Working Paper Series 10/2021

Does Sentiment Affect Stock Returns? A Meta-analysis Across Survey-based Measures

Zuzana Gric, Josef Bajzík, Ondřej Badura

Czech National Bank —— Working Paper Series —— 10/2021

CNB CZECH NATIONAL BANK

www.cnb.cz

Reviewed by:	Evžen Kočenda (Charles University, Prague)
	Adam Kučera (Czech National Bank)
Project Coordinator:	Jan Janků

Does Sentiment Affect Stock Returns? A Meta-analysis Across Survey-based Measures

Zuzana Gric, Josef Bajzík, and Ondřej Badura *

Abstract

We are the first to analyse the literature on the relationship between sentiment and stock returns, a topic which reacts to the history of systemic events causing asset bubbles in financial markets. Our meta-analysis of 1311 point estimates from 30 primary studies suggests that the true effect of an improvement in sentiment is significant and negative. In the majority of specifications, researchers tend to report this effect as being much stronger than it actually is, but in specific subsamples we also find positive publication bias driving the results into less negative or even positive territory. We reveal that the effect of sentiment on future returns is significantly stronger in the case of individual investors compared with large institutions and in the case of stock markets in the US compared with Europe. The effect also depends on several data and model characteristics. We propose an implied estimate which suggests that a one standard deviation increase in sentiment reduces future monthly returns by 0.198 standard deviations on average. This result may help enhance the predictive power of stock market models and be useful in conducting stress tests of financial markets and assessments of risks to financial stability.

Abstrakt

Jako první analyzujeme literaturu o vztahu mezi sentimentem a akciovými výnosy. Toto téma reaguje na skutečnost, že systémové události opakovaně vytvářejí bubliny na finančních trzích. Naše metaanalýza 1311 odhadů ze 30 primárních studií naznačuje, že skutečný efekt zlepšení sentimentu je statisticky významný a záporný. Ve většině specifikací mají výzkumníci tendenci uvádět tento efekt jako mnohem silnější, než ve skutečnosti je, ale ve specifikacích pro dílčí vzorky dat nacházíme také kladný vliv publikační selektivity, který výsledky posouvá na méně záporné nebo dokonce kladné hodnoty. Ukazujeme, že vliv sentimentu na budoucí výnosy je výrazně silnější v případě individuálních investorů než u velkých institucí a také na akciových trzích v USA oproti v Evropě. Efekt je také závislý na několika charakteristikách dat a modelu. Předkládáme implikovaný odhad, podle něhož zvýšení sentimentu o jednu směrodatnou odchylku snižuje budoucí měsíční výnosy v průměru o 0,198 směrodatné odchylky. Tento výsledek může pomoci zlepšit predikční schopnost modelů akciových trhů a najít využití při provádění zátěžových testů finančních trhů a hodnocení rizik pro finanční stabilitu.

JEL Codes: C11, G12, G15, G23, G41.
Keywords: Bayesian model averaging, individual and institutional investors, meta-analysis, publication bias, stock returns, survey-based sentiment.

* Zuzana Gric, Czech National Bank and Masaryk University in Brno, zuzana.gric@cnb.cz
Josef Bajzík, Czech National Bank and Charles University in Prague, josef.bajzik@cnb.cz
Ondřej Badura, Czech National Bank and VSB - Technical University of Ostrava, ondrej.badura@vsb.cz
The authors note that the paper represents their own views and not necessarily those of the Czech National Bank. We want to thank Simona Malovaná, who helped us with methodology, Adam Kucera, Evžen Kocenda, Dominika Ehrenbergerová, and participants at the Czech National Bank's seminar for valuable comments. All errors and omissions remain the fault of the authors.

1. Introduction

Well-known financial market events such as the Black Monday crash, the dot-com bubble, and the Global Financial Crisis, as well as more recent episodes linked to the ups and downs of cryptoassets and the success of zero-fee trading applications such as Robinhood, including the late-2020 meme stock craze (Bloomberg, 2020; NY Times, 2021), have repeatedly demonstrated that market fundamentals are not the only variables lying behind movements in stock markets. The increased popularity of behavioral finance (and economics) has inspired many authors to consider behavioral concepts when studying the dynamics of stock returns. Most commonly, they have turned to sentiment – an elusive concept that can be defined as a set of beliefs "that are not fully justified by fundamental news" (Shleifer and Summers, 1990, p. 19) or "optimism or pessimism about stocks in general" (Baker and Wurgler, 2006, p. 1649). Hence, an increase or positive change in sentiment implies increased optimism, while a decrease or negative change in sentiment suggests a decline in optimism.

Many studies have been conducted in the last few decades to verify the power of sentiment to explain stock returns. Generally, there is a consensus that this relationship is indeed significant (Schmeling, 2007; Ben-Rephael et al., 2012; Zhou, 2018; Gao et al., 2020). However, the specific properties of the relationship often remain ambiguous, obviously being subject to other related characteristics.

Above all, as explained by Baker and Wurgler (2006), the sentiment–stock return relationship works differently over different time horizons. The contemporaneous effect of increased sentiment is presumed to be positive, as it can temporarily boost investor demand. Stocks thus become overvalued (Barberis et al., 1998) and earn positive returns. However, the longer-run effect of sentiment is negative. Due to limits on arbitrage (De Long et al., 1990b; Shleifer and Summers, 1990), the initial overvaluation effect does not disappear immediately. Only after some time do stock prices converge to their intrinsic values, causing price reversals and thus negative future returns. Many studies present further empirical evidence of this pattern (Hengelbrock et al., 2013; Rakovská, 2021), while Brown and Cliff (2005) and Schmeling (2009), for instance, cast a little doubt on its full validity, as they report almost universally negative coefficients.

Considerable variation is also found in the effects of sentiment, as revealed by two types of market participants – individual and institutional investors. As suggested by Schmeling (2007), institutional sentiment (representing firms, mutual funds, and other large investors) is based on relevant information, is more closely related to movements in fundamental factors, and thus predicts future stock returns correctly (positively) on average. In contrast, the sentiment of individual investors is perceived as noise, and thus, due to temporary mispricing, the predictive effect for future returns is negative (Schmeling, 2007; Bathia and Bredin, 2013; Fernandes et al., 2013).

Geographical and cultural aspects also seem to matter. Schmeling (2007) argues that sentiment effects are stronger in countries that have higher levels of individualism and herding investment behavior, while Schmeling (2009) shows that countries with high institutional integrity are less prone to sentiment. Nevertheless, not all studies share the same conclusion (Bathia and Bredin, 2013; Gao et al., 2020).

Obviously, the relationship between investor sentiment and stock markets calls for a rigorous analysis of the existing literature in order to shed light on the properties of the sentiment effect under different conditions. We employ meta-analysis as the most suitable method for such an evaluation (Imai et al., 2021).[1] As far as we know, no meta-analysis has ever been conducted in this field. Thus, the main goal of this paper is to fill the literature gap and, by doing so, to reveal patterns that could not be observed in single studies.

On top of that, we have another motive. Quantifying the true sentiment effect and identifying the drivers of heterogeneity might not only help researchers in their quest for rigorous specification of the relationship studied, but also aid professionals in implementing the results in their investment strategies. Furthermore, policymakers concerned with the amplitude of financial and business cycles and risks to financial stability may also benefit from our research. The global growth of the investment fund sector, together with the increasing share of equity holdings in households' balance sheets,[2] makes the stock market a relevant channel of transmission of potential adverse shocks to the financial sector and the real economy. An influential stream of literature also relates sentiment to systematic events such as asset bubbles (Akerlof and Shiller, 2010; Shefrin, 2016) and to systemic risks as a whole (Borovkova et al., 2017). Moreover, local returns have been shown to be more sensitive to outside sentiment (Baker et al., 2012; Corredor et al., 2015), which can threaten the stability of a national financial system even outside its boundaries. Thus, sentiment as a likely measure of financial fragility (Dow, 2011) may well also serve macroprudential policy.

In our meta-analysis, we focus on three main characteristics which may determine the sentiment effect. They are the effect of the horizon and the type of sentiment, as described above, and additionally the high-level type of return series. With the first characteristic, we aim to recognize diverse effects of sentiment on contemporaneous and future returns, while the second characteristic differentiates between the sentiment of individual and institutional investors. The last aspect distinguishes between aggregate market returns, equity index returns, and portfolio returns, and has previously been given hardly any attention. Beside this, we look at many other properties that may have influenced the published results. We categorize them into data characteristics (different categorizations and transformations of sentiment and return series, including information on whether sentiment was orthogonalized or not, region of data, sample size, etc.), model specification (data frequency, use of control variables, method of estimation, etc.), publication characteristics (impact factor of journal, number of citations, year of publication, etc.), and also external information related to the region studied rather than directly to the given article (CPI growth, financial development, savings-to-GDP ratio, etc.).

Furthermore, we focus solely on survey-based sentiment, that is, on sentiment derived directly from questionnaires. The reasons for looking only at sentiment from surveys are threefold: (1) these measures serve as a direct quantification of sentiment, while other types of sentiment are just indirect approximations; (2) among the various types of sentiment, survey-based sentiment has the longest history of empirical results in the academic literature; and (3) the economic interpretation of different survey-based sentiment indicators is similar, while this might not be the case for other types of sentiment.

[1] Meta-analysis is based on all the empirical results related to a given topic and reviews them both quantitatively and qualitatively. It systematically investigates the differences in empirical findings and draws real inferences. It corrects for sample selection bias and measurement errors, which may ruin the primary studies. Moreover, by taking into account a large number of independent variables, meta-analysis addresses the endogeneity problem. It shows the magnitude by which each factor of the study design and cross-country variation affects the estimates reported in the primary studies (Kim et al., 2019).

[2] In the period 2012–2020, equity and investment fund shares in households' balance sheets grew by about 40% in the EU as a whole and by more than 100% in some countries (for example, the Czech Republic and Denmark). Source: https://ec.europa.eu/eurostat.

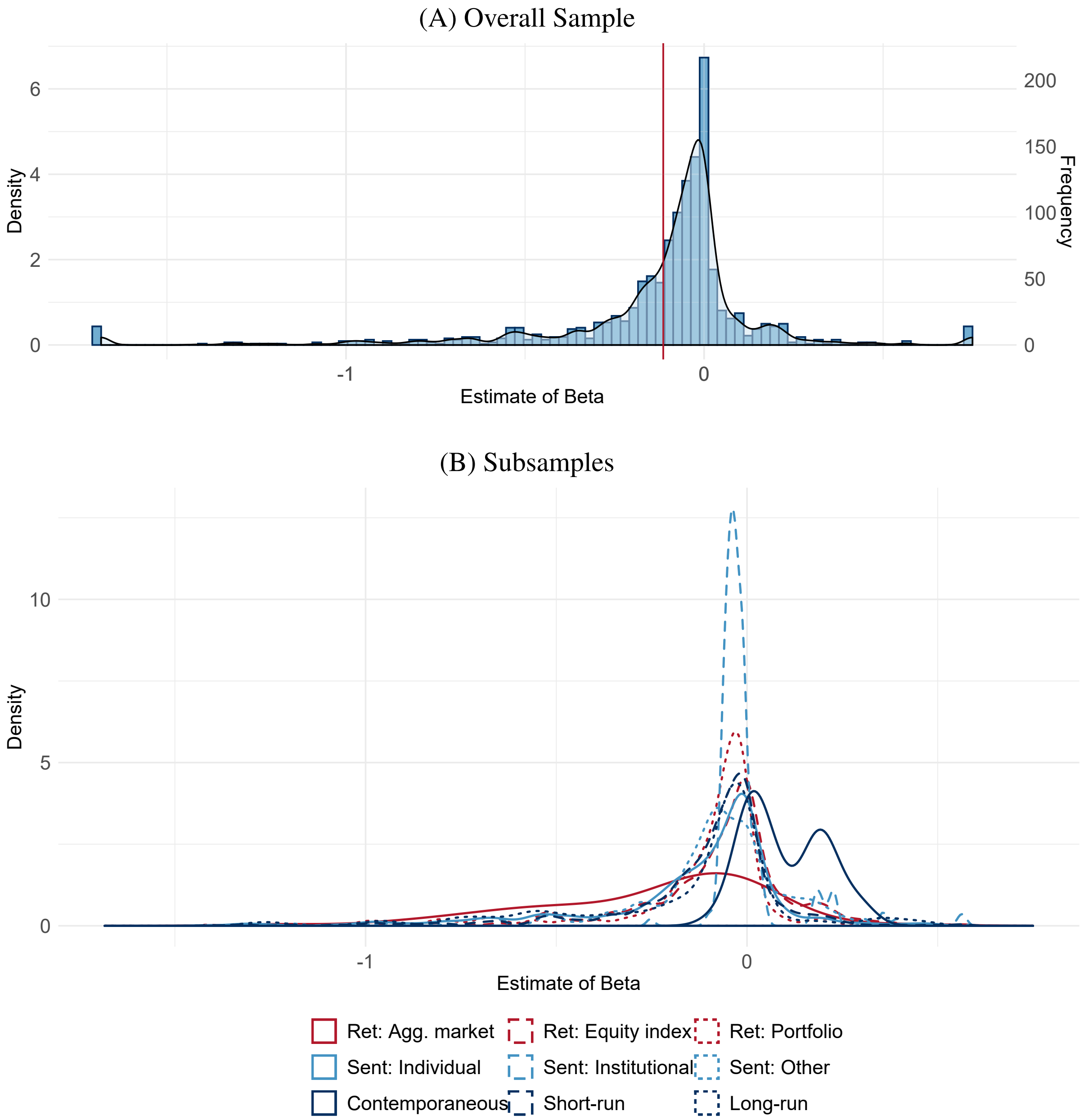
Even though we restrict the sentiment measures to survey-based indexes, we still deal with substantial heterogeneity in the data as regards its range and units of measurement. To overcome this issue, we apply "full standardization" of elasticities, as described in Bowman (2012). After full standardization, all the elasticities represent the estimated effect on (contemporaneous or future) returns measured in standard deviations implied by a one standard deviation increase in sentiment. Being aware of the shortcomings of such transformation (standard deviations are not as easy to understand as levels or percentages), we also propose an approach for recalculating the implied effects back into more familiar units.

We collect 1311 estimates (and their statistics) from 30 primary studies. To investigate the heterogeneity among the published results, we further gather almost 50 variables covering the topics mentioned above. Figure 1 provides a quick overview of the distribution of the estimated effects of sentiment on stock returns in the current literature. Two initial findings can be drawn from these plots. Panel A shows that most of the estimates are close to zero, but the left tail of the distribution is longer than the right one. Second, Panel B demonstrates slight differences in the distributions for different subsamples based on our main characteristics (type of return, type of sentiment, and effect horizon). More specifically, a large part of the left tail is related to aggregate market estimates, the sentiment of institutional investors is mostly distributed around zero while the effect of individuals has a much fatter left tail, and the contemporaneous effect of sentiment on stock returns even has the opposite (positive) direction.

The present meta-analysis addresses both publication bias and the drivers of heterogeneity. We follow the key studies in the field. When studying publication bias, we use funnel plots (Egger et al., 1997) and simple formal tests, both unweighted and weighted (Stanley and Doucouliagos, 2012). In addition, we follow the most recent literature (Havranek et al., 2017; Gechert et al., 2020) and employ the non-linear techniques developed in Bom and Rachinger (2019) and Furukawa (2019). To study the heterogeneity among the estimates, we use Bayesian model averaging (BMA) with a dilution prior (Raftery et al., 1997; George, 2010) and frequentist model averaging (FMA, Amini and Parmeter, 2012).

We find publication bias across the various types of return series, individual and institutional sentiment, and the contemporaneous and long-term effect horizons. Interestingly though, the direction of the bias is not the same in all these categories of data. In the case of institutional sentiment and the contemporaneous effect horizon, the bias has a positive sign, while in all other cases it is significantly negative. This pattern implies that researchers tend to report more positive results relative to the true effects in the case of the two positively biased categories, while for the negatively biased categories they exaggerate them in opposite direction. Altogether, we find that after removing the publication bias, the true effect of sentiment on stock returns is not negligible and that it is mostly negative, although in some specifications it is not significant. For the subsample of longer horizons, this echoes the overreaction hypothesis (Barberis et al., 1998), which posits that sentiment has negative effects on *future* returns. Nevertheless, we also find a significantly negative true effect for the *contemporaneous* horizon specification, which, together with the positive publication bias, suggests that only in theory do stocks become immediately overvalued when sentiment improves. All in all, the bias-free impact of sentiment on stock returns, whether contemporaneous or future, is negative. Hence, our meta-analysis suggests that sentiment stands as a contrarian predictor of stock returns.

Figure 1: (Smoothed) Frequency of Elasticities



Note: Panel A depicts the histogram (right y-axis) and the density (left y-axis) of the estimated elasticities. The solid vertical line indicates the mean value. Panel B depicts the densities of the estimated elasticities in selected subsamples of data. All the elasticities are transformed to monthly effects, fully standardized (see Section 2), and winsorized at 1% from each side.

The results for the drivers of heterogeneity confirm and expand this conclusion. Publication bias remains the most prominent driver of the heterogeneity in the estimates. Furthermore, (1) the sentiment of individual investors has a much stronger negative effect on future returns than institutional sentiment; (2) the contemporaneous effect affects markets to a lesser degree than the longer-lasting one; (3) the type of returns the researcher applies does not play a role – the mean sentiment effect is not affected by this choice; (4) US markets are more prone to sentiment behavior than European ones; and (5) data and model characteristics also matter; for example, higher frequency data or models estimated as a system of equations induce a more negative effect of sentiment. These findings are stable across various robustness checks.

Finally, based on our knowledge and the current state of the literature we propose implied estimates that might be taken into account in trading strategies, model calibrations, and financial market research in the future. We find that a one standard deviation increase in survey-based sentiment reduces future stock market returns by 0.198 standard deviations on average.

The remainder of the article is structured as follows. Section 2 explains the data collection process and discusses the main characteristics of the data, including the theoretical foundations for why they should matter. Section 3 explores the magnitude of the publication bias in the literature on the effect of sentiment on stock returns. Section 4 examines the drivers of heterogeneity in the primary studies and provides the main findings of our research. In Section 5 we present the implied estimates and formulate several implications for the current sentiment–return literature. A summary is provided in Section 6.

2. Data

The sentiment–return relationship has been studied for more than three decades (Clarke and Statman, 1998), but there seems to be no quantitative and qualitative overview of its results. It is time to fill this gap by means of meta-analysis, the most suitable approach for conducting such an evaluation (Imai et al., 2021).

