The Relationships between Sentiment, Returns and Volatility

February 2005

Yaw-huei Wang
National Central University
yhwang@mgt.ncu.edu.tw

Aneel Keswani
Cass Business School
a.keswani@city.ac.uk

Stephen J Taylor[*]
Lancaster University
s.taylor@lancaster.ac.uk

[*] Corresponding author. Address: Department of Accounting and Finance, Lancaster University, Lancaster LA1 4YX, UK. Phone: +44 (0)1524 593624. Fax: +44 (0)1524 847321. We thank two anonymous referees, Roy Batchelor, Gordon Gemmill, Barbara Ostdiek and seminar participants at the 2004 EFMA Basel meetings for helpful comments. This work was commenced while all the authors were at Lancaster University.

The Relationships between Sentiment, Returns and Volatility

Abstract

Previous papers that test whether sentiment is useful for predicting volatility ignore whether lagged returns information might also be useful for this purpose. By doing so, these papers potentially overestimate the role of sentiment in predicting volatility. In this paper we test whether sentiment is useful for volatility forecasting purposes. We find that most of our sentiment measures are caused by returns and volatility rather than vice versa. In addition, we find that lagged returns cause volatility. All sentiment variables have extremely limited forecasting power once returns are included as a forecasting variable.

Key words: Causality; Investor Surveys; Market based sentiment measures; Realized volatility; Stock index returns

JEL classification: G12; G14

1. Introduction

Whilst earlier papers have underplayed the importance of noise traders, more recent analysis has discussed how such traders acting on a noisy signal, such as sentiment, can induce systematic risk and affect asset prices in equilibrium. For example, De Long, Shleifer, Summers and Waldmann (1990) demonstrate that if risk averse arbitrageurs know that prices may diverge further away from fundamentals before they converge closer, they may take smaller positions when betting against mis-pricing. Thus if such uninformed noise traders base their trading decisions on sentiment, then measures of it may have predictive power for asset price behavior.

Most papers that test whether sentiment can predict returns or volatility motivate the relationship through the role of noise traders who respond to changes in sentiment influencing subsequent returns and volatility. If this is in fact what happens in practice, then it might be possible to use sentiment to forecast returns and volatility.[1]

Causality must run from sentiment to market behavior if we accept the noise trader explanation. If we step back from the noise trader framework, however, and ask how sentiment might be generated, it is quite natural to expect that market behavior should influence sentiment. Evidence of this was found by Solt and Statman (1988) and Brown and Cliff (2004) who document that returns cause sentiment rather than vice versa. If returns have a strong impact on sentiment then it is also possible that volatility influences sentiment as well. If this is the case we might observe a much stronger link between sentiment and returns or volatility if we do not assume that sentiment is the causal variable. Thus it is clearly important to test for the direction of causality.

A failure to recognize the impact of market behavior on sentiment may also explain why all previous studies that test the predictive power of sentiment fail to include lagged volatility when predicting returns and omit lagged returns as an additional variable when predicting volatility. However, if sentiment responds to lagged volatility or lagged returns then it makes sense to include these variables to supplement any forecasting tests of sentiment. Doing so is likely to avoid overestimating the true forecasting power of sentiment.

We test these ideas at a market-wide level by first looking at whether aggregate sentiment measures cause the returns and the realized volatility of the S&P 100 index as predicted by the noise trader literature or whether sentiment simply responds to market behavior. In addition we test whether returns cause volatility. [2]After deciding on the variables that cause returns and volatility, we use these variables for forecasting. This allows us to determine the incremental contribution of sentiment for forecasting.

The analysis is conducted on both a daily and weekly basis. In the daily analysis, the sentiment indicators used include the S&P 100 (OEX) put-call trading volume ratio (PCV), the OEX put-call open interest ratio (PCO), and the NYSE ARMS index.[3] In the weekly analysis, the sentiment indicators used consist of PCO and PCV and two sentiment ratios gathered through surveys by two different investment information providers.[4] As a number of papers have found a significant relationship between changes in sentiment and returns or volatility, we investigate both sentiment and its first differences.

Overall it is found that all sentiment measures are Granger-caused by returns and that many measures of sentiment are also caused by realized volatility. We show that the one sentiment measure, the ARMS index, which appears to consistently Granger-cause volatility has only limited predictive power once returns are included.

This study makes two particular contributions. Firstly, it indicates that research that seeks to exploit the potential market impact of noise traders is unlikely to be successful for returns and volatility forecasting. Secondly, it clarifies the relationship between returns, sentiment and realized volatility. In particular our results show that it is returns rather than sentiment that contain useful information for volatility forecasting purposes.

This paper is arranged as follows. Section 2 discusses why sentiment, returns and realized volatility might be related and how this relationship might manifest itself. Section 3 presents the data and explains the choice of variables chosen to proxy for investor sentiment. The methods used to test whether market behavior causes investor sentiment or vice versa are presented in Section 4 with the results of these Granger causality tests. Section 5 presents the results of the volatility forecasting analysis. Finally, conclusions are stated in Section 6.

2. Theory & literature

De Long *et al* (1990) construct a model that explains why noise trader risk in financial markets is priced. They argue that whilst prices will revert to their fundamental values in the long term, this process may not be smooth and may take a long time. As a result, arbitrageurs can lose out if prices diverge further away from fundamentals before they get closer. Their model makes predictions about the relationship between sentiment and price volatility at the level of individual securities: more noise trading is associated with increased price volatility. Furthermore, sentiment will affect returns via its impact on volatility. If the signal that drives noise trading is sentiment, then we would expect to see a link between measures of sentiment and returns and volatility.

The precise form in which sentiment will affect returns or volatility is not clear ex ante. If noise traders are sensitive to sentiment changes, then sentiment changes should drive returns and volatility. Alternatively, if noise traders only trade if sentiment is extreme (either high or low) relative to previous levels, then it might be expected that it is sentiment levels that influence returns and volatility.

The predictive power of sentiment for returns has been explored in a number of papers. The results that have been found are mixed.

Neal and Wheatley (1998), Wang (2001) and Simon and Wiggins (2001) find that sentiment can predict returns. Neal and Wheatley (1998) find that two measures of individual investor sentiment, one compiled from the discounts on closed-end funds and the other redemptions of mutual funds, predict equity returns. Wang (2001) uses the positions held by large traders in the futures markets as a proxy for sentiment and discovers that they are useful for predicting the returns on futures in a subsequent period. Simon and Wiggins (2001) also find that sentiment measures are able to predict returns on futures.

However, not all papers that have studied the relationship between sentiment and returns have come to these conclusions. Fisher and Statman (2000) find that the causality between equity returns and sentiment can be significant in both directions. Brown and Cliff (2004) use a large number of sentiment indicators to investigate the relationship between sentiment and equity returns and find much stronger evidence that sentiment is caused by returns. Solt and Statman (1988) also make similar findings. Both papers tell us that returns may be important for sentiment determination.

