Journal of Monetary Economics 108 (2019) 156–161

Contents lists available at ScienceDirect

Journal of Monetary Economics

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

Comment on "Monetary Policy Communication, Policy Slope, and the Stock Market" by Andreas Neuhierl and Michael Weber☆

Emanuel Moench*, Tobias Stein

Deutsche Bundesbank, Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 27 August 2019
Accepted 29 August 2019
Available online 2 September 2019

Keywords:
Return predictability
Central bank communication
Content analysis

ABSTRACT

This note discusses the article "Monetary Policy Communication, Policy Slope, and the Stock Market" by Andreas Neuhierl and Michael Weber. The authors document that the slope of the fed funds futures curve predicts stock returns one week ahead and is correlated with the tone of Fed Board member speeches over the period 1994–2007. We show that this return predictability is restricted to the subsample from 1999 to 2001, with no evidence of predictability outside of this period. We further point out some issues with the proposed measure of monetary policy tone and show that the positive correlation between tone and the fed funds futures slope is driven by two speeches that, when analyzed in detail, do not appear particularly hawkish.

1. Introduction

A growing recent literature has shown that Federal Open Market Committee (FOMC) announcements, and the anticipation thereof, are a key driver of equity returns (see, e.g. Bernanke and Kuttner (2005); Lucca and Moench (2015)). This is consistent with the view that market expectations of Federal Reserve policy are an important component of the risk-return tradeoff that stock market investors need to consider. But expectations about future monetary policy actions are also shaped through central bankers' communication with the public in the period between policy meetings. This raises the question whether such intermeeting communication also has an impact on stock returns. The paper by Neuhierl and Weber (2019) argues that it does. The authors provide two pieces of evidence. First, they show that the slope of the term structure of fed funds futures rates between one and three months out predicts the excess return on the stock market one week ahead. Second, they document that the tone of speeches by Federal Reserve officials is positively correlated with the slope of the fed funds futures curve.

In this comment, we focus on three issues in the empirical analysis provided by Neuhierl and Weber (2019). We first show that the predictability of stock returns via the fed funds futures slope is limited to the sample from 1999 to 2001. We document that the coefficient for slope is insignificant for the periods from 1994 to 1998 and 2002 to 2007. Second, we highlight some issues related to the measurement of the tone of speeches by Fed officials and propose an alternative index

☆ We thank Michael Weber for kindly sharing data with us. The views expressed here are those of the authors and do not necessarily reflect those of Deutsche Bundesbank or the Eurosystem.

* Corresponding author.
E-mail address: emanuel.moench@bundesbank.de (E. Moench).

https://doi.org/10.1016/j.jmoneco.2019.08.020
0304-3932/

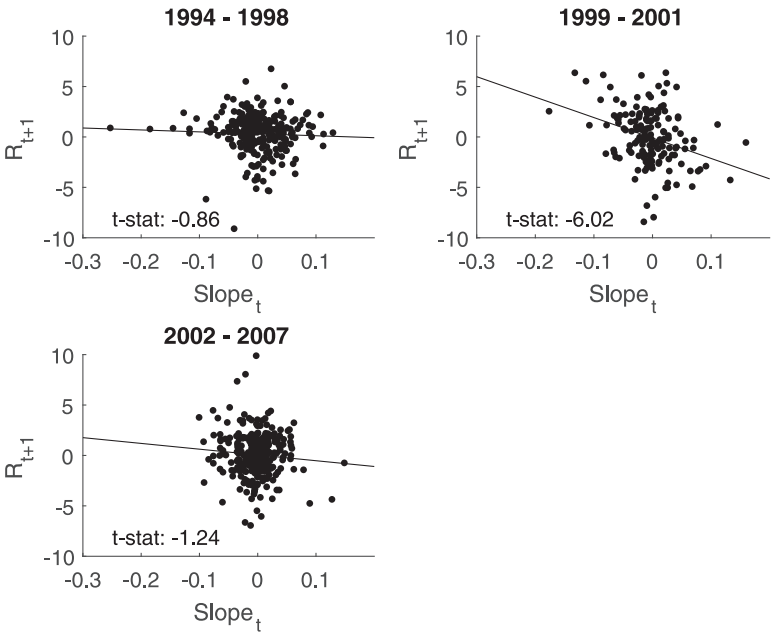

Fig. 1. Plot of Weekly Excess Return on Lagged Slope Factor. This figure plots the weekly excess return of the CRSP value-weighted index against the lagged slope factor (regression residual) for three subsamples. The first sample period (top left) is from 1994 to 1998, the second sample (top right) from 1999 to 2001, and the last sample (bottom left) from 2002 to 2007. The displayed t-statistic corresponds to the coefficient of the slope factor when regressing excess returns on a constant and the lagged slope factor (Newey-West standard errors with four lags).

