Dispersion in options investors' versus analysts' expectations: Predictive inference for stock returns*

Panayiotis C. Andreou†‡ Anastasios Kagkadis§ Paulo Maio¶ Dennis Philip‡

August 2019

Abstract

We create a market-wide measure of dispersion in options investors' expectations by aggregating across all stocks the dispersion in trading volume across moneynesses (DISP). DISP exhibits strong negative predictive power for future market returns and its information content is not subsumed by several alternative equity premium predictors. Consistent with the implications of theoretical models that link dispersion to overpricing, the predictive power of DISP is particularly pronounced in relatively optimistic periods. Although an aggregate analysts' forecasts dispersion (AFD) measure also performs well in optimistic periods, it delivers insignificant overall predictability. This is because in the aftermath of the 2008 financial crisis, AFD was heavily driven by pessimistic forecasts and hence its increase did not reflect a true overpricing. As a result, AFD does not appear to be a robust equity premium predictor in recent years.

JEL Classification: G10, G11, G12

Keywords: Dispersion in beliefs; Predictability of stock returns; Equity premium; Trading volume dispersion; Out-of-sample predictability; Economic significance

*We would like to thank Ivo Welch (the Editor), Yakov Amihud, Kevin Aretz, Turan Bali, Suleyman Basak, Stephen Brown, Frankie Chau, George Constantinides, Damian Damianov, Gregory Eaton, Amit Goyal, Bruce Grundy, Hui Guo, Peter Hansen, Fuwei Jiang, Madhu Kalimipalli, Neil Kellard, Eirini Konstantinidi, Robert Korajczyk, Alex Kostakis, Yuehao Lin, Hao Luo, Hong Luo, James Mitchell, Bradley Paye, Ilaria Piatti, David Rapach, Esther Ruiz, Fabio Trojani, Radu Tunaru, Masahiro Watanabe, Hao Zhou and seminar and conference participants at Athens University of Economics and Business, Nottingham University Business School, Arne Ryde Workshop (Lund University), Eurofidai Paris Finance Meeting, CICF, and the annual meetings of NFA, FMA, MFA, and EFMA for helpful comments and suggestions. We are also grateful to Robert Shiller, Jeffrey Wurgler, and Hao Zhou for making data available on their websites. An earlier version of the paper was circulated under the title "Dispersion in options investors' exectations and stock return predictability". The DISP time-series and the codes used for its construction are available on the authors' websites.

†Department of Commerce, Finance and Shipping, Cyprus University of Technology, Lemesos, Cyprus; Email: panayiotis.andreou@cut.ac.cy

‡Durham University Business School, Mill Hill Lane, Durham DH1 3LB, UK; Email: dennis.philip@durham.ac.uk

§Department of Accounting and Finance, Lancaster University Management School, Lancaster LA1 4YX, UK; Email: a.kagkadis@lancaster.ac.uk

¶Hanken School of Economics, P.O. Box 479, 00101 Helsinki, Finland; Email: paulofmaio@gmail.com

1 Introduction

A number of papers, such as Diether et al. (2002), Park (2005), and Yu (2011), among others, have studied dispersion in investors' beliefs via the dispersion in analysts' forecasts. These papers show that various such measures significantly forecast lower future stock returns at both individual and aggregate levels. This negative relation can be explained either by the short-sales constraints framework of Miller (1977) or the theoretical model of Atmaz and Basak (2018). Both theories predict that dispersion in beliefs heightens the effect of optimistic views on stock prices and hence leads to overpricing and subsequent low returns.

In our paper, we construct an alternative aggregate measure of belief dispersion by using the value-weighted aggregate dispersion in trading volume across different moneyness levels of individual stock option contracts (DISP). Such a dispersion measurement relies on the empirical fact that options end-users establish positions that are mainly related to directional expectations about the future price of the underlying asset (Lakonishok et al., 2007). Within such an environment, more optimistic (pessimistic) investors are expected to buy more out-of-the-money call (put) options, to benefit from the higher leverage, or sell more in-the-money put (call) options, to benefit from the higher premium. Therefore, the dispersion in trading volume across moneyness levels can be seen as a natural proxy for investors' diverge expectations (Andreou et al., 2018).

Our in-sample predictive analysis establishes a significantly negative relationship between DISP and future market returns that is particularly strong for horizons up to a quarter ahead. Out-of-sample predictive analysis shows that DISP has significantly higher forecasting power than the historical average – with out-of-sample R^2 values ranging between 2.95% and 5.18% – and that this forecasting power is also economically important. A market-timing trading strategy offers excess certainty equivalent returns that range from 5.53% to 7.30%. We further show that the information content of DISP is not subsumed by a wide range of prominent predictors. Finally, consistent with Yu (2011), we find that the aggregate analysts' forecasts dispersion (AFD) is also negatively related to future market returns. However, unlike DISP, the predictive power of AFD has become statistically insignificant during our sample period, 1996-2017.

In accordance with the implications of the theories of Miller (1977) and Atmaz and Basak (2018), we find that both DISP and AFD perform better in relatively optimistic periods. More importantly, we show that the source of the overall insignificant predictive power of AFD is its pattern in the pessimistic period following the 2008 financial crisis. In particular, AFD increased due to some highly pessimistic forecasts. Because an increase in dispersion of this nature does not lead to an optimistic bias on the part of the marginal investor, it is highly unlikely that under such circumstances high dispersion is associated with concurrent overpricing. Therefore, AFD does not appear to be a robust predictor in recent years. In contrast, DISP was never heavily driven by pessimistic views in the post-crisis period. It is more likely that, when increasing, it correctly anticipates overpricing. As a result, DISP provides a consistently strong predictive power throughout the sample period.

The remainder of the paper is structured as follows. Section 2 describes the data and the construction of the main variable used in the study. Section 3 provides the empirical results. Finally, Section 4 concludes.

2 Data and Variables

Our paper proxies for the dispersion in options investor's expectations via the dispersion in trading volume across moneyness levels. In particular, we use an aggregated version of the measure described in Andreou et al. (2018). First, for each month t and for each firm i, we calculate the average daily dispersion in volume-weighted moneyness levels:

$$\mathrm{IDISP}_{i,t} = \frac{1}{N}\sum_{n=1}^{N}\left\{\sqrt{\sum_{j=1}^{J} w_{i,n,j}\left(Y_{i,n,j} - \sum_{j=1}^{J} w_{i,n,j}Y_{i,n,j}\right)^2}\right\}, \tag{1}$$

where $w_{i,n,j}$ is the proportion of the total daily trading volume attributed to the j^{th} moneyness $Y_{i,n,j} = \frac{K_{i,n,j}}{S_{i,n}}$, $S_{i,n}$ is the stock price, $K_{i,n,j}$ for $j = 1, ..., J$ is the series of available strike prices, and $n = 1, ..., N$ is the number of days within a month.[1] Next, we estimate the monthly aggregate dispersion measure as:

$$\mathrm{DISP}_t = \sum_{i=1}^{I} v_{i,t}\mathrm{IDISP}_{i,t}, \tag{2}$$

[1]We use the last-but-one trading day as the end of the month to eliminate the effect of non-synchronous trading in the stock versus the options market (Battalio and Schultz, 2006). Moreover, in the estimation of Equation (1) we consider only those days when there are at least five contracts with positive trading volume and we require that $N \geq 10$ within the given month.

where $v_{i,t}$ is the weight attributed to each stock based on its market capitalization at the end of the month.

Our data are from OptionMetrics and cover the period 1996:01 to 2017:12. We keep only options with maturities between 5 and 60 calendar days, because short-term options tend to be the most liquid. We also discard near-the-money options with moneyness between 0.95 and 1.05. Such near-the-money options exhibit the highest sensitivity to volatility and are likely to be traded as part of straddles, strangles and delta-hedged positions, plausibly reflecting investors' expectations about future market volatility rather than returns (Ni et al., 2008).[2]

The main alternative predictor is the analysts' forecasts dispersion (AFD) of Yu (2011). Figure 1 plots DISP and AFD, both standardized to have zero mean and variance one for easier comparison. While AFD is smoother than DISP, there is a clear similarity in the evolution of the two variables, especially until the 2008 financial crisis. In response to the subprime mortgage crisis, DISP starts increasing in late 2007 and peaks in the period around the Lehman Brothers' collapse, while it reverts back to relatively low levels in the latest part of the sample. On the other hand, AFD increases but not substantially during the 2008 financial crisis, while it remains at high levels during the whole period of the European sovereign debt crisis. As expected, DISP and AFD are positively correlated in the 1996:01-2007:12 period with a coefficient of 65%, while they are negatively correlated in the 2008:01-2017:12 period with a coefficient of –10%.

In addition, we control for some other predictors, namely the variance risk premium (VRP; Bollerslev et al., 2009), tail risk (TAIL; Kelly and Jiang, 2014), the aggregate dividend-price ratio (d-p; Campbell and Shiller, 1988a,b), the aggregate earnings-price ratio (e-p; Campbell and Shiller, 1988a), the market dividend-payout ratio (d-e; Campbell and Shiller, 1988a), the yield term spread (TERM; Fama and French, 1989), the default spread (DEF; Fama and French, 1989) and the relative short-term risk-free rate (RREL; Campbell, 1991).

[2]In the Supplementary Material online we show that various alternative DISP measures that rely on different construction methods or filtering rules provide very similar results to those presented in the paper. We further show that a sophisticated DISP measure that utilizes signed volume data (only out-of-the-money options purchases and in-the-money options sales) provides almost identical information content with that of the main DISP.

