiment (the second approach). More specifically, I calculate the returns on long-short portfolios based on size characteristics, that is, portfolios in which I go long on the smaller-cap indices and short on the larger-cap indices. Then I use these return series as a dependent variable to test the hypothesis that smaller stocks are more prone to sentiment, similarly to Baker and Wurgler (2006) and Han and Li (2017). Besides constructing a small-minus-big portfolio (SMB), which takes into account only the small-cap (SDAX) and large-cap (DAX) indices (analogous to the SMB factor of Fama and French (1993)), I also construct small-minus-medium (SMM) and medium-minus-big (MMB) portfolios, which employ returns on the medium-cap index – the MDAX. By doing so, I aim to fill a gap in the existing literature, which concentrates only on SMB.

For each stock market index k and for each month t in the sample period, I compute continuously compounded stock market returns, R_t^k , using end-of-month values based on the formula $R_t^k = \ln(P_t^k/P_{t-1}^k)$. Using these return series for $k = DAX, MDAX, SDAX$, I then calculate the returns on

¹⁵ For example, Corredor et al. (2013) used the same number of observations and Baker et al. (2012) used 25 yearly observations for each European country.

¹⁶ More information about these indices can be found on the Deutsche Börse AG website (<https://deutsche-boerse.com>).

the SMB, SMM, and MMB portfolios as described above. Table A1 provides the summary statistics, based on which I can conclude that all the return series are stationary at the 1% confidence level.

4.2 Economic Variables

The last set of variables that enter the analysis is formed of four controls that are directly linked to economic fundamentals and that have been considered in the existing literature as predictors of stock returns (included in parentheses below). Those economic variables will be included in the main analysis as controls to capture some of the equity market dynamics. These variables are as follows (see also Table A1):

- The yearly growth rate of the dividend yield for the given index k (dy_t^k), where the monthly dividend yield is computed as the aggregate gross dividend per share in the last twelve months divided by the price of the broad stock market index k in the given month (Verma and Soydemir, 2006; Schmeling, 2009).
- The yearly growth of the consumer price index lagged by one month (cpi_{t-1}) as a proxy for inflation (Brown and Cliff, 2005; Huang et al., 2014). Following Welch and Goyal (2008), the inflation is lagged by two months relative to the stock market return to account for the delay in consumer price index releases.
- The term spread, defined as the difference between ten-year and two-year German government bond yields smoothed by its 12-month moving average (ts_t), as a proxy for future economic expectations (Brown and Cliff, 2005; Schmeling, 2007; Baker et al., 2012).
- A dummy variable indicating a recession month (bc_t).

5. Empirical Design

To estimate the effects of the composite survey-based sentiment indicator $sent$ on the future returns on various German DAX-family stock market indices, I decided to follow the empirical design suggested in Brown and Cliff (2005),¹⁷ that is, for each stock market index k , I estimate a series of regressions with future h -month average returns as the dependent variable and sentiment in month t as the explanatory variable. Similarly to Schmeling (2009), I consider the equations for the individual horizons of the future h months, where $h = 1, 2, 3, 6, 9, 12$, to be the system, and estimate them jointly. This empirical design allows me not only to test the null hypothesis that sentiment does not affect future returns separately for each horizon h , but also to test these hypotheses jointly for all the horizons h at once.¹⁸ Therefore, for each stock market index k , I run the *naïve* specification – the system of six equations, one for each horizon h – as follows:

$$\frac{1}{h} \sum_{h_0=1}^h R_{t+h_0}^k = \alpha_0^{k(h)} + \alpha_1^{k(h)} sent_t + \xi_{t+1 \rightarrow t+h}^{k(h)} \quad (2)$$

¹⁷ This empirical design was also applied in Schmeling (2007).

¹⁸ Based on Mark (1995) and Schmeling (2009), the joint test of multiple horizons is more reasonable than single-period tests of sentiment-return predictability.

where $\frac{1}{h} \sum_{h_0=1}^h R_{t+h_0}^k$ is the simple arithmetic average of the continuously compounded returns in the 1 to h -months which follow after month t , and $sent_t$ is the composite sentiment variable derived in Section 3.

As the next step, I re-estimate Equation (2) with the set of four control variables to form the baseline specification as:

$$\frac{1}{h} \sum_{h_0=1}^h R_{t+h_0}^k = \beta_0^{k(h)} + \beta_1^{k(h)} sent_t + \Theta^{k(h)} Z_t + \varepsilon_{t+1 \rightarrow t+h}^{k(h)} \quad (3)$$

where Z_t^k represents the matrix of control variables, one of which is specific to the stock market index k . Specifically, it contains three controls that are used for all indices k – the yearly growth rate of the consumer price index ($cpit$), the term spread smoothed by its 12-month moving average (ts_t), and a dummy variable which indicates a recession month according to the Centre for Economic Policy Research (CEPR)¹⁹ (bc_t). The last control is index-specific and represents the yearly growth rate of the dividend yield for the given index k (dy_t^k).

The system is estimated using ordinary least squares (OLS), for two reasons. First, according to Greene (2012, p. 294) and Cameron and Trivedi (2005, p. 210), it is not outperformed in terms of efficiency by seemingly unrelated regression (SUR, Zellner, 1962) when all the equations in the system have identical explanatory variables.²⁰ Second, it is computationally faster than other methods, a property which was of great benefit in the bootstrap simulation described in Appendix B.

The functional specification of Equations (2) and (3) and the high persistence of the explanatory variables (see Table A1) introduce two well-known econometric problems into the analysis. First, OLS estimation produces consistent but biased coefficients in finite samples when the regressors are persistent and their innovations are correlated with those of the dependent variable (Stambaugh, 1999; Ferson et al., 2003). Second, the design of the dependent variable, more specifically its overlapping-period structure induced by averaging across h months, introduces autocorrelation into the residuals of order $(h-1)$. It is possible to circumvent this problem by using bias-adjusted standard errors as suggested by Hansen and Hodrick (1980) and Newey and West (1987), but these corrections do not perform well in finite samples (Hodrick, 1992).

To overcome the above-mentioned problems, I decided to adopt the methodology suggested in Brown and Cliff (2005) and later employed also in Schmeling (2007).²¹ These authors addressed both problems using bootstrap simulation. They ran restricted vector autoregression (VAR) under

¹⁹ <https://cepr.org/>

²⁰ Taking into account the design of the dependent variables, it is reasonable to assume that the disturbance terms $\varepsilon_{t+1 \rightarrow t+h}^{k(h)}$ are contemporaneously correlated across horizons h . The problem of selecting the appropriate multivariate regression model would immediately point to SUR. Nevertheless, if all the equations in the system have identical explanatory variables, which is my case, SUR does not yield more efficiency than equation-by-equation ordinary least squares (see the proof in Greene, 2012).

²¹ Several other approaches have been developed and used in the literature. Schmeling (2009) and Corredor et al. (2013) employed a moving-block bootstrap as a device to reduce the small-sample bias induced by highly persistent regressors. In addition, Corredor et al. (2013) followed the procedure described in Jagadeh and Titman (2001) and formed long-short portfolios to avoid issues connected with highly autocorrelated disturbance terms. This procedure was also used in Baker and Wurgler (2006), accompanied by bootstrapped standard errors to mitigate the Stambaugh (1999) small-sample bias. A different approach was used in Han and Li (2017), who followed Amihud et al. (2009) to build a multipredictor augmented regression method (mARM), a multipredictor extension of the ARM model suggested by Amihud and Hurvich (2004), to circumvent the small-sample bias in OLS-estimated

the null of no return predictability and used the bootstrapped residuals to form new time series that were, in turn, used to run the original overlapping-period estimation. This procedure was then repeated (simulated) a sufficient number of times (10,000 in the case of the simulation in Brown and Cliff, 2005) to obtain the distribution of the coefficients. I believe that this framework reliably mitigates both problems – by producing appropriate critical values for inference in finite samples with persistent regressors, and by correcting the bias induced by the overlapping-period dependent variable. A detailed description of the bootstrap simulation employed in this paper can be found in Appendix B.

6. Results of the h -period Average Return Regressions

The results of the h -period future returns regressions for the *naïve* specification (Equation (2)) and the baseline specification (Equation (3)) for each index k are collected in Table 2. The aim of this paper is to examine the forecasting power of public sentiment for equity returns at a horizon of up to 12 months, so only the results for *sent* are reported and the estimation outputs for other explanatory variables are omitted.²² Panel A contains the estimated coefficients $\alpha_1^{k(h)}$ and $\beta_1^{k(h)}$, which are derived from the bootstrap simulation described in Appendix B. I used the same number of simulations as Brown and Cliff (2005) – 10,000 – to build the empirical distribution for the inference. The coefficients, as they are, represent the effect (in %) of a one standard deviation increase in *sent* on the h -period average future returns on equity index k . To obtain the overall effect for the h -horizons, the coefficients could be multiplied by the respective horizon h . The last row reports the mean of all the coefficients across all horizons. The corresponding p-values that arise from the simulation can be found in Panel B, with the last row representing the p-value for the F-test that $\alpha_1^{k(h)}$ or $\beta_1^{k(h)}$ for the given k are jointly zero for all horizons h . Finally, Panel C reports $\Delta adj R^2$, which equals the difference between the adjusted R^2 from estimating Equation (3) and the adjusted R^2 from regressing the h -period average future returns only on the constant and control variables. This measure simply shows how much explanatory power sentiment brings to the particular regression.

The outcome for the *naïve* specification does not differ much from the results produced in the baseline model in the case of the first four equity indices. This implies that the pattern, as discussed below, is little affected by controlling for the yearly change in the dividend yield, growth of consumer prices, the term spread, and recession months. Nevertheless, the results for the technology

predictive slope coefficients. Note that Han and Li (2017) studied the effects not on the accumulated h -period returns, but rather on the future one-month returns up to horizon h . As a result, the authors did not have to overcome the problem of overlapping periods. Similarly, by analyzing “only” the future h -horizon one-month returns, Corredor and Santamaria (2015) employed a single corrective device – Newey-West standard errors (Newey and West, 1987) – to avoid heteroscedasticity and autocorrelation. Last but not least, Huang et al. (2014) employed a combination of the wild-bootstrap technique and the mARM model of Amihud et al. (2009) to address both problems. Finally, there are also a number of studies which did not focus on persistent regressors, small-sample bias, or the problem of overlapping periods at all – see, for example, Perez-liston et al. (2018) and Concetto and Ravazzolo (2019) as the most recent examples.

²² Given the bootstrap simulation design (see Appendix B), it is not possible to make a proper inference about the return predictability of variables other than sentiment. The simulation is based on restricted vector autoregression under the null of no predictability of sentiment. As such, it is possible to draw appropriate critical values for inferences about the effects of sentiment, but it is not possible to use the results for other variables without the risk of biased results. Moreover, the analysis involves five equity indices, each analyzed in two systems of six-horizon equations (Equations (2) and (3)). This yields 60 sets of coefficients. Reporting the estimation outputs for all the explanatory variables included in these 60 sets would be exhausting, not to mention the fact that such tables might be confusing for the reader. Brown and Cliff (2005) and Schmeling (2007) did not report estimates for other explanatory variables for the same reasons.

segment index – TecDAX – suggest that the population-wide sentiment captured by the variable *sent* lacks the capacity to explain its future returns. One possible explanation is that, historically, the TecDAX contains the largest portion of foreign member companies compared to the other four indices. Even though this share has been shrinking in recent years (from 17% in 2005 to just 10% in 2018; see Table 5), this characteristic might have caused the returns on the TecDAX to be more prone to sentiment which extends beyond the boundaries of Germany. This topic is further elaborated in Section 6.2.

Table 2: Sentiment Coefficients and Significance from the h -period Average Returns Regressions

	Equation (2)					Equation (3)				
	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}
Panel A: Coefficients on <i>sent</i>										
$h = 1$	-0.251	-0.299	-0.504	-0.54	0.184	-0.113	-0.193	-0.310	-0.370	0.501
$h = 2$	-0.224	-0.262	-0.484	-0.538	0.141	-0.060	-0.133	-0.234	-0.319	0.492
$h = 3$	-0.315	-0.351	-0.548	-0.636	0.066	-0.090	-0.162	-0.212	-0.367	0.495
$h = 6$	-0.565	-0.576	-0.822	-0.922	-0.276	-0.348	-0.390	-0.482	-0.634	0.178
$h = 9$	-0.696	-0.692	-0.957	-1.039	-0.423	-0.380	-0.412	-0.491	-0.593	0.139
$h = 12$	-0.704	-0.696	-0.931	-1.022	-0.402	-0.320	-0.351	-0.397	-0.491	0.208
Mean	-0.459	-0.479	-0.708	-0.783	-0.118	-0.219	-0.274	-0.354	-0.462	0.336
Panel B: Significance of <i>sent</i>										
$h = 1$	0.411	0.388	0.243	0.150	0.927	0.704	0.650	0.470	0.310	0.503
$h = 2$	0.479	0.467	0.268	0.162	0.961	0.797	0.753	0.520	0.356	0.562
$h = 3$	0.311	0.287	0.163	0.071	0.933	0.690	0.640	0.463	0.233	0.557
$h = 6$	0.011	0.007	0.002	0.001	0.291	0.073	0.059	0.025	0.007	0.801
$h = 9$	0.000	0.000	0.000	0.000	0.072	0.005	0.005	0.001	0.002	0.534
$h = 12$	0.000	0.000	0.000	0.000	0.038	0.001	0.001	0.000	0.001	0.579
Mean	0.180	0.167	0.106	0.081	0.534	0.455	0.432	0.331	0.209	0.887
Panel C: $\Delta adjR^2$										
						-0.005	-0.004	-0.004	-0.001	-0.005
						-0.003	-0.002	-0.001	0.002	-0.005
						0.005	0.006	0.008	0.013	-0.005
						0.040	0.042	0.048	0.058	0.002
						0.066	0.070	0.068	0.077	0.012
						0.064	0.067	0.066	0.074	0.014

Note: The table reports the estimated bootstrap simulation coefficients $\alpha_1^{k(h)}$ and $\beta_1^{k(h)}$ (Panel A) together with the respective p-values for each horizon h (Panel B) from regressing the composite sentiment indicator $sent_t$ on the average future returns on equity index k according to Equations (2) and (3), respectively. The last rows in Panel A and Panel B list the mean of the bootstrap coefficients across all the horizons h and the p-value from the test that the coefficients across the forecast horizons are jointly equal to zero. Numbers in bold in Panel A identify significance of up to 10% according to the simulated p-values listed in Panel B. Panel C reports the difference between the $adjR^2$ from estimating Equation (3) and the $adjR^2$ from regressing the h -period average returns on the constant and control variables only. Due to the estimation of the system of equations with h -period future returns, the sample period was reduced to February 2005–May 2018.

