Yield Curve Trading Strategies Exploiting Sentiment Data*

Francesco Audrino & Jan Serwart

December 12, 2022

Abstract

This paper builds upon previous research findings that show macro sentiment data-augmented models are better at predicting the yield curve. We extend the dynamic Nelson-Siegel model with macro sentiment data from either Twitter or RavenPack. Vector autogressive (VAR) models and Markov-switching VAR models are used to predict changes in the shape of the yield curve. We build bond butterfly trading strategies that exploit our yield curve shape change predictions. Although the economic returns from our trading strategies based upon models exploiting macro sentiment data do not statistically significantly differ from those which do not rely on it, we find some evidence that models exploiting inflation sentiment are economically useful when trading the curvature of the yield curve.

JEL classifications: G12, E43, C32, E52

Keywords: bond butterflies, yield curve, sentiment data

*We thank participants of various seminars for valuable comments. Any remaining errors are our own.

Address: Jan Serwart, University of St.Gallen, Faculty of Mathematics and Statistics, Bodanstrasse 6, 9000 St.Gallen, Switzerland. E-Mail: jan.serwart@unisg.ch.

1 Introduction

The idea that asset prices are in part driven by investors' psychology has been under academic scrutiny since the early 1990s; see e.g. Daniel et al. (2002) for a literature review. Recent advances in data sciences and the ubiquitous use of social media platforms by financial market participants allowed for the creation of novel market sentiment data. These data complement the survey data on consumers' expectations, which had been available long before the advent of the Internet. The way sentiment data can be used to explain financial market developments has attracted significant scholarly attention in previous years. The potential of sentiment data to explain stock market developments is much better understood compared to fixed income markets. A better understanding of how sentiment data can describe fixed income market developments is warranted given its significance in terms of overall value, connectedness to the real economy and its effects on governments and private entities with interest rate exposure.

This paper extends the existing literature on the applicability of sentiment data in fixed income markets. Audrino and Offner (2022) show that sentiment data statistically significantly increase the predictive power of simple Taylor rule-based interest rate models as well as term structure models of the Nelson-Siegel family and arbitrage-free, short rate-based models. However, thus far no concrete economic application has been tested, that would show that the use of sentiment data also translates into significantly higher economic returns. This paper fills the gap.

We use a term-structure model of the Nelson-Siegel family to estimate yield curve factors. Applying simple vector autoregressive models (VAR) and Markov-switching VAR (MSVAR) models augmented with macroeconomic sentiment data and monetary policy indicators, we forecast the Nelson-Siegel yield curve factors to predict changes in the shape of the yield curve. We exploit these forecasts using a bond butterfly trading strategy.

The empirical results show that the inclusion of sentiment data in models to forecast changes in the shape of the yield curve can have economic benefits. However, concerning bets on the slope of the yield curve, the outperformance of sentiment-augmented models over non-sentiment-augmented models is neither statistically significant nor robust over time. Nevertheless, we find mild evidence that models exploiting inflation sentiment are economically useful when trading the curvature of the yield curve. Inflation sentiment-augmented models consistently rank among the top models for all considered trading periods and deliver economic returns that outperform the benchmark, especially during the Coronavirus pandemic. This paper proceeds as follows: Section 2 summarizes related literature, followed by Section 3 describing the data. Section 4 presents the statistical models, which is followed by Section 5 introducing the trading strategy. Finally, Section 6 discusses the empirical results. Section 7 concludes.

2 Literature Review

This paper contributes generally to the term structure literature given impetus by Diebold and Li (2006), who first showed that the time series of the factors from a dynamic Nelson-Siegel model (DNS) contain information that can be exploited for yield curve forecasting purposes using low order vector autoregressions (VAR). Diebold et al. (2006) extended the basic Nelson-Siegel approach to include observable macroeconomic variables. However, the evidence concerning the superior forecasting power of these macroeconomic variable-augmented DNS models remains somewhat mixed (Duffee, 2011). Recently, a series of models was introduced to account for the well-known stylized fact that yields are subject to regime shifts; for an overview, see Filipova et al. (2014). For example, Xiang and Zhu (2013) show that Markov regime-switching DNS models for the U.S. yield curve have superior predictive performance compared to single-state DNS models. Hevia et al. (2014) showed similarly in a no-arbitrage context that allowing for Markov regimes in DNS models can entail superior predictive performance for some parametrizations compared to single regime DNS VAR models. Guidolin and Pedio (2019) drop the no-arbitrage requirements and show that Markov-switching VAR models augmented with monetary policy indicators outperform models without such indicators in terms of the forecast accuracy of U.S. yield curve. So far, however, little research has appeared on whether these models can benefit from the inclusion of macro sentiment data. Modern sentiment data is often constructed using methods from the field of natural language processing (NLP).

The advances in the field of NLP have been significant in past years due to increasing computational power. One stream of NLP research is devoted to the development of methods and models to harness quantitative sentiment data from raw text to be used in empirical analysis. Recent survey articles on the developments in this field include Algaba et al. (2020) and Renault (2019). The progress in the field of NLP has led to an increase in studies researching the potential of sentiment data to predict asset price movements. Most notably, the effectiveness of sentiment data created on the basis of text from social media platforms like Twitter or Stocktwits has gained a lot of attention (Audrino et al. (2020); Bollen et al. (2011); Nofer & Hinz (2015); Oliveira et al. (2013); Renault (2017); Sul et al. (2016); Sun et al. (2016); Tan & Tas (2020)).

