Journal of Economic Behavior and Organization 188 (2021) 379–393

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Volatility expectations and disagreement☆

Ronald Huisman[a,b], Nico L. Van der Sar[a], Remco C.J. Zwinkels[c,*]

[a] Erasmus School of Economics, Erasmus University Rotterdam, Netherlands
[b] Kyiv-Mohyla Business School (KMBS), National University of Kyiv Mohyla Academy (NaUKMA), Ukraine
[c] Vrije Universiteit Amsterdam and Tinbergen Institute, Netherlands

ARTICLE INFO

Article history:
Received 8 June 2020
Revised 22 April 2021
Accepted 17 May 2021
Available online 11 June 2021

JEL classification:
C42
C53
G12

Keywords:
Volatility forecasting
Disagreement
Survey data

ABSTRACT

This paper examines the use of survey-based measures in volatility forecasting. We argue that an aggregate volatility forecast built up from individual forecasts should be the sum of individual expected volatilities and the dispersion in mean return forecasts. We use data coming from a repeated survey to capture volatility expectations and mean returns of investors, and to produce aggregate volatility forecasts. Our survey-based volatility forecasts are consistent and quantitatively similar with forecasts based on GARCH and implied volatility models. This result is robust to both in-sample and out-of-sample comparisons and in response to news.

1. Introduction

Research in volatility forecasting has flourished over the previous decades. Using historical price information, methods based on generalized autoregressive conditional heteroskedasticity (GARCH), stochastic volatility (SV), and realized volatility, capture salient features of financial time series to provide backward-looking volatility forecasts (e.g. Andersen and Bollerslev, 1998; Andersen et al., 2006). The option-implied volatility method is a popular alternative using market expectations as implied from option prices to produce a forward-looking volatility estimate (e.g. Poon and Granger, 2003; Poon and Granger, 2005). In contrast, survey methods are not commonplace, although they could be potentially important (Pesaran and Weale, 2006; Ang et al., 2007). In fact, Faust and Wright (2013) show that survey-based forecasts often outperform model-based approaches in macro-economics.

In this paper, we focus on investor expectations of returns volatility extracted from investor survey data. We analyze these survey-based volatility expectations and compare them with model-based expected volatilities computed from stock index data as well as option pricing data. In addition, we explore the ability of these survey-based measures of volatility expectations to forecast actual future volatilities compared to the predictive ability of the model-based expected volatilities measures. Our evidence suggests that investor survey forecasts can be used to reflect aggregated beliefs about future

☆ We thank the editor, two anonymous reviewers, Bart Frijns, Andrey Ukhov, Glenn Boyle, Marc Schauten, Albert Menkveld, Lars Bannenberg, Gregory Duffee, Stephan Siegel, Wayne Ferson, William Goetzmann, Werner DeBondt, participants of the AFBC, the AFM, as well as participants of seminars at Erasmus University for all the helpful comments and suggestions.

* Corresponding author at: Erasmus School of Economics, Erasmus University Rotterdam, Netherlands.
E-mail addresses: rhuisman@ese.eur.nl (R. Huisman), vandersar@ese.eur.nl (N.L. Van der Sar), r.zwinkels@vu.nl (R.C.J. Zwinkels).

https://doi.org/10.1016/j.jebo.2021.05.020
0167-2681/

returns volatilities concurrent with expected volatilities models. This contributes to an earlier literature that focused on survey measures of investor expectations of future returns (e.g., Greenwood and Shleifer, 2014; Barberis et al., 2015). In addition, we provide empirical evidence in support of dispersion in return expectations being a part of risk (volatility, or standard deviation) instead of expressing uncertainty. Hence, prior evidence of the average volatility forecast systematically underestimating market volatility can therefore be reconciled with rational expectations models rather than with the behavioral alternative of overconfidence.

This paper uses survey data on beliefs of active and affluent retail investors about future stock market returns to construct a market volatility forecast. The data come from a survey we repeated frequently on a regular and consistent basis. It was explicitly designed to capture volatility expectations and mean returns of investors, and used to extract an aggregate volatility forecast. The primary goal of our empirical investigation is to compare the volatility forecasts based on survey expectations with forecasts derived from existing methods. As such, we test whether our proposed method is a viable alternative to the existing models.

Drawing out an aggregate volatility forecast from survey expectations requires information on individual expected volatilities. We use investors' assessments of the maximum and minimum bounds around their point forecast to fit a triangular-distribution and obtain volatility estimates (Engelberg et al., 2009). Upper and lower bounds that indicate the likely range of outcomes are widely used to measure, for example, 80% or 90% confidence intervals and capture beliefs precision (e.g. DeBondt, 1998; Deaves et al., 2010; Ben-David et al., 2013; Jain et al., 2013). The proportion of actual realizations outside the confidence interval, tends to be much higher than, for example, 20% or 10%. This is an indication of overconfidence, which seems pertinent as an empirical finding in finance and psychology studies (Barberis and Thaler, 2003; Daniel et al., 1998; Alpert and Raiffa, 1982). Volatility forecasts that are recovered from confidence intervals consistently show too precise beliefs (e.g. Graham and Harvey, 2005; Glaser and Weber, 2005; Glaser and Weber, 2007). This implies that whereas at the individual level the volatility of returns is persistently underestimated, based on market information time-series and option-implied methods generally produce accurate forecasts. To take a step in explaining this issue, we analyze overconfidence, or rather beliefs precision, in relation to market volatility and do so jointly with disagreement in beliefs.

Heterogeneity in beliefs is a prevailing feature in the marketplace. Disagreement could be driven by differences in opinion that are not data-based but arise due to investors interpreting the same information differently (e.g. Varian, 1989; Harris and Raviv, 1993). Another possible source is asymmetric information (e.g. Shalen, 1993). The empirical literature on the nature of the relationship between beliefs disagreement and asset prices and returns is extensive. Methods of assessing the level of disagreement include analysis based on active holdings or trades of investors (e.g. Goetzmann and Massa, 2005; Garfinkel, 2009; Jiang and Sun, 2014), analysis of analysts earnings forecasts (e.g. Diether et al., 2002; Johnson, 1957–1978; Boehme et al., 2006; Anderson et al., 2005; Anderson et al., 2009, and direct use of active investor forecasts (e.g. Beber et al., 2010; Carlin et al., 2014). Our paper fits in the last category. This allows us to avoid possible confounding effects of other factors on investor actions as well as possible biases in analyst-based data (Daniel et al., 2002). The setting of stock market returns minimizes information asymmetry. While it is generally accepted that heterogeneous beliefs and asymmetric information are important in financial markets (e.g. He and Wang, 1995; Wang, 1998; Bachetta and Wincoop, 2006), our paper is among the few that investigate individual beliefs and their role in a context where private information is limited or absent (e.g. Patton and Timmermann, 2010; Beber et al., 2010).

Most existing work finds that beliefs disagreement matters for return volatility, but the literature is inconclusive about the interpretation. For example, Johnson (1957–1978) argues that dispersion is a proxy for unpriced idiosyncratic parameter risk, whereas Anderson et al. (2005) establish empirically that a higher dispersion implies both higher expected returns and volatility beyond the effect of traditional risk factors. However, Anderson et al. (2009) make an explicit link to uncertainty, which puts them in a different position. Assuming that risk refers to possible deviations from the mean return, they take the practical view that the true mean is unknown and that uncertainty can be quantified as the dispersion in predictions of market returns. Risk is empirically identified as the conditional volatility of the market. In their framework for asset pricing, they find that uncertainty matters in addition to risk. Ter Ellen et al. (2019) study to what extent disagreement captures uncertainty, and conclude that it is a better proxy for heterogeneity rather than uncertainty. To contribute to this debate, we analyze beliefs dispersion in relation to market volatility and do so jointly with beliefs precision.

The key finding of our paper is a survey-based forecast of market volatility, which conveys investor beliefs directly and has good empirical performance. More specifically, the forecast for the market is constructed exclusively from data on individual beliefs relating to volatility expectations and mean return of the Dutch AEX index. These data come from 79 bi-weekly repeated surveys of a pool of active retail investors between December 2009 and March 2013.

However, investors' volatility expectations cannot be directly aggregated because each individual volatility, or rather belief precision, is attached to a different distribution. To create an individual volatility forecast that is meaningful in an aggregated context, we reduce each volatility to the same scale. We do this per individual by adding in his belief dispersion (i.e., the distance between his point forecast and the mean forecast). With it, we bridge the gap between the individual and the aggregated mean return forecast and provide a theoretical reason for disagreement to be an integral part of risk. We then use the finite mixture distribution which, in line with the point forecast pooling literature (Wallis, 2005; French, 1985; French, 2011), aggregates individual investors beliefs precision and dispersion. That is, to produce a well-calibrated volatility forecast at each survey date, we do not simply take the average belief precision only but add in the dispersion of individual means. Whereas this approach is not new, we are, to the best of our knowledge, the first to apply it in an equity market setting, and also the first to measure the individual components directly.