The signs and magnitudes of the estimated coefficients in Table 2 are not surprising given the results in the stream of influential literature (Brown and Cliff, 2005; Baker and Wurgler, 2006; Lemmon and Portniaguina, 2006; Schmeling, 2009). However, a significant effect was observed only for the horizon of six months or more, while the impact of sentiment in the short run of one to three months is negligible. This pattern holds for all indices k except for the TecDAX (see the discussion above). Strong negative predictability of sentiment for horizons longer than six months was found in Brown and Cliff (2005). Support for the present results can also be found in Schmeling (2007), who showed that the sentiment of individual investors negatively affects future returns of up to one year. Moreover, he argues that it takes more time for the sentiment of individuals to become a significant

predictor of returns. In their study, Han and Li (2017) showed that sentiment acts as a contrarian predictor of future returns for the majority of horizons between nine and 60 months. However, they assert that in the short run, sentiment is a strong momentum signal in the Chinese stock market. This finding was exposed to a critique published in Cheema et al. (2018), which contends that the positive momentum predictability of sentiment found in Han and Li (2017) is caused by the boom and bust period of 2006–2008. Finally, Corredor and Santamaria (2015) demonstrated on a sample of three Central European markets – the Czech Republic, Hungary, and Poland – that sentiment derived from consumer confidence has no predictive capacity for selective index returns over the horizon of one month. Similar results can be found in less recent studies, such as Clarke and Statman (1998), Fisher and Statman (2000), and Brown and Cliff (2004).

In contrast, Schmeling (2009), who employed a survey-based sentiment measure, report a significantly negative effect even for one lead of returns, which contradicts the findings discussed above. Nevertheless, it can be argued, in line with the critique in Sibley et al. (2016), that the results in Schmeling (2009) are not directly comparable with those presented in this study, because the authors did not employ the “orthogonalization” procedure to remove the business cycle component from the sentiment measure. The author did control for various risk factors in the regressions, but the estimated effect might well be attributable to that part of consumer confidence which is driven by fundamental factors.²³

The intuition for the observed pattern is provided by the overreaction hypothesis (Barberis et al., 1998), based on which uninformed market participants overreact to a string of news with the same sign, which, in turn, leads to short-term departures of stock prices from their intrinsic values. The sentiment of overreacting market participants might be captured well in the population-wide surveys that are used in the construction of *sent*. Based on the theory, the subsequent correction of prices back to their fundamental level is then ensured by the actions of sophisticated traders in the long run. In other words, the sentiment of *noise* traders might serve as a contrarian predictor of future stock returns.²⁴ The overreaction hypothesis builds on the concepts of noise trader risk and limits to arbitrage (De Long et al., 1990b; Shleifer and Summers, 1990), which describes lagged mean-reversion as being a result of sophisticated traders being risk averse in arbitraging out market overvaluation in the short run. Inconclusive evidence for the other side of the story – the momentum predictability of sentiment – has been presented in number of empirical studies (see above). Moreover, Brown and Cliff (2005) argue that it is a natural outcome, because short-term predictability of sentiment would cause abnormal market returns. The possibly diverse effects of “noise traders” versus “smart-money” (Shleifer and Summers, 1990), or shaky versus firm hands (Kostolany, 1990), are further examined in Section 6.3.

²³ For example, using a VAR framework, Barsky and Sims (2012) showed that the popular consumer confidence measures for the US exhibit both fundamental and “noise” factors that represent just a reflection of news about “apparently permanent” current and future economic fundamentals. Moreover, Hodula et al. (2019) constructed a composite measure of 13 macroeconomic variables (the Households Macroeconomic Conditions Index, H-MCI) and showed on a large international sample that this index co-moves with the business cycle as well as with consumer confidence indicators.

²⁴ The ambiguous result for returns on the TecDAX equity index might also be due to the fact that technology firms have become very popular in recent years. As a result, the general public might overreact to good news connected to this sector for considerably longer compared to traditional sectors. The prescribed mean-reverting effect of sentiment might therefore happen after the window of 12 months considered in this paper.

6.1 Cross-Sectional Effects of Sentiment

The mean coefficients reported in the last row of Panel A of Table 2 represent the average of the sentiment effects estimated for horizons of $h = 1, 2, 3, 6, 9,$ and 12 future months. When considering only the first four indices k (all except the TecDAX), the mean coefficients exhibit a negative sign and their absolute magnitude decreases with rising market capitalization of the index. This provides the intuition for what was already shown in the pioneering study of Baker and Wurgler (2006) – increased sentiment affects small stocks negatively, while the effect decreases in intensity or completely vanishes once the size of the stocks goes up. To demonstrate this pattern empirically for the returns on German equity indices, I applied a similar method to the one used in Baker and Wurgler (2006), that is, I calculated the returns on long-short portfolios based on size characteristics (I go long on the smaller-capitalization indices and short on the larger-capitalization indices) and supplied these series in Equations (2) and (3) as the dependent variable. The nature of the DAX, MDAX, and SDAX equity indices allows me to do so because these markets are mutually exclusive as regards their constituents. In contrast to the existing literature, which examines the size effects of sentiment only within small-minus-big (SMB) portfolios based on empirical breakpoints of market capitalization, for example, at the top and bottom 20% (Corredor et al., 2013) or 30% (Baker and Wurgler, 2006; Han and Li, 2017), I also assess two other natural scenarios – small-minus-medium (SMM) and medium-minus-big (MMB) portfolios.

Table 3: Cross-Sectional Effects of Sentiment

	Equation (2)			Equation (3)		
	SMB	SMM	MMB	SMB	SMM	MMB
Panel A: Coefficients on <i>sent</i>						
$h = 1$	-0.209	0.002	-0.153	-0.119	0.025	-0.072
$h = 2$	-0.249	-0.048	-0.189	-0.124	-0.026	-0.098
$h = 3$	-0.259	-0.087	-0.169	-0.128	-0.071	-0.046
$h = 6$	-0.325	-0.103	-0.227	-0.140	-0.053	-0.080
$h = 9$	-0.327	-0.088	-0.249	-0.080	-0.008	-0.057
$h = 12$	-0.304	-0.095	-0.222	-0.035	0.004	-0.024
Mean	-0.279	-0.070	-0.202	-0.104	-0.022	-0.063
Panel B: Significance of <i>sent</i>						
$h = 1$	0.367	0.749	0.269	0.440	0.836	0.357
$h = 2$	0.246	0.632	0.207	0.363	0.660	0.319
$h = 3$	0.158	0.446	0.166	0.253	0.426	0.372
$h = 6$	0.017	0.250	0.006	0.040	0.273	0.044
$h = 9$	0.004	0.210	0.000	0.023	0.286	0.013
$h = 12$	0.003	0.121	0.000	0.034	0.210	0.046
Mean	0.179	0.491	0.202	0.3972	0.655	0.490
Panel C: $\Delta adjR^2$						
				0.000	-0.005	-0.003
				0.002	-0.004	0.001
				0.009	-0.001	0.003
				0.033	0.004	0.020
				0.028	0.005	0.020
				0.026	0.011	0.020

Note: The table reports the estimated bootstrap simulation coefficients $\alpha_1^{k(h)}$ and $\beta_1^{k(h)}$ (Panel A) together with the respective p-values for each horizon h (Panel B) from regressing the composite sentiment indicator $sent_t$ on the average future returns on long-short portfolios based on size characteristics – SMB (small-minus-big), SMM (small-minus-medium), and MMB (medium-minus-big). See also the note in Table 2.

Table 4: Holdings of German Equity and Investment Fund Shares

	2013	2014	2015	2016	2017	2018
Panel A: Holdings of equity (in %)						
Residents	61.4	60.1	58.6	58.5	56.4	56.3
Non-residents	38.6	39.9	41.4	41.5	43.6	43.7
Panel B: Holdings of investment fund shares (in %)						
Residents	91.2	91.5	92.1	92.8	93.1	93.4
Non-residents	8.8	8.5	7.9	7.2	6.9	6.6

Source: Deutsche Bundesbank

Table 3 lists the results. Firstly, the pattern of significance is robust to the inclusion of control variables even though the coefficients stemming from the baseline specification are less negative than those from the *naïve* model. Second, the future average returns on the SMB and MMB portfolios for the horizon of six months or more are shown to be negatively affected by increases in *sent*, while the effect on the *h*-period average returns on small-minus-medium portfolios is ambiguous for all the horizons *h*. This suggests that the size story apparent in the results of Baker and Wurgler (2006) can also be applied well to the MMB portfolio, but there is no significant difference in sentiment predictability when it comes to the future returns of the smallest (SDAX) and smaller (MDAX) firms.

Not surprisingly, the term structure of significance in Table 3 mimics the one exhibited in Table 2. Naturally, a null result for the sentiment-return predictability for the horizon of up to three months is not the size story. Nevertheless, the presence of insignificant coefficients for SMB at the horizon of one month contradicts the results established in Baker and Wurgler (2006). Even though the magnitude of the coefficient for SMB for $h = 1$ almost matches the magnitude of the related coefficient in Baker and Wurgler (2006) (-0.2), it is not possible to reject the null of no predictability at a sufficient confidence level. One possible explanation is that it is easier to observe the sentiment effect in extreme characteristic portfolios. In addition, the results in Baker and Wurgler (2006) were exposed to a critique formulated in Sibley et al. (2016), who suggested that their famous composite index (BW) was not “cleaned” enough. As a result, the strong effects of the famous BW sentiment are attributable mainly to economic fundamentals and risk factors.²⁵

6.2 Global and Local Effects of Sentiment

The observation period in this study is greatly affected by globalization, which is characterized by tighter economic and financial integration among countries and regions. The share of equity held by non-residents in Germany increased from 38.6% in 2013 to 43.7% in 2018 (see Table 4).²⁶ Similarly, the share of companies which are domiciled abroad and belong to the main equity indices in Germany - the DAX, the MDAX, and the SDAX – increased during the last decade. Between 2005 and 2018, the share of such firms included in the MDAX and the SDAX increased from 4% and 2% to 10% and 7.1%, respectively (Table 5). Even the main German equity index, the DAX, in 2018 experienced the inclusion of the first company domiciled in a foreign country – Linde plc (Great Britain). It is therefore reasonable to assume that besides country-specific sentiment, sentiment linked to countries with strong financial ties to Germany also affects German equity markets.

²⁵ The average R^2 from the “orthogonalization” equations as described in Section 3 is 0.43, which quite closely matches the prescribed R^2 of the two-factor approach in Sibley et al. (2016).

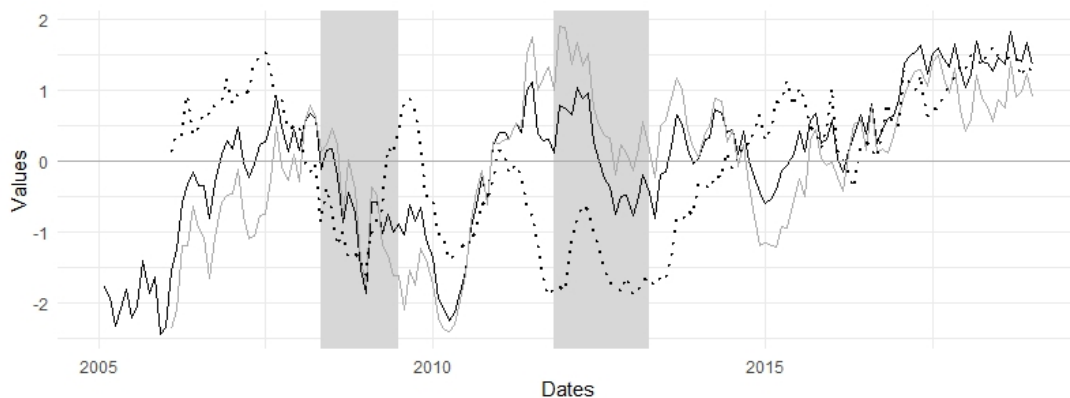
²⁶ Interestingly, a similar pattern is not present for investment fund shares, as the proportion of foreigners holding them decreased from 8.8% to 6.6% between 2013 and 2018.

Table 5: Average Share of Index Members Domiciled Abroad (in %)

	CDAX	DAX	MDAX	SDAX	TDAX
2005–2009	0.02	0.00	4.40	4.00	16.00
2010–2014	0.28	0.00	4.80	4.40	6.70
2015–2019	0.66	1.32	7.26	6.62	9.34

Source: Bloomberg

To examine how irrational beliefs formed in foreign countries affect returns on the German equity indices, I follow the methodology proposed in Baker et al. (2012) and construct a “global” (outside-world) and “local” (German-specific) measure of sentiment. The “global” sentiment measure is formed by aggregating the sentiment series of other countries that are relevant to the German financial market. Nowzohour and Stracca (2020) posit that such a global component exists because the cross-country correlation of sentiment measures is rather high. The “local” sentiment measure results from orthogonalizing *sent* to “global” sentiment. These two measures are then used instead of *sent* in the h -period average returns regression as characterized in the baseline Equation (3).