The sentiment literature pertaining to interest rate developments is less developed. Early work focused on sentiment analysis of central bank communication mediums such as FOMC statements. For example, Lucca and Trebbi (2009) find that the tone of policy communications influences longer-dated treasury yields. Other non-central bank-related sources for sentiment analysis in the realm of interest rate research include the general news. In this regard, RavenPack data have become frequently used. RavenPack marks every news item from a broad universe of news sources with a sentiment value. For example, Erlwein-Sayer (2017) using RavenPack data investigates the relationship between news sentiments and European sovereign yield spreads. Audrino et al. (2022) show that augmented Nelson-Siegel (NS) models with sentiment data from RavenPack have superior predictive performance when forecasting the U.S. yield curve for different maturities.

Moreover, other studies focus directly on specific macroeconomic data releases and their effects on interest rates. In this respect, Edison (1997) researches the reaction of short- and long-term treasury yields to unemployment and inflation-related news. She finds that positive inflation deviations from expectations increase yields while positive unemployment deviations from expectations lower yields. Goldberg and Leonard (2005) show that the largest moves in yields are driven by U.S. announcements on labor market conditions, real GDP growth, and consumer sentiment.

Concerning yield curve movements, the contribution by Wright (2011) found that the term structure of yields reacts to changes in inflation uncertainty as measured by variation in monetary policy frameworks. Furthermore, Goldberg and Grisse (2013) find evidence that the reaction of the U.S. Treasury yield curve to economic data is not constant over time. Recently, Gotthelf and Uhl (2018) introduced the idea of a novel sentiment factor besides the well-known level-slope-curvature factors in a Nelson-Siegel context. They show that sentiment data improve the prediction accuracy of yields especially at the short end of the curve.

While the literature on forecasting treasury yield curves is well established, few authors have developed concrete trading strategies that exploit the forecasting power of their models. Fabozzi et al. (2005) use the forecasts on changes in the shape parameters of the Nelson-Siegel model in order to build swap butterflies. These are among the most commonly used active trading strategies to exploit predictions concerning the shape of the yield curve. They show that swap butterflies based on their model earn economically significant returns, but they do not statistically benchmark their results against other forecasting models. Recently, Guidolin and Pedio (2019) used a bond butterfly trading strategy to show that the superior statistical performance of their yield curve forecasting models that include monetary policy indicators to predict changes in the shape of the yield curve can translate into higher trading profits. However, the realized returns are very uncertain, and as a result they cannot statistically confirm return differences across models. This paper extends the literature with a concrete trading application based on DNS factor forecasts from MSVAR models augmented with macro sentiment data from diverse sources.

3 Data

3.1 Measuring Sentiment

We use three different sources to create macroeconomic sentiment data: RavenPack, Twitter, and the Michigan consumer sentiment surveys.

RavenPack News Analytics provides a global macro dataset, of which the Event Sentiment Score (ESS) is used as sentiment variable. The Event Sentiment Score is calculated by applying a supervised Bayes classifier trained with news articles labelled by financial experts as having either a positive or negative short-term effect on the economy. The algorithm is applied on news articles published by Dow Jones Newswires, Wall Street Journal, Barron's, and Market Watch. The ESS sentiment scores range from 0 (negative) to (100) positive. The RavenPack sentiment scores have been found to contain information exploitable for forecasting purposes for German Bund yields, various European sovereign spreads, and CDS spreads (Erlwein-Sayer, 2017, 2018; Yang, Liu, & Wang, 2020).

The macroeconomic variables assumed to have an effect on the yield curve are inflation, unemployment and the Federal Funds Rate (FFR), which we also denote short rate. We build sentiment variables concerning inflation, unemployment and the short-term interest rate. For this choice, we follow Audrino and Offner (2022), who have shown that such sentiment variables have predictive power for the yield curve. For the sake of comparability, we use the same procedure as Audrino and Offner (2022) to construct the sentiment variables. Specifically, for all articles related to one of the macroeconomic variables, we download its ESS scores from RavenPack. In a next step, we aggregate the ESS scores for each macroeconomic variable monthly by taking the mean. Thus, the sentiment score can be stated as:

$$S_h(t) = \frac{1}{N} \sum_{j=1}^{N} SS(k_{j,h}) \mathbb{1}_{\{Month(k_{j,h})=t\}}, \qquad \text{for } h = i, \pi, u, \tag{1}$$

where SS stands for one ESS sentiment score, $k_{j,h}$ corresponds to one news item about h, where i, π, and u stand for interest rates, inflation and unemployment rate, respectively. N is the number of observed articles in a month, and t denotes the respective month. The applied keywords and summary statistics are given in Appendix A.1 in Table A.1.