Comparing our survey-based forecasts of market volatility in-sample to expected volatilities based on GARCH, stochastic volatility, realized volatility, and implied volatility, we find some striking results. The survey-based volatility forecasts are statistically equivalent to the GARCH model forecasts as well as those based on implied volatility. Both the average belief precision and, to a lesser extent, dispersion account for a certain share of future volatility. With it, we provide an empirical reason for disagreement to be a constituent of risk. Neglecting the dispersion of individual means would therefore lead to the underestimation of the volatility forecast and, correspondingly, to the unjustified interpretation of overconfident return expectations. When examining the two components separately, we find that the results are mainly driven by precision rather than dispersion.[1] This suggests that the dispersion alone cannot be regarded as a good proxy for predictive variance (see, e.g., Krüger and Nolte, 2016).

We proceed by focusing on the ability of the survey-based measure to forecast actual return volatility. Assessing its predictive ability in comparison with the out-of-sample forecasting performance of the set of existing volatility models, we find that the performance of the survey-based forecasts is statistically equivalent to the GARCH and implied volatility forecasts.

Furthermore, we show that the survey-based volatility response to news at the market level is approximately the same in strength and direction as the reactions of the GARCH and implied volatilities. Our findings are largely consistent with the leverage effect (e.g., Nelson, 1991; Bollerslev et al., 2006), though its impact is less important in the last part of the sample period. It cannot be excluded that there may be other factors driving the results (e.g., Bekaert and Wu, 2000; Hibbert et al., 2008), but the estimation of alternative causes and effects is beyond the scope of our paper. Instead, the evidence being consistent across all volatility forecasts allows us to confirm the robustness of the survey-based results and to solidify the use of volatility expectations next to or in place of model-based expected volatilities.

An important insight from our analysis is that individual level data can be aggregated to the market level, without additional information or indirect measurements being required. Our study relates to a stream of research that examines inflation and GDP growth forecasts (e.g. Zarnowitz and Lambros, 1987; Giordani and Soderlind, 2003; Boero et al., 2008; Rich and Tracy, 2010; Krüger and Nolte, 2016). In this, however, the use of aggregate volatility forecasts is ignored or discarded on information-based arguments and the results are mixed.

Our study implies that beliefs precision as well as beliefs dispersion may be considered components of the volatility forecast for the stock market. Dispersion in beliefs regarding the mean return then causes beliefs precision to be smaller than market volatility. This explains why smaller individual variances, in the presence of dispersed mean returns, should not be taken as indicating overconfidence (see Huisman et al., 2012). Correspondingly, Knüppel et al. (2019) show that in the case individual variances *are* unbiased (i.e., there is no overconfidence), the dispersion term leads to bias in the aggregate volatility forecast.

Next, our result on the nature of the relationship between beliefs dispersion and volatility forecasts strengthens the basis for its use in research on risk (see Anderson et al., 2005; Anderson et al., 2009). Note that characterizing beliefs dispersion as a component of volatility puts it on a level with volatility forecasts. This is consistent with the test results of Frankel and Froot (1990), who study forecasts of future exchange rates. They find that dispersion Granger-causes volatility and vice versa.

The remainder of the paper is organized as follows. In Section 2 we present the method that combines investor beliefs precision and disagreement in an aggregate forecast for market volatility. Section 3 introduces the data that we collect and describes the methodology that we use to produce volatility forecasts based on survey expectations, GARCH and option-implied models. Section 4 discusses our empirical results on forecast comparison and predictive ability. Section 5 concludes.

2. An aggregate volatility forecast

In this section, we develop a theoretical measure for an aggregate (market-wide) volatility forecast built up from individual beliefs. Specifically, we show that aggregate volatility expectation is the mean of individual's expected volatility (precision, or PREC) plus the dispersion in their mean expectations (DISP). This approach is not new; in fact, it is essentially a definition (Zarnowitz and Lambros, 1987). The main purpose of this framework is to explain how the PREC-DISP distinction can be interpreted within the equity market context, as the existing literature tends to focus on macro-economic variables like inflation expectations. Subsequently, it motivates how the survey data nicely fits this approach. Whereas the distinction between PREC and DISP has indeed been described in the literature before, the unique feature in our approach is that we build up the relationship from the individual to the aggregate (as we need for our data) rather than top-down.

Assume that n investors frame their beliefs about the stock market return r at some future time as a random variable with probability density functions $f_i(r)$, $i = 1, \ldots, n$. For notational convenience, time subscripts are suppressed. We denote the corresponding expected returns as $\bar{r}_i$, $i = 1, \ldots, n$, and the average over investors as $\bar{r}$, such that:

$$\bar{r}_i = \int_{-\infty}^{\infty} r f_i(r)\, \mathrm{d}r, \tag{1}$$

[1] In fact, the two components are highly correlated, making it challenging to disentangle their respective added values. The contribution of dispersion, however, appears to be time-varying. This is consistent with the findings of Deaves et al. (2010) and Glaser and Weber (2005), who show time variation in overconfidence.

and

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n}\bar{r}_i. \tag{2}$$

Now define investor i's belief precision as the expected squared deviation of possible returns r from his expected return $\bar{r}_i$:

$$prec_i = \int_{-\infty}^{\infty}(r-\bar{r}_i)^2 f_i(r)\,\mathrm{d}r, \tag{3}$$

and his belief dispersion as the expected squared deviation of his expected return $\bar{r}_i$ from the mean $\bar{r}$:

$$disp_i = (\bar{r}_i - \bar{r})^2. \tag{4}$$

We can interpret $prec_i$ as a measure of investor i's (lack of) confidence in his return forecast $\bar{r}_i$ by itself and $disp_i$ as that in $\bar{r}_i$ being the mean $\bar{r}$. A straightforward way to measure the extent to which investor i can(not) rely on his belief about market return to forecast the mean then is an additional sum: $prec_i + disp_i$. This supposedly captures how (little) certain he believes his forecast $\bar{r}_i$ is. Taking Eqs. (3) and (4) together, the sum can be written in the form:

$$prec_i + disp_i = \int_{-\infty}^{\infty}(r-\bar{r}_i)^2 f_i(r)\,\mathrm{d}r + (\bar{r}_i - \bar{r})^2. \tag{5}$$

A method to combine investors' beliefs is by taking the finite mixture:

$$f_A(r) = \frac{1}{n}\sum_{i=1}^{n} f_i(r), \tag{6}$$

where $f_A(r)$ denotes the aggregate density. That is, $f_A(r)$ is defined as an equally weighted arithmetic average of the individual densities. Then the mean of $f_A(r)$ is $\bar{r}$. To derive the volatility of $f_A(r)$, we define *PREC* and *DISP* as the average belief precision and dispersion:

$$PREC = \frac{1}{n}\sum_{i=1}^{n} prec_i, \tag{7}$$

and

$$DISP = \frac{1}{n}\sum_{i=1}^{n} disp_i. \tag{8}$$

From Eq. (5) it can easily be seen that the additional sum of these composite measures is equal to the variance of the aggregate density $f_A(r)$:

$$PREC + DISP = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(r-\bar{r})^2 f_i(r)\,\mathrm{d}r. \tag{9}$$

Denoting the aggregate variance, the term on the right-hand side of Eq. (9), as *AVAR*, we can write

$$AVAR = PREC + DISP. \tag{10}$$

In capturing how (little) certain the collectivity of investors are about forecasting market return, *AVAR* is a natural metric for volatility forecast in the market. In the extreme case of entire agreement on the expected return, beliefs about the shape of the densities can be aggregated in a meaningful way by averaging the individual volatilities since these are attached to the same mean. In practice, however, investors disagree on the location of these densities. That is, the expected returns are spread around the mean, i.e. $DISP \neq 0$. As a consequence, the individual volatilities are attached to different expected returns, such that simply averaging is no longer sufficient. Adding in the individual levels of dispersion resolves the problem.

Several features of our approach are worth mentioning. First, combining forecasts by taking averages across individuals does not require the introduction of a conditioning variable. In this respect our approach differs from that of Giordani and Soderlind (2003) where information sets are aggregated. Their model assumption of individual forecasters facing different information sets involves the problem of how to interpret the aggregate distribution. Instead, we allow investors to have different probabilities when faced with the same information. Morris (1995) argues in favour of this personalistic or subjectivist Bayesian view of probabilities (see also Savage, 1954) in research on financial markets. According to Wallis (2005) the aggregate density forecast is a combined forecast in the tradition of the point forecasting pooling literature. Granger (1989) states that "*aggregating forecasts is not the same as aggregating information sets*". These considerations motivate us to use the finite mixture distribution defined by Eq. (6) for aggregating investor beliefs about the density's shape and location.