A few papers have also investigated the relationship between sentiment and volatility. Brown (1999) looks at whether investor sentiment levels are related to the volatility of closed-end fund returns. As measures of sentiment he uses both investor survey data and closed-end fund discounts. His results show that deviations from the mean level of sentiment are positively and significantly related to volatility during trading hours.

Lee *et al* (2002) look at the relationship between volatility, returns and sentiment. They estimate a GARCH-in-mean model which includes contemporaneous shifts in investor sentiment in the mean equation and lagged shifts in sentiment in the conditional volatility equation. They use the survey indicator provided by Investor's Intelligence to examine the impact of changes in investor sentiment on the conditional volatilities of the DJIA, S&P 500, and NASDAQ indices, which are estimated from the GJR-GARCH model. They find that bullish (bearish) changes in sentiment result in downward (upward) adjustments in volatility.

In summary therefore the literature tells us that sentiment may be useful for forecasting volatility. It also tells us that this relationship may be influenced by the behavior of returns. In our empirical analysis we do two things. Firstly, we examine the causality relationship between returns, sentiment and volatility. Secondly, we examine whether sentiment measures are useful for forecasting returns and volatility. In contrast to previous studies our analysis uses realized volatility rather than a latent volatility measure estimated using a time series model.

3. Data

The sample period used for daily data is from 1 February 1990 until 31 December 2001. Results are obtained for the full period and also for two sub-samples given by dividing the sample period into two equal parts around 11 January 1996. This allows us to assess whether the results are robust through time. The weekly data is for the slightly shorter period from 6 April 1990 to 28 December 2001 and it is also divided into two equal sub-samples.

We study measures of realized volatility, returns and indicators of market participants' sentiment at the daily and weekly frequencies. The methods used to gather this data and the measures of sentiment used are explained in this section.

3.1. Realized volatility

Andersen and Bollerslev (1998) show that the squared return can be a highly noisy measure of the realized variance of a financial asset's return. However they also show that using the cumulative sum of high-frequency squared intraday returns can greatly mitigate the noisy component[5]. Five-minute S&P 100 index returns are used to calculate a measure of daily realized volatility in this paper, as the 5-minute frequency provides the best measure in Andersen and Bollerslev (1998). The latest observations available before 5-minute marks from 9:30 EST until 16:00 EST are used to calculate 5-minute returns. To construct the measure of daily realized variance, we sum the 78 squared intraday 5-minute returns and the previous squared overnight return. For the weekly realized variance we average the daily realized variances by the number of trading days in the week. This procedure is used to avoid the bias induced by variations in the number of trading days in a week.

3.2. Sentiment indicators

The sentiment indicators used are different for daily and weekly returns due to data availability. The daily sentiment indicators used consist of the OEX put-call trading volume ratio, the OEX put-call open interest ratio and the NYSE ARMS index. Whilst the put-call trading volume and open interest ratios are available on a weekly basis,[6] the ARMS index is not collated at the weekly frequency. As a result a weekly ARMS measure is not used in the analysis. This study does however use two additional sentiment indicators available on a weekly basis that are compiled from surveys by the AAII (American Association for Individual Investors) and II (Investor Intelligence).

3.2.1. Put-call trading volume and open interest ratios

The put-call trading volume ratio (PCV) is a measure of market participants' sentiment derived from options and equals the trading volume of put options divided by the trading volume of call options. When market participants are bearish, they buy put options either to hedge their spot positions or to speculate bearishly. Therefore, when the trading volume of put options becomes large in relation to the trading volume of call options, the ratio goes up, and vice versa.

Another measure of the put-call volume ratio can be calculated using the open interest of options instead of trading volume. This ratio can be calculated on a daily basis using the open interest of options at the end of the day or on a weekly basis using the open interest of options at the end of the week. This might be a preferred measure of sentiment as it may be argued that the open interest of options is the final picture of sentiment at the end of the day or the week and is therefore likely to have better predictive power for volatility in subsequent periods. This measure of sentiment is therefore used as well. The put-call ratio calculated in this way is labeled the PCO ratio.

3.2.2. ARMS index

The ARMS index on day *t* is equal to the number of advancing issues scaled by the trading volume (shares) of advancing issues divided by the number of declining issues scaled by the trading volume (shares) of declining issues. It is calculated as:

$$ARMS_t = \frac{\#Adv_t \,/\, AdvVol_t}{\#Dec_t \,/\, DecVol_t} = \frac{DecVol_t \,/\#Dec_t}{AdvVol_t \,/\#Adv_t}$$

where $\#Adv_t$, $\#Dec_t$, $AdvVol_t$, and $DecVol_t$ respectively denote the number of advancing issues, the number of declining issues, the trading volume of advancing issues, and the trading volume of declining issues.

ARMS can be interpreted as the ratio of volume per declining issue to the volume in each advancing issue. If the index is greater than one, more trading is taking place in declining issues whilst if it is less than one more volume in advancing stocks outpaces the volume in each declining stock. Its creator, Richard Arms, argued that if the average volume in declining stocks far outweighs the average volume in rising stocks then the market is oversold and that this should be treated as a bullish sign. Likewise he argued that if the average volume in rising stocks far outweighs the average volume in falling stocks then the market is overbought and that this should be treated as a bearish sign. [7]

3.2.3. AAII and II ratios

Surveys of the bullishness or bearishness of investors provide an alternative way to measure investor sentiment.

The American Association for Individual Investors (AAII) has conducted a sentiment survey by polling a random sample of its members each week since 1987. The respondents are asked whether they are bullish, bearish, or neutral about the future condition of the stock market in six months. Only subscribers to AAII are eligible to vote and they can only vote once during the survey period[8]. As the respondents to this survey are individuals, this can be interpreted as a measure of individual sentiment. The ratio of the bearish percentage to the bullish percentage is used as a measure of investor sentiment in this paper. [9]

Investor Intelligence (II) has compiled its sentiment data weekly by categorizing approximately 150 market newsletters since 1964. Newsletters are read and marked starting on Friday each week. The results are reported as percent bullish, bearish, or neutral on the following Wednesday.[10] Since many of the writers of these newsletters are current or past market professionals, the ratio of bullish to bearish responses compiled by II can be considered as a proxy of institutional investors' sentiment.[11]

3.3. Summary Statistics

Table 1 contains summary statistics of all the variables discussed in this section.[12] The statistics are presented for the full period and for two sub-periods of equal duration. The daily series of log realized volatility has high autocorrelations with a first-lag correlation of 0.73 for the full period. The weekly series of log realized volatility has a similar distribution to the daily series but has less kurtosis. Both daily and weekly returns display excess kurtosis, negative skewness and almost no serial correlation.