In general, meta-analysis compares differences in results for homogeneous variables of interest. It is desirable for the studies to be heterogeneous to some degree so that sources of potential bias in the reported results can be examined. With this in mind, we decided to focus only on sentiment and stock return series that exhibit certain characteristics.

First, this study centers on survey-based sentiment, that is, sentiment derived from a broad range of questionnaires, such as consumer surveys directed at the general public and investment surveys aimed at narrower groups of investors. Overall, another three groups of sentiment measures can be considered: (i) financial proxies for sentiment, such as market turnover (Baker and Stein, 2004) and closed-end fund discounts (Lee et al., 1991), often combined into a composite sentiment index (Baker and Wurgler, 2006; Han and Li, 2017); (ii) search- or attention-based sentiment, which leverages information on the public's interest in a specific online topic or area (Kristoufek, 2013; Behrendt et al., 2020); and (iii) sentiment extracted from text, ranging from formal newspapers (Tetlock, 2007) to informal social media (Bollen et al., 2011).

Second, as for the characteristics of the return series, we only considered those studies which employed simple returns on stock market indexes, portfolios, and aggregate markets, calculated as either growth rates or log differences. We accepted both simple returns and returns in excess of the risk-free rate.[3] Put differently, we disregarded all studies that employed simple prices or price differences as the dependent variable and studies that used too specific return series, such as IPO returns and returns on long-short portfolios. The rationale for this is straightforward – we limited our data collection process to the most frequently used return series in order to gather as many estimates as possible while maintaining the overall comparability of the results.

According to our search, the general (reduced) regression equation can be formulated as:

$$Return_{t+h} = \beta_0 + \beta_1 Sentiment_{t-k} + \beta_2 X_t + \varepsilon_t. \tag{1}$$

[3] There is even a third category of returns – abnormal returns, but this category was not covered in our set of primary studies so we do not discuss it.

Our elasticity of interest is represented by β_1. We collect this coefficient from all relevant primary studies. The sentiment variable is typically lagged by one period ($k = 1$), but it may be lagged by more periods or be contemporaneous. Similarly, returns may enter the regression either at time t or with some positive lead ($t+h$) in order to capture the future dynamics of the series. Moreover, a number of studies model the relation for returns averaged or accumulated across h periods. We collect all of these forms. While the specific transformations of stock returns and sentiment,[4] the vector of controls (X), the data frequency, and other model specifications may differ, we use standard meta-analysis tools (described in the following sections) to make all the results comparable.

2.1 Data Collection Process

In the data collection process, we closely followed the guidelines set out by Havránek et al. (2020) for meta-analysis in economics. As a first step, we searched for relevant studies in Google Scholar, which provides a broad, open-access database of research papers. We used the query: *("return" OR "stock market") AND ("sentiment" OR "belief" OR "confidence")* and downloaded the first 600 articles. This search was made on November 25, 2020, and any studies added beyond that date were only included using the "snowballing" method.[5] This procedure allowed us to identify an additional 123 studies, yielding a total of 723 articles for initial screening.

As a next step, we reviewed the abstract of each study and eliminated 344 of them because the topic was not relevant. We read the rest of the articles and reduced their number based on the purpose of our analysis. In addition to the envisioned characteristics of sentiment and the return series discussed at the beginning of Section 2, that is, survey-based sentiment and simple or excess stock return series, our main criteria were the following: (1) the article must report standard errors or other statistics from which the standard error can be calculated; (2) sentiment must not be defined as a dummy variable (e.g., Akhtar et al., 2011); (3) sentiment must not interact with any other continuous variable (e.g., Ahmed, 2020) but may interact with a dummy variable (e.g., Smales, 2017); (4) the effect must not be a spillover,[6] meaning that the estimate of sentiment in one country must not be connected with the stock market in a different country (e.g., Sayim and Rahman, 2015); (5) information about the standard deviations of both sentiment and the return series has to be available either in the primary study or via external sources (see the discussion below); and finally, (6) estimates based on impulse-response functions returned by VAR-type models must not have been considered.[7] All six points of our selection procedure represent a trade-off between having a sufficient number of observations for meta-analysis and our desire to keep the collected elasticities mutually comparable.

[4] As discussed above, we collected either simple returns or simple returns in excess of the risk-free rate. The possible transformations of the independent variable - sentiment – are: simple level, difference between two periods, logarithm, and growth rate. We collect all these types of sentiment transformations.

[5] *Snowballing* is a practice in meta-analysis which involves reviewing the references in each of the relevant studies returned by the initial search and identifying additional articles for screening. The aim of this method is to enlarge the set of estimates collected by reviewing articles that were not delivered by the initial search.

[6] During the screening process, we identified five studies which analyzed spillover effects of sentiment, and the number of point estimates was too small to consider it as a separate characteristic. Nevertheless, we admit that by filtering out spillover effects, we removed one part of the story. Specifically, Baker et al. (2012) showed that "global" sentiment is to a large extent responsible for individual countries' movements in returns, while country-specific sentiment – "local" sentiment – does not play that much of a role.

[7] During the paper collection process, we identified only a few articles that reported impulse-response functions returned by VAR-type models to describe the underlying relationship. In order to keep the heterogeneity of the estimates at a sustainable level, we decided not to harvest the elasticities from these methodologically different models.

Our final dataset contains 1311 estimates from 30 articles.[8] Before finalizing the dataset, we made sure that our primary studies are represented by journal publications wherever possible. In other words, if the relevant study was a working paper, we checked for a corresponding journal publication, and if there was one, we replaced the former with the latter. See Figure A1 for a brief flow diagram of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The list of primary studies that form our final dataset is available in Table 1. Nine articles contribute more than 50 observations to the overall sample size. More importantly, the mean elasticities of all nine are negative, suggesting that improvements in sentiment have adverse effects on returns. Nevertheless, the story is not that simple. As shown in Figure 2, the elasticities reported in the majority of articles span negative and positive values. Finally, it is notable that only two out of the 30 studies in our sample were published before 2000. This relatively short time span is due to the fact that the research field of behavioral finance, to which the examination of sentiment effects clearly corresponds, started to gain in importance mainly in the last two decades.

Table 1: Articles Included in Meta-analysis

Primary Study	**Obs.**	**Mean**	**Primary Study**	**Obs.**	**Mean**
Bathia and Bredin (2013)	99	-0.210	Hengelbrock et al. (2013)	52	-0.180
Ben-Rephael et al. (2012)	9	0.002	Ho and Hung (2012)	28	-0.001
Bremmer (2008)	9	0.207	Jiang et al. (2019)	4	-0.105
Brown and Cliff (2005)	144	-0.035	Kale and Akkaya (2016)	5	0.035
Chan and Fong (2004)	12	-0.769	Liston (2016)	13	0.253
Charoenrook (2003)	38	-0.068	Otoo (1999)	3	0.079
Clarke and Statman (1998)	3	-0.018	Perez-Liston et al. (2018)	2	0.229
Concetto and Ravazzolo (2019)	7	-0.069	Rakovská (2021)	210	-0.010
Corredor et al. (2015)	3	-0.078	Rashid et al. (2014)	4	0.145
Fernandes et al. (2013)	100	-0.288	Schmeling (2007)	5	0.321
Fisher and Statman (2000)	10	0.014	Schmeling (2009)	74	-0.517
Fisher and Statman (2003)	54	-0.005	Smales (2017)	168	-0.065
Gao et al. (2020)	15	0.023	Stambaugh et al. (2012)	19	-0.327
Grigaliūnienė and Cibulskienė (2010)	208	-0.111	Sum et al. (2014)	1	0.003
Güneş and Çelik (2009)	6	-0.003	Zhou (2018)	6	-0.084

Note: This table lists all the primary studies employed in the meta-analysis, together with the number of observations per study and the study-level mean elasticities.

The data collection was performed by all three authors. Each author was assigned one third of the articles, from which (s)he harvested the elasticities together with the corresponding standard errors (or p-values and t-statistics) and a wide range of other characteristics underlying the data, the model specification, the estimation design, and information about the publication itself. This initial data collection by each author was then cross-examined and validated by the other two authors in several rounds to reveal potential systematic errors and to ensure the overall comparability of the data compiled.

Before conducting the analysis proper, we made several adjustments to the data. First, we calculated standard errors for all the elasticities for which this information was not available, using the t-statistics or p-values that were reported instead. Second, to cope with data that were relatively

[8] Ioannidis et al. (2017), who investigated 159 meta-analytical articles in economics, reported a mean number of estimates in meta-analyses of 400 and a mean number of studies of about 42. In this sense, the number of observations in our analysis may be considered sufficient, while the number of studies is slightly below the average. Nevertheless, there are other published meta-analyses that work with lower numbers of observations and primary studies. For instance, Imai et al. (2021) use only 220 observations from 28 studies and Zigraiova and Havranek (2016) employ 598 observations from 31 studies.

heterogeneous in terms of the typology of the sentiment and return series (see Section 2.2), we standardized the elasticities and the corresponding standard errors using the "full standardization" methodology described in Bowman (2012). Specifically, we multiplied each elasticity and standard error by a factor equal to the ratio SD_x/SD_y, where SD_x is the standard deviation of the independent variable (=sentiment) and SD_y is the standard deviation of the dependent variable (=returns).[9] After full standardization, all the elasticities – regardless of their original units of measurement (levels or percentages) and, in the case of sentiment, also regardless of their original range[10] – represent the estimated effect on returns measured in standard deviations implied by a one standard deviation increase in sentiment. Third, we divided all the elasticities and their standard errors that corresponded to cumulative effects by the relevant number of periods.[11] Fourth, we transformed all the standardized elasticities and their standard errors so that they capture monthly changes. For example, the statistics originally reflecting weekly (quarterly) changes were multiplied by four (one third), and so on. Finally, we winsorized both the elasticities and the standard errors at 1% from each side to remove the noisy effects of outliers.

2.2 Main Categories of Data

As we have already mentioned, we need to keep some level of data heterogeneity in our meta-analysis so that we can investigate what drives the different results across studies. Naturally, some data characteristics are more powerful in this regard than others. Yet the importance of such characteristics does not need to be confirmed statistically. Intuition and theoretical underpinnings are often equally important. In this section, we introduce three main data characteristics that we assume to be influential in driving the heterogeneity. Later on, we segment our dataset based on these characteristics and inspect whether the elasticities differ in the resulting subsamples and whether there are any signs of publication bias conditional on the characteristics themselves.

The first way we look at the data is linked to the typology of the return series. Specifically, we distinguish between returns on the aggregate market (such as in Schmeling, 2009), returns computed for a specific portfolio of stocks, for example, growth stocks or stocks of small companies only (such as in Stambaugh et al., 2012), and returns on publicly published equity indexes, for example, the S&P 500 and the Euro Stoxx 50 (such as in Hengelbrock et al., 2013).

The existing literature provides empirical support for diverse effects of sentiment in the cross-section of returns. In their pioneering study, Baker and Wurgler (2006) showed that future-period returns on portfolios that are formed by stocks with a certain characteristic (small, young, highly volatile, non-dividend paying, or of extreme growth) are more prone to sentiment than portfolios formed by stocks of the opposite nature. Our typology of returns is different in the sense that we group various portfolios into one category – *Portfolio* – and add two additional "high-level" return categories – *Aggregate market* and *Equity index*. We do not distinguish between various types of portfolios in our analysis, because the data would get too granular,

[9] In approximately one third of cases, we were not able to gather information about SD_x (20% of the elasticities) and SD_y (45% of the elasticities) from the primary studies. In all of these cases, we calculated the standard deviations manually, that is, we downloaded the series described in the primary study using Datastream, employed the prescribed transformation, and calculated SD.

[10] For example, the most commonly employed sentiment variable in our dataset – The University of Michigan Consumer Confidence Index, ranges from 2 to 150, while the Consumer Confidence Index calculated by the Directorate-General for Economic and Financial Affairs spans between -100 and 100. Lastly, the bull-bear spread calculated from Investor Intelligence data may be expressed in the interval of <-1, 1> or <-100, 100> depending on the preferences of the researchers.

[11] We also collected elasticities that correspond to average effects across *h* periods. Nevertheless, these elasticities are already "single-period" measures, so we did not apply any adjustments in this case. We further control for these characteristics in the analysis of the drivers of heterogeneity.

Figure 2: Variation of Elasticities for Sentiment–return Relationship in Primary Studies

Note: The figure depicts box plots of the elasticities across the individual primary studies. The elasticities are transformed to monthly effects, fully standardized (see Section 2), and also winsorized at 1% from each side. The red vertical line represents the mean across the whole sample.

leaving us with only a few observations per portfolio category.[12] The first part of Table 2 shows the mean elasticities calculated for subsamples determined by the three return categories. The mean coefficients are negative for all of them. The strongest effect (-0.26) is observed for the aggregate market, while the weakest effect stems from the equity index returns (-0.08). With weighted data (weighted by the inverse of the number of estimates reported per study) the difference is even bigger (-0.31 and -0.02, respectively). This finding suggests either that there is significant publication bias (researchers tend to publish stronger results for the aggregate market) or that this effect is truly stronger when one considers aggregate market returns. However, this category of

[12] In fact, we collected the data on the different categories of portfolios in the same way as the characteristics of the stocks from which they are created. We also included the variables underlying this typology in the analysis, but we obtained inconclusive results that were mostly non-robust. For example, the theory would suggest that the returns on small stocks are more sensitive to changes in sentiment than those on large stocks (Baker and Wurgler, 2006, 2007). Nevertheless, even the average elasticities calculated for these cases showed that this premise does not hold in our dataset: the mean elasticity was -0.06 for portfolios of small stocks and -0.07 for large stocks.

returns is represented relatively weakly in the data (four studies and 5% of the observations only), so we must take this finding with caution.

Table 2: High-level Summary of Standardized Elasticities in Subsamples

			Unweighted			**Weighted**		
	Obs	Studies	Mean	5%	95%	Mean	5%	95%
All elasticities	1311	30	-0.11	-0.66	0.19	-0.05	-0.67	0.35
Different returns:								
Ret: Aggregate market	67	4	-0.26	-0.86	0.11	-0.31	-0.86	0.12
Ret: Equity index	687	23	-0.08	-0.51	0.24	-0.02	-0.55	0.36
Ret: Portfolio	557	10	-0.14	-0.77	0.05	-0.12	-0.82	0.23
Sentiment of different investors:								
Sent: Individual	735	25	-0.17	-0.83	0.11	-0.12	-0.97	0.28
Sent: Institutional	198	12	0.04	-0.07	0.6	0.14	-0.06	0.75
Sent: Other	378	6	-0.08	-0.38	0.18	-0.09	-0.36	0.15
Different effect horizons:								
Contemporaneous	63	8	0.12	-0.02	0.30	0.15	0.00	0.30
Short-term (up to 1Y)	1041	25	-0.11	-0.55	0.16	-0.10	-0.74	0.23
Long-term (more than 1Y)	207	6	-0.20	-0.99	0.28	-0.09	-1.16	0.39

Note: This table presents the mean effect of sentiment on monthly stock returns based on different data categories. In the second part of the table, we use the inverse of the number of estimates reported per study as weights to assign each study the same importance. The estimates are fully standardized and winsorized at 1% from each side.

The second categorization is related to the independent variable from Equation 1 – sentiment. As we have already highlighted, in this study we focus solely on sentiment derived from surveys. Even though it may seem relatively homogeneous at first glance, survey-based sentiment can be further categorized into subgroups based on the type of survey respondent.[13] In our framework, we focus on sentiment classes based on the type of respondent, distinguishing between the sentiment of *institutional* and *individual* investors/respondents.

We employed a two-step procedure to categorize the sentiment measures into groups. First, we reviewed all the primary studies and marked a given measure as *individual* or *institutional* when at least one primary study identified it as so. Second, we went through all the remaining sentiment measures and marked them as *individual* when they came from consumer surveys and as *institutional* when they came from investor or business surveys. We classified all the sentiment measures that remained uncategorized after this second step as *other*.[14]

Motivated by the existing theoretical and empirical literature, we expect the estimated effects of the two measures of sentiment to differ. The competing behavior of individual investors (irrational noise traders, or "dumb money") and their sophisticated institutional counterparts (rational arbitrageurs, or "smart money") was discussed long ago in the classical texts (Shleifer and Summers, 1990; De Long et al., 1990b). The actions of individual investors are consistent with the

[13] In fact, the wide variety of existing questionnaires causes immense heterogeneity in the sentiment measures that can be constructed from them. The resulting sentiment measures vary mainly in scale. This is why we applied the full standardization method described in Section 2.1.

[14] The category *other* comprises either the Economic Sentiment Index constructed by the Directorate-General for Financial and Economic Affairs (ESI DG ECFIN) or a composite-type index such as Zhou (2018) or Rakovská (2021). The first represents the general economic situation and the latter typically combines both individual and institutional sentiment measures.

overreaction hypothesis (Barberis et al., 1998), which states that uninformed investors overreact to consistent patterns of news with the same sign – good news induces optimism, while bad news induces pessimism. What is more, this overreaction on the markets creates a negative relation between sentiment and future returns in the medium to long term, because prices eventually get back to their intrinsic values. In contrast, institutional investors form fully rational expectations about the future dynamics of the stock market (Shleifer and Summers, 1990) and hence correctly predict future returns – they realize that prices are away from their intrinsic values and become pessimistic when noise traders are optimistic, and vice versa (Schmeling, 2007). Such tendencies are supported by the descriptive statistics in Table 2, where the mean elasticity for individual sentiment is negative (-0.17, or -0.12 in the weighted data), while the mean effect of institutional sentiment is positive (0.04 and 0.14, respectively). Nevertheless, simple density plots for these categories (Figure 1, Panel B) show that the positive mean effect for institutional sentiment might be a bare representation of a few strongly positive observations. Of course, the theory also offers an explanation for the negative effect of institutional sentiment. This explanation lies in the positive feedback theory (De Long et al., 1990a), in which sophisticated investors join individual ones instead of betting against them just to skim off the first profits. This implies that both individual and institutional sentiment may have negative effects on future stock returns due to price reversal.

The third and the last categorization focuses on the length of the horizon through which the author calculates the future returns and hence on the time horizon at which the sentiment effect works. We define three classes of effect horizon: (i) *contemporaneous* – when sentiment and returns enter Equation 1 at time t; (ii) *short-term* – when the delay between sentiment and the response variable is between one and twelve months; and (iii) *long-term* – when the delay is larger than one year.