Figure 2: Total, Local, and Global Sentiment

Note: This figure depicts the dynamics of the “total” composite sentiment indicator – *sent* (the black line), its “local” component (the grey line), and the “global” sentiment indicator using the data for 14 important export destinations for Germany (the dotted line). Shaded areas represent CEPR recession months.

I approximated the economies that might matter for German financial markets by the economies that represent the largest export destinations for German goods. Arguably, companies in these countries form large supply chains with local companies, which might, in the end, be reflected in international trade ties. I assert that this approach might provide more reliable information about irrational beliefs shared globally than the approach of considering only a few selected economies (Baker et al., 2012; Corredor et al., 2013). Since 2005, the same set of countries has formed the list of the top 14 export partners for Germany based on the World Integrated Trade Solution (WITS).²⁷ These are: the United States, France, China, the Netherlands, the United Kingdom, Italy, Austria, Poland, Switzerland, Belgium, Spain, the Czech Republic, the Russian Federation, and Sweden.

²⁷ The cutoff of the first 14 export destinations was selected because the range of a wider set of countries exhibited heterogeneity across the years of the study. See <https://wits.worldbank.org/> for more information on WITS statistics.

Table 6: Global and Local Sentiment

	Sentiment	$\ln R^{CDAX}$	$\ln R^{DAX}$	$\ln R^{MDAX}$	$\ln R^{SDAX}$	$\ln R^{TDAX}$
Panel A: Coefficients on <i>global</i> and <i>local</i> sentiment						
$h = 1$	<i>global</i>	-0.440	-0.420	-0.999	-1.145	0.206
	<i>local</i>	0.452	0.414	1.000	0.805	0.904
$h = 2$	<i>global</i>	-0.670	-0.645	-1.191	-1.315	-0.209
	<i>local</i>	0.921	0.845	1.483	1.157	1.455
$h = 3$	<i>global</i>	-0.798	-0.764	-1.288	-1.392	-0.608
	<i>local</i>	1.150	1.046	1.763	1.268	1.914
$h = 6$	<i>global</i>	-1.076	-1.053	-1.348	-1.634	-0.888
	<i>local</i>	0.852	0.784	0.953	0.911	1.495
$h = 9$	<i>global</i>	-0.968	-0.930	-1.203	-1.415	-0.990
	<i>local</i>	0.479	0.405	0.561	0.518	1.116
$h = 12$	<i>global</i>	-0.887	-0.832	-1.096	-1.208	-1.113
	<i>local</i>	0.375	0.298	0.517	0.366	1.090
Mean	<i>global</i>	-0.807	-0.774	-1.188	-1.352	-0.600
	<i>local</i>	0.705	0.632	1.046	0.838	1.329
Panel B: Significance of <i>global</i> and <i>local</i> sentiment						
$h = 1$	<i>global</i>	0.712	0.745	0.278	0.222	0.726
	<i>local</i>	0.963	0.980	0.622	0.815	0.381
$h = 2$	<i>global</i>	0.380	0.423	0.147	0.114	0.871
	<i>local</i>	0.406	0.461	0.271	0.416	0.098
$h = 3$	<i>global</i>	0.188	0.215	0.073	0.059	0.423
	<i>local</i>	0.186	0.214	0.114	0.279	0.019
$h = 6$	<i>global</i>	0.014	0.011	0.014	0.002	0.086
	<i>local</i>	0.322	0.322	0.527	0.335	0.054
$h = 9$	<i>global</i>	0.006	0.006	0.003	0.001	0.016
	<i>local</i>	0.925	0.973	0.989	0.755	0.158
$h = 12$	<i>global</i>	0.002	0.001	0.000	0.000	0.001
	<i>local</i>	0.781	0.718	0.996	0.941	0.065
Mean	<i>global</i>	0.275	0.267	0.149	0.087	0.349
	<i>local</i>	0.848	0.869	0.759	0.781	0.470
Panel C: $\Delta adj R^2$						
$h = 1$		-0.012	-0.012	-0.009	-0.002	-0.003
$h = 2$		-0.010	-0.010	-0.004	0.007	0.004
$h = 3$		-0.007	-0.008	0.001	0.017	0.006
$h = 6$		0.037	0.039	0.047	0.089	-0.010
$h = 9$		0.065	0.071	0.07	0.107	0.005
$h = 12$		0.070	0.074	0.072	0.095	0.031

Note: The table reports the estimated bootstrap simulation coefficients (Panel A) together with the respective p-values for each horizon h (Panel B) from running the baseline specification as described in Equation (3) in which $sent_t$ is replaced by both “local” and “global” sentiment (see Section 6.2). See also the note in Table 2.

The average cumulative share of these countries in total German exports is more than 70% and the marginal share of all other destinations is less than 2%. As the next step, for each of these countries I constructed a composite measure of its “total” *irrational* sentiment using the same procedure as the one described in Section 3. Due to data limitations on survey-based sentiment measures in such an international sample, I employed only two proxies for sentiment per country: the Business Confidence Index and the Consumer Confidence Index, both extracted from the OECD Statistical Data Warehouse. Further, I incorporated all 14 “total” *irrational* composite sentiment indicators into the principal component analysis in order to produce a “global” measure of population beliefs. Finally, I approximated the “local” measure of German irrational beliefs by the residuals which

result from regressing *sent* (the German analogy of “total” sentiment) on the “global” sentiment measure. Figure 2 depicts all three series.

The results in Table 6 confirm the earlier findings in Baker et al. (2012) and Corredor and Santamaria (2015) for *sent*. Irrational sentiment of the global population seems to drive the negative effect of sentiment reported in Table 2. Moreover, for the smaller capitalization indices, the MDAX and the SDAX, one observes significant estimates of global sentiment even at the horizon of three months, which was not the case before. The estimated coefficients for global sentiment in the case of these two indices are even larger in absolute magnitude than the estimates for future returns on the CDAX and DAX. Besides the size effect already demonstrated in the previous section, one might argue that the share of foreign issuers included in the index also matters. This share is virtually zero for the CDAX and the DAX and around 7% in the case of the other two indices (Table 5).

Nevertheless, the results for the TecDAX – the equity index with the highest share of foreign issuers – erode this assumption. Even though global sentiment proved to be a significant contrarian predictor of its future returns for horizons of six months or more, the magnitude of the mean effect is smaller in absolute terms than the effect for the CDAX and the DAX. Moreover, German-specific (local) population beliefs seem to matter as well. The sign of this effect suggests that local sentiment serves as a momentum signal when it comes to future returns on the technology sector index. This might be because technology firms have become very popular in recent years. The general public might overreact to good news connected to this sector for considerably longer compared to traditional sectors, which further lags the subsequent mean-reversion.

6.3 Robustness Considerations

I performed several tests to confirm the robustness of the results reported in Table 2. First, I constructed two alternative versions of the variable which captures the irrational beliefs of the general public. These alternative indices were then incorporated into Equation (3) instead of the original composite sentiment indicator *sent*. First, following Han and Li (2017), I generated the variable $sent_t^{-1/+1}$ such that it is equal to 1 if the sentiment captured by $sent_t$ is positive and -1 if $sent_t$ is negative. In other words, I treated all the positive observations as strictly positive and all the negative observations as strictly negative ($sent_t$ was constructed such that it has a zero mean and a standard deviation of unity).

Second, I calculated a composite sentiment index, *sentAlt*, which in a sense represents a “cherry-picking” alternative to *sent*. To do so, I employed the following multi-step procedure:

1. The set of 13 irrational sentiment measures (see Section 3) was complemented by the set of 13 *lagged* irrational sentiment proxies that result from regressing the original sentiment variables at time t on the same set of macroeconomic controls as before but now at time $t-1$. The reason for including this additional set of measures is that one might argue that the state of the economy affects the population’s beliefs with a lag; for example, not only current, but also past economic prospects affect sentiment.
2. Because it is reasonable to assume that some of the sentiment variables manifest beliefs that are strongly related to future market returns, while others are not related to stock markets at all, I examined the simple bivariate Granger causality²⁸ between the individual irrational and

²⁸ The Granger-causality test (Granger, 1969) is a simple device for testing whether one time series is useful in forecasting another. Each variable entering the Granger-causality test was tested for the presence of a unit root

lagged irrational sentiment measures and the monthly returns on all five equity indices (see the results in Table A2). Sentiment indices that were found to Granger-cause the returns on at least one stock market index entered the next steps, while the rest were discarded.²⁹

3. Both versions of the irrational counterparts – irrational sentiment and lagged irrational sentiment – of the five survey-based measures were indicated to Granger-cause returns. In order to select only one of them, I ran the auxiliary principal component analysis (PCA) on the set of all variables that displayed Granger-causality for returns and selected only that version of the variables which demonstrated a higher correlation with the resulting first principal component. This scheme yielded a set of nine variables that entered the final step.
4. All nine variables were standardized to have zero mean and unit variance and then incorporated into the PCA. The resulting *alternative* composite survey-based sentiment indicator, *sentAlt*, represents the first principal component of the underlying analysis. It explains 52.1% of the sample variance and its correlation with the indicator from the auxiliary PCA is 99%; therefore, only a little information is lost by omitting the lower-correlation individual sentiment indicators. The following equation characterizes the alternative index:

$$\begin{aligned} sentAlt_t = & 0.897bci_t^{ir,l} + 0.915cci_t^{ir} + 0.950esi_t^{ir,l} + 0.915gri fpbus_t^{ir} + \\ & + 0.763gri fpex_t^{ir} + 0.814gei fombc_t^{ir,l} - 0.127sntmdxh_t^{ir,l} - \\ & - 0.106sntmdxi_t^{ir} - 0.193sntmdxp_t^{ir,l} \end{aligned} \quad (4)$$

where ir,l characterizes the lagged irrational version of sentiment.

Table A3 reports the results of the two robustness estimations. Both $sent^{-1/+1}$ and *sentAlt* exhibit qualitatively and quantitatively the same outcome as the results for the baseline specification in Table 2. The significance of the coefficients for future returns on the TecDAX at horizons of six and nine months in Panel A suggests that the returns of the technology sector exhibit sentiment predictability only when one considers the extreme values of sentiment captured by $sent^{-1/+1}$. It can be concluded that the negative relationship observed between composite sentiment drawn from population surveys and the average future returns on equity indices holds regardless of how the sentiment variable is constructed.

As a next robustness exercise, I examined the effects of a composite sentiment indicator formed of the sentiment proxies that were not orthogonalized prior to the PCA as described in Section 3. This “uncleaned” sentiment – $sent^u$ – represents a common component stemming from the *original* (“uncleaned”) 13 sentiment variables. It explains 51.2% of the sample variance and its correlation with *sent* is 59.3%. The aim of this exercise is to demonstrate that the selection of macroeconomic variables employed in the orthogonalization as described in Section 3 does not drive the results. Conversely, the cleaning procedure was applied to mitigate the usual argument that PCA cannot distinguish between a common component driven by sentiment and a common component driven by the business cycle (Baker and Wurgler, 2006). Table A4 lists the results for $sent^u$. It shows that the estimates for “uncleaned” sentiment – $sent^u$ – are smaller than those reported in Table 2,

using the augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. No sign of a unit root was detected in any of the variables. The resulting statistics of the unit-root tests and the F-statistics from testing the null hypothesis of sentiment NOT Granger-causing returns are not reported but are available upon request.

²⁹ The cut-off rule of at least one Granger-causing relation between sentiment and return turned out to be a robust criterion. I also constructed the index with a cut-off value of two Granger-causing relations, but the results did not change significantly.

but in general they are very similar. This outcome suggests that even though the orthogonalization removed a portion of the fundamental factors embedded in the individual proxies, these fundamental factors play a minor role when one examines the effects of composite sentiment.

The variable *sent*, or the irrational beliefs of the general public, is constructed as the first principal component from running the PCA on the set of 13 sentiment proxies. The loadings listed in Equation (1), however, suggest that there might be a difference between the sentiment revealed by households or company management and the sentiment expressed by the expert community, for example, investors and bankers. On the one hand, *sent* is driven predominantly by the sentiment embedded in consumer and business surveys; hence, it is reasonable to refer to *sent* as the sentiment of the general public. On the other hand, *sent* is represented by a factor which captures 46.7% of all the variance explained by the 13 sentiment variables, and this leaves quite a lot of room for the second principal component. Moreover, this second principal component might correspond to the sentiment which is common among experts and, as such, it might offer supporting evidence for the overreaction hypothesis. In line with this motivation, I performed an additional robustness test in which I supplemented the *sent* in Equation (3) with the first two principal components stemming from running the PCA on the set of 13 irrational sentiment measures and applying varimax rotation.³⁰ The two components explain cumulatively 74.1% of the sample variance. Not surprisingly, the first principal component closely resembles *sent*, and their pairwise correlation is 93%. The second principal component is then factored by those individual sentiment proxies which showed the opposite sign to the loadings in *sent*. As a result, I refer to these two indices as the sentiment of *the general public* and the sentiment of *experts*, respectively.

The results of the underlying robustness test are reported in Table 7. In general, they confirm the outcome of my baseline specification, that is, the irrational beliefs of the general public affect future returns negatively and this effect is most significant for longer horizons. In addition, the effect is robust even when the sentiment of experts is included in the regressions. The effects of the sentiment of experts have a positive sign for all the horizons studied, but the significance of the estimates seems rather inconclusive across the equity indices. Naturally, and in line with the overreaction hypothesis (Barberis et al., 1998) and the shaky and firm hands hypothesis (Kostolany, 1990), the positive coefficients suggest that experts are able to forecast future movements of equity markets correctly and act as rational arbitragers in the economy. These results are comparable to those reported in Schmeling (2007) for the sentiment of private and institutional investors.