Twitter is a social media network with roughly 400 million active daily users. On the platform, people can share thoughts, ideas, and opinions in the form of short messages currently consisting of currently 280 characters.[1] The data are obtained using an academic API terminal provided by Twitter. For each of the macroeconomic variables inflation, interest rates and unemployment, respectively, we download all tweets that contain specific keywords related to the underlying macroeconomic variables. The keywords and sentiment summary statistics are given in Appendix A.1 in Table A.2. Due to academic Twitter API limitations, the entirety of keywords used for the RavenPack data set could not be used for the Twitter dataset. However, since at least the eponymous macroeconomic variable keyword was used, we capture the macroeconomic sentiment sufficiently well. The econometric results do not contradict this assumption. Overall, 4.35 million tweets were analyzed. In order to assign a sentiment value to each tweet, we use a dictionary approach. Dictionaries are a set of words labelled to express a certain sentiment. Specifically, the sentiment of a text is determined by counting the words therein that also appear in the dictionary. In this paper, we use the dictionary provided by Loughran and McDonald (2011), henceforth LM. In an extensive analysis of different approaches to gauging investor sentiment, Ballinari and Behrendt (2021) show that the easy to implement and transparent dictionary approach performs best in an asset pricing application. Moreover, the models estimated using the LM dictionary belong to Hansen model confidence set together with those using the finance-specific dictionary by Renault (2017). For every tweet, we count the number of positive and negative words used in the Twitter post as given by the LM dictionary. The sentiment of a given social media message is calculated as the difference between the share of positive and negative words occuring in the text data. As for the RavenPack dataset, we aggregate the individual tweet sentiment scores for each macroeconomic variable monthly by taking the mean. Hence, the sentiment can be calculated analogously as in equation (1).

Finally, we use the *University of Michigan Consumer Sentiment Index* as a third sentiment dataset. This dataset contains information pertaining to consumer expectations for the economy. The sentiment score reflects people's expectations concerning their current financial health, as well as the health of the economy in the short- and long-term. The sentiment scores cannot be disaggregated at the level of our macroeconomic variables but reflect sentiment towards the economy as a whole.

[1] Before November 7, 2017, the limit was 140 characters.

3.2 Relation to Macroeconomic data

The potential of sentiment data to improve the forecast accuracy of our analysis depends crucially on the assumption that the extracted sentiment data bear a relationship to the underlying macroeconomic variables that affect interest rates. Figure 1 shows the relationship between the macroeconomic variables and their Twitter sentiment counterpart. We observe a statistically significant positive correlation between the change of short-term interest rates and its contemporaneous sentiment counterpart as depicted in Table 1. This is in sharp contrast to the results when using RavenPack data as depicted in Appendix A.2, where we find a negative relationship between those two variables. Audrino and Offner (2022) argue that the correlation is negative for the RavenPack data because "[the RavenPack] sentiment variable is mainly based on the perception of news based directly on interest rates. As declining interest rates boost the performance of fixed income securities, it is possible that investors will rate as positive news which reports falling interest rates" (p. 20). They reason in such a way because other studies such as those by Gotthelf and Uhl (2018), who find a positive relationship, are based on all articles spanning the entire Thomson Reuters News Analytics universe, which are both mostly related to business cycles. Following this logic, the tweets from Twitter on interest rates must also be related (mostly) to business cycles. Furthermore, we find a negative correlation between inflation and unemployment growth and their contemporaneous counterparts, which is the same for the RavenPack data. These results indicate that our sentiment variables are reasonably connected to the underlying macroeconomic variables and thus should contain exploitable information for the purpose of yield curve forecasting.

Concerning the lead-lag structure between macroeconomic variables and their sentiment counterparts, we find that the direction of the correlation between the macroeconomic variables and their lead or lagged sentiment counterpart is the same as the contemporaneous correlation. Furthermore, we observe that contemporaneous inflation is statistically significantly and positively correlated with lead and lagged unemployment sentiment.

Figure 1: Sentiment vs. Macroeconomic Variables

This figure shows different macroeconomic variables in blue against their sentiment counterpart in yellow. All time series are standardized. The data cover the period from January 2008 to May 2022.

	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$	$S_{i,t+1}$	$S_{\pi,t+1}$	$S_{u,t+1}$	$S_{i,t-1}$	$S_{\pi,t-1}$	$S_{u,t-1}$
Δi_t	**0.22**	0.05	-0.15	**0.17**	0.02	-0.12	**0.26**	0.04	-0.04
π_t	-0.06	**-0.24**	**0.31**	-0.05	**-0.23**	**0.34**	-0.06	**-0.22**	**0.3**
u_t	-0.11	0.08	-0.11	-0.15	0.13	**-0.21**	-0.15	0.05	-0.06

Table 1: Correlation Analysis Sentiment vs. Macroeconomic Variables
The table depicts the correlation between the Twitter sentiment variables (contemporaneously, lead one period and lagged one period) and their macroeconomic counterparts. Δi_t, π_t and u_t represent the first difference in the short rate, inflation and unemployment growth, respectively. S_i, S_π, S_u denote the sentiment on each macroeconomic variable. Bold font indicates a p-value lower than 0.05, testing whether the true correlation coefficient is equal to zero. Due to the relatively small number of posts during the initial years of Twitter, the time period ranges from January 2011 to May 2022.