A second feature of our approach is the use of equal weights in combining density forecasts. In the absence of information differences across investors, there is no knowledge as to which investor will produce a more accurate forecast. The appropriate method then is to treat each investor equal, give each forecast the same weight, and average out random errors. Empirical evidence on the accuracy of a simple average as an aggregation method is provided by several studies (e.g.

Figlewski and Ulrich, 1983; Clemen and Winkler, 1986; Stock and Watson, 2004; Smith and Wallis, 2009; Kenny et al., 2014). In our paper on aggregating expectations with regard to a stock market index, there is no reason to depart from this practice of equally weighted averaging.

A third feature of our approach is that we focus on individual investors and aggregate from there. That is, we use $prec_i$ and $disp_i$ to arrive at a result for the aggregate, *AVAR*, being the average of $prec_i + disp_i$ $(i = 1, \ldots, n)$, as expressed by Eq. (10). Other studies start with the aggregate distribution and use Eq. (10) to decompose its variance, *AVAR*, into the average of the individual variances, *PREC*, and the variance of the individual expected returns, *DISP*. They leave the matter there, not wondering whether it is possible to make a split-up into portions at the individual level and why this is of the importance. Quite differently, we explicate the position of individual expectations, $\bar{r}_i$ and $prec_i$, relative to aggregate expectations, $\bar{r}$ and *AVAR*, by showing that on an individual basis $prec_i$ attached to $\bar{r}_i$ is meaningful in an aggregated context only if $disp_i$ is added since it bridges the gap between $\bar{r}_i$ and $\bar{r}$. *AVAR* then captures how (little) certain investors are, on average, about forecasting a return in itself and about this return being the mean. By demonstrating that our method of aggregation is consistent with the rescaling of volatilities and the decomposition at the individual level, we provide a theoretical reason for dispersion to be a component part of market volatility.

The result of our framework on the complementary roles played by individual volatilities and dispersion in forecasting market volatility has some important implications. First, it offers a potential explanation for the empirical finding that estimates at an individual level tend to be smaller than at the market level. Eq. (10) demonstrates that $PREC < AVAR$ whenever $DISP \neq 0$. This means that the empirical outcome arises out of beliefs dispersion and is not necessarily fully due to overconfidence.

Second, our result is important for empirical research that links volatility forecasts, e.g., based on market information time-series or option-implied methods, to beliefs precision and dispersion based on survey expectations. Most likely, the market's expected volatility, beliefs precision and dispersion are endogenously determined. Eq. (10) does not imply causality but suggests that *PREC* and *DISP*, as additional components, come through as communicating vessels. This is consistent with an observation made by Glaser and Weber (2005), who find that individual volatility forecasts after September 11, are up to about the subsequently realized volatility while differences of opinion are minimized, in contrast with before September 11.

Third, our result establishes that the relationship between the market's expected variance, beliefs precision, and dispersion is contemporaneous in nature. This strengthens the basis for using dispersion in research on risk. This is consistent with Frankel and Froot (1990)'s empirical results on forecasts of future exchange rates, who find that Granger-causality runs from dispersion to volatility and vice versa.

3. Data and methodology

This section first provides an overview of the survey data. We next consider the approach to derive measures of individual expected returns and variances that are used to calculate *AVAR*, *PREC* and *DISP*. We then discuss the methodology applied to compare the survey-based measures with model-based expected volatilities.

3.1. Data

To construct the survey-based measures that are central to our empirical analysis, we use data on investor expectations and adopt the setup of Huisman et al. (2012).[2] We ran a repeated survey among retail clients of a large Dutch commercial bank. The group of clients we analyze are serviced by a department on the trading floor of the bank and belong to the group of mass affluent clients, with amounts of invested wealth between 100,000 and 1,000,000 EUR, who trade at least once per month on financial markets and decide on their portfolio themselves (possibly based on advice provided by the bank). Anecdotal evidence suggests that the investors are especially active in trading options, suggesting that they have a certain degree of knowledge on derivatives and have incentives to focus on market volatility.

Our empirical investigation is based on the AEX Index which is the leading stock market index in the Netherlands, consisting of 25 large cap stocks. It is widely used by practitioners as a benchmark. It is easy to invest in and essentially free from trading frictions. Information asymmetry is absent or only little and all data are readily available.

Every other week on Friday afternoon after market close an e-mail was sent to the pool of investors containing the most recent closing price of the AEX Index and five questions. The first three questions are (translated from Dutch):[3]

"Today, the AEX Index closed at XXX,

- on what level will the AEX Index end on dd-mm-yyyy?
- on what level will the AEX Index end maximally on dd-mm-yyyy?
- on what level will the AEX Index end minimally on dd-mm-yyyy?"

[2] The Huisman et al. (2012) paper is based on data from 21 repeated surveys. This paper utilizes the same surveys plus another 58 repeated surveys.

[3] The original text in Dutch is: Vandaag is de AEX Index gesloten op XXX. - Op welk niveau denkt u dat de AEX Index zal sluiten op dd/mm/yyyy? - Op welk niveau denkt u dat de AEX Index minimaal zal sluiten op dd/mm/yyyy? - Op welk niveau denkt u dat de AEX Index maximaal zal sluiten op dd/mm/yyyy?

Table 1
Descriptive statistics.

	P	n	$E(P)$	$E(H)$	$E(L)$
average	329.43	87.06	331.29	338.79	319.51
median	333.11	81.00	334.56	340.44	322.33
min	276.10	47.00	279.51	290.42	262.44
max	369.65	192.00	370.81	376.34	360.27
st.dev.	22.50	26.84	22.72	21.10	24.53

Notes: This table presents the descriptive statistics of the survey data. P represents the actual index level; n the number of respondents; $E(P)$ the expected index level in two weeks; $E(H)$ the expected maximum index level; and $E(L)$ the expected minimum index level.

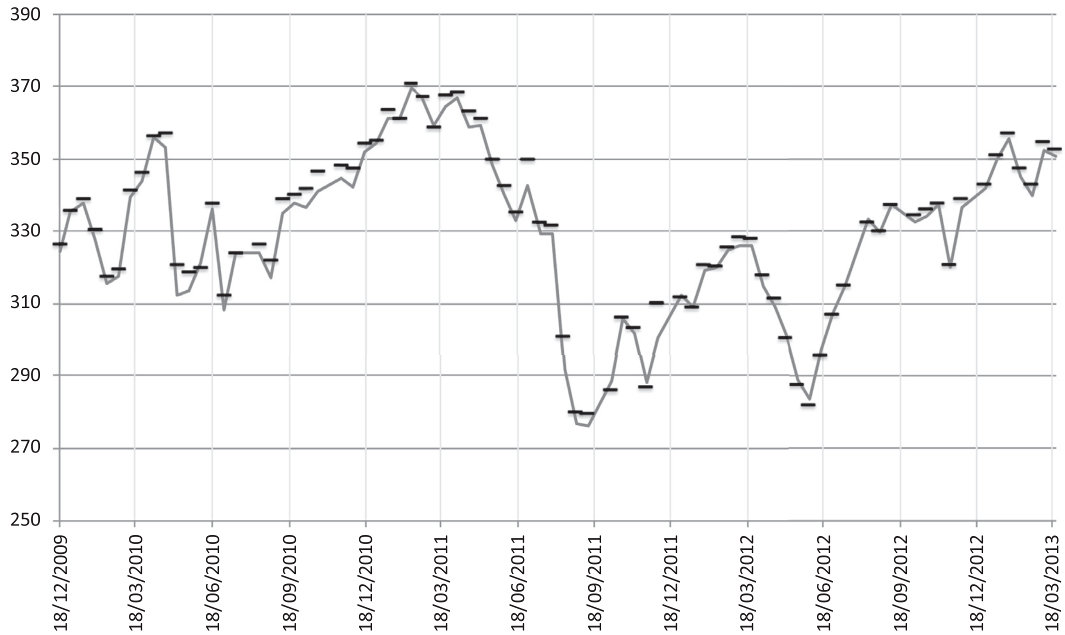

Fig. 1. Index levels and expectations. *Notes:* The figure presents the spot price (line) and average expectation per survey date (bars).

with XXX replaced by the exact closing price of the AEX Index and with dd-mm-yyyy being a specific date of the Friday two weeks after the survey, or a Thursday in case the specific Friday is a holiday.[4]

We obtained data from 79 repeated surveys. Table 1 presents information and summary statistics of these surveys. The first survey was sent out on 18 December 2009. The first closing price of the AEX Index that we communicated to the respondents, was 324.63. A total of 180 participants responded to the first survey. We repeated sending the surveys bi-weekly until 22 March 2013. In total, we have 79 surveys. A few weeks / surveys are missing due to random technical reasons[5]. On average, $n = 87.06$ investors responded; the standard deviation of the number of respondents over the surveys was 26.84. The minimum number of respondents was 47 (survey 77 on 22 February 2013). The maximum was 192 (survey 2 on 31 December 2009). For reasons of privacy we only have anonymous observations and, as a consequence, we therefore cannot track the responses of an individual investor over time. The third column of Table 1 shows that the respondents expected the level of the AEX Index after two weeks to be $E(P) = 331.29$ on average. The fourth and fifth columns show for the maximum (highest) and minimum (lowest) level that the AEX Index might obtain after two weeks average values of $E(H) = 338.79$ and $E(L) = 319.51$, respectively.