The levels of all the sentiment indicators display a skewed and leptokurtic pattern, whilst the first differences of all the indicators are also skewed and most are leptokurtic. All levels of sentiment indicators, except the ARMS index, have substantial positive autocorrelations, while the first differences have significant first-lag autocorrelations that are negative except for the II ratio. [13]

Table 2 contains the contemporaneous correlations between the sentiment measures and the other variables, namely returns and realized volatility. We find that ARMS has a substantial negative correlation with returns, between –0.7 and –0.8, for all periods considered. ARMS also has a small positive correlation with log realized volatility. The correlations are reduced when the first differences of ARMS are used.

As regards the put-call volume ratios, we find that they are more correlated with returns than volatility. The correlations between the volume ratio, PCV, and returns are more substantial for the daily frequency than the weekly frequency and they are similar for either the level or the change in PCV. The open interest ratio, PCO, has more substantial correlations for the weekly frequency than the daily frequency, that are negative with returns and positive with volatility.

There is evidence of non-negligible correlation between our survey based measures of sentiment and both returns and realized volatility. Small correlations are observed for both the levels and the first differences of the survey variables.

4. Granger causality tests

4.1. Methodology

On the way to investigate the predictive power of sentiment for returns and realized volatility, first we run Granger causality tests to determine whether there exists any Granger-causality relationship among them. The results are given in this section. Then, in the next section, we try to discover if the sentiment measures that have a causal effect can be used for forecasting purposes. This requires us to look more deeply at the relationship between returns, volatility and sentiment.

We test for Granger causality between sentiment and returns by estimating bivariate VAR models. We test for causality in both directions. We also test for causal relationships between sentiment and realized volatility, and between returns and volatility.

To decide whether or not sentiment causes returns we estimate two models, one restricted and the other unrestricted. For the restricted model we regress returns on lagged values of returns alone. For the unrestricted model we regress returns on lagged values of returns and lagged values of sentiment. A standard likelihood ratio is used to see whether we have significant evidence to reject the restricted form of the model, i.e. whether we have evidence to reject the null hypothesis that sentiment does not cause returns. We use an identical methodology to decide if returns do or do not cause sentiment. The same test procedure is also employed to test for the causality relationships between sentiment and realized volatility, and between returns and realized volatility.

The degrees-of-freedom of the LR test depend on the number of lags used in the vector autoregressions. To determine the appropriate number of lags, we optimize the Akaike Information Criterion. The optimal number of lags depends on the pair of variables used in the causality tests; it varies between 2 and 12 for the daily data and between 2 and 6 for the weekly data.

4.2. Results

The results of the Granger causality tests using sentiment measures and returns are presented in Table 3. There is very limited evidence that sentiment, however measured, Granger-causes returns at either the daily or the weekly frequency. However, we find strong and consistent evidence that all sentiment measures, in levels and first differences, are Granger-caused by returns; all the likelihood ratio statistics are significant at the 1% level for the full sample. Thus, we find stronger evidence that sentiment measures are not causal variables but are in fact the variables being caused. These results confirm the findings of Brown and Cliff (2004) who also show that returns cause sentiment.

The results of the Granger causality tests using sentiment measures and volatility are presented in Table 4. First, consider the daily data. There is no significant evidence that the levels or first differences of either PCV or PCO Granger-cause realized volatility. However, there is compelling evidence that the levels and first differences of these sentiment measures are caused by realized volatility, with all four likelihood ratios significant at the 1% level. ARMS produces very different results to all the other sentiment measures. There is significant evidence of two-way causality, with stronger evidence for causality running from sentiment to volatility than from volatility to sentiment. Next, consider the weekly data. The null hypothesis that sentiment does not cause volatility is accepted for the AAII, PCV and PCO variables at the 10% level (full sample) but is rejected at the 1% level by the II survey variable. However, there is again much more evidence for causality in the other direction: the null hypothesis that volatility does not cause sentiment is rejected at the 1% level for the AAII, II and PCO variables (full sample). [14]

We have also tested for causal relationships between returns and realized volatility. The results of these tests are presented in Table 5 and show that returns strongly Granger-cause volatility rather than vice versa.

Three conclusions can be drawn from Tables 3, 4 and 5, with causality defined by Granger's methodology. Firstly, sentiment does not cause returns but rather returns cause sentiment. Secondly sentiment variables apart from ARMS do not consistently cause realized volatility. Our findings suggest that most of the sentiment measures used here should not be used for realized volatility forecasting purposes. All sentiment measures apart from ARMS appear to be caused by realized volatility. Thirdly returns cause realized volatility.

5. Tests of the forecasting power of ARMS for realized volatility

Two of the most frequently used variables for forecasting realized volatility are historical volatility and volatilities implied from options. Numerous papers, surveyed by Poon and Granger (2003), have examined the forecasting power of these variables. The general conclusion of papers such as Fleming (1998), Christensen and Prabhala

(1998) and Blair *et al* (2001) is that both implied and lagged volatility have considerable forecasting power, with implied volatility being the more accurate predictor.

To see whether ARMS could be a useful forecasting variable we therefore decided to examine whether it could enhance forecasts of the realized volatility of S&P 100 index returns that are computed from either lagged realized volatility or implied volatility represented by the VIX index.[15] We therefore estimated two benchmark regressions for logarithmic variables, one that used five lags of lagged realized volatility[16] to forecast realized volatility and another that used lagged implied volatility to forecast realized volatility. After estimating these benchmark models we investigated whether adding the level of ARMS and then its first differences can enhance forecasting power. In the first difference regressions, two dummy variables are used to see whether positive and negative changes in ARMS might have an asymmetric impact on realized volatility.

The following equation is estimated when the level of ARMS is included in the benchmark model with lagged volatility:

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \gamma ARMS_{t-1} + \varepsilon_t \tag{1}$$

where V_t is the realized volatility at time t. When VIX is used instead as the benchmark forecast the following equation is estimated:

$$LogV_t = K + \beta LogVIX_{t-1} + \gamma ARMS_{t-1} + \varepsilon_t \,. \tag{2}$$

When the first differences of ARMS are included, the regression equation is specified as:

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t , \qquad (3)$$

for the case where lagged realized volatility defines the benchmark forecast and where $D_{1,t-1}$ and $D_{2,t-1}$ are dummy variables that are respectively one if the change in ARMS is positive and one if the change in ARMS is negative. In the case where VIX defines the benchmark forecast, the regression equation is

$$LogV_t = K + \beta LogVIX_{t-1} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t . \qquad (4)$$

From Table 6, we find that ARMS does consistently enhance the benchmark models in a statistically significant manner. The null hypothesis that ARMS can not improve forecasts is rejected at the 1% level in all cases, although the increment in the adjusted R^2 is small (between 0.24% and 1.24% for the full sample). Thus it appears that ARMS does contain useful statistical information for forecasting purposes. The sign of the ARMS coefficient 'γ' in equations 1 and 2 indicates that as ARMS rises and the market becomes more bearish, future realized volatility rises. [17]

The leverage (or asymmetric volatility) effect is well documented in the volatility literature and describes the fact that as prices or returns fall volatility is more likely to rise. It is therefore possible that the relationship between ARMS and future realized volatility that we detect could be spurious and may merely be a consequence of the leverage effect, because ARMS reflects the market direction which is demonstrated by its correlation of –0.7 with returns as seen in Table 2. To assess the hypothesis of a spurious effect we added the S&P 100 return into our forecasting equations. These equations are formulated recognizing that the leverage effect relates volatility shocks to an asymmetric function of returns. For the case that uses the level of sentiment, we estimate the following equation:

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma ARMS_{t-1} + \varepsilon_t, \quad (5)$$

and for the case that includes the first differences of sentiment we estimate the specification

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t \quad (6)$$

with $D_{3,t-1}$ equal to one if r_{t-1} is negative and zero otherwise.