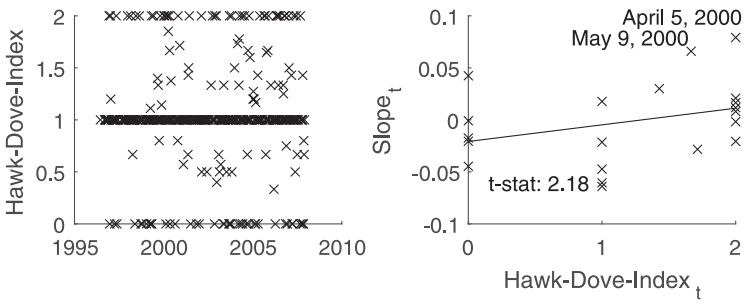

Fig. 2. Plot of Hawk-Dove-Index over Time and against Slope. This figure plots the Hawk-Dove-Index over time for the sample from June 1996 to the last week of 2007 (left), and the slope factor against the Hawk-Dove-Index for the subsample from 1999 to 2001 (right), including only speeches by the chair or vice chair with at least one classification of a hawkish or dovish word. The displayed t-statistic corresponds to the coefficient of the Hawk-Dove-Index when regressing slope on a constant and the Hawk-Dove-Index (Newey-West standard errors with four lags).

which partly addresses them. Third, we show that the correlation between the fed funds futures slope and the tone index in the crucial period from 1999 to 2001 is driven by two speeches which, when analyzed carefully, do not appear particularly hawkish.

2. Local return predictability

Neuhierl and Weber (2019) define the fed funds futures slope factor as the residual of a regression of changes in the three-month federal funds futures-implied rate on a constant and changes in the one-month federal funds futures-implied rate. They show that the slope factor predicts one-week ahead excess returns of the CRSP value-weighted index over the period 1994–2007 both in-sample and out-of-sample. Fig. 2 in their article plots the out-of-sample performance of slope relative to the performance of the historical mean. It shows that the superior predictability of the slope factor is limited to the period from 1999 to 2001. For the periods from 1994 to 1998 and from 2002 to 2007 the cumulative sum of squared forecast errors differential is about flat, indicating that during these periods the slope factor does not have predictive power for the stock market over and above the historical mean return of the market portfolio. To further illustrate that the years from 1999 to 2001 are influential in this relationship, Fig. 1 shows scatter plots of the excess market return and the lagged slope factor for the different subsamples. In a linear regression of returns on a constant and lagged slope only the coefficient

Table 1
Predictive Regressions. This table shows results for predictive regressions of weekly excess returns of the CRSP value-weighted index on a constant, the lagged CRSP value-weighted return (R_t), a dummy variable that is set to one for the period from 1999 to 2001 ($\mathbf{1}_{1999-2001}$), the slope factor (Slope_t), the dividend-price ratio (DP_t), the VIX (VIX_t), the term spread (TERM_t), and interaction terms of the dummy variable with slope, dividend-price ratio, VIX, and term-spread. Column (1) repeats the baseline regression of Neuhierl and Weber (2019), in column (2) we include the dummy variable and an interaction term of the dummy variables with slope. In columns (3) to (5) we replace slope by dividend-price ratio, VIX, and term spread, respectively. In column (6) we include all regressors and interaction terms of dummy variable with regressors simultaneously. We report t-statistics based on Newey-West standard errors with four lags in parentheses. The sample period is from the first week of 1994 to the last week of 2007. The R^2 is reported in % and the number of observations is reported in the last row (Num. Obs).