Fig. 1 plots the price of the AEX Index and the average price expectations per survey over time. Consider the last survey on the right (sent out on 22 March 2013). On that day, the AEX Index closed at 350.74. On average, the respondents expected the AEX Index to close at 352.45 after two weeks. The time series of the average expected price per survey appears to be

[4] The fourth question asks for individuals' expectations of the average expectation. The fifth question varies over the surveys and is chosen by the bank. Examples are questions about the oil price, or the level of the index at the end of the year.

[5] The missing surveys are from 30 July 2010, 5 November 2010, 23 September 2011, 16 December 2011, 27 July 2012, 21 September 2012 and 14 December 2012.

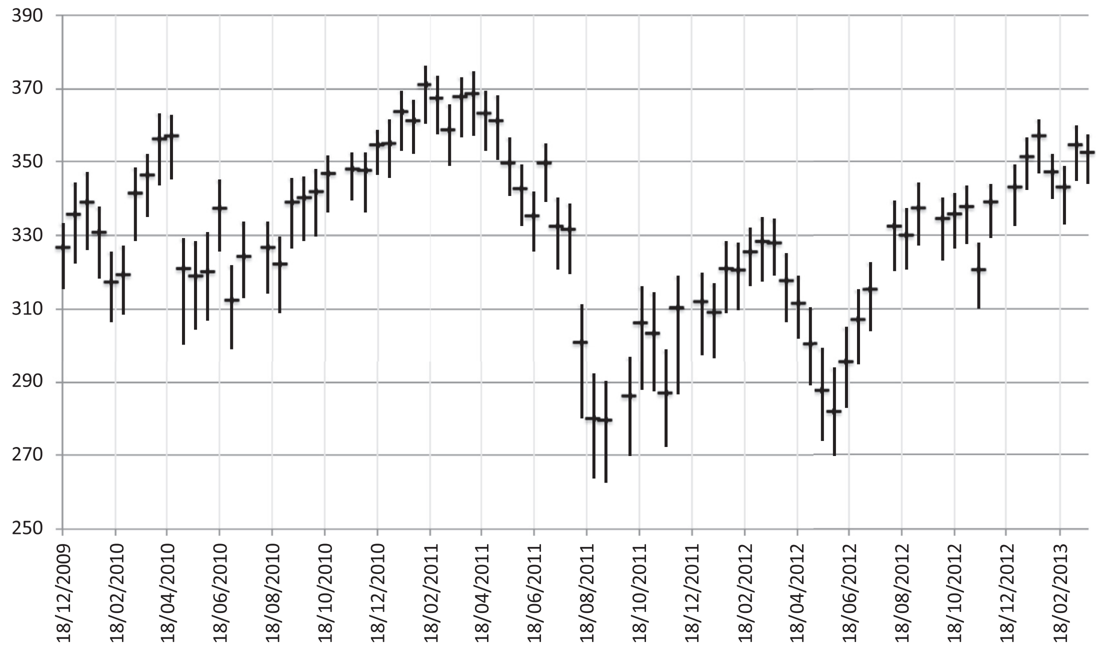

Fig. 2. Level, minimum, and maximum expectation. *Notes:* The figure presents the average expectation and the minimum-maximum range per survey date.

rather close to the price series of the AEX Index. The closing of the AEX Index on the survey date was communicated at the beginning of our set of survey questions to minimize information asymmetry, and the respondents apparently expect that the level of the AEX Index will not change that much in two weeks.

Fig. 2 shows the time-series plots of the expected, the maximum and the minimum level after two weeks for the AEX Index on average per survey. For instance, the last survey on the right (sent out on 22 March 2013) shows that, on average, the respondents expected after two weeks a price of 352.45, a maximum price of 357.63 and a minimum price of 344.08.

3.2. Methodology

Our baseline approach is to calculate the survey-based measures of volatility expectations and compare these with model-based expected volatilities in-sample, out-of-sample, and in relation to news.

For each survey, we obtain the aggregate variance of return expectations *AVAR* by summing *PREC* and *DISP* as in Eq. (10).

To calculate the average belief precision *PREC* of Eq. (7), we measure the individually expected volatility by fitting a triangular distribution to the minimum, maximum, and expected values as in Engelberg et al. (2009) and Christelis et al. (2020). Specifically,

$$\sigma_i^2 = \frac{E(r_h)^2 + E(r_l)^2 + E(r)^2 - E(r_h)E(r) - E(r_h)E(r_l) - E(r)E(r_l)}{18}, \tag{11}$$

in which $E(r)$, $E(r_h)$, and $E(r_l)$ are the expected return, maximum expected return, and minimum expected return of individual i based on the survey data (Kotz and van Dorp, 2004).

We then calculate *PREC* as the average of the individual volatility estimates $\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$, where n is the number of survey respondents. Using only three observations (H_i, L_i, and $E(P)$), the measure is very sensitive to outliers. Poon (2005) suggests to apply trimming procedures in order to reduce the impact of destabilising large values on the volatility estimates. We observe that survey respondents sometimes provided huge differences between minimum and maximum expected prices. As such expectations most likely yield too extreme volatility estimates, we follow Poon (2005) and chose to winsorize the raw survey data at a 10% level.[6]

The individual expected returns are measured by taking the price forecast of each survey participant about the AEX Index after two weeks relative to the price on the survey date. We use these return forecasts per survey to construct the variance of individual expected returns which defines the average beliefs dispersion *DISP* of Eq. (8).

The survey-based volatility forecasts are compared in-sample with different model-based volatility forecasts by estimating univariate regressions of these expected volatilities on *AVAR* and bivariate regressions on *PREC* and *DISP*. The regression results are used to test whether the beliefs of investors about future volatility reflect expected volatilities. To generate ex-

[6] The results are robust to the exact choice of percentage; results available on request.

pected volatilities, we consider four standard measures based on (1) squared two-weekly returns, (2) the GARCH(1,1) model of Bollerslev (1986), (3) the SV model of Harvey et al. (1994), and (4) implied volatilities.

For the first future volatility estimate, we use the squared returns calculated from the closing prices of the AEX Index between two survey dates. To generate one-step ahead forecasts based on the second and third methods, we estimate the GARCH(1,1) and SV model on 10-day returns calculated from the closing prices of the AEX Index using an expanding window. That is, we estimate the model from 14 January 2000 through 4 December 2009, create a two-week volatility forecast for the first survey date (18 December 2009), re-estimate the model including one additional observation to obtain a forecast for the succeeding survey date and we continue to do so until the last survey date. For the fourth future volatility estimate, we use the volatility index VAEX, which reflects the implied volatility of the AEX Index and is reported to be forward looking. The VAEX can be regarded as the most common estimator of future volatility for investors as it is timely available and freely published on several websites.[7] The VAEX is constructed similarly as the VIX index for the S&P500 and represents the expected volatility for the coming month. Since we are asking survey participants for two-week returns, we implicitly assume a flat term structure of implied volatilities from two to four weeks.

The survey-based measures are compared out-of-sample with the four model-based expected volatilities by inspecting how well they predict actual volatility. The results are used to assess their predictive abilities. To produce the forecast errors, we rely on realized volatility (*RV*) as the proxy for actual volatility to be forecasted, which is in line with the literature (e.g., Andersen et al., 2001; Andersen et al., 2010). More specifically, *RV* is computed as the sum of 10 squared historical daily returns between two survey dates. The mean squared error criterion is applied to evaluate the forecast performance (e.g., Koopman et al., 2005). Furthermore, we calculate the significance of the difference in forecasting ability between our survey-based *AVAR* measure and the other models using the Diebold-Mariano statistic; see Diebold and Mariano, 1995.

Finally, we compare the survey-based and model-based volatility forecasts in their response to news at the market level. The results are used to examine whether changes in beliefs of investors about volatility induced by news are consistent with the reactions of model-based expected volatilities. To capture the impact of new information and estimate the responsiveness of the different forecasts, we use univariate regressions of the volatility expectations and the four expected volatilities on prior market return as measured by the percentage price change of the AEX Index over the two weeks before the survey was sent.

4. Empirical results

4.1. In-sample comparison

This paper uses results for the three survey-based measures: *AVAR*, *PREC* and *DISP*, the four expected volatility measures: *GARCH* based on the GARCH(1,1) model, *VAEX* based on implied volatilities, *SV* based on the stochastic volatility model, and *RR* based on squared returns, and the one measure chosen as a proxy to stand in for actual volatility and not used here for forecasting, namely the realized volatility *RV*. Table 2 contains summary statistics of these measures. Mean, standard deviation and extreme values are shown for the first surveys 1–40 in Panel A, the last surveys 41–79 in Panel B, and all the surveys 1–79 in Panel C.