3.3 Yield Curve Data

To construct the U.S. Treasury zero-coupon yield curve, we use the factors made available by Gürkaynak et al. (2007). We obtain the time series of yields with maturities of 3-, 12-, 24-, 36-, 48-, 60-, 72-, 84-, 96-, 108-, and 120-months for the time period January 2008 until May 2022. We provide summary statistics for the constant maturity yields in Table 2. On average, the yield curve is increasing and concave, consistent with established stylized facts. Furthermore, we see that the standard deviation is on average decreasing with maturity, which is consistent with the idea that future yields contain expected values (thus average) of future short-term yields (Audrino & Offner, 2022). We also document non-normality of the yields in terms of negative excess kurtosis and positive skewness for the yields at all maturities. The finding that the serial correlation on average decreases with maturity differs from the results in Diebold and Li (2006). Finally, in Panel *B*, we show that the correlations of yields across maturities is high and on average the highest for yields with adjacent maturities, ranging between 0.92 and 0.99.

Panel A

	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year
Mean	0.72	0.72	0.77	0.96	1.19	1.66	2.06	2.49
Median	0.27	0.28	0.35	0.71	1.03	1.65	2.07	2.4
Maximum	2.71	2.66	2.7	2.86	2.92	3.46	3.89	4.84
Minimum	0.03	0.08	0.07	0.11	0.14	0.24	0.37	0.55
Std. Deviation	0.76	0.77	0.78	0.77	0.74	0.74	0.79	0.9
Skewness	1.18	1.14	1.09	0.98	0.72	0.17	0.08	0.21
Kurtosis	2.93	2.81	2.75	2.78	2.65	2.38	2.48	2.6
ACF 1	0.96	0.97	0.96	0.95	0.94	0.94	0.94	0.95

Panel B

Correlations	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year
3-month	1							
6-month	0.99	1						
1-year	0.96	0.99	1					
2-year	0.89	0.93	0.98	1				
3-year	0.81	0.86	0.91	0.98	1			
5-year	0.6	0.64	0.71	0.83	0.92	1		
7-year	0.4	0.44	0.5	0.63	0.76	0.95	1	
10-year	0.22	0.24	0.29	0.41	0.57	0.83	0.96	1

Table 2: U.S. Treasury Yield Curve Summary Statistics
This table shows in Panel *A* summary statistics for the U.S. Treasury yield curve at various maturities. The data are expressed in annualized percentage terms. We use monthly data for the sample period January 2008 - May 2022. Panel *B* depicts the correlations between pairs of Treasury yields.

In addition, by analyzing first differences of yields, it becomes evident that the volatility is not constant. In Figure 2, we depict the path of the first difference of the 3-month and the 10-year yield. For both series, we see spikes in volatility around the Great Financial Crisis (GFC) in 2008. Furthermore, the outbreak of the Coronavirus in spring 2020 led to another visible spike in the volatility of the 3-month yield. Such volatility clusters in the yields have attracted scholarly attention (Audrino, 2006; Audrino & Medeiros, 2010) and additionally justify the implementation of regime-switching models for yield curve modelling.

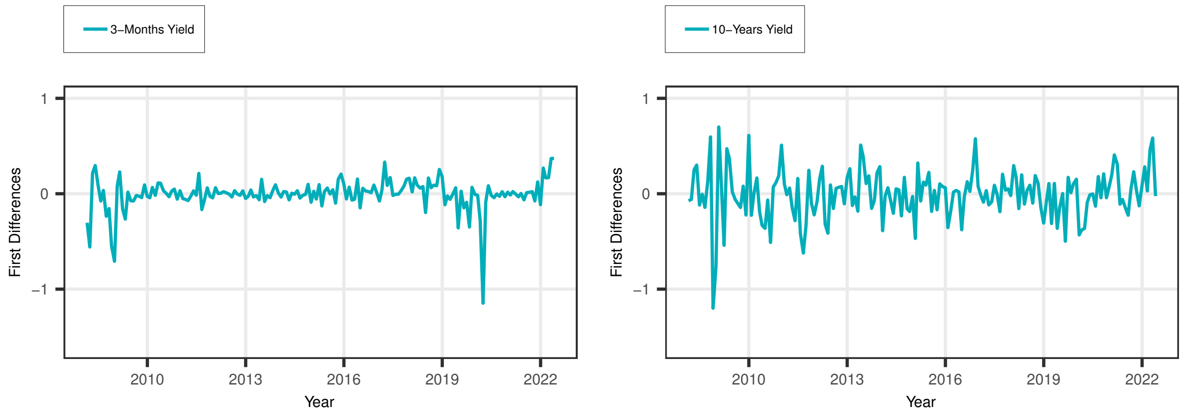

Figure 2: Yields First Differences

This figure shows the first differences of 3-month and 10-year U.S. Treasury yields from January 2008 to May 2022.

3.4 Monetary policy indicators

Guidolin and Pedio (2019) show that variables capturing the stance of monetary policy add predictive power to the standard DNS factors when forecasting the U.S. Treasury yield curve. We extend our benchmark analysis by also comparing our sentiment-augmented models against the monetary policy indicator-augmented models presented in Guidolin and Pedio (2019). Specifically, we include the same five variables that represent the features of monetary policy adopted by the Fed after the Great Financial Crisis in 2008. Two variables are chosen to represent the size of the Fed's balance sheet. The first is the level of the Fed's total asset, for which we use the data series published on FRED by the Federal Reserve Bank of St. Louis.[2]. The second is a divisia money aggregate index. We use the log level of the M3 Index for the United States obtainable on FRED.[3,4] Next, we include two variables that approximate the composition of the Fed's balance sheet.