Lastly, I looked at the performance of the original composite sentiment index, *sent*, in subsamples. Specifically, I divided the sample into two equal halves. Subsample 1 (Table A5, Panel A) captures the period from 2005 to 2011 and, as such, encompasses the whole Global Financial Crisis (GFC).³¹ Subsample 2 (Table A5, Panel B) spans the period from 2012 to 2018. Studies such as Gadea Rivas and Perez-Quiros (2015) and Duprey and Klaus (2017) suggest that most of the in-sample predictive performance of leading indicators is implied by including observations linked to the GFC. Consequently, my aim is to show that the inclusion of these data points does not alter the main findings.

³⁰ Varimax rotation (Kaiser, 1958) is the most common method used to obtain a clear pattern of loadings. The rotation itself does not affect the sum of the eigenvalues but does affect the eigenvalues of particular factors and changes the factor loadings.

³¹ I also experimented with dividing the sample into 2005–2012 and 2013–2018 subsamples so that the first subsample contains the major portion of the European Debt Crisis. The results are qualitatively and quantitatively similar.

Table 7: Diverse Effects of Public and Expert Sentiment

	Sentiment	$\ln R^{CDAX}$	$\ln R^{DAX}$	$\ln R^{MDAX}$	$\ln R^{SDAX}$	$\ln R^{TDAX}$
Panel A: Coefficients on public and expert sentiment						
$h = 1$	<i>general public</i>	-0.405	-0.436	-0.428	-0.450	0.646
	<i>experts</i>	0.530	0.401	1.329	1.568	0.459
$h = 2$	<i>general public</i>	-0.296	-0.311	-0.395	-0.358	0.630
	<i>experts</i>	0.558	0.458	1.054	1.605	0.533
$h = 3$	<i>general public</i>	-0.232	-0.242	-0.313	-0.375	0.676
	<i>experts</i>	0.656	0.543	1.237	1.655	0.826
$h = 6$	<i>general public</i>	-0.495	-0.469	-0.668	-0.821	0.147
	<i>experts</i>	0.585	0.497	0.931	1.168	0.495
$h = 9$	<i>general public</i>	-0.486	-0.465	-0.614	-0.784	0.058
	<i>experts</i>	0.405	0.338	0.674	0.685	0.274
$h = 12$	<i>general public</i>	-0.349	-0.333	-0.446	-0.618	0.182
	<i>experts</i>	0.387	0.327	0.594	0.584	0.179
Mean	<i>general public</i>	-0.377	-0.376	-0.477	-0.591	0.342
	<i>experts</i>	0.518	0.427	0.97	1.211	0.461
Panel B: Significance of public and expert sentiment						
$h = 1$	<i>general public</i>	0.281	0.217	0.325	0.251	0.463
	<i>experts</i>	0.267	0.384	0.045	0.007	0.313
$h = 2$	<i>general public</i>	0.331	0.286	0.283	0.257	0.571
	<i>experts</i>	0.209	0.293	0.054	0.003	0.275
$h = 3$	<i>general public</i>	0.311	0.262	0.236	0.133	0.510
	<i>experts</i>	0.146	0.220	0.018	0.001	0.107
$h = 6$	<i>general public</i>	0.012	0.013	0.003	0.001	0.482
	<i>experts</i>	0.074	0.112	0.012	0.000	0.137
$h = 9$	<i>general public</i>	0.001	0.000	0.000	0.000	0.165
	<i>experts</i>	0.067	0.110	0.014	0.005	0.170
$h = 12$	<i>general public</i>	0.001	0.000	0.000	0.000	0.200
	<i>experts</i>	0.052	0.096	0.013	0.008	0.159
Mean	<i>general public</i>	0.296	0.297	0.196	0.110	0.756
	<i>experts</i>	0.325	0.435	0.125	0.052	0.453
Panel C: $\Delta adj R^2$						
$h = 1$		-0.005	-0.006	0.009	0.024	-0.006
$h = 2$		-0.009	-0.010	0.006	0.033	-0.011
$h = 3$		-0.007	-0.010	0.018	0.049	-0.003
$h = 6$		0.022	0.019	0.054	0.083	-0.027
$h = 9$		0.034	0.033	0.062	0.082	-0.033
$h = 12$		0.023	0.021	0.050	0.066	-0.050

Note: The table reports the estimated bootstrap simulation coefficients (Panel A) together with the respective p-values for each horizon h (Panel B) from running the baseline specification as described in Equation (3) in which $sent_t$ is replaced by the first two principal components (PCs) stemming from running the PCA on the irrational parts of the 13 sentiment proxies. The first PC represents the sentiment of the *general population*, while the second PC mimics *experts'* sentiment. See also the note in Table 2.

On the qualitative side of the story, the results show that the direction of the effect is stable over time for all the estimates in Subsample 2. In the case of Subsample 1, this holds only for R^{SDAX} as a whole and for “longer-horizon” estimates in the case of the returns on the other equity indices. However, the quantitative side suggests some interesting aspects. Firstly, the coefficient estimates for the period until 2011 are, in general, less negative than the estimates for the whole sample. Moreover, one even observes a significantly positive relationship for some horizons for the CDAX, the DAX (the 9-month horizon), the MDAX (the 6-month horizon), and the TDAX (the 12-month

horizon). Subsample 1 contains the financial crisis, which is accompanied by a rather sharp decline in optimism after its recovery following the dot.com bubble (see Figure 1), and these extreme events in both the sentiment and return measures probably erode the stability of part of the baseline results. In fact, this pattern supports the findings in Cheema et al. (2018), who argue that the positive effect of sentiment estimated in Han and Li (2017) at horizons of up to nine months is attributable to the boom and bust period of 2006–2008.

Secondly, all the estimates for the second subsample are negative, and the power of this negative sign is also more pronounced in all cases compared to the results in Table 2. The returns on the first four equity indices even display significance for periods that have not been indicated as significant before. Note that the period under examination is 2012–2018, which, in general, exhibits a consistent pattern of increasing optimism and economic growth. Hence, the result for Subsample 2 suggests that the population’s irrational beliefs proxied by *sent* might affect average future returns on equity markets conditional on the stage of the business cycle (Moller et al., 2014).

7. Comparison of the Composite Survey-Based Sentiment Indicator with Consumer Confidence: Evidence from Out-Of-Sample Forecasts

To underline the degree of improvement achieved by combining several survey-based measures into one composite sentiment indicator, I decided to compare the out-of-sample capacity to predict future returns of the composite index *sent* with the capacity of the Consumer Confidence Indicator (CCI) calculated for Germany by DG ECFIN. The CCI was selected as the benchmark for the evaluation of forecasting performance because it is a popular proxy measure in academic circles and there are several studies that show that consumer confidence is a significant predictor of future stock market returns (Lemmon and Portniaguina, 2006; Schmeling, 2009; Chang et al., 2011; Corredor and Santamaria, 2015; Perez-liston et al., 2018). Prior to the analysis, the CCI measure was transformed to have a zero mean and unit variance.³² This adjusted consumer confidence, *cci*, was then used in the final analysis. Besides the out-of-sample evidence, I also used *cci* in the long-term return regressions outlined in Equations (2) and (3). The results of these regressions can be found in Table A6 in the Appendix. This in-sample analysis shows weaker effects of *cci* compared to *sent*, which suggests that the effects of the wide-range population beliefs embedded in the composite sentiment indicator on future returns are indeed stronger than the effects of consumer confidence.

I employed the recursive estimation (RE) scheme for the out-of-sample analysis, similarly to Welch and Goyal (2008), Huang et al. (2014), and Han and Li (2017). Within this framework, I used the period between January 2005 and April 2013 as the startup sample of $s = 100$ observations to initiate the model and obtain the first parameters needed for the forecasts of period $t + 1$. The remaining sample of $p = (T - s) = 74$ months was used for the out-of-sample analysis. The selection of the startup period is a discretionary task. Nevertheless, the trade-off between having a sufficient number of observations to run a reliable estimation and having a sufficient number of out-of-sample

³² The measure was not orthogonalized using the methodology described in Section 3. Sibley et al. (2016) showed that the sentiment predictability observed in Baker and Wurgler (2006) is mainly attributable to the fact that their BW sentiment index was not “cleaned” enough. Moreover, Moller et al. (2014) showed that the distribution of expected returns is better captured by fundamental risk factors than by consumer confidence. Therefore, any forecasting improvement achieved by *sent* over *cci* would mean that *sent* is a more powerful predictor of future equity returns than consumer confidence. Nevertheless, I also run the out-of-sample forecasting exercise for consumer confidence orthogonalized to fundamental news, and the results (available upon request) are qualitatively the same.

forecasts to perform the evaluation should be considered. Given the relatively short length of the sample employed in this study, my choice for s seems to balance the two considerations.

To build the sample of the forecast for the out-of-sample evaluation, I used the single-factor return regression model specified in Equation (2). The horizon of one month ($h = 1$) is selected as the forecasting period. Hence, for each out-of-sample period p , I forecast the future 1-month average returns on stock market index k as

$$E(R_{t+1}^k) = \hat{\alpha}_{0,t}^k + \hat{\alpha}_{1,t}^k S_t \quad (5)$$

where S_t represents either the composite indicator $sent_t$ or the scaled German consumer confidence indicator cci_t . To avoid the forward-looking bias that might arise in both measures from the “orthogonalization” procedure and subsequent PCA in the case of $sent$ and from scaling in the case of cci , both variables are estimated recursively each time using real-time information up to period t . $\hat{\alpha}_{0,t}^k$ and $\hat{\alpha}_{1,t}^k$ are the regression coefficients that were obtained from the bootstrap simulation using the information up to t . The bootstrap simulation was performed with 5,000 repetitions each time the model was re-estimated (see Appendix B for more details on the bootstrap simulation procedure).

To evaluate the results, I employed the normalized mean absolute error (NMAE) measure for each stock market index k as follows:

$$NMAE^k = \frac{\sum_{t=p}^{T-11} |R_{t+1}^k - E(R_{t+1}^k)|}{R_H^k - R_L^k} \quad (6)$$

where R_H^k and R_L^k represent, respectively, the highest and the lowest value of the monthly returns for index k during the entire sample of observations. I decided to normalize the mean absolute error (MAE) because this transformation ensures scale independence, which is of great benefit considering the design of the present analysis. Nevertheless, the MEA could be easily obtained by multiplying the NMEA by the respective value of $R_H^k - R_L^k$.³³ Quantitatively similar results are obtained from the mean square forecast error (MSFE) metrics.

Moreover, following Huang et al. (2014) and Han and Li (2017) I also compare the forecasts resulting from Equation (5) with the historical average benchmark that corresponds to the constant expected return model.³⁴ The historical average takes into account the return series from the beginning up to month t . To evaluate the out-of-sample performance of $sent$ and cci compared to the historical benchmark, I employ the Campbell and Thompson (2008) R_{os}^2 statistic, which measures the proportional reduction in the MSFE for the predictive regression forecast relative to the historical average benchmark. In addition, using the modified Diebold and Mariano (1995) test proposed by Harvey et al. (1997) – the DM test – I test the null that the predictive regression forecast and the historical average benchmark have the same predictive accuracy against the one-sided alternative

³³ The return ranges $range^k = R_H^k - R_L^k$ for indices k are: $range^{CDAX} = 35.86$, $range^{DAX} = 36.81$, $range^{MDAX} = 46.33$, $range^{SDAX} = 35.61$, and $range^{TDAX} = 45.36$

³⁴ Goyal and Welch (2008) showed that the typical economic variables usually fail to outperform the historical average in out-of-sample analysis.

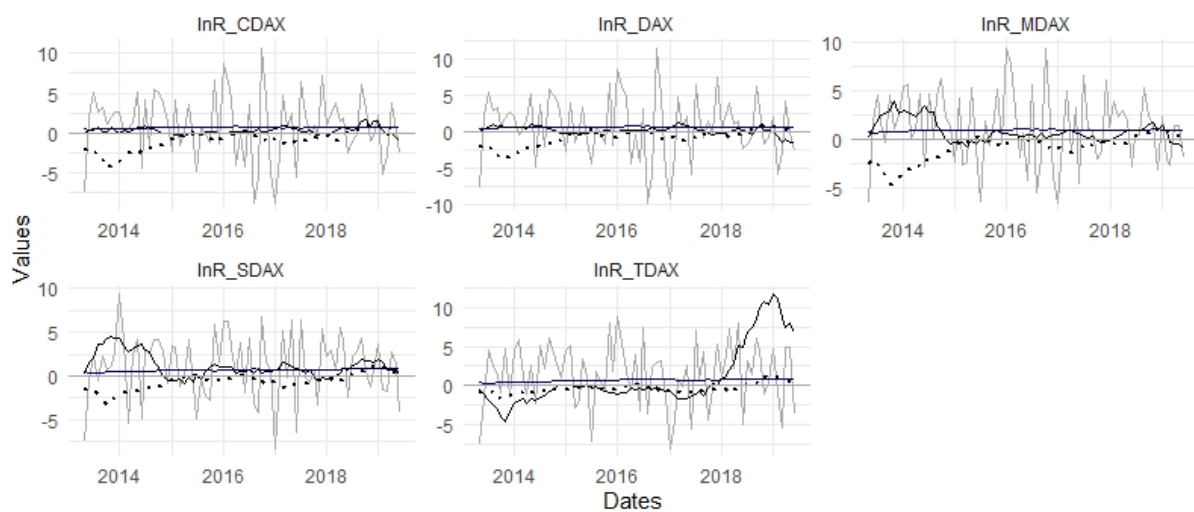
that the historical benchmark is *less* accurate than the out-of-sample forecast obtained from running Equation (5). I further employ both R_{os}^2 and the DM test to evaluate the accuracy of *sent* compared to *cci*, that is, I treat the predictive regression forecast obtained for *cci* as a benchmark.