We see that the mean value 0.197 of *AVAR* (with a standard deviation of 0.134) for all the surveys is rather close to the corresponding values 0.161 of *GARCH* (with a standard deviation of 0.114) and 0.196 of *VAEX* (with a standard deviation of 0.143). The mean and standard deviation of *AVAR* and both *GARCH* and *VAEX* are higher in the period covered by the last surveys 41–79 than in the period covered by the first surveys 1–40. When considering the mean, standard deviation and extreme values of *AVAR* in comparison with *SV* and *RR*, the results show relatively large differences. This suggests that *GARCH* and *VAEX* are more appropriate, here, for the forecasting of volatility than *SV* and *RR* are.

Fig. 3 plots the series of *AVAR* (dots) alongside the series of *GARCH* (solid line) on the left and the series of *VAEX* (solid line) on the right, in the top graphs for the first surveys 1–40, the middle graphs for the last surveys 41–79, and the bottom graphs for all the surveys 1–79. As shown, the volatility expectations measured by *AVAR* track the expected volatilities measured by *GARCH* and *VAEX* closely. Although it is not a formal test, it is seen that the path of the *AVAR* series comes nearest to that of *VAEX*.[8] This evidence is consistent with the view that *AVAR* tracks better with *VAEX* because *AVAR* and *VAEX* are both forward looking.

We also compare the volatility expectations with the four model-based expected volatilities by using regression analysis. In the spirit of Mincer and Zarnowitz (1969), we project the different model-based volatility forecasts on a constant and the survey-based forecasts. Table 3 reports the OLS estimates obtained from the in-sample regressions of y on a constant and *AVAR*, where y denotes *GARCH*, *VAEX*, *SV*, or *RR* on the date when survey $s = 1, \ldots, 79$ was sent and ϵ is the error term:

$$y_s = c + aAVAR_s + \epsilon_s. \tag{12}$$

[7] See https://indices.nyx.com/en/products/indices/QS0011052147-XAMS for more information. We obtained historical VAEX close observations from the website www.iex.nl.

[8] The bottom graph at the right of Fig. 3 illustrates that *AVAR* and *VAEX* even peak at the same time, though there are some differences in magnitude. For that matter, the highest *VAEX* observations (surveys 11 and 12 in May 2010) are around 0.006. This is equivalent to an annualized volatility estimate of 39% (annualized standard deviation is equal to $\sqrt{26 \times 0.006} = 0.39$, since 26 two-week periods make up a year and the variance forecast of 0.006 concerns a two-week period), whereas the annualized volatility forecast of *AVAR* then is 33%.

Table 2
Summary statistics volatility.

	AVAR	*PREC*	*DISP*	*GARCH*	*VAEX*	*RV*	*SV*	*RR*
Panel A				Surveys 1–40				
mean	0.159	0.124	0.035	0.151	0.185	0.123	0.120	0.121
st.dev.	0.082	0.062	0.023	0.113	0.123	0.140	0.036	0.272
min.	0.053	0.047	0.006	0.060	0.062	0.016	0.056	0.000
max.	0.480	0.386	0.094	0.541	0.654	0.849	0.186	1.523
Panel B				Surveys 41–79				
mean	0.236	0.178	0.058	0.171	0.208	0.170	0.121	0.132
st.dev.	0.164	0.122	0.047	0.115	0.162	0.163	0.026	0.237
min.	0.068	0.048	0.009	0.060	0.049	0.006	0.066	0.000
max.	0.573	0.478	0.175	0.521	0.656	0.651	0.183	1.448
Panel C				Surveys 1–79				
mean	0.197	0.151	0.046	0.161	0.196	0.147	0.121	0.126
st.dev.	0.134	0.100	0.038	0.114	0.143	0.153	0.031	0.253
min.	0.053	0.047	0.006	0.060	0.049	0.006	0.056	0.000
max.	0.573	0.478	0.175	0.541	0.656	0.849	0.186	1.523

Notes: This table presents the summary statistics of the volatility measures. *AVAR* represents our aggregate variance measure as given by Eq. (10); *PREC* represents the average belief precision as given by Eq. (7); *DISP* represents the average belief dispersion as given by Eq. (8); *GARCH* represents the generalized autoregressive conditional heteroskedasticity volatility estimate; *VAEX* represents the implied volatility index; *RV* represents the realized volatility; *SV* is the stochastic volatility estimate; and *RR* is the squared return.

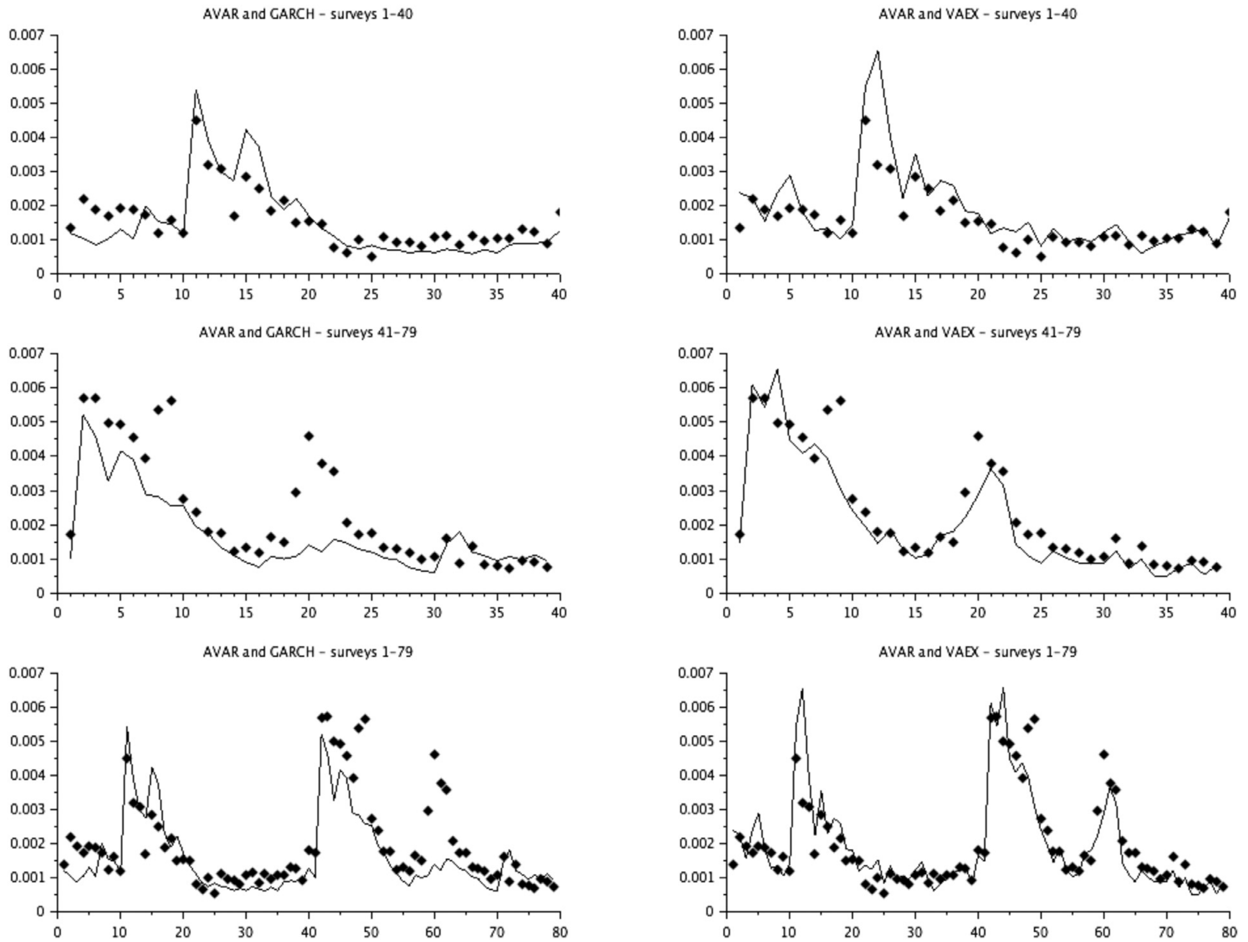

Fig. 3. Volatility measures. *Notes:* The figure presents the volatility measures per survey date. *AVAR* is represented by the dots, while *VAEX* and *GARCH* are given by the lines.

Table 3
Estimation results.