3 Empirical Results

3.1 In-sample analysis

We run multiple-horizon in-sample regressions of excess annualized CRSP value-weighted index returns on a set of predictive variables. We report Newey and West (1987) standard errors with lag length equal to the forecasting horizon. In addition, we account for the Stambaugh (1999) small-sample bias, using bootstrapped p-values as in Neely et al. (2014) and Huang et al. (2015). Details for the bootstrap procedure are provided in the Supplementary Material online. The slope coefficients reported in the subsequent tables have been multiplied by the standard deviation of the respective predictors and can be interpreted as the percentage annualized excess market return associated with a one standard deviation increase in each regressor.

Panel A of Table 1 provides the results of 1- and 3-month horizon univariate predictive regressions for the dispersion in options investors' expectations and the dispersion in analysts' forecasts. DISP is statistically significant at the 5% level for the 1-month horizon and at the 1% level for the 3-month horizon. Unreported results suggest that its forecasting power gradually tapers off for longer horizons. The slope estimates are negative and economically significant: for example, a one standard deviation increase in DISP predicts a negative annualized monthly market excess return of 10% and a negative annualized quarterly market excess return of 8%. The respective R^2 values are 3.45% for the 1-month horizon and 6.04% for the 3-month horizon.

Unlike DISP, AFD exhibits a negative but statistically insignificant relationship with future market returns. At a first glance, this result is in line with Yu (2011), who finds a significant predictability only for horizons of one to three years ahead. However, Table 2 shows that in our sample period AFD does not exhibit any long-horizon predictive power either.[3]

Finally, the results of bivariate regression analyses are presented in Panel B of Table 1. They show that the information content of DISP is not subsumed by AFD or by any of the alternative predictors used in the study. In particular, DISP remains significant, mostly at the 5% or 1% level, in all cases. Only VRP and to a lesser extent TAIL exhibit similar significance.

[3] It is important to note that we are able to replicate Yu's (2011) results for his sample period.

3.2 Out-of-sample analysis

Following Goyal and Welch (2003, 2008), at the end of each month and for each predictive model we form an out-of-sample (OS) forecast of the equity premium using only data that are available at the time the forecast is made. We use five years of data as the initial in-sample (IS) period and hence the first prediction is made for 2001:01.[4] This way we create a series of OS forecasts that is compared to a series of recursively estimated historical averages. A visual representation of the OS forecasting performance of the two dispersion proxies is given by Figure 2. The red solid lines show the difference in cumulative squared OS prediction error between the historical average model and the DISP and AFD models. We observe that DISP exhibits positive values at the end of the sample period indicating that it outperforms the historical average model in terms of OS predictive performance. In contrast, AFD exhibits negative values indicating that it underperforms the historical average model.

Table 3 assesses the significance of the OS predictive power of all the predictive variables examined. It displays the OS R^2 of Goyal and Welch (2008), the MSE-F test of McCracken (2007), the ENC-NEW test of Clark and McCracken (2001) and the MSE-Adjusted test of Clark and West (2007). We further evaluate the economic significance of the OS predictability results by considering a mean-variance investor who allocates her wealth every month between the market index and the risk-free asset using the OS forecast of each predictive model and an estimate of the market return variance based on historical data. The risk aversion coefficient is set equal to three. Following Campbell and Thompson (2008), we impose realistic leverage values by constraining the portfolio weight on the market index to lie between 0 and 1.5. For each trading strategy, Table 3 presents the Sharpe ratio (SR) and the certainty equivalent return (ΔCER) in excess of the historical average (HAV) strategy.

The results in Table 3 demonstrate that DISP exhibits positive and quite large R^2_{OS} estimates, ranging between 2.95% at the 1-month horizon and 5.18% at the 3-month horizon. Moreover, the three statistical tests reveal that the outperformance of DISP compared to the historical average is statistically significant in all cases, mostly at the 5% level. On the other hand, AFD provides

[4] A five-year initial IS period provides a sufficiently long series of data for accurately estimating each predictive model. The OS performance for alternative initial IS periods can be inferred from Figure 2. Moreover, the Hansen and Timmermann (2012) split mining robust test shows that the OS forecasting power of DISP is robust at the 1% level to the OS split point chosen for all horizons examined.

consistently negative R^2_{OS} values and AFD's OS performance is significantly better than that of the historical average model only on the ENC-NEW test and only for the 3-month horizon.

Turning to the other predictors, only VRP provides positive R^2_{OS} values, ranging between 5.51% and 7.07%, with the MSE-F, ENC-NEW and MSE-Adjusted tests rejecting the respective null hypotheses at the 5% level in all cases. Therefore, DISP performs not only better than AFD but also better than most of the other popular equity premium predictors.

Table 3 further shows that the strategy associated with DISP exhibits an annualized SR of 1.00 at the 1-month horizon and an annualized SR of 0.73 at the 3-month horizon. These values are higher than the SR values of the HAV strategy (0.37 and 0.28, respectively; unreported). Moreover, DISP exhibits a ΔCER of 7.30% at the 1-month horizon and a ΔCER of 5.53% at the 3-month horizon. These values show that the utility provided by the DISP strategy is higher than the utility of the strategy associated with the recursively estimated historical average. Although the AFD strategy also performs better than the HAV strategy, AFD's performance is worse than that of DISP. AFD has a SR of 0.79 (ΔCER of 3.33%) at the 1-month horizon and a SR of 0.64 (ΔCER of 3.17%) at the 3-month horizon. DISP also outperforms the alternative equity premium predictors in terms of economic gains, ΔCER and SR.[5]

3.3 Interpretation of the negative predictability

The documented negative relationship between dispersion in beliefs and the equity premium can be explained in the context of Miller's (1977) theory. In the presence of short-sale constraints,[6] asset prices may reflect only the views of the most optimistic investors, because pessimistic investors sit out of the market. Therefore, higher dispersion in beliefs is associated with increased optimism on the part of those investors who hold the stock. Hence, it is associated with overpricing and a subsequent lower return.[7] A direct implication of this mechanism is that the generated overpricing

[5]The Supplementary Material online provides additional trading strategy results using binary weights. DISP continues to perform better than the HAV and most of the alternative strategies. By contrast, AFD provides Sharpe ratios that are lower than those of the HAV strategy and negative ΔCERs.

[6]A series of studies shows that in practice short-selling is relatively limited. See, for example, D'Avolio (2002), Almazan et al. (2004) and Barber and Odean (2008).

[7]It is important to note that while pessimistic investors may migrate to the options market in order to obtain negative exposure to the risky asset, this will not necessarily prevent the appearance of an overpricing. For the overpricing to be eliminated there must be selling or shorting activity in the equity market. For example, Grundy et al. (2012) show that during the period of the 2008 short-sales ban there were substantial put-call parity deviations in banned stocks, due to the fact that the underlying stocks were persistently overpriced. The reason that these mispricings were not eliminated is because the necessary arbitrage strategies required selling the underlying asset and

should be more pronounced in periods of high investor sentiment (Yu, 2011; Stambaugh et al., 2012).

Alternatively, the negative relation can be interpreted in the context of Atmaz and Basak's (2018) theoretical model of belief dispersion without short-sales constraints. They demonstrate that the effect of dispersion in expectations can be dual. First, dispersion represents an additional risk factor and hence is related to a decreased stock price and a higher required return. Second, it enhances the effect of news on investor sentiment and hence leads to a higher (lower) stock price and a lower (higher) expected return following good (bad) news. In periods of high investor sentiment the news effect can dominate the risk effect and higher levels of dispersion can be associated with an increased stock price and a lower expected return.

3.4 Informativeness of DISP and AFD across time and sentiment periods

Figure 2 helps to gauge the informativeness of DISP and AFD for future market returns across time. Intuitively, whenever the curve in each of the graphs increases, i.e. it is positively sloped, the respective dispersion in beliefs measure outperforms the historical average. By contrast, when the curve in each of the graphs decreases, i.e. it is negatively sloped, the dispersion in beliefs measure underperforms the historical average. The black dashed lines refer to the IS predictive analysis, while the red solid lines refer to the OS predictive analysis. Shaded areas represent relatively optimistic months based on whether the Baker and Wurgler (2006) index is above its sample median.

First, in line with the implications of the theoretical models discussed above, both dispersion measures exhibit strong predictive power for the largely optimistic period from 1996 to 2007. Their performance is particularly strong for the years 2000-2001 and 2007-2008.[8] Second, DISP performs satisfactorily for the largely pessimistic period that follows the 2008 financial crisis. In particular, the DISP curves are mostly positively-sloped in the periods from mid-2009 to mid-2011, from mid-2012 until the first quarter of 2015, and from the second quarter of 2016 until the end of the sample. Negatively-sloped curves are limited to the periods from mid-2011 to mid-2012 and from the second quarter of 2015 to the first quarter of 2016 (mainly for the 3-month horizon). Unlike DISP, AFD

hence they were open only to investors who already owned the respective stocks.

[8] However, in the Supplementary Material online we present additional empirical evidence showing that the predictive power of DISP is not driven by the crisis observations.

exhibits a poor performance for the period from mid-2009 until 2012 and never really recovers.

Because AFD performed poorly in the post-2008 period, it is interesting to investigate the divergence of the two variables. A dispersion measure can increase either because pessimistic views become even more pessimistic, or optimistic views become even more optimistic, or both. However, overpricing leading to subsequent low returns can be generated only if an increase in dispersion is at least partly driven by more optimistic views. If a rise in dispersion is heavily driven by pessimistic views, then it is highly unlikely to be associated with an overpricing.