The results are shown in Table 7. We find that the adjusted R^2 of the benchmark models that contain returns are significantly higher than those that do not contain returns: the null hypothesis $\alpha = 0$ is always rejected at the 1% level. Furthermore, these R^2 values (for models that include returns) always exceed the corresponding values of R^2 in Table 6 (where ARMS replaces returns). The predictive power of ARMS becomes very limited when returns are included in the benchmark models. No matter in which form and with which benchmark variables, the incremental R^2 for the ARMS variable is then between 0.04% and 0.17% for the full sample. Although some of the coefficients of ARMS are still statistically significant in Table 7, for forecasting purposes the improvement made by ARMS is negligible and therefore of no economic significance. Nevertheless, we can conclude that a non-linear function of returns may enhance forecasts of realized volatility that are calculated from implied volatility and/or lagged realized volatility.

6. Concluding remarks

Risk managers and regulators are periodically required to forecast volatility whilst those working in the fund management industry frequently attempt to predict security returns. In this paper, we look at whether sentiment, measured using information from derivatives, spot markets and surveys, can be used to enhance these forecasts. In addition, recognizing that sentiment itself is affected by recent market behavior, we seek to determine the direction of any causal relationships.

Our analysis is conducted in two steps using equity market data. In the first step, we investigate the direction of causality between various measures of sentiment, returns and realized volatility to determine which of these variables might be useful for forecasting purposes. We find that most sentiment indicators, except ARMS, the ratio of the average volume of advancing versus declining issues, are caused by realized volatility rather than vice versa. We also detect that sentiment indicators are caused by returns and that returns predict realized volatility.

We test whether these causal relationships can be exploited in the second step by examining if ARMS and returns are of use for realized volatility forecasting purposes. As the commonly used benchmark models for predicting realized volatility use either lagged realized volatility or implied volatility, we test for the incremental predictive power of ARMS and returns to these benchmark models. We find that ARMS has predictive power for future realized volatility but that this is limited once returns are included. However, equity returns systematically improve the prediction of future realized equity market volatility.

Thus we do not observe a visible link between sentiment measures and realised volatility or returns as predicted by the theoretical literature. Our research design and results lend no support to the hypothesis that noise traders influence either returns or volatility.

To conclude, there is very limited evidence that sentiment, however measured, provides incremental information for forecasts of returns and volatility. Any such incremental information is unlikely to be economically significant. By contrast, all sentiment measures are caused by returns and volatility. Our results also indicate that returns may be useful in predicting realized volatility.

References

Andersen, T.G., & Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 885-905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., & Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics*, 61, 43-76.

Arms, R.W. (1989). *The Arms Index (TRIN): An introduction to the volume analysis of stock and bond markets*: McGraw-Hill Companies.

Blair, B.J., Poon, S., & Taylor, S.J. (2001). Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency index returns, *Journal of Econometrics*, 105, 5-26.

Brown, G.W. (1999). Volatility, sentiment, and noise traders, *Financial Analysts Journal*, 55, 82-90.

Brown, G.W., & Cliff, M.T. (2004). Investor sentiment and the near-term stock market, *Journal of Empirical Finance*, 11, 1-27.

Christensen, B.J., & Prabhala, N.R. (1998). The relation between implied and realized volatility, *Journal of Financial Economics*, 50, 125-150.

De Long, J.B., Shleifer, A., Summers, L.H., & Waldmann, R.J. (1990). Noise trader risk in financial markets, *Journal of Political Economy*, 98, 703-738.

Fisher, K.L., & Statman, M. (2000). Investor sentiment and stock returns, *Financial Analysts Journal*, 56, 16-23.

Fleming, J. (1998). The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance*, 5, 317-345.

Fleming, J., Ostdiek, B., & Whaley, R.E. (1995). Predicting stock market volatility: a new measure, *Journal of Futures Markets*, 15, 265-302.

Lee, W.Y., Jiang, C.X., & Indro, D.C. (2002). Stock market volatility, excess returns, and the role of investor sentiment, *Journal of Banking and Finance*, 26, 2277-2299.

Neal, R., & Wheatley, S.M. (1998). Do measures of sentiment predict returns? *Journal of Financial Quantitative Analysis*, 33 (4), 523-547.

Poon, S., & Granger, C.W.J. (2003). Forecasting financial market volatility: a review, *Journal of Economic Literature*, 41, 478-539.

Simon, D.P., & Wiggins III, R.A. (2001). S&P futures returns and contrary sentiment indicators, *Journal of Futures Markets*, 21, 447-462.

Solt, M.E., & Statman, M. (1988). How useful is the sentiment index? *Financial Analysts Journal*, 44, 45-55.

Wang, C. (2001). Investor sentiment and return predictability in agricultural futures markets, *Journal of Futures Markets*, 21, 929-952.

Endnotes:

[1] Forecasting realized volatility is important for a number of reasons. Firstly, the future behavior of realized volatility has an impact on current derivatives prices. Secondly, it is a required input for many models that calculate value at risk. For example, *Riskmetrics* requires a volatility estimate to calculate value at risk.

[2] Our paper focuses on sentiment measured at the aggregate level rather than the security specific level.

[3] The ARMS index is named after its creator Richard Arms and is defined in Section 3.

[4] We were unable to use a weekly measure of the ARMS index as the data is not compiled.

[5] Andersen and Bollerslev (1998) show that the more frequent the observations, the more accurate the measure in theory. It is impossible in reality to obtain a continuous dataset because of the discontinuities in the price process and the market microstructure effects such as bid-ask spreads and nonsynchronous trading effects.

[6] Weekly PCV is calculated as the sum of daily put trading volume over the week divided by the sum of daily call trading volume over the week. Weekly PCO is the open interest calculated on the last trading day of the week.