Based on the famous concepts of behavioral finance – noise trader risk (De Long et al., 1990b) and limits to arbitrage (Shleifer and Summers, 1990) – price reversal does not happen immediately and prices can remain away from their fundamental value for a protracted period of time. Hence, during this period, we observe a positive relation between sentiment and future (and contemporaneous) stock returns. According to the underreaction hypothesis (Barberis et al., 1998; Hong and Stein, 1999) this positive association lasts only in the short run (up to twelve months), because in this time window investors are too conservative in changing their strategies – they tend to underreact to new evidence. However, the empirical support for a positive effect in the short run is inconclusive, as some authors report negative results for this time frame (Brown and Cliff, 2005; Schmeling, 2007; Rakovská, 2021). This is probably the main reason why Table 2 shows a negative mean elasticity for the short term and why Panel B in Figure 1 demonstrates a fatter left tail for the particular subsample of elasticities. In contrast, longer time horizons provide more space for price corrections, yielding net negative effects of sentiment. Consistent with the overreaction hypothesis of Barberis et al. (1998), uninformed traders extrapolate the current situation to the future, push prices far away from their intrinsic values, and then face adverse consequences in the form of price reversals. This *contrarian view* enjoys ample empirical evidence in the literature (Brown and Cliff, 2005; Baker and Wurgler, 2006; Schmeling, 2009). The mean elasticities listed in Table 2 for long-term effects also support this view.

3. Publication Bias

Publication bias means that the results reported in the literature deviate systematically from the results that are actually achieved. It is typical of all research fields, and economics (Doucouliagos and Paldam, 2011; Havranek and Kokes, 2015; Ugur et al., 2018; Campos et al., 2019; Blanco-Perez and Brodeur, 2020; Duan et al., 2020; Brown et al., 2021) and finance (Zigraiova and Havranek, 2016; Geyer-Klingeberg et al., 2018; Astakhov et al., 2019; Kim et al., 2019) are no exceptions. Typically, as documented by Ioannidis et al. (2017), the results published in economic articles are overestimated to such a degree that they are twice as large as the true effects.

In general, researchers should not be influenced by what they wish to see in the results. However, given the typically high level of freedom in empirical analysis, including the vast range of possible model specifications, researchers can eventually end up with the results that they desired prior to the analysis. There are several reasons behind the publication bias phenomena. An important one is the procedure used to evaluate published articles. Typically, published articles are more likely to report statistically significant results, often with the desired signs, while there is not much desire to publish insignificant outcomes (Franco et al., 2014). As a result, insignificant estimates are generally underrepresented in the empirical literature, and this creates a bias.

As far as we know, there is no meta-analytical study that assesses the relationship between sentiment and stock market returns. While we cannot draw on other systematic review studies on this topic, we are still able to formulate three reasons why we believe this particular research field may be prone to publication bias. The first was already discussed in Section 2.2, where we documented differences in the reported estimates (including their theoretical underpinnings) conditional on selected characteristics of the dependent and independent variables and the effect horizon. The second involves the famous critique of some well established empirical articles on this topic. The classical example is the response of Chen et al. (1993a) to Lee et al. (1991) about the effects of sentiment proxied by closed-end fund discounts in the sector of smaller capitalization stocks. The critique was based on the premise that the original results were sensitive to the inclusion of specific calendar months.[15] More recently, Cheema et al. (2020) issued a critique of Han and Li (2017) in which they argue that the strong momentum signal of sentiment reported for the Chinese stock market is not robust and in fact highly affected by the crisis years of 2006–2008. Lastly, there is no reason to believe that among all the topics in economics this one is an exception.[16] Sentiment–return research is mostly conducted by the same researchers, reviewed by the same referees, and published by the same editors, and their motivation to prefer some results to others may also the same.

3.1 Preliminary Analysis and Methodology

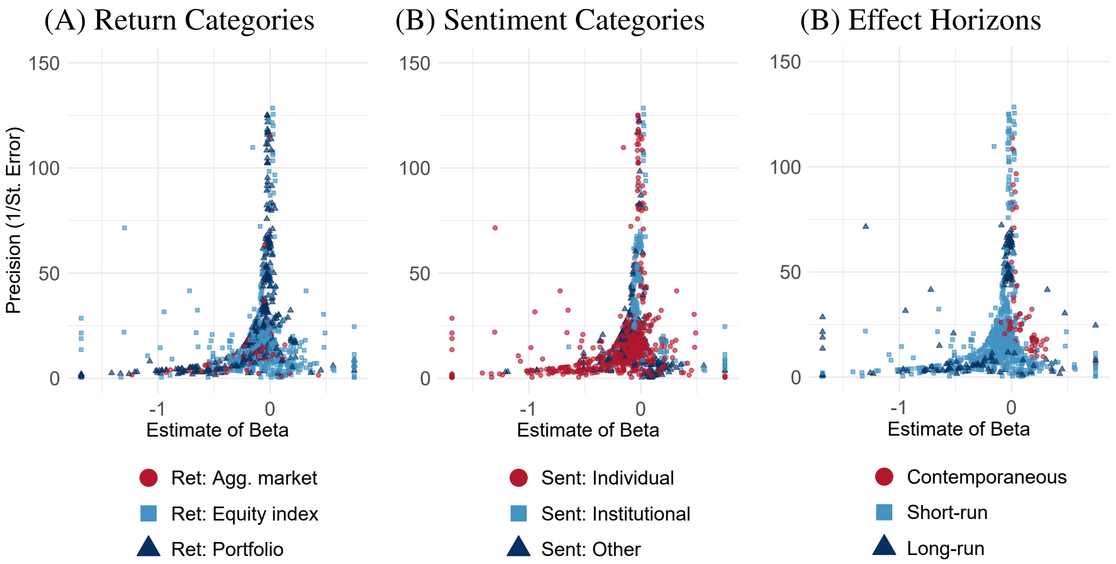
There are several established ways of assessing possible publication bias. First, we use a simple graphical assessment – a funnel plot, proposed by Egger et al. (1997), which plots standardized elasticities on the horizontal axis and their precision on the vertical axis. The precision of each elasticity is defined as the inverse of the corresponding standard error. If there are no selection tendencies, there is also no relationship between the elasticity and its precision. In other words, if publication bias is absent, we should observe scatter plots symmetrically and randomly distributed around the mean effect. In addition, as the measured elasticities get closer to the mean (true) effect, the precision should increase, forming the graph into an inverted funnel. In the presence of selection bias, the funnel is asymmetric (one direction is preferred) or shows some hollow parts (imprecise estimates are discarded).

[15] The public discussion then continued in Chopra et al. (1993b) and Chen et al. (1993b) and ended with an official statement that the original results were weak but still significant enough compared with similar empirical frameworks (Chopra et al., 1993a).

[16] For example, publication bias was found recently by Astakhov et al. (2019) in a similar area: the effect of firm size on stock returns.

Figure 3 depicts funnel plots in which we distinguish between the three main categories described in Section 2.2. If we focus on the overall shape of the funnel plots without any categorization, it seems that the estimates are distributed relatively symmetrically around the mean, which is very close to zero. The only disturbance is that the range to the left of the mean is clearly longer than that to the right. However, the funnel shape under no publication bias is violated much more obviously when we look at particular categories. The effect is skewed to the left for portfolio returns (Panel A), sentiment of individual investors (Panel B), and for long-run estimates (Panel C). In other words, it seems that the primary studies have a slight preference for reporting negative estimates for these characteristics, a tendency which corresponds to the theoretical underpinnings outlined in Section 2.2. In contrast, the contemporaneous effects are skewed slightly to the right, with almost no representation in the negative part of the x-axis (Panel C). Thus, positive estimates for the contemporaneous effects of sentiment seem to be preferred in the literature. This again fits the theory summarized in Section 2.2, namely, the overreaction hypothesis (Barberis et al., 1998).

Figure 3: Funnel Plots for Different Subsamples



Note: In the absence of publication bias the plots should resemble inverted funnels symmetric around the most precise estimates. The individual panels depict the funnel plots for the subsample groups: Panel A – categories of returns, Panel B – categories of sentiment, and Panel C – categories of effect horizons. Estimates with a precision greater than 150 are excluded from the graph for ease of exposition but are included in all the statistical analysis.

Obviously, funnel plots serve only as an initial visual assessment of publication bias. Still, they can provide us with a useful clue about its direction. As mentioned above, in the absence of systematic selection bias there should be no relationship between elasticity and its precision. Thus, a more sophisticated formal test for the presence of publication bias can easily be derived from a simple regression analysis based on the following equation:

$$\beta_{ij} = \alpha_0 + \sigma_0 SE(\beta_{i,j}) + z_{ij}, \tag{2}$$

where the reported elasticity i in study j – β_{ij} – is explained by its standard error $SE(\beta_{i,j})$. Coefficient σ_0 measures the magnitude of the publication bias, while coefficient α_0 shows the effect beyond this bias. If there is no selection bias in the data, σ_0 is statistically equal to zero. If

σ_0 is different from zero, there is a relationship between β_{ij} and its standard error which must have been caused by a systematic preference in the result-reporting process.

As a robustness exercise, we estimate Equation 2 using several additional methods. We start with linear techniques. First, we try to control directly for distortion via similarities in a given study by adding study-level fixed effects and, for comparison, also study-level random effects. Second, we investigate how the results change when greater importance is attached to more precise results. Specifically, we run OLS with the weights set as the inverse of the standard errors.

A strong assumption of all the aforementioned methods is that the relationship between β_{ij} and its standard error is linear. However, if there is a significant relationship between these variables and it is not driven by a linear function, those methods either would not identify publication bias at all, or would generate a bias of their own. Therefore, we also employ several non-linear procedures that have been developed recently. These methods do not show the magnitude of the publication bias, but display the value of the corrected mean beyond the bias, for example, what the relationship would look like if there was no publication bias in the estimates.

First, we run the top 10 method proposed by Stanley et al. (2010). It is called top 10 because it runs the underlying regression only for the 10% of observations with the highest precision and discards the rest. This step should theoretically reduce the publication bias and thus enable the mean effect to be estimated more efficiently. Second, we use the stem-based method by Furukawa (2019). This procedure assumes that the publication bias decreases with rising precision of the results, and as such it calculates the mean effect considering the trade-off between the bias and the variance. Finally, we employ the kinked method (Bom and Rachinger, 2019), which aims to find the threshold between a linear and non-linear relationship in publication bias.[17]

3.2 Results

The outcomes of estimating Equation 2 using linear and non-linear techniques are presented in Table 3. In the analysis of publication bias, we distinguish between the three main categories introduced in Section 2.2. Therefore, we provide results not only for the full dataset, but also for the subsamples determined by those categories. A bird's-eye view of the results of the non-linear techniques (Panel B) suggests that the mean effect of sentiment on stock market returns adjusted for publication bias is near zero, even though it is significant in the majority of cases. We included the non-linear techniques because they are state-of-the-art methods of meta-analysis. Nevertheless, we acknowledge that in our case, namely, when the effect of sentiment on stock returns is rather small or close to zero (Table 2), linear techniques might be better than non-linear ones (see the discussion in Havranek and Sokolova, 2020). It is also apparent that the linear techniques are more conservative and show smaller publication bias, as the corrected mean effect for the non-linear methods is even closer to zero than that for the linear ones. Having this in mind, we narrow the subsequent discussion to the linear techniques (Panel A).

When considering the whole sample of standardized elasticities (column 1), we can say that the mean effect beyond bias (the true effect) of sentiment is negative and in the majority of cases also significant. Conversely, publication bias is significant only when higher weight is given to data with higher precision. However, a more detailed examination of the publication bias tendencies in the subsamples reveals some interesting features.

[17] Cazachevici et al. (2020) and Gechert et al. (2020) offer more discussion of these non-linear methods.

Table 3: Publication Bias

	Return Categories				Sentiment Categories			Effect Horizons		
	All	Aggregate	Equity Index	Portfolio	Individual	Institutional	Other	Contemp.	Short-term	Long-term
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Linear Techniques										
Study-level fixed effects										
Effect beyond bias	-0.095**	-0.138***	-0.089*	0.031**	-0.070*	-0.050	-0.088.	-0.078**	-0.101**	-0.078
	(0.034)	(0.031)	(0.041)	(0.011)	(0.033)	(0.032)	(0.050)	(0.023)	(0.032)	(0.051)
Publication bias	-0.152	-0.699**	0.083.	-1.985***	-0.767*	1.963**	0.054	4.277***	-0.089	-0.972***
	(0.208)	(0.229)	(0.047)	(0.207)	(0.334)	(0.750)	(0.065)	(0.334)	(0.171)	(0.063)
Study-level random effects										
Effect beyond bias	-0.046	-0.103.	-0.042	0.054	-0.031	0.023	-0.103*	-0.062*	-0.103*	-0.058
	(0.037)	(0.052)	(0.035)	(0.033)	(0.038)	(0.039)	(0.050)	(0.026)	(0.041)	(0.128)
Publication bias	-0.204	-0.892***	-0.036	-1.982***	-0.597*	1.317.	0.054	4.221***	-0.149	-0.954***
	(0.200)	(0.180)	(0.134)	(0.185)	(0.235)	(0.727)	(0.065)	(0.359)	(0.179)	(0.097)
Weighted OLS										
Effect beyond bias	-0.004*	-0.001	-0.005.	-0.001	-0.004*	-0.046.	-0.005	-0.004	-0.004**	-0.012
	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.022)	(0.005)	(0.006)	(0.001)	(0.017)
	[-0.025, 0.000]	[-0.055, 0.042]	[-0.032, 0.000]	[-0.048, -0.005]	[-0.017, 0.001]	[-0.094, 0.006]	[-0.211, 0.363]	[-0.083, 0.218]	[-0.028, 0.000]	[-0.359, 0.002]
Publication bias	-0.891**	-1.454**	-0.492	-1.613***	-1.258***	1.865*	-0.513	2.671**	-0.841*	-1.511**
	(0.291)	(0.179)	(0.290)	(0.144)	(0.266)	(0.815)	(0.498)	(0.625)	(0.315)	(0.296)
	[-1.521, -0.342]	[-1.786, -1.288]	[-1.461, -0.022]	[-1.804, -1.175]	[-1.793, -0.688]	[-0.108, 3.230]	[-2.118, 1.449]	[1.048, 3.775]	[-1.641, -0.271]	[-6.497, -0.763]
Panel B: Non-linear Techniques										
Top 10 method (Stanley et al., 2010)										
Effect beyond bias	-0.005.	-0.001***	-0.003***	-0.011***	-0.003***	0.016***	-0.003**	0.016**	-0.003***	-0.089
	(0.003)	(0.000)	(0.000)	(0.002)	(0.000)	(0.005)	(0.001)	(0.006)	(0.000)	(0.063)
Stem-based method (Furukawa, 2019)										
Effect beyond bias	0.000	-0.001	0.000	0.000	0.000	0.006	-0.001	0.004	0.000	-0.030*
	(0.004)	(0.001)	(0.005)	(0.003)	(0.005)	(0.023)	(0.001)	(0.023)	(0.003)	(0.011)
Kinked method (Bom and Rachinger, 2019)										
Effect beyond bias	-0.003***	-0.001**	-0.002***	-0.004***	-0.003***	-0.003	0.000	0.003***	-0.003***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.001)	(0.000)	(0.002)
No. of observations.	1311	67	687	557	735	198	378	63	1041	207
Studies	30	4	23	10	25	12	6	8	25	6

Note: **Panel A:** Standard errors, clustered at the study level, are reported in parentheses. Whenever possible, 90% confidence intervals from wild bootstrap clustering are reported in square brackets; the procedure was implemented via the *boottest* procedure in R. The weighted OLS method uses the inverse of the standard errors as weights. **Panel B:** The number of observations for the top 10 method (Stanley et al., 2010) decreases to only 10% of the observations for the other methods. For example, it is 130 for the total sample and only 7 for the subsample of aggregate market returns.

First, when we look at the results for the subsamples determined by return characteristics (columns 2–4), we spot clear cases of negative publication bias for the *aggregate market* and *portfolio* returns. In other words, researchers who employ these types of return series tend to publish more negative results relative to the true effect. The evidence of publication bias for *equity index* returns is weak. Moreover, the only true effect beyond bias was identified by the OLS model with study-level fixed effects. This may suggest that if there is a real effect outside of selection bias in the return categories, it is visible only after controlling directly for heterogeneity among the studies investigated.

Second, as suggested by our preliminary funnel plot analysis (Figure 3), the elasticities for the *individual* and *institutional* categories of sentiment apparently suffer from publication bias (columns 5 and 6). Nonetheless, the direction of the bias differs. This points to an interesting phenomenon. When researchers employ the sentiment of individual investors as a proxy for sentiment, they tend to report more negative elasticities relative to the true effect. In contrast, when they use institutional sentiment as an independent variable, the reported effect is predisposed to be more positive than the true effect (which in this case is insignificant and hence probably not different from zero). This suggests that authors may tend to report predominantly negative (positive) elasticities for individual (institutional) sentiment because the theory and previous empirical studies suggest so (see Section 2.2).

Third, similarly to the previous typology, different effect horizons also show compelling patterns in terms of selection bias (columns 8–10). Again, we observe bias of opposite directions, this time for the *contemporaneous* and *long-term* effects of sentiment. Specifically, the use of the former specification category yields more positive elasticities in the studies, while the use of the latter gives rise to a tendency to publish negative estimates even though the true effect seems to be negligible. The results for the category of *short-run* effects are inconclusive, as only the weighted OLS yields a significant estimate of the bias. The different effect horizons offer an interesting story for the effects beyond bias. On the one hand, the true effect for future horizons, both short-term and long-term, is negative, but it is significant only in the case of the latter. In other words, the longer the horizon, the stronger the contrarian view underpinned by the overreaction hypothesis (Barberis et al., 1998). Moreover, one might view this result as a manifestation of limited arbitrage (De Long et al., 1990b; Shleifer and Summers, 1990) in the short run. On the other hand, the true contemporaneous effect seems to contradict the classical literature, as it turns negative and significant in two out of the three linear specifications. Given the positive sign of the publication bias for this subsample, our results suggest that only in theory do stocks become immediately overvalued when sentiment increases. Nevertheless, this outcome has to be taken with caution, as there are only 63 observations behind it.

All in all, we find that the true effect of sentiment on stock returns is non-negligible and negative. Moreover, in the majority of specifications, researchers report exaggerated estimates. A more detailed analysis also reveals that in two specific specifications – the use of institutional sentiment and the contemporaneous effect specification – researchers tend to report positive elasticities, while the true effects are near zero and negative, respectively. The reason for this selection bias might be that authors are willing to publish results that fit the theory, but discard unsatisfactory outcomes.