To see this, consider a simplified example of an investing environment à la Miller (1977). Assume that there are eleven investors with increasing subjective valuations about the stock price from 5 to 15 dollars (mean and median of 10, standard deviation of 3.16) and there is a fixed supply of three shares. If each investor is able to purchase only one share, then the marginal investor is the third most optimistic investor and the stock price is 13. An increase of the three most optimistic beliefs by one, combined with a decrease of the three most pessimistic beliefs by one (a new standard deviation of 4.57), would lead to a new share price of 15. In the above scenario, the increase in divergence of beliefs is symmetric and hence the mean and median valuations continue to be equal to 10. However, it is possible to have the same increase in standard deviation with a change only in the three most optimistic beliefs. In this case, the valuation of the marginal investor will be even higher, calculated at 16.73, with the mean belief higher than the median belief by 1.02. If instead we have the same increase in standard deviation with a change only in the three most pessimistic beliefs, the valuation of the marginal investor remains at 13 and the mean is lower than the median by 1.02. Overall, this example illustrates that overpricing is expected to be more pronounced when high dispersion is mainly driven by optimistic views, i.e. when dispersion is positively related to the difference between the mean and the median of the distribution. In contrast, it is very unlikely to have an overpricing when high dispersion is mostly driven by pessimistic views, i.e. when dispersion is negatively related to the difference between the mean and the median of the distribution.[9]

Motivated by this, we investigate whether DISP and AFD were affected consistently by op-

[9]The model of Atmaz and Basak (2018) relies on the assumption that investors' beliefs are normally distributed and hence cannot account for a scenario where dispersion is driven asymmetrically either by optimistic or pessimistic views. However, the mechanism of the model is conceptually very similar to the above discussion. In particular, the stock price is determined by the mean bias in belief, which is positive (negative) after good (bad) news. A higher dispersion of beliefs amplifies the above effect, in the sense that for the same level of good (bad) news the belief distribution is moved further to the right (left) when dispersion is higher. Consequently, an overpricing due to higher dispersion is always associated with a predominant increase in optimistic views.

timistic and pessimistic views across time. To this end, we compare DISP with an aggregate asymmetry in options investors' expectations measure, ASYM, which captures the difference between the mean and the median of the volume-weighted moneyness levels. Analogously, we compare AFD with an aggregate analysts' forecasts asymmetry measure, AFA, which captures the difference between the mean and median of the distribution of forecasts. The precise definitions of the two asymmetry measures are described in Appendix A. Intuitively, those two variables are high when large positive beliefs dominate and low when large negative beliefs dominate.

We observe that the 1996-2017 correlation between AFD and AFA was –10%, but the 1996-2007 correlation was 48%. By contrast, the 1996-2017 correlation between DISP and ASYM was 36%, while the correlation in the 1996-2007 period was 53%. This pattern is evident by looking also at the time-series plot of the variables which is provided in the Supplementary Material online. Therefore, it seems that both dispersion measures exhibited strong predictive power when their increases were accompanied by an asymmetric increase in optimistic views, i.e. a higher difference between the mean and the median. AFD failed to predict after the 2008 financial crisis because its increases were driven by some very pessimistic analysts' forecasts. An increase of this nature cannot be related to a concurrent overpricing. DISP remained a robust equity premium predictor because, unlike AFD, it was never predominantly driven by pessimistic views.

4 Conclusion

In this paper, we capture the dispersion in options investors' expectations by the dispersion in trading volume across various moneyness levels. Aggregating across all stocks, yields an aggregate dispersion proxy. This proxy exhibits significant in-sample and out-of-sample predictive power for the equity premium. Its information content is not subsumed by the dispersion in analysts' forecasts, the variance risk premium, tail risk, various market valuation ratios and macroeconomic variables.

Moreover, an aggregate measure of dispersion in analysts' forecasts no longer exhibits significant predictive power. Dispersion in beliefs is theoretically linked to overpricing and hence both measures tend to perform particularly well in optimistic periods. The dispersion in options investors' expectations performed reasonably well also in the largely pessimistic period following the 2008 financial crisis. This is because it was never driven by pessimistic views. In contrast, the dispersion in analysts' forecasts performed very poorly in the post-crisis period because it increased mainly due to pessimistic forecasts.

Overall, it seems that in the last two decades the dispersion in options investors' expectations has been a more reliable equity premium predictor than the dispersion in analysts' forecasts. More research is needed in the future to investigate whether this finding will continue to hold or not.

Appendix A: Description of Variables

Alternative predictors

AFD (Yu, 2011): The cross-sectional value-weighted average of the individual-stock standard deviation of analysts' forecasts. The forecasts refer to the long-term growth rate of each stock's earnings per share. The data on analysts' forecasts are obtained from I/B/E/S, while the market capitalization for each NYSE/AMEX/NASDAQ firm (with share code either 10 or 11) at the end of each month is calculated using data from CRSP.

VRP (Bollerslev, Tauchen and Zhou, 2009): The difference between the expected 1-month ahead stock return variance under the risk neutral measure and the expected 1-month ahead variance under the physical measure. The past 1-month realized variance is used as the expected 1-month ahead variance under the physical measure. Monthly VRP data are obtained from Hao Zhou's website.

TAIL (Kelly and Jiang, 2014): The probability of extreme negative market returns constructed by applying a power law estimator to the whole NYSE/AMEX/NASDAQ cross-section (share codes 10 and 11) of daily returns within a given month. Data are from CRSP.

d-p (Campbell and Shiller, 1988a,b): The aggregate dividend-price ratio constructed as the difference between the log aggregate annual dividends and the log level of the S&P 500 index. Data are obtained from Robert Shiller's website.

e-p (Campbell and Shiller, 1988a): The aggregate earnings-price ratio constructed as the difference between the log aggregate annual earnings and the log level of the S&P 500 index. Data are obtained from Robert Shiller's website.

d-e (Campbell and Shiller, 1988a): The market dividend-payout ratio constructed as the difference between the log aggregate annual dividends and the log aggregate annual earnings. Data are obtained from Robert Shiller's website.

TERM (Fama and French, 1989): The difference between the 10-year bond yield and the 1-year bond yield. Data are obtained from the FRED database of the Federal Reserve Bank of St. Louis.

DEF (Fama and French, 1989): The difference between BAA and AAA corporate bonds yields from Moody's. Data are obtained from the FRED database of the Federal Reserve Bank of St. Louis.

RREL (Campbell, 1991): The difference between the 3-month t-bill rate and its moving average over the preceding twelve months. Data are obtained from the FRED database of the Federal Reserve Bank of St. Louis.

Other variables

ASYM: It is constructed using exactly the same data used for the DISP estimation and following the same procedure. The only difference is that the standard deviation of volume-weighted moneyness levels in Equation (1) is replaced by the difference between the mean and the median of the volume-weighted moneyness levels (Scherbina, 2008).

AFA: It is constructed using exactly the same data used for the AFD estimation and following the same procedure. The only difference is that the standard deviation of the distribution of analysts' forecasts is replaced by the difference between the mean and the median of the distribution (Scherbina, 2008).

References

[1] Almazan, A., K. C. Brown, M. Carlson and D. A. Chapman. 2004. Why constrain your mutual fund manager? Journal of Financial Economics 73, 289-321.

[2] Andreou, P. C., A. Kagkadis, D. Philip and R. Tuneshev. 2018. Differences in options investors' expectations and the cross-section of stock returns. Journal of Banking and Finance 94, 315–336.

[3] Atmaz, A. and S. Basak. 2018. Belief dispersion in the stock market. Journal of Finance 73, 1225-1279.

[4] Baker, M. and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645-1680.

[5] Barber, B. M. and T. Odean. 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785-818.

[6] Battalio, R. and P. Schultz. 2006. Options and the bubble. Journal of Finance 61, 2071-2102.

[7] Bollerslev, T., G. Tauchen and H. Zhou. 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22, 4463-4492.

[8] Campbell, J. Y. 1991. A variance decomposition for stock returns. Economic Journal 101, 157-179.

[9] Campbell, J. Y. and R. J. Shiller. 1988a. Stock prices, earnings and expected dividends. Journal of Finance 43, 661-676.

[10] Campbell, J. Y. and R. J. Shiller. 1988b. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195-228.

[11] Campbell, J. Y. and S. Thompson. 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies 21, 1509-1531.

[12] Clark, T. and M. McCracken. 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85-110.

[13] Clark, T. and K. West. 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

[14] D'Avolio, G. 2002. The market for borrowing stock. Journal of Financial Economics 66, 271-306.

[15] Diether, K. B., C. J. Malloy and A. Scherbina. 2002. Differences of opinion and the cross section of stock returns. Journal of Finance 57, 2113-2141.

[16] Fama, E. and K. R. French. 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

[17] Goyal, A. and I. Welch. 2003. Predicting the equity premium with dividend ratios. Management Science 49, 639-654.

[18] Goyal, A. and I. Welch. 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455-1508.

[19] Grundy, B. D., B. Lim and P. Verwijmeren. 2012. Do option markets undo restrictions on short sales? Evidence from the 2008 short-sale ban. Journal of Financial Economics 106, 331-348.

[20] Hansen, P. R. and A. Timmermann. 2012. Choice of sample split in out-of-sample forecast evaluation. Working paper, European University Institute.

[21] Huang, D., F. Jiang, J. Tu and G. Zhou. 2015. Investor sentiment aligned: a powerful predictor of stock returns. Review of Financial Studies 28, 791-837.

[22] Kelly, B. and H. Jiang. 2014. Tail risk and asset prices. Review of Financial Studies 27, 2841-2871.