[2]The series is found using the label *WALCL*.

[3]The series is labelled *MABMM301USM189S*.

[4]We do not use the MZM Index as in Guidolin and Pedio (2019) since its publication terminated in 2021.

The first variable describes the percentage of Treasuries on the Fed's total balance sheet. The total amount of securities outstanding in any given month is obtained on FRED. For the analysis, we use the corresponding end of month values and divide them with the size of the total balance sheet.[5] The second variable describing the average maturity of the Fed's portfolio of Treasuries is constructed by taking the sum of the weighted maturities, where the weights correspond to the percentage of Treasuries with a specific maturity of the total amount of Treasuries outstanding. Finally, we use the Fed funds rate to proxy the interest rate policy of the Fed.[6] For a detailed description of how these variables relate to monetary policy, we refer to Guidolin and Pedio (2019).

4 Statistical Models

4.1 Baseline Model

Our starting point is the standard Nelson and Siegel (1985) model to fit the forward yield curve in a static setting:

$$f(\tau) = \beta_1 + \beta_2 e^{-\tau\lambda} + \beta_3 \lambda\tau e^{-\tau\lambda}, \tag{2}$$

where τ is the maturity, and $\beta_1, \beta_2, \beta_3$ and λ are function parameters. The connection between the instantaneous forward rate and the yield is derived from the no-arbitrage theory and given by:

$$y(\tau) = \frac{1}{\tau}\int_0^\tau f(u)du. \tag{3}$$

Equation (4) represents the dynamic Nelson-Siegel model (DNS) which obtains from inserting (2) into (3) and allowing for time dynamics as in Diebold and Li (2006):

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right) + \beta_{3,t}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\right). \tag{4}$$

$y_t(\tau)$ represents the yield to maturity τ at time t and $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ are the latent factors, which control the dynamics of $y_t(\tau)$ for all maturities. 1, $\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)$, and $\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\right)$

[5] The FRED label for the data series describing the total amount of Treasuries outstanding is *TREAST*.

[6] The data are downloadable from FRED using the label *FEDFUNDS*.

are the factor loadings. In terms of interpretation of the DNS factors, we follow Diebold and Li (2006) and rename $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ as level factor L_t, slope factor S_t and curvature factor C_t. Since the implied term structure gets steeper the lower the slope factor is, it is actually a (negative) slope factor. Following the established literature — Byrne et al. (2017), Coroneo et al. (2016), Diebold and Li (2006), Guidolin and Pedio (2019) — we fix the decay parameter λ at 0.0609. Consequently, the curvature factor is maximized at exactly 30-months, consistent with the discussion by Yu and Zivot (2011). We use the state-space representation introduced by Diebold et al. (2006) to get $N \cdot T$ measurement equations. Specifically, the vector of measurement equations reads:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_T(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_T(\tau_N) \end{pmatrix}, \tag{5}$$

where $\epsilon(\tau)$ represents maturity-specific errors, which are IID over time and simultaneously cross-sectionally uncorrelated as in Diebold, Rudebusch, and Aruoba (2006). Fixing the decay parameter λ allows us to estimate (5) using OLS. For every point in time t, we fit the riskless yield curve obtaining the three-dimensional time series of DNS factors, $\left\{\hat{L}_t, \hat{S}_t, \hat{C}_t\right\}_t^T$. The latter are used to build a multivariate Gaussian VAR(p) model that constitutes our baseline model. Notationwise, we follow Guidolin and Pedio (2019) and describe the model as:

$$\boldsymbol{f_{t+1}} = \begin{pmatrix} L_{t+1} \\ S_{t+1} \\ C_{t+1} \end{pmatrix} = \begin{pmatrix} \mu_L \\ \mu_S \\ \mu_C \end{pmatrix} + \sum_{j=1}^{p} \begin{pmatrix} \phi_{1,1}^j & \phi_{1,2}^j & \phi_{1,3}^j \\ \phi_{2,1}^j & \phi_{2,2}^j & \phi_{2,3}^j \\ \phi_{3,1}^j & \phi_{3,2}^j & \phi_{3,3}^j \end{pmatrix} \begin{pmatrix} L_{t+1-j} \\ S_{t+1-j} \\ C_{t+1-j} \end{pmatrix} + \begin{pmatrix} \eta_{L,t+1} \\ \eta_{S,t+1} \\ \eta_{C,t+1} \end{pmatrix}, \tag{6}$$

which we abbreviate as follows:

$$\boldsymbol{f}_{t+1} = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{f}_{t+1-j} + \boldsymbol{\eta}_{t+1}, \tag{7}$$

where $\boldsymbol{\mu}$ represents the vector of intercepts, $\boldsymbol{\Phi}_j$ is a 3x3 matrix that determines factor dynamics, and $\boldsymbol{\eta}$ is the innovation vector for which we assume normality such that $\boldsymbol{\eta}_{t+1} \sim N(0, \boldsymbol{\Sigma})$. We do not impose restrictions on $\boldsymbol{\Sigma}$ such that shocks to the DNS factors may exhibit non-zero contemporaneous correlations.