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.13	0.18**	−0.17	0.02	0.38***	−0.73
	[1.51]	[2.00]	[−0.39]	[0.08]	[2.87]	[−1.14]
R_t	−0.09**	−0.07*	−0.10**	−0.08***	−0.11***	−0.05
	−[2.00]	[−1.67]	[−2.40]	[−1.95]	[−2.71]	[−1.13]
$\mathbf{1}_{1999-2001}$		−0.28	−1.94*	−3.70***	−0.76**	−0.22
		[−1.12]	[−1.71]	[−2.79]	[−2.50]	[−0.13]
Slope_t	−7.04***	−2.92				−2.71
	[−3.39]	[−1.31]				[−1.26]
$\text{Slope}_t \times \mathbf{1}_{1999-2001}$		−16.08***				−20.45***
		[−3.81]				[−4.23]
DP_t			20.09			43.93*
			[0.85]			[1.84]
$\text{DP}_t \times \mathbf{1}_{1999-2001}$			147.83*			−481.83**
			[1.69]			[−2.49]
VIX_t				0.01		0.02
				[0.43]		[1.00]
$\text{VIX}_t \times \mathbf{1}_{1999-2001}$				0.14**		0.20***
				[2.53]		[2.72]
TERM_t					−0.13*	−0.17**
					[−1.74]	[−2.21]
$\text{TERM}_t \times \mathbf{1}_{1999-2001}$					0.54***	1.23***
					[2.86]	[2.94]
R^2 (in %)	2.63	4.47	1.93	2.61	2.27	7.92
Num. Obs.	724	724	724	723	724	723

Table 2
Regressions of Slope on Hawk-Dove Index and Tone-Index. This table shows results for regressions of the slope factor on a constant, the Hawk-Dove Index, and the Tone-Index. In columns (1) and (2) we include all observations from the first week of June 1996 to the last week of 2007. In columns (3) and (4) we only include speeches by the chair or vice chair. In column (5) we only include observations with at least one hawkish or dovish classification. Column (6) shows results for at least one classification and speeches by the chair of vice chair. In column (7) we further exclude speeches from April 5, 2000 and May 9, 2000. We report t-statistics based on Newey-West standard errors with four lags in parentheses. The R^2 is reported in %.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	−0.32	−0.16	−1.56***	−0.81***	−0.02	−1.09	−1.04
	[−0.79]	[−0.70]	[−2.58]	[−2.86]	[−0.04]	[−1.24]	[−1.18]
Hawk-Dove Index	0.33		1.06**		0.30	0.98*	0.75
	[1.00]		[2.27]		[0.85]	[1.95]	[1.48]
Tone-Index		0.39*		0.81***			
		[1.87]		[3.29]			
R^2 (in %)	0.21	0.63	2.98	3.84	0.42	6.29	4.33
Num. Obs.	380	380	173	173	157	68	66
Only Speeches by Chair			x	x		x	x
Without influential obs.							x
At least one classification					x	x	x

for the period from 1999 to 2001 is highly significant with a t-statistic of −6.02. This is in sharp contrast to t-statistics of −0.86 and −1.24 in the other two periods.

To document this subsample instability more formally, we reestimate Table 2 from Neuhierl and Weber (2019) adding an interaction term between the slope factor and a dummy variable that equals one for the period 1999–2001. Results are shown in Table 1. Column (1) repeats the baseline regression with a constant, lagged excess returns, and the slope factor. The coefficient for the slope factor is highly significant and the R^2 is 2.6%. When we add the subsample dummy and interaction term in column (2), only the coefficient for the interaction term is highly statistically significant. Moreover, the coefficient

more than doubles from the baseline regression to the subsample interaction while it drops sharply without interaction. At the same time, the R^2 increases to 4.5%. This is strong evidence for the documented return predictability being restricted to the short subsample from 1999 to 2001. In columns (3) to (5) we run the same regression with DP, VIX, and TERM instead of the slope factor and find that – interestingly – all of the predictors are significant for the subsample from 1999 to 2001. In column (6) we include all regressors and interaction terms jointly. The R^2 jumps to 7.9% and slope, VIX, and TERM have significant coefficients for the interaction terms. The coefficient for dividend-price ratio is significant but has the wrong sign, which we attribute to the high correlation with VIX (53%) and TERM (69%) between 1999 and 2001.[1] In light of these results, it thus appears premature to conclude that the slope of the fed funds futures curve systematically predicts equity returns.

3. Policy speeches: Hawk-Dove-Index and slope factor

In the second part of their analysis, Neuhierl and Weber (2019) show that the tone of speeches by Fed board members correlates with the slope factor.[2] To this end, they construct a Hawk-Dove-Index following Apel and Grimaldi (2012) as

$$\text{Hawk}-\text{Dove}-\text{Index}_t = \begin{cases} \left[\frac{\#\text{hawk}_t - \#\text{dove}_t}{\#\text{hawk}_t + \#\text{dove}_t}\right] + 1 = \frac{2\times\#\text{hawk}_t}{\#\text{hawk}_t + \#\text{dove}_t} & \text{if } \#\text{hawk}_t + \#\text{dove}_t > 0 \\ 1 & \text{otherwise} \end{cases}$$

Here, #hawk and #dove are the number of words identified as hawkish and dovish according to a list of bigrams (combination of two words). For example, "employment declined" is considered a dovish bigram. The index is designed to measure the "net hawkishness" of each of the analyzed texts. For a value above one the speech consists of more hawkish than dovish words, and the authors expect that this signals faster future monetary policy tightening. Fig. 2 plots the Hawk-Dove-Index over time for the sample from June 1996 through 2007. The chart gives rise to several observations.