	GARCH				VAEX				SV				*RR*			
Panel A	Surveys 1–40				Surveys 1–40				Surveys 1–40				Surveys 1–40			
c	−0.382***	0.422***	0.111	0.349***	0.247	0.283	0.281	0.190	0.801***	0.777***	0.935***	0.773***	-2.340**	-2.720***	-0.788	-2.986***
	(0.138)	(0.125)	(0.171)	(0.132)	(0.199)	(0.198)	(0.269)	(0.160)	(0.168)	(0.179)	(0.128)	(0.185)	(0.935)	(0.943)	(0.652)	(0.938)
AVAR	1.191***				1.323***				0.250***				2.239***			
	(0.120)				(0.138)				(0.094)				(0.624)			
PREC		1.556***		1.048***		1.721***		1.076***		0.339***		0.371*		3.171***		5.016***
		(0.164)		(0.291)		(0.185)		(0.302)		(0.132)		(0.198)		(0.775)		(1.822)
DISP			4.036***	1.606			4.536***	2.041*			0.759***	-0.101			5.782**	-5.844
			(0.548)	(0.999)			(0.906)	(1.261)			(0.244)	(0.474)			(2.311)	(4.345)
R^2	0.734	0.707	0.647	0.729	0.765	0.731	0.691	0.764	0.313	0.326	0.214	0.309	0.452	0.517	0.220	0.573
P_{Wald}	0.090	0.760	0.000		0.746	0.013	0.000		0.285	0.002	0.000		0.000	0.000	0.446	
Panel B		Surveys 41–79				Surveys 41–79				Surveys 41–79				Surveys 41–79		
c	0.322	0.290	0.689***	0.290	-0.063	-0.085	0.439***	-0.088	1.129***	1.128***	1.148***	1.128***	-0.364	-0.448	0.180	-0.442
	(0.218)	(0.227)	(0.170)	(0.229)	(0.205)	(0.248)	(0.166)	(0.244)	(0.090)	(0.091)	(0.083)	(0.092)	(0.607)	(0.698)	(0.256)	(0.687)
AVAR	0.586***				0.908***				0.036				0.712**			
	(0.127)				(0.129)				(0.032)				(0.350)			
PREC		0.797***		0.822***		1.217***		1.085***		0.048		0.040		0.992*		1.281
		(0.161)		(0.259)		(0.195)		(0.424)		(0.045)		(0.086)		(0.511)		(1.075)
DISP			1.748***	0.076			2.816***	0.409			0.1145	0.0260			1.9513**	-0.8929
			(0.551)	(0.654)			(0.411)	(0.732)			(0.090)	(0.165)			(0.825)	(1.784)
R^2	0.694	0.716	0.494	0.708	0.839	0.842	0.652	0.841	0.024	0.023	0.014	-0.003	0.216	0.236	0.121	0.224
P_{Wald}	0.254	0.145	0.000		0.056	0.371	0.000		0.000	0.000	0.000		0.055	0.302	0.083	
Panel C		All 79 surveys				All 79 surveys				All 79 surveys				All 79 surveys		
c	0.285*	0.236	0.652***	0.239	0.127	0.075	0.604***	0.093	1.059***	1.052***	1.102***	1.051***	-0.600	-0.765	0.117	-0.840
	(0.170)	(0.179)	(0.164)	(0.187)	(0.179)	(0.205)	(0.155)	(0.217)	(0.090)	(0.093)	(0.080)	(0.095)	(0.531)	(0.614)	(0.256)	(0.673)
AVAR	0.671***				0.933***				0.075**				0.947***			
	(0.114)				(0.120)				(0.035)				(0.362)			
PREC		0.910***		0.881***		1.254***		1.092***		0.102**		0.110		1.348***		2.046
		(0.145)		(0.268)		(0.173)		(0.336)		(0.049)		(0.089)		(0.525)		(1.124)
DISP			2.062***	0.086			2.940***	0.492			0.223**	-0.024			2.478***	-2.110
			(0.520)	(0.698)			(0.453)	(0.627)			(0.104)	(0.173)			(0.917)	(1.900)
R^2	0.622	0.632	0.476	0.627	0.759	0.759	0.612	0.760	0.091	0.094	0.062	0.083	0.242	0.272	0.129	0.290
P_{Wald}	0.873	0.199	0.000		0.565	0.008	0.000		0.000	0.000	0.000		0.003	0.074	0.000	

Notes: This table presents the estimation results of Equations (12) - (15). Newey-West standard errors are in parentheses. R^2 indicates adjusted R^2. P_{Wald} represents the P-value for the joint hypotheses $c = 0, a/p/d = 1$.

Results are presented for the first surveys 1–40 in Panel A, the last surveys 41–79 in Panel B, and all the surveys 1–79 in Panel C. We find that in all six cases with *GARCH* and *VAEX*, the Wald test fails to reject the joint hypothesis that $c = 0$ and $a = 1$ at the 10% level. This confirms the previous evidence that investor beliefs about future volatility consistently reflect expected volatilities as measured by *GARCH* and *VAEX*. In contrast, for all but two regressions with *SV* and *RR*, the joint null hypothesis is rejected and the coefficient estimates are significantly different from the hypothesized values $c = 0$ and $a = 1$. This evidence is supportive of the previous interpretation that our *AVAR* measure is more akin to expected volatilities measured by *GARCH* and *VAEX* than measured by *SV* and *RR*.

We proceed to compare the model-based expected volatilities with the investor beliefs about future volatility when decomposed into separate parts. Table 3 reports the estimation results obtained from regressing y on the component parts of *AVAR*, namely on *PREC*, *DISP*, and *PREC* and *DISP* as given by the equations:

$$y_s = c + pPREC_s + \epsilon_s, \tag{13}$$

$$y_s = c + dDISP_s + \epsilon_s, \tag{14}$$

$$y_s = c + pPREC_s + dDISP_s + \epsilon_s. \tag{15}$$

For the univariate regressions with *GARCH* and *VAEX*, the estimates of Eqs. (13) and (14) indicate that the coefficients p on *PREC* and d on *DISP* are typically positive and significant. They are, however, generally not equal to 1.0. From the theory, we expect that the coefficient on *AVAR*, a, is closest to one, whereas the coefficients p and d on *PREC* and *DISP* are larger than one.[9] Therefore, we explicitly focus on the results of the Wald-test rather than the significance of the individually estimated coefficients. Looking at the results for *GARCH* and *VAEX*, we observe that the Wald p-values are highest for the model including *AVAR* in four out of six cases. In the other cases, *PREC* gives the highest p-values. For *SV* and *RR*, the Wald p-values are substantially lower across the board, confirming our earlier observation that the survey-based volatility forecast most closely resembles the *GARCH* and *VAEX* models.

On inclusion of *PREC* and *DISP* in the same regression, the estimated effects of *PREC* and *DISP* on y change. The result that the estimate of the coefficient d on *DISP* in Eq. (15) appears to be insignificant in most cases is most likely caused by the fact that there is a high correlation between *PREC* and *DISP* (0.87), leading to a multicollinearity issue. Furthermore, when looking at the model fit as measured by the adjusted R^2, we observe that the model including only *PREC* tends to have the best fit in all cases, although the differences with *AVAR* are minimal.

Taken together, the evidence shows that the volatility levels provided by *AVAR* closely resemble those of the GARCH and VAEX models. However, whereas *PREC* alone provides a worse fit of the level of volatility, it does better capture the variation in volatility than *AVAR*. *DISP* and *PREC* are highly correlated, and *DISP* appears to be dominated by *PREC*, but *AVAR* is the survey-based volatility forecast that is closest to the canonical empirical volatility forecasting models.

4.2. Out-of-sample forecasting

The previous subsection involves in-sample comparisons to test and check whether what is measured by volatility expectations consistently reflect expected volatilities. This subsection centers on out-of-sample forecasts to assess whether volatility expectations and expected volatilities perform comparatively well in forecasting actual volatility. We evaluate each measure y against *RV* by computing the mean squared error (*MSE*) as the average of the squared differences between the actual value, proxied by realized volatility, and the prediction made for it:

$$MSE = \frac{1}{S}\sum_{s=1}^{S}(RV_s - y_s)^2, \tag{16}$$

where y denotes *AVAR*, *GARCH*, *VAEX*, *SV*, or *RR* on the date when survey $s = 1, \dots, 79$ was sent, and S is the number of surveys or observations in the forecast sample.

Table 4 reports *MSE*s of each forecast measure for various sets of surveys. The values produced by *AVAR* compare closely with those of *GARCH* and *VAEX*. This indicates that these three distinct measures are practically equally accurate in forecasting actual volatility here. The same is not true of *SV* and especially *RR*, which provide clearly less accurate forecasts and show low predictive ability. In terms of significance, we observe that typically all models perform equally strong as *AVAR* in predicting volatility.[10] This is a confirmation that the survey-based *AVAR* measure is a valid alternative to the well-known volatility forecasting models. There are two exceptions to this general result. *PREC* does significantly better than *AVAR* in the second subsample. This might be explained by the fact that the underestimation of individual volatilities is lower in the second subsample (Knüppel et al., 2019). Second, *RR* does significantly worse than *AVAR* in the first subsample as well as the full sample, at the 10% significance level.