4. Drivers of Heterogeneity

In the previous section, we showed that there is significant bias in the empirical literature on the sentiment–stock return relationship. That is, the authors of primary articles prefer some estimates to others. Nevertheless, one may simply object that Equation 2 is subject to omitted variable bias, because many other factors might influence the relation between survey-based sentiment and (contemporaneous and future) stock market returns. Therefore, we now focus on the reasons why those elasticities differ in the first place. For this purpose, we collected 47 additional variables (beside the elasticity and its standard error) that may explain the differences in the reported sentiment effects.[18] We focus on the drivers of heterogeneity commonly used in meta-analysis and also on the characteristics that are specific to our topic.

We have already described the first group of variables that will enter the subsequent analysis in Section 2.2. Besides these three main categories (returns, sentiment, and the horizon of the effect), the primary studies offer other interesting specifics, and the standard meta-analytical literature motivated us to collect other variables related to different study designs with respect to data frequency and methodology (Bajzik et al., 2020; Zigraiova et al., 2021). In addition to variables directly harvested from primary studies, we collect external variables (Ehrenbergerova et al., 2021). Descriptive statistics for all those variables are given in Tables 4 and 5. For clarity and better orientation, we categorize them into the classes described below.

4.1 Variables from Primary Studies and External Variables

Data Characteristics. In this category, we focus on specific properties of the dependent and independent variables entering Equation 1 in addition to those covered in Section 2.2, as well as the identification of the region in question. Beginning with the dependent variable, we distinguish between two categories that determine whether the author used *excess* returns or not. Excess returns are defined as returns minus some risk-free rate, usually represented by the 3-month or 1-month government bond yield. The majority of the primary studies use simple returns (around 90% of our data). As excess returns are generally smaller than simple returns,[19] it is evident that the effect of a change in sentiment on excess returns is also generally smaller than the effect on simple returns. This conclusion is supported by the results in one of our primary studies, Liston (2016), which reports estimates for both simple and excess returns when studying the effect of sentiment in the US sin stock market. The mean standardized elasticity is 0.46 for the specifications with simple returns and only 0.16 for those with excess returns.

As for the independent variable, we also investigate whether the index construction methodology matters. Specifically, we distinguish between sentiment measured as the difference between the shares of bullish and bearish respondents in total respondents (*bull-bear spread*) or just the share of bullish respondents in the total (*bullish*).[20] The former can be simply interpreted as the respondents' optimism *net* of their pessimism, while the latter represents raw optimistic beliefs. The vast majority of the analyses are based on bull-bear spread measures of sentiment.[21] However, when a study does not account for bearish respondents (such as in Chan and Fong, 2004), it misses the information on pessimistic beliefs in the market and the final effect may be different.[22] This is supported by

[18] In fact, our original data table included more than 200 columns in which we stored the characteristics underlying each estimate. In the end, some of the characteristics were merged, some of them were discarded as they turned out to be irrelevant, and others were removed in the later stages of the analysis, due mainly to multicollinearity.

[19] It is worth mentioning that when the government bond yield is negative (not that unusual a phenomenon over the last decade), the excess return is obviously higher than the simple one.

[20] In the data collection process, we discarded the elasticities that corresponded to sentiment measured as the share of bearish respondents in the total. The reason is twofold. First, there were only three such elasticities, and second, we believe that merging it with either the *bull-bear spread* or the *bullish* category would be erroneous, as the *bearish* sentiment captured opposite beliefs to the other categories.

[21] For example, we assume all the consumer confidence indexes to be constructed as bull-bear spreads.

[22] As documented by Greenwood and Shleifer (2014) and Zhou (2018), measures based on bull responses are highly correlated, as are measures based on bear responses.

the weighted and unweighted mean standardized elasticities for these two categories reported in Table 4.[23]

Table 4: Description and Summary Statistics of Main Variables – Part 1

Variable	**Short description**	**Mean**	**S.D.**	**Elast. Unw.**	**Elast. W.**
Estimate	The reported estimate of the beta coefficient.	-0.11	0.3	-	-
St. Error	The reported standard error of the beta coefficient.	0.12	0.26	-	-
Data characteristics:					
Ret: Simple	= 1 if the DV represents returns that are not abnormal or in excess of the risk-free rate.	0.9	0.3	-0.12	-0.05
Ret: Excess	= 1 if the DV represents returns that are in excess of the risk-free rate (e.g., the policy rate).	0.1	0.3	-0.06	-0.06
Ret: Agg. market	= 1 if the DV represents returns for an overall/aggregate market, e.g., the aggregate returns of all the stocks traded on the stock exchange.	0.05	0.22	-0.26	-0.31
Ret: Equity index	=1 if the DV represents equity index returns (e.g., S&P 500, DAX, Russell 2000).	0.52	0.5	-0.08	-0.02
Ret: Portfolio	= 1 if the DV represents returns on a portfolio (e.g., a portfolio formed based on size, book-to-market ratio, or other characteristics calculated manually by the author).	0.42	0.49	-0.14	-0.12
Sent: Individual	= 1 if the IV represents the sentiment of individual investors.	0.56	0.5	-0.17	-0.12
Sent: Institutional	= 1 if the IV represents the sentiment of institutional investors.	0.15	0.36	0.04	0.14
Sent: Other	= 1 if the IV represents the sentiment of neither individual nor institutional investors.	0.29	0.45	-0.08	-0.09
Sent: Bull-bear spread	= 1 if the IV is calculated as the difference between the shares of bullish and bearish respondents.	0.98	0.14	-0.11	-0.03
Sent: Bullish	= 1 if the IV represents the share of bullish respondents.	0.02	0.14	-0.37	-0.32
Sent: Orthogonalized	= 1 if the IV was orthogonalized to/adjusted for macroeconomic/fundamental factors (e.g., if it was decomposed into a "rational" and "irrational" part).	0.27	0.45	-0.09	-0.04
Sent.: Level	= 1 if the IV is in levels.	0.98	0.15	-0.12	-0.08
Sent.: Log	= 1 if the IV is in logs.	0.02	0.15	0.09	0.11
Sent.: Diff. or growth	= 1 if the IV is in differences or growth rates.	0.14	0.35	0.01	0.09
US	= 1 if the study only covers the US.	0.42	0.49	-0.07	-0.02
Europe	= 1 if the study only covers a country/countries from Europe.	0.5	0.5	-0.12	-0.09
Asia	= 1 if the study only covers a country/countries from Asia.	0.04	0.19	-0.21	-0.12
Other regions	= 1 if the study covers regions other than the US, Europe, or Asia.	0.05	0.21	-0.4	-0.02
No. of obs.	The natural logarithm of the number of observations in the study.	5.46	0.71	-	-
Model specification (Part 1):					
Cumulative	= 1 if the estimate represents a cumulative effect (e.g., the cumulative effect of four consecutive quarters).	0.05	0.23	-0.16	-0.14
Average	= 1 if the estimate represents an average effect (e.g., the average effect over four consecutive quarters).	0.35	0.48	-0.12	-0.08
Contemporaneous	= 1 if the elasticity corresponds to the contemporaneous effect.	0.05	0.21	0.12	0.15
Short-run	= 1 if the elasticity corresponds to the effect in future periods of up to one year.	0.79	0.4	-0.11	-0.1
Long-run	= 1 if the elasticity corresponds to the effect in future periods more than one year ahead.	0.16	0.36	-0.2	-0.09
Other Sent. in Eq.	= 1 if there is an additional sentiment variable in the equation. The additional sentiment variable might be the additional lag of the IV, the interaction of the IV with another binary or continuous variable, or another sentiment variable.	0.12	0.32	0.01	0.07
No controls	= 1 if the model does not contain macroeconomic, corporate finance, or financial market control variables.	0.16	0.36	-0.13	-0.07

Note: See the note to Table 5

[23] In our set of primary studies, we again find one study – Zhou (2018) – which reports elasticities for both categories of sentiment measures. However, the mean standardized elasticities for these two cases are very close to each other – -0.09 for the bull-bear spread and -0.08 for bullish sentiment.

Table 5: Description and Summary Statistics of Main Variables – Part 2

Variable	Short description	Mean	S.D.	Elast. Unw.	Elast. W.
Model specification (Part 2):					
At least 1 control	= 1 if the model contains at least one control variable from any of the categories: macroeconomic, corporate finance, financial market.	0.84	0.36	-0.11	-0.04
Time	= 1 if time series data were used.	0.94	0.23	-0.1	-0.05
Panel	= 1 if panel data were used.	0.06	0.23	-0.42	-0.2
Data freq: D/W	= 1 if the data frequency is daily or weekly.	0.06	0.23	-0.23	-0.16
Data freq: M	= 1 if the data frequency is monthly.	0.92	0.27	-0.11	-0.05
Data freq: Q/S/Y	= 1 if the data frequency is quarterly, semi-annual, or yearly.	0.02	0.15	0	0.05
One eq.	= 1 if a one-equation model is used.	0.66	0.47	-0.1	-0.06
More eq. & Other	= 1 if a multi-equation model (e.g., VAR, VECM) or any other model is used.	0.05	0.21	0.04	0.07
OLS	= 1 if the OLS estimator is used.	0.8	0.4	-0.11	-0.07
GMM	= 1 if the GMM estimator is used.	0.06	0.24	-0.14	-0.14
Panel FE	= 1 if panel fixed effects are used.	0.03	0.16	-0.67	-0.68
Other est.	= 1 if a method other than OLS, GMM, or FE is used.	0.11	0.31	-0.02	0.03
Dynamic	= 1 if the model is dynamic (e.g., if it includes a lagged DV).	0.1	0.29	-0.09	-0.06
System of eg.	= 1 if a system of equations is used.	0.29	0.46	-0.17	-0.26
Publication characteristics:					
Published	= 1 if the primary study was published in a journal with an impact factor.	0.96	0.2	-0.12	-0.07
Impact	The recursive impact factor.	0.19	0.4	-0.13	-0.1
Citations	The logarithm of the number of citations divided by the number of years from its publication until 2021.	3.74	2.36	-	-
Publication year	The logarithm of the publication year of the primary study minus the earliest publication year in our dataset plus one.	2.64	0.46	-	-
Preferred	= 1 if the reported estimate is preferred in the study (e.g., if it is the baseline estimate).	0.2	0.4	-0.08	0.04
Not preferred	= 1 if the reported estimate is not preferred in the study (e.g., if the preferred estimate is a robustness/sensitivity/additional/alternative estimate).	0.8	0.4	-0.12	-0.11
External variables:					
Ext: CPI growth	Average CPI inflation.	2.78	1.62	-	-
Ext: Unemployment	The average unemployment rate in %.	6.53	1.43	-	-
Ext: Credit to GDP	The average credit-to-GDP ratio.	133	24.78	-	-
Ext: Savings to GDP	The average savings-to-GDP ratio.	22.23	4.41	-	-
Ext: Fin. openness	The average financial openness index (Chinn and Ito, 2008).	2.2	0.42	-	-
Ext: Fin. development	The average financial development index (Garriga, 2016).	0.7	0.09	-	-

Note: Variables in *italic font* are baseline variables that do not enter the analysis of heterogeneity drivers because of the dummy variable trap. The last two columns represent unweighted and weighted mean elasticities, respectively. We use the inverse of the number of estimates reported per study as weights. IV = independent variable; DV = dependent variable.

We also focus on rather technical properties of the sentiment variable. We look at the possible impacts of data size and transformation of sentiment measures from levels into logs or growths/differences,[24] as a number of meta-analytical studies show that such characteristics might play a role (Cazachevici et al., 2020; Bajzik et al., 2020). One third of the elasticities collected from our primary studies correspond to the effect induced by "orthogonalized" sentiment (Table 4). Orthogonalization is a statistical procedure which simply removes part of the information from a variable, leaving it cleaned of (or not related to) a pre-selected set of factors. Several primary studies in our sample followed the pioneering studies in this field (Baker and Wurgler, 2006; Lemmon and Portniaguina, 2006) and engaged in orthogonalization. They cleaned sentiment of

[24] For example, Smales (2017) states that the impact on stock returns is stronger for sentiment changes than for levels.

macroeconomic/fundamental factors, believing that only the residual (irrational) component is the true sentiment measure, while the predicted values (the rational component) of the process represent just general economic conditions. Whether orthogonalization affects the estimated elasticities is still an open question. Baker and Wurgler (2006) argue that not removing fundamental factors from sentiment may increase the probability that it is connected to systematic risk. Nevertheless, the authors state that their cleaned and non-cleaned composite sentiment indexes induce similar effects on the cross-section of returns. In addition, Corredor et al. (2015), Liston (2016), and Rakovská (2021) assert that they employed the orthogonalization procedure because the pioneering studies did so, and do not provide results for unadjusted sentiment measures.

The region of the data may give rise to a significant difference in the sentiment effect. The majority of collected elasticities is related to Europe and the US (see Table 4), therefore, we relate the subsequent discussion only on these regions. The reasons why sentiment should be different in different locations is obvious. Chui et al. (2010) states that the main drivers lie in cultural differences. Moreover, Schmeling (2007) argues that irrational sentiment behaviour is stronger in the US compared to Europe and that it is mainly due to higher level of individualism in the US. Supporting this conclusion, Schmeling (2009) explains that sentiment effect is generally stronger in countries that are more prone to herding investment behaviour and weaker in countries with high institutional integrity. Ho and Hung (2012) even finds that taking Europe as a whole, the sentiment effect is not significant (although it may be a driver of future returns in some particular countries). However, not all researchers come to the same conclusion. For example, Hengelbrock et al. (2013) find only minor differences between the sentiment effects in Europe and US, while Bathia and Bredin (2013) do not reveal any significant differences.

Model Specification. The main determinant of interest to us in this class is the horizon of the sentiment effect, described in Section 2.2. Besides this factor, we investigate a wide range of characteristics. First, we distinguish between elasticities that represent a cumulative or average effect.[25] Second, we identify cases in which the equation from which we harvested the elasticity included some other variable connected to the main independent variable, such as another lag of sentiment (Otoo, 1999), another survey-based measure of sentiment (Schmeling, 2007), or another non-survey-based measure of sentiment (for example, the Baker and Wurgler (2006) composite sentiment index), or simply when sentiment was interacted with some dummy variable, such as a dummy for recession (Smales, 2017). Each of these cases was represented by only a few observations in our dataset, so we combined all of them into one indicator variable – *Other Sent. in Eq.*[26] Third, we also differentiate between cases where the estimation equation contained other important determinants of stock returns and where the researcher did not include any such control variables. Last but not least, we control for different types of data – be they *Time* series or *Panel* – and for different time frequencies.[27]

[25] In general, the authors of our primary studies do not provide a rationale for using single-period, cumulative multi-period, or average returns. Because the last two categories represent approximately 40% of the data (see Table 4), we are interested whether the different ways of expressing return series matter or not. Note that we divided elasticities which captured cumulative effects by the respective number of periods to obtain the effect for one "average" period (Section 2.1).

[26] Additional sentiment-related variables in one model equation may worsen the effect studied. For example, the inclusion of another sentiment variable may cause the harvested elasticity to capture only a part of the effect, while the elasticity for sentiment interacted with a dummy only describes the effect conditional on the characteristic analyzed.

[27] We transferred all the data in the same - monthly – frequency (see Section 2.1). However, the frequency of the data in the original study may still drive the heterogeneity of the estimates. As demonstrated by Rodriguez et al.

Finally, we also collected variables which allow us to examine the impact of various estimation techniques. Specifically, we distinguish between one-equation models and other econometric models, including models based on multiple equations. We further distinguish between four groups of estimation methods – simple Ordinary Least Squares (OLS), the Generalized Method of Moments (GMM), panel fixed effects (FE), and all others (*Otherest*).[28] As part of this categorization, we also differentiate between two specific characteristics: dynamic nature of the model (the inclusion of a lagged dependent variable) and models that were estimated as a system of equations.[29] Some of those technical characteristics (such as the estimation method) do not lead directly to economic implications and thus are usually not given much attention. In contrast, the results of every empirical analysis are, in some respects, determined by the properties of the estimation method selected.[30] For these reasons, it is useful to investigate these technical aspects in our case as well.

Publication Characteristics. In this group of variables, we consider aspects that should in some sense be correlated with the quality of an article. It is normal to control for such aspects in meta-analysis research, as shown, for example, by Cazachevici et al. (2020) and Matousek et al. (2021). First, we consider whether and when the article was published in a journal with an impact factor. Second, we try to measure article quality directly using the discounted recursive IDEAS/RePEc impact factor (IDEAS/RePEc, 2021). Further, we employ the smoothed volume of citations, assuming that higher quality articles are generally cited more.

Besides controlling for article quality, we also check whether there is any systematic difference in the reported coefficients based on the time of publication. Newer studies tend to use more sophisticated methods and procedures and can moreover gain from all the previous research. Therefore, it can be assumed that the results of recent studies are more precise than those of older research. We also consider whether the results were reported as the study's baseline outcome, or whether they were only reported as part of the robustness exercise. Interestingly, the weighted mean elasticity of the sentiment effect for the former is positive (0.04, see Table 5). This suggests that the overall negative effect (Table 4) may be driven by the robustness results.

External Variables. In addition to the characteristics collected directly from primary studies, we include six external variables. These variables are not only country- or region-specific, but also specific to the given estimate.[31] In this sense, we are able to control for cross-regional differences more carefully.

(2014), stock market efficiency positively depends on the time scale. Along the same lines, Kim et al. (2019) shows that stock markets are more prone to inefficiencies when measured with higher frequency data.

[28] OLS is by far the most common estimation technique, covering 80% of the elasticities collected. This might seem surprising, as economic research (like other research fields) is gradually shifting to more complex econometric models. Many authors of our primary studies are aware of the critique of OLS (such as in Stambaugh, 1999) and, in response, apply bias-correction methods (71% of the observations for the OLS technique), such as Newey and West (1987) standard errors (Corredor et al., 2015; Smales, 2017) and bootstrap simulation (Brown and Cliff, 2005; Schmeling, 2009). However, when analyzing the drivers of heterogeneity, we were forced to discard the variable which indicated the use of these bias-correction methods due to multicollinearity.

[29] The variable *System* does not coincide with the variable *More eq. & Other*, even though the two might seem similar at first sight. For example, Brown and Cliff (2005) and Rakovská (2021) use a one-equation model estimated as a system for *h* future horizons of returns. We coded the elasticities that come from these studies as *One eq.* = 1 and *System* = 1 and left the *More eq. & Other* category for models such as VAR- or ARCH-type models.