First, the index is quite volatile, with tone changing substantially from week to week. This is in contrast to Fig. 1 from Apel and Grimaldi (2012) which shows a rather smooth measure of tone with lower frequency swings. Hansen and McMahon (2016) analyze FOMC statements with computational linguistics. Their FOMC communication index also fluctuates rather slowly, in line with the notion that the tone of monetary policy communication does not move strongly from week to week.[3]

One potential reason for the substantial variation is the small number of observed hawkish or dovish classifications. The bigrams in the word list always combine a noun with an adjective or a verb, this is the key input for the search-and-count approach. Hansen and McMahon (2016) show that frequently used stemmed words in FOMC statements are "improv" and "foster" (for expansion words) or "slow" and "weak" (for contraction words). Yet, none of these words are included in the word list in Neuhierl and Weber (2019). Even further, while according to Hansen and McMahon (2016) "moder" is the most frequently used stemmed adjective in FOMC statements, the word "moderate" only appears in the word list of Neuhierl and Weber (2019) in one combination. An alternative approach would be to use dictionaries and search for individual words rather than bigrams. A commonly used dictionary is the one of Loughran and McDonald (2011), see Tietz (2019) for an adjusted version of the dictionary for the context of speeches by Fed officials. How a more comprehensive word list or the dictionary approach affect correlation between tone and slope remain questions for future research.

Second, the chart shows a clustering of the tone measure at values of 0, 1, and 2. Many of the observations where tone equals 1 refer to weeks where there was neither a hawkish nor a dovish term in Fed policy makers' speeches. As the index is not defined in such cases, Neuhierl and Weber (2019) set it equal to one. This is the case for 223 of the 380 weekly observations that they use in their baseline sample.[4] Moreover, the authors implicitly assume that a period without any hawkish or dovish word has a positive "net hawkishness". This is in contrast to other tone measures that set speeches with neutral tone to zero (see e.g. Solomon (2012)).

Other commonly used tone measures share the characteristic of being zero for texts without any classification (Ehrmann and Talmi, 2019; Hansen and McMahon, 2016). In line with these approaches, we propose an alternative index that simply sets all periods without a classification to zero:

$$\text{Tone}-\text{Index}_t = \begin{cases} \frac{2\times\#\text{hawk}_t}{\#\text{hawk}_t + \#\text{dove}_t} & \text{if } \#\text{hawk}_t + \#\text{dove}_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

The correlation between the Hawk-Dove-Index and the alternative Tone-Index is 72%. Table 2 shows results for the regression of slope on both indices separately. Column (1) replicates the baseline regression of Neuhierl and Weber (2019) and shows an insignificant coefficient for the Hawk-Dove index. Replacing the Hawk-Dove index with the Tone-Index, the coefficient is significant at the 10% level (see column 2). Columns (3) and (4) repeat the regression including only speeches

[1] Note that here we simply capture time-variation in the coefficient by including a dummy variable. A more sophisticated approach such as nonparametric regression estimation is likely to uncover "pockets of predictability" in a more systematic fashion, see Farmer et al. (2019).

[2] Even though they write that they "collect all speeches for members of the FOMC", in practice they only consider the governors of the Federal Reserve Board. In addition to these, the FOMC comprises the presidents of the twelve regional Federal Reserve Banks as members.

[3] A novelty of their approach is the combination of topic and tone, they measure sentiment separately for the state of the economy and for monetary policy decisions.

[4] Apel and Grimaldi (2012) construct the index for 82 periods with a total of 819 hawkish words and 609 dovish words and do not have zero classifications.