Table 8: Out-of-Sample Forecasting Results

<i>Dep.Var.</i>	<i>Sentiment</i>	<i>NMAE</i>	Historical Average Benchmark		<i>cci</i> Benchmark	
			R_{os}^2	<i>DMtest</i>	R_{os}^2	<i>DMtest</i>
Panel A: Results for <i>sent</i>						
R^{CDAX}	<i>sent</i>	0.088	-0.009	0.29	0.267	-3.60***
R^{DAX}	<i>sent</i>	0.089	0.008	-0.23	0.228	-3.27***
R^{MDAX}	<i>sent</i>	0.062	0.038	-0.64	0.389	-3.43***
R^{SDAX}	<i>sent</i>	0.081	0.008	-0.09	0.248	-1.86***
R^{TDAX}	<i>sent</i>	0.099	-0.879	3.46	-0.581	2.67
Panel B: Results for <i>cci</i>						
R^{CDAX}	<i>cci</i>	0.108	-0.377	3.30		
R^{DAX}	<i>cci</i>	0.107	-0.285	2.99		
R^{MDAX}	<i>cci</i>	0.080	-0.573	3.52		
R^{SDAX}	<i>cci</i>	0.099	-0.319	2.86		
R^{TDAX}	<i>cci</i>	0.081	-0.189	2.82		

Note: The table reports the normalized mean absolute error (NMAE) metric as defined in Equation (6) for either the composite sentiment indicator *sent* (Panel A) or the scaled consumer confidence indicator for Germany *cci* (Panel B). Individual NMAEs are produced for each stock market index k at a forecasting horizon of one month ($h = 1$) using the recursive estimation out-of-sample framework. R_{os}^2 represents the statistic proposed in Campbell and Thompson (2008) and DM test represents the modified Diebold and Mariano (1995) test proposed by Harvey et al. (1997). ***, **, and * denote the 1%, 5%, and 10% significance levels, respectively, for the DM test under the null that the historical average benchmark (columns 4–5) or *cci* (the last two columns) has a smaller or equivalent MSFE compared to *sent*.

Figure 3: Equity Return Forecasts of *sent* and *cci* – May 2013 to May 2019



Note: This figure depicts for each equity index k the realized monthly returns (the grey solid line), the out-of-sample predictive regression forecast produced by the recursively constructed *sent* (the black solid line), and the recursively scaled consumer confidence index *cci* (the black dotted line). The historical average benchmark is depicted by the blue solid line.

Table 8 summarizes the outcome of the underlying analysis, while Figure 3 depicts the out-of-sample forecast series together with the realized monthly returns and the historical average benchmark. Firstly, based on the NMAE, the composite sentiment measure derived in this paper outperforms *cci* in the out-of-sample predictive regression forecasts of all the German equity returns except for the returns on the TecDAX. However, R_{os}^2 , which shows how much the MSFE of the predictive regression forecast is reduced compared to the historical average benchmark, only provides support for returns on the traditional equity indices: the DAX, the MDAX, and the SDAX. Moreover, the increased accuracy of *sent* compared to the historical average benchmark was not confirmed by the DM test for all the return series studied. The same holds for the out-of-sample predictive regression forecast produced by *cci*. The last two columns in Panel A of Table 8 show interesting results. Based on the R_{os}^2 statistic, which now considers the predictive regression forecast produced by *cci* as a benchmark, *sent* has much stronger out-of-sample predictive ability for equity market returns. However, the strong significance of the DM test statistic is attributable mainly to the high correlation between the forecasting errors of the two series (almost 81% on average).

To sum up, the out-of-sample analysis yields an important implication. It suggests that the popular measure of sentiment, the consumer confidence indicator, which has been employed by numerous authors as a valid and significant predictor of future stock market returns, could be improved by considering information from other available surveys conducted on the international level.

8. Conclusion

This paper contributes to the literature examining the relationship between sentiment and stock market returns. A wide range of survey-based sentiment measures available for Germany were employed to form a novel composite sentiment index that captures the irrational beliefs of the general public about the future state of the German equity markets. Each sentiment variable was cleaned of macroeconomic factors prior to entering the composite index to ensure that the inherent information mimics sentiment and not extrapolated good or bad economic prospects.

The resulting sentiment indicator was used in h -period average return regressions with the goal to estimate its predictive power for future returns on the aggregate German equity market and on more narrow segments of that market 1, 2, 3, 6, 9, and 12 months ahead. The analysis showed that the irrational beliefs commonly embodied in surveys of the population negatively affect future returns on four out of the five equity indices considered in the analysis but that this effect is significant only at horizons of six months or more. This outcome was verified by several robustness tests. Although the impact of sentiment in the shorter term remained negligible, the outcome of negative predictability at longer horizons supports the overreaction hypothesis (Barberis et al., 1998) and also verifies the results obtained in related studies. More importantly, it shows that the sentiment of the general public as manifested in a variety of surveys, many of which are not primarily intended to capture market sentiment, serves as a reliable predictor of future market returns.

Next, an analysis of the cross-sectional effects of public sentiment showed that its in-sample predictability is strongly linked to the equity size effect. Even though this tendency has already been demonstrated in Baker and Wurgler (2006), the results in the present paper suggest that the effect is not evident immediately at horizons of one to three months but is rather lagged. A complementary analysis of the global and local effects of sentiment à la Baker et al. (2012) showed a similar pattern. The sentiment of the general public does indeed spill over from the global market (approximated by the main export destinations for German goods), but the significance of this effect is observed with a lag of three to six months depending on the equity market considered.

Moreover, local sentiment was shown to have near zero predictive power for returns on almost all the stock indices. Therefore, similarly to Corredor and Santamaria (2015), one can conclude that the majority of the impact estimated for the sentiment-return relation is attributable to “a variable with components reaching beyond the domestic market environment.”

Last, an out-of-sample forecasting framework was established to evaluate the benefits of the wide-range survey-based composite index compared to the popular individual sentiment measure – the consumer confidence indicator (CCI) – at a forecasting horizon of one month. The results of this analysis showed that the composite survey-based sentiment measure provides a more accurate forecast of returns than the CCI for almost all the equity indices considered. This suggests that the well-known consumer confidence indicator might not be the immediate choice when searching for a reliable proxy for sentiment and that information from other available surveys might enrich its forecasting performance for stock market returns. This finding might be of interest not only to individual investors who are building their trading strategies, but also to central authorities who are concerned with identifying the causes of systemic events or quantifying risks to financial stability.

The underlying analysis covers the dynamics of the equity market in Germany, the largest European economy and the most important trading partner for the majority of EU countries. The period under observation is characterized by increasing economic and financial integration among countries and regions. Moreover, as discussed in Section 2.1, empirical findings related to the international spillover effects of sentiment and its role in systemic events suggest that the effects of sentiment in one country might well be observed abroad. As such, apart from the direct implications for German financial markets, the results of this study also have implications for financial market performance and financial stability policy in the surrounding European economies. Nevertheless, it has to be acknowledged that the power of the sentiment-return relation is subject to cross-country heterogeneity (Schmeling, 2009; Corredor and Santamaria, 2015),³⁵ so my results might not be directly applicable to countries other than Germany.

³⁵ Schmeling (2009) showed that the strength of the sentiment-return relation depends negatively on market integrity and positively on behavioral biases which prevail in the economy. Similarly, Corredor and Santamaria (2015) suggest that the degree of stock market openness might cause cross-country heterogeneity of sentiment effects.

References

- AISSIA, D. B. (2016): “Home and Foreign Investor Sentiment and the Stock Returns.” *Quarterly Review of Economics and Finance*, 59:71–77.
- AKERLOF, G. A. AND R. J. SHILLER (2010): *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton university press.
- AMIHUD, Y. AND C. M. HURVICH (2004): “Predictive regressions: a reduced-bias estimation method.” *Journal of Financial and Quantitative Analysis*, 39(4):813–841.
- AMIHUD, Y., C. M. HURVICH, AND Y. WANG (2009): “Multiple-predictor regressions: Hypothesis testing.” *Review of Financial Studies*, 22(1):413–434.
- BAKER, M. AND J. C. STEIN (2004): “Market liquidity as a sentiment indicator.” *Journal of Financial Markets*, 7:271–299.
- BAKER, M. AND J. WURGLER (2006): “Investor sentiment and the cross-section of stock returns.” *The Journal of Finance*, 61(4):1645–1680.
- BAKER, M. AND J. WURGLER (2007): “Investor sentiment in the stock market.” *Journal of Economic Perspectives*, 21(2):129–151.
- BAKER, M., J. WURGLER, AND Y. YUAN (2012): “Global, local, and contagious investor sentiment.” *Journal of Financial Economics*, 104(2):272–287.
- BARBERIS, N., A. SHLEIFER, AND R. VISHNY (1998): “A model of investor sentiment.” *Journal of Financial Economics*, 49:307–343.
- BARONE-ADESI, G., L. MANCINI, AND H. SHEFRIN (2012): “Sentiment, Asset Prices, and Systemic Risk.” *Swiss Finance Institute Research Paper*, (11-50).
- BARSKY, R. B. AND E. R. SIMS (2012): “Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence.” *American Economic Review*, 102(4):1343–77.
- BLOOMBERG (2020): “The Robinhood Craze Is Now Moving Stocks Everywhere.” Available at (<https://www.bloombergquint.com/markets/robinhood-craze-born-in-america-is-now-moving-stocks-everywhere>). Accessed: 16.12.2020
- BOROVKOVA, S., E. GARMAEV, P. LAMMERS, AND J. RUSTIGE (2017): “SenSR: A Sentiment-Based Systemic Risk Indicator.” DNB Working Paper No. 553, De Nederlandsche Bank
- BRENNER, M. AND D. GALAI (1989): “New Financial Instruments for Hedge Changes in Volatility.” *Financial Analysts Journal*, 45(4):61–65.
- BROWN, G. W. AND M. T. CLIFF (2004): “Investor Sentiment and the Near-Term Stock Market.” *Journal of Empirical Finance*, 11(1):1–27.
- BROWN, G. W. AND M. T. CLIFF (2005): “Investor Sentiment and Asset Valuation.” *The Journal of Business*, 78(2):405–440.
- BROWNLEES, C. AND R. F. ENGLE (2017): “SRISK: A Conditional Capital Shortfall Measure of Systemic Risk.” *The Review of Financial Studies*, 30(1):48–79.
- CAMERON, A. C. AND P. K. TRIVEDI (2005): *Microeconometrics: Methods and Application*. Cambridge University Press, New York, 1st edition.

- CAMPBELL, J. AND S. THOMPSON (2008): “Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies*, 21(1):1509—1531.
- CHANG, Y. Y., R. FAFF, AND C.-Y. HWANG (2011): “Local and Global Sentiment Effects , and the Role of Legal , Information and Trading Environments.” *Trading and Information Environments*.
- CHEEMA, M. A., Y. MAN, AND K. R. SZULCZYK (2018): “Does Investor Sentiment Predict the Near-Term Returns of the Chinese Stock Market?” *International Review of Finance*, 1–9.
- CHU, X., C. WU, AND J. QIU (2016): “A Nonlinear Granger Causality Test between Stock Returns and Investor Sentiment for Chinese Stock Market: A Wavelet-Based Approach.” *Applied Economics*, 48(21):1915–1924.
- CLARKE, R. G. AND M. STATMAN (1998): “Bullish or Bearish?” *Financial Analysts Journal*, 54 (3):63–72.
- CONCETTO, C. L. AND F. RAVAZZOLO (2019): “Optimism in Financial Markets: Stock Market Returns and Investor Sentiments.”
- CORREDOR, P. AND R. SANTAMARIA (2015): “The Impact of Investor Sentiment on Stock Returns in Emerging Markets: The Case of Central European Markets.” *Eastern European Economics*.
- CORREDOR, P., E. FERRER, AND R. SANTAMARIA (2013): “Investor Sentiment Effect in Stock Markets: Stock Characteristics or Country-Specific Factors?” *International Review of Economics and Finance*, 27:572–591.
- DE LONG, J. B., A. SHLEIFER, AND L. H. SUMMERS (1990): “Positive Feedback Investment Strategies and Destabilizing Rational Speculation.” *Journal of Finance*, 45(2):379–395.
- DE LONG, J. B., A. SHLEIFER, L. H. SUMMERS, AND R. J. WALDMANN (1990): “Noise Trader Risk in Financial Markets.” *Journal of Political Economy*, 98(4):703–738.
- DIEBOLD, F. AND R. MARIANO (1995): “Comparing Predictive Accuracy.” *Journal of Business and Economic Statistics*, 13:253—263.
- DOW, S. C. (2011): “Cognition, Market Sentiment and Financial Instability.” *Cambridge journal of economics*, 35(2):233–249.
- DUNNE, P. G., J. FORKER, A. ZHOLOS, ET AL. (2011): “The Value Relevance of Sentiment.” Research Technical Paper 05/RT/11, Central Bank of Ireland
- DUPREY, T. AND B. KLAUS (2017): “How to Predict Financial Stress? An Assessment of Markov Switching Models.” ECB Working Paper No. 2057, European Central Bank
- EASLEY, D. AND M. O’HARA (1991): “Order Form and Information in Securities Markets.” *The Journal of Finance*, 46(3):905–927.
- ERGUNGOR, O. E. (2016): “Where the Wild Things Are: Measuring Systemic Risk through Investor Sentiment.” Working Paper 16-08R, Federal Reserve Bank of Cleveland
- EUROPEAN COMMISSION (2007): “Joint Harmonised EU Programme of Business and Consumer Surveys: User Guide.”
- FAMA, E. F. AND K. R. FRENCH (1993): “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, 33:5–56.