[23] Lakonishok, J., I. Lee, N. D. Pearson and A. M. Poteshman. 2007. Option market activity. Review of Financial Studies 20, 813-857.

[24] McCracken, M. 2007. Asymptotics for out of sample tests of Granger causality. Journal of Econometrics 140, 719-752.

[25] Miller, E. M. 1977. Risk, uncertainty, and divergence of opinion. Journal of Finance 32, 1151-1168.

[26] Neely, C. J., D. E. Rapach, J. Tu and G. Zhou. 2014. Forecasting the equity risk premium: the role of technical indicators. Management Science 60, 1772–1791.

[27] Newey, W. and K. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

[28] Ni, S. X., J. Pan and A. M. Poteshman. 2008. Volatility information trading in the option market. Journal of Finance 63, 1059-1091.

[29] Park, S. 2005. Stock return predictability and the dispersion in earnings forecasts. Journal of Business 78, 2351-2376.

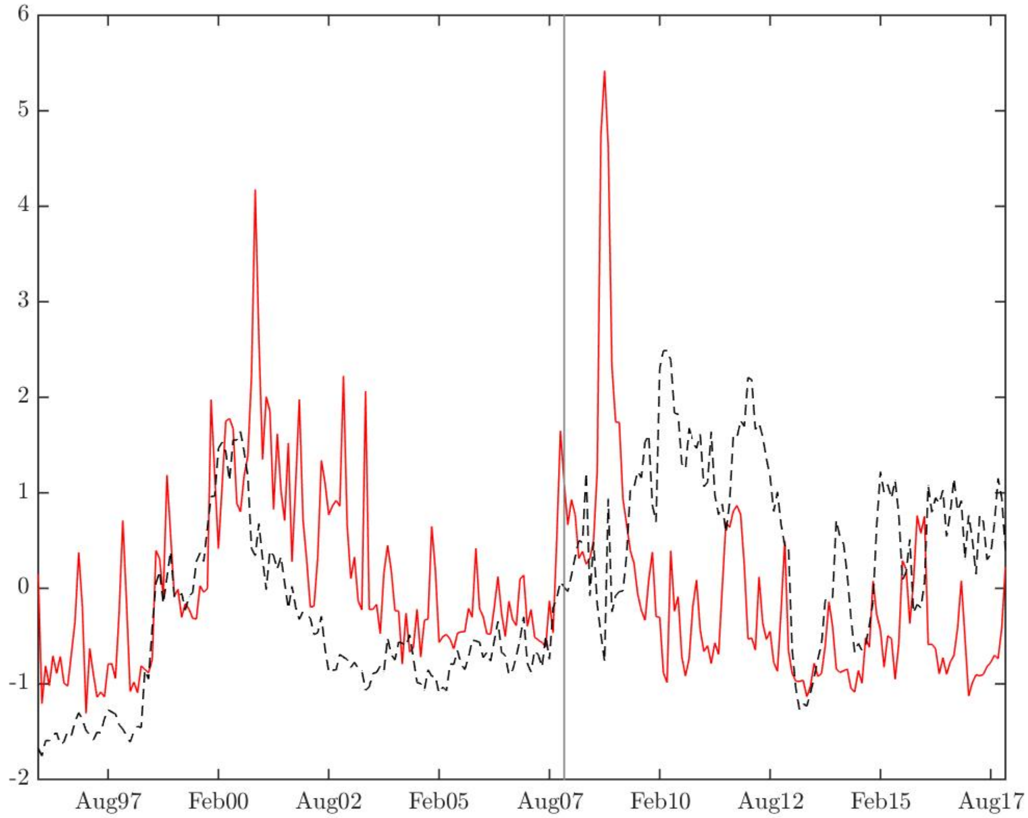
[30] Scherbina, A. 2008. Suppressed negative information and future underperformance. Review of Finance 12, 533–565.

[31] Stambaugh, R. 1999. Predictive regressions. Journal of Financial Economics 54, 375-421.

[32] Stambaugh, R., J. Yu and Y. Yuan. 2012. The short of it: investor sentiment and anomalies. Journal of Financial Economics 104, 288-302.

[33] Yu, J. 2011. Disagreement and return predictability of stock portfolios. Journal of Financial Economics 99, 162-183.

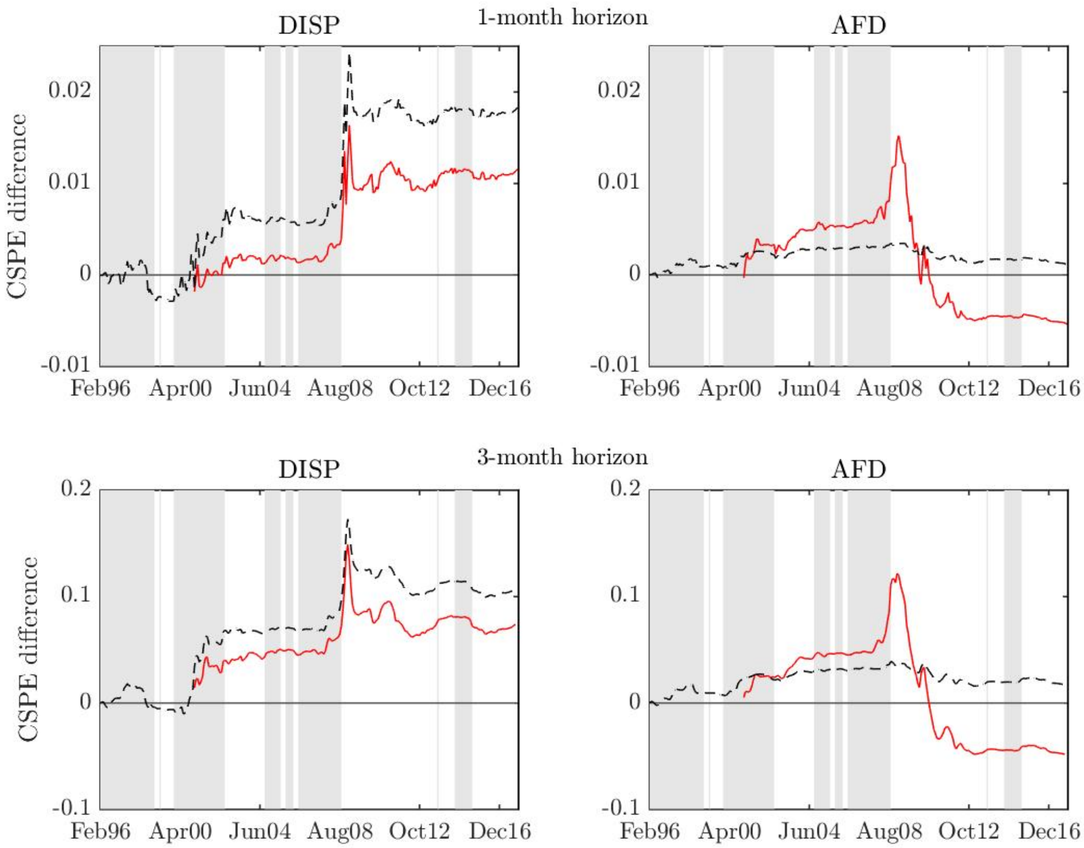
Figure 1: Dispersion in options investors' expectations versus dispersion in analysts' forecasts



Description: This figure plots the monthly time series of the dispersion in options investors' expectations (DISP, red solid line) versus the dispersion in analysts' forecasts (AFD, black dashed line) for the period 1996:01-2017:12. Both variables have been standardized to have zero mean and variance one. The grey vertical line denotes the month 2007:12.

Interpretation: DISP and AFD tended to comove until 2008 but differed in their responses to the financial crisis and in the post-2008 period.

Figure 2: Differences in cumulative squared prediction error



Description: This figure plots the differences between the cumulative squared prediction error (CSPE) of the historical average model and each of the models based on the dispersion in options investors' expectations (DISP) or the dispersion in analysts' forecasts (AFD). The forecasting horizons are one and three months ahead. The black dashed lines correspond to the CSPE from the in-sample analysis, while the red solid lines correspond to the CSPE from the out-of-sample analysis. The total sample period is 1996:01-2017:12 and the out-of-sample period is 2001:01-2017:12. Shaded areas denote optimistic periods based on the Baker and Wurgler (2006) index being above its sample median.

Interpretation: Both DISP and AFD performed well in the largely optimistic period from 1996 to 2007. Their predictive performance was exceptional in the years 2000-2001 and 2007-2008. DISP also exhibited a reasonably good predictive ability in the largely pessimistic period that followed the 2008 financial crisis. By contrast, the predictive power of AFD collapsed after the crisis.