Table 3
Linguistic Analysis for the Subsample from 1999 to 2001. This table shows results for regressions of the slope factor on a constant, the Hawk-Dove-Index, and the Tone-Index for the sample period from the first week of 1999 to the last week of 2001. We only include speeches by the chair or vice chair. In columns (1) and (2) we separately regress slope on the Hawk-Dove Index and Tone-Index. In columns (3) and (4) we repeat the regressions but exclude speeches from April 5, 2000 and May 9, 2000. In column (5) we drop periods without any hawkish or dovish classification and in column (6) we further exclude the two outliers. We report t-statistics based on Newey-West standard errors with four lags in parentheses. The R^2 is reported in %, the number of observations is given by Num. Obs., and the last three rows indicate when we only include speeches by the chair or vice chair, drop two influential observations, or only include periods with at least one hawkish or dovish classification.

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−2.67***	−1.40**	−1.99*	−1.36**	−2.08	−1.88
	[−2.66]	[−2.27]	[−1.86]	[−2.20]	[−1.24]	[−1.20]
Hawk-Dove Index	1.73***		0.80		1.60**	0.79
	[2.62]		[1.25]		[2.18]	[1.09]
Tone-Index		1.20**		0.48*		
		[2.47]		[1.69]		
R^2 (in %)	6.25	6.79	1.57	1.21	10.97	4.28
Num. Obs.	48	48	46	46	20	18
Only Speeches by Chair	x	x	x	x	x	x
Without influential obs.			x	x		x
At least one classification					x	x

by the chair or vice chair of the Federal Reserve Board which are arguably more influential in shaping market expectations about monetary policy. Both indices are positive and statistically significant at the 5% and 1% level, respectively. Repeating the baseline regression for periods with at least one classification in column (5), however, we see that the coefficient becomes insignificant.[5] This indicates that the large number of speeches without any hawkish or dovish classification affects the results.[6] The R^2 is only 0.4% in this case. In column (6) we only include speeches by the chair or vice chair with at least one classification. The coefficient is weakly significant with an R^2 of 6.3%. When we drop two influential observations in column (7), the coefficient becomes insignificant and the R^2 drops from 6.3% to 4.3%. We analyze these observations in more detail in the next section. In sum, the provided evidence shows at best a statistically weak correlation between the hawkishness of Fed governors' speeches and the slope of the fed funds futures curve.

Third, common measures of tone are defined as $\frac{\#\text{Negative Words}}{\#\text{Total Words}}$ or $\frac{\#\text{Positive Words}-\#\text{Negative Words}}{\#\text{Total Words}}$, thus dividing the net number of hawkish words by the total number of words in a speech rather than the sum of hawkish and dovish words. The normalisation by the total word count accommodates texts of different lengths, see Ehrmann and Talmi (2019). Acosta and Meade (2015) show that both the structure and length of FOMC statements has changed over time, from an average word count between 100 and 200 until 2008 to more than 500 words between 2013 and 2015.[7] To what extent this may be an issue for the analysis of Fed Governor speeches remains to be explored.

4. A closer look at fed governor speeches between 1999 and 2001

We have shown in the previous section that the predictability of slope for excess returns is mainly driven by the period 1999 to 2001. A natural follow-up question is thus whether the correlation between the measured tone of speeches by Fed Board governors and the slope of the fed funds futures curve is also more pronounced during this period. In the first two columns of Table 3 we show regressions of the slope factor on the Hawk-Dove-Index and the Tone-Index for the sample 1999–2001, including only speeches by the chair and vice chair. The coefficient is positive and significant in both cases with an R^2 of 6.3% for the Hawk-Dove Index and 6.8% for the Tone-Index. In Fig. 2 we plot the slope factor against the Hawk-Dove-Index, dropping weeks without classification. While there is a positive relation, this is primarily driven by two observations which represent a speech by Alan Greenspan (then chairman of the Federal Reserve Board of Governors) on April 5, 2000 and a speech by Roger W. Ferguson, Jr. (then vice chairman) on May 9, 2000. In columns (3) and (4) we repeat the regression for speeches by the chair or vice chair that drop these two observations. The coefficient for the Hawk-Dove Index becomes insignificant and the R^2 drops from 6.3% to 1.6%. While the Tone-Index remains weakly significant, the R^2 drops sharply to 1.2%. In column (5) we only include observations from speeches of the chair or vice chair with at least one classification. The picture is the same as before, the coefficient is highly significant but when dropping the two observations in column (6) the significance vanishes and the R^2 drops from 11% to 4.3%.

[5] The definition of both indices is identical for $\#\text{hawk}_t + \#\text{dove}_t > 0$ and therefore the coefficient estimates are identical.

[6] Neuhierl and Weber (2019) identify 43 weeks with speeches by chair or vice share and at least one classification. With the data at hand we get 68 observations, see column (6).