[9] Assuming a positive correlation between *PREC* and *DISP*. Note that our setup is not suited for a more in-depth investigation into the relation between *PREC* and *DISP*.

[10] Given the quite large differences in RMSE, this might be explained by the relatively short time-series of our sample.

Table 4
Forecasting ability.

	surveys		
	1–40	41–79	1–79
AVAR	0.104	0.238	0.170
PREC	0.098	0.166	0.132
	(0.763)	(2.074)	(2.180)
GARCH	0.105	0.188	0.146
	(0.234)	(1.008)	(1.014)
VAEX	0.157	0.184	0.170
	(1.127)	(1.489)	(0.155)
SV	0.167	0.314	0.240
	(0.710)	(0.719)	(1.011)
RR	0.317	0.683	0.498
	(1.943)	(1.330)	(1.876)

Notes: This table presents the mean squared error *MSE* of each forecast measure computed as the average of the squared differences between the realized volatility *RV* and the forecast made for it. The numbers in parentheses represent Diebold-Mariano statistics, with $h = 1$, testing for equal forecasting ability between *AVAR* and the respective model.

Table 5
Volatility and news.

	Surveys 1–40			Surveys 41–79			All surveys		
	c	b	R^2	c	b	R^2	c	b	R^2
AVAR	0.158***	0.946***	0.146	0.238***	0.746	0.004	0.198***	0.802**	0.036
	(0.012)	(0.342)		(0.026)	(0.692)		(0.015)	(0.406)	
PREC	0.124***	0.730***	0.155	0.179***	0.561	0.005	0.151***	0.614**	0.039
	(0.009)	(0.256)		(0.020)	(0.517)		(0.011)	(0.302)	
DISP	0.035***	0.216**	0.091	0.059***	0.185	0.003	0.046***	0.188	0.020
	(0.003)	(0.098)		(0.008)	(0.198)		(0.004)	(0.117)	
GARCH	0.151***	0.920*	0.059	0.172***	0.380	0.010	0.161***	0.622	0.028
	(0.017)	(0.496)		(0.019)	(0.488)		(0.013)	(0.346)	
VAEX	0.185***	1.337***	0.125	0.210***	1.096	0.042	0.198***	1.196***	0.082
	(0.018)	(0.521)		(0.025)	(0.671)		(0.015)	(0.423)	
SV	0.120***	0.042	-0.025	0.121***	0.295***	0.157	0.120***	0.178*	0.031
	(0.006)	(0.163)		(0.004)	(0.104)		(0.003)	(0.095)	
RR	0.120***	3.879***	0.243	0.137***	2.541***	0.142	0.130***	3.157***	0.199
	(0.037)	(1.055)		(0.036)	(0.940)		(0.025)	(0.699)	

Notes: This table presents the estimation results of Equation (17) using the distinct volatility measures *AVAR*, *DISP*, *PREC*, *GARCH*, *VAEX*, *SV*, or *RR*.

4.3. Impact of news

The previous evidence involves in-sample and out-of-sample comparisons of volatility expectations and expected volatilities concerning consistent measurement and predictive ability. We now consider the similarity of these survey-based and model-based measures in their relation to news. One way to appraise how news at the market level affects the distinct volatility forecasts is by using prior market return directly as a regressor in the volatility equation. The regression is specified as:

$$y_s = c + br_{s-1} + \epsilon_s, \tag{17}$$

where y denotes *AVAR*, *PREC*, *DISP*, *GARCH*, *VAEX*, *SV*, or *RR* on the date when survey $s = 1, \ldots, 79$ was sent, r denotes the return for the AEX Index over the two-week period prior to the survey date and ϵ is the error term. Table 5 reports the OLS regression estimates of Eq. (17). Results are presented for the first surveys 1–40 in the three most-left columns, the last surveys 41–79 in the three center columns, and all the surveys 1–79 in the three most-right columns.

We find a significantly negative coefficient on prior market return in the first subsample as well as the total sample when *AVAR* is used as the dependent variable. This suggests that good news about the market reduces volatility expectations whereas, in contrast, bad news leads to an increase. Results using *GARCH* and *VAEX* are quantitatively of the same nature, although the coefficient for *VAEX* is also significant in the second subsample. The same findings hold for the two components of *AVAR*, *PREC* and *DISP*. The results for *SV* are quite different, as the coefficient on lagged return is positive. *RR*, finally, shows a very strong negative correlation with lagged returns. This might be explained by the fact that *RR* is constructed using returns only.

Taken together, the evidence on the asymmetric volatility response to market returns is consistent with the leverage hypothesis (e.g., Nelson, 1991; Bollerslev et al., 2006).[11] Furthermore, the findings support the use of survey-based investor ex-

[11] There may be other mechanisms than the leverage effect behind the asymmetric volatility phenomenon (e.g., Bekaert and Wu, 2000; Hibbert et al., 2008).

pectations as a valid measuring-device serving to determine volatility forecasts equivalent to model-based expected volatilities measured by *GARCH* and *VAEX*.

5. Conclusion

In this paper, we create a survey-based volatility forecast directly from individual beliefs about stock market returns. It is an aggregate of the individual volatility expectations, and the dispersion in mean expected returns which reflect beliefs precision and disagreement, respectively. This provides a theoretical reason for dispersion to be a component part of market volatility and contributes to the literature on the role of disagreement in relation to risk. Our empirical investigation uses data coming from repeated surveys of mass affluent retail investors to produce a volatility forecast at each survey date. A comparison with expected volatilities based on GARCH and implied volatility models shows a strong empirical similarity in how they evolve over time, how they perform in predicting future volatility, and how they respond to news.

Most likely, individual expected volatilities are attached to different distributions and can therefore not to be compared directly. A natural way to reduce each volatility to the same scale is by adding the dispersion of individual means. That is not to say that an individual investor knows the average mean. He does not, but probably understands that his mean is an estimate. This is indirectly supported by our empirical evidence on the suitability of the average individual volatility jointly with the dispersion for the forecasting of market volatility. We find that the dispersion bridges the gap between the individual and aggregate context in an appropriate manner and thus induces the average individual volatility forecast to be smaller than the aggregate forecast. Hence, the empirical finding of individual estimates being smaller than market volatility is in itself no proof of overconfidence, which is a novel result.

Our empirical analysis confirms the robustness of using survey-based investor expectations for the determination of volatility forecasts consistent and quantitatively similar with model-based expected volatilities. An important implication is that the dispersion in predictions of the market return may be utilized and estimated as one of the component parts of the volatility prediction for the stock market. This puts dispersion on a level with volatility forecasts and strengthens the basis for its use in research on risk rather than on auxiliary variables.

The empirical contribution of dispersion does not always appear to be equally strong. At the same time, given the theoretical considerations as well as parts of the empirical results, we are of the opinion that it plays a crucial role. Specifically, the theoretical measure suggests that disagreement is part of volatility. The theoretical construct is a definition without any restrictive assumptions. Related, our survey measures match the theoretical constructs one-on-one. When looking at the in-sample results, we observe that the model fit is typically slightly higher for beliefs precision alone than for the aggregate measure. At the same time, we observe that the point estimates are typically better for the aggregate measure than for precision alone. Regarding the out-of-sample results as well as news results, we essentially see no difference in performance. What explains these seemingly contradictory results? We can think of three reasons: (1) there is a high correlation between precision and disagreement, which distorts the tests. (2) the sample period is long enough to perform these fairly basic tests, but not terribly long. This could inflate the standard errors. (3) the survey respondents might not be a representative sample. Altogether, given these caveats, we believe it is remarkable to find that the results for the aggregate survey-based measure are as strong as they are, providing support for the approach.

Our findings could be of importance in situations in which there is no traded asset, such as when valuing real options. The results indicate that the survey-based volatility forecast is statistically equivalent to expected volatilities based on GARCH and implied volatility models. A useful area for future research is to sort out whether the survey-based method provides a viable alternative measuring-device when market data are not available and model-based expected volatilities cannot be determined.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.05.020.