[30] That the results in the financial markets field might be affected by the estimation methodology selected was mentioned, for example, in Bajzik (2021).

[31] More specifically, three external variables are constructed as simple averages calculated for the same time period as that linked to the given estimate. Hence, an external variable for a primary study which returns ten elasticities

First, we include a set of variables that have been extensively used as factors explaining stock returns dynamics in the existing literature, namely, CPI growth as a measure of inflation and the unemployment rate. As described by modern asset pricing theory (Bekaert and Engstrom, 2010), stock market returns should be strongly correlated with expected inflation, which tends to co-move with economic growth uncertainty and risk aversion, both of which lead to higher returns. Thus, some studies (Subeniotis et al., 2011; Grigaliūnienė and Cibulskienė, 2010) also control for the factor of inflation when estimating the sentiment effect. As the labor market influences the stock market typically via expected equity risk premiums (Atanasov, 2021), the unemployment rate is also often controlled for (Gao et al., 2020; Smales, 2017).

Second, we consider an additional four variables that we hypothesize to have an effect on the sentiment–return relationship. These are the ratio of credit to GDP, the ratio of savings to GDP, the financial openness index (Chinn and Ito, 2008), and the index of financial development (Garriga, 2016). The first variable can be considered a general proxy for the size of the financial sector, while the second might well approximate the size of the wealth available for investment. As documented by Garcia and Liu (1999), both may considerably influence the stock market. Lee et al. (2002), for instance, show that the price for time-varying risk theory (determining investors' decision-making) is likely to be negative when the savings rate is high. The remaining two variables also affect the relationship studied. Schmeling (2009) shows that in more integrated markets, the relationship between sentiment and returns weakens. This is because such markets are more developed as regards financial services and institutions. The information there flows faster, and this, in turn, improves information efficiency, allowing less space for noise.

4.2 Methodology

Our next goal is to find out whether and how the variables mentioned above systematically affect the reported estimates of the effect of survey-based sentiment on stock returns. The simplest way to achieve our objective would be to regress the estimated coefficients on all the proposed variables, including the standard errors. However, given the large number of explanatory variables (48 including the standard errors) such an estimation would be inefficient. Even if we decided to employ just some of the proposed characteristics, the problem would not be solved, because we lack prior knowledge of which variables belong in the baseline model. Besides this model uncertainty issue, the omitted variables problem may arise if we exclude some of the explanatory variables in advance. However, all of these issues have solutions, since each of them can be addressed by Bayesian model averaging (BMA, Steel, 2020).

BMA is aimed at finding the best approximation of the distribution for each regression parameter. This weighted model approach does not exclude any of the explanatory variables in advance. Moreover, it uses all the possible model combinations and compares them directly. Since our 48 explanatory variables give rise to 2^{48} possible model combinations, in our BMA analysis we employ the Markov chain Monte Carlo (MCMC) process with the Metropolis Hastings algorithm to cut the very time-consuming estimation procedure to more manageable proportions. The algorithm searches only the most probable models (Zeugner et al., 2015). Moreover, based on a comparison of the various models, it assigns a posterior model probability (PMP) to each of them. This measure indicates how good each model is in comparison with the others. It also derives the posterior inclusion probability (PIP) based on the weights of the models and the variables included in them. The PIP equals one when the corresponding explanatory variable is included in every

for country X, each estimated on a different length of sample data, would have ten different values, even though it concerns the same region and the same external variable.

model. Conversely, it is zero if the explanatory variable is of no relevance in explaining the sentiment–return relationship. The rest of the variables are sorted between a PIP of 0 and 1 according to their significance in explaining the underlying relationship. Moreover, using the PMPs and PIPs, it is possible to compute the coefficients and standard deviations of all the explanatory variables.

When we discuss the MCMC process, one question naturally arises: what are the most probable models and how can we find them? To answer this question, BMA requires a set of two priors. The first (the g-prior) provides information about the regression coefficients. The second (the model prior) expresses our prior knowledge regarding the distribution of the models. Following Eicher et al. (2011) and George (2010), we employ the unit information g-prior (UIP) and the dilution model prior as our baseline. The use of the UIP g-prior shows that we have no prior knowledge regarding the probabilities of inclusion and the importance of the individual regressors. In this setting, the zero regression coefficient has the same weight as one observation in the data. This choice of dilution model prior has become very popular recently (see, for example, Bajzik et al. 2020). It re-weighs the model probabilities by the determinant of the correlation matrix of variables in each model, giving the highest weights to the model with the lowest multicollinearity.

Having the baseline model specification, we now have to define the reference set of variables that enter the model in order to avoid the dummy variable trap. We dropped the most frequent variables for each dummy variable group (for example, *Europe* for the group of different regions and *Short-run* for the group of different effect horizons). We ended up with a total of 40 regressors including standard errors and external variables (see Tables 4 and 5; the variables in *italics* are those that were dropped). When the results for a specific characteristic are interpreted, the remaining variables in the given group are compared with the dropped reference variable.

4.3 Baseline Results

The results of our baseline procedure (BMA) are graphically represented by Figure 4 and summarized numerically in the left part of Table A1. The reported figure shows the significance of all the possible drivers of heterogeneity in the sentiment–return relationship, with the most significant at the top and the least significant at the bottom (in descending order based on the variables' PIPs). Each column stands for one estimated model. The models are ordered from left to right based on the PMP (measured on the horizontal axis). Red color (lighter in grayscale) for a given row means that the BMA coefficient is negative, that is, the given driver strengthens the negative effect of sentiment on stock returns. Blue color (darker in grayscale) conversely indicates that the driver has a weakening effect on the relationship of interest. If the cell is blank, the given determinant is excluded from the corresponding model. According to Eicher et al. (2011), variables with a PIP above 0.99 are decisive, those with a PIP between 0.95 and 0.99 are strong, those with a PIP in the range of 0.75 to 0.95 are substantial, and those with a PIP of 0.5 to 0.75 are weak.

Figure 4: Model Inclusion in Bayesian Model Averaging – Baseline Specification

Note: The figure shows the outcomes of Bayesian model averaging with the unit information (UIP) g-prior and the dilution model prior to account for potential multicollinearity (Eicher et al., 2011; George, 2010). Each column represents an individual model. The y-axis shows the cumulative posterior model probability (PMP). The variables are sorted in descending order based on posterior inclusion probability (PIP). Only the first 10,000 models are shown. We used 3 million iterations and 1 million burn-ins. Blue color (darker in greyscale): the estimated effect of the corresponding explanatory variable is positive. Red color (lighter in greyscale): the estimated effect of the corresponding explanatory variable is negative. No color: the corresponding explanatory variable is not included in the model.

The results in Table 3 for the overall sample point to a negative true effect of sentiment on stock returns. However, this result relies on the simple Equation 2, which probably suffers from the omitted variable bias problem. Nevertheless, the same result is achieved even after controlling for numerous possible drivers of heterogeneity, as evidenced by the negative coefficient on *intercept* in Table A1. Hence, an increase in survey-based sentiment reduces future stock market returns on average. The results of BMA also reveal strong publication bias in the topic studied. Once additional drivers of heterogeneity are considered in the model, the publication bias turns out to be significant and negative. In other words, researchers tend to report the sentiment effect as being stronger than it actually is. This finding is confirmed by various robustness checks (see Section 4.4).

There are aspects other than publication bias which also affect the elasticities studied. Almost one half of our explanatory variables have a PIP above 0.75 and thus affect the sentiment–return relationship substantially. These variables include excess returns (suggesting that the way the return is measured matters), the sentiment of individual investors (indicating that individual and institutional investors behave differently), the US dummy (indicating different behavior in the US market than in the European market), and the contemporaneous dummy (implying differences in the effect of sentiment at different effect horizons). In the following discussion, we first pay attention to our main data categorization (Section 2.2) and then move on to other factors.

Return characteristics. First, the effect of sentiment is more negative for excess returns than for simple returns. This result is in accordance with our expectations derived in Section 4.1. For the negative mean effect, it means that when a researcher employs excess returns instead of simple returns, she obtains a stronger estimate. Second, the results for the three different types of returns – index returns, portfolio returns, and aggregate market returns – imply a clear and straightforward conclusion: it does not matter which type of return a researcher applies, the mean sentiment effect is not affected by this choice. This outcome at least explains why the existing literature does not offer any hypothesis about possibly different effects across different types of returns, apart from the specific portfolio characteristics outlined in Baker and Wurgler (2006). A similar finding can be found in Bajzik (2021), who studied the relationship between trading volume and stock returns.

Sentiment characteristics. Interesting results are obtained for the categorization of sentiment. We find both the sentiment of institutional investors and the sentiment of individuals have a negative effect on stock returns, but that of individual investors is much stronger. This outcome contradicts the influential theoretical studies of De Long et al. (1990b) and Shleifer and Summers (1990) and the empirical work of Schmeling (2007), which suggest that the two groups of agents affect stock markets in opposite directions. Instead, it echoes the *positive feedback hypothesis* (De Long et al., 1990a), which states that in the short run, institutional investors join the herd of individual investors simply because they are aware of their overreaction and expect stock market gains in the foreseeable future. Besides these results, we find no other sentiment characteristics to be important drivers of heterogeneity. Even the distinction between bullish sentiment and sentiment constructed as the bull-bear spread has no role in the analysis.

Effect horizon. The contemporaneous effect of sentiment is less negative than the short-run (up to one year) effect, while the long-run (more than one year) effect basically does not matter when compared with the short-run category. This finding suggests that if there is a difference in the lagged effect of sentiment on returns, it is driven by time horizons of less than one year. In fact, this result contradicts the overreaction hypothesis (Barberis et al., 1998), which prescribes strictly positive effects of sentiment in the first year and negative effects in the medium to long term (see also Section 2.2). Nevertheless, our results do not deny the presence of overreaction that temporarily pushes prices up (creating positive returns). We just limit the temporal shift to the contemporaneous time frame and show that over a horizon of less than one year, the average effect turns negative, reflecting prices returning to their intrinsic values.

Data characteristics. Another significant determinant of the sentiment–return relationship is the region of the data. Our results suggest that the sentiment effect tends to be more negative in the US than in the European region. This outcome supports the assumption that countries with a higher level of individualism engage more extensively in irrational behavior on the financial markets (Schmeling, 2007). This seems natural, as the share ownership structure in the US differs from that in Europe, in the sense that US households and individuals are much more engaged in the stock market. For example, in 2018, households in Germany and the UK owned only 13.4% and 13.5% of stocks, respectively, while individuals in the US held 61.09% of stocks in the same year.[32] Nevertheless, we admit that the result pointing to stronger effects of sentiment in the US than in the European region might have been affected by the sixth filtering criterion (see Section 2.1), based on which we discarded elasticities representing spillover effects. Baker et al. (2012) found that the sentiments in individual regions are correlated and that US sentiment displays the highest pairwise correlations. Further, they showed that the common component across the regions studied – "global" sentiment – affects local stock returns to a greater extent than region-specific sentiment.[33] Importantly, they argued that this result is not affected by removing the US from the sample. This suggests that stock markets in non-US regions might exhibit a less strong response to improvements in their regional ("local") sentiment simply because regional measures of sentiment do not contain a "global" component. Regarding the *Other regions* category, we identify weaker sentiment effects than in Europe. However, this category contains a mix of unrelated regions and is relatively poorly represented in our data (accounting for only 5% of the observations; see Table 4), so one must be careful in making any strong conclusions.

Next, we find that the sentiment effect tends to be more negative when researchers use higher frequency data compared with monthly data.[34] However, applying data of lower than monthly frequency has no impact in the primary studies. We also note that based on our baseline, the studied effect is influenced neither by the number of observations, nor by orthogonalization or other transformations of sentiment.

Model specification. We observe several interesting findings in this class. First, when researchers estimate their model as a system of equations, the effect of sentiment turns more negative.[35] Second, compared with one-equation models, other model specifications seem indifferent, suggesting that more complex model specifications are not necessary. Third, as for the method of estimation, there is a significant difference in the use of OLS (the reference category) and other estimation techniques (*Other est.*). On the other hand, GMM coefficients seem to be biased in the opposite direction, so one need to be careful to choose appropriate methods that compensate for the original OLS bias. Different results are also obtained using panel fixed effects. Nonetheless, this is difficult to attribute

[32] By stocks, we mean equity and investment fund shares. The data for Germany comes from Deutsche Bundesbank, the data for the UK from the Office for National Statistics, and the data for the US from the Federal Reserve System.

[33] Similar results were found in Corredor et al. (2015) and Han and Li (2017).

[34] Since the studied effect might be diluted over longer time periods in the case of monthly data frequency, one might suggest including a robustness check using data with daily or higher frequency only. We consider such a check in our sample to be not feasible, as it would use only 6% of the data collected, i.e., about 78 observations.

[35] The majority of studies that applied such an estimation framework also worked with average returns as well as bootstrap simulation to adjust biased estimates (Brown and Cliff, 2005; Schmeling, 2009; Rakovská, 2021). Therefore, the significant negative effect may well be due to a mixture of these characteristics.

only to the estimation method itself, as all the elasticities that are estimated with the use of panel fixed effects come from panel datasets – a characteristic which we also control for in the analysis.

Fourth, we can conclude that the sentiment effect is more negative when the model does not contain controls capturing macroeconomic, corporate finance, or general financial market characteristics. This result may be driven by omitted variable bias in the equations of the primary studies not containing these explanatory variables. However, if there is an additional sentiment variable in the model, the sentiment coefficient tends to be less negative. As the inclusion of other sentiment variables is highly heterogeneous (the specification may include lagged sentiment, interaction with a dummy variable, or a different survey-based or non-survey-based sentiment measure), it is difficult to make any general conclusion at this point. Finally, it does not matter whether the sentiment effect is averaged or cumulative or whether the model has a panel or dynamic nature.[36]

Publication characteristics. As regards publication characteristics, the results are slightly conflicting. On the one hand, it does not matter whether or not the article is published in a journal with an impact factor. Next, if the article is published in a journal with an impact factor, a higher impact is associated with less negative sentiment coefficients. On the other hand, the sentiment coefficients are stronger for newer and more cited publications. This inconsistency might be caused by the "Prometheus effect" (Ioannidis, 2008), i.e., the situation where empirical effects decrease over time after an initial novel finding. In order to support their findings, or in effort to delineate themselves from the current literature, newer studies tend to cite articles which in the past found a strong negative effect of sentiment on stock returns. This is where the negative effect of more cited studies arises. In contrast, journals with a high impact factor may have a tendency to publish something unusual. Therefore, they are likely to publish studies with lower effects in order to bring attention to something new. This is where the positive effect of *Impact* might stem from. Last but not least, the fact that the results were published as preferred ones or as part of a robustness check within one article does not matter at all.

External variables. Only one external variable appeared to be a strong determinant of heterogeneity. Higher financial development of the region studied weakens the negative effect produced by our baseline, that is, it moves it closer to zero. This result may seem to conflict with the results for the regions. Specifically, we find that the studied effect is stronger in the US region, which, however, is more financially developed than Europe (the average Garriga, 2016 index is 0.75 for the US and only 0.41 for Europe). A possible solution to this is offered by Schmeling (2009), who showed that market integrity (including the quality of institutions) matters and that countries with high market integrity are less prone to sentiment, simply because such markets are more efficient. Other external variables turn out to be weak or insignificant in our analysis.

4.4 Robustness Checks

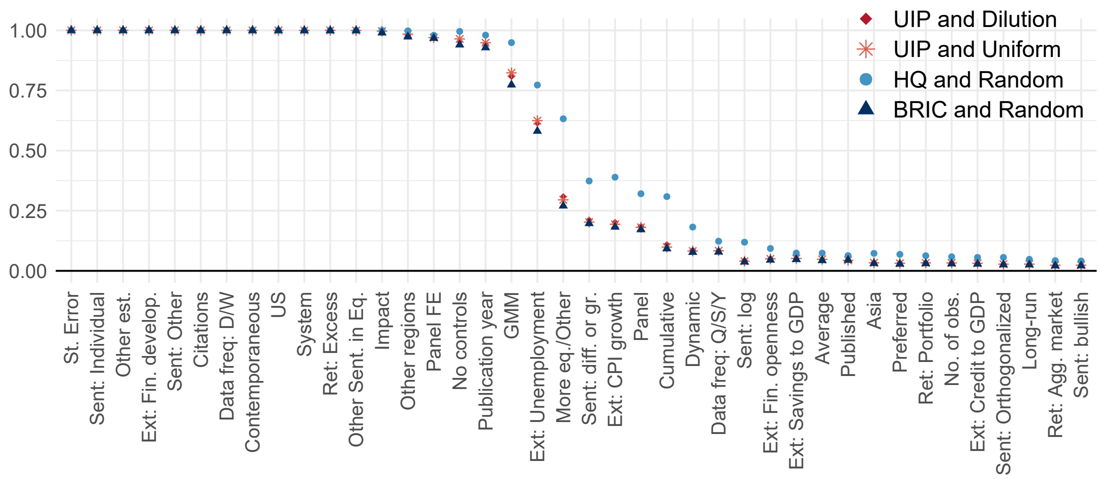
To verify the stability of the results described in the previous section, we run several robustness checks. First of all, we run frequentist model averaging (FMA), as proposed by Hansen (2007); Amini and Parmeter (2012), and a simple frequentist check (OLS) for our baseline specification. In the simple OLS exercise, we only employ regressors with a PIP above 0.5 , that is, regressors with at least a weak effect according to Eicher et al. (2011). The results are listed in the right part of Table A1. Noticeably, our previous conclusion about negative publication bias is fully confirmed.

[36] However, the relationship between averaged returns and the system-of-equations framework and the relationship between the panel data type and the panel fixed effects estimation technique are not negligible.

However, unlike BMA, both OLS and FMA show no significant sentiment effect beyond this bias. In other words, the reported influence of sentiment on stock returns may be driven by systematic selectivity bias. As for the results of heterogeneity drivers, there are some minor differences relative to our baseline, but none of them significantly change the conclusions derived above.

Second, we experiment with different priors for BMA. The additional pairs of priors are the UIP and uniform priors (Eicher et al., 2011), the Hannan-Quinn (HQ) and random priors (Zigraiova et al., 2021), and the BRIC and random priors (Gechert et al., 2020). The HQ g-prior adjusts data quality (Feldkircher and Zeugner, 2012). The BRIC g-prior minimizes the prior effect on the results (Zeugner et al., 2015). The random model prior expresses our lack of prior knowledge about the model distribution by assigning an equal prior probability to every model size (Gechert et al., 2020). We provide the outcome of this robustness check only graphically in Figure 5. In general, we can conclude that our baseline results are not affected by the initial selection of g-prior and model prior.