[7] See FEDS Notes from September 30, 2015: https://www.federalreserve.gov/econresdata/notes/feds-notes/2015/semantic-analysis-of-the-FOMCs-postmeeting-statement-20150930.html.

Interestingly, U.S. stock investors did not react strongly to these two speeches. The S&P500 index saw moderate intraday losses of 0.49 and 0.85 percent, respectively. Given the importance of these two speeches for the documented empirical relation and the muted immediate stock market response, it is instructive to study them in somewhat greater detail. We focus on the May 9, 2000 speech by former vice chairman Roger W. Ferguson, Jr. which was given at the New Economy Forum at Haas School of Business, University of California, Berkeley. With the Hawk-Dove Classification of Neuhierl and Weber (2019, Table A.13) at hand, the identified one dovish and five hawkish sentences in this speech are the following:

> "This is significantly above the pace in the previous five years, and you have to go back to the decade of the 1960s to find even closely comparable periods of consistently robust economic expansion. In this environment, the **unemployment rate has fallen** to 4 percent, and the underlying rate of price inflation has slowed, on net, despite very high rates of resource utilization. Even the most optimistic of forecasters could not have anticipated such a favorable confluence of economic events." (hawkish)

> "In many industries, investments in information technologies have helped firms to cut back on the volume of inventories that they hold as a precaution against glitches in their supply chain or as a hedge against unexpected **increases in aggregate demand**." (hawkish)

> "About one-half of the 1 percentage point **increase in productivity growth** over the 1995–1999 period can be attributed to so-called 'capital deepening'." (hawkish)

> "But technological waves ebb and flow, and it is natural to ask whether we can count on such **rapid productivity growth** in the future." (hawkish)

> "Are new technologies emerging from R&D that have the realistic potential to **increase productivity growth** in the economy even further?" (hawkish)

> "This is the process through which an innovation on the supply side of the economy generates a comparable **increase in aggregate demand**." (dovish)

All of these sentences either refer to past experiences or reflect general statements or questions about the importance of technology for macroeconomic fluctuations. None appears in a paragraph that directly addresses monetary policy. While we think that the linguistic analysis of speeches is a useful tool to establish connections between policy communication and market expectations, these examples show that a simple count of (a short list of) bigrams may possibly lead to premature conclusions.

5. Conclusion

We have shown that the slope of the fed funds futures curve has significant predictive power for stock returns only from 1999 to 2001 where it outperforms other predictors such as the dividend-price ratio, the VIX, or the Treasury term spread. For the periods from 1994 to 1998 and 2002 to 2007 there is no evidence of such predictability. In addition, the documented statistical relation between the measured tone of Fed Board governors' speeches as well as the fed funds futures slope is sensitive to the inclusion of two individual speeches, which, reading in detail and given the stock market's immediate reaction, do not appear particularly hawkish. Combined, our findings and those of Neuhierl and Weber (2019) call for further analysis of the link between monetary policy communication, market expectations of the policy stance, and the stock market.

References

Acosta, M., Meade, E.E., 2015. Hanging on every word: semantic analysis of the FOMC's postmeeting statement. FEDS Notes, September. Board of Governors of the Federal Reserve System, Washington. doi:10.17016/2380-7172.1580.
Apel, M., Grimaldi, M. B., 2012. The information content of central bank minutes. Sveriges Riksbank Working Paper Series No. 261.
Bernanke, B.S., Kuttner, K.N., 2005. What explains the stock market's reaction to federal reserve policy? J. Finance 60 (3), 1221–1257.
Ehrmann, M., Talmi, J., 2019. Starting from a blank page? semantic similarity in central bank communication and market volatility. J. Monet. Econ. forthcoming.
Farmer, L., Schmidt, L., Timmermann, A., 2019. Pockets of predictability. doi:10.2139/ssrn.3152386.
Hansen, S., McMahon, M., 2016. Shocking language: understanding the macroeconomic effects of central bank communication. J. Int. Econ. 99, S114–S133.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. J. Finance 66 (1), 35–65.
Lucca, D.O., Moench, E., 2015. The pre-FOMC announcement drift. J. Finance 70 (1), 329–371.
Neuhierl, A., Weber, M., 2019. Monetary policy communication, policy slope, and the stock market. J. Monet. Econ. 108.
Solomon, D.H., 2012. Selective publicity and stock prices. J. Finance 67 (2), 599–637.
Tietz, R., 2019. How does the fed manage interest rate expectations? doi:10.2139/ssrn.3254094.