Table 1: In-sample stock return predictability

Panel A: Predictive power of DISP and AFD

Predictor	1-month horizon $\widehat{\beta}$	1-month horizon R^2 (%)	3-month horizon $\widehat{\beta}$	3-month horizon R^2 (%)
DISP	–10.02 (–2.59)**	3.45	–8.14 (–2.87)***	6.04
AFD	–2.46 (–0.79)	0.21	–3.24 (–1.28)	0.96

Panel B: Predictive power of DISP controlling for alternative predictors

Alternative Predictor	1-month horizon $\widehat{\beta}_{DISP}$	1-month horizon $\widehat{\beta}_{ALT}$	1-month horizon R^2 (%)	3-month horizon $\widehat{\beta}_{DISP}$	3-month horizon $\widehat{\beta}_{ALT}$	3-month horizon R^2 (%)
AFD	–9.87 (–2.53)**	–1.03 (–0.33)	3.48	–7.83 (–2.70)**	–2.06 (–0.83)	6.42
VRP	–8.52 (–2.16)**	12.45 (3.63)**	8.68	–7.01 (–2.86)**	9.20 (3.27)**	13.62
TAIL	–10.79 (–2.66)**	–1.75 (–0.49)	3.53	–10.32 (–3.58)***	–4.99 (–2.18)**	7.87
d-p	–9.62 (–2.30)**	5.93 (1.26)	4.65	–7.70 (–2.36)**	6.45 (1.65)	9.81
e-p	–11.17 (–2.76)**	–2.70 (–0.56)	3.65	–9.57 (–3.50)***	–3.36 (–0.75)	6.88
d-e	–11.84 (–3.14)***	5.39 (1.17)	4.33	–10.24 (–4.17)***	6.21 (1.74)	9.15
TERM	–10.01 (–2.58)**	0.86 (0.26)	3.47	–8.12 (–2.87)**	1.07 (0.38)	6.14
DEF	–12.09 (–2.92)***	3.76 (0.63)	3.79	–10.94 (–3.89)***	5.12 (1.05)	7.71
RREL	–8.62 (–2.25)**	3.35 (0.85)	3.76	–6.20 (–2.14)*	4.69 (1.39)	7.69

Description: Panel A reports the in-sample regression results for a univariate predictive model with either the dispersion in options investors' expectations (DISP) or the dispersion in analysts' forecasts (AFD). Panel B reports the in-sample regression results for a bivariate predictive model with DISP and each of the alternative predictors used in the study. The forecasted variable is the CRSP value-weighted index excess return. The remaining forecasting variables are the variance risk premium (VRP), tail risk (TAIL), dividend-price ratio (d-p), earnings-price ratio (e-p), dividend payout ratio (d-e), yield term spread (TERM), default spread (DEF) and relative short-term risk-free rate (RREL). The sample period is 1996:01-2017:12. Reported coefficients indicate the percentage annualized excess return resulting from a one standard deviation increase in each predictor variable. Newey and West (1987) t-statistics with lag length equal to the forecasting horizon are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels based on a wild bootstrap experiment.

Interpretation: DISP is negatively and significantly related to monthly and quarterly future market excess returns. Its predictive power is not subsumed by several alternative predictors. By contrast, AFD is negatively but insignificantly related to future market returns.

Table 2: Long-horizon predictive power of the dispersion in analysts' forecasts

	6-month horizon	12-month horizon	24-month horizon	36-month horizon
AFD	–3.47	–3.37	–1.91	–0.33
	(–1.37)	(–1.25)	(–0.63)	(–0.11)
R^2 (%)	1.98	3.47	2.12	0.10

Description: This table reports the in-sample results for the predictive regressions of the CRSP value-weighted index excess return on the dispersion in analysts' forecasts (AFD). The sample period is 1996:01-2017:12. Reported coefficients indicate the percentage annualized excess return resulting from a one standard deviation increase in each predictor variable. Newey and West (1987) t-statistics with lag length equal to the forecasting horizon are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels based on a wild bootstrap experiment.

Interpretation: AFD does not provide significant predictive power for future market excess returns at long horizons.

Table 3: Out-of-sample stock return predictability

	DISP	AFD	VRP	TAIL	d-p	e-p	d-e	TERM	DEF	RREL
	1-month horizon									
R^2_{OS} (%)	2.95	–1.37	5.51	–1.19	–0.99	–3.22	–7.26	–4.02	–4.40	–0.80
MSE-F	6.20**	–2.76	11.89**	–2.39	–1.99	–6.36	–13.81	–7.89	–8.59	–1.61
ENC-NEW	7.27**	0.66	13.82**	–0.10	0.87	2.18*	–0.89	–2.70	3.59**	0.11
MSE-Adj	1.86**	0.26	1.87**	–0.05	0.30	0.41	–0.13	–1.66	0.48	0.06
SR	1.00	0.79	0.43	0.36	0.61	0.85	0.61	0.29	0.47	0.68
ΔCER (%)	7.30	3.33	0.82	–0.10	3.95	4.87	2.40	–0.77	0.80	3.55
	3-month horizon									
R^2_{OS} (%)	5.18	–3.38	7.07	–1.87	–4.72	–13.88	–36.93	–15.06	–24.55	–4.81
MSE-F	10.92**	–6.53	15.23**	–3.66	–9.02	–24.37	–53.94	–26.17	–39.42	–9.19
ENC-NEW	14.76**	2.54**	24.55**	–1.65	0.25	–1.73	–12.38	–8.97	0.91	–0.42
MSE-Adj	1.62*	0.44	2.25**	–1.44	0.03	–0.18	–1.04	–1.51	0.06	–0.07
SR	0.73	0.64	0.48	0.19	0.56	0.63	0.29	0.15	0.49	0.63
ΔCER (%)	5.53	3.17	2.37	–0.82	3.95	3.65	0.17	–1.62	1.69	3.84

Description: This table reports the results of out-of-sample predictability of the CRSP value-weighted index excess return. The total sample period is 1996:01-2017:12 and the forecasting period begins in 2001:01. The forecasting variables are the dispersion in options investors' expectations (DISP), analysts' forecasts dispersion (AFD), variance risk premium (VRP), tail risk (TAIL), dividend-price ratio (d-p), earnings-price ratio (e-p), dividend payout ratio (d-e), yield term spread (TERM), default spread (DEF) and relative short-term risk-free rate (RREL). R^2_{OS} is the out-of-sample coefficient of determination, MSE-F is the McCracken (2007) F-statistic, ENC-NEW is the encompassing test of Clark and McCracken (2001) and MSE-Adj is the MSE-Adjusted statistic of Clark and West (2007). ** and * denote significance at the 5% and 10% levels. The critical values for the MSE-F test are 1.518 and 0.616, respectively, while the critical values for the ENC-NEW test are 2.374 and 1.442, respectively. These critical values are based on Monte-Carlo simulations and are provided by the respective studies. We also report the annualized Sharpe ratio (SR) and certainty equivalent return (ΔCER) of a market-timing strategy that is based on each of the predictive models and utilizes mean-variance weights. The benchmark strategy follows the historical average model.

Interpretation: DISP exhibits significant out-of-sample predictive power for monthly and quarterly future market excess returns. AFD does not provide significant out-of-sample predictive power. DISP also outperforms the historical average strategy in terms of economic gains. Overall, DISP is the second best OS predictor, after VRP, at both horizons in terms of statistical significance. It is the best OS predictor at both horizons in terms of economic significance.

Supplementary Material to "Dispersion in options investors' versus analysts' expectations: Predictive inference for stock returns"

Not for publication

August 2019

Abstract

This supplementary material includes additional results not presented in the main body of the paper.

A DISP with Signed Volume

In the main paper, DISP is estimated using raw options volume data. This is supported by the empirical evidence provided by Hu (2014) that end-users are typically net buyers of out-of-the-money (OTM) options and net sellers of in-the-money (ITM) options.

In this section, we compare the main DISP measure with a sophisticated DISP measure that utilizes only the end-users' buy-side volume of OTM options and the end-users' sell-side volume of ITM options (DISP_SIGNED). In essence, for the construction of DISP_SIGNED we retain only OTM call purchases and ITM put sales, which are undoubtedly optimistic trades related to positive expectations, and OTM put purchases and ITM call sales, which are undoubtedly pessimistic trades related to negative expectations. To do so, we obtain signed volume data from the International Securities Exchange (ISE) Trade Profile. This dataset contains all end-users' trades disaggregated by whether each trade is a buy or a sell order. While the ISE options volume data represent about 30% of the total individual stock options trading volume across all options exchanges, Ge et al. (2016) show that the data are quite representative of the total options volume provided by OptionMetrics. Unfortunately, the ISE data are only available from May 2005 onwards, thus limiting considerably the period for which we can estimate DISP.

Figure 1 depicts DISP (red solid line) and DISP_SIGNED (black dashed line) for the period 2005:05-2015:08, both standardized to have zero mean and variance equal to one. It can be seen that DISP behaves very similarly to DISP_SIGNED. In fact, the correlation between the two variables is 95%. The above evidence shows that a simple DISP measure that uses unsigned volume data captures almost exactly the same information with a sophisticated DISP measure that uses signed volume data. Therefore, given the considerably longer sample period covered by OptionMetrics than by ISE Trade Profile, it is natural that our analysis is conducted with unsigned volume data. It is also important to note that, unlike signed volume data, daily unsigned volume data are publicly available and hence easily accessible to investors. This means that the trading strategy based on the out-of-sample predictive power of DISP would be relatively cheap and implementable by an investor in real time.

B Alternative DISP Measures

In this section, we provide in-sample, out-of-sample and economic significance results for eight alternative dispersion measures. In particular, DISP_{I} and DISP_{II} are constructed using the volume-weighted mean absolute deviation and interquartile range of moneyness levels, respectively. DISP_{III} follows conceptually the dispersion in beliefs measure of Diether et al. (2002) and utilizes the volume-weighted standard deviation of the strike prices scaled by the volume-weighted average strike price. For DISP_{IV} we remove options with moneyness between 0.975 and 1.025, while for DISP_{V} we use all available moneyness levels. DISP_{VI} employs only the end-of-month DISP value for each stock rather than the average value of the given month. Finally, DISP_{VII} uses only options that expire on the next standard expiration date,[1] while $\text{DISP}_{\text{VIII}}$ uses only options that expire one month after the next available standard expiration date. This way our dispersion measure is estimated using always only options that have the same expiration date. The results provided in Tables 1 - 2 are very similar to those provided for the main DISP measure and show that the predictive power of the dispersion in options investors' expectations is robust to alternative specifications and filtering rules.