- FERSON, W. E., S. SARKISSIAN, AND T. T. SIMIN (2003): “Spurious Regressions in Financial Economics?” *The Journal of Finance*, 58(4):1393–1413.
- FISHER, K. L. AND M. STATMAN (2000): “Investor Sentiment and Stock Returns.” *Financial Analysts Journal*, 56(2):16–23.
- GADEA RIVAS, M. D. AND G. PEREZ-QUIROS (2015): “The Failure to Predict the Great Recession – a View Through the Role of Credit.” *Journal of the European Economic Association*, 13(3):534–559.
- GOYAL, A. AND I. WELCH (2008): “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *The Review of Financial Studies*, 21(4):1455—1508.
- GRANGER, C. W. J. (1969): “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica*, 37(3):424–438.
- GREENE, W. (2012): *Econometric analysis*, number 7. of., Upper Saddle River, 7 edition.
- HAN, X. AND Y. LI (2017): “Can Investor Sentiment Be a Momentum Time - Series Predictor? Evidence from China.” *Journal of Empirical Finance*, 42:212–239.
- HANSEN, L. P. AND R. J. HODRICK (1980): “Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis.” *Journal of Political Economy*, 88(5): 829–853.
- HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): “Testing the Equality of Prediction Mean Squared Errors.” *International Journal of Forecasting*, 13(2):281—291.
- HENGELBROCK, J., E. THEISSEN, AND C. WESTHEIDE (2013): “Market Response to Investor Sentiment.” *Journal of Business Finance and Accounting*, 40(7-8):901–917.
- HODRICK, R. J. (1992): “Dividend Yields and Stock Returns: Alternative Procedures for Inference and Measurement.” *Review of Financial Studies*, 5(3):357–386.
- HODULA, M., S. MALOVANÁ, AND J. FRAIT (2019): “Introducing a New Index of House-holds’ Macroeconomic Conditions.” CNB Working Paper No. 10/2019, Czech National Bank, Prague.
- HUANG, D., F. JIANG, J. TU, AND G. ZHOU (2014): “Investor Sentiment Aligned: A Powerful Predictor of Stock Returns.” *Review of Financial Studies*, 28(3):791–837.
- JAGADESH, N. AND S. TITMAN (2001): “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations.” *The Journal of Finance*, 56(2):699–720.
- KAHNEMAN, D. AND A. TVERSKY (1974): “Judgment under Uncertainty: Heuristics and Biases.” *Science (New York, N.Y.)*, 185(4157):1124–1131.
- KAISER, H. F. (1958): “The varimax criterion for analytic rotation in factor analysis.” *Psychometrika*, 23(3):187–200.
- KOSTOLANY, A. (1990): *Börsenseminar*. Heyne Verlag, München.
- KUROV, A. (2010): “Investor Sentiment and the Stock Market’s Reaction to Monetary Policy.” *Journal of Banking and Finance*, 34(1):139–149.
- LEMMON, M. AND E. PORTNIAGUINA (2006): “Consumer Confidence and Asset Prices: Some Empirical Evidence.” *Review of Financial Studies*, 19(4):1499–1529.

- MACKINNON, J. (2007): “Bootstrap Hypothesis Testing.” Queen’s Economics Department Working Paper No. 1127
- MARK, B. N. C. (1995): “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability.” *The American Economic Review*, 85(1):201–218.
- MARTIN, A. AND J. VENTURA (2011): “Theoretical Notes on Bubbles and the Current Crisis.” *IMF Economic Review*, 59(1):6–40.
- MINSKY, H. P. (1977): “The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to “Standard” Theory.” *Challenge*, 20(1):20–27.
- MOLLER, S. V., H. NORHOLM, AND J. RANGVID (2014): “Consumer Confidence or the Business Cycle: What Matters More for European Expected Returns?” *Journal of Empirical Finance*, 28:230—248.
- NEWKEY, W. K. AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55(3):703–708.
- NOWZOHOUR, L. AND L. STRACCA (2020): “More than a Feeling: Confidence, Uncertainty, and Macroeconomic Fluctuations.” *Journal of Economic Surveys*, 34(4):691–726.
- OLIVEIRA, N., P. CORTEZ, AND N. AREAL (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.
- OSLER, C. AND T. SAVASER (2011): “Extreme Returns: The Case of Currencies.” *Journal of Banking and Finance*, 35(11):2868–2880.
- PEREZ-LISTON, D., D. HUERTA-SANCHEZ, AND J. GUTIERREZ (2018): “Do Domestic Sentiment and the Spillover of US Investor Sentiment Impact Mexican Stock Market Returns?” *Journal of Emerging Market Finance*, 17(2):185–212.
- RAKOVSKA, Z. AND M. SVOBODA (2016): “Practical Application of Sentiment Indicators in Financial Analysis.” In Krajíček, J., J. Nešleha, and K. Urbanovský, editors, *In Proceedings of European Financial Systems 2016*, Brno. Masaryk University.
- RUTKOWSKA, A. (2017): “The influence of investor sentiment on sector indices - the Ini index analysis.” *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (482):227–239.
- SAUER, S. AND K. WOHLRABE (2018): “The new ifo Business Climate Index for Germany.” *CESifo Forum*, 19(2):59–64.
- SAVASER, T. (2011): “Exchange Rate Response to Macronews: Through the Lens of Microstructure.” *Journal of International Financial Markets, Institutions and Money*, 21(1):107–126.
- SCHMELING, M. (2007): “Institutional and Individual Sentiment: Smart Money and Noise Trader Risk?” *International Journal of Forecasting*, 23(1):127–145.
- SCHMELING, M. (2009): “Investor Sentiment and Stock Returns: Some International Evidence.” *Journal of Empirical Finance*, 16(3):394–408.
- SCHMITZ, P., M. GLASER, AND M. WEBER (2007): “Individual investor sentiment and stock returns - What do we learn from warrant traders?” *Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung*, 6:1–36.

- SHEFRIN, H. (2016): *Behavioral Risk Management: Managing the Psychology that Drives Decisions and Influences Operational Risk*. Palgrave Macmillan.
- SHLEIFER, A. AND L. H. SUMMERS (1990): “The Noise Trader Approach to Finance.” *Journal of Economic Perspectives*, 4(2):19–33.
- SIBLEY, S. E., Y. WANG, Y. XING, AND X. ZHANG (2016): “The Information Content of the Sentiment Index.” *Journal of Banking and Finance*, 62:164–179.
- SMALES, L. A. (2014): “News Sentiment and Measures of Bank Credit Risk.”
- STAMBAUGH, R. F. (1999): “Predictive Regressions.” *Journal of Financial Economics*, 54(3): 375–421.
- TANG, D. Y. AND H. YAN (2010): “Market Conditions, Default Risk and Credit Spreads.” *Journal of Banking & Finance*, 34(4):743–753.
- THALER, R. H. AND C. R. SUNSTEIN (2008): *Nudge: Improving Decisions about Health, Wealth, and Happiness*, volume 1 of., New Haven & London.
- VERMA, R. AND G. SOYDEMIR (2006): “The Impact of U . S . Individual and Institutional Investor Sentiment on Foreign Stock Markets.” *The Journal of Behavioral Finance*, 7(3): 128–144.
- WELCH, I. AND A. GOYAL (2008): “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *Review of Financial Studies*, 21(4):1455–1508.
- ZELLNER, A. (1962): “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias.” *Journal of the American Statistical Association*, 57(298): 348–368.

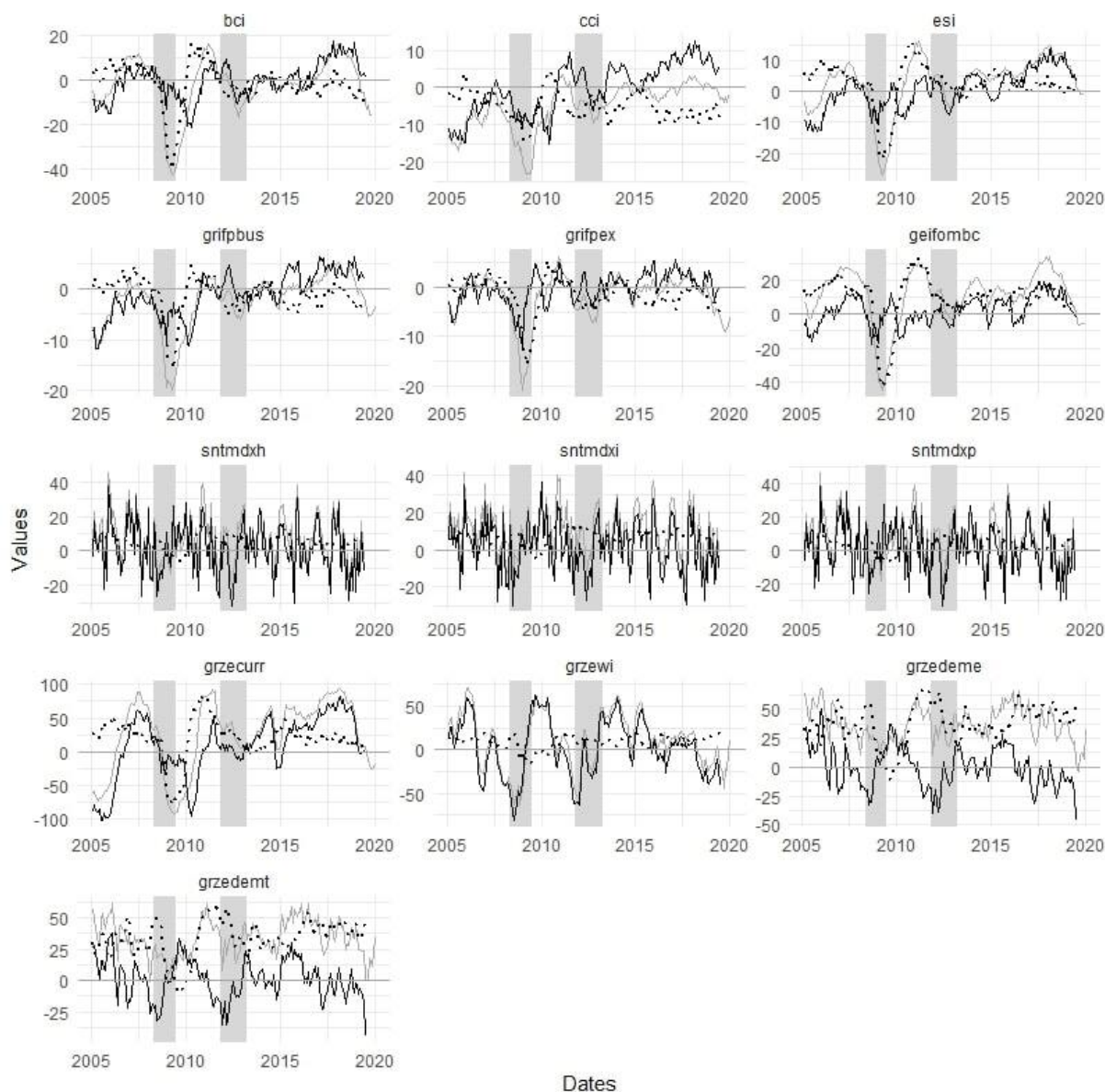
Appendix A: Additional Figures and Tables

Table A1: Summary Statistics

	Summary statistics						Unit root test statistics			Source
	Mean	S.D.	Min	Max	Skew.	Kurt.	ADF	KPSS	ρ_1	
Panel A: Returns										
R^{CDAX}	0.622	5.109	-19.780	16.080	-0.826	2.067	-5.604***	0.052	0.140	Bloomberg
R^{DAX}	0.586	5.152	-21.310	15.497	-0.802	2.130	-5.662***	0.050	0.107	Bloomberg
R^{MDAX}	0.883	5.545	-23.432	22.897	-0.770	3.798	-5.358***	0.058	0.156	Bloomberg
R^{TDAX}	0.837	5.813	-24.748	20.611	-0.925	3.037	-6.225***	0.072	0.151	Bloomberg
R^{SDAX}	0.715	5.323	-23.312	12.298	-1.171	2.770	-5.198***	0.074	0.220	Bloomberg
<i>SMB</i>	0.100	3.239	-9.622	9.342	0.011	0.241	-5.910***	0.118	-0.052	-
<i>SMM</i>	-0.177	2.483	-10.599	7.309	-0.086	1.734	-5.655***	0.119	-0.187	-
<i>MMB</i>	0.277	2.967	-12.493	7.968	-0.753	2.785	-5.779***	0.075	-0.171	-
Panel B: Control variables										
dy^{CDAX}	6.664	30.272	-49.423	132.836	1.851	4.756	-3.989**	0.224	0.943	Bloomberg
dy^{DAX}	6.352	28.677	-46.314	122.712	1.726	4.311	-3.937**	0.272	0.940	Bloomberg
dy^{MDAX}	9.247	43.469	-69.085	225.281	2.422	7.999	-4.817***	0.164	0.942	Bloomberg
dy^{TDAX}	13.179	37.215	-38.776	114.286	0.837	-0.052	-4.369***	0.461*	0.864	Bloomberg
dy^{SDAX}	8.547	44.496	-55.326	163.348	1.114	1.262	-3.399*	0.166	0.930	Bloomberg
<i>ip</i>	1.634	6.508	-23.622	14.82	-1.452	3.533	-3.621**	0.142	0.940	Eurostat
<i>ts</i>	-0.03	0.353	-0.654	1.217	1.221	2.106	-3.330*	0.194	0.897	Bloomberg
<i>fce</i>	2.294	1.016	-0.895	4.228	-0.767	0.874	-2.901	0.440*	0.950	Eurostat
<i>neer</i>	0.068	2.911	-8.337	5.129	-0.599	-0.260	-4.296***	0.111	0.928	BIS
<i>cpi</i>	1.411	0.802	-0.404	3.397	-0.089	-0.115	-3.524**	0.513**	0.948	Eurostat
<i>bc</i>	0.191	0.394	0.000	1.000	1.561	0.438	-2.968	0.394*	0.925	CEPR
Panel C: Original survey-based sentiment variables										
<i>bci</i>	-1.475	12.263	-42.200	16.400	-1.366	2.212	-3.365*	0.234	0.979	EC DG ECFIN
<i>cci</i>	-5.138	6.232	-23.300	3.600	-0.952	0.273	-3.547**	1.880***	0.970	EC DG ECFIN
<i>esi</i>	2.924	8.690	-26.900	16.200	-1.425	2.314	-3.427*	0.864***	0.981	EC DG ECFIN
<i>grifpbu</i>	-2.362	5.260	-19.900	5.200	-1.446	1.901	-3.480**	1.004***	0.976	Bloomberg
<i>grifpex</i>	-1.747	4.563	-20.800	6.100	-1.900	4.378	-3.318*	0.184	0.955	Bloomberg
<i>geifombc</i>	11.873	15.881	-44.900	34.100	-1.602	2.937	-3.464**	0.232	0.978	Bloomberg
<i>sntmdxh</i>	6.405	14.935	-25.044	45.665	0.048	-0.642	-5.313***	0.160	0.314	Sentix
<i>sntmdxi</i>	8.995	13.986	-21.513	41.89	0.005	-0.571	-5.091***	0.104	0.318	Sentix
<i>sntmdxp</i>	5.543	15.616	-26.394	46.893	0.045	-0.662	-5.362***	0.185	0.316	Sentix
<i>grzecurr</i>	27.582	52.266	-92.800	95.200	-0.811	-0.449	-3.666**	1.866***	0.980	Bloomberg
<i>grzewi</i>	10.852	31.179	-63.900	71.000	-0.292	-0.583	-4.260***	0.630**	0.935	Bloomberg
<i>grzedeme</i>	38.309	13.545	10.000	68.100	0.098	-0.792	-2.822	1.566***	0.889	Bloomberg
<i>grzedemt</i>	34.462	13.229	-0.400	61.700	-0.073	-0.697	-3.110	0.700**	0.849	Bloomberg
Panel D: Composite sentiment indicator										
<i>sent</i>	0.000	1.000	-2.441	1.819	-0.522	-0.242	-3.308*	1.767***	0.929	-
<i>sent.local</i>	0.000	1.000	-2.402	1.903	-0.487	-0.436	-2.371	1.088***	0.893	-
<i>sent.global</i>	0.000	1.000	-1.899	1.61	-0.319	-1.075	-2.745	0.809**	0.956	-