[7] The relationship between ARMS and whether the market is bearish or bullish may not be clear cut. Let us suppose the market has been falling broadly across the majority of stocks and ARMS has risen. It is only if market participants perceive that the level of the market has reached a low enough point that a recovery will follow and only then can ARMS be treated as a bullish measure. Before that point is reached, high trading volume in declining stock may simply be treated as a sign that the market will continue to fall.

[8] The average response rate of the AAII survey is about 50% with a standard deviation of 15%.

[9] AAII mails the questionnaires, and members fill them out and return them via US mail. Each week AAII collects responses from Friday to the following Thursday and reports the results on Thursday or Friday.

[10] In the case of both the AAII and the II measures, there is a time lag between responses and reporting. If we want to look at the true relationship between sentiment in week t and subsequent market behavior it might be argued that we should actually work with the AAII or II measures reported one or two weeks ahead to overcome this reporting lag. Whilst these measures reported in week $t+1$ or

week *t+2* might more accurately reflect sentiment at week *t*, market participants would not have such information to hand in week *t* to predict subsequent market behavior. Hence in our analysis of the forecasting role of sentiment that follows we do not temporally adjust our AAII or II measures.

[11] This point is made by Solt and Statman(1988).

[12] In our analysis we work with the logarithm of realized volatility as in log form it is much closer to being normally distributed than the original variable (Andersen *et al* (2001)).

[13] All the sentiment time series appear to be stationary, and all reject the unit root null hypothesis at the 1% level (augmented Dickey Fuller test, with four lags). Interestingly the ARMS index has a low level of autocorrelation and so appears to be close to a white noise process. Thus it is not surprising that the first-lag autocorrelation of the first differences is close to –0.5.

[14] In some of our sub samples the II index does significantly Granger-cause realized volatility. These results are similar to the results shown in Lee et al (2002). However as the causality is not consistent across all measurement periods we do not use II as a forecasting variable.

[15] For the implied volatility measure, the VIX index of Fleming *et al* (1995) is used. It is a weighted index of eight American option implied volatilities calculated from the closest to at-the-money call and put options of the two most nearby expiration months. These eight implied volatilities are weighted so that VIX represents the average implied volatility of an at-the-money option one month before expiration.

[16] Five lags were selected by optimizing the Akaike Information Criterion.

[17] As there was weak evidence that the II index Granger caused realized volatility in the first step of our analysis (which was not consistent across periods) we decided to run similar tests to those carried out above using the II index instead of ARMS. We found that the incremental adjusted R^2s of the II index when forecasting realized volatility range from -0.18% to 1.44% and most of them are less than 1%. Thus we conclude that the II index is not a reliable indicator for volatility forecasting purposes.

Table 1: Summary Statistics

This table shows summary statistics for the logarithm of Realised volatility (Log V), returns and various sentiment measures. These are PCV (the put-call volume ratio), PCO the put-call open interest ratio, the ARMS ratio and the survey based measures of the American Association for Individual Investors (AAII) and Investor Intelligence (II) defined in section 3. The full daily sample that contains 3005 observations from 1 February 1990 to 31 December 2001, sub-sample 1 that contains 1503 observations from 1 February 1990 to 11 January 1996, and sub-sample 2 that contains 1502 observations from 12 January 1996 to 31 December 2001. The symbol Δ represents the first difference.

Panel A: Daily Data									
Variable	Mean	Standard Deviation	Skewness	Kurtosis	Autocorrelation Lag 1	2	3	4	5
Log V									
Full sample	- 2.1363	0.2083	0.2936	3.3698	0.73	0.69	0.67	0.64	0.64
Sub-sample 1	- 2.2565	0.1516	0.6928	4.0401	0.51	0.46	0.41	0.37	0.36
Sub-sample 2	- 2.0160	0.1869	- 0.2175	5.7222	0.64	0.58	0.56	0.53	0.52
Returns									
Full sample	4.44e-4	0.0103	- 0.1854	6.5675	0.01	- 0.03	- 0.04	0.02	- 0.04
Sub-sample 1	4.17e-4	0.0075	- 0.0378	5.7471	0.00	- 0.02	- 0.06	0.03	- 0.01
Sub-sample 2	4.71e-4	0.0125	- 0.2050	5.3608	0.01	- 0.04	- 0.04	0.02	- 0.05
PCV									
Full sample	1.1820	0.2956	1.1053	5.6590	0.41	0.29	0.28	0.26	0.22
Sub-sample 1	1.1337	0.2416	0.7595	3.9536	0.47	0.31	0.31	0.26	0.26
Sub-sample 2	1.2303	0.3344	1.0355	5.2241	0.35	0.25	0.23	0.22	0.16
ΔPCV									
Full sample	3.63e-4	0.3212	- 0.0378	6.0137	- 0.40	- 0.10	0.01	0.01	- 0.04
Sub-sample 1	2.26e-4	0.2486	0.2293	4.7079	- 0.34	- 0.16	0.05	- 0.05	0.00
Sub-sample 2	4.99e-4	0.3803	- 0.1103	5.2721	- 0.42	- 0.07	0.00	0.04	- 0.05
PCO									
Full sample	1.2212	0.2609	1.1159	5.3453	0.90	0.85	0.81	0.77	0.74
Sub-sample 1	1.1903	0.2271	1.3049	6.0821	0.86	0.81	0.77	0.73	0.70
Sub-sample 2	1.2521	0.2875	0.9041	4.6644	0.93	0.88	0.83	0.79	0.75
ΔPCO									
Full sample	9.61e-5	0.1146	- 2.7782	42.0705	- 0.25	- 0.03	- 0.01	- 0.04	- 0.01
Sub-sample 1	2.75e-4	0.1207	- 2.9944	48.9912	- 0.33	- 0.04	- 0.01	- 0.01	- 0.02
Sub-sample 2	- 8.25e-5	0.1081	- 2.4515	30.2010	- 0.14	- 0.03	- 0.02	- 0.08	- 0.01
ARMS									
Full sample	0.9913	0.3524	2.0358	12.4300	0.06	0.06	0.01	0.04	0.05
Sub-sample 1	0.9741	0.3324	1.7813	10.5652	0.01	0.02	- 0.01	0.02	0.07
Sub-sample 2	1.0085	0.3706	2.1911	13.3334	0.10	0.09	0.02	0.05	0.02
ΔARMS									
Full sample	5.36e-4	0.4829	- 0.3993	8.7152	- 0.50	0.02	- 0.04	0.01	0.02
Sub-sample 1	1.66e-4	0.4682	- 0.2502	7.9815	- 0.50	0.02	- 0.02	- 0.02	0.07
Sub-sample 2	9.06e-4	0.4973	- 0.5235	9.2572	- 0.50	0.03	- 0.05	0.03	- 0.02

Table 1: Summary Statistics (continued)

The full weekly sample contains 613 observations from 6 April 1990 to 28 December 2001, sub-sample 1 contains 307 observations from 6 April 1990 to 16 February 1996, and sub-sample 2 contains 306 observations from 23 February 1996 to 28 December 2001.