C Description of the Bootstrap Method

This section describes the wild bootstrap procedure for computing the empirical p-values. Similar procedures are also followed by Neely et al. (2014) and Huang et al. (2015).

We begin by estimating the error terms from a regression of the future market return on the set of predictors used in the study:

$$\widehat{\varepsilon}_{t,t+h} = re_{t,t+h} - \left(\widehat{\alpha}_h + \widehat{\boldsymbol{\beta}}_h' \mathbf{z}_t\right), \tag{1}$$

where $re_{t,t+h}$ is the h-month excess market return, $\mathbf{z}_t$ is the vector of predictors and $\widehat{\alpha}_h$ and $\widehat{\boldsymbol{\beta}}_h'$ are the estimated OLS parameters.

Following convention, each predictor i included in the model of equation (1) is assumed to follow

[1] A standard maturity option is one that expires on the third Friday of a given month.

an AR(1) process:

$$z_{i,t+1} = \rho_{i,0} + \rho_{i,1} z_{i,t} + \phi_{i,t+1}. \tag{2}$$

For each predictor i define:

$$\widehat{\phi}^{c}_{i,t+1} = z_{i,t+1} - \left(\widehat{\rho}^{c}_{i,0} + \widehat{\rho}^{c}_{i,1} z_{i,t}\right), \tag{3}$$

where $\widehat{\rho}^{c}_{i,0}$ and $\widehat{\rho}^{c}_{i,1}$ are reduced-bias estimates of the respective AR(1) parameters in (2). The reduced-bias estimates are computed by iterating on the analytical second-order bias expression for the OLS estimates.

Using these reduced-bias AR(1) parameters and the fitted error terms from (1) and (3) we build up a pseudo-sample for the excess market return under the null of no predictability and for each of the predictive variables:

$$\widetilde{re}_{t,t+h} = \overline{re}_{t,t+h} + \widehat{\varepsilon}_{t,t+h} w_{t+1}, \tag{4}$$

$$\widetilde{z}_{i,t+1} = \widehat{\rho}^{c}_{i,0} + \widehat{\rho}^{c}_{i,1} \widetilde{z}_{i,t} + \widehat{\phi}^{c}_{i,t+1} w_{t+1}, \tag{5}$$

where $\overline{re}_{t,t+h}$ is the sample mean of the market excess return, w_{t+1} is a draw for the standard normal distribution and $\widetilde{z}_{i,t}$ for $t = 0$ is the initial value $z_{i,0}$ for each predictor. By multiplying $\widehat{\varepsilon}_{t,t+h}$ and each predictor's $\widehat{\phi}^{c}_{i,t+1}$ with the same draw from the standard normal distribution w_{t+1}, we are able to account for the cross-correlation between the market returns and the innovations in the predictive variables (Stambaugh, 1999) as well as for general forms of conditional heteroskedasticity. In addition, the reduced-bias AR(1) estimates ensure that the high persistence of several predictive variables is properly captured.

Finally, for each regression model examined in the paper, we estimate the Newey-West t-statistics using the equity premium and appropriate predictor time-series from the constructed pseudo-sample. By repeating the process 2,000 times, we obtain an empirical distribution for each of the t-statistics. The empirical p-value for each predictor in each regression model is the proportion of the bootstrapped t-statistics that exceed in absolute terms the respective Newey-West t-statistic from the original sample.

D Alternative Market-Timing Strategies

In the main paper, we present the results from a market-timing strategy that utilizes mean-variance portfolio weights. In this section, we provide additional results from a strategy that utilizes binary weights. In particular, we consider two scenarios: one where short-sales are not allowed and one where short-sales are allowed. In the first scenario, the investor allocates 100% of her wealth in the market index (risk-free asset) every time the predicted equity premium is positive (negative). In the second scenario, the investor allocates 150% of her wealth in the market index (risk-free asset) and –50% in the risk-free asset (market index) every time the predicted equity premium is positive (negative).

The historical average (HAV) strategy delivers a Sharpe ratio ranging from 0.48 to 0.51 when short-sales are not allowed and from 0.45 to 0.48 when short-sales are allowed. It can be seen that DISP clearly outperforms the HAV strategy in all scenarios since it provides higher Sharpe ratios and positive ΔCER values. Moreover, unlike the case of the mean-variance strategy discussed in the main paper, in the case of binary strategies AFD performs worse than the historical average delivering lower Sharpe ratios and negative ΔCER values.

E Sample without the financial crisis

In this section, we examine whether the strong predictive power of DISP for the equity premium is driven by the financial crisis period. In particular, we consider the pre-crisis period (1996:01-2008:06), as well as the whole sample period excluding the financial crisis (1996:01-2008:06 and 2009:07-2017:12). Tables 4 and 5 provide the results, while Figure 2 plots the respective cumulative square prediction error differences.

Overall, DISP continues to be a strong predictor of future market returns even when the financial crisis period is not taken into consideration. One exception is the 1-month ahead predictability in the pre-crisis period where the results are somewhat weaker. However, a closer inspection of Figure 2 reveals that even in this case DISP consistently outperforms the historical average model across the majority of the months. The low statistical significance is driven by one observation, namely DISP in 1998:07 predicting the return of 1998:08. This observation corresponds to the outburst of the Russian financial crisis, when the US stock market experienced a monthly return of almost –16% (the second lowest in our sample after the return of 2008:10). By removing this one observation from the pre-crisis sample, DISP becomes significant at the 5% level with a t-stat of –2.33 and a p-value from the wild bootstrap experiment of 0.024.

F Time-series plots of DISP versus ASYM and AFD versus AFA

In the main paper, we report the correlations between the dispersion measures under examination (DISP and AFD) and the corresponding asymmetry measures (ASYM and AFA). Figure 3 compares the time-series pattern of DISP versus ASYM, and AFD versus AFA. It is evident that after 2007 AFD and AFA tend to move in opposite directions.

References

[1] Clark, T. and M. McCracken. 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85-110.

[2] Clark, T. and K. West. 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

[3] Diether, K. B., C. J. Malloy and A. Scherbina. 2002. Differences of opinion and the cross section of stock returns. Journal of Finance 57, 2113-2141.

[4] Ge, L., T. Lin and N. D. Pearson. 2016. Why does the option to stock volume predict stock returns? Journal of Financial Economics 120, 601-622.

[5] Hu, J. 2014. Does option trading convey stock price information? Journal of Financial Economics 111, 625-645.

[6] Huang, D., F. Jiang, J. Tu and G. Zhou. 2015. Investor sentiment aligned: a powerful predictor of stock returns. Review of Financial Studies 28, 791-837.

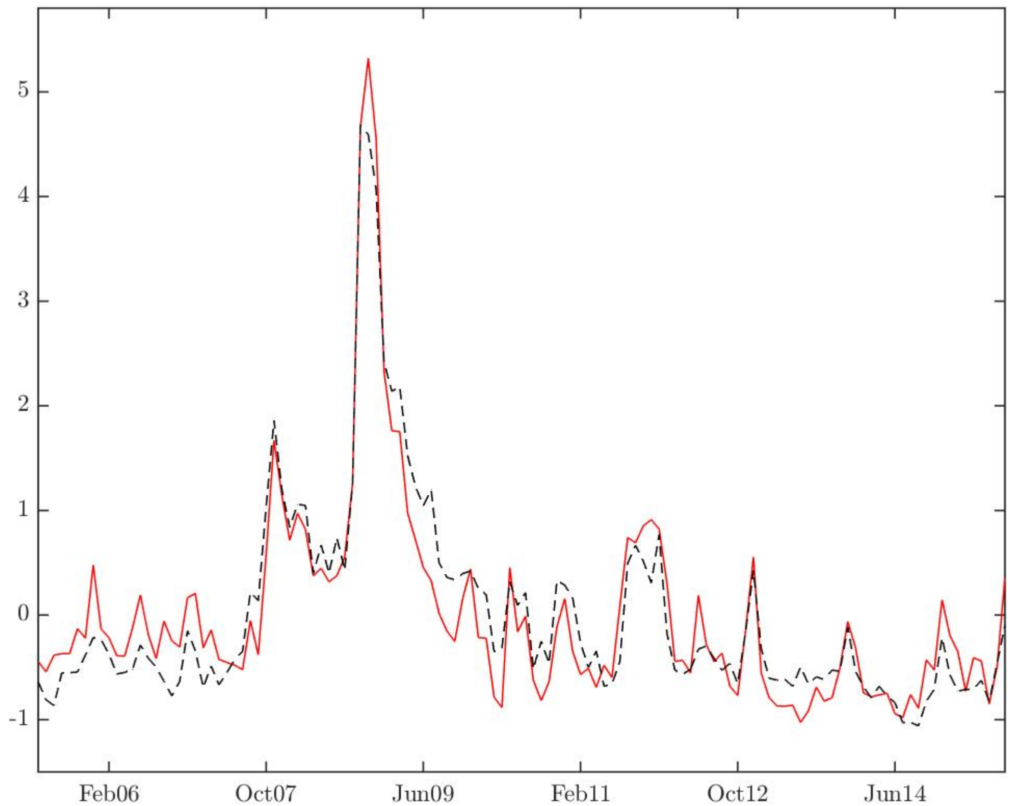
[7] McCracken, M. 2007. Asymptotics for out of sample tests of Granger causality. Journal of Econometrics 140, 719-752.

[8] Newey, W. and K. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

[9] Neely, C. J., D. E. Rapach, J. Tu and G. Zhou. 2014. Forecasting the equity risk premium: the role of technical indicators. Management Science 60, 1772–1791.