Figure 5: Sensitivity of Posterior Inclusion Probabilities – Baseline Specification



Note: The figure depicts the posterior inclusion probabilities for various selections of g-priors and model priors. Our baseline employs the unit information g-prior (UIP) and the dilution model prior. UIP: the g-prior assigns to each model the same information as one observation from the data. HQ: the g-prior asymptotically mimics the Hannan-Quinn criterion. BRIC: the g-prior minimizes the effect of the prior on the result. Dilution model prior: the prior weight of each model is proportional to the determinant of the correlation matrix. Uniform model prior: the prior weight of each model is the same. Random model prior: models of same size are assigned the same prior.

Last, we use an alternative set of regressors and run all the BMAs with them. The alternative set of regressors excludes the respondent-driven categorization of sentiment (see Section 2.2) and instead uses an alternative categorization that focuses on types of survey.[37] Specifically, we replace the variables *Sent: Individual*, *Sent: Institutional*, and *Sent: Other* with a new set of variables: (1) *Sent: Cons./Business* – which indicates sentiment measures constructed from consumer- or business-type surveys,[38] (2) *Sent: Investor* – which is based on sentiment from questionnaires directed at stock

[37] Due to high collinearity, those classes could not be simultaneously added to the same model with the previous sentiment categories.

[38] We merged business sentiment with consumer sentiment into one category, because there were only six observations for business-type sentiment and because they are constructed using the same methodology (European Commission, 2018). Consumer-type sentiment can be represented by the University of Michigan Consumer Confidence Index (University of Michigan, 2021) or the DG ECFIN Consumer Confidence Index (European Commission, 2018). Business-type sentiment is covered, for example, by the DG ECFIN Industrial Confidence Index (European Commission, 2020) or the Ifo Business Climate Index (ifo institute, 2021).

market investors,[39] and (3) *Sent: Complex* – which in a certain way combines all three types of sentiment described above.[40]

The intuition behind the alternative categorization of sentiment driven by the type of survey is closely linked to the original, respondent-driven categorization. As described by Ho and Hung (2012), expected returns should theoretically be related not only to the allocation of consumers' expenditure, but also to the potential for investment growth. Thus, complex sentiment measures such as the Economic Sentiment Index (ESI, European Commission, 2020) which contain all this information may have more significant predictive power. However, the empirical results tend to indicate the opposite (Concetto and Ravazzolo, 2019; Ho and Hung, 2012). On the contrary, Gao et al. (2020) concludes that investor-type sentiment is the best possible proxy for market sentiment, as it contains unique information directly from a stock market. This statement can be supported by calculating the mean elasticity for subsamples created based on these characteristics – the effect stemming from investor-type sentiment is, on average, twice as strong (-0.16) as the consumer (-0.07) or complex (-0.08) sentiment proxy.

The results of this robustness exercise using the baseline combination of the g-prior and the model prior are depicted in Figure 6. The results for other combinations of priors can be found in Figure A2. Regarding the types of sentiment, *Consumer/Business* sentiment and *Complex* sentiment yield substantially stronger negative effects on returns than *Investor* sentiment. Furthermore, there are no material changes in the results for the determinants described in the baseline model. Naturally, the PIP changes, as manifested in a decline in the significance of the GMM and a rise in the significance of multi -equation modeling. Nonetheless, as the sign of the effect remains the same for all the main determinants and the significance does not change dramatically, there is no reason to question the general conclusions derived so far.

In a nutshell, after running several robustness checks, we arrive at the following conclusions. The robustness tests reveal only slight deviations from the results achieved in the baseline estimation. As a result, we can conclude that the results derived in Section 4.3 are robust.

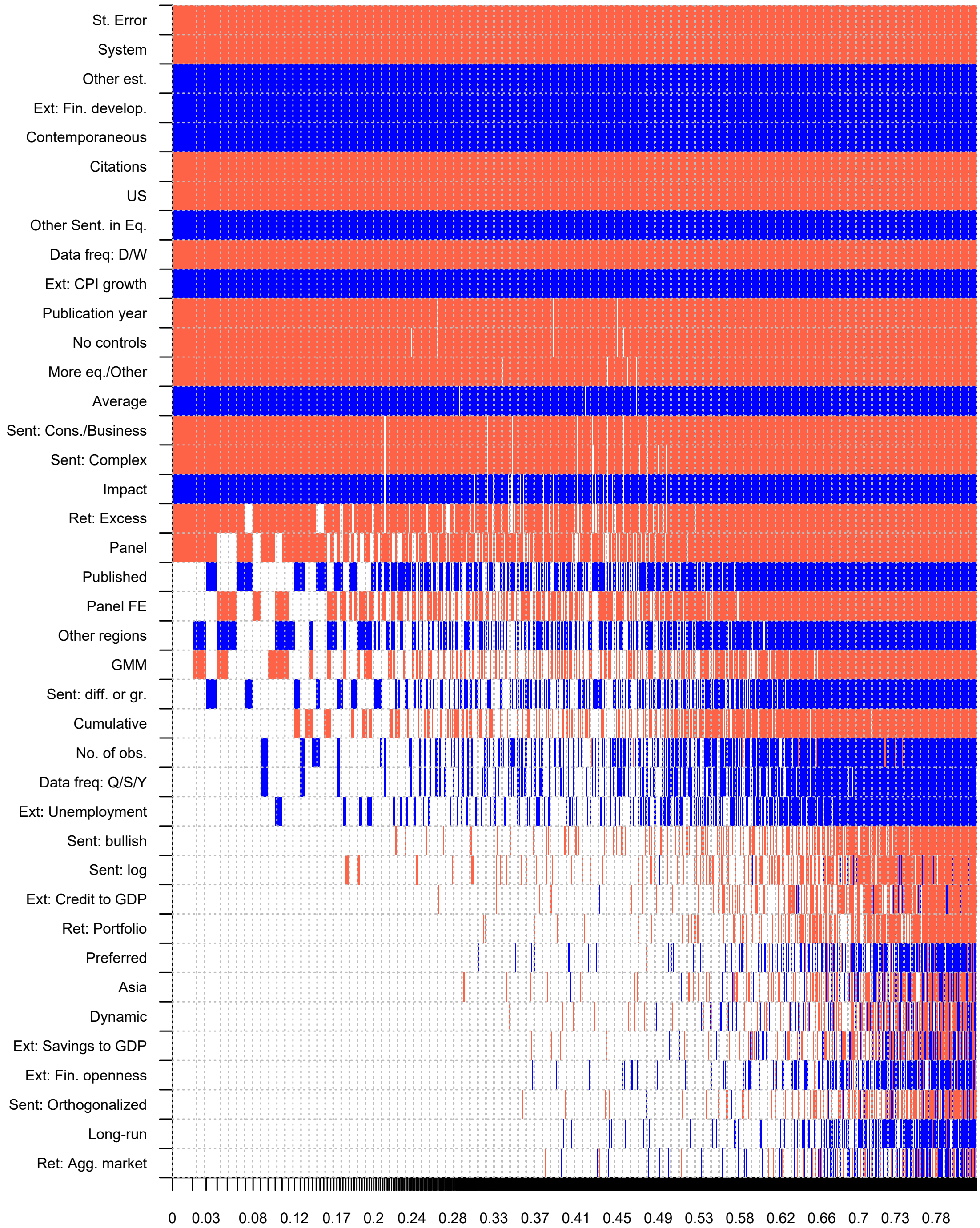
5. Implied Elasticity and Economic Significance of the Results

What do the above results imply for the true effect of sentiment on stock returns? We can formulate the following facts: (1) publication bias distorts the literature negatively in the majority of specifications, but also distorts it positively when one looks only at institutional sentiment and contemporaneous effects (see Table 3), (2) the sentiment of individual investors affects the market negatively and to a greater degree than institutional sentiment, (3) the contemporaneous effect of sentiment on returns is less negative than the effect in future periods of up to one year, and (4) the US region seems to be more prone to sentiment than the European region. Based on this knowledge, combined with the methodological setup of the best and most recent articles, we are able to propose implied estimates across the different categories and thereby aid future model calibration.

[39] Examples of such sentiment measures include sentiment derived from Investor Intelligence (Investors Intelligence, 2020) and SENTIX sentiment (Sentix, 2021).

[40] In the European environment, the most famous is probably the DG ECFIN Economic Sentiment Index (European Commission, 2020). Another example of a *complex* sentiment index is the composite survey-based sentiment constructed in Rakovská (2021).

Figure 6: Model Inclusion in Bayesian Model Averaging – Alternative Sentiment Groups



Note: The figure shows the outcomes of Bayesian model averaging with the unit information (UIP) g-prior and the dilution model prior to account for potential multicollinearity (Eicher et al., 2011; George, 2010). We replaced the original groups of sentiment variables used in our baseline with their alternatives. **Sent: Cons./Business** = 1 if the independent variable (IV) is sentiment derived from consumer or business surveys; **Sent: Investor** = 1 if the IV is sentiment derived from investor surveys (this variable did not enter the analysis of heterogeneity due to the dummy variable trap); and **Sent: Complex** = 1 if the IV is sentiment derived from sources other than consumer, business, or investor surveys or is sentiment constructed as a combination of sentiments from various types of surveys. See also the note to Figure 4.

We identify Stambaugh et al. (2012), Ben-Rephael et al. (2012), and Jiang et al. (2019) as our sample research studies for the analysis of implied elasticity. All three studies were published in the top three journals in the financial field and have a large number of citations and a relatively recent publication year. Based on their study designs, we suggest the use of monthly data frequency, the common OLS approach,[41] and an equation with macroeconomic, corporate finance, or financial market controls on the right-hand side. In addition, we plug in zero as the standard error effect to express no publication bias in our implied estimates. Moreover, we enter sample means for the external variables as *Ext: Fin. develop* and *Ext: Unemployment*. Regarding publication quality, we set the 90th percentile value for the impact factor and the number of citations, with the aim to mimic high-quality, peer-reviewed studies. We prefer newer studies by setting the publication year to its 90th percentile value as well. Finally, we propose the use of simple returns instead of excess returns, even though our sample studies used excess returns in their estimation framework. The newest literature seems to be heading toward more extensive use of excess returns (all the primary studies that employed excess returns were published after 2012). However, the ratio is still dominated by simple returns (even after 2012 the ratio is 9 nine to 10 on behalf one in favor of simple returns).

Table 6 summarizes the implied estimates calculated for several scenarios. The estimates vary between -0.378 and 0.069, meaning that a one standard deviation change in sentiment causes a change in *excess* returns ranging from -0.378 standard deviation in the research design implied by our selection (Ben-Rephael et al., 2012; Stambaugh et al., 2012; Jiang et al., 2019) to 0.069 standard deviation in the design in which we consider a scenario with worse publication settings (where the impact factor and the number of citations are set to the 10th percentile value). All the results differ significantly from zero, except for the implied estimates for institutional investors, the contemporaneous time horizon, other regions, and other-than-OLS estimation. Further, the effects for individual and institutional investors still differ greatly, with the effect for the former being negative and that for the latter being almost negligible. An analogous conclusion can be derived for the effect horizon categories. Interestingly, the analysis returned very similar implied elasticities for the US and Europe, even though we showed previously that the US is more prone to sentiment than Europe. This outcome is due to the fact that we use region-specific values of external variables in these scenarios. Specifically, for each region, we calculated the implied elasticity not only by setting the respective regional dummy to 1, but also by plugging in the real 2018 measures of external variables (*Fin. development* and *Unemployment*) instead of their sample means. Because the values of these external variables do not differ much, the implied elasticities reported in Table 6 do not differ much either. Lastly, we suggest that the overall implied estimate is that a one standard deviation change in sentiment causes a drop in returns of 0.198 standard deviations.

What does this effect mean? As discussed in Section 2.1, we employ the "full standardization" methodology described in Bowman (2012) to produce elasticities that are comparable across studies. The method involves multiplying each raw estimate collected from the primary literature by the corresponding standard deviation for the sentiment measure (the independent variable) and dividing it by the standard deviation of the return series (the dependent variable). On the one hand, the standardization solved the initial problem of having the variables heterogeneous but not too much so. Even though sentiment is a single concept, and even when one considers only survey-based sentiment, there is a wide range of surveys and sentiment measures derived from them, each having

[41] Surprisingly, OLS is the most common technique in the primary studies. Nevertheless, 71% of the elasticities that come from OLS estimation are bias-adjusted in some way (see also Section 4.1). Moreover, the median number of observations in our sample excluding the panel studies is 252, which suggests that the small sample property is not necessarily an issue in this field.

genuine characteristics. On the other hand, full standardization converted the unit of measurement of the elasticities into standard deviations, which is not that easy to understand compared with the effects expressed in levels or percentage points.

Table 6: Implied Estimates

Specification	Estimate	5% CI	95% CI
Implied estimate	-0.198	-0.320	-0.077
Sentiment of different investors			
Sent: Individual investors	-0.276	-0.373	-0.178
Sent: Institutional investors	-0.003	-0.094	0.089
Sent: Other	-0.261	-0.364	-0.158
Different horizons of future returns			
Contemporaneous	-0.039	-0.209	0.132
Short-term (up to one year)	-0.217	-0.337	-0.097
Different regions			
Region: US	-0.198	-0.353	-0.044
Region: Europe	-0.168	-0.294	-0.043
Region: Other	-0.007	-0.279	0.265
Other characteristics			
Excess returns instead of simple returns	-0.378	-0.593	-0.164
GMM instead of OLS	-0.272	-0.413	-0.132
Other estimation instead of OLS	-0.038	-0.158	0.082
Worse publication	0.069	0.003	0.136

Note: The values represent the standard deviation response of returns to a one standard deviation increase in sentiment. They correspond to the mean estimates implied by Bayesian model averaging and conditional on the empirical design implied by the selected trio of articles (Ben-Rephael et al., 2012; Stambaugh et al., 2012; Jiang et al., 2019).

Moreover, as shown in Table 7, the average monthly absolute change in the selected sentiment variables measured in standard deviations (column *Mean/SD* in Panel A) is always less than one, which suggests that the modeled effect of a one standard deviation increase in sentiment is rather overestimated.[42] Instead, sentiment changes on average by only 0.2 SD in the case of sentiment from consumer surveys and by more than 0.5 SD in the case of sentiment from investor surveys. This is an important observation, because it suggests that some sentiment measures are more sensitive than others, i.e., their month-on-month fluctuations are different. This feature might stand as an additional explanation of why we find that the sentiment categories based on survey type are statistically significant drivers of heterogeneity (Section 4.4). The situation is different for the return series, as their average monthly absolute change seems to be well approximated by their one standard deviation.

To improve the overall understanding of our results, we propose a simple recalculation of the first implied estimate in Table 6 into more familiar units, reversing the full standardization described above. We use the first implied estimate as an illustrative example, but the method can be applied to the other implied estimates as well. The last column in Table 7 summarizes the approach. Each row presents the implied estimate multiplied by the ratio of the standard deviation of the return series to the standard deviation of the corresponding sentiment measure. Hence, the result in the

[42] Note that all the statistics reported in Table 7 come from external data (i.e., they were not collected from primary studies) and that the primary study listed for each sentiment–return pair is just a sample study that employs this pair of variables, but the sample of observations or any other characteristics may be different.

first row: -0.07 pp $(= -0.198 \times (4.41/12.5))$, suggests that under the preferred empirical setup (see the description at the beginning of this section), and if one considers the University of Michigan Consumer Confidence Index sentiment and the monthly excess market return from the CRSP value-weighted index taken from the Kenneth R. French data library as the explanatory and dependent variable, respectively (see the first row in Table 7), then a one unit increase in sentiment reduces future monthly returns by 0.07 pp. A similar explanation holds for the other sentiment–return pairs mentioned. Note that this exercise is only an attempt to quantify the effects in more familiar units and does not replace the findings from Table 6.

Table 7: Illustrative Mapping of Implied Estimate

			Abs. Monthly Diff.					**Abs. Monthly Diff.**			
Study	Measure	1SD	Mean	Mean/SD	Freq.≥ 1SD	Measure	1SD	Mean	Mean/SD	Freq.≥ 1SD	Effect size
	Panel A: Sentiment					**Panel B: Returns**					
(1)	**MICH CCI**	12.5	2.96	0.24	0.43	**Exc. Mkt. Ret.**	4.41	4.59	1.04	40.47	**-0.07pp**
(2)	**CB CCI**	25.16	4.5	0.18	0	**S&P500 TR exc.**	4.66	4.66	1	35.97	**-0.04pp**
(3)	**CCI (ECFIN)**	7.62	1.75	0.23	1.23	**Mkt. Ret.**	5.33	5.7	1.07	41.33	**-0.14pp**
(4)	**ESI (ECFIN)**	10.8	2.15	0.2	0	**MSCI PI**	5.88	5.76	0.98	36.22	**-0.11pp**
(5)	**II**	17.76	9.44	0.53	13.92	**S&P500**	4.23	4.43	1.05	40.9	**-0.05pp**
(6)	**AAII**	17.40	14.66	0.84	31.65	**S&P500 exc.**	4.24	4.43	1.04	40.9	**-0.05pp**

Note: MICH CCI – University of Michigan Consumer Confidence Index; CB CCI – Conference Board Consumer Confidence Index; CCI (ECFIN) – Consumer Confidence Index calculated by the Directorate-General for Economic and Financial Affairs (ECFIN); ESI (ECFIN) – Economic Sentiment Index calculated by the DG ECFIN; II – bull-bear spread calculated from Investor Intelligence data (in %); AAII – bull-bear spread calculated from American Association of Individual Investors survey data (in %).
Exc. Mkt. Ret. – the monthly excess market return from the CRSP value-weighted index taken from the Kenneth R. French data library; S&P500 TR exc. – returns of the S&P 500 Total Return Index (including dividends) – risk-free rate (Welch and Goyal, 2008); MSCI PI – MSCI Portugal Price Index; Mkt. Ret. – aggregate market returns taken from the Kenneth R. French data library; S&P500 – S&P 500 Index; S&P500 exc. – returns of the S&P 500 Index – risk-free rate (Welch and Goyal, 2008). All the return series are in %.
Sample studies: (1) – Ben-Rephael et al. (2012); (2) – Jiang et al. (2019); (3) – Schmeling (2009); (4) – Fernandes et al. (2013); (5) – Fisher and Statman (2000); (6) – Zhou (2018). The first column gives an example of a primary study which employed the given combination of sentiment and return variables, but the table itself does not use any information from these studies. By contrast, the data used for the calculations was downloaded from external sources such as Datastream, the DG ECFIN website, and the website of Professor Kenneth R. French.
The table quantifies the standard deviation for the selected sentiment variables (Panel A) and return series (Panel B) and shows how it relates to the absolute value of their month-on-month changes. The last column – *Effect size* – provides an illustrative mapping of the standardized implied estimate from the first row of Table 6 to these series. It represents the recalculation of the "fully standardized" implied estimate measured in standard deviations to the more familiar effect in percentage points corresponding to a one unit increase in the selected sentiment series. *Mean* stands for the average absolute month-on-month change of the given variable, while *Mean/SD* provides information on how many standard deviations the variable changes by on average on a month-on-month basis. *Freq.* $\geq$ *1SD* shows the frequency (in %) of the times when the absolute month-on-month difference of the given series is equal to or above the standard deviation. We calculated the SDs for the majority of the variables using the sample period from January 1980 to December 2018. Data on some of the variables were not available starting from January 1980. In all of these cases, we used the maximum sample size. Specifically, the sample for CCI (ECFIN) starts in January 1985, the sample for ESI (ECFIN) starts in January 1987, the sample for AAII starts in August 1987, and lastly the sample for MSCI PI starts in February 1987.

6. Conclusions

We conducted the first meta-analysis of the effect of survey-based sentiment on stock returns. We collected 1311 estimates from 30 studies, together with almost 50 possible explanatory variables that can clarify our relationship of interest. Our analysis provides a deeper insight into the structure of the financial markets, and our results might enhance investment model calibrations and also serve as useful information for policymakers concerned with the causes of systemic events such as asset bubbles.

We proved that publication bias is present in several specifications. In most of them, especially individual sentiment and the long-term effect horizon, the bias appeared with a negative sign, exaggerating the true effect. In two related specifications, namely, the sentiment of institutional investors and the contemporaneous effect horizon, however, the selection bias turned positive. This suggests that, conditional on these characteristics, researchers tend to report less negative (or even positive) estimates, probably in a quest to conform with the theoretical (but also empirical) literature. Altogether, we find that the true effect of sentiment on (contemporaneous and future) stock returns after correction for publication bias is negative, although in some specifications it is not significant.

The analysis of the drivers of heterogeneity in our relationship of interest confirmed the previous result, namely, that publication bias is a key driver of the magnitude of the sentiment effect. Further, we revealed several factors that shape this relationship. First, there is a significant difference in the sentiment effect between institutions and individual investors, with the latter being more negative. Therefore, one must be careful about what types of sentiment are included in the analysis, as their predictive power for the stock market are not the same. Second, the immediate effect of sentiment was shown to be less negative than the longer-lasting effect of up to one year. However, we did not find any significant difference between the short-term (up to one year) and longer-term (more than one year) effect horizon. Third, the type of return series does not matter, but we cannot reject the possibility of further breakdown of portfolio returns generating different results. Because the portfolio category contains many very different classes (small/large, value/growth stocks, etc.), this area could be investigated in more detail as a new potential driver of sentiment effect heterogeneity. Fourth, sentiment effects turned out to be stronger in the US than in Europe. On the one hand, this might be explained by cultural differences, mainly by individualism (Schmeling, 2007). On the other, it may simply reflect the fact that we did not consider spillover effects in our analysis and, as a result, future returns in regions other than the US might exhibit lower sensitivity to regional sentiment, which, by definition, does not contain a "global" component (Baker et al., 2012). Lastly, we found, for instance, that the use of higher frequency data pushes the sentiment coefficients into more negative territory and that the estimation results are also influenced by the choice of model and estimation methodology. In addition, we showed that among the external variables employed, the degree of financial development of the country studied weakens the estimated elasticities. All our main findings were successfully challenged across various robustness exercises.

We closed our analysis by offering scenarios for the implied estimates, motivated by the knowledge gained and the common setup in the influential literature. Our results show that a one standard deviation increase in sentiment reduces future monthly returns by 0.198 standard deviations. Throughout the paper, we worked with fully standardized elasticities expressing the sentiment effect in standard deviations. To improve the readability of our results, we also proposed an approach for recalculating the estimated elasticities into more understandable units. The outcomes of these exercises could be taken into account by market professionals in developing trading strategies, and also by policymakers in enhancing stress-testing scenarios, assessing business/financial cycles, and examining the causes of financial crises (Næs et al., 2011; Angelidis et al., 2015).

Our meta-analysis and all the implications derived in it are limited to the use of survey-based sentiment. In order to keep the analysis and its conclusions straightforward, we did not consider other types of sentiment measures. We acknowledge that their potential publication bias and hence their true predictive power – especially when related to other determinants (heterogeneity drivers) – may be very different. Meta-analysis of non-survey-based types of sentiment thus remains a challenge for further research.

References

AHMED, B. (2020): "Understanding the Impact of Investor Sentiment on the Price Formation Process: A Review of the Conduct of American Stock Markets." *The Journal of Economic Asymmetries,* 22:e00172.

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.

AKHTAR, S., R. FAFF, B. OLIVER, AND A. SUBRAHMANYAM (2011): "The Power of Bad: The Negativity Bias in Australian Consumer Sentiment Announcements on Stock Returns." *Journal of Banking & Finance,* 35(5):1239–1249.

AMINI, S. M. AND C. F. PARMETER (2012): "Comparison of Model Averaging Techniques: Assessing Growth Determinants." *Journal of Applied Econometrics,* 27(5):870–876.

ANGELIDIS, T., A. SAKKAS, AND N. TESSAROMATIS (2015): "Stock Market Dispersion, the Business Cycle and Expected Factor Returns." *Journal of Banking & Finance,* 59:265–279.

ASTAKHOV, A., T. HAVRANEK, AND J. NOVAK (2019): "Firm Size And Stock Returns: A Quantitative Survey." *Journal of Economic Surveys,* 33(5):1463–1492.

ATANASOV, V. (2021): "Unemployment and Aggregate Stock Returns." *Journal of Banking & Finance,* 129:106159.

BAJZIK, J. (2021): "Trading Volume and Stock Returns: A Meta-analysis." *International Review of Financial Analysis,* 78:101923.

BAJZIK, J., T. HAVRANEK, Z. IRSOVA, AND J. SCHWARZ (2020): "Estimating the Armington Elasticity: The Importance of Study Design and Publication Bias." *Journal of International Economics,* 127:103383.

BAKER, M. AND J. C. STEIN (2004): "Market Liquidity as a Sentiment Indicator." *Journal of Financial Markets,* 7(3):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):129–152.

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 Finacial Economics,* 49:307–343.

BATHIA, D. AND D. BREDIN (2013): "An Examination of Investor Sentiment Effect on G7 Stock Market Returns." *The European Journal of Finance,* 19(9):909–937.

BEHRENDT, S., F. J. PETER, AND D. J. ZIMMERMANN (2020): "An Encyclopedia for Stock Markets? Wikipedia Searches and Stock Returns." *International Review of Financial Analysis,* 72:101563.

BEKAERT, G. AND E. ENGSTROM (2010): "Inflation and the Stock Market: Understanding the "Fed Model"." *Journal of Monetary Economics,* 57(3):278–294.

BEN-REPHAEL, A., S. KANDEL, AND A. WOHL (2012): "Measuring Investor Sentiment with Mutual Fund Flows." *Journal of Financial Economics,* 104(2):363–382.

BLANCO-PEREZ, C. AND A. BRODEUR (2020): "Publication Bias and Editorial Statement on Negative Findings." *The Economic Journal,* 130(629):1226–1247.

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: 11.9.2021

BOLLEN, J., H. MAO, AND X. ZENG (2011): "Twitter Mood Predicts the Stock Market." *Journal of Computational Science,* 2(1):1–8.

BOM, P. R. AND H. RACHINGER (2019): "A Kinked Meta-regression Model for Publication Bias Correction." *Research Synthesis Methods,* 10(4):497–514.

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

BOWMAN, N. A. (2012): "Effect Sizes and Statistical Methods for Meta-analysis in Higher Education." *Research in Higher Education,* 53(3):375–382.

BREMMER, D. (2008): "Consumer Confidence and Stock Prices." In *72nd Annual Meeting of the Midwest Economics Association Hyatt Regency, Chicago, Illinois.* Citeseer.

BROWN, A. L., T. IMAI, F. VIEIDER, AND C. CAMERER (2021): "Meta-analysis of Empirical Estimates of Loss-aversion." CESifo Working Paper 8848.

BROWN, G. W. AND M. T. CLIFF (2005): "Investor Sentiment and Asset Valuation." *The Journal of Business,* 78(2):405–440.

CAMPOS, N. F., J. FIDRMUC, AND I. KORHONEN (2019): "Business Cycle Synchronisation and Currency Unions: A Review of the Econometric Evidence Using Meta-analysis." *International Review of Financial Analysis,* 61:274–283.

CAZACHEVICI, A., T. HAVRANEK, AND R. HORVATH (2020): "Remittances and Economic Growth: A Meta-analysis." *World Development,* 134:105021.

CHAN, S. Y. AND W.-M. FONG (2004): "Individual Investors' Sentiment and Temporary Stock Price Pressure." *Journal of Business Finance & Accounting,* 31(5-6):823–836.

CHAROENROOK, A. (2003): "Does Sentiment Matter." Unpublished Working Paper, Vanderbilt University.

CHEEMA, M. A., Y. MAN, AND K. R. SZULCZYK (2020): "Does Investor Sentiment Predict the Near-term Returns of the Chinese Stock Market?" *International Review of Finance,* 20(1): 225–233.

CHEN, N.-F., R. KAN, AND M. H. MILLER (1993): "Are the Discounts on Closed-end Funds a Sentiment Index?" *The Journal of Finance,* 48(2):795–800.

CHEN, N.-F., R. KAN, AND M. H. MILLER (1993): "A Rejoinder." *The Journal of Finance,* 48 (2):809–810.

CHINN, M. D. AND H. ITO (2008): "A New Measure of Financial Openness." *Journal of Comparative Policy Analysis,* 10(3):309–322.

CHOPRA, N., C. M. LEE, A. SHLEIFER, AND R. H. THALER (1993): "Summing up." *The Journal of Finance,* 48(2):811–812.

CHOPRA, N., C. M. LEE, A. SHLEIFER, AND R. H. THALER (1993): "Yes, Discounts on Closed-end Funds Are a Sentiment Index." *Journal of Finance,* 48(2):801–808.

CHUI, A. C., S. TITMAN, AND K. J. WEI (2010): "Individualism and Momentum around the World." *The Journal of Finance,* 65(1):361–392.

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." *Journal of Risk and Financial Management,* 12(2):85.

CORREDOR, P., E. FERRER, AND R. SANTAMARIA (2015): "The Impact of Investor Sentiment on Stock Returns in Emerging Markets: The Case of Central European Markets." *Eastern European Economics,* 53(4):328–355.

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.

DOUCOULIAGOS, H. AND M. PALDAM (2011): "The Ineffectiveness of Development Aid on Growth: An Update." *European Journal of Political Economy,* 27(2):399–404.

DOW, S. C. (2011): "Cognition, Market Sentiment and Financial Instability." *Cambridge Journal of Economics,* 35(2):233–249.

DUAN, J., K. K. DAS, L. MERILUOTO, AND W. R. REED (2020): "Estimating the Effect of Spillovers on Exports: A Meta-analysis." *Review of World Economics,* 156(2):219–249.

EGGER, M., G. D. SMITH, M. SCHNEIDER, AND C. MINDER (1997): "Bias in Meta-analysis Detected by a Simple, Graphical Test." *BMJ,* 315(7109):629–634.

EHRENBERGEROVA, D., J. BAJZIK, AND T. HAVRANEK (2021): "When Does Monetary Policy Sway House Prices? A Meta-analysis." CEPR Discussion Paper DP16196, CEPR.

EICHER, T. S., C. PAPAGEORGIOU, AND A. E. RAFTERY (2011): "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics,* 26(1):30–55.

EUROPEAN COMMISSION (2020): "The Joint Harmonised EU Programme of Business and Consumer Surveys." (https://ec.europa.eu/info/sites/default/files/bcs_user_guide.pdf).

EUROPEAN COMMISSION (2018): "A Revised Consumer Confidence Indicator." (https://ec.europa.eu/info/sites/info/files/new_cci.pdf).

FELDKIRCHER, M. AND S. ZEUGNER (2012): "The Impact of Data Revisions on the Robustness of Growth Determinants–A Note on 'Determinants of Economic Growth: Will Data Tell?'." *Journal of Applied Econometrics,* 27(4):686–694.

FERNANDES, C. M. D. A., P. M. M. G. GONÇALVES, AND E. F. S. VIEIRA (2013): "Does Sentiment Matter for Stock Market Returns? Evidence from a Small European Market." *Journal of Behavioral Finance,* 14(4):253–267.

FISHER, K. L. AND M. STATMAN (2000): "Investor Sentiment and Stock Returns." *Financial Analysts Journal,* 56(2):16–23.

FISHER, K. L. AND M. STATMAN (2003): "Consumer Confidence and Stock Returns." *The Journal of Portfolio Management,* 30(1):115–127.

FRANCO, A., N. MALHOTRA, AND G. SIMONOVITS (2014): "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science,* 345(6203):1502–1505.

FURUKAWA, C. (2019): "Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method." *Evidence, and a New Correction Method (March 29, 2019).*

GAO, X., C. GU, AND K. KOEDIJK (2020): "Institutional Investor Sentiment and Aggregate Stock Returns." *European Financial Management.*

GARCIA, V. F. AND L. LIU (1999): "Macroeconomic Determinants of Stock Market Development." *Journal of Applied Economics,* 2(1):29–59.

GARRIGA, A. C. (2016): "Central Bank Independence in the World: A New Data Set." *International Interactions,* 42(5):849–868.

GECHERT, S., T. HAVRANEK, Z. IRSOVA, AND D. KOLCUNOVA (2020): "Measuring Capital-Labor Substitution: The Importance of Method Choices and Publication Bias." Working Paper, CNB.

GEORGE, E. I. (2010): *Dilution Priors: Compensating for Model Space Redundancy.* In *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown,* pages 158–165. Institute of Mathematical Statistics.

GEYER-KLINGEBERG, J., M. HANG, M. WALTER, AND A. RATHGEBER (2018): "Do Stock Markets React to Soccer Games? A Meta-regression Analysis." *Applied Economics,* 50(19): 2171–2189.

GREENWOOD, R. AND A. SHLEIFER (2014): "Expectations of Returns and Expected Returns." *The Review of Financial Studies,* 27(3):714–746.

GRIGALIŪNIENĖ, Ž. AND D. CIBULSKIENĖ (2010): "Investor Sentiment Effect on Stock Returns in Scandinavian Stock Market." *Economics and Management,* 15:929–940.

GÜNEŞ, H. AND S. ÇELIK (2009): "Consumer Confidence and Financial Market Variables in an Emerging Economy: The Case of Turkey." In *Society for Computational Economics 15th International Conference on Computing in Economics and Finance University of Technology, Sydney.* Citeseer.

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, B. E. (2007): "Least Squares Model Averaging." *Econometrica,* 75(4):1175–1189.

HAVRANEK, T. AND O. KOKES (2015): "Income Elasticity of Gasoline Demand: A Meta-analysis." *Energy Economics,* 47:77–86.

HAVRANEK, T. AND A. SOKOLOVA (2020): "Do Consumers Really Follow a Rule of Thumb? Three Thousand Estimates from 144 Studies Say "Probably not"." *Review of Economic Dynamics,* 35:97–122.

HAVRANEK, T., M. RUSNAK, AND A. SOKOLOVA (2017): "Habit Formation in Consumption: A Meta-analysis." *European Economic Review,* 95:142–167.

HAVRÁNEK, T., T. STANLEY, H. DOUCOULIAGOS, P. BOM, J. GEYER-KLINGEBERG, I. IWASAKI, W. R. REED, K. ROST, AND R. VAN AERT (2020): "Reporting Guidelines for Meta-analysis in Economics." *Journal of Economic Surveys,* 34(3):469–475.

HENGELBROCK, J., E. THEISSEN, AND C. WESTHEIDE (2013): "Market Response to Investor Sentiment." *Journal of Business Finance & Accounting,* 40(7-8):901–917.

HO, J. C. AND C.-H. HUNG (2012): "Predicting Stock Market Returns and Volatility with Investor Sentiment: Evidence from Eight Developed Countries." *Journal of Accounting and Finance,* 12(4):49.

HONG, H. AND J. C. STEIN (1999): "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets." *The Journal of Finance,* 54(6):2143–2184.

IDEAS/REPEC (2021): "IDEAS/RePEc Recursive Discounted Impact Factors for Journals." (https://ideas.repec.org/top/top.journals.rdiscount.html).

IFO INSTITUTE (2021): "Ifo Business Climate Index for Germany." (https://www.ifo.de/en/survey/ifo-business-climate-index).

IMAI, T., T. A. RUTTER, AND C. F. CAMERER (2021): "Meta-analysis of Present-bias Estimation using Convex Time Budgets." *The Economic Journal,* 131(636):1788–1814.

INVESTORS INTELLIGENCE (2020): "US Advisors' Sentiment Report - Signals when You Need Them – Near Important Market Tops and Bottoms." (https://www.investorsintelligence.com/x/us_advisors_sentiment.html).

IOANNIDIS, J. P. (2008): "Why Most Discovered True Associations Are Inflated." *Epidemiology,* 640–648.

IOANNIDIS, J. P., T. D. STANLEY, AND H. DOUCOULIAGOS (2017): "The Power of Bias in Economics Research." *The Economic Journal,* 127(605):F236–F265.

JIANG, F., J. LEE, X. MARTIN, AND G. ZHOU (2019): "Manager Sentiment and Stock Returns." *Journal of Financial Economics,* 132(1):126–149.

KALE, S. AND M. AKKAYA (2016): "The Relation between Confidence Climate and Stock Returns: The Case of Turkey." *Procedia Economics and Finance,* 38:150–162.

KIM, J., H. DOUCOULIAGOS, AND T. STANLEY (2019): *Market Efficiency in Asian and Australasian Stock Markets: A Fresh Look at the Evidence.* In *International Financial markets,* pages 382–419. Routledge.

KRISTOUFEK, L. (2013): "BitCoin Meets Google Trends and Wikipedia: Quantifying the Relationship between Phenomena of the Internet Era." *Scientific Reports,* 3(1):1–7.

LEE, C. M., A. SHLEIFER, AND R. H. THALER (1991): "Investor Sentiment and the Closed-end Fund Puzzle." *The Journal of Finance,* 46(1):75–109.

LEE, W. Y., C. X. JIANG, AND D. C. INDRO (2002): "Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment." *Journal of Banking & Finance,* 26(12):2277–2299.

LEMMON, M. AND E. PORTNIAGUINA (2006): "Consumer Confidence and Asset Prices: Some Empirical Evidence." *Review of Financial Studies,* 19(4):1499–1529.

LISTON, D. P. (2016): "Sin Stock Returns and Investor Sentiment." *The Quarterly Review of Economics and Finance,* 59:63–70.

MATOUSEK, J., T. HAVRANEK, AND Z. IRSOVA (2021): "Individual Discount Rates: A Meta-analysis of Experimental Evidence." *Experimental Economics,* 1–41.

NÆS, R., J. A. SKJELTORP, AND B. A. ØDEGAARD (2011): "Stock Market Liquidity and the Business Cycle." *The Journal of Finance,* 66(1):139–176.

NEWEY, W. K. AND K. D. WEST (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica,* 55(3):703–708.

NY TIMES (2021): "'Dumb Money' Is on GameStop, and It's Beating Wall Street at Its Own Game." Available at (https://www.nytimes.com/2021/01/27/business/gamestop-wall-street-bets.html). Accessed: 11.9.2021

OTOO, M. W. (1999): "Consumer Sentiment and the Stock Market." (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=205028).

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_suppl):S185–S212.

RAFTERY, A. E., D. MADIGAN, AND J. A. HOETING (1997): "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association,* 92(437): 179–191.

RAKOVSKÁ, Z. (2021): "Composite Survey Sentiment as a Predictor of Future Market Returns: Evidence for German Equity Indices." *International Review of Economics & Finance,* 73: 473–495.

RASHID, M., M. K. HASSAN, AND N. Y. YEIN (2014): "Macroeconomics, Investor Sentiment, and Islamic Stock Price Index in Malaysia." *Journal of Economic Cooperation and Development,* 35(4):219–234.

RODRIGUEZ, E., M. AGUILAR-CORNEJO, R. FEMAT, AND J. ALVAREZ-RAMIREZ (2014): "US Stock Market Efficiency Over Weekly, Monthly, Quarterly and Yearly Time Scales." *Physica A: Statistical Mechanics and Its Applications,* 413:554–564.

SAYIM, M. AND H. RAHMAN (2015): "An Examination of US Institutional and Individual Investor Sentiment Effect on the Turkish Stock Market." *Global Finance Journal,* 26:1–17.

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.

SENTIX (2021): "Sentix Sentiment Index." (https://www.sentix.de/index.php/en/item/sntm.html).

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.

SMALES, L. A. (2017): "The Importance of Fear: Investor Sentiment and Stock Market Returns." *Applied Economics,* 49(34):3395–3421.

STAMBAUGH, R. F. (1999): "Predictive Regressions." *Journal of Financial Economics,* 54(3): 375–421.

STAMBAUGH, R. F., J. YU, AND Y. YUAN (2012): "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics,* 104(2):288–302.

STANLEY, T. D. AND H. DOUCOULIAGOS (2012): *Meta-regression Analysis in Economics and Business.* Routledge, New York, USA.

STANLEY, T. D., S. B. JARRELL, AND H. DOUCOULIAGOS (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox." *The American Statistician,* 64(1):70–77.

STEEL, M. F. (2020): "Model Averaging and Its Use in Economics." *Journal of Economic Literature,* 58(3):644–719.

SUBENIOTIS, D. N., D. L. PAPADOPOULOS, I. A. TAMPAKOUDIS, AND A. TAMPAKOUDI (2011): "How Inflation, Market Capitalization, Industrial Production and the Economic Sentiment Indicator Affect the EU-12 Stock Markets." *European Research Studies Journal,* 14(1):103–118.

SUM, V. ET AL. (2014): "Effects of Business and Consumer Confidence on Stock Market Returns: Cross-sectional Evidence." *Economics, Management, and Financial Markets,* 9(1):21–25.

TETLOCK, P. C. (2007): "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance,* 62(3):1139–1168.

UGUR, M., S. AWAWORYI CHURCHILL, AND E. SOLOMON (2018): "Technological Innovation and Employment in Derived Labour Demand Models: A Hierarchical Meta-regression Analysis." *Journal of Economic Surveys,* 32(1):50–82.

UNIVERSITY OF MICHIGAN (2021): "Surveys of Consumers: University of Michigan." (https://data.sca.isr.umich.edu/).

WELCH, I. AND A. GOYAL (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *The Review of Financial Studies,* 21(4):1455–1508.

ZEUGNER, S., M. FELDKIRCHER, ET AL. (2015): "Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R." *Journal of Statistical Software,* 68(4):1–37.

ZHOU, G. (2018): "Measuring Investor Sentiment." *Annual Review of Financial Economics,* 10: 239–259.

ZIGRAIOVA, D. AND T. HAVRANEK (2016): "Bank Competition and Financial Stability: Much Ado about Nothing?" *Journal of Economic Surveys,* 30(5):944–981.

ZIGRAIOVA, D., T. HAVRANEK, Z. IRSOVA, AND J. NOVAK (2021): "How Puzzling Is the Forward Premium Puzzle? A Meta-analysis." *European Economic Review,* 134:103714.

Appendix A

Figure A1: Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Flow Diagram

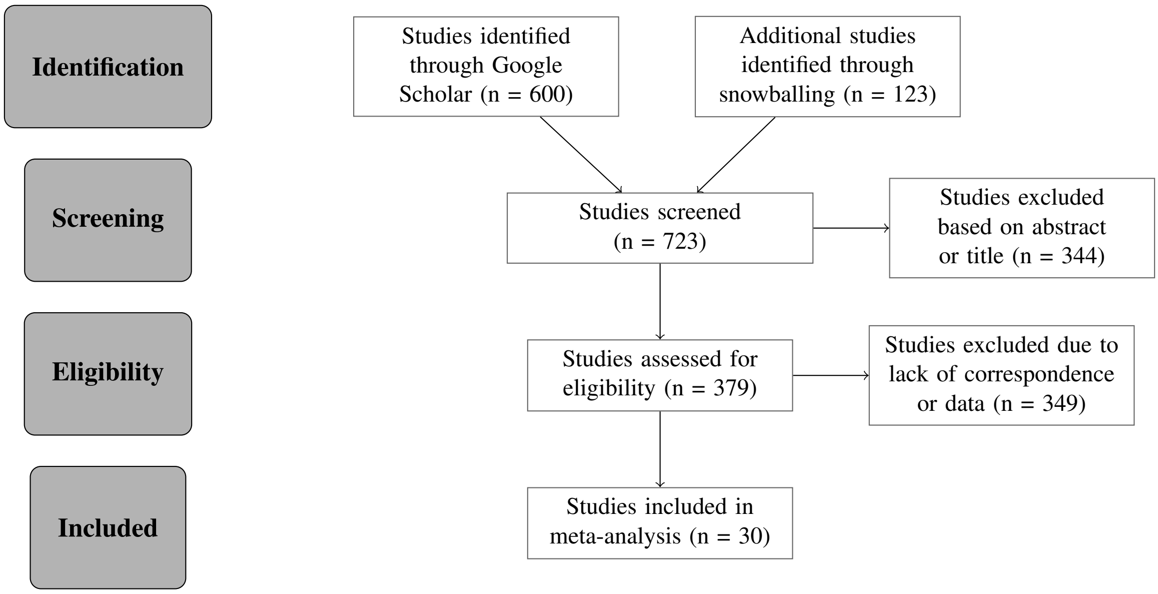
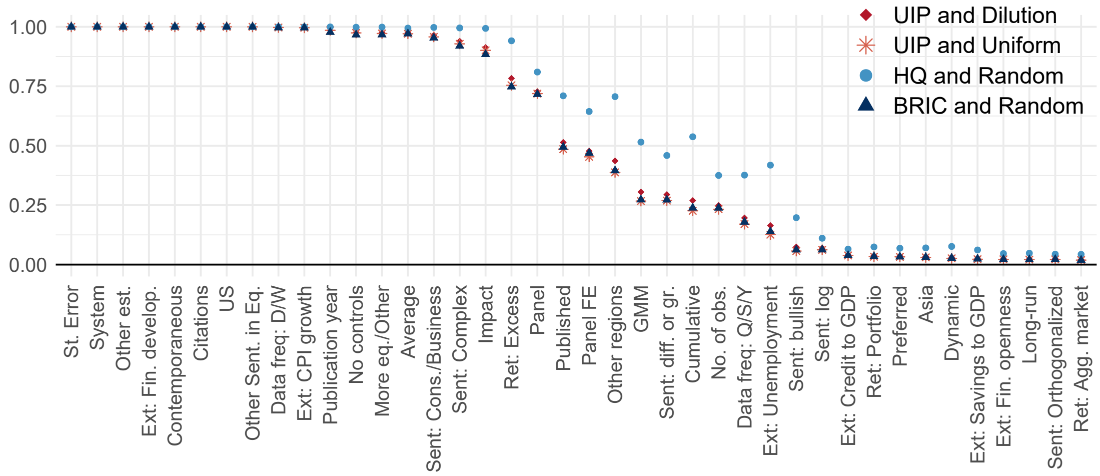


Table A1: What Drives the Heterogeneity of the Estimates Collected – Baseline Specification

	Bayesian Model Averaging			**Frequentist Check (OLS)**			**Frequentist Model Averaging**		
	PIP	P. mean	P. SD	Coef.	SE	p-value	Coef.	SE	p-value
Intercept	1	-0.29		-0.12	0.19	0.55	-0.32	0.28	0.26
St. Error	1	-0.21	0.03	-0.25	0.14	0.08	-0.22	0.04	0
Data characteristics									
Ret: Excess	1	-0.19	0.04	-0.19	0.05	0	-0.22	0.04	0
Ret: Agg. market	0.02	0	0				0.01	0.04	0.88
Ret: Portfolio	0.03	0	0.01				-0.02	0.02	0.40
Sent: Individual	1	-0.50	0.04	-0.29	0.05	0	-0.31	0.03	0
Sent: Other	1	-0.44	0.05	-0.29	0.07	0	-0.30	0.04	0
Sent: Bullish	0.02	0	0.01				0	0.08	0.98
Sent: Orthogonalized	0.03	0	0.01				-0.04	0.03	0.27
Sent: Log	0.04	0	0.01				-0.11	0.07	0.10
Sent: Diff. or gr.	0.21	0.01	0.03				0.06	0.03	0.04
US	1	-0.27	0.06	-0.15	0.10	0.14	-0.24	0.04	0
Asia	0.03	0	0.01				0.08	0.05	0.16
Other regions	0.98	0.11	0.03	0.21	0.14	0.15	0.17	0.05	0
No. of obs.	0.03	0	0.01				0.04	0.03	0.19
Model specification and estimation									
Cumulative	0.11	-0.01	0.03				-0.14	0.05	0.01
Average	0.05	0	0.01				-0.01	0.03	0.82
Contemporaneous	1	0.12	0.02	0.17	0.04	0	0.20	0.04	0
Long-run	0.03	0	0.01				0	0.02	0.92
Other Sent. in Eq.	1	0.12	0.03	0.11	0.06	0.08	0.09	0.02	0
No controls	0.95	-0.10	0.04	-0.09	0.05	0.06	-0.11	0.03	0
Panel	0.19	-0.01	0.03				-0.14	0.06	0.02
Data freq: D/W	1	-0.20	0.03	-0.26	0.07	0	-0.34	0.06	0
Data freq: Q/S/Y	0.08	0	0.02				0.11	0.08	0.17
More eq./Other	0.31	-0.03	0.05				-0.22	0.06	0
GMM	0.81	-0.08	0.05	-0.12	0.07	0.12	-0.10	0.05	0.03
Panel FE	0.97	-0.15	0.04	-0.32	0.08	0	-0.27	0.09	0
Other est.	1	0.23	0.04	0.21	0.04	0	0.33	0.05	0
Dynamic	0.09	0.01	0.03				0.11	0.06	0.06
System	1	-0.24	0.04	-0.16	0.04	0	-0.16	0.04	0
Publication characteristics									
Published	0.04	0	0.02				0.13	0.10	0.16
Impact	0.99	0.15	0.04	0.11	0.03	0	0.17	0.03	0
Citations	1	-0.42	0.05	-0.06	0.01	0	-0.07	0.01	0
Publication year	0.94	-0.15	0.06	-0.12	0.06	0.07	-0.13	0.04	0
Preferred	0.03	0	0.01				0	0.02	0.99
External variables									
Ext: CPI growth	0.20	0.02	0.04				0.02	0.01	0.13
Ext: Unemployment	0.61	0.05	0.04	0.02	0.01	0.24	0.01	0.01	0.05
Ext: Credit to GDP	0.03	0	0.01				0	0	0.42
Ext: Savings to GDP	0.05	0	0.01				0	0	0.22
Ext: Fin. develop.	1	0.34	0.05	1.15	0.53	0.04	1.24	0.17	0
Ext: Fin. openness	0.05	0	0.01				-0.01	0.03	0.68
Adj.R^2 (OLS only)					0.446				

Note: The table reports the estimation results for Bayesian model averaging (BMA), the frequentist check (OLS), and frequentist model averaging (FMA). PIP: posterior inclusion probability; P.mean: posterior mean; P.SD: posterior standard deviation. BMA employs the unit information (UIP) g-prior and the dilution model prior to account for potential multicollinearity (Eicher et al., 2011; George, 2010). OLS only includes those variables which received a PIP higher than 0.5 and uses study-level clustered standard errors. FMA employs Mallow's weights (Hansen, 2007) using orthogonalization of the covariate space as suggested by Amini and Parmeter (2012) to reduce the number of models estimated. A description of all the variables can be found in Tables 4 and 5.

Figure A2: Sensitivity of Posterior Inclusion Probabilities – Alternative Sentiment Groups



Note: See the note to Figure 5.

CNB Working Paper Series (since 2020)

WP 10/2021	Zuzana Gric Josef Bajzík Ondřej Badura	*Does sentiment affect stock returns? A meta-analysis across survey-based measures*
WP 9/2021	Jan Janků Ondřej Badura	*Non-linear effects of market concentration on the underwriting profitability of the non-life insurance sector in Europe*
WP 8/2021	Simona Malovaná Martin Hodula Josef Bajzík Zuzana Gric	*A tale of different capital ratios: How to correctly assess the impact of capital regulation on lending*
WP 7/2021	Volha Audzei	*Learning and cross-country correlations in a multi-country DSGE model*
WP 6/2021	Jin Cao Valeriya Dinger Tomás Gómez Zuzana Gric Martin Hodula Alejandro Jara Ragnar Juelsrud Karolis Liaudinskas Simona Malovaná Yaz Terajima	*Monetary policy spillover to small open economies: Is the transmission different under low interest rates?*
WP 5/2021	Martin Hodula Ngoc Anh Ngo	*Does macroprudential policy leak? Evidence from non-bank credit intermediation in EU countries*
WP 4/2021	Tomáš Adam Ondřej Michálek Aleš Michl Eva Slezáková	*The Rushin index: A weekly indicator of Czech economic activity*
WP 3/2021	Michal Franta Jan Libich	*Holding the economy by the tail: Analysis of short- and long-run macroeconomic risks*
WP 2/2021	Jakub Grossmann	*The effects of minimum wage increases in the Czech Republic*
WP 1/2021	Martin Časta	*Deriving equity risk premium using dividend futures*
WP 15/2020	Petr Polák Nikol Poláková Anna Tlustá	*How bad are trade wars? Evidence from tariffs*
WP 14/2020	Dominika Ehrenbergerová Josef Bajzík	*The effect of monetary policy on house prices – How strong is the transmission?*
WP 13/2020	Zuzana Rakovská	*Composite survey sentiment as a predictor of future market returns: Evidence for German equity indices*
WP 12/2020	Vojtěch Siuda	*A top-down stress-testing framework for the nonfinancial corporate sector*

WP 11/2020	Hana Hejlová Michal Hlaváček Blanka Vačkova	*Estimating commercial property price misalignment in the CEE countries*
WP 10/2020	Zuzana Rakovská Dominika Ehrenbergerová Martin Hodula	*The power of sentiment: Irrational beliefs of households and consumer loan dynamics*
WP 9/2020	Ivan Sutóris	*The intertemporal cost of living and dynamic inflation: The case of the Czech Republic*
WP 8/2020	Martin Hodula Jan Janků Martin Časta Adam Kučera	*On the determinants of life and non-life insurance premiums*
WP 7/2020	František Brázdik Tibor Hlédik Zuzana Humplová Iva Martonosi Karel Musil Jakub Ryšánek Tomáš Šestořád Jaromír Tonner Stanislav Tvrz Jan Žáček	*The g3+ model: An upgrade of the Czech National Bank's core forecasting framework*
WP 6/2020	Volha Audzei Jan Brůha	*A model of the Euro Area, China and the United States: Trade links and trade wars*
WP 5/2020	Dominika Ehrenbergerová Martin Hodula Zuzana Rakovská	*Does capital-based regulation affect bank pricing policy?*
WP 4/2020	Alexis Derviz	*Sovereign capital, external balance, and the investment-based Balassa-Samuelson effect in a global dynamic equilibrium*
WP 3/2020	Milan Szabo	*Growth-at-risk: Bayesian approach*
WP 2/2020	Martin Hodula Ngoc Anh Ngo	*Finance, growth and (macro)prudential policy: European evidence*
WP 1/2020	Michal Franta Ivan Sutóris	*Dynamics of Czech inflation: The role of the trend and the cycle*

CNB Research and Policy Notes (since 2020)

RPN 3/2021	Martin Hodula Jan Janků Lukáš Pfeifer	*Interaction of cyclical and structural systemic risks: Insights from around and after the global financial crisis*
RPN 2/2021	Karel Musil Stanislav Tvrz Jan Vlček	*News versus surprise in structural forecasting models: Central bankers' practical perspective*

RPN 1/2021	Miroslav Plašil	*Designing macro-financial scenarios: The New CNB framework and satellite models for property prices and credit*
RPN 3/2020	Simona Malovaná Martin Hodula Zuzana Rakovská	*Researching the research: A central banking edition*
RPN 2/2020	Simona Malovaná Josef Bajzík Dominika Ehrenbergerová Jan Janků	*A prolonged period of low interest rates: Unintended consequences*
RPN 1/2020	Simona Malovaná	*How to organize research in central banks: The Czech National Bank's experience*

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