Notes: Columns ADF and KPSS report unit-root test statistics based on the Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The number of lags in each of the two tests was selected to minimize the AIC. Column ρ_1 lists the first-order autocorrelation coefficient. Panel C reports the statistics for the original sentiment indicators that were not “orthogonalized” to the fundamental factors. Sample period: January 2005–March 2019. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Figure A1: Decomposition of Sentiment



Note: This figure depicts 13 plots, each representing the dynamics of one of the 13 initial survey-based measures (the grey solid line), its irrational part (the black solid line), and the rational part (the black dotted line). The rational and irrational parts designate the fitted values and the residuals, respectively, from regressing each sentiment variable on the yearly change of industrial production (ip_t), the yearly change of the final consumption expenditure of households and NPISHs (fce_t), the term spread, defined as the difference between ten-year and two-year German government bond yields smoothed by its 12-month moving average (ts_t), and the yearly change in the nominal effective exchange rate ($neer_t$). Shaded areas represent CEPR recession months. A description of the sentiment measures can be found in Table 1.

Table A2: Bivariate Granger-Causality Tests

	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{TDAX}	R^{SDAX}	GC
Panel A: Granger-causality from irrational sentiment						
$bci^{ir} \rightarrow$	0.000	0.009	2.749*	0.980	2.677*	yes
$cci^{ir} \rightarrow$	1.478	1.336	1.136	5.14**	1.881	yes
$esi^{ir} \rightarrow$	0.141	0.191	0.003	0.871	2.43*	yes
$grifpbus^{ir} \rightarrow$	0.593	0.741	0.349	0.128	2.099**	yes
$grifpex^{ir} \rightarrow$	0.817	0.647	0.934	2.618	1.607*	yes
$geifombc^{ir} \rightarrow$	0.648	0.517	1.377	0.340	1.715	no
$sntmdxh1^{ir} \rightarrow$	3.739**	4.645**	0.009	1.221	0.006	yes
$sntmdxi1^{ir} \rightarrow$	2.299	3.041*	0.055	2.354	0.412	yes
$sntmdxp1^{ir} \rightarrow$	3.799**	4.65**	0.027	0.733	0.057	yes
$grzrecurr^{ir} \rightarrow$	0.161	0.22	1.104	0.228	0.956	no
$grzewi^{ir} \rightarrow$	0.255	0.143	1.036	0.963	1.426	no
$grzedeme^{ir} \rightarrow$	1.239	1.583	0.940	1.998	1.207	no
$grzedemi^{ir} \rightarrow$	0.081	0.223	0.405	1.464	1.066	no
Panel B: Granger-causality from lagged irrational sentiment						
$bci^{ir,l} \rightarrow$	0.179	0.210	0.123	0.918	3.485**	yes
$cci^{ir,l} \rightarrow$	2.445	2.063	2.125*	4.049**	2.371*	yes
$esi^{ir,l} \rightarrow$	0.696	0.703	0.132	1.906*	3.661**	yes
$grifpbus^{ir,l} \rightarrow$	0.009	0.044	0.009	0.506	0.000	no
$grifpex^{ir,l} \rightarrow$	1.687	1.732	1.690	2.724	1.842	no
$geifombc^{ir,l} \rightarrow$	0.115	0.032	0.001	0.196	2.241*	yes
$sntmdxh1^{ir,l} \rightarrow$	3.024*	3.903**	0.014	1.117	0.017	yes
$sntmdxi1^{ir,l} \rightarrow$	1.959	2.675	0.136	1.883	0.597	no
$sntmdxp1^{ir,l} \rightarrow$	3.013*	3.844**	0.005	0.718	0.000	yes
$grzrecurr^{ir,l} \rightarrow$	0.099	0.060	0.994	0.322	1.429	no
$grzewi^{ir,l} \rightarrow$	0.702	0.438	1.385	0.399	2.391	no
$grzedeme^{ir,l} \rightarrow$	0.900	1.051	0.711	0.774	2.223	no
$grzedemi^{ir,l} \rightarrow$	0.046	0.000	1.151	0.535	2.24	no

Note: The table reports the F-statistics from testing the null hypothesis of irrational sentiment (subscript ir , Panel A) or lagged irrational sentiment (subscript ir,l , Panel B) NOT Granger-causing the returns on equity index k . The number of lags for each bivariate Granger-causality test was chosen to minimize the AIC. The last column reports *yes* if the given sentiment variable Granger-causes the returns on at least equity index k and *no* otherwise. ***, **, and * denote the 1%, 5%, and 10% significance levels, respectively.

Table A3: Robustness – Effects of the Sign of Sentiment and the Alternative Sentiment Indicator

	Sign of Sentiment					Alternative Sentiment Index				
	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}
Panel A: Coefficients on sentiment										
$h = 1$	-0.650	-0.674	-0.759	-0.529	-0.281	0.149	0.071	0.035	0.059	0.865
$h = 2$	-0.426	-0.441	-0.454	-0.434	-0.128	0.120	0.050	0.041	0.059	0.820
$h = 3$	-0.420	-0.424	-0.445	-0.553	-0.182	-0.037	-0.094	-0.079	-0.11	0.683
$h = 6$	-0.451	-0.424	-0.551	-0.597	-0.422	-0.392	-0.414	-0.472	-0.501	0.203
$h = 9$	-0.436	-0.416	-0.523	-0.581	-0.356	-0.464	-0.473	-0.516	-0.525	0.087
$h = 12$	-0.275	-0.253	-0.368	-0.459	-0.108	-0.382	-0.387	-0.409	-0.427	0.184
Mean	-0.397	-0.439	-0.517	-0.526	-0.246	-0.168	-0.208	-0.233	-0.241	0.474
Panel B: Significance of sentiment										
$h = 1$	0.164	0.140	0.178	0.274	0.582	0.804	0.895	0.962	1.000	0.207
$h = 2$	0.281	0.255	0.299	0.293	0.701	0.846	0.896	0.942	0.966	0.259
$h = 3$	0.193	0.174	0.197	0.112	0.520	0.885	0.830	0.696	0.649	0.336
$h = 6$	0.042	0.040	0.026	0.014	0.083	0.102	0.093	0.046	0.035	0.767
$h = 9$	0.009	0.008	0.005	0.001	0.043	0.007	0.004	0.003	0.005	0.331
$h = 12$	0.020	0.020	0.005	0.001	0.164	0.003	0.002	0.002	0.002	0.367
Mean	0.198	0.190	0.140	0.083	0.412	0.460	0.448	0.347	0.306	0.719

Note: The table reports the estimated bootstrap simulation coefficients $\beta_1^{k(h)}$ together with the respective p-values for each horizon h from regressing the sign of the composite sentiment index, $sent_t^{-1/+1}$ (left panel), and the alternative sentiment index, $sentAlt_t$ (right panel), on the average future returns on equity index k according to Equation (3). See also the note in Table 2.

Table A4: Robustness – Effects of “Uncleaned” Sentiment

	Equation (2)					Equation (3)				
	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}
Panel A: Coefficients on $sent_t^u$										
$h = 1$	-0.153	-0.207	-0.159	-0.081	-0.123	-0.408	-0.493	-0.172	0.260	0.260
$h = 2$	-0.250	-0.289	-0.310	-0.230	-0.258	-0.220	-0.279	-0.120	0.399	0.378
$h = 3$	-0.326	-0.356	-0.422	-0.353	-0.357	-0.160	-0.212	-0.092	0.343	0.511
$h = 6$	-0.519	-0.528	-0.683	-0.640	-0.587	-0.471	-0.499	-0.512	-0.211	0.023
$h = 9$	-0.597	-0.596	-0.765	-0.756	-0.643	-0.530	-0.560	-0.519	-0.379	-0.024
$h = 12$	-0.588	-0.578	-0.767	-0.796	-0.592	-0.352	-0.373	-0.365	-0.329	0.197
Mean	-0.437	-0.426	-0.559	-0.497	-0.427	-0.410	-0.371	-0.332	-0.035	0.217
Panel B: Significance of $sent_t^u$										
$h = 1$	0.815	0.740	0.826	0.899	0.867	0.447	0.362	0.697	0.846	0.806
$h = 2$	0.673	0.609	0.644	0.698	0.706	0.594	0.516	0.699	0.790	0.825
$h = 3$	0.531	0.469	0.480	0.509	0.563	0.575	0.485	0.642	0.881	0.733
$h = 6$	0.106	0.081	0.065	0.058	0.144	0.122	0.088	0.085	0.207	0.416
$h = 9$	0.001	0.001	0.000	0.000	0.015	0.015	0.011	0.016	0.020	0.179
$h = 12$	0.000	0.000	0.000	0.000	0.002	0.016	0.010	0.010	0.003	0.265
Mean	0.304	0.174	0.124	0.157	0.293	0.390	0.332	0.364	0.400	0.742
Panel C: $\Delta adj R^2$										
						-0.004	-0.003	-0.007	-0.007	-0.008
						-0.011	-0.009	-0.011	-0.010	-0.012
						-0.015	-0.012	-0.014	-0.017	-0.016
						0.001	0.007	0.004	-0.013	-0.022
						0.019	0.029	0.015	0.004	-0.022
						0.000	0.008	0.002	0.002	-0.044

Note: The table reports the estimated bootstrap simulation coefficients $\alpha_1^{k(h)}$ and $\beta_1^{k(h)}$ (Panel A) together with the respective p-values for each horizon h (Panel B) from regressing the “uncleaned” composite sentiment indicator $sent_t^u$ on the average future returns on equity index k according to Equations (2) and (3), respectively. See also the note in Table 2.

Table A5: Robustness - Effects of the Composite Sentiment Indicator in Subsamples

	<i>Subsample 1 (2005–2011)</i>					<i>Subsample 2 (2012–2018)</i>				
	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}
Panel A: Coefficients on <i>sent</i>										
$h = 1$	0.803	0.711	0.281	-0.170	1.728	-0.577	-0.621	-0.469	-0.500	-0.387
$h = 2$	0.977	0.923	0.406	-0.062	1.562	-0.717	-0.796	-0.488	-0.605	-0.220
$h = 3$	1.134	1.094	0.718	0.075	1.611	-1.007	-1.096	-0.834	-0.937	-0.327
$h = 6$	0.587	0.620	0.293	-0.296	0.780	-1.013	-1.095	-0.898	-1.145	-0.372
$h = 9$	0.005	0.021	-0.136	-0.528	0.258	-0.452	-0.504	-0.381	-0.634	-0.038
$h = 12$	-0.069	-0.054	-0.216	-0.540	0.077	-0.695	-0.786	-0.369	-0.565	-0.118
Mean	0.573	0.553	0.224	-0.254	1.003	-0.744	-0.816	-0.573	-0.731	-0.244
Panel B: Significance of <i>sent</i>										
$h = 1$	0.226	0.215	0.927	0.455	0.130	0.302	0.305	0.250	0.351	0.414
$h = 2$	0.275	0.237	0.964	0.422	0.276	0.099	0.076	0.140	0.205	0.422
$h = 3$	0.243	0.192	0.909	0.340	0.268	0.029	0.027	0.029	0.052	0.293
$h = 6$	0.518	0.731	0.068	0.006	0.911	0.001	0.001	0.001	0.001	0.110
$h = 9$	0.006	0.010	0.000	0.000	0.125	0.042	0.025	0.044	0.038	0.367
$h = 12$	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.006	0.007	0.132
Mean	0.442	0.444	0.343	0.215	0.633	0.182	0.169	0.199	0.102	0.734

Note: The table reports the estimated bootstrap simulation coefficients $\beta_1^{k(h)}$ together with the respective p-values for each horizon h from regressing the composite sentiment indicator $sent_t$ on the average future returns on equity index k according to Equation (3). The left panel lists the results for the February 2005–December 2011 subsample and the right panel for the January 2012–December 2018 subsample. See also the note in Table 2.