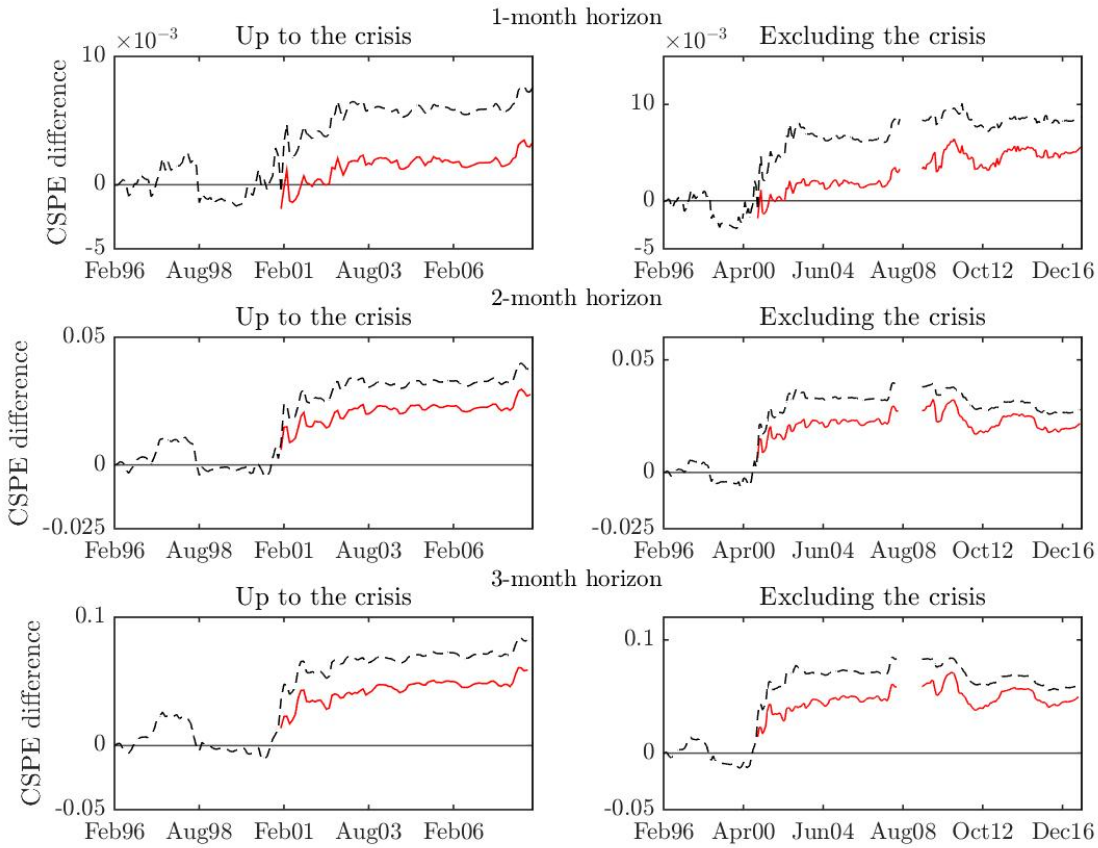
[10] Stambaugh, R. 1999. Predictive regressions. Journal of Financial Economics 54. 375-421.

Figure 1: DISP versus DISP_SIGNED



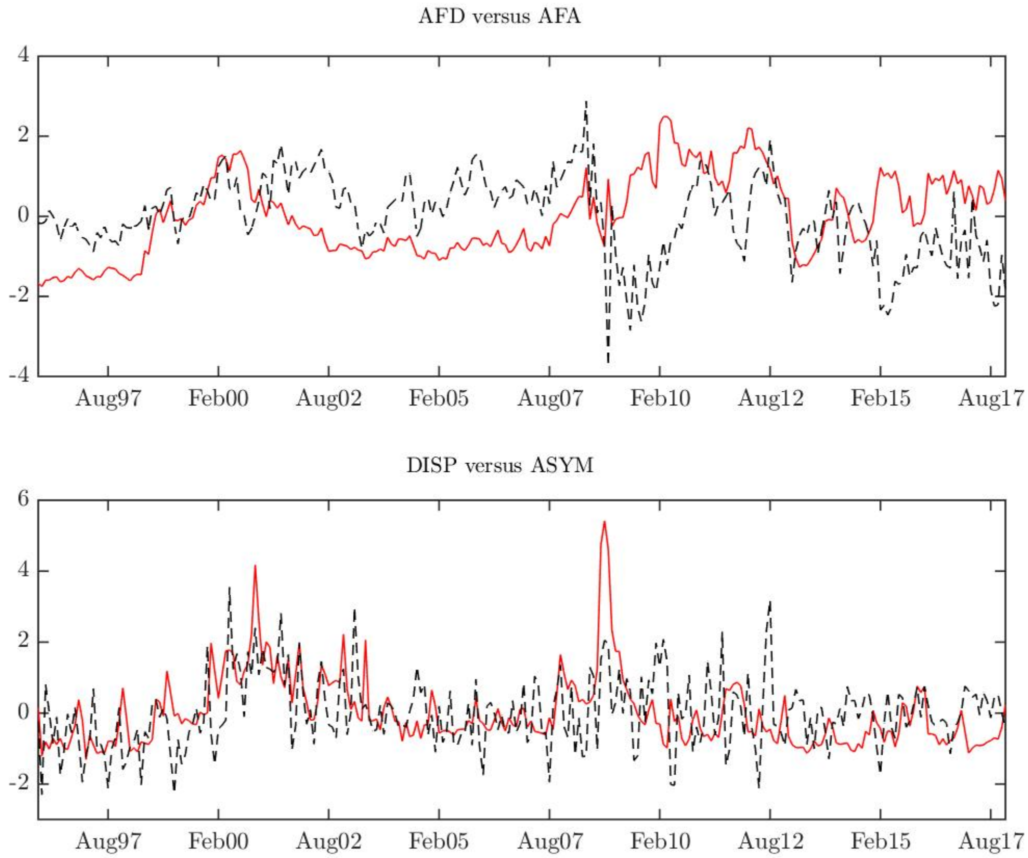
This figure plots the monthly time series of the dispersion in options investors' expectations (DISP, red solid line) versus the dispersion in options investors' expectations with signed volume data (DISP_SIGNED, black dashed line) for the period 2005:05-2015:08. Both variables have been standardized to have zero mean and variance one.

Figure 2: Differences in cumulative squared prediction error for DISP without the financial crisis



This figure plots the differences between the cumulative squared prediction error (CSPE) of the historical average model and the model based on the dispersion in options investors' expectations (DISP). The forecasting horizons are one, two and three months ahead. The black dashed lines correspond to the CSPE from the in-sample analysis, while the red solid lines correspond to the CSPE from the out-of-sample analysis. In the left panels the sample period is 1996:01-2008:06. In the right panels the sample spans the periods 1996:01-2008:06 and 2009:07-2017:12. The out-of-sample period begins in 2001:01 in both cases.

Figure 3: DISP versus ASYM and AFD versus AFA



The top panel of this figure plots the monthly time series of AFD (red solid line) versus AFA (black dashed line). The bottom panel plots the monthly time series of DISP (red solid line) versus ASYM (black dashed line). Both variables have been standardized to have zero mean and variance one. The sample period is 1996:01-2017:12.

Table 1: In-sample predictive power of alternative DISP measures

	1-month horizon		2-month horizon		3-month horizon	
Predictor	$\widehat{\beta}$	R^2 (%)	$\widehat{\beta}$	R^2 (%)	$\widehat{\beta}$	R^2 (%)
DISP_{I}	–10.06	3.47	–8.71	4.66	–8.16	6.06
	(–2.58)**		(–2.91)***		(–2.79)**	
DISP_{II}	–10.14	3.52	–8.49	4.43	–8.09	5.96
	(–2.66)**		(–2.88)**		(–2.75)**	
DISP_{III}	–8.84	2.68	–7.69	3.63	–6.67	4.06
	(–2.35)**		(–2.46)**		(–2.21)**	
DISP_{IV}	–9.91	3.37	–8.63	4.58	–7.96	5.77
	(–2.53)**		(–2.83)**		(–2.66)**	
DISP_{V}	–9.50	3.09	–8.36	4.29	–7.90	5.69
	(–2.37)**		(–2.61)**		(–2.51)**	
DISP_{VI}	–9.18	2.89	–7.42	3.38	–7.44	5.04
	(–2.55)**		(–2.88)***		(–2.97)***	
DISP_{VII}	–10.31	3.64	–8.63	4.58	–7.43	5.02
	(–2.52)**		(–2.66)**		(–2.60)**	
$\text{DISP}_{\text{VIII}}$	–7.50	1.93	–7.15	3.14	–7.25	4.79
	(–1.97)*		(–2.34)**		(–2.52)**	

This table reports the in-sample results for the predictive regressions of the CRSP value-weighted index excess return on alternative dispersion in options investors' expectations (DISP) estimates. DISP_{I} uses the mean absolute deviation, DISP_{II} uses the interquartile range, DISP_{III} uses the standard deviation of normalized strike prices, DISP_{IV} removes options with moneyness between 0.975 and 1.025, DISP_{V} uses options from all available moneyness levels, DISP_{VI} uses only end-of-month values, DISP_{VII} uses only options of the next available standard expiration date and $\text{DISP}_{\text{VIII}}$ uses only options that expire one month after the next available standard expiration date. The sample period is 1996:01-2017:12. Reported coefficients indicate the percentage annualized excess return resulting from a one standard deviation increase in each predictor variable. Newey and West (1987) t-statistics with lag length equal to the forecasting horizon are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels based on a wild bootstrap experiment.