4.2 Extension to include sentiment variables and monetary policy indicators

We extend the baseline model in (7) to include sentiment variables related to the state of the macroeconomy and variables pertaining to monetary policy as in Guidolin and Pedio (2019). The additional Q variables are collected in a Q x 1 vector times series denoted $\boldsymbol{m}_{t+1}$. The additional variables in $\boldsymbol{m}_{t+1}$ are included in a bi-directional manner, such that they can affect the dynamics of the DNS factors and vice-versa be predicted by the DNS factors. For a discussion, see for example Aguiar-Conraria et al. (2012), who provide evidence for shifting relationships between the shape of the yield curve and macroeconomic determinants. Let $\boldsymbol{x}_{t+1} \equiv \left[\boldsymbol{f}'_{t+1}\boldsymbol{m}'_{t+1}\right]$ denote the entire vector consisting of the DNS factors and the additional Q variables at time $t+1$. The augmented baseline model from equation (7) is adjusted to:

$$\boldsymbol{x}_{t+1} = \boldsymbol{c} + \sum_{j=1}^{p} \boldsymbol{A}_j \boldsymbol{x}_{t+1-j} + \boldsymbol{\zeta}_{t+1}, \tag{8}$$

where $\boldsymbol{c}$ represents the new vector of intercepts, $\boldsymbol{A}_j$ is a $(3+Q)$ x $(3+Q)$ matrix that governs the dynamics between the DNS factors and the additional variables, and $\boldsymbol{\zeta}$ represents the innovation vector for which we also assume normality such that $\boldsymbol{\zeta}_{t+1} \sim N\left(0, \boldsymbol{\Omega}\right)$. We do not impose restrictions on $\boldsymbol{\Omega}$, such that shocks to the variables can have non-zero contemporaneous correlations.

4.3 Markov switching vector autoregressive models

We extend the analysis and allow for regime switches in the analysis since the literature has shown their benefits when forecasting the dynamics of the riskless interest rate curve: see, among others Audrino and Offner (2022), Filipova et al. (2014), Guidolin and Timmermann (2009). For Markov-switching vector autoregressions with K states and p lags, MSVAR(K,p), we assume that there is a latent state-variable $S_{t+1} = 1, ..., K$, which determines the state of the system and thus causes shifts in the parameters of the model. We only allow for one lag $p = 1$ and two states $K = 2$ in order to avoid over-parametrization. Our MSVAR(2,1) model can be written as:

$$\boldsymbol{x}_{t+1} = \boldsymbol{c}_{S+1} + \boldsymbol{A}_{S_{t+1}}\boldsymbol{x}_t + \boldsymbol{\Omega}^{1/2}_{S_{t+1}}\boldsymbol{\nu}_{t+1} \qquad \boldsymbol{\nu}_t \sim IID \quad N(0, \boldsymbol{I}_M), \tag{9}$$

where $\boldsymbol{\nu}_{t+1} \equiv \boldsymbol{\Omega}^{-1/2}_{S_{t+1}}\boldsymbol{\zeta}_{t+1}$. Furthermore, note that $\boldsymbol{\Omega}^{1/2}_{S_{t+1}}\left(\boldsymbol{\Omega}^{1/2}_{S_{t+1}}\right)' = Var\left(\boldsymbol{\zeta}_{t+1} \mid S_{t+1}\right)$. Concerning the Markov property, we assume that the latent state variable S_{t+1} is the result of a discrete, time homogeneous, irreducible and ergodic first-order Markov chain such that $Pr\left(S_t = j \mid \{S_j\}^{t-1}_{\tau=1}, \{\boldsymbol{x}_l\}^{t-1}_{l=1}\right) = Pr\left(S_t = j \mid S_{t-1} = i\right) = p_{i,t}$, where $p_{i,j}$ is the $[i, j]$ element of the K x K transition matrix, $\boldsymbol{P}$ as in (Guidolin & Pedio, 2019). We estimate the MSVAR models using maximum likelihood, applying the blockwise optimization technique introduced by Sims et al. (2008). In order to forecast the yield curve factors in the MSVAR models, we follow Hamilton (1994). The forecasts can be derived as:

$$\hat{\boldsymbol{x}}_{t+1} = \sum_{S=1}^{K} \hat{\xi}_{S_{t+1}}\hat{\boldsymbol{A}}_{S_{t+1}}\boldsymbol{x}_t, \tag{10}$$

where $\xi_{S_{t+1}}$ denotes the probability that at $t + 1$ state S will prevail. Furthermore, we use $E\left[\boldsymbol{\xi}_{t+1} \mid \boldsymbol{\xi}_t\right] = \boldsymbol{P}'\boldsymbol{\xi}_t$: see Hamilton (1994).

5 Trading Strategy Description

In this section, we dig deeper into earlier results and investigate whether and to what extent sentiment-augmented models used to predict the interest term structure can be applied to implement a systematic trading strategy. Specifically, we build bond butterfly strategies and exploit our macro sentiment data. We explain how to build butterfly strategies and illustrate how to measure their economic value.