Panel B: Weekly Data									
Variable	Mean	Standard Deviation	Skewness	Kurtosis	Autocorrelation Lag 1	2	3	4	5
Log V									
Full sample	- 1.7622	0.1878	0.4155	2.5781	0.82	0.75	0.72	0.70	0.69
Sub-sample 1	- 1.8859	0.1257	0.7877	3.6671	0.60	0.52	0.49	0.46	0.46
Sub-sample 2	- 1.6381	0.1550	0.2355	2.9476	0.73	0.60	0.53	0.48	0.45
Returns									
Full sample	2.12e-3	0.0218	- 0.5237	5.2161	- 0.08	0.04	0.02	0.02	- 0.05
Sub-sample 1	2.16e-3	0.0161	- 0.4397	5.8075	- 0.06	0.05	0.02	- 0.02	- 0.11
Sub-sample 2	2.09e-3	0.0263	- 0.4956	4.1387	- 0.08	0.03	0.02	0.02	- 0.03
AAII									
Full sample	0.8350	0.5754	2.8314	15.6196	0.68	0.63	0.55	0.56	0.55
Sub-sample 1	1.1012	0.7022	2.4285	11.2036	0.68	0.63	0.56	0.59	0.58
Sub-sample 2	0.6592	0.3283	1.0743	3.9604	0.50	0.39	0.25	0.16	0.13
ΔAAII									
Full sample	- 7.28e-4	0.4599	0.2061	10.0813	- 0.42	0.03	- 0.12	0.04	0.07
Sub-sample 1	- 1.46e-3	0.5624	0.2507	8.5421	- 0.43	0.03	- 0.15	0.08	0.10
Sub-sample 2	8.35e-6	0.3274	- 0.1114	4.7991	- 0.39	0.02	- 0.04	- 0.06	- 0.02
II									
Full sample	0.8262	0.3265	1.5098	5.3271	0.95	0.87	0.80	0.73	0.66
Sub-sample 1	0.9720	0.3732	0.9871	3.5065	0.94	0.86	0.79	0.72	0.65
Sub-sample 2	0.6799	0.1770	1.2026	4.5997	0.89	0.74	0.57	0.40	0.26
ΔII									
Full sample	- 7.76e-4	0.1065	- 0.2942	8.2748	0.20	0.02	- 0.04	- 0.08	- 0.15
Sub-sample 1	- 1.43e-3	0.1249	- 0.2848	7.5897	0.19	- 0.02	- 0.05	- 0.06	- 0.17
Sub-sample 2	- 1.15e-4	0.0842	- 0.2360	5.8695	0.16	0.09	0.00	- 0.13	- 0.12
PCV									
Full sample	1.1641	0.2052	0.8466	4.9898	0.45	0.40	0.34	0.28	0.27
Sub-sample 1	1.1238	0.1807	0.4290	2.8043	0.51	0.47	0.46	0.47	0.38
Sub-sample 2	1.2045	0.2200	0.9626	5.4540	0.38	0.32	0.20	0.21	0.17
ΔPCV									
Full sample	9.26e-4	0.2130	0.0298	4.2075	- 0.45	0.03	- 0.08	0.06	- 0.04
Sub-sample 1	1.65e-4	0.1796	0.0281	3.4004	- 0.46	- 0.03	- 0.01	0.10	- 0.06
Sub-sample 2	1.69e-3	0.2422	0.0250	4.0453	- 0.44	0.06	- 0.12	0.04	- 0.03
PCO									
Full sample	1.2363	0.2826	1.1681	5.1223	0.73	0.59	0.53	0.55	0.46
Sub-sample 1	1.2204	0.2692	1.5662	6.8146	0.71	0.57	0.52	0.57	0.50
Sub-sample 2	1.2522	0.2949	0.8455	4.0160	0.73	0.60	0.52	0.52	0.41
ΔPCO									
Full sample	4.89e-5	0.2090	- 0.9561	6.9610	- 0.26	- 0.13	- 0.16	0.21	- 0.01
Sub-sample 1	4.33e-3	0.1902	- 0.9695	6.3796	- 0.28	- 0.12	- 0.19	0.30	- 0.06
Sub-sample 2	- 4.25e-3	0.2266	- 0.9188	6.9846	- 0.23	- 0.11	- 0.10	0.17	- 0.04

Table 2: Correlation Coefficients

The correlations are between sentiment variables and either returns or the logarithm of realized volatility. Sentiment is measured as either the put-call volume ratio (PCV) the put-call open interest ratio (PCO) the ARMS ratio or the survey based measures provided by the (AAII) American Association for Individual Investors or (II) Investor Intelligence.

Panel A: Daily Level Data				
Variable	PCV	PCO	ARMS	
Log V				
Full sample	0.0634	- 0.1750	0.1704	
Sub-sample 1	- 0.0104	- 0.0771	0.1740	
Sub-sample 2	- 0.0549	- 0.4432	0.1752	
Returns				
Full sample	- 0.1721	- 0.0077	- 0.7260	
Sub-sample 1	- 0.3102	- 0.0349	- 0.7724	
Sub-sample 2	- 0.1166	- 0.0046	- 0.7225	
Panel B: Daily Change Data				
Variable	△PCV	△PCO	△ARMS	
Log V				
Full sample	- 0.0091	- 0.0350	- 0.0226	
Sub-sample 1	0.0301	- 0.0253	0.0012	
Sub-sample 2	- 0.0336	- 0.0576	- 0.1107	
Returns				
Full sample	- 0.1915	0.0001	- 0.5447	
Sub-sample 1	- 0.3642	0.0028	- 0.5539	
Sub-sample 2	- 0.1234	- 0.0017	- 0.5590	
Panel C: Weekly Level Data				
Variable	AAII	II	PCV	PCO
Log V				
Full sample	0.0078	- 0.2700	0.0759	- 0.2476
Sub-sample 1	0.3698	0.0468	- 0.0487	- 0.1210
Sub-sample 2	0.2489	0.0349	- 0.0900	- 0.5734
Returns				
Full sample	- 0.0479	0.0408	- 0.0241	0.2850
Sub-sample 1	- 0.0067	0.0212	- 0.3154	0.2606
Sub-sample 2	- 0.1322	0.0953	0.1229	0.3078
Panel D: Weekly Change Data				
Variable	△AAII	△II	△PCV	△PCO
Log V				
Full sample	0.0583	0.0909	- 0.0314	- 0.0974
Sub-sample 1	0.0836	0.0716	0.0040	- 0.0968
Sub-sample 2	0.0786	0.1803	- 0.0745	- 0.1219
Returns				
Full sample	- 0.0436	- 0.1192	- 0.0324	0.3885
Sub-sample 1	0.0642	- 0.0984	- 0.3773	0.4191
Sub-sample 2	- 0.1698	- 0.1606	0.1252	0.3791

Table 3: Granger Causality Tests between Returns and Sentiment

Results of Granger causality tests between returns and sentiment indicators. The tabulated statistics are the p-values of the test statistics, that are twice the likelihood ratio and have an asymptotic chi-squared distribution when the null hypothesis holds. The numbers of lagged terms in the VAR models are decided by the minimum Akaike Information Criterion.