Table 2: Out-of-sample predictive power of alternative DISP measures

	DISP_{I}	DISP_{II}	DISP_{III}	DISP_{IV}	DISP_{V}	DISP_{VI}	DISP_{VII}	$\text{DISP}_{\text{VIII}}$
				1-month horizon				
R^2_{OS} (%)	3.04	3.22	2.17	2.66	2.24	2.54	2.97	1.34
MSE-F	6.41**	6.80**	4.52**	5.58**	4.67**	5.31**	6.24**	2.78**
ENC-NEW	7.53**	7.81**	5.72**	7.72**	7.21**	5.64**	7.86**	4.02**
MSE-Adj	1.80**	1.73**	1.58*	1.76**	1.65**	2.04**	1.82**	1.44*
SR	0.99	0.97	0.88	0.99	0.95	0.84	0.98	0.84
ΔCER (%)	7.08	6.73	5.88	7.39	6.99	5.64	7.68	5.53
				2-month horizon				
R^2_{OS} (%)	4.34	4.03	3.09	4.24	3.72	3.63	3.74	2.82
MSE-F	9.17**	8.48**	6.44**	8.94**	7.81**	7.62**	7.86**	5.85**
ENC-NEW	12.22**	11.73**	9.26**	12.37**	11.90**	8.05**	10.95**	7.97**
MSE-Adj	1.64*	1.60*	1.41*	1.59*	1.51*	1.77**	1.66**	1.45*
SR	0.82	0.84	0.69	0.80	0.78	0.59	0.72	0.77
ΔCER (%)	6.31	6.35	4.92	6.30	6.07	3.87	5.84	5.70
				3-month horizon				
R^2_{OS} (%)	5.06	4.81	2.74	4.54	4.10	4.26	3.85	3.28
MSE-F	10.65**	10.11**	5.64**	9.52**	8.55**	8.89**	8.00**	6.78**
ENC-NEW	15.45**	15.65**	9.76**	14.80**	15.10**	12.22**	11.72**	12.20**
MSE-Adj	1.59*	1.51*	1.28	1.53*	1.49*	1.64*	1.64*	1.55*
SR	0.77	0.80	0.63	0.79	0.80	0.53	0.65	0.75
ΔCER (%)	5.91	6.15	4.22	6.12	6.39	3.26	4.85	5.81

This table reports the results of out-of-sample predictability of the CRSP value-weighted index excess return. The total sample period is 1996:01-2017:12 and the forecasting period begins in 2001:01. The forecasting variables are alternative estimates of the dispersion in options investors' expectations (DISP). DISP_{I} uses the mean absolute deviation, DISP_{II} uses the interquartile range, DISP_{III} uses the standard deviation of normalized strike prices, DISP_{IV} removes options with moneyness between 0.975 and 1.025, DISP_{V} uses options from all available moneyness levels, DISP_{VI} uses only end-of-month values, DISP_{VII} uses only options of the next available standard expiration date and $\text{DISP}_{\text{VIII}}$ uses only options that expire one month after the next available standard expiration date. R^2_{OS} is the out-of-sample coefficient of determination, MSE-F is the McCracken (2007) F-statistic, ENC-NEW is the encompassing test of Clark and McCracken (2001) and MSE-Adj is the MSE-Adjusted statistic of Clark and West (2007). ** and * denote significance at the 5% and 10% levels. The critical values for the MSE-F test are 1.518 and 0.616, respectively, while the critical values for the ENC-NEW test are 2.374 and 1.442, respectively. These critical values are based on Monte-Carlo simulations and are provided by the respective studies. We also report the annualized Sharpe ratio (SR) and certainty equivalent return (ΔCER) of a market-timing strategy that is based on each of the predictive models and utilizes mean-variance weights. The benchmark strategy follows the historical average model.

Table 3: Binary market-timing strategies

	DISP	AFD	VRP	TAIL	d-p	e-p	d-e	TERM	DEF	RREL
	Without Short-Sales – 1-month horizon									
SR	0.77	0.37	0.61	0.55	0.53	0.73	0.78	0.39	0.70	0.59
ΔCER (%)	2.50	–1.96	0.74	0.18	0.29	2.50	2.72	–1.70	1.78	0.93
	Without Short-Sales – 2-month horizon									
SR	0.68	0.44	0.61	0.47	0.46	0.64	0.70	0.39	0.72	0.54
ΔCER (%)	2.24	–0.76	1.34	–0.12	–0.28	1.98	2.32	–1.30	2.32	0.75
	Without Short-Sales – 3-month horizon									
SR	0.63	0.34	0.65	0.45	0.50	0.62	0.51	0.38	0.78	0.59
ΔCER (%)	1.73	–2.04	1.84	–0.66	0.24	1.65	0.20	–1.68	2.66	1.31
	With Short-Sales – 1-month horizon									
SR	0.73	0.24	0.53	0.47	0.50	0.72	0.75	0.29	0.65	0.55
ΔCER (%)	5.02	–3.94	1.46	0.35	0.59	5.03	5.47	–3.43	3.56	1.85
	With Short-Sales – 2-month horizon									
SR	0.64	0.32	0.55	0.43	0.42	0.61	0.65	0.31	0.66	0.49
ΔCER (%)	4.51	–1.57	2.67	–0.26	–0.58	3.99	4.66	–2.64	4.65	1.50
	With Short-Sales – 3-month horizon									
SR	0.59	0.21	0.61	0.39	0.48	0.59	0.45	0.29	0.71	0.55
ΔCER (%)	3.48	–4.16	3.71	–1.36	0.49	3.33	0.37	–3.43	5.34	2.63

This table reports the results of market-timing strategies with binary weights based on the out-of-sample predictability of the CRSP value-weighted index excess return. The total sample period is 1996:01-2017:12 and the forecasting period begins in 2001:01. The forecasting variables are the dispersion in options investors' expectations (DISP), analysts' forecasts dispersion (AFD), variance risk premium (VRP), tail risk (TAIL), dividend-price ratio (d-p), earnings-price ratio (e-p), dividend payout ratio (d-e), yield term spread (TERM), default spread (DEF) and relative short-term risk-free rate (RREL). SR stands for the annualized Sharpe ratio and ΔCER is the certainty equivalent return in excess of the historical average strategy.

Table 4: In-sample predictive power of DISP without the financial crisis

	1-month horizon		2-month horizon		3-month horizon	
Predictor	$\widehat{\beta}$	R^2 (%)	$\widehat{\beta}$	R^2 (%)	$\widehat{\beta}$	R^2 (%)
Sample period ending before the financial crisis						
DISP	–8.57 (–1.85)*	2.57	–9.63 (–2.95)***	6.24	–9.48 (–3.50)***	9.30
Total sample period excluding the financial crisis						
DISP	–7.09 (–2.07)**	2.07	–6.38 (–2.37)**	3.34	–6.30 (–2.64)**	5.09

This table reports the in-sample results for the predictive regressions of the CRSP value-weighted index excess return on the dispersion in options investors' expectations (DISP). In the top panel the sample period is 1996:01-2008:06. In the bottom panel the sample spans the periods 1996:01-2008:06 and 2009:07-2017:12. Reported coefficients indicate the percentage annualized excess return resulting from a one standard deviation increase in each predictor variable. Newey and West (1987) t-statistics with lag length equal to the forecasting horizon are reported in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% levels based on a wild bootstrap experiment.

Table 5: Out-of-sample predictive power of DISP without the financial crisis

	DISP - Sample period ending before the financial crisis	DISP - Total sample period excluding the financial crisis
	1-month horizon	
R^2_{OS} (%)	2.15	1.98
MSE-F	1.98**	3.85**
ENC-NEW	1.95*	4.55**
MSE-Adj	1.50*	2.25**
SR	0.81	1.05
ΔCER (%)	7.79	6.53
	2-month horizon	
R^2_{OS} (%)	8.02	3.73
MSE-F	7.67**	7.28**
ENC-NEW	6.26**	9.04**
MSE-Adj	2.11**	2.04**
SR	0.80	0.89
ΔCER (%)	8.62	5.31
	3-month horizon	
R^2_{OS} (%)	11.64	5.92
MSE-F	11.33**	11.64**
ENC-NEW	9.03**	13.15**
MSE-Adj	2.61**	2.34**
SR	0.74	0.85
ΔCER (%)	7.64	4.71

This table reports the results of out-of-sample predictability of the CRSP value-weighted index excess return. In the left panel the sample period is 1996:01-2008:06. In the right panel the sample spans the periods 1996:01-2008:06 and 2009:07-2017:12. The forecasting period begins in 2001:01 in both cases. The forecasting variable is the dispersion in options investors' expectations (DISP). R^2_{OS} is the out-of-sample coefficient of determination, MSE-F is the McCracken (2007) F-statistic, ENC-NEW is the encompassing test of Clark and McCracken (2001) and MSE-Adj is the MSE-Adjusted statistic of Clark and West (2007). ** and * denote significance at the 5% and 10% levels. The critical values for the MSE-F test are 1.518 and 0.616, respectively, while the critical values for the ENC-NEW test are 2.374 and 1.442, respectively. These critical values are based on Monte-Carlo simulations and are provided by the respective studies. We also report the annualized Sharpe ratio (SR) and certainty equivalent return (ΔCER) of a market-timing strategy that is based on each of the predictive models and utilizes mean-variance weights. The benchmark strategy follows the historical average model.