The choice of bond butterflies as a trading strategy is motivated by their ubiquitous use in practice to profit from views on changes in the shape of the yield curve. We follow Guidolin and Pedio (2019) and Fabozzi et al. (2005), who have shown how to use them within a Nelson-Siegel framework. Specifically, we build semi-hedged, long-short, self-financing strategies in which we are exposed either to the slope or curvature factor, while staying neutral to the other factors. For example, if we predict and bet on a change in the slope of the yield curve, we neutralize our exposure of the portfolio to level and curvature movements. Concretely, we build long and short positions at different maturities to hedge against level and curvature movements of the yield curve. For example, we predict a change in the slope of the yield curve between t and $t+1$, but we have no view concerning the development of the curvature and the level of the curve. We build a duration-weighted hedged and self-financing bond butterfly using three bonds with short (2-years), medium (5-years), and long maturity (10-years). We obtain the portfolio weights, from the following system of equations:

$$\begin{cases} q_s\, MD_s\, L_s + q_m\, MD_m\, L_m + q_l\, MD_l\, L_m = 0, \\ q_s\, MD_s\, C_s + q_m\, MD_m\, C_m + q_l\, MD_l\, C_m = 0, \\ q_s + q_m + q_l = 1, \end{cases} \tag{11}$$

where q_s, q_m, and q_l are the weights of the short-, medium-, and long-term bonds, MD_s; MD_m and MD_l are the modified durations of the short-, medium-, and long-term bonds; L_s, L_m and L_l are the sensitivities of the short-, medium-, and long-term yields to the level factor as calculated in Nelson-Siegel's model; and finally, C_s, C_m and C_l, are the exposures of the short-, medium-, and long-term yields to the curvature factor (i.e. the loadings of curvature).

We build butterflies based on 2-, 5-, and 10-year bonds.[7] The direction of the butterfly trades can take two distinct forms: barbell and bullet. In a barbell butterfly, we go long on the wings and short the body. This means that we buy the bonds with 2-years and 10-years maturity and sell bonds with 5-years maturity. In a bullet butterfly we buy the body and sell the wings. Which strategy is executed depends on our yield curve shape forecasts generated by the models described in Section 4. We re-estimate the model every month and use a rolling window for the estimation. As in Fabozzi et al. (2005), when we predict an increase (decline) of the slope or curvature factor, we implement the bullet (barbell) butterfly if the sensitivity is positive. The sensitivity of a Nelson-Siegel-weighted butterfly is derived in Martellini et al. (2003, p. 197). The

[7] We do not consider other butterfly schemes such as the (2/5/7) or (2/10/30) butterfly in order to avoid multiple testing issues.

positions are held for a period of one month. As in Fabozzi et al. (2005) we find negative (positive) sensitivities of our butterflies to the slope (curvature) factor.

In order to measure the economic performance of the strategy outlined above, we determine the monthly returns and average them over the entire out-of-sample period. Monthly returns are calculated using the definition of the modified duration of a bond. The percentage return of a given bond between t and $t+1$ is commonly approximated by the negative of the product of the modified duration and the change in the yield between t and $t+1$. The return of a bond $r_{i,t}$ in a given month t is given by:

$$r_{i,t} = -MD_i \cdot \Delta Y_i, \tag{12}$$

where MD_i is the modified duration of bond i and ΔY_i represents the change in the yield of bond i in percentage points between period t and $t+1$. The return approximation in equation (12) neglects convexity effects on returns. If changes in the yields are small, which they usually are on a monthly level, the approximation is justified. We weight the monthly returns using the weights q obtained in equation (11). Hence, the return of the butterfly in a given month can be stated as:

$$R_t = \sum_{j=s,m,l} q_{j,t} r_{j,t} \tag{13}$$

where s, m, l represent short-, medium-, and long-term maturity. Furthermore, for the case when we short the butterfly in order to execute a bullet trade, we invest the proceeds for one month at the risk-free interest rate. Our trading period ranges from January 2016 to May 2022. Hence, in total we have 77 monthly butterfly returns, of which we calculate the monthly average performance.

Concerning transaction costs, we follow the literature (Guidolin & Pedio, 2019) and deduct from the monthly returns the costs for funding the margin positions, assumed to be ten percent of the current repo rate. The repo rates are given by the Federal Reserve Bank of New York.

Furthermore, we calculate the Sharpe ratio SR_p of the returns for trading period p, which is given by:

$$SR_p = \frac{\frac{1}{P}\sum_{t=1}^{P} R_t}{\sqrt{\frac{1}{P}\left(\sum_{t=1}^{P} R_t - \frac{1}{P}\sum_{t=1}^{P} R_t\right)^2}}, \tag{14}$$

where the numerator does not include the risk-free rate adjustment because the strategy derived in equation (11) is zero net-outlay, which means that any funding costs are deducted from the total return; see Guidolin and Pedio (2019).

6 Empirical Results

6.1 Discussion Term Structure Models

Table 3 shows the results of the sentiment data-augmented dynamic Nelson-Siegel model described in equation (8). All three factors (level, slope, curvature) are highly persistent, while the degree of persistence varies across factors. In line with stylized facts, the model implies that long-term yields are less persistent than short-term yields. This is because the long-term yields are mainly driven by the least persistent level factor, whereas short-term yields are affected both by the level and the slope factor.