Test 1: H_0: Granger-noncausality from sentiment to returns, i.e. sentiment does not cause returns.
Test 2: H_0: Granger-noncausality from returns to sentiment.
Test 3: H_0: Granger-noncausality from sentiment change to returns.
Test 4: H_0: Granger-noncausality from returns to sentiment change.

Panel A: Daily Data				
Sentiment	Test 1	Test 2	Test 3	Test 4
PCV				
Full Sample	0.0821	<0.0001	0.1359	<0.0001
Sub-sample 1	0.1550	<0.0001	0.4049	<0.0001
Sub-sample 2	0.1406	<0.0001	0.3817	<0.0001
PCO				
Full Sample	0.7999	<0.0001	0.6784	<0.0001
Sub-sample 1	0.0643	<0.0001	0.0115	<0.0001
Sub-sample 2	0.8991	<0.0001	0.9020	<0.0001
ARMS				
Full Sample	0.0314	<0.0001	0.0212	<0.0001
Sub-sample 1	0.1219	0.0695	0.3038	<0.0001
Sub-sample 2	0.2011	0.0932	0.1947	<0.0001
Panel B: Weekly Data				
Sentiment	Test 1	Test 2	Test 3	Test 4
AAII				
Full Sample	0.1866	<0.0001	0.2106	<0.0001
Sub-sample 1	0.4179	<0.0001	0.3537	<0.0001
Sub-sample 2	0.0349	<0.0001	0.3187	0.0007
II				
Full Sample	0.1174	<0.0001	0.3394	<0.0001
Sub-sample 1	0.5877	<0.0001	0.9187	<0.0001
Sub-sample 2	0.0004	<0.0001	0.0203	<0.0001
PCV				
Full Sample	0.9223	0.0002	0.7095	0.0005
Sub-sample 1	0.7526	0.0037	0.1998	0.0010
Sub-sample 2	0.7064	0.0163	0.2846	0.0199
PCO				
Full Sample	0.7317	<0.0001	0.5848	<0.0001
Sub-sample 1	0.1346	<0.0001	0.0645	<0.0001
Sub-sample 2	0.7333	<0.0001	0.6289	0.0044

Table 4: Granger Causality Tests between Volatility and Sentiment

Results of Granger causality tests between log realized volatility and sentiment indicators. The tabulated statistics are the p-values of the test statistics, that are twice the likelihood ratio and have an asymptotic chi-squared distribution when the null hypothesis holds. The numbers of lagged terms in the VAR models are decided by the minimum Akaike Information Criterion.

Test 1: H_0: Granger-noncausality from sentiment to realized volatility, i.e. sentiment does not cause realized volatility.
Test 2: H_0: Granger-noncausality from realized volatility to sentiment.
Test 3: H_0: Granger-noncausality from sentiment change to realized volatility.
Test 4: H_0: Granger-noncausality from realized volatility to sentiment change.

Panel A: Daily Data				
Sentiment	Test 1	Test 2	Test 3	Test 4
PCV				
Full Sample	0.2894	0.0025	0.4144	0.0045
Sub-sample 1	0.3631	0.0940	0.4285	0.3072
Sub-sample 2	0.1909	<0.0001	0.2616	<0.0001
PCO				
Full Sample	0.4232	<0.0001	0.7265	<0.0001
Sub-sample 1	0.6845	0.2522	0.2624	0.1870
Sub-sample 2	0.0207	<0.0001	0.0249	<0.0001
ARMS				
Full Sample	<0.0001	0.0030	<0.0001	<0.0001
Sub-sample 1	<0.0001	0.0715	<0.0001	0.0102
Sub-sample 2	<0.0001	0.1265	<0.0001	0.0026

Panel B: Weekly Data				
Sentiment	Test 1	Test 2	Test 3	Test 4
AAII				
Full Sample	0.1450	<0.0001	0.6078	<0.0001
Sub-sample 1	0.8221	0.0019	0.8059	0.0038
Sub-sample 2	0.3680	0.0080	0.5152	0.0052
II				
Full Sample	0.0052	<0.0001	0.1572	<0.0001
Sub-sample 1	0.5613	<0.0001	0.5985	<0.0001
Sub-sample 2	0.0118	0.0003	0.1945	<0.0001
PCV				
Full Sample	0.4682	0.4999	0.4716	0.5789
Sub-sample 1	0.0030	0.0155	0.0048	0.0490
Sub-sample 2	0.3832	0.3157	0.3664	0.3944
PCO				
Full Sample	0.1711	0.0108	0.1440	0.0040
Sub-sample 1	0.6303	0.0799	0.3263	0.0479
Sub-sample 2	0.5820	0.0330	0.5842	0.0370

Table 5: Granger Causality Tests between Returns and Volatility

Results of Granger causality tests between returns and log realized volatility. The tabulated statistics are the p-values of the test statistics, that are twice the likelihood ratio and have an asymptotic chi-squared distribution when the null hypothesis holds. The numbers of lagged terms in the VAR models are decided by the minimum Akaike Information Criterion.

Test 1: H_0: Granger-noncausality from returns to volatility, i.e. returns do not cause volatility.
Test 2: H_0: Granger-noncausality from volatility to returns.

	Panel A: Daily Data	
	Test 1	Test 2
Full Sample	<0.0001	0.9495
Sub-sample 1	<0.0001	0.6634
Sub-sample 2	<0.0001	0.9596
	Panel B: Weekly Data	
	Test 1	Test 2
Full Sample	0.0001	0.7148
Sub-sample 1	0.0120	0.4231
Sub-sample 2	0.0120	0.4213

Table 6: Incremental Predictive Power of ARMS for Realized Volatility (Without returns)

This table shows the incremental contribution of ARMS for realized volatility forecasting. The current log realized volatility is regressed on either lagged log realized volatility (Panel A) or the VIX index (Panel B) and either the level or first difference of ARMS. IR is the incremental adjusted R-square relative to the benchmark model.

Panel A: Based on Lagged Realized Volatility

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \gamma ARMS_{t-1} + \varepsilon_t,$$

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t,$$

where $D_{1,t-1}$: 1 if $\Delta S_{t-1}>0$, otherwise 0. $D_{2,t-1}$: 1 if $\Delta S_{t-1}<0$, otherwise 0.
F test: H_0: The incremental explanatory power of ARMS or ΔARMS is zero.