Table A6: Effects of the Consumer Confidence Index

	Equation (2)					Equation (3)				
	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}	R^{CDAX}	R^{DAX}	R^{MDAX}	R^{SDAX}	R^{TDAX}
Panel A: Coefficients on $sent$										
$h = 1$	0.298	0.25	0.263	0.355	0.350	-0.459	-0.521	-0.347	0.093	0.251
$h = 2$	0.196	0.148	0.142	0.222	0.243	-0.322	-0.359	-0.262	0.175	0.319
$h = 3$	0.118	0.067	0.042	0.14	0.214	-0.268	-0.309	-0.202	0.188	0.489
$h = 6$	-0.162	-0.199	-0.277	-0.232	-0.070	-0.509	-0.536	-0.497	-0.273	0.094
$h = 9$	-0.342	-0.368	-0.429	-0.388	0.029	-0.596	-0.628	-0.465	-0.348	0.051
$h = 12$	-0.397	-0.409	-0.485	-0.464	0.237	-0.452	-0.479	-0.311	-0.254	0.287
Mean	-0.048	-0.085	-0.124	-0.061	0.181	-0.434	-0.472	-0.347	-0.07	0.249
Panel B: Significance of $sent$										
$h = 1$	1.000	0.937	0.971	0.935	0.547	0.410	0.349	0.548	0.857	0.894
$h = 2$	0.870	0.806	0.835	0.882	0.745	0.554	0.496	0.643	0.933	0.885
$h = 3$	0.754	0.677	0.696	0.748	0.816	0.575	0.491	0.642	0.923	0.736
$h = 6$	0.329	0.261	0.268	0.254	0.433	0.236	0.178	0.236	0.314	0.653
$h = 9$	0.051	0.032	0.052	0.041	0.376	0.065	0.044	0.116	0.108	0.395
$h = 12$	0.003	0.001	0.005	0.003	0.820	0.041	0.023	0.093	0.049	0.624
Mean	0.356	0.319	0.334	0.375	0.906	0.469	0.402	0.541	0.563	0.917
Panel C: $\Delta adjR^2$										
						-0.004	-0.004	-0.005	-0.008	-0.008
						-0.011	-0.010	-0.011	-0.012	-0.013
						-0.014	-0.013	-0.015	-0.017	-0.016
						-0.010	-0.006	-0.010	-0.016	-0.028
						0.003	0.012	-0.010	-0.013	-0.032
						-0.010	-0.001	-0.022	-0.022	-0.047

Note: The table reports the estimated bootstrap simulation coefficients $\alpha_1^{k(h)}$ and $\beta_1^{k(h)}$ (Panel A) together with the respective p-values for each horizon h (Panel B) from regressing the scaled Consumer Confidence Indicator cci on the average future returns on equity index k according to Equations (2) and (3), respectively. See also the note in Table 2.

Appendix B: Bootstrap Simulation Procedure

In this paper, I employ non-parametric bootstrap simulation as proposed in Brown and Cliff (2005) and further applied in Schmeling (2007) to deal with two well-known problems connected with the h -period average return regressions as specified in Equations (2) and (3). The first problem is linked to the small sample bias that is caused by persistent regressors in finite samples (Stambaugh, 1999; Ferson et al., 2003). The second problem arises from the autoregressive residuals of order $(h-1)$ that are induced by the overlapping periods in the dependent variable, which could not be corrected by the bias-adjusted standard errors suggested by Hansen and Hodrick (1980) and Newey and West (1987) because these corrections do not perform well in finite samples (Hodrick, 1992).

As the first step, I run the restricted vector autoregression with one lag (VAR(1))³⁶ for $y_t = [R_t^k, Sent_t, Z_t^k]$, where R_t^k are the monthly returns on stock market index k , $Sent_t$ is the sentiment variable, and Z_t^k is the matrix of controls. The restriction is imposed so that it represents the null of no return predictability by sentiment. The reduced form of VAR(1) with M endogenous variables $y_t = [y_{1,t}, \dots, y_{M,t}]$ has the form:

$$y_t = A_0 + A_1 y_{t-1} + \varepsilon_t, \quad (B1)$$

where A_0 is an $(M \times 1)$ vector of constants, A_1 is an $(M \times M)$ coefficient matrix, and ε_t is an M -dimensional process with $E(u_t) = 0$ and time-invariant positive definite covariance matrix $E(\varepsilon_t \varepsilon_t^T) = \Sigma_\varepsilon$. The null hypothesis that sentiment does not predict returns is imposed on the coefficient matrix A_1 by setting the corresponding elements to zero. Without loss of generality, suppose from now on that the matrix Z_t^k only contains one control variable. The restriction on matrix A_1 then has the form:

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Rewriting Equation (B1) so that it embodies the null hypothesis of no return predictability of VAR(1) with $y_t = [R_t^k, Sent_t, Z_t^k]$ yields the following specification:

$$R_t^k = \alpha_{1,0} + \alpha_{1,1} R_{t-1}^k + \alpha_{1,3} Z_{t-1}^k + \varepsilon_{1,t} \quad (B2)$$

$$Sent_t = \alpha_{2,0} + \alpha_{2,1} R_{t-1}^k + \alpha_{2,2} Sent_{t-1} + \alpha_{2,3} Z_{t-1}^k + \varepsilon_{2,t} \quad (B3)$$

$$Z_t = \alpha_{3,0} + \alpha_{3,1} R_{t-1}^k + \alpha_{3,2} Sent_{t-1} + \alpha_{3,3} Z_{t-1}^k + \varepsilon_{3,t} \quad (B4)$$

Now let $\hat{\alpha} = (\hat{\alpha}_{1,0}, \hat{\alpha}_{2,0}, \dots, \hat{\alpha}_{3,3})$ be a vector of coefficient estimates and $\hat{\varepsilon} = (\hat{\varepsilon}_{1,t}, \hat{\varepsilon}_{2,t}, \hat{\varepsilon}_{3,t})$ be a vector of estimated residuals from Equations (B2), (B3), and (B4). Further, assume that there are T observations in the sample studied. The bootstrap distribution is then generated by running the following steps B times:³⁷

1. Resample the observations in $\hat{\varepsilon}$ with replacement to obtain bootstrapped samples ε^* of length T .

³⁶ One lag was selected based on the AIC criteria. Moreover, Brown and Cliff (2005) used VAR(1) as well.

³⁷ Brown and Cliff (2005) used $B = 10,000$ replications.

2. Using the bootstrapped residuals from the previous step and the estimated coefficients ($\hat{\alpha}$), generate the sequence of observations for R_t^{k*} , $Sent_t^*$, and Z_t^* according to Equations (B2), (B3), and (B3), respectively. Use the sample means for the initial values in starting up the autoregression.
3. Use the generated R_t^{k*} to calculate the h -horizon average returns on stock market index k .
4. With the new series in hand, estimate the original system of equations (such as Equations (2) and (3)) and save coefficients $\tilde{\beta}^{k(h)}$ along with the corresponding Hansen and Hodrick (1980) standard errors that correct for heteroscedasticity and autocorrelation in the residuals. In addition, save Theil's (1971) \hat{F} -statistic to test the joint linear restriction that sentiment does not predict stock market returns at all horizons h .

Following Brown and Cliff (2005) and Greene (2012), I calculated the t -statistic for the test that sentiment does not predict the future h -horizon average returns for each iteration $i, \forall i, \dots, B$, as the difference between the coefficient estimate resulting from iteration i and the bias (which, in the present bootstrap simulation under the null of no return predictability, is simply the mean of the estimated coefficients across B iterations) divided by the standard deviation of the B coefficient estimates:

$$t^{i,k(h)} = \frac{\tilde{\beta}^{i,k(h)} - \overline{\tilde{\beta}^{i,k(h)}}}{std[\tilde{\beta}^{i,k(h)}]} \quad (B5)$$

The corresponding p-values of the two-tailed test³⁸ are then calculated as described in MacKinnon (2007):

$$\hat{p}(\hat{t}^{k(h)}) = \frac{1}{B} \sum_{i=1}^B I(|t^{i,k(h)}| > |\hat{t}^{k(h)}|) \quad (B6)$$

where $I(\cdot)$ represents the indicator function, which is equal to 1 when its argument is true and 0 otherwise. Next, I calculated the bias-adjusted coefficients $boot\beta^{k(h)}$ as the difference between the original coefficient estimates and the mean of the estimated coefficients across B iterations.³⁹ Finally, I calculated the bootstrap \hat{F} -statistic of the joint test of significance across the six horizons as the bias-adjusted Theil (1971) \hat{F} -statistic, where the bias is defined in the same fashion as that in the coefficient estimates. The corresponding p-values are then computed similarly to the p-values in Equation (B6).

³⁸ I assume that the t -statistics calculated in Equation (B5) closely follow the normal distribution and are symmetrically distributed around zero.

³⁹ Mark (1995) used the median values for bias adjustment, but I follow Brown and Cliff (2005) and use the bootstrap simulation means instead.

CNB Working Paper Series (since 2019)

WP 13/2020	Zuzana Rakovská	<i>Composite survey sentiment as a predictor of future market returns: Evidence for German equity indices</i>
WP 12/2020	Vojtěch Siuda	<i>A top-down stress-testing framework for the nonfinancial corporate sector</i>
WP 11/2020	Hana Hejlová Michal Hlaváček Blanka Vačkova	<i>Estimating commercial property price misalignment in the CEE countries</i>
WP 10/2020	Zuzana Rakovská Dominika Ehrenbergerová Martin Hodula	<i>The power of sentiment: Irrational beliefs of households and consumer loan dynamics</i>
WP 9/2020	Ivan Sutóris	<i>The intertemporal cost of living and dynamic inflation: The case of the Czech Republic</i>
WP 8/2020	Martin Hodula Jan Janků Martin Časta Adam Kučera	<i>On the determinants of life and non-life insurance premiums</i>
WP 7/2020	František Brázdík Tibor Hlédik Zuzana Humplová Iva Martonosi Karel Musil Jakub Ryšánek Tomáš Šestořád Jaromír Tonner Stanislav Tvrz Jan Žáček	<i>The g3+ model: An upgrade of the Czech National Bank's core forecasting framework</i>
WP 6/2020	Volha Audzei Jan Brůha	<i>A model of the Euro Area, China and the United States: Trade links and trade wars</i>
WP 5/2020	Dominika Ehrenbergerová Martin Hodula Zuzana Rakovská	<i>Does capital-based regulation affect bank pricing policy?</i>
WP 4/2020	Alexis Derviz	<i>Sovereign capital, external balance, and the investment-based Balassa-Samuelson effect in a global dynamic equilibrium</i>
WP 3/2020	Milan Szabo	<i>Growth-at-risk: Bayesian approach</i>
WP 2/2020	Martin Hodula Ngoc Anh Ngo	<i>Finance, growth and (macro)prudential policy: European evidence</i>
WP 1/2020	Michal Franta Ivan Sutóris	<i>Dynamics of Czech inflation: The role of the trend and the cycle</i>
WP 12/2019	Josef Bajzík Tomáš Havránek Zuzana Iršová Jiří Schwarz	<i>The elasticity of substitution between domestic and foreign goods: A quantitative survey</i>

WP 11/2019	Martin Hodula Simona Malovaná Jan Frait	<i>Too much of a good thing? Households' macroeconomic conditions and credit dynamics</i>
WP 10/2019	Martin Hodula Simona Malovaná Jan Frait	<i>Introducing a new index of households' macroeconomic conditions</i>
WP 9/2019	Jiří Panoš Petr Polák	<i>How to improve the model selection procedure within a stress testing framework?</i>
WP 8/2019	Sebastian Gechert Tomáš Havránek Zuzana Iršová Dominika Kolcunová	<i>Death to the Cobb-Douglas production function? A quantitative survey of the capital-labor substitution elasticity</i>
WP 7/2019	Alexis Derviz	<i>Coexistence of physical and crypto assets in a stochastic endogenous growth model</i>
WP 6/2019	Dominika Ehrenbergerová Simona Malovaná	<i>Introducing macro-financial variables into semi-structural model</i>
WP 5/2019	Martin Hodula	<i>Monetary policy and shadow banking: Trapped between a rock and a hard place</i>
WP 4/2019	Simona Malovaná Žaneta Tesařová	<i>Banks' credit losses and provisioning over the business cycle: Implications for IFRS 9</i>
WP 3/2019	Aleš Bulíř Jan Vlček	<i>Monetary policy is not always systematic and data-driven: Evidence from the yield curve</i>
WP 2/2019	Dominika Kolcunová Simona Malovaná	<i>The effect of higher capital requirements on bank lending: The capital surplus matters</i>
WP 1/2019	Jaromír Baxa Tomáš Šestořád	<i>The Czech exchange rate floor: Depreciation without inflation?</i>

CNB Research and Policy Notes (since 2019)

RPN 3/2020	Simona Malovaná Martin Hodula Zuzana Rakovská	<i>Researching the research: A central banking edition</i>
RPN 2/2020	Simona Malovaná Josef Bajzík Dominika Ehrenbergerová Jan Janků	<i>A prolonged period of low interest rates: Unintended consequences</i>
RPN 1/2020	Simona Malovaná	<i>How to organize research in central banks: The Czech National Bank's experience</i>

RPN 4/2019	Tomáš Konečný Lukáš Pfeifer	<i>Macprudential ring-fencing</i>
RPN 3/2019	Petr Polák Jiří Panoš	<i>Impact of expectations on IFRS 9 loan loss provisions</i>
RPN 2/2019	Jan Filáček Ivan Sutóris	<i>Inflation targeting flexibility: The CNB's reaction function under scrutiny</i>
RPN 1/2019	Iveta Polášková Luboš Komárek Michal Škoda	<i>The contemporary role of gold in central banks' balance sheets</i>

CZECH NATIONAL BANK

Na Příkopě 28

115 03 Praha 1

Czech Republic

ECONOMIC RESEARCH DIVISION

Tel.: +420 224 412 321

Fax: +420 224 412 329

<http://www.cnb.cz>

e-mail: research@cnb.cz

ISSN 1803-7070