References

Alpert, M., Raiffa, H., 1982. A progress report on the training of probability assessors. In: Kahneman, D., Slovic, P., Tverski, A. (Eds.), Judgement Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, pp. 294–305.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39 (4), 885–905.
Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Volume 1. Elsevier, Amsterdam, pp. 777–878.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2010. Parametric and nonparametric volatility measurement. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, Volume 1. Elsevier, Amsterdam, pp. 67–137.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. J. Financ. Econ. 61 (1), 43–76.
Anderson, E.W., Ghysels, E., Juergens, J.L., 2005. Do heterogeneous beliefs matter for asset pricing? Rev. Financ. Stud. 18 (3), 875–924.
Anderson, E.W., Ghysels, E., Juergens, J.L., 2009. The impact of risk and uncertainty on expected returns. J. Financ. Econ. 94 (2), 233–263.
Ang, A., Bekaert, G., Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better? J. Monet. Econ. 54 (4), 1163–1212.
Bachetta, P., Wincoop, E.V., 2006. Can information heterogeneity explain the exchange rate determination puzzle? Am. Econ. Rev. 96 (3), 552–576.
Barberis, N., Greenwood, R., Jin, L., Shleifer, A., 2015. X-CAPM: an extrapolative capital asset pricing model. J. Financ. Econ. 115 (1), 1–24.
Barberis, N., Thaler, R.H., 2003. A survey of behavioral finance. In: Constantinides, G.M., Harris, M., Stulz, R. (Eds.), Handbook of Economics and Finance, Volume 1b. Elsevier, Amsterdam, pp. 1053–1123.
Beber, A., Breedon, F., Buraschi, A., 2010. Differences in beliefs and currency risk returns. J. Financ. Econ. 98 (3), 415–438.
Bekaert, G., Wu, G., 2000. Differences in beliefs and currency risk returns. Rev. Financ. Stud. 13 (1), 1–42.
Ben-David, I., Graham, J.R., Harvey, C.R., 2013. Managerial miscalibration. Q. J. Econ. 128 (4), 1547–1584.
Boehme, R.D., Danielson, B.R., Sorescu, S.M., 2006. Short-sale constraints, differences of opinion, and overvaluation. J. Financ. Quant. Anal. 41 (2), 455–487.
Boero, G., Smith, J., Wallis, K.F., 2008. Uncertainty and disagreement in economic prediction: the bank of england survey of external forecasters. Econ. J. 118, 1107–1127. (July)
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31 (3), 307–327.
Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. J. Financ. Econom. 4 (3), 353–384.
Carlin, B.I., Longstaff, F.A., Matoba, K., 2014. Disagreement and asset prices. J. Financ. Econ. 114 (2), 226–238.
Christelis, D., Georgarakos, D., Jappelli, T., & van Rooij, M. (2020). Trust in the central bank and inflation expectation. ECB Working Paper, No. 2375.
Clemen, R.T., Winkler, R.L., 1986. Combining economic forecasts. J. Bus. Econ. Stat. 4 (1), 39–46.
Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. J. Financ. 53 (6), 1839–1885.
Daniel, K., Hirshleifer, D., Teoh, S., 2002. Investor policy in capital markets: evidence and policy implications. J. Monet. Econ. 49 (1), 139–209.
Deaves, R., Luders, E., Schroder, M., 2010. The dynamics of overconfidence: evidence from stock market forecasters. J. Econ. Behav. Organ. 75 (3), 402–412.
DeBondt, W.F.M., 1998. A portrait of the individual investor. Eur. Econ. Rev. 42 (3–5), 831–844.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.
Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. J. Financ. 57 (5), 2113–2141.
Engelberg, J., Manski, C.F., Williams, J., 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. J. Bus. Econ. Stat. 27 (1), 30–41.
Faust, J., Wright, J.H., 2013. Forecasting inflation. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 2a. Elsevier, Amsterdam, pp. 3–56.
Figlewski, S., Ulrich, T., 1983. Optimal aggregation of money supply forecasts: accuracy, profitability and market efficiency. J. Financ. 28 (3), 695–710.
Frankel, J.A., Froot, K.A., 1990. The rationality of the foreign exchange rate: chartists, fundamentalists, and trading in the foreign exchange market. Am. Econ. Assoc. Pap. Proc. 80 (2), 181–185.
French, S., 1985. Group consensus probability distributions: a critical survey. In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics, Volume 2. North-Holland, Amsterdam, pp. 183–201.
French, S., 2011. Aggregating expert judgement. RACSAM 105 (1), 181–206.
Garfinkel, J.A., 2009. Measuring investors' opinion divergence. J. Account. Res. 47 (5), 1317–1348.
Giordani, P., Soderlind, P., 2003. Inflation forecast uncertainty. Eur. Econ. Rev. 47 (6), 1037–1059.
Glaser, M., Weber, M., 2005. September 11, and stock return expectations of individual investors. Rev. Financ. 9 (2), 243–279.
Glaser, M., Weber, M., 2007. Overconfidence and trading volume. Geneva Risk Insur. Rev. 32 (1), 1–36.
Goetzmann, W.N., Massa, M., 2005. Dispersion of opinion and stock returns. J. Financ. Mark. 8 (3), 325–350.
Graham, J.R., Harvey, C.R., 2005. The long-run equity risk premium. Financ. Res. Lett. 2, 185–194.
Granger, C.W.J., 1989. Combining forecasts - twenty years later. J. Forecast. 8, 167–173.
Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Rev. Financ. Stud. 27 (3), 714–746.
Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. Rev. Financ. Stud. 6 (3), 473–506.
Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Econ. Stud. 61, 247–264.
He, H., Wang, J., 1995. Differential information and dynamic behavior of stock trading volume. Rev. Financ. Stud. 8 (4), 919–972.
Hibbert, A.M., Daigler, R.T., Dupoyet, B., 2008. A behavioral explanation for the negative asymmetric return-volatility relation. J. Bank. Financ. 32 (10), 2254–2266.
Huisman, R., Sar, N.L.V.d., Zwinkels, R.C.J., 2012. A new measurement method of investor overconfidence. Econ. Lett. 114 (1), 69–71.
Jain, K., Mukherjee, K., Bearden, J.N., Gaba, A., 2013. Unpacking the future: a nudge toward wider subjective confidence intervals. Manag. Sci. 59 (9), 1970–1987.
Jiang, H., Sun, Z., 2014. Dispersion in beliefs among active mutual funds and the cross-section of stock returns. J. Financ. Econ. 114 (2), 341–365.
Johnson, T.C., 2004. Forecast dispersion and the cross section of expected returns. J. Financ. 59 (5), 1957–1978.
Kenny, G., Kostka, T., Masera, F., 2014. How informative are the subjective density forecasts of macroeconomists? J. Forecast. 33, 163–185.
Knüppel, M., & Krüger, F.. Forecast uncertainty, disagreement, and the linear pool. Deutsche Bundesbank Discussion Paper, No. 28/,. 2019.
Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the s&p100 stock index using historical, realised, and implied volatility measurements. J. Empir. Financ. 12, 445–475.
Kotz, S., van Dorp, J., 2004. Beyond Beta: Other Continuous Families of Distributions With Bounded Support and Applications. World Scientific Publishing Co., Singapore.
Krüger, F., Nolte, I., 2016. Disagreement versus uncertainty: evidence from distribution forecasts. J. Bank. Financ. 72, S172-S186.
Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance. National Bureau of Economic Research, New York, pp. 3–46.
Morris, S., 1995. The common prior assumption in economic theory. Econ. Philos. 11 (2), 227–253.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59 (2), 347–370.
Patton, A.J., Timmermann, A., 2010. Why do forecasters disagree? lessons from the term structure of cross-sectional dispersion. J. Monet. Econ. 57 (7), 803–820.
Pesaran, M.H., Weale, M., 2006. Survey expectations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Volume 1. Elsevier, Amsterdam, pp. 715–776.
Poon, S.H., 2005. A Practical Guide to Forecasting Financial Market Volatility. John Wiley & Sons Ltd, Chichester.
Poon, S.H., Granger, C.W.J., 2003. Forecasting volatility in financial markets: a review. J. Econ. Lit. 41 (2), 478–539.
Poon, S.H., Granger, C.W.J., 2005. Practical issues in forecasting volatility. Financ. Anal. J. 61 (1), 45–56.
Rich, R., Tracy, J., 2010. The relationships among expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. Rev. Econ. Stat. 92 (1), 200–207.
Savage, L.J., 1954. The Foundations of Statistics. John Wiley & Sons Ltd.
Shalen, C.T., 1993. Volume, volatility, and the dispersion of 'beliefs'. Rev. Financ. Stud. 6 (2), 405–434.
Smith, J., Wallis, K.F., 2009. A simple explanation of the forecast combination puzzle. Oxf. Bull. Econ. Stat. 6 (2), 331–355.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. J. Forecast. 23, 405–430.
Ter Ellen, S., Verschoor, W.F.C., Zwinkels, R.C.J., 2019. Agreeing on disagreement: heterogeneity or uncertainty? J. Financ. Mark. 44, 17–30.

Varian, H.R., 1989. Differences of opinion in financial markets. In: Stone, C.S. (Ed.), Financial risk: Theory, Evidence and Implications. Kluwer, Boston, pp. 3–37.
Wallis, K.F., 2005. Combining density and interval forecasts: a moderate proposal. Oxf. Bull. Econ. Stat. 67, 983–994. (supplement).
Wang, F.A., 1998. Strategic trading, asymmetric information and heterogeneous prior beliefs. J. Financ. Mark. 1 (3–4), 321–352.
Zarnowitz, V., Lambros, L.A., 1987. Consensus and uncertainty in economic prediction. J. Polit. Econ. 95 (3), 591–621.