	K	$\Sigma\ \beta_i$	γ	λ	IR	F Test
Benchmark					(Adj. R^2)	
Full Sample	- 0.2293***	0.8927***			61.54%	
Sub-sample 1	- 0.5812***	0.7425***			33.63%	
Sub-sample 2	- 0.3136***	0.8445***			49.88%	
ARMS						
Full Sample	- 0.3234***	0.8800***	0.0675***		+1.24%	100.56***
Sub-sample 1	- 0.6584***	0.7315***	0.0538***		+1.29%	30.69***
Sub-sample 2	- 0.4344***	0.8251***	0.0810***		+2.45%	77.77***
ΔARMS						
Full Sample	- 0.2664***	0.8809***	0.0543***	- 0.0140	+0.43%	17.95***
Sub-sample 1	- 0.6839***	0.7044***	0.0595***	- 0.0396***	+0.97%	12.14***
Sub-sample 2	- 0.3668***	0.8254***	0.0699***	- 0.0128	+0.93%	27.07***

Panel B: Based on Lagged VIX

$$LogV_t = K + \beta LogVIX_{t-1} + \gamma ARMS_{t-1} + \varepsilon_t,$$

$$LogV_t = K + \beta LogVIX_{t-1} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t,$$

where $D_{1,t-1}$: 1 if $\Delta S_{t-1}>0$, otherwise 0. $D_{2,t-1}$: 1 if $\Delta S_{t-1}<0$, otherwise 0.
F test: H_0: The incremental explanatory power of ARMS or ΔARMS is zero.

	K	β	γ	λ	IR	F Test
Benchmark					(Adj. R^2)	
Full Sample	0.0651**	1.1472***			61.00%	
Sub-sample 1	- 0.5725***	0.8353***			37.09%	
Sub-sample 2	0.4952***	1.3784***			49.64%	
ARMS						
Full Sample	- 0.0060	1.1301***	0.0388***		+0.41%	32.46***
Sub-sample 1	- 0.6415***	0.8184***	0.0359***		+0.56%	13.88***
Sub-sample 2	0.3899***	1.3420***	0.0386***		+0.51%	16.49***
ΔARMS						
Full Sample	- 0.0183	1.1286***	0.0378***	- 0.0274***	+0.24%	10.22***
Sub-sample 1	- 0.6698***	0.7943***	0.0441***	- 0.0434***	+0.70%	9.52***
Sub-sample 2	0.4103***	1.3390***	0.0474***	- 0.0265***	+0.40%	6.99***

* : Significance at 10% level.
** : Significance at 5% level.
***: Significance at 1% level.

Table 7: Incremental Predictive Power of ARMS for Realized Volatility (With returns)

This table shows the incremental contribution of ARMS for realized volatility forecasting, when returns are included in the benchmark models. The current log realized volatility is regressed on either lagged log realized volatility (Panel A) or the VIX index (Panel B) and either the level or first difference of ARMS. IR is the incremental adjusted R-square relative to the benchmark model.

Panel A: Based on Lagged Realized Volatility

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma ARMS_{t-1} + \varepsilon_t,$$

$$LogV_t = K + \sum_{i=1}^{5} \beta_i LogV_{t-i} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t,$$

where $D_{1,t-1}$: 1 if $\Delta S_{t-1}>0$, otherwise 0. $D_{2,t-1}$: 1 if $\Delta S_{t-1}<0$, otherwise 0.
$D_{3,t-1}$: 1 if $r_{t-1}<0$, otherwise 0.
F test: H_0: The incremental explanatory power of ARMS or ΔARMS=0.

	K	$\Sigma \beta_i$	α	γ	λ	IR	F Test
Benchmark						(Adj. R^2)	
Full Sample	- 0.2569***	0.8888***	- 0.0494***			63.49%	
Sub-sample 1	- 0.6278***	0.7287***	- 0.0400***			35.96%	
Sub-sample 2	- 0.3383***	0.8440***	- 0.0609***			53.62%	
ARMS							
Full Sample	- 0.2796***	0.8858***	- 0.0417***	0.0194**		+0.04%	4.63**
Sub-sample 1	- 0.6330***	0.7284***	- 0.0378***	0.0054		- 0.04%	0.15
Sub-sample 2	- 0.3766***	0.8373***	- 0.0497***	0.0288**		+0.16%	5.95**
ΔARMS							
Full Sample	- 0.2703***	0.8847***	- 0.0551***	- 0.0093	- 0.0227***	+0.10%	5.10***
Sub-sample 1	- 0.6704***	0.7050***	- 0.0454***	0.0029	- 0.0451***	+0.54%	7.41***
Sub-sample 2	- 0.3614***	0.8356***	- 0.0644***	0.0012	- 0.0248**	+0.09%	2.40*

Panel B: Based on Lagged VIX

$$LogV_t = K + \beta LogVIX_{t-1} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma ARMS_{t-1} + \varepsilon_t,$$

$$LogV_t = K + \beta LogVIX_{t-1} + \alpha D_{3,t-1} \frac{r_{t-1}}{\sqrt{V_{t-1}}} + \gamma D_{1,t-1} \Delta ARMS_{t-1} + \lambda D_{2,t-1} \Delta ARMS_{t-1} + \varepsilon_t,$$

where $D_{1,t-1}$: 1 if $\Delta S_{t-1}>0$, otherwise 0. $D_{2,t-1}$: 1 if $\Delta S_{t-1}<0$, otherwise 0.
$D_{3,t-1}$: 1 if $r_{t-1}<0$, otherwise 0.
F test: H_0: The incremental explanatory power of ARMS or ΔARMS=0.

	K	β	α	γ	λ	IR	F Test
Benchmark						(Adj. R^2)	
Full Sample	0.0284	1.1334***	- 0.0264***			61.55%	
Sub-sample 1	- 0.6333***	0.8101***	- 0.0258***			38.05%	
Sub-sample 2	0.4262***	1.3475***	- 0.0323***			50.65%	
ARMS							
Full Sample	0.0087	1.1298***	- 0.0203***	0.0154*		+0.03%	2.78*
Sub-sample 1	- 0.6387***	0.8096***	- 0.0236***	0.0056		- 0.04%	0.17
Sub-sample 2	0.4107***	1.3428***	- 0.0292***	0.0081		- 0.02%	0.44
ΔARMS							
Full Sample	0.0006	1.1222***	- 0.0302***	0.0029	- 0.0320***	+0.17%	7.56***
Sub-sample 1	- 0.7021***	0.7814***	- 0.0313***	0.0052	- 0.0476***	+0.64%	8.81***
Sub-sample 2	0.3729***	1.3231***	- 0.0346***	0.0116	- 0.0335***	+0.20%	4.07**

*   : Significance at 10% level.
** : Significance at 5% level.
***: Significance at 1% level.