We observe statistically significant autoregressive coefficients at the 1% level for all variables except inflation sentiment, which is significant at the 10% level. Cross-sentiment dynamics is as follows: The sentiment on interest rates influences negatively the sentiment on unemployment and vice versa. Concerning sentiment-factor dynamics, we find that the inflation sentiment positively affects the level factor and that the unemployment sentiment negatively affects the slope factor. Concerning lagged effects of factors on sentiment, we find statistically significant impacts of the lagged level factor on the unemployment sentiment as well as from the lagged curvature factor on current interest and unemployment sentiment.

	Dependent Variables					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_{t-1}	0.907^{***}	0.075	-0.137	0.170	-0.050	0.245^{**}
	(0.042)	(0.048)	(0.111)	(0.154)	(0.152)	(0.103)
S_{t-1}	-0.057^{*}	0.975^{***}	-0.007	-0.175	-0.028	0.066
	(0.030)	(0.035)	(0.090)	(0.114)	(0.121)	(0.086)
C_{t-1}	0.005	0.042^{***}	0.935^{***}	0.077^{*}	0.051	0.073^{**}
	(0.014)	(0.014)	(0.043)	(0.042)	(0.043)	(0.031)
$S_{i,t-1}$	0.002	-0.004	0.0001	0.216^{***}	-0.027	-0.105^{*}
	(0.032)	(0.028)	(0.082)	(0.083)	(0.097)	(0.058)
$S_{\pi,t-1}$	0.040^{*}	-0.034	-0.067	-0.039	0.140^{*}	-0.092
	(0.023)	(0.025)	(0.052)	(0.096)	(0.084)	(0.075)
$S_{u,t-1}$	0.037	-0.075^{**}	0.058	-0.226^{**}	-0.187	0.588^{***}
	(0.027)	(0.030)	(0.088)	(0.104)	(0.124)	(0.089)
μ	0.192^{**}	-0.136	0.206	-0.691^{**}	0.332	-0.365^{*}
	(0.092)	(0.099)	(0.262)	(0.297)	(0.302)	(0.213)

	Estimated Correlation Matrix					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_t	-	-	-	-	-	-
S_t	-0.82***	-	-	-	-	-
C_t	-0.30***	-0.01	-	-	-	-
$S_{i,t}$	-0.10	0.14	0.06	-	-	-
$S_{\pi,t}$	-0.09	0.07	0.11	0.09	-	-
$S_{u,t}$	0.12	-0.21**	-0.02	-0.21***	-0.14	-
Observations	172	172	172	172	172	172
Adjusted R^2	0.935	0.951	0.915	0.182	0.055	0.554

Table 3: Regression Output of the Dynamic Nelson Siegel Model
The upper table shows the coefficients of the first-order VAR process for the yield curve factors and sentiment from January 2008 to May 2022. The factors are estimated under the Dynamic Nelson-Siegel model. HC3 standard errors are shown in parentheses. The middle table reports the residual correlation matrix. Inference is made under the assumption of white-noise residuals. ***, **, and * denote significance at the 1%, 5% and 10% level, respectively. The adjusted R squared describes the model fit for the respective dependent variable.

In Table 4 and 5 the coefficients of the MSVAR model from equation (9) are depicted. The first (second) table depicts the coefficients of regime 1 (2). We observe statistically significant autoregressive coefficients for the DNS at the 1% level. Regime 1 coefficients for the DNS factors are very similar to those found in the simple VAR model in Table 3. However, in regime 2, we observe a much lower level factor, which determines yield level dynamics. Thus, regime 2 exhibits lower serial persistency with respect to the yield level. Consistent with this finding, Figure 3 reveals that regime 2 is instigated, *inter alia*, when there are abrupt changes in the yield levels as during the Great Financial Crisis or as during the onset of the COVID-19 pandemic in March 2020.[8] Since we are estimating a MSVAR model not only for the DNS factors, the shape factors of the yield curve, but also for sentiment variables, changes to the sentiment variables can also induce regime shifts. The tendency towards regime 2 during the year 2021 can be explained by lower levels of inflation sentiment. The interest rate sentiment and unemployment sentiment do not show any noticeable regime-shifting dynamics during that period, as shown in Figure 1. However, inflation sentiment reverts to the mean in 2022, while regime 2 still dominates during 2022, which defies interpretation.

Furthermore, we find that sentiment variables have a larger impact on the DNS factors and thus the shape of the yield curve in regime 2, as depicted in Table 5. For example, in regime 2, a one standard deviation increase in lagged interest rate sentiment statistically significantly decreases the level and increases the slope and curvature factor by -0.177, 0.094, 0.633, respectively. However, no impact of lagged interest rate sentiment on the yield curve factors was found in regime 1. Similarly, a one standard deviation increase in lagged inflation sentiment increases the level and decreases the slope and curvature factor by 0.270, -0.242, -0.442, respectively. In regime 1, lagged inflation sentiment has no statistically significant impact on the DNS factors. Finally, lagged unemployment sentiment statistically significantly influences level and slope dynamics in regime 1. While the coefficients are statistically significant, they are comparably small in economic terms. However, in regime two, lagged unemployment sentiment impacts the curvature factor statistically and economically significantly. A one standard deviation increase in lagged unemployment sentiment increases the curvature factor by 0.233. Consequently, we argue

[8] We use the Fed Funds Rate for illustrative purposes as it impacts Treasury yields at all maturities.

that sentiment variables matter, especially in regime 2, which characterizes times of