// NO.20-092 | 04/2022

DISCUSSION PAPER

// FRANK BRÜCKBAUER

Do Financial Market Experts Know Their Theory? New Evidence From Survey Data

Leibniz Association

ZEW

Do Financial Market Experts Know Their Theory? New Evidence From Survey Data*

Frank Brueckbauer†

This draft: April 2022
First draft: December 2020

Abstract

Using a unique survey dataset, I study how financial market experts form their stock market expectations. I document a strong disagreement among experts about how important macroeconomic and financial variables are related to stock returns. The results of an analysis of the relationships between my main survey measure of expected returns and measures of economic conditions are largely consistent with the view that expected returns are counter-cyclical. In particular, I find a positive relationship between expected returns and the dividend–price ratio, which is at odds with the findings of previous papers studying survey measures of expected returns. Finally, I find that an aggregated measure of the financial market experts' stock return forecasts has weak predictive power for actual returns, but is a less precise forecast than a simple average of historical stock returns.

Keywords: stock market expectations, survey data, macro-finance, stock return predictability
JEL codes: D84, G12

*I would like to thank Erik Theissen and Tabea Bucher-Koenen for their valuable advice and support. I would also like to thank Carolin Schmidt, Michael Schroeder and seminar participants at the University of Mannheim and ZEW Mannheim for helpful comments and suggestions.
†ZEW Mannheim, University of Mannheim. E-mail: frank.brueckbauer@zew.de

1 Introduction

Expected excess returns on risky assets, in particular on stocks, play a pivotal role in finance theory and practice. A good understanding of the properties of expected stock returns is, for example, required in the areas of portfolio management and corporate finance, where return forecasts are an important input to decisions on optimal portfolios and on whether a corporate project is a worthwhile investment (Cochrane, 2011). The existing empirical evidence based on realized stock returns suggests that expected stock returns are time-varying and counter-cyclical (Fama and French, 1989; Cochrane, 2011, 2017). Expected stock returns are considered as time-varying, because realized stock returns seem to be predictable by several (time-varying) macro-financial variables, one of the most prominent variable being the dividend–price ratio of the equity market (see e.g. Campbell (2000) and Welch and Goyal (2008) for a list of forecasting variables).[1] Because most of the variation in the dividend–price ratio of the equity market seems to be unrelated to the variation in dividends, the dividend–price ratio and related variables are interpreted as proxies for expected stock returns (Cochrane, 2011). Expected stock returns are considered to be counter-cyclical, because proxies for expected stock returns seem to be negatively correlated with measures of economic conditions (Fama and French, 1989).

However, evidence based on survey data, which has for a long time been regarded as unreliable and redundant (Gennaioli et al., 2016; Manski, 2018; Giglio et al., 2019), is largely at odds with the evidence based on realized stock returns. The evidence suggests that expected stock returns are negatively correlated with variables that positively predict subsequent realized returns (e.g. Bacchetta et al., 2009; Greenwood and Shleifer, 2014; Amromin and Sharpe, 2014; Adam et al., 2017), that they negatively predict actual stock returns (e.g. Bacchetta et al., 2009; Greenwood and Shleifer, 2014; Amromin and Sharpe, 2014), that they are pro-cyclical (Amromin and Sharpe, 2014) and that they are extrapolative in recent returns on the stock market or returns on the portfolios of the respondents (e.g. Hurd et al., 2011; Greenwood and Shleifer, 2014; Barberis et al., 2015; Hoffmann and Post, 2017). Many authors in this strand of the literature therefore draw the conclusion that survey data of stock return expectations are inconsistent with the assumption of rational expectations.

Motivated by this contradictory evidence on the time-variation in expected stock returns, I use a unique survey dataset to study how financial market experts form their stock market expectations. I study survey measures of stock return expectations, because I regard them as more precise measures of expected stock returns than realized returns, given that the latter can be very noisy, for example, due to information surprises (Elton, 1999). I focus on financial market experts, because I expect their understanding of stock returns to be superior to that of households or individual investors, whose expectations are studied in the large majority of papers in the literature. I also expect the expectations of financial market experts to matter more for asset prices, given that institutional

[1] The issue whether stock returns are predictable has not been settled yet. Welch and Goyal (2008), for example, argue that most of the financial variables that are considered to be predictors of stock returns fail to predict stock returns in out-of-sample tests of predictive power. Examples of papers defending stock return predictability are Campbell and Thompson (2008), Cochrane (2008), Rapach et al. (2010) and Ferreira and Santa-Clara (2011).

investors usually have a bigger impact on asset markets than private investors. Another reason is that the dataset that includes the stock market expectations of financial market experts has additional features that set it apart from other survey datasets studied in the existing literature. More specifically, the dataset is based on micro data from the ZEW Financial Market Survey (ZEW FMS, hereafter), which is a survey among German financial market experts, including professional stock market forecasters. The survey combines questions on macroeconomic and financial developments in Germany and other important economies, which makes it possible to study how the respondents' stock market expectations co-vary with their macroeconomic expectations. Moreover, in the survey, the respondents are asked to provide both qualitative and quantitative forecasts for the German DAX index in six months. This allows me to explore whether the question type matters for the results. Finally, the data on stock market expectations can be combined with personal information about the respondents. The information includes gender, age and indicators of the respondents' skill in forecasting stock returns, for example the respondents' main occupations or whether or not they are professional stock market forecasters.

The aim of this paper is threefold. First, I aim to get a better understanding of the sources of the variation in expected returns. I therefore follow Giglio et al. (2019) and decompose the variance of my quantitative survey measure of expected returns into three components: a component that captures the common time-series variation, a component that captures the variation across respondents and a component that captures the residual variance. The result of the variance decomposition indicates that respondents differ considerably in how they incorporate macroeconomic and financial information into their DAX forecasts. More specifically, I find that the component that captures the common time-series variation is the least important for explaining the total variation in my quantitative survey measure of expected returns, followed by the component that captures the variation across respondents. Together, these two components account for only a third of the total variation, implying that the remaining two thirds are idiosyncratic.

I then move on to study each of the three components in detail. I first explore to what extent the variation across respondents can be traced back to differences in the respondents' personal characteristics. The results suggest that all but one of the studied variables, i.e. birth year, career entry year, main occupation and whether the respondent is or has been a professional DAX forecaster, cannot account for this variation. The only characteristic that seems to be related to the variation across respondents is the self-assessed level of expertise in conducting DAX forecasts. To get a better understanding of the underlying drivers of the common time-series variation, I study the informational overlap with a set of macroeconomic and financial state variables I expect the respondents to consider when they forecast DAX returns. While the informational overlap ranges from non-existent to moderate, when each of the variables are evaluated on their own, they overlap strongly with the common time-series variation in expected returns when evaluated together. Surprisingly, the variable that shows the highest informational overlap is the return of the DAX over the month prior to the survey period. Finally, to illustrate the heterogeneity of how respondents incorporate information into their DAX forecasts, I exploit the long individual time-series in the ZEW FMS dataset and run respondent-level regressions of the quantitative survey measure of expected returns on my set of potential macroeconomic and financial determinants of DAX expectations. For the variables studied, I document considerable differences in R^2 statistics and a disagreement about the direction of the relationships with expected returns among respondents. Put differently, the respondents disagree about the importance of the variables for DAX returns and also about how these variables affect DAX returns.

Second, I aim to provide new evidence on the relationship between expected returns and economic conditions. More specifically, I explore whether expected returns are counter-cyclical, i.e. whether they are higher when economic conditions are bad and vice versa. As measures of economic conditions in Germany, I use the dividend–price ratio of the CDAX, the earnings–price ratio of the CDAX, the respondents' own assessments of the current economic situation in Germany and a composite economic indicator constructed from monthly indicators of German economic conditions. For comparability with the results of other studies, I first explore whether expected returns are counter-cyclical on average, i.e. I initially ignore the heterogeneity of the respondents' expectations. Motivated by the observation that previous studies in the literature are based on both types of expectation data, I also study both the respondents' quantitative and qualitative DAX expectations. An additional benefit of using both variables is that it allows me to investigate whether the result on the relationship between expected returns and economic conditions depends on the type of the expectation data used.

First, I find that, for some variables, the direction of the estimated relationship between economic conditions and DAX expectations depends on whether I use the qualitative DAX return forecasts or the quantitative DAX return forecasts. For example, the dividend–price ratio of the CDAX has a positive coefficient in the regression on the qualitative DAX return forecasts and a negative coefficient in the regression on the qualitative DAX return forecasts. As I am able to rule out that these differences either arise because respondents give answers to both questions that are inconsistent with each other or are the implication of outliers in the qualitative DAX return forecasts, the only remaining interpretation of the evidence is that the scale of the variable, i.e. metric vs. ordinal, matters strongly for the measured relationship between economic conditions and stock return expectations.

Second, focusing on the results for the quantitative forecasts, I find that the survey data is largely consistent with the hypothesis that stock return expectations are counter-cyclical. More specifically, I find that, for three out of the four considered measures of economic conditions, expected returns are on average higher when the measures indicate that economic conditions are bad, all else equal. Somewhat surprisingly, the only measure for which this is not the case, is the only subjective measure of economic conditions, which is the respondents' own assessments of current economic conditions in Germany. Furthermore, although it is only a control variable in the regressions, I also document a negative relationship between expected returns and the DAX return over the month prior to the survey. The evidence presented in previous studies, in contrast, suggests that stock return expectations are extrapolative in recent stock returns (see e.g. Greenwood and Shleifer, 2014; Barberis et al., 2015).

Third, I document minor differences in the relationships between DAX expectations and economic conditions across respondents. When I differentiate by age, I find that the correlation between the earnings–price ratio of the CDAX (which is higher when economic conditions are bad) and expected returns is decreasing with age. When I differentiate by the respondents' self-reported interest in the stock market results of the ZEW FMS, I find that the correlation between the composite economic indicator (which is lower when economic conditions are bad) and expected returns is only negative if the respondents report that they are interested. Lastly, when I differentiate by main occupation, I document that financial market experts across occupations seem to use different combinations of the measures of economic conditions when forecasting DAX returns, suggesting that, of all the characteristics explored, main occupation is the best differentiator when it comes to the relationship between DAX return expectations and measures of economic conditions.

Finally, the third aim of this paper is to evaluate the accuracy of the financial market experts' DAX return forecasts. An evaluation of the forecast accuracy is the natural next step, after I have studied how financial market experts form their stock return expectations. I begin by studying the aggregated quantitative forecast, which is the average expected DAX return by survey wave and the aggregated qualitative DAX return forecast, which is calculated as the difference between the shares of respondents that expect the DAX to increase and decrease, respectively, i.e. a so-called bull–bear spread. I find that the aggregate quantitative forecast is positively correlated with actual returns and explains about 6% of the variation in the latter. The aggregate qualitative forecast, in contrast, seems to be uncorrelated with actual returns. Both results are at odds with the finding of previous studies that survey measures of expected returns are negatively correlated with realized returns (see e.g. Greenwood and Shleifer, 2014).

Having shown that it is positively associated with realized returns, I next ask whether the aggregated quantitative DAX return forecast is superior in terms of forecast accuracy to the average historical return, the latter being an often used benchmark which stock return forecasts are compared to in the literature (see e.g. Welch and Goyal, 2008; Campbell and Thompson, 2008). I find that this is not the case, i.e. the use of the average historical DAX return produces DAX return forecasts that are at least as good as the aggregated DAX forecast from the ZEW FMS. As a final step, I explore whether there are differences in forecast accuracy within subgroups of the ZEW FMS panel formed by the various personal characteristics available to me. Most comparisons yield that the forecasts are equivalent in terms of accuracy. Interestingly, my results suggest that respondents who regularly conduct DAX forecasts outside of the ZEW FMS underperform those who do so only irregularly. In some cases, I also document differences in forecast accuracy when I distinguish by the respondents' main occupations. For example, during the sample period, respondents who have worked in "Trading" have provided DAX return forecasts that were closer to the actual realized returns than respondents who have worked in "Management". In all cases, however, the differences in forecast accuracy cannot be attributed to differences in how the respective groups form their DAX return expectations conditional on economic conditions.

To sum up, I document a strong disagreement among respondents about how important macroeconomic and financial variables are related to DAX returns. Despite this strong heterogeneity, the empirical evidence is largely in support of the view that expected returns are counter-cyclical. The two findings that weaken my results in this respect are that the respondents' own assessments of current economic conditions – the only subjective measure of economic conditions – are on average positively associated with expected returns and that the relationship between expected returns and economic conditions is not negative for all respondents. A methodological result is that the measured relationship between expected returns and economic conditions depends on whether I study qualitative or quantitative DAX return forecasts. Lastly, I find that the average quantitative DAX return forecast has predictive power for actual DAX returns, but is not superior to a simple average of historical DAX returns. However, because it is positively correlated with realized returns, the aggregated quantitative DAX return forecast from the ZEW FMS panel is a more accurate forecast than the survey measures of expected returns studied in the previous literature (e.g. Greenwood and Shleifer, 2014), which were found to be negatively correlated with realized returns.

The paper proceeds as follows. Section 2 gives an overview of the literature to which this paper contributes. Section 3 introduces the ZEW Financial Market Survey, which is the main data source for this study and describes the composition of the associated panel of financial market experts. Section 4 gives more details about my two survey measures of stock return expectations and the other macroeconomic and financial variables studied in this paper. Section 5 contains the analysis of the sources of the variation in the quantitative DAX return forecasts. Section 6 explores whether expected returns are countercyclical. Section 7 studies the accuracy of the financial market experts' DAX return forecasts. Section 8 concludes.

2 Related Literature

This paper contributes to different strands of the literature studying the determinants of stock return expectations using survey data. Table 1 gives an overview of different surveys studied in this literature. First, my paper contributes to the literature that is concerned with the relationship between survey measures of expected stock returns and variables that are considered to be proxies for expected returns. There is extensive empirical evidence in this literature that suggests that survey measures of expected returns are negatively correlated with these proxies. One of the first studies in this strand is Vissing-Jorgensen (2003), which documents that expected returns were high when the US stock market was high between 1998 and 2003. Follow-up studies by Bacchetta et al. (2009), Greenwood and Shleifer (2014), Amromin and Sharpe (2014) and Adam et al. (2017) find that survey measures of expected returns are negatively correlated with the dividend–price ratio of the stock market, the negative of surplus consumption[2] and the consumption–wealth ratio[3]. All three variables are, however, considered to be positively correlated with subsequent realized returns (Campbell and Cochrane, 1999; Lettau and Ludvigson, 2001; Cochrane, 2011). As shown by Greenwood and Shleifer (2014), this result holds for different survey measures of stock return expectations. They study six different survey measures which they find to be highly positively correlated with each other and negatively correlated with proxies for expected stock market returns. Lastly, Amromin and Sharpe (2014) document that survey expectations of stock returns are pro-cyclical, i.e.

[2]See Campbell and Cochrane (1999) for the definition of surplus consumption.
[3]See Lettau and Ludvigson (2001) for the definition of the consumption–wealth ratio.

they are higher when economic conditions are good and vice versa, which is at odds with empirical evidence based on realized returns (see e.g. Fama and French (1989))) and the implications of consumption-based asset pricing models (e.g. Campbell and Cochrane (1999)). Söderlind (2010), in contrast, finds that survey forecasts of economists are higher in recessions. However, he also finds that expectations are negatively correlated with the dividend–price ratio.

My paper also contributes to the literature that documents systematic differences in stock return expectations across individuals. Dominitz and Manski (2004, 2007) study stock market expectation data from the Michigan Survey of Consumers and the Health and Retirement Study and find that expectations differ systematically by sex, age and schooling. Using the Michigan Survey of Consumers, Dominitz and Manski (2011) further show that the respondents differ in how they use available information to forecast stock returns.[4] Using data from the Health and Retirement Survey, Hudomiet et al. (2011) document an increase in the cross-sectional heterogeneity of expected returns after the US stock market crash of 2008, where the increase of the heterogeneity has been the highest for respondents who own stocks, for those who follow the stock markets and for those with higher average cognitive capabilities. Hurd et al. (2011) study data from the centER Panel and report lower expected returns for females and higher expected returns for active traders. Finally, Giglio et al. (2019) administer a survey among randomly selected U.S. based clients of Vanguard to study the link between the respondents' expectations and their portfolio holdings. They decompose the variance of their measure of stock return expectations and find that the majority of the variation is explained by person fixed effects. Giglio et al. (2019) further explore whether the person fixed effects in expected returns can be traced back to observable personal characteristics like gender or age etc., but find that this is not the case.

Lastly, my paper is related to research that evaluates the predictive power of stock market forecasts obtained from survey data. Bacchetta et al. (2009) study survey data from UBS/Gallup and the ICF of the Yale School of Management and find that variables that forecast realized returns – the dividend–price ratio of the stock market in particular – are negatively correlated with the forecast errors made by the respondents. Deaves et al. (2010) use the ZEW FMS dataset to study 90% confidence intervals for stock returns. They find that, during the sample period between 2003 and 2005, the percentage of respondents, for which the respective confidence interval contained the realization of the DAX, ranges from around 10% to about 80%. In a follow-up study, Deaves et al. (2019) document that the mean forecast for the excess DAX return explains about 6% of the variation in actual DAX returns out-of-sample. Söderlind (2010) analyzes the forecasting performance of the Livingston Survey and reports that the median forecast has no explanatory power for realized returns out-of-sample. Finally, Greenwood and Shleifer (2014) document a weak and negative relationship between the survey measures of expected returns studied by them and subsequent realized returns. They attribute this result to the negative relationship between survey expected returns and proxies for expected return.

[4] Glaser et al. (2019) show that the way how individuals use the information available to them to make a forecast also depends on how the information is presented to them.

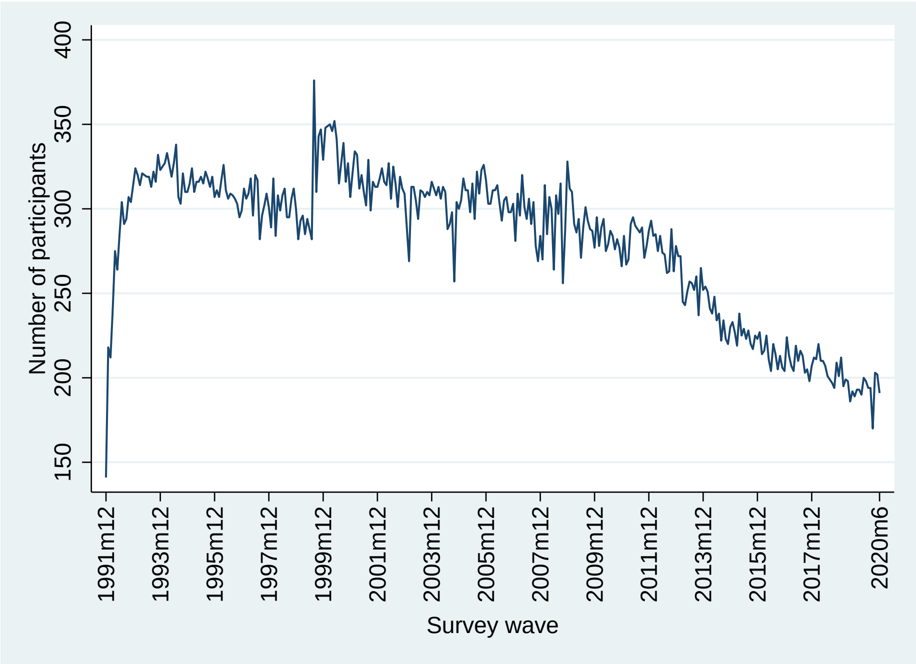
Table 1: Overview of surveys studied in the literature

Survey	Participants	Availability of stock market expectations and frequency	Country of respondents / stock markets	Type of forecast	Authors
UBS/Gallup Investor Survey	households	1996–2007, monthly	USA/USA	quantitative	Vissing-Jorgensen (2003); Bacchetta et al. (2009); Greenwood and Shleifer (2014); Amromin and Sharpe (2014); Adam et al. (2017)
Stock Market Confidence Indices, International Center for Finance, Yale School of Management	wealthy individuals, institutional investors	1989–, monthly	USA/USA, Japan/Japan	qualitative	Bacchetta et al. (2009)
The CFO Survey (Graham and Harvey)	financial professionals	1998–, quarterly	USA/USA	quantitative	Greenwood and Shleifer (2014)
The American Association of Individual Investors Investor Sentiment Survey	individual investors	1987–, weekly	USA/USA	qualitative	Greenwood and Shleifer (2014)
Survey of Consumers, Michigan University	households	2000–2005, monthly	USA/USA	quantitative	Dominitz and Manski (2004, 2011); Greenwood and Shleifer (2014); Amromin and Sharpe (2014)
Livingston Survey	economists	1946–, monthly	USA/USA	quantitative	Söderlind (2010)
centER	households	2004, 2006	Netherlands/Netherlands	quantitative	Hurd et al. (2011)
Health and Retirement Study	households	2002–, bi-annually	USA/USA	quantitative	Dominitz and Manski (2007); Hudomiet et al. (2011)
Giglio et al. (2019)	clients of Vanguard	2017–2019, bi-monthly	USA/USA	quantitative	Giglio et al. (2019)

3 The ZEW Financial Market Survey

My main data source is the ZEW Financial Market Survey (ZEW FMS). The ZEW FMS is a monthly panel survey among German financial market experts that covers macroeconomic and financial developments in Germany and other important countries. The panel members are predominantly Germans who work in financial institutions and corporate finance departments of non-financial companies in Germany. The ZEW FMS was first conducted in December 1991 and is still running. Until the end of 2019, the length of each survey period was two weeks. Since 2019, it has been one week. As of June 2020, the number of monthly participants has ranged from 141 to 376 since the beginning of the survey. The time series of the monthly participants is depicted in Figure 1.

Figure 1: Monthly number of participants of the ZEW FMS



Important features of the survey design are that the survey is anonymous and that the participants receive as a non-monetary compensation for taking part the aggregated results, as well as a short report with comments on the most important results. The anonymity of the participants is important because the participants might otherwise be discouraged from reporting their true expectations (Croushore, 1993). Given that the ZEW FMS has a high international media coverage and is closely followed by economists and by finance practitioners, receiving the survey results for free likely is sufficiently valuable to motivate the financial experts to participate. Moreover, the participants receive the results prior to the release on the ZEW website.

The survey questions cover macroeconomic and financial developments in Germany, France, Italy, the Eurozone, Great Britain, the USA and Japan. The questionnaire consists of a set of regular questions and one or more extra questions, with varying topics. In the regular macroeconomic questions, the participants are asked to provide their assessments of current economic conditions, as well as their medium-term expectations regarding economic growth and inflation. The regular financial questions cover the participants' medium-term expectations with respect to short-term and long-term interest rates, exchange rates, the price of oil and important stock market indices.

The questionnaire includes three questions about the German DAX index. The results to these questions are the focus of this paper. The first question asks the participants to provide a qualitative forecast of the level of the DAX in six months. More specifically, the participants are asked whether they expect the DAX to "increase", "not change" or "decrease". This question has been asked since 1991. The second question asks the participants to provide a point forecast, as well as the lower and upper bounds of a 90% confidence interval, for the DAX in six months. This question was added to the questionnaire in 2003. The third question is concerned with the current level of the DAX and was added in 2011. In this question, the participants are asked whether they think that DAX is currently "fairly-valued", "over-valued" or "under-valued" in view of the current fundamentals of the DAX companies.

3.1 Panel Composition

On entry to the ZEW FMS panel, the participants are asked to provide details about themselves. These details are only available to researchers. The personal details include gender, age, career entry year and the highest achieved educational degree. Personal characteristics are occasionally also collected retrospectively. Examples are the respondents professional occupation, whether the respondents are currently or have been professional DAX forecasters in the past and the participants' self-assessed levels of expertise in answering the ZEW FMS questions. Unfortunately, not all of these details are available for every panel member. Reasons are that the collection of personal details only began after the start of the survey and that the panel members do not have to answer these questions, so some decide not to.

Figures 2 and 3 illustrate how the ZEW FMS panel is composed in terms of gender, birth year, main occupation and professional experience in stock market forecasting. Since the group of respondents fluctuates from month to month, I document both the composition of the full panel, i.e. that of all current and past participants, as well as the composition by survey wave. As of June 2020, the dataset includes responses of a total of 1,971 different participants. Panels 2a and 2b of Figure 2 show the panel composition by gender. For about 70% of the panel members, gender is unknown. As Panel 2b reveals, the information about gender is mainly missing for panel members that were active before the year 2010.[5] It is also revealed that gender is highly unevenly distributed in the panel: of the 30% of panel members with known gender, about 93% are male.

Panels 2c and 2d of Figure 2 depict the panel composition by birth year. The distribution of birth years ranges from 1938 to 1990, with a median of 1965 (Panel 2c). Over the years, the distribution of birth years has moved upwards, i.e. the median birth year has increased from around 1955 to 1965, while the

[5]However, it should be possible to infer the gender from the panel members' names, which are available. I leave this for future projects.

differences between the 25th percentile, the median and the 75th percentile have remained largely stable (Panel 2d). The upward movement suggests that participants exiting the panel are usually replaced by younger participants. The share of participants for which the birth year is unknown, is also very high in the beginning of the sample period and decreases to under 50% over time.

Panels 3a and 3b of Figure 3 display the panel composition by main occupation. The variables for main occupation combine the results of special surveys from 2011 and 2020. More specifically, if respondents had a given main occupation in either 2011 or 2020, I assume that they had this main occupation during the full sample period. A respondent thus can have multiple main occupations. The information on occupation is available for about 17% of the panel members. As can be seen in Panel 3b, the availability of the information on main occupation is mainly restricted to the current field of participants. The three most frequent main occupations are "Fund Management", "Economic Research" and "Wealth Management".

Finally, Panels 3c and 3d of Figure 3 present the panel composition by professional experience in stock forecasting. The variable can take the values "regular", "sometimes" and "never", which refers to the frequency of the respondents' DAX forecasting activities outside of the scope of the ZEW FMS. The variable combines the results of special surveys from 2013 and 2020. Similar to the assumptions with respect to main occupation, I assume that respondents had a high professional experience, i.e. regularly forecasted the DAX outside of the scope of the ZEW FMS, when they answered so in either 2013 or 2020.[6] As Panel 3c reveals, details on professional stock market forecasting activities are available for about 43% of the panel members. While 20% of panel members have never conducted DAX forecasts outside of the scope of the ZEW FMS, 16% and 7% have done so regularly or irregularly, respectively. In recent years, the share of participants that regularly and professionally forecasts the DAX has fluctuated around 30% (see Panel 3d).

4 Data And Data Preparation

This section introduces the variables I use in this paper. These include my two survey measures of DAX return expectations, the macroeconomic and financial variables used to measure economic conditions and macroeconomic and financial control variables.

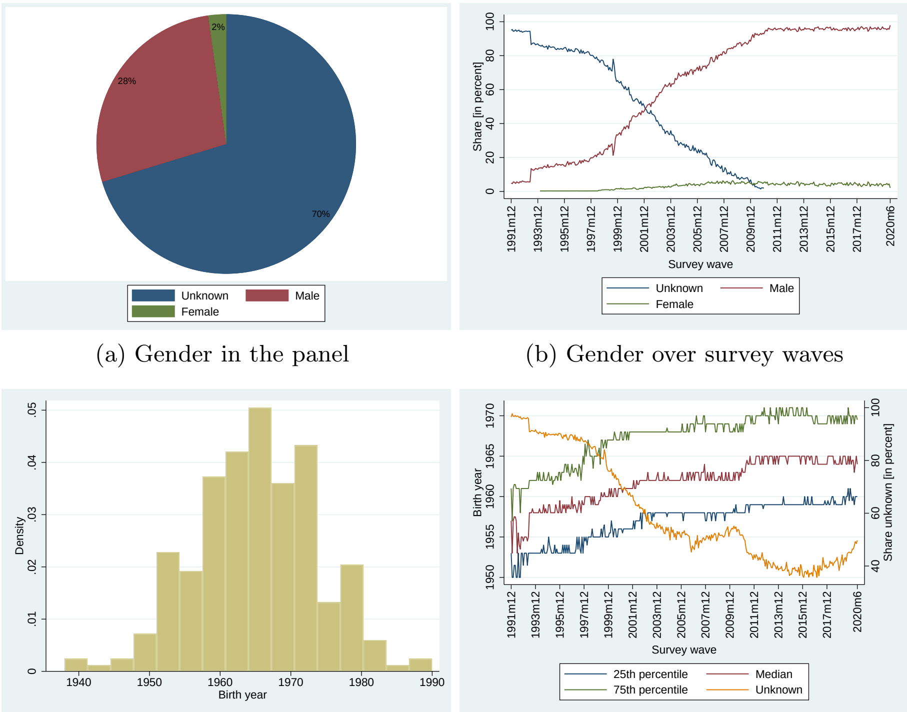
4.1 Survey Measures Of DAX Return Expectations

My main variable of interest is the expectation of the return of the DAX over the next six months, obtained from the ZEW FMS. Because the respective question in the survey asks the participants to provide a forecast of the level of the DAX in six months, the level forecast needs to be transformed into a return first. I define the return forecast implied by the level forecast as

$$expret_{i,s,t} = \frac{E_{i,t}[P^{DAX}_{t+6m}]}{P^{DAX}_t} - 1, \tag{1}$$

[6]The option "sometimes" was only available in the surveys in 2020.

Figure 2: Panel composition: gender and birth year



(a) Gender in the panel

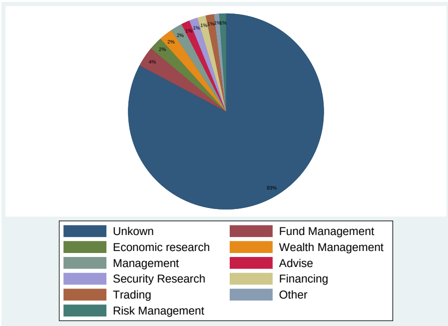
(b) Gender over survey waves

(c) Birth year in the panel

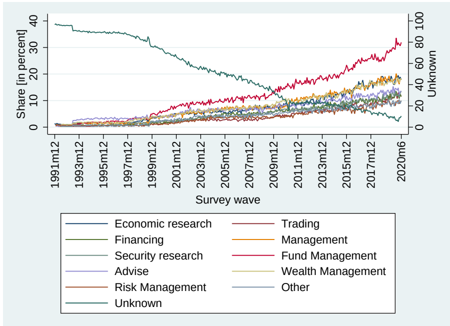
(d) Birth year over survey waves

Notes: These figures illustrate the compostion of the ZEW FMS in terms of gender and birth year. The figures on the left show the composition of the full panel, i.e. all current and past participants of the ZEW FMS. The figures on the right show how the composition has evolved over time.

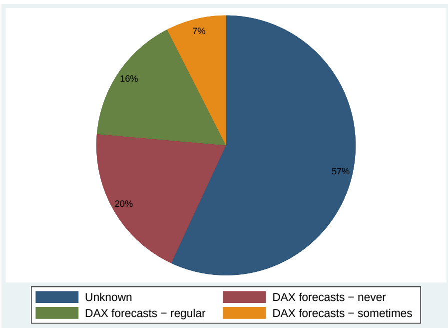
Figure 3: Panel composition: main occupation and professional experience in stock market forecasting



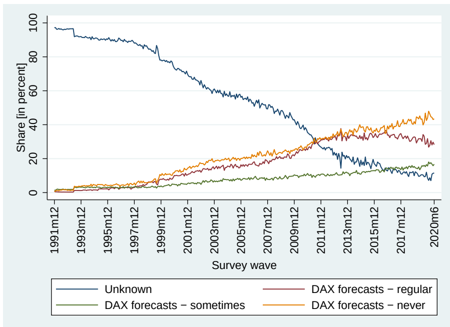
(a) Main occupation in the panel



(b) Main occupation across survey waves



(c) Professional experience in stock market forecasting in the panel



(d) Professional experience in stock market forecasting across survey waves

Notes: These figures illustrate the compostion of the ZEW FMS in terms of main occupation and professional experience in stock market forecasting. The figures on the left show the composition of the full panel, i.e. all current and past participants of the ZEW FMS. The figures on the right show how the composition has evolved over time.

where $E_{i,t}[P^{DAX}_{t+6m}]$ is the point forecast of the level of the DAX in six months of respondent i on date t and P^{DAX}_t is the latest closing level of the DAX available at date t. In some cases, it was necessary to clean the DAX forecasts $E_{i,t}[P^{DAX}_{t+6m}]$ prior to the calculation of the implied return. In these cases, I have applied the following adjustment rules to the raw data. First, if a respondent abbreviated numbers, the forecast was multiplied by an appropriate factor. A forecast of 12.5, for example, was multiplied by the factor thousand, resulting in the forecast 12,500. Second, if the 90% confidence interval for the DAX in six months provided by a respondent did not contain his or her DAX expectation, it was assumed that middle response of the three values is the actual DAX expectation. To minimize the effect that these manual adjustments have on my results, I include the variable *corrected* in all of my analyses, which takes the value of one if the original DAX expectation has been corrected and zero otherwise.

For comparability with the results of other studies, I also study the financial market experts' qualitative forecasts of the DAX in six months, which I refer to as *expdir*. In the respective question, the survey participants are asked whether they expect the DAX to "increase", "not change" or "decrease" over the next six months, thus $expdir_{i,s,t} \in \{increase, notchange, decrease\}$. Qualitative stock return forecasts of this type are usually aggregated by calculating the difference between the share of respondents who expect the DAX to increase and the share of respondents who expect the DAX to decrease, i.e. a so-called bull–bear spread. I will follow this convention when I study survey expectations of DAX returns at the aggregated level in Section 7.1.

4.2 Other Data

To obtain a better understanding of the determinants of the respondents' DAX expectations, I relate my survey measures of expected returns to a set of macroeconomic and financial state variables, as well as to the respondents' answers to other questions from the ZEW FMS. My variable selection is thereby guided by asset pricing theory and empirical evidence. I distinguish between four groups of explanatory variables. Table 2 contains a list of all variables used in the empirical analyses.

The first group includes variables that are considered to be predictive for realized returns. The two variables in this group are the dividend–price ratio (*dp*, hereafter) and the earnings–price ratio (*ep*, hereafter) of the equity market. As the relevant measure of the German equity market, I use the CDAX index. I consider *dp*, because the dividend–price ratio is one of the most studied proxy variable for expected stock returns in the literature (see e.g. Cochrane, 2008, 2011) and therefore also studied in Greenwood and Shleifer (2014) and Amromin and Sharpe (2014). I additionally consider the earnings–price ratio, because there seems to have been a disconnect between earnings and dividends before the financial crisis of 2007–2009 (see Section 6.1), which has implications for my results in Section 6. Two other, potentially interesting forecasting variables studied in Greenwood and Shleifer (2014) are the consumption-wealth-ratio (Lettau and Ludvigson, 2001) and the surplus–consumption ratio (Campbell and Cochrane, 1999). These are, however, unavailable to me.[7]

[7]With respect to the consumption–wealth ratio, I lack the information about the wealth

Table 2: List of macroeconomic and financial covariates

Variable	Abbreviation	Source	Comments
Log dividend–price ratio, CDAX	*dp*	Eikon Datastream (CDAXGEN)	Datatype: DSDY
Log earnings–price ratio, CDAX	*ep*	Eikon Datastream (CDAXGEN)	Datatype: DSPE
Industrial production, Germany	*ipgrowth*	Eikon Datastream (BDIPTOT.G)	Year-on-year growth rate; inflation, calendar and seasonal adjusted; publication lag: 2 months
Employment rate, Germany	*empl*	Eikon Datastream (BDUN%TOTR)	Calculated as $1-$unemployment rate; in percent of civilian labor force; publication lag: 1 month
Inflation rate, Germany	*infl*	Eikon Datastream (BDCPANNL)	Year-on-year change in prices; publication lag: 1 month
Consumer confidence, Germany	*conf*	Eikon Datastream (BDCNFCONQ)	Based on European Commission consumer survey; publication lag: 1 month
Price of crude oil	*oil*	Eikon Datastream (CRUDBFO)	European Brent Spot
Exchange rate US-Dollar to Euro	*exchrate*	Eikon Datastream (USEURSP)	
DAX return - 12 months to 3 months prior to date t	*dax12to3*	Eikon Datastream (DAXINDX)	Datatype: RI
DAX return - 3 months to 1 month prior to date t	*dax3to1*	Eikon Datastream (DAXINDX)	Datatype: RI
DAX return - 1 month prior	*dax1to0*	Eikon Datastream (DAXINDX)	Datatype: RI
Assessment of the current economic situation, Germany	*sit*	ZEW FMS dataset	Ordinal variable, response options: “good”, “normal”, “bad”
Outlook for economic situation, Germany (6 months)	*expsit*	ZEW FMS dataset	Ordinal variable, response options: “improve”, “not change”, “worsen”
Outlook for inflation rate, Germany (6 months)	*expinfl*	ZEW FMS dataset	Ordinal variable, response options: “improve”, “not change”, “decrease”
Outlook for short-term interest rates, Eurozone (6 months)	*expint_st*	ZEW FMS dataset	Suggested reference in questionnaire: “3-month interbank rates”; ordinal variable, response options: “improve”, “not change”, “decrease”
Outlook for long-term interest rates, Germany (6 months)	*expint_lt*	ZEW FMS dataset	Suggested reference in questionnaire: “yields on 10-year bonds”; ordinal variable, response options: “improve”, “not change”, “decrease”

The second group includes variables that contain information about the current state of the German economy. Because variables considered to be predictive for stock returns seem to move with business cycles (see e.g Fama and French, 1989; Cochrane, 2017), Amromin and Sharpe (2014) study the correlations between their survey measure of expected returns and measures of economic conditions. Following Amromin and Sharpe (2014), I study the respondents' own assessment of the current economic situation in Germany from the ZEW FMS dataset. I also consider the following economic indicators for Germany: the year-on-year growth rate of industrial production (*ipgrowth*), the employment rate (*empl*), the year-on-year growth rate of the German Consumer Price Index (*infl*), a consumer confidence indicator (*conf*), the exchange rate between US dollars and the euro (*exchrate*) and the price of crude oil (*oil*). Most of the economic indicators have a publication lag, meaning that the respondents learn the realizations of these variables only after one or two months. Since I would otherwise compare the DAX expectations to the realizations of the economic indicators that were unknown to the respondents at the time of the response, I shift these variables by their respective publication lag.

The third group encompasses the respondents' answers to forward-looking questions regarding the German economy from the ZEW FMS dataset. These are the respondents' outlooks with respect to the general economic situation, the inflation rate, short-term interest rates and long-term interest rates. I include these variables because they are likely correlated with the respondents' assessment of the current economic situation, which is a key explanatory variable for expected returns in Section 6. Moreover, the results of Amromin and Sharpe (2014) suggest that these variables might themselves be important explanatory variables for DAX expectations.

The fourth group includes past DAX returns. Past returns have been shown to explain survey expectation of stock returns (see e.g. Greenwood and Shleifer, 2014; Barberis et al., 2015). Here I consider the return of the DAX up to 12 months prior to each response and split the 12–month return into three parts: the return from $m_s - 12m$ to $m_s - 3m$ (*dax12to3*), the return from month $m_s - 3m$ to $m_s - 1m$ (*dax3to1*) and the return from month $m_s - 1m$ up to the day of the response (*dax1to0*), where m_s is the month of survey wave s.

5 Understanding The Sources Of The Variation In Expected Stock Returns

In this section, I study the sources of the variation in my quantitative survey measure of DAX return expectations, *expret*. I follow Giglio et al. (2019) and decompose the variance of *expret* into three components. The first component captures the common variation in *expret* over time, for example, due to changes in the general macroeconomic and financial environment and is obtained by regressing *expret* on either survey fixed effects or time fixed effects, where time fixed effects are fixed effects for the specific days on which the participants completed the questionnaire. While survey fixed effects only capture the time-series variation across survey waves, time fixed effects additionally capture the

of German households. To obtain the surplus–consumption ratio, it is necessary to calibrate the habit-model to the German economy. The benefit of this calibration is only minor.

time-series variation within survey waves. Under the assumption of rational expectations and the absence of private information, most of the variation in expected returns is driven by this component (Manski, 2018). The corresponding regression models are

$$expret_{i,s,t} = \sum_{s=1}^{S} \phi_s D_s + \epsilon_{i,t} \tag{2}$$

$$expret_{i,s,t} = \sum_{t=1}^{T} \phi_{s,t} D_{s,t} + \epsilon_{i,t}, \tag{3}$$

where $expret_{i,s,t}$ is the implied quantitative DAX return expectation of respondent i on date t in survey wave s for a horizon of six months and ϕ_s and $\phi_{s,t}$ are the survey and time fixed effects, respectively. Note that date t is always uniquely associated with a survey wave (e.g. June 2020) which is indexed by s. The indices s and i thereby run from 1 to the number of survey waves, S and the number of survey days, T, respectively. To avoid that my results on the importance of time fixed effects are driven by days with low numbers of responses, I exclude all survey days where the number of responses is lower than 30 when estimating Equation (3).

The second component captures systematic differences in the overall level of *expret* in the cross-section of respondents, for example because some respondents are generally optimistic or pessimistic and is obtained by regressing *expret* on respondent fixed effects. The corresponding regression model is

$$expret_{i,s,t} = \sum_{i=1}^{I} \phi_i D_i + \epsilon_{i,s,t}, \tag{4}$$

where ϕ_i is the fixed effect of respondent i and I is the total number of respondents in the ZEW FMS panel.

The third component is the residual variance in a regression of *expret* on survey and respondent fixed effects or time and respondent fixed effects. The residual variance can be attributed to either idiosyncratic changes in expectations over time or noise (Giglio et al., 2019). The corresponding regression models are

$$expret_{i,s,t} = \sum_{s=1}^{S} \phi_s D_s + \sum_{i=1}^{I} \phi_i D_i + \epsilon_{i,t} \tag{5}$$

$$expret_{i,s,t} = \sum_{t=1}^{T} \phi_{s,t} D_t + \sum_{i=1}^{I} \phi_i D_i + \epsilon_{i,t}. \tag{6}$$

Table 3 reports the R^2 statistics from estimated models (2) to (6). Columns 1 and 2 of Table 3 reveal that survey and time fixed effects account for only about 10.5% and 12.7%, respectively, of the variation in *expret*, adjusted for the degrees of freedom. The result that time fixed effects explain a larger share of the variance of *expret* than survey fixed effects indicates that the respondents' information sets relevant for DAX forecasts change on a daily basis and may change considerably during a given survey period, which has to be considered when aggregating forecasts. Column 3 shows that the adjusted R^2 statistic for respondent fixed effects is 23.4%. While respondent fixed effects explain a larger share of the variance of *expret* than survey or time fixed effects, the share explained is significantly lower than that measured in other survey datasets. Giglio et al. (2019), for example, find that person fixed effects account for nearly 60% of the variation in their survey measure of expected returns. Finally, the (adjusted) R^2 statistics reported in columns 4 and 5 imply that the majority of the variation in *expret* has to be attributed to idiosyncratic changes in expectations and noise: the combinations of survey and respondent fixed effects, as well as time and respondent fixed effects, explain only 33.3% and 36.3%, respectively, of the variance in *expret*.

Table 3: Variance decomposition of *expret*

Dependent variable: *expret*	Survey fixed effects	Time fixed effects	Respondent fixed effects	Survey & respondent fixed effects	Time & respondent fixed effects
R^2	10.9%	14.7%	24.7%	34.7%	39.5%
Adj. R^2	10.5%	12.7%	23.4%	33.3%	36.3%
N	45,605	26,251	45,605	45,605	26,251
Comments		#responses ≥30			#responses ≥30

Notes: This table reports the results of separate regressions of *expret* on survey fixed effects, time fixed effects and respondent fixed effects. The dependent variable *expret* has been orthogonalized with respect to the variable *corrected*. In the regressions that include time fixed effects as independent variables, all observations were excluded for which the number of total responses on the day on which the respective response was submitted is below 30.

5.1 Decomposing Respondent Fixed Effects

Having quantified the relative importance of the three components of the variance of *expret*, I move on to study the three components in detail. To shed more light on the component that captures the variation in *expret* across respondents, I ask to what extent the variation in the estimated respondent fixed effects, $\hat{\phi}_i$, are explainable by differences in the respondents' observable characteristics. Available characteristics are the respondents' birth years, career entry years, their main professional occupations, whether they are currently or were professional DAX forecasters in the past and their own assessments of their level of expertise in forecasting the DAX. Because this information has been collected in different surveys, the number of observations for each characteristic varies considerably. The corresponding regression model is

$$\hat{\phi}_i = \alpha + x_i\beta_i + \epsilon_i, \tag{7}$$

where $\hat{\phi}_i$ is the estimate of respondent i's fixed effect, x_i is a row vector holding the characteristic and β_i is a column vector holding the coefficient.

Table 4 documents the R^2 statistics from separate regressions of the estimated respondent fixed effects on respondent characteristics. The first two columns reveal that the variables *birth year* and *career entry year* do not explain the variation in respondent fixed effects. The occupation variables in the third column imply a R^2 statistic of about 11%, which shrinks to almost 0% when it is adjusted for the number of variables. With an adjusted R^2 statistic of about 3%, the categorical variable indicating whether the respondent is currently or was a professional DAX forecaster has small explanatory power for the variation in the estimated respondent fixed effects (fourth column). The respondents' own assessment of their level of expertise in forecasting the DAX produces an adjusted R^2 statistic of 6.6% (fifth column) and is therefore the variable that explains the largest share of the cross-sectional variance of respondent fixed effects. Finally, the model that includes all variables yields a R^2 statistic of about 55% (sixth column). However, the high R^2 statistic is mainly the implication of the large number of variables relative to the number of observations (only 58). Adjusted for the number of variables, the R^2 statistic is about 14%. In summary, differences in the respondents' observable characteristics account for only a small share of the variation in *expret* across respondents. Variables that proxy for the respondents' experience in conducting DAX forecasts have the highest explanatory power for the cross-sectional variance of respondent fixed effects.

5.2 Common Time-series Variation

The results of the variance decomposition of *expret* indicate that between about 10.5% (for survey fixed effects) and about 12.7% (for time fixed effects) of its variation can be attributed to common times-series variation. In this section, I attempt to identify the macroeconomic and financial determinants of *expret* that are captured by this component. I consider a variable as a potential driver of the common variation in *expret* if there is a considerable informational overlap between the variable and survey or time fixed effects. To quantify the informational overlap, I compare the adjusted R^2 statistic from the regression of *expret* on the candidate variable to that from the regression of *expret* on the candidate variable plus survey or time fixed effects. The difference in adjusted R^2 then indicates how much of the common time-series variation in *expret* is explained by the candidate variable. In other words, the smaller the increase in adjusted R^2 when survey or time fixed effects are added to the model, the higher is the informational overlap and the more important is the variable for explaining the common variation in *expret*.

I consider the following macroeconomic and financial variables as potential drivers of the common variation in *expret* over time. The macroeconomic variables are *ipgrowth*, *empl*, *infl* and *conf*. Given that these variables do not vary within surveys, I only consider them in the analysis of survey fixed effects.The financial variables are *dp*, *ep*, *exchrate*, *oil*, *dax12to3*, *dax3to1* and *dax1to0*. Since

Table 4: Variance decomposition of respondent fixed effects

Dependent variable: respondent fixed effects	Birth year	Career entry year	Occupation	Professional forecasting activities	Expertise DAX forecasts	All variables
R^2	0.0%	0.0%	10.7%	3.4%	10.9%	54.5%
Adj. R^2	-0.04%	-0.04%	0.8%	2.7%	6.6%	13.6%
N	256	253	191	281	132	58

Note: This table reports the results of separate regressions of estimates of the respondents' fixed effects on their personal characteristics. The dependent variable *expret* has been orthogonalized with respect to the variable *corrected*.

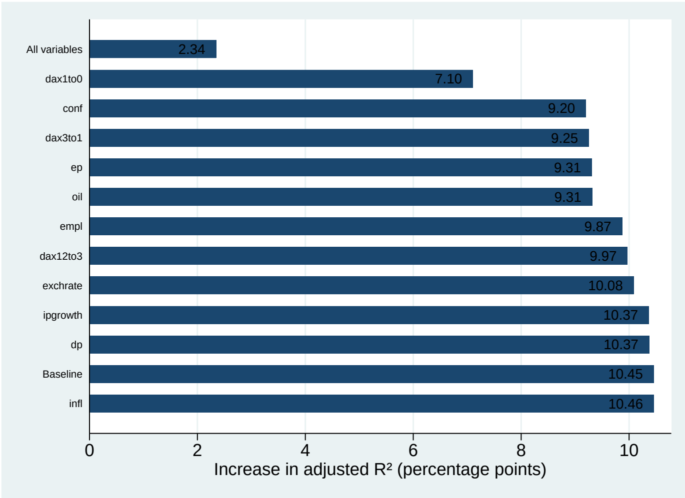
these financial variables have a daily frequency and thus vary also within survey waves, I use survey averages in my analysis of survey fixed effects.

Figure 4 documents the informational overlap between the macroeconomic and financial variables and survey and time fixed effects. The overlap with survey fixed effects is depicted in Figure 4a. Each bar represents the increase in adjusted R^2 when survey fixed effects are added to a regression of *expret* on the respective variable(s). The baseline model only includes survey fixed effects and is the benchmark against which the other models are compared. I also evaluate the model that includes all considered candidate variables. The results are the following. The variable with the least overlap with survey fixed effects is *infl*. When survey fixed effects are added to the regression of *expret* on *infl*, the adjusted R^2 increases by about 10.46 percentage points. The result that the increase is larger than the adjusted R^2 statistic of the baseline model indicates that variation in *infl* is unrelated to the time-series variation in *expret*. With an increase of about 7.10 percentage points, the variable with the highest overlap with survey fixed effects is the return of the DAX over the month prior to dates when the responses are submitted, averaged by survey wave, *dax1to0*. As the first bar illustrates, the model that includes all variables has the highest overlap with survey fixed effects. When survey fixed effects are added to this model, the increase in the adjusted R^2 is only about 2.34 percentage points, suggesting that these variables are direct or indirect drivers of the common time-series variation in *expret*. However, the large difference between the increase in adjusted R^2 for the full model and the increases in the adjusted R^2 for the individual variables, suggests that informational overlap across the considered macroeconomic and financial variables is rather small. Interestingly, *dp*, which should be one of the most important variables, ranks very low and is able to explain only a very small share of the common time-series variation in *expret*. The earnings–price ratio (*ep*), in contrast, does relatively better, but still has a smaller informational overlap with survey fixed effects than, for example, *conf*.

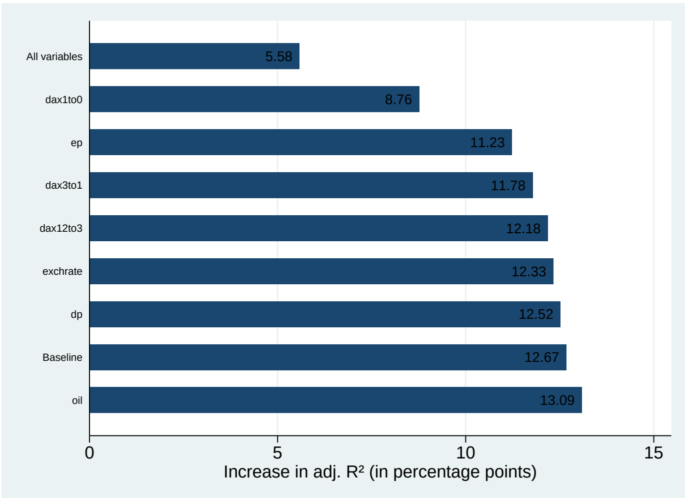
Figure 4b documents the overlap between the financial candidate variables and time fixed effects. As before, I dropped all survey period days, on which the total number of responses is below 30. Qualitatively, the results are similar to those from the analysis of survey fixed effects. Again, *dax1to0* is the variable with the highest overlap with time fixed effects. The variable *ep* performs better than *dp*, the latter showing only little overlap with time fixed effects. Finally, the combination of all investigated variables shows a sizable informational overlap with time fixed effects. Adding these fixed effects to the full model leads to an increase in the adjusted R^2 of about 5.58 percentage points versus an increase of about 12.67 percentage points for the baseline model. The overlap is, however, smaller than in the analysis of survey fixed effects, in which the increase for the full model was only about 2.34 percentage points, but where the full model also includes the macroeconomic variables.

To sum up, none of the considered variables shows a significant informational overlap with survey and time fixed effects when considered on their own. The variable with the highest overlap is *dax1to0*. Only when considered together, the variables account for the majority of the common variation in *expret* over time.

Figure 4: Measuring the informational overlap between survey and time fixed effects and potential determinants of *expret*



(a) Survey fixed effects



(b) Time fixed effects

Note: Figures 4a and 4b document the increases in adjusted R^2 when survey fixed effects (Figure 4a) and time fixed effects (Figure 4b), respectively, are added to regressions of *expret* on the variables on the vertical axes. Lower values are interpreted as a higher informational overlap between the respective variables and survey or time fixed effects. *Baseline* refers to the model that only includes survey fixed effects (Figure 4a) or time fixed effects (Figure 4b). The variables *dax1to0*, *dax3to1*, *dax12to3*, *ep*, *dp*, *oil* and *exchrate* in Figure 4a are survey wave averages. The variables *ipgrowth*, *infl*, *conf* and *empl* in Figure 4a have been shifted by their respective publication lags. Figure 4b only reports the increases in adjusted R^2 for values that have a daily frequency. Moreover, all observations for which the number of total responses on the days, on which the respective response was submitted is below 30 have been excluded.

5.3 Idiosyncratic Variation

Finally, I turn to the idiosyncratic component, which accounts for the highest share of the variance of *expret*. The high importance of this component indicates a large heterogeneity of how respondents incorporate information into their DAX forecasts. To shed more light on this heterogeneity, I exploit the long respondent-level time series available in the ZEW FMS dataset and run separate respondent-level regressions of *expret* on the macroeconomic and financial variables already studied in Section 5.2, as well as the respondents' own assessments of the current and future situation of the German economy from the ZEW FMS dataset. For a better comparability with other variables, I treat categorical ZEW FMS variables as continuous variables. Moreover, to obtain meaningful estimates, I exclude all respondents that have responded less than 30 times in total. For the remaining sample of respondents, the number of responses ranges from 30 to 202, with an average of about 101. In total, I run 409 times 20 regressions, where the former is the number of respondents and the latter is the number of variables.

Table 5 shows the results from these regressions. The results suggest that the determinants of *expret* indeed differ considerably across respondents. Columns 2–5 report the most relevant properties of the distribution of adjusted R^2 across respondents for each of the considered variables. Over all variables, adjusted R^2 statistics range from slightly negative to up to about 72%, suggesting that, for each variable, there exist respondents who do not consider the variable at all, while others assign a very high importance to it when forecasting the DAX. The variable with the highest average adjusted R^2 across respondents is *dax1to0*, which was also the variable that showed the highest overlap with survey and time fixed effects (see Section 5.2). The variable for which the importance varies the most across respondents is *conf*.

The last three columns of Table 5 document the heterogeneity of the correlation coefficients between the variables and *expret* across respondents. The way how the estimated coefficients are distributed between having a positive sign and having a negative sign provides insight into the idiosyncratic variation in *expret*. It is also informative about why some variables have a higher overlap with survey and time fixed effects than others.[8] In this analysis, I do not consider whether the coefficients are statistically significant or not, given that the focus is only the variance of *expret*.[9] The three columns reveal that the degree of heterogeneity of the correlation between each variable and *expret* across respondents is relatively high. One can distinguish between two groups of variables. In the first group, the estimated coefficients show the same sign for the large majority of the respondents. The variable with the highest agreement across respondents is *dax1to0*, for which I measure a negative relationship with *expret* for about 86% of respondents. Other examples are *ep* (26.65% positive vs. 73.35% negative) and *dax3to1* (27.63% positive vs. 72.37% negative). In the second group, the estimated coefficients are more or less evenly balanced be-

[8] The sign alone is of course not sufficient to explain the overlap of a variable with the common time-series variation in *expret*. The degree of overlap also depends on the average magnitude of the coefficients in both groups.

[9] When I consider statistical significance, I find that the correlations with *expret* are statistically insignificant at the 5% level for the majority of respondents and variables. This is also true when I restrict the sample to respondents with at least 100 observations or when I use a 10% threshold instead of the 5% threshold for statistical significance.

Table 5: Idiosyncratic variance of *expret*

Variable	N	Min R^2	Avg. R^2	Std. dev. R^2	Max R^2	Neg. coefficient (%)	Zero coefficient (%)	Pos. coefficient(%)
dax1to0	409	-0.0323	0.1011	0.1134	0.6004	86.06	0.00	13.94
dax12to3	409	-0.0336	0.0547	0.0946	0.5902	61.86	0.00	38.14
dax3to1	409	-0.0323	0.0514	0.0723	0.3528	27.63	0.00	72.37
dp	409	-0.0302	0.0945	0.1241	0.723	35.7	0.00	64.3
ep	409	-0.0346	0.0791	0.1333	0.6498	26.65	0.00	73.35
ipgrowth (shifted)	409	-0.0332	0.0466	0.0811	0.6019	47.68	0.00	52.32
conf (shifted)	409	-0.0339	0.0901	0.1175	0.5486	64.79	0.00	35.21
empl (shifted)	409	-0.0327	0.0504	0.0872	0.4834	57.46	0.00	42.54
oil	409	-0.0357	0.0253	0.0536	0.3054	60.39	0.00	39.61
exchrate	409	-0.0321	0.0468	0.0893	0.477	59.90	0.00	40.10
infl (shifted)	409	-0.0351	0.0374	0.0765	0.5178	54.28	0.00	45.72
expsit	409	-0.0336	0.023	0.0558	0.2963	32.52	0.24	67.24
sit	409	-0.0333	0.0178	0.0483	0.3068	52.57	0.00	47.43
expint_lt	409	-0.0332	0.0216	0.0588	0.4029	36.19	0.49	63.33
expinfl	409	-0.0334	0.0136	0.0518	0.4452	50.12	0.00	49.88
expint_st	409	-0.0336	0.0154	0.0483	0.341	56.72	0.24	43.03

This table illustrates the heterogeneity of the respondents' DAX expectations. The dependent variable *expret* has been orthogonalized with respect to the variable *corrected*. Columns 3–6 report characteristics of the distribution of adj. R^2 statistics from regressions of *expret* on the respective variables listed in the first column. Columns 7–9 show how coefficients on the respective variables are distributed across having a negative sign, being exactly zero or having a positive sign.

tween having a positive and having a negative sign. Examples are *infl* (54.28% positive vs 45.72% negative), which is also the variable with the lowest overlap with survey fixed effects (see Section 5.2) and *ipgrowth* (47.68% positive vs. 52.32% negative), which also ranks very low in Section 5.2.

6 Expected Returns And Economic Conditions

In this section, I explore whether my survey measures of stock return expectations are consistent with macro-financial theory and the empirical evidence based on realized returns. The predominant view in the macro-financial literature is that expected excess returns on stocks vary with economic conditions and are counter-cyclical, i.e. they are higher when economic conditions are bad and vice versa. This view goes back to Fama and French (1989), who, using data for the US economy between 1927 and 1987, document that variables that are considered to be positively correlated with subsequent realized returns, e.g. the dividend–price ratio, were high when economic conditions were bad and low when economic conditions were good.[10] In contradiction to this view, previous studies using US survey data have found that survey measures of stock return expectations are both positively correlated with proxies for expected returns (e.g. Vissing-Jorgensen, 2003; Greenwood and Shleifer, 2014; Amromin and Sharpe, 2014) and economic conditions (e.g. Amromin and Sharpe, 2014). Using a dataset which has not been used to study this question before, covers Germany instead of the US, combines stock market and macroeconomic expectations and features long, respondent-level time-series on a monthly frequency, I present more evidence on the relationship between stock market expectations and economic conditions.

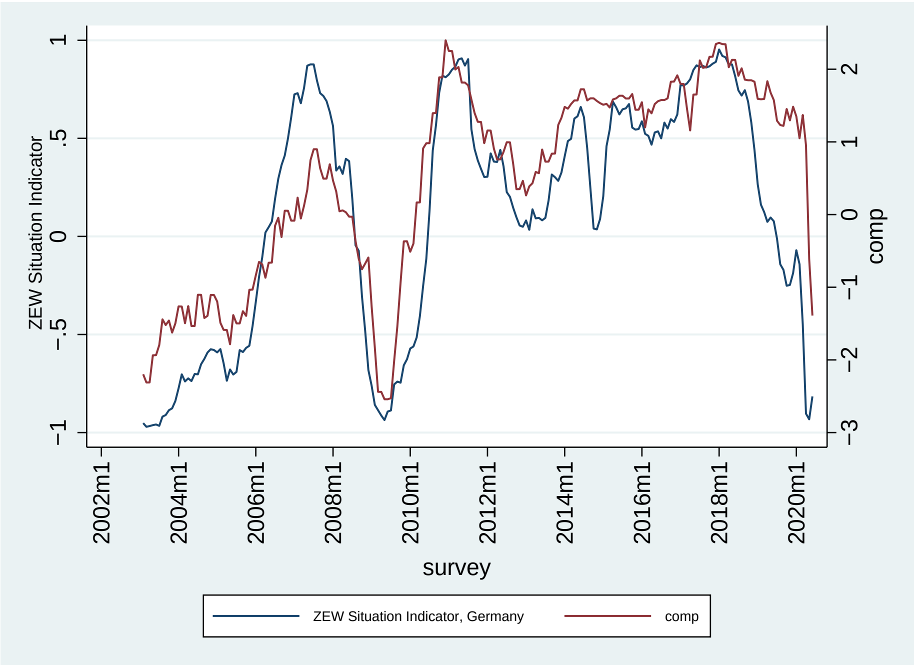
6.1 Measuring Economic Conditions

I use four different variables to measure economic conditions – two direct, economic measures and two indirect, financial measures. The first direct, economic measure is a composite economic indicator for Germany (*comp*). The composite indicator *comp* is the first principal component resulting from a principal component analysis of *ipgrowth*, *empl*, *conf* and *infl* (see Section 4.2). The first component explains about 43% of the variables' total variation and is positively correlated with all variables but *infl*. The use of a composite economic indicator simplifies my analysis because I have to consider only one variable that proxies for economic conditions instead of four. The second direct measure is the respondents' own subjective assessment of the current economic situation in Germany (*sit*, hereafter) from the ZEW FMS dataset. In the survey, the participants of the ZEW FMS are asked whether they think that the current economic situation in Germany is "good", "normal" or "bad". The variable thus already provides the respondents' subjective classifications of survey periods. Figure 5 compares the time-series of *comp* and the ZEW Situation Indicator Germany, where the latter is the difference between the share of respondents who assess the situation as "good" and the share of respondents who assess the situation as "bad". Interestingly, although there are short-term deviations, e.g. in the year

[10] A search of the literature has not yielded more recent empirical results on the relationship between expected excess returns on stocks and economic conditions.

2015, both time-series broadly show the same cyclical pattern. The similarity between the two time-series suggests that the interpretation of the most recent economic data differs systematically between respondents. These differences, however, cancel out when the individual assessments of the current economic situation in Germany are aggregated in this way.

Figure 5: Comparison of the two direct measures of German economic conditions



Note: This figure compares the ZEW Situation Indicator and *comp*. The ZEW Situation Indicator is calculated as the difference between the shares of respondents who assess the current economic situation in Germany as "good" and who assess the current economic situation in Germany as "bad".

The indirect, financial measures are the log dividend–price ratio (*dp*) and the log earnings–price ratio (*ep*) of the CDAX, which are considered to be counter-cyclical in the literature (see e.g. Cochrane, 2017). As Figure 6 shows, this is only partially the case for Germany during the sample period. Figure 6a, which plots the deciles of *dp* against the respective averages of *comp* and the *ZEW Situation Indicator*, reveals that the relationship between *dp* and economic conditions is inversely U-shaped, i.e. both low and high dividend–price ratios occurred when the two direct measures of economic conditions were low. The inverse U-shape has important implications for the relationship between *dp* and *expret*, because if *expret* are indeed counter-cyclical, I will not be able to validate this with *dp*. As Figure 6b illustrates, the relationship between *ep* and my direct measures of economic conditions is less ambiguous. With exception of the first and last *ep*-deciles, the relationship can be described as linear and downward-sloping. The difference between both figures suggests that the payout ratios of the CDAX companies are unusually low or high relative to economic conditions during the sample period. Figure 7, which compares the time-series of *dp* and *ep*, confirms this, as it shows a disconnect between dividends and earnings before and to a smaller extent, during the financial crisis of 2007–2009. This disconnect coincides with the economic boom before the financial crisis, which generates the ambiguous relationship between *dp* and economic conditions. I will therefore choose *ep* over *dp* whenever I have to choose between the two measures.

6.2 Are Expected Returns Counter-cyclical?

If expected returns are counter-cyclical, I should be able to detect positive relationships between my survey measures of stock return expectations and *ep* and, to a lesser extent, *dp* and negative relationships between my survey measures of stock return expectations and *sit* and *comp*. To rule out that the regression results depend on the question format, I consider both available measures of DAX expectations, i.e. the quantitative forecast *expret* and the qualitative forecast *expdir*. An additional advantage of using the qualitative forecast is that the results are robust to large outliers in *expret*. In the analysis, I treat the qualitative forecast *expdir* as a continuous variable, which allows me to use the OLS estimator, facilitating the comparison between the results for both measures of DAX expectations. I also re-define *expdir* as

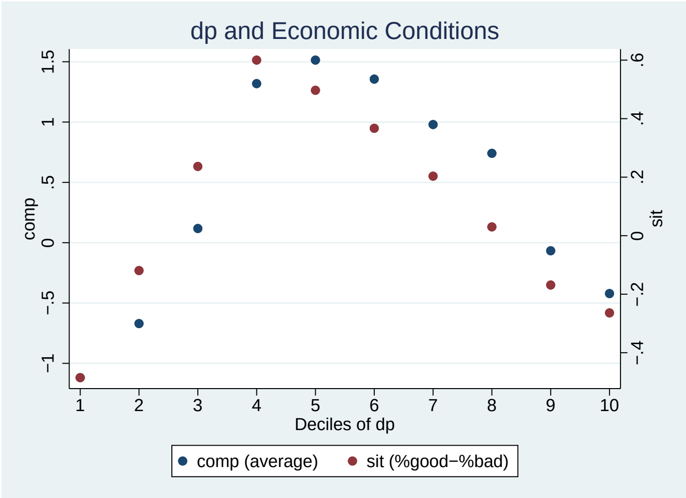
$$\widetilde{expdir}_{i,s,t} = \begin{cases} 1 & \text{if} \quad expdir_{i,s,t} = \text{“increase”} \\ 0 & \text{if} \quad expdir_{i,s,t} = \text{“not change”} \\ -1 & \text{if} \quad expdir_{i,s,t} = \text{“decrease”} \end{cases} . \tag{8}$$

To test my hypothesis of counter-cyclical stock return expectations, I run regressions of *expret* and *expdir* on my four different measures of economic conditions and control variables. As control variables, I include the respondents' own outlooks for the macroeconomy (*expsit*), inflation (*expinfl*), short-term (*expint_st*) and long-term interest rates (*expint_lt*), as well as the prior one-month return of the DAX (*dax1to0*). I control for the respondents' economic outlook, because Amromin and Sharpe (2014) have shown that it matters for stock market expectations. Moreover, Greenwood and Shleifer (2014) and Barberis et al. (2015) document that the recent returns of the equity market are positively correlated with survey stock market expectations.

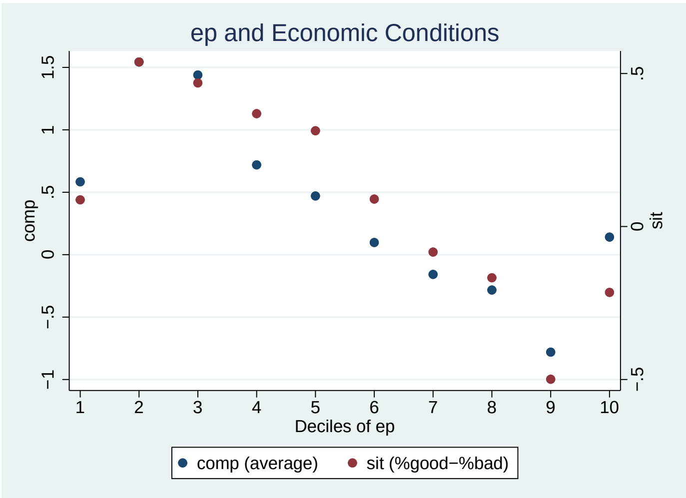
Tables 6 reports the regression results. The results are not unanimously in support of my hypothesis and, in some cases, reverse when I study *expdir* instead of *expret*. Consider, for example, the results for the regressions on *dp* documented in columns 1 and 2. Whereas the coefficient on *dp* in the regression of *expret* is positive (column 1), it is negative in the regression of *expdir* (column 2). The estimates of specifications 3–4 suggest that the contradictory relationship between *dp* and my two survey measures of stock return expectations might at least in part be an implication of the disconnect between dividends and earnings described in the previous section: consistent with my hypothesis, the coefficients on *ep* are both positive and also highly statistically significant.[11] In contrast, the results for *sit*, documented in columns 5 and 6, are both not

[11] If I restrict the sample to the years after 2010 (not shown), i.e. after dividends and earnings have started to move together (see Figure 7), I find a positive but statistically insignificant coefficient in the equivalent of specification 2. This result gives additional support to my side-hypothesis that the contradictory results in specification 1–2 can be attributed to the disconnect between dividends and earnings.

Figure 6: The dividend–price ratio, the earnings–price ratio and economic conditions



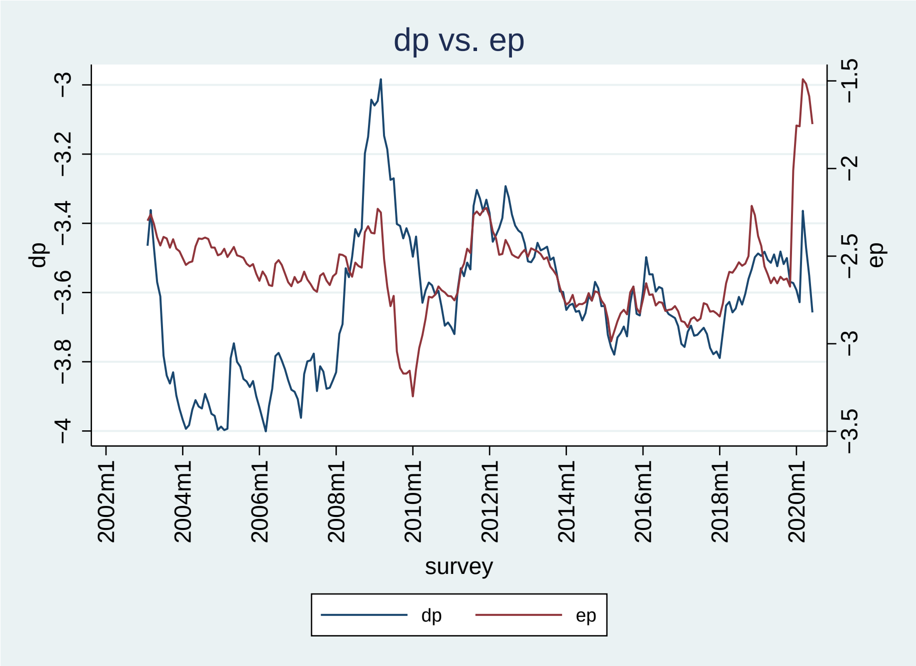
(a) *dp* and economic conditions



(b) *ep* and economic conditions

Note: This figure shows how the two indirect measures of economic conditions, *dp* and *ep*, are related to the two direct measures of economic conditions, *comp* and *sit*. Figure 6a plots the average of *comp* and an aggregated measure of *sit* against *dp*. The aggregated measure of *sit* is calculated as the difference between the shares of responses where *sit* = "good" and *sit* = "bad", respectively, i.e. the ZEW Situation Indicator. Figure 6b plots the average of *comp* and the aggregated measure of *sit* against *ep*.

Figure 7: Development of dividends and earnings of CDAX companies



Note: This figure compares the developments of *dp* and *ep* over time.

in support of my hypothesis. More specifically, I neither find that *expret* is on average higher when respondents assess the current situation as "bad" nor that the respondents are more likely to expect the DAX to increase (*expdir*). On the contrary, the respondents are actually less likely to expect the DAX to increase when they think the current economic situation is "bad" (column 6). Columns 7 and 8 reveal that *comp* shows the same contradictory pattern as *dp* and, to a lesser extent, *sit*. While *comp* is negatively associated with *expret* (column 7), which is in support of my hypothesis, its correlation with *expdir* is statistically insignificant (column 8), which is not in support of my hypothesis. The most supportive for my hypothesis of counter-cyclical stock market expectations are specifications 9 and 10, in which my survey measures of stock return expectations are regressed on all measures of economic conditions but *dp*.[12] Whereas *sit* is still negatively associated with return expectations, the coefficients on *ep* and *comp* are in line with my hypothesis, i.e. return expectations are on average negatively correlated with economic conditions.

To sum up, when I study *expret*, the results are largely in support of my hypothesis that stock market expectations are counter-cyclical. For three out of the four measures of economic conditions studied, return expectations are on average higher when economic conditions are lower. The one exception is the

[12] *dp* is highly correlated with *ep* and, based on my analysis in Section 6.1, an inferior measure of the valuation of the CDAX. The results from a test for multicollinearity suggest that it is unproblematic to include the remaining three measures of economic conditions simultaneously: variance inflation factors range from 1.31 (*ep*) to 3.39 (*sit*).

Table 6: Are expected returns counter-cyclical?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	expret	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*
dp	0.0037***	-0.0425***								
	(0.0012)	(0.0088)								
ep			0.0074***	0.0346***					0.0071***	0.0412***
			(0.0006)	(0.0058)					(0.0006)	(0.0057)
sit = normal					-0.0018	-0.0418***			-0.0078***	-0.0763***
					(0.0012)	(0.0119)			(0.0013)	(0.0127)
sit = bad					0.0026	-0.1070***			-0.0138***	-0.2021***
					(0.0024)	(0.0197)			(0.0023)	(0.0205)
comp							-0.0057***	0.0006	-0.0078***	-0.0449***
							(0.0012)	(0.0099)	(0.0013)	(0.0114)
expsit = not change	-0.0170***	-0.1890***	-0.0167***	-0.1888***	-0.0165***	-0.2110***	-0.0148***	-0.1898***	-0.0167***	-0.2120***
	(0.0015)	(0.0133)	(0.0015)	(0.0133)	(0.0016)	(0.0132)	(0.0015)	(0.0131)	(0.0015)	(0.0129)
expsit = worsen	-0.0476***	-0.5255***	-0.0479***	-0.5481***	-0.0454***	-0.5694***	-0.0434***	-0.5407***	-0.0475***	-0.5815***
	(0.0027)	(0.0247)	(0.0027)	(0.0253)	(0.0028)	(0.0253)	(0.0027)	(0.0254)	(0.0028)	(0.0253)
expinfl = not change	0.0023**	0.0159	0.0002	0.0038	0.0024**	0.0251**	0.0011	0.0146	-0.0001	0.0106
	(0.0012)	(0.0108)	(0.0012)	(0.0107)	(0.0012)	(0.0107)	(0.0012)	(0.0106)	(0.0012)	(0.0105)
expinfl = decrease	0.0056***	0.0362*	0.0019	-0.0025	0.0066***	0.0391**	0.0044**	0.0212	0.0010	0.0068
	(0.0021)	(0.0203)	(0.0020)	(0.0191)	(0.0021)	(0.0190)	(0.0020)	(0.0188)	(0.0020)	(0.0186)
expint_st = not change	0.0035***	0.0042	0.0046***	-0.0136	0.0047***	-0.0031	0.0056***	-0.0121	0.0067***	0.0084
	(0.0013)	(0.0124)	(0.0013)	(0.0128)	(0.0013)	(0.0130)	(0.0013)	(0.0126)	(0.0013)	(0.0123)
expint_st = decrease	0.0144***	0.0424**	0.0151***	-0.0207	0.0179***	0.0195	0.0165***	-0.0046	0.0156***	0.0063
	(0.0022)	(0.0177)	(0.0024)	(0.0183)	(0.0022)	(0.0181)	(0.0022)	(0.0183)	(0.0022)	(0.0180)
expint_lt = not change	-0.0076***	-0.0859***	-0.0077***	-0.0853***	-0.0075***	-0.0860***	-0.0076***	-0.0854***	-0.0075***	-0.0859***
	(0.0012)	(0.0110)	(0.0012)	(0.0110)	(0.0012)	(0.0111)	(0.0012)	(0.0111)	(0.0012)	(0.0110)
expint_lt = decrease	-0.0278***	-0.2354***	-0.0270***	-0.2256***	-0.0282***	-0.2282***	-0.0286***	-0.2312***	-0.0272***	-0.2224***
	(0.0028)	(0.0224)	(0.0027)	(0.0219)	(0.0028)	(0.0223)	(0.0028)	(0.0224)	(0.0027)	(0.0219)
dax1to0	-0.0181***	-0.0479***	-0.0176***	-0.0361***	-0.0188***	-0.0402***	-0.0190***	-0.0412***	-0.0179***	-0.0354***
	(0.0006)	(0.0045)	(0.0006)	(0.0046)	(0.0006)	(0.0046)	(0.0006)	(0.0046)	(0.0006)	(0.0047)
corrected	-0.0149***		-0.0141***		-0.0151***		-0.0139***		-0.0137***	
	(0.0033)		(0.0033)		(0.0033)		(0.0033)		(0.0033)	
Constant	0.0425***	0.6129***	0.0431***	0.6473***	0.0405***	0.6785***	0.0399***	0.6352***	0.0486***	0.7254***
	(0.0016)	(0.0145)	(0.0016)	(0.0148)	(0.0020)	(0.0164)	(0.0016)	(0.0143)	(0.0020)	(0.0177)
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44,311	44,311	44,311	44,311	44,311	44,311	44,311	44,311	44311	44311
R^2	0.1107	0.0893	0.1178	0.0888	0.1095	0.0894	0.1131	0.0864	0.1225	0.0948
Adj. R^2	0.1105	0.0891	0.1176	0.0886	0.1092	0.0891	0.1128	0.0862	0.1222	0.0946

Note: This table documents the results of regressions of *expret* and *expdir* on measures of economic conditions and control variables. All dependent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

respondents' own assessment of current economic conditions, which is the only subjective measure of economic conditions considered in the analysis. In contrast, when I study *expdir*, I find that economic conditions are either unrelated or even positively associated with stock market expectations. The discrepancies between the results for *expret* and *expdir*, however, vanish when all measures of economic conditions are considered together.

The results for *expret* and to a limited extent for *expdir*, differ from central findings of the previous literature on survey measures of stock market expectations. In contrast to Greenwood and Shleifer (2014) and Amromin and Sharpe (2014), I find that the valuation of the stock market, proxied by *dp* and *ep*, is on average positively associated with the DAX return expectations of the survey respondents. Moreover, in contrast to Greenwood and Shleifer (2014), Amromin and Sharpe (2014) and Barberis et al. (2015), I do not find evidence for an extrapolation of past returns. Both survey measures of DAX return expectations are negatively correlated with the DAX return over the previous month in all specifications.

Why are the results for *expret* and *expdir* qualitatively different in some cases? There are three possible explanations. First, the qualitative differences might arise because the respondents give answers to the question asking for a point forecast of the DAX (i.e. *expret*) that contradict their answers to the question asking for a directional forecast of the DAX (i.e. *expret*). Second, outliers in *expret* might impact the estimates such that the direction of the measured relationship between economic conditions and *expret* differs from that of the respective relationship with *expdir*. Finally, the qualitative differences might be the result of the different scales of the two survey measures of DAX expectations, i.e. metric for *expret* and ordinal for *expdir*.

I first turn to inconsistent answers. Table 7 reports features of the distributions of *expret* conditional on *expdir*. These statistics suggest that the respondents' quantitative forecasts are largely consistent with their respective qualitative forecasts.[13] More specifically, *expret* is on average positive, close to zero and negative, if respondents answer "increase", "not change" and "decrease", respectively. However, there are also a few inconsistent answers. For example, the smallest value for *expret* in the category "increase" is -91%, which, in addition to having the "wrong" sign, is also very large in magnitude. To quantify the extent to which inconsistent answers are responsible for the differences between the results for the qualitative and quantitative forecasts, I drop all inconsistent answers and re-run my regressions of both survey measures of DAX return expectations on my measures of economic conditions. I also drop all observations in the category "not change", given that there are no observations for which *expret* is exactly 0. Table 8 reports the results from these regressions. Although the exclusion of inconsistent answers produces stronger results, i.e. coefficients of variables that are hypothesized to be positively associated with DAX expectations become larger and vice versa, it does not solve the problem of contradicting results in regressions of *expret* vs. *expdir*. In particular, the coefficient on *dp* is still positive in specification (1) and negative in specification (2) and the coefficient on *comp* is still negative in specification (7) and positive but statistically insignificant in specification (8).

[13]The order of the questions in the questionnaire is the following: First qualitative, then quantitative.

Table 7: Distributions of *expret* conditional on *expdir*

expdir	Min	p10	p25	p50	Mean	p75	p90	Max
"increase"	-91.07	1.47	3.33	5.78	7.01	9.32	14.20	80.97
"not change"	41.80	-3.79	-1.47	0.22	0.29	2.05	4.43	41.30
"decrease"	-87.64	-16.16	-11.07	-7.32	-8.35	-4.41	-2.18	47.47

Note: This table reports characteristics of the conditional distributions of *expret* (in percent), conditional on *expdir*. The labels p10, p25, p50, p75 and p90 refer to the 10th, 25th, 50th, 75th and 90th percentile of the overall distribution of *expret*.

I next turn to the role of outliers in *expret*. To quantify the effect that outliers have on my estimates, I re-run all regression with a winsorized version of *expret*. The winsorization is done by replacing the 5% smallest and the 5% largest values of *expret* by the variable's 5th and 95th percentile, respectively, where both percentiles are calculated from the distributions of *expret* specific to each survey wave. Table 9 documents the regression results. Again, the coefficient on *dp* is positive in specification (1) but negative in specification (2) and the coefficient on *comp* is negative in specification (7) and positive but statistically insignificant in specification (8). Thus, outliers in *expret* are not the reason for why the results are qualitatively different.

Having ruled out both inconsistent responses and outliers as the causes of the qualitative differences between the results for *expret* and *expdir*, the only remaining explanation is that the differences are due to the different scales of the two variables. Given that the respondents can only choose between "increase", "not change" and "decrease" when answering the question asking for a directional DAX forecast and that they are able to provide an exact forecast in the question asking for a point forecast, *expdir* co-varies less with perceived economic conditions than *expret* by construction.

6.3 Expected Returns, Economic Conditions And The Respondents' Personal Characteristics

I now turn to the relationship between the respondents' personal characteristics and their DAX expectations, which I have ignored so far. As summarized in the literature overview, the empirical evidence suggests that the individual characteristics of respondents matter for their expectations. The focus of my analysis is whether the respondents' characteristics affect the relationships between *expret* and economic conditions and, if this is the case, whether these characteristics are associated with pro-cyclical or counter-cyclical DAX expectations. Differences in personal characteristics might thus explain why the correlations between economic conditions and *expret* vary extensively across respondents (see Section 5.3).

I study how the relationship between *expret* and economic conditions depends on the following characteristics: age and age cohort, indicators of the levels of expertise in conducting DAX forecasts and main occupation.[14] I restrict my analysis to *expret*, because the results presented in the previous section

[14] I do not consider the career entry year or the number of years of working experience, because these variables are highly correlated with age – the coefficient of correlation is 0.9 – and thus imply very similar results.

Table 8: The effect of inconsistent answers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	expret	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*
dp	0.0039**	-0.0429***								
	(0.0015)	(0.0112)								
ep			0.0090***	0.0429***					0.0085***	0.0514***
			(0.0008)	(0.0078)					(0.0009)	(0.0076)
sit = normal					-0.0015	-0.0320**			-0.0085***	-0.0640***
					(0.0016)	(0.0147)			(0.0017)	(0.0162)
sit = bad					0.0036	-0.1223***			-0.0154***	-0.2072***
					(0.0031)	(0.0247)			(0.0030)	(0.0272)
comp							-0.0067***	0.0103	-0.0090***	-0.0358**
							(0.0015)	(0.0122)	(0.0016)	(0.0141)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	30,442	30,442	30,442	30,442	30,442	30,442	30,442	30,442	30,442	30,442
R^2	0.1272	0.1058	0.1366	0.1064	0.1261	0.1069	0.1304	0.1033	0.1417	0.1129
Adj. R^2	0.1269	0.1055	0.1363	0.1061	0.1257	0.1066	0.1302	0.1030	0.1414	0.1125

Note: This table documents the results of regressions of *expret* and *expdir* on measures of economic conditions and control variables. All observations for which respondents have provided quantitative and qualitative DAX forecasts which are inconsistent with each other were dropped from the regression. Observations were classified as inconsistent if *expret* < 0 and *expdir* = "increase" or *expret* > 0 and *expdir* = "decrease". Also, all observations, for which *expdir* = "not change", were dropped from the regression. Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

Table 9: The effect of outliers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	expret	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*	*expret*	*expdir*
dp	0.0030***	-0.0425***								
	(0.0010)	(0.0088)								
ep			0.0072***	0.0346***					0.0068***	0.0412***
			(0.0006)	(0.0058)					(0.0006)	(0.0057)
sit = normal					-0.0013	-0.0418***			-0.0069***	-0.0763***
					(0.0010)	(0.0119)			(0.0011)	(0.0127)
sit = bad					0.0037*	-0.1070***			-0.0118***	-0.2020***
					(0.0021)	(0.0197)			(0.0019)	(0.0205)
comp							-0.0057***	0.0006	-0.0073***	-0.0449***
							(0.0010)	(0.0099)	(0.0011)	(0.0114)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44,311	44,311	44,311	44,311	44,311	44,311	44,311	44,311	44,311	44,311
R^2	0.1254	0.0893	0.1347	0.0888	0.1247	0.0894	0.1292	0.0864	0.1400	0.0948
Adj. R^2	0.1252	0.0891	0.1345	0.0886	0.1245	0.0891	0.1289	0.0862	0.1398	0.0946

Note: This table documents the results of regressions of *expret* and *expdir* on measures of economic conditions and control variables. The variable *expret* was winsorized by replacing the 5% smallest and the 5% largest values of *expret* by its 5th and 95th percentiles, respectively. Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

suggest that it is a more precise measure of the respondents' DAX expectations than *expdir*. Moreover, because these specifications did not show contradictory results and to save space, I report only the results of the regressions in which all measures of economic conditions are included simultaneously (the equivalents of specifications 9 and 10 in Table 6).

Table 10 documents how the relationships between *expret* and *ep*, *comp* and *sit* vary with the respondents' age and their age cohorts. For a reduced complexity of the analysis, I have divided the group of respondents who provided their birth date into four groups along the distribution of the respondents' birth years. The breakpoints for the four age cohorts are the three quartiles of the distribution of birth years. The distribution of birth years ranges from 1938 to 1990 and the three quartiles are 1958, 1963 and 1969. Column 2 (specification 1) of Table 10 reveals that the results for the relationships between *expret* and the three measures of economic conditions documented in Table 6 also hold in the sub-sample, for which the birth years of the respondents are available. Interestingly, when I include *age* into the model (column 3), which is negatively associated with *expret*, the relationships between *expret* and *ep* and *sit* remain largely unchanged, whereas the coefficient on *comp* loses its statistical significance. The most likely explanation for the latter finding is that *comp* and age are spuriously correlated in the sample, i.e. the upward movement of the distribution of age (see Figure 2d) during the sample period happens to coincide with an upward trend in economic conditions as measured by *comp* (see Figure 5).[15] Column 4 (specification 3) reports the result of the regression, in which I interact my three measures of economic conditions with *age*. The estimated model suggests that the relationship between *expret* and *ep* depends on age, while those of *comp* and *sit* do not. More specifically, the coefficient on *ep* decreases when age increases. As can be seen from Figure 8, which plots the coefficient on *ep* against age, the estimated model implies that the association between *ep* and *expret* is positive (i.e. counter-cyclical) if age is below 69 and statistically insignificant if age is 69 or higher (judged by a 95% confidence interval). Given that only a small minority of the financial market experts in the sample has reached the age of 69, this threshold holds no economic significance. Finally, when I interact the three measures of economic conditions with the respondents' age cohorts (column 5), I do not find any differences across age cohorts.

I next explore whether the relationships between *expret* and *ep*, *comp* and *sit* depend on the respondents' level of expertise in conducting DAX forecasts. There are four self-reported measures of expertise available to me. These are the respondents' own assessments of their levels of expertise in the areas of stock forecasts in general, in conducting DAX point and interval forecasts and in assessing the fundamental value of the DAX, as well as the respondents' professional experience in conducting DAX forecasts, i.e. whether and how often respondents have conducted DAX forecasts outside of the context of the ZEW FMS before.

A natural hypothesis is that a high level of expertise is associated with counter-cyclical DAX expectations. Experts should know that subsequent realized returns are on average higher when economic conditions are bad, given that this is well documented in the literature. Moreover, I also test whether taking

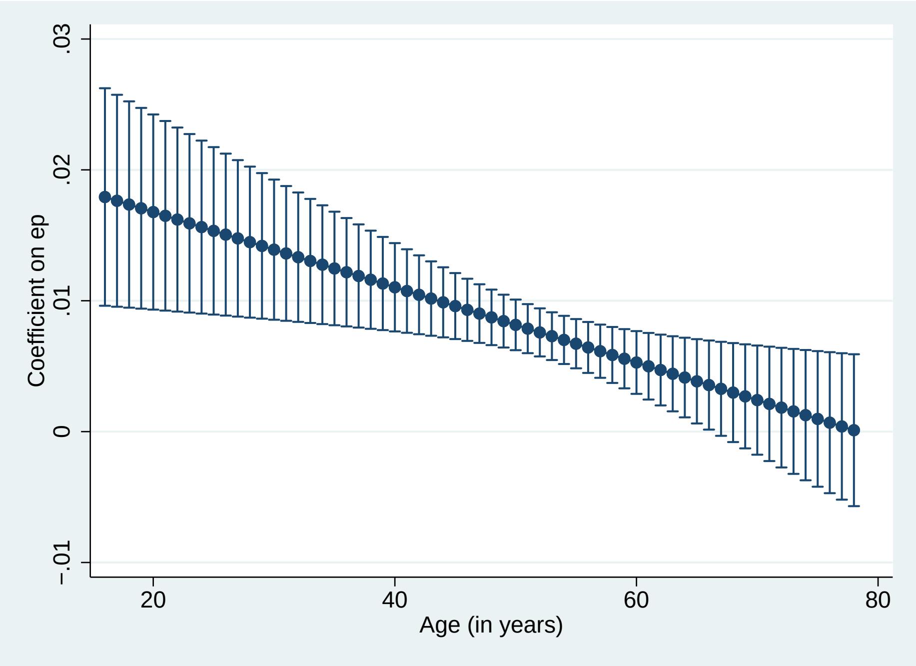
[15] While *comp* seems to follow an upward trend during the sample period, its components are all stationary variables. Hence, detrending *comp* to remove the spurious correlation between it and age would be inappropriate.

Table 10: Expected returns, economic conditions and age

	(1)	(2)	(3)	(4)
	expret	*expret*	*expret*	*expret*
ep	0.0070***	0.0080***	0.0225***	0.0088***
	(0.0009)	(0.0009)	(0.0060)	(0.0022)
ep × age			-0.0003**	
			(0.0001)	
ep × cohort = 2				-0.0004
				(0.0029)
ep × cohort = 3				-0.0021
				(0.0031)
ep × cohort = 4				-0.0044
				(0.0029)
comp	-0.0075***	0.0011	-0.0112	-0.0066*
	(0.0015)	(0.0018)	(0.0083)	(0.0033)
comp × Age			0.0003	
			(0.0002)	
comp × Cohort = 2				-0.0015
				(0.0042)
comp × Cohort = 3				-0.0013
				(0.0050)
comp × Cohort = 4				-0.0005
				(0.0038)
sit = good × Cohort = 2				0.0112
				(0.0074)
sit = good × Cohort = 3				0.0116
				(0.0076)
sit = good × Cohort = 4				0.0081
				(0.0083)
sit = normal	-0.0067***	-0.0050***	-0.0160*	-0.0036
	(0.0016)	(0.0016)	(0.0085)	(0.0025)
sit = normal × Age			0.0002	
			(0.0002)	
sit = normal × Cohort = 2				0.0084
				(0.0067)
sit = normal × Cohort = 3				0.0052
				(0.0070)
sit = normal × Cohort = 4				0.0055
				(0.0072)
sit = bad	-0.0101***	-0.0060**	-0.0138	-0.0023
	(0.0028)	(0.0029)	(0.0163)	(0.0057)
sit = bad × age			0.0002	
			(0.0003)	
Age		-0.0022***	-0.0021***	
		(0.0004)	(0.0004)	
Constant	0.0475***	0.1402***	0.1366***	0.0397***
	(0.0024)	(0.0176)	(0.0199)	(0.0055)
Controls	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes
N	24,611	24,611	24,611	24,611
R^2	0.1275	0.1341	0.1361	0.1285
Adj. R^2	0.1270	0.1336	0.1354	0.1276

Note: This table reports the results of regressions of *expret* on measures of economic conditions, the respondents' age and their age cohort. Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

Figure 8: The coefficient on *ep* conditional on age



Note: This figure shows how the measured relationship between *expret* and *ep* depends on the respondents' age. The plot is based on the regression results documented in Table 10 and shows the estimates of the coefficient on *ep* and the accompanying 95% confidence intervals.

an interest in the results of the ZEW FMS on stock markets in general matters for the respondents' DAX expectations conditional on economic conditions. Respondents who take interest in stock market forecasts might, for example, put more effort in making their own forecasts than those who are not.

Tables 11 and 12 report the regression results. The results on the interactions between my measures of expertise and my measures of economic conditions do not suggest that a higher level of expertise is associated with more counter-cyclical DAX expectations: the interactions between expertise and economic conditions in specifications (1)–(3) reported in Table 11 as well as the interactions between professional forecasting activities and economic conditions reported in specification (2) in Table 12 all are statistically not significant. When I differentiate by whether a respondent takes interest in the results of the ZEW FMS on stock markets in general, I find that the coefficient on *comp* is only statistically significant (i.e. counter-cyclical) if the respondents report that they are interested.

Lastly, I differentiate by the respondents' self-reported main occupation. Table 13 reports the regression results for each of the ten categories. While not all are statistically significant, the coefficients on *ep* and *comp* across all main occupations have the same sign and also the same sign as the respective coefficients from the main regression reported in Table 6. With two exceptions, i.e. specifications 5 and 9, this is also the case for *sit*. Financial market experts with different main occupations thus mainly seem to differ with respect to whether they consider a given measure of economic conditions when forecasting the DAX or not. The results suggest that, of all the characteristics explored in this section, main occupation is the best differentiator when it comes to the relationship between DAX return expectations and measures of economic conditions.

7 Evaluating Forecasting Performance

In this section, I evaluate the forecast performance of the respondents to the ZEW FMS. I am interested in two characteristics of the respondents' DAX forecasts. First, I explore whether their forecasts are predictive for subsequent realized returns and, if this is the case, whether the forecasts are positively or negatively correlated with them. The sign of the correlation is of particular interest, given that Greenwood and Shleifer (2014) find that their survey measures of expected stock returns are negatively correlated with subsequent realized returns. They explain this puzzling finding with their result that proxies for expected excess stock returns and their survey measures of expected returns are negatively correlated. As I document relationships between *expret* and proxies for expected stock returns (e.g. the dividend–price ratio) that differ from those reported in Greenwood and Shleifer (2014), I expect to find that my survey measures of stock return expectations are positively correlated with subsequent realized returns. Second, I test whether the respondents' DAX forecasts are more accurate than the historical average realized return, the latter being an often used benchmark which stock market forecasts are compared to in the literature (see e.g. Welch and Goyal, 2008).

I begin by studying the predictive power of aggregated versions of my two survey measures of DAX return expectations. These are an equally-weighted average of *expret* and the bull-bear spread, the latter being the difference be-

Table 11: Expected returns, economic conditions, stock market expertise and taking interest in ZEW FMS results on stock markets

Dependent variable: *expret*	(1) Expertise: stocks	(2) Expertise: quantitative DAX forecasts	(3) Expertise: current valuation DAX	(4) Interest: stocks
ep	0.0095**	0.0076***	0.0099***	0.0077***
	(0.0041)	(0.0021)	(0.0029)	(0.0025)
ep × expertise = medium	-0.0014	0.0002	-0.0048	
	(0.0045)	(0.0026)	(0.0034)	
ep × expertise = high	-0.0041	-0.0042	-0.0029	
	(0.0044)	(0.0030)	(0.0033)	
ep × interested = yes				-0.0012
				(0.0028)
comp	-0.0052*	-0.0073**	-0.0064	-0.0039
	(0.0028)	(0.0031)	(0.0046)	(0.0024)
comp × expertise = medium	-0.0031	-0.0028	0.0000	
	(0.0040)	(0.0037)	(0.0051)	
comp × expertise = high	-0.0048	-0.0050	-0.0080	
	(0.0034)	(0.0054)	(0.0054)	
comp × interested = yes				-0.0076**
				(0.0031)
sit = good × expertise = medium	0.0115	0.0028	-0.0124	
	(0.0099)	(0.0079)	(0.0089)	
sit = good × expertise = high	-0.0084	-0.0115	-0.0087	
	(0.0085)	(0.0101)	(0.0087)	
sit = good × × interested = yes				-0.0085
				(0.0105)
sit = normal	-0.0085	-0.0142***	-0.0141*	-0.0097***
	(0.0065)	(0.0050)	(0.0079)	(0.0026)
sit = normal × expertise = medium	0.0087	0.0102	-0.0040	
	(0.0077)	(0.0071)	(0.0076)	
sit = normal × expertise = high	-0.0064	-0.0020	-0.0037	
	(0.0060)	(0.0099)	(0.0079)	
sit = normal × interested = yes				-0.0076
				(0.0093)
sit = bad	-0.0129*	-0.0138**	-0.0218***	-0.0181*
	(0.0075)	(0.0060)	(0.0069)	(0.0096)
Constant	0.0504***	0.0523***	0.0594***	0.0556***
	(0.0078)	(0.0066)	(0.0082)	(0.0086)
Controls	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes
N	16,342	16,536	16,536	14,517
R^2	0.1400	0.1423	0.1433	0.1405
Adj. R^2	0.1388	0.1412	0.1421	0.1394

Note: This table reports the results of regressions of *expret* on measures of economic conditions, three indicators of the respondents' level of expertise in conducting stock market forecasts and an indicator of whether the respondents take interest in the results of the ZEW FMS on stock markets. The labels "stocks", "quantitative DAX forecasts" and "current valuation DAX" refer to the respondents' levels of expertise in the areas of stock market forecasting in general, of making quantitative DAX forecasts and of DAX valuation, respectively. The expertise variables can take the values "low", "medium" and "high". The variable that indicates whether the respondents take interest in the results on stock markets can take the values "yes" or "no". Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

Table 12: Expected returns, economic conditions and professional experience in conducting DAX forecasts

	(1)	(2)
	expret	*expret*
ep	0.0067***	0.0067***
	(0.0008)	(0.0014)
ep × DAX forecasts = sometimes		0.0034
		(0.0023)
ep × DAX forecasts = never		-0.0013
		(0.0018)
comp	-0.0082***	-0.0101***
	(0.0014)	(0.0024)
comp × DAX forecasts = sometimes		0.0055
		(0.0039)
comp × DAX forecasts = never		0.0022
		(0.0031)
Economic situation Germany = good × DAX forecasts = sometimes		0.0040
		(0.0063)
Economic situation Germany = good × DAX forecasts = never		-0.0035
		(0.0058)
Economic situation Germany = normal	-0.0078***	-0.0078***
	(0.0015)	(0.0020)
Economic situation Germany = normal × DAX forecasts = sometimes		0.0025
		(0.0067)
Economic situation Germany = normal × DAX forecasts = never		-0.0034
		(0.0052)
Economic situation Germany = bad	-0.0136***	-0.0143***
	(0.0026)	(0.0041)
Constant	0.0482***	0.0493***
	(0.0026)	(0.0040)
Controls	Yes	Yes
Person FE	Yes	Yes
N	30,765	30,765
R^2	0.1299	0.1312
Adj. R^2	0.1295	0.1306

Note: This table reports the results of regressions of *expret* on measures of economic conditions and an indicator of the respondents' professional experience in conducting DAX forecasts, *DAX forecasts*. The variable *DAX forecasts* can take the values "regular", "sometimes" and "never". Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

Table 13: Expected returns, economic conditions and the respondents' main occupation

Dependent variable: *expret*	(1) Economic research	(2) Trading	(3) Financing	(4) Management	(5) Security research	(6) Fund/portfolio management	(7) Investment advice	(8) Wealth management	(9) Risk management	(10) Other
ep	0.0070***	0.0044**	0.0073***	0.0032	0.0055**	0.0056***	0.0045***	0.0056***	0.0057**	0.0073**
	(0.0017)	(0.0019)	(0.0018)	(0.0019)	(0.0024)	(0.0013)	(0.0015)	(0.0015)	(0.0023)	(0.0027)
comp	-0.0069	-0.0059	-0.0098**	-0.0070**	-0.0012	-0.0064*	-0.0065**	-0.0069	-0.0051*	-0.0112*
	(0.0050)	(0.0050)	(0.0037)	(0.0028)	(0.0058)	(0.0033)	(0.0031)	(0.0042)	(0.0024)	(0.0057)
sit = normal	-0.0038	-0.0057	-0.0078*	-0.0054	-0.0023	-0.0036	-0.0134***	-0.0050	-0.0080**	-0.0052
	(0.0034)	(0.0040)	(0.0044)	(0.0042)	(0.0048)	(0.0033)	(0.0037)	(0.0042)	(0.0038)	(0.0063)
sit = bad	-0.0104	-0.0154*	-0.0219***	-0.0098	0.0021	-0.0067	-0.0061	-0.0095	0.0035	-0.0133
	(0.0065)	(0.0082)	(0.0065)	(0.0074)	(0.0050)	(0.0046)	(0.0086)	(0.0061)	(0.0059)	(0.0128)
Constant	0.0453***	0.0489***	0.0498***	0.0431***	0.0482***	0.0448***	0.0464***	0.0504***	0.0373***	0.0633***
	(0.0063)	(0.0083)	(0.0066)	(0.0083)	(0.0057)	(0.0043)	(0.0049)	(0.0056)	(0.0057)	(0.0148)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,309	2,636	3,236	5,047	3,206	7,775	4,403	4,867	2,549	2,652
R^2	0.1429	0.1073	0.1714	0.1046	0.1286	0.1484	0.1523	0.1633	0.2110	0.1054
Adj. R^2	0.1401	0.1026	0.1678	0.1021	0.1247	0.1468	0.1496	0.1609	0.2067	0.1007

Note: This table reports the results of regressions of *expret* on measures of economic conditions, conditional on the respondents' main occupation. Control variables are *expsit*, *expinfl*, *expint_st*, *expint_lt* and *dax1to0*. All independent variables were standardized. Standard errors were clustered on the respondent-level and are reported in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively.

tween the shares of respondents who expect the DAX to increase and decrease, respectively, over the course of the next six months. I then explore whether there are differences in forecast performance between subgroups formed by the various personal characteristics available to me.

7.1 Aggregated Forecasts

To evaluate the predictive power of the aggregated return forecasts, I run separate regressions of an aggregated measure of realized six-month DAX returns on the average of *expret* and on the bull-bear spread. The regression model is

$$\bar{r}_s^{DAX,Q} = \alpha + \beta f_s^Q + \epsilon_s, \tag{9}$$

where $\bar{r}_s^{DAX,Q}$ is the aggregated measure of realized six-month DAX returns for survey wave s and f_s^Q is either the bull-bear spread ($bullbear_s$) or the average of *expret* ($quant_s$) in survey wave s. As the index $Q \in \{bullbear, quant\}$ indicates, $\bar{r}_s^{DAX,Q}$ depends on whether I study *bullbear* or *quant*. More specifically, I define the aggregated realized return $\bar{r}_s^{DAX,Q}$ as

$$\bar{r}_s^{DAX,Q} = (N_s)^{-1} \sum_{i=1}^{N_s} D_{i,s,t}^{Q} r_{s,t;t+6m}^{DAX}, \tag{10}$$

where N_s is the number of respondents in survey wave s, i indexes the respondents of survey wave s, $r_{s,t;t+6m}^{DAX}$ is the realized six-month DAX return associated with a DAX forecast made on survey day t during survey wave s and $D_{i,s,t}^{Q}$ is an indicator variable which takes the value of 1 if respondent i provided a forecast for forecast Q on survey day t during survey wave s and 0 otherwise. By only considering the realized returns specific to respondents who actually provided forecasts, I ensure that the aggregated measure of realized returns better aligns with the aggregated forecasts. The aggregated forecasts are calculated as

$$bullbear_s = (N_s)^{-1} \sum_{i=1}^{N_s} D_{i,s,t}^{Q} \widetilde{expdir}_{i,s,t} \tag{11}$$

and

$$quant_s = (N_s)^{-1} \sum_{i=1}^{N_s} D_{i,s,t}^{Q} expret_{i,s,t}, \tag{12}$$

respectively, where $\widetilde{expdir}_{i,s,t}$ is the continuous version of the directional DAX forecast defined in Equation (8).

Table 14 reports the regression results. I first regress $\bar{r}^{DAX,Q}$ on *bullbear*. I do this both for the whole time-series starting in 1991 (specification (1)), as well as the period starting in 2003 (specification (2)), which is the period for which *quant* is available. Realized six-month DAX returns are available until survey wave February 2020. To account for heteroskedasticity and autocorrelation in the error term, I use the Newey and West (1987) estimator to estimate standard errors. Because the forecast horizon is six months, I follow Greenwood and Shleifer (2014) and set the maximum lag in the Newey-West estimation to six. The results in columns 2 and 3 (specifications (1) and (2)) suggest that *bullbear* is not predictive for realized returns. For both models,

Table 14: Evaluating predictive power

	(1)	(2)	(3)
	$\bar{r}^{DAX,bullbear}$	$\bar{r}^{DAX,bullbear}$	$\bar{r}^{DAX,quant}$
bullbear	-0.0406	0.1618	
	(0.1161)	(0.1049)	
quant			1.3422**
			(0.6085)
Constant	0.0650	-0.0116	0.0182
	(0.0503)	(0.0489)	(0.0212)
N	338	205	205
R^2	0.0017	0.0340	0.0676
Adj. R^2	-0.0012	0.0292	0.0630

Note: This table documents the results of regressions of average realized six-month DAX returns, $\bar{r}_s^{DAX,Q}$, $Q \in \{bullbear, quant\}$, on the two aggregated DAX forecasts *bullbear* and *quant*. Specification (1) is estimated on the full sample, i.e. December 1991–June 2020, whereas specifications (2) and (3) are estimated on the sample, for which *quant* is available, i.e. February 2003–June 2020. Newey–West standard errors in parantheses. ***, ** and * refer to significance levels of 1%, 5% and 10%, respectively. R^2 and adjusted R^2 statistics are taken from separate OLS regressions of $\bar{r}_s^{DAX,Q}$ on *bullbear* and *quant*.

the null hypothesis that the coefficients on *bullbear* are 0 cannot be rejected at a reasonable significance level. The variable also does not explain much of the variation in returns, whereby R^2 seems to depend strongly on the sample period. More specifically, while *bullbear* explains only about 0.17% (-0.12% adjusted) of the variation in realized returns in the full sample, it explains about to 3.40% (2.92% adjusted) in the sample starting in 2003. In contrast, I find strong evidence that the variable *quant* has predictive power for realized returns. As documented in the last column of Table 14 (specification (3)), the coefficient on *quant* is positive, larger than 1 and has a p-value of 2.9% (not reported). Moreover, the variation in *quant* accounts for about 6.76% of the variation in realized returns, which is nearly two times the share explained by *bullbear* in column 3 (specification (2)). The results remain qualitatively unchanged when I consider excess returns, i.e. when I subtract the risk-free rate at the time of the forecasts from realized returns and the quantitative DAX forecasts (not reported). This result contradicts the finding of Greenwood and Shleifer (2014) that survey measures of expected return are negatively correlated with actual returns.

Having shown that *quant* is predictive for realized returns, I next compare the forecast accuracy of the variable to that of the historical average realized return. I use end-of-month values of the DAX index to calculate the historical average six-month DAX return prevailing in survey wave s, which began in month m as

$$\bar{r}_s^{DAX} = (m-1)^{-1} \sum_{i=1}^{m-1} r_{i-6;i}^{DAX}, \tag{13}$$

where i indexes months since the start of the calculation of the DAX index, which is December 1964 in my data source Eikon Datastream. Since I use all available six-months returns since December 1964, $\bar{r}_s^{DAX}$ changes only moderately between survey waves. Between December 1991 and February 2020, the historical average ranges from 4.58% to 6.04%, with a mean of 5.36% and a standard deviation of 0.27%. To compare accuracies of the two forecasts, I follow the approach outlined in Diebold and Mariano (1995), Harvey et al. (1997) and Rapach and Zhou (2013) and test whether the forecast errors made by the respondents of the ZEW FMS are smaller than those for DAX forecasts made with the historical average. My null hypothesis thus is

$$H_0 : MSFE^{histavg} \leq MSFE^{quant},$$

where $MSFE^{FE}$ is the mean squared forecast error of forecast $FE \in \{quant, histavg\}$. To carry out this test, I calculate the modified Diebold–Mariano test statistic (Equation (8) in Harvey et al., 1997, p. 283). In my case, the parameters n (the number of periods) and h (the forecast horizon), are 205 and 6, respectively. The data implies a test statistic of 0.4644. According to Harvey et al. (1997), the modified Diebold–Mariano test statistic follows a Student-t distribution. Using the cumulative distribution function of the Student-t distribution, I arrive at a p-value of 32.14%. The null hypothesis thus cannot be rejected, suggesting that the forecast accuracy of *quant* is not higher than that of the historical average.

7.2 Cross-sectional Differences In Forecast Accuracy

Having shown that the aggregate quantitative DAX forecast has predictive power for realized six-months DAX returns, I next explore whether there are differences in forecast accuracy between subgroups of the ZEW FMS panel formed by the personal characteristics available to me. To ensure that a forecaster always belongs to exactly one group in the comparisons, I only distinguish by time-invariant characteristics. I distinguish by age cohort, professional experience in conducting DAX forecasts, the self-assessed level of expertise in conducting DAX forecasts, whether the respondents take interest in the ZEW FMS results on stock markets in general and the respondents' main occupation. This allows me to relate potential differences in forecast accuracy to the respective differences in the documented relationships between economic conditions and DAX return expectations documented in Section 6.3. For the comparisons of forecast accuracy, I use the same approach as in Section 7.1, i.e. I calculate the subgroup-specific averages of realized returns and *expret* as in Equations (10) and (12) and use the adjusted Diebold–Mariano test statistic to evaluate whether one forecast is better than another. Given that the availability of the personal characteristics is concentrated at the end of the sample period (see Section 3.1), I might face a problem with small group sizes, implying that the average DAX return forecasts of some groups are very volatile. To alleviate this problem, I restrict the sample used to evaluate differences in forecast accuracy to the years 2012–2020. In this subsample, the personal characteristics of interest are available for at least 50% of the panel members and the minimum size per group and survey wave is not smaller than 15 for the large majority of groups.

Table 15: Differences in forecast accuracy: professional experience in conducting DAX forecasts

B → A ↓	DAX forecasts: regular	DAX forecasts: sometimes	DAX forecasts: never
DAX forecasts: regular	-	-1.4864 (92.99%)	0.2544 (39.99%)
DAX forecasts: sometimes		-	0.6928 (24.50%)
DAX forecasts: never			-

Note: This table reports the results of pairwise comparisons of the mean squared forecast errors (MSFE) made by three subsets of the ZEW FMS panel: respondents who regulary conduct DAX forecasts outside of the ZEW FMS, respondents who sometimes conduct DAX forecasts outside of the ZEW FMS and respondents for never conduct DAX forecasts outside of the ZEW FMS. The table reports the adj. Diebold–Mariano statistic and the corresponding p-value for each comparison. The null hypotheses of the tests are given by $H_0 : MSFE^B \leq MSFE^A$, where the row determines A and the column determines B. Given that the tests are symmetric, I report only one result for each pair. For a pair (A,B), the adj. Diebold–Mariano statistic and p-value of a test of whether A is a more precise forecast than B, are the inverse of the the adj. Diebold–Mariano statistic and 1 minus the p-value, respectively, of the test of whether B is a more precise forecast than A.

Tables 15 to 18 report the adjusted Diebold–Mariano statistics and the corresponding p-values for the pairwise comparisons of mean squared forecast errors. The null hypotheses of the respective tests are $H_0 : MSFE^B \leq MSFE^A$, where the rows determine A and the columns determine B. Since the Diebold–Mariano test statistic of a test with $H_0 : MSFE^B \leq MSFE^A$ has the opposite sign as the test statistic of that with $H_0 : MSFE^A \leq MSFE^B$, I choose to report only the result of one of the two comparisons between A and B.[16]

Table 15 reports the results from the pairwise comparisons of the three categories of professional DAX forecasting experience. For all pairwise comparisons, judged by a 95% threshold for statistical significance, the evidence suggests that these forecasts are equivalent in terms of forecast accuracy. The difference in mean squared forecast errors is the largest between regular and irregular DAX forecasters. The respective test statistic implies a p-value of about 7%.

The results of the pairwise comparisons of the three groups of self-assessed expertise in conducting quantitative DAX forecasts reported in Table 16 suggest that forecast accuracy increases with expertise, albeit only when a 10% threshold for statistical significance is used. In terms of forecast accuracy, a high level of expertise dominates both medium and low levels of expertise and a medium level of expertise dominates a low level of expertise. The differences in forecast accuracy cannot be attributed to differences in how the groups form their DAX expectations conditional on economic conditions (see column 3 of Table 11).

When I compare the forecasts of those respondents who report to taking interest in the results of the ZEW FMS on stock markets in general to those who are not, I find the forecast accuracy to be equivalent. The adjusted Diebold–Mariano statistic and the implied p-value for the respective test are -0.5684 and 71.45%, respectively. When I re-estimate specification (4) of Table 11 for the

[16]The p-value of a test of whether A is a more precise forecast than B is 1 minus the p-value of the test of whether B is a more precise forecast than A.

Table 16: Differences in forecast accuracy: expertise in conducting quantitative DAX forecasts

B → A ↓	Low expertise	Medium expertise	High expertise
Low expertise	-	-1.4696 (92.76%)	-1.4939 (93.08%)
Medium expertise		-	-1.3788 (91.44%)
High expertise			-

Note: This table reports the results of pairwise comparisons of the mean squared forecast errors (MSFE) made by three subsets of the ZEW FMS panel: respondents who assess their own expertise in conducting quantitative DAX forecasts as low, respondents who assess their own expertise in conducting quantitative DAX forecasts as medium and respondents who assess their own expertise in conducting quantitative DAX forecasts as high. The table reports the adj. Diebold–Mariano statistic and the corresponding p-value for each comparison. The null hypotheses of the tests are given by $H_0 : MSFE^B \leq MSFE^A$, where the row determines A and the column determines B. Given that the tests are symmetric, I report only one result for each pair. For a pair (A,B), the adj. Diebold–Mariano statistic and p-value of a test of whether A is a more precise forecast than B, are the inverse of the the adj. Diebold–Mariano statistic and 1 minus the p-value, respectively, of the test of whether B is a more precise forecast than A.

subsample covering the years 2012–2020, I also do not find any differences in DAX expectations conditional on economic conditions.

The results documented in Table 17 suggest that the respondents' age cohorts do not matter for forecast accuracy.[17] Of the six possible pairwise comparisons, none of the respective null hypotheses can be rejected at the 5% level. Only the null hypothesis of the test of whether the forecasts of age cohort 4 are more precise than those of age cohort 2 can be rejected at the 10% significance level. Consistent with the notion that forecast accuracy and the relationships between DAX expectations and economic conditions are related, the absence of heterogeneity of forecast accuracy across age cohorts coincides with the absence of heterogeneity of DAX expectations conditional on economic conditions, the latter being valid both in the full sample (see column 5 of Table 10) and the subsample from 2012–2020.

Lastly, Table 18 reports the results of pairwise comparisons of the forecast accuracy of the different main occupations represented in the ZEW FMS panel. Given that there are 10 different groups, the issue with too small group sizes is the most pronounced for this personal characteristic, which should be kept in mind when interpreting the results. There are three comparisons for which the null hypotheses can be rejected at the 5% threshold. These are "Trading" vs. "Management", "Financing" vs. "Management", and "Security Research" vs. "Wealth Management". While Table 13 suggests that there are differences with respect to which variables these occupations consider when forecasting DAX returns, these differences are small and unsystematic (e.g. "Trading" and "Financing" seem to consider *ep* while "Management" seems not, whereas "Financing" and "Management" seem to consider *comp* while "Trading" seems not). The results reported in Table 13 thus do not suggest that the detected differences in forecast accuracy across main occupations can be traced back

[17]See Section 6.3 for the definition of age cohorts.

Table 17: Differences in forecast accuracy: age cohorts

B → A ↓	Cohort 1	Cohort 2	Cohort 3	Cohort 4
Cohort 1	-	-1.0402 (84.96%)	-0.5727 (71.59%)	0.3846 (35.07%)
Cohort 2		-	0.9567 (17.06%)	1.3367 (9.22%)
Cohort 3			-	1.0612 (14.56%)
Cohort 4				-

Note: This table reports the results of pairwise comparisons of the mean squared forecast errors (MSFE) made by three subsets of the ZEW FMS panel: respondents who assess their own expertise in conducting quantitative DAX forecasts as low, respondents who assess their own expertise in conducting quantitative DAX forecasts as medium and respondents who assess their own expertise in conducting quantitative DAX forecasts as high. The table reports the adj. Diebold–Mariano statistic and the corresponding p-value for each comparison. The null hypotheses of the tests are given by $H_0 : MSFE^B \leq MSFE^A$, where the row determines A and the column determines B. Given that the tests are symmetric, I report only one result for each pair. For a pair (A,B), the adj. Diebold–Mariano statistic and p-value of a test of whether A is a more precise forecast than B, are the inverse of the the adj. Diebold–Mariano statistic and 1 minus the p-value, respectively, of the test of whether B is a more precise forecast than A.

to differences in how they forecast DAX returns conditional on measures of economic conditions.

8 Summary And Discussion

Motivated by the contradictory empirical evidence on the time-variation in expected stock returns, I have studied the stock market expectations of German financial market experts. My aim was to get a better understanding of the sources of the variation in expected returns, to provide new evidence on the relationship between expected returns and economic conditions and to evaluate the financial experts' forecasting performance. My main findings are that i) respondents strongly disagree about how important macroeconomic and financial variables are related to DAX returns, ii) the measured relationships between my quantitative survey measure of DAX return expectations and measures of economic conditions are largely consistent with the view that expected returns are counter-cyclical, iii) in some cases, the scale of the expectation variable, i.e. metric resulting from a quantitative forecast or ordinal resulting from a qualitative forecast, matters for the measured direction of the relationship between DAX expectations and economic conditions and iv) an aggregated version of my quantitative survey measure of DAX return expectations positively predicts an aggregated measure of realized returns, but is not superior to a simple average of historical DAX returns.

These results contradict the empirical findings from the literature studying expected returns via survey data, which raises the question of why this is the case. From my results, I am not able to give a definite answer to this question. Two explanations are, however, plausible. First, as my results indicate, a potential explanation for why previous studies have documented pro-cyclical

Table 18: Differences in forecast accuracy: main occupation

B → A ↓	Economic research	Trading	Financing	Management	Security research	Fund/portfolio management	Investment advice	Wealth manage-ment	Risk manage-ment	Other
Economic research	-	0.2243	0.6003	1.2187	0.4773	0.4779	0.8814	1.5694	0.7747	1.0172
		(41.15%)	(27.49%)	(11.30%)	(31.71%)	(31.69%)	(19.01%)	(5.99%)	(22.02%)	(15.58%)
Trading		-	1.1061	1.8370	0.0482	0.2174	0.3349	0.7200	1.0715	1.1418
			(13.57%)	(3.46%)	(48.08%)	(41.42%)	(36.92%)	(23.66%)	(14.33%)	(12.82%)
Financing			-	1.8257	-0.3785	-0.5906	-0.2143	0.1300	0.9456	0.7286
				(3.55%)	(64.71%)	(72.19%)	(58.46%)	(44.84%)	(17.34%)	(23.40%)
Management				-	-1.0269	-1.5793	-0.9112	-0.5968	-0.9162	-0.3923
					(84.65%)	(94.12%)	(81.78%)	(72.40%)	(81.91%)	(65.21%)
Security research					-	0.0779	0.6539	1.6743	0.5989	0.8086
						(46.90%)	(25.73%)	(4.86%)	(27.53%)	(21.04%)
Fund/portfolio management						-	0.2676	0.9053	0.8449	1.1609
							(39.48%)	(18.38%)	(20.01%)	(12.43%)
Investment advice							-	0.9214	0.4961	0.6367
								(17.96%)	(31.05%)	(26.29%)
Wealth management								-	0.2346	0.3831
									(40.75%)	(35.12%)
Risk management									-	0.1360
										(44.61%)
Other										-

Note: This table reports the results of pairwise comparisons of the mean squared forecast errors (MSFE) made the different main occupations in the ZEW FMS panel. The table reports the adj. Diebold–Mariano statistic and the corresponding p-value for each comparison. The null hypotheses of the tests are given by $H_0 : MSFE^B \leq MSFE^A$, where the row determines A and the column determines B. Given that the tests are symmetric, I report only one result for each pair. For a pair (A,B), the adj. Diebold–Mariano statistic and p-value of a test of whether A is a more precise forecast than B, are the inverse of the the adj. Diebold–Mariano statistic and 1 minus the p-value, respectively, of the test of whether B is a more precise forecast than A.

expected returns might be measurement error, for example, because the researchers study a qualitative measure of stock return expectations. The list of surveys used in the literature on stock return expectations compiled in Table 1, however, reveals that most studies are based on quantitative measures of stock return expectations. Measurement error might thus only play a minor role here.

The second possible explanation might be that the differences in the results are due to the differing backgrounds of the respondents. Table 1 shows that most studies are based on data from surveys among households or individual investors, whereas my results are based on data from a survey among financial market experts. It is reasonable to assume that financial market experts form stock return expectations that are more in line with the empirical evidence from studies based on realized stock returns, either because they know the literature or, because they have learned the relationship between stock returns and economic conditions while working in the financial sector. The findings of Söderlind (2010), who studies the expectations of economists, point into this direction. Although he also finds that it is negatively correlated with the dividend–price ratio, Söderlind (2010) documents that his survey measure of stock return expectations is higher in recession periods, which is in line with what I find. Interesting questions for future research are thus how the format of the survey question used to measure expected returns affects the measured relationship between expected returns and proxies for expected returns and whether individuals with a background in economics or finance hold systematically different stock return expectations than households or individual investors.

References

Adam, K., A. Marcet, and J. Beutel (2017). Stock price booms and expected capital gains. *American Economic Review 107*(8), 2352–2408.

Amromin, G. and S. A. Sharpe (2014). From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science 60*(4), 845–866.

Bacchetta, P., E. Mertens, and E. Van Wincoop (2009). Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance 28*(3), 406–426.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-capm: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance 55*(4), 1515–1567.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), 1509–1531.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *The Review of Financial Studies 21*(4), 1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cochrane, J. H. (2017). Macro-finance. *Review of Finance 21*(3), 945–985.

Croushore, D. D. (1993). Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia 6*, 3.

Deaves, R., J. Lei, and M. Schröder (2019). Forecaster overconfidence and market survey performance. *Journal of Behavioral Finance 20*(2), 173–194.

Deaves, R., E. Lüders, and M. Schröder (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization 75*(3), 402–412.

Diebold, F. M. and R. Mariano (1995). R.(1995). comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1).

Dominitz, J. and C. F. Manski (2004). How should we measure consumer confidence? *Journal of Economic Perspectives 18*(2), 51–66.

Dominitz, J. and C. F. Manski (2007). Expected equity returns and portfolio choice: Evidence from the health and retirement study. *Journal of the European Economic Association 5*(2-3), 369–379.

Dominitz, J. and C. F. Manski (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics 26*(3), 352–370.

Elton, E. J. (1999). Presidential address: expected return, realized return, and asset pricing tests. *The Journal of Finance 54*(4), 1199–1220.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Ferreira, M. A. and P. Santa-Clara (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics 100*(3), 514–537.

Gennaioli, N., Y. Ma, and A. Shleifer (2016). Expectations and investment. *NBER Macroeconomics Annual 30*(1), 379–431.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2019). Five facts about beliefs and portfolios. Technical report, National Bureau of Economic Research.

Glaser, M., Z. Iliewa, and M. Weber (2019). Thinking about prices versus thinking about returns in financial markets. *The Journal of Finance 74*(6), 2997–3039.

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *The Review of Financial Studies 27*(3), 714–746.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Hoffmann, A. O. and T. Post (2017). How return and risk experiences shape investor beliefs and preferences. *Accounting & Finance 57*(3), 759–788.

Hudomiet, P., G. Kézdi, and R. J. Willis (2011). Stock market crash and expectations of american households. *Journal of Applied Econometrics 26*(3), 393–415.

Hurd, M., M. Van Rooij, and J. Winter (2011). Stock market expectations of dutch households. *Journal of Applied Econometrics 26*(3), 416–436.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance 56*(3), 815–849.

Manski, C. F. (2018). Survey measurement of probabilistic macroeconomic expectations: progress and promise. *NBER Macroeconomics Annual 32*(1), 411–471.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation. *Econometrica 55*(3), 703–708.

Rapach, D. and G. Zhou (2013). Forecasting stock returns. In *Handbook of Economic Forecasting*, Volume 2, pp. 328–383. Elsevier.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies 23*(2), 821–862.

Söderlind, P. (2010). Predicting stock price movements: regressions versus economists. *Applied Economics Letters 17*(9), 869–874.

Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does" irrationality" disappear with wealth? evidence from expectations and actions. *NBER Macroeconomics Annual 18*, 139–194.

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.

Download ZEW Discussion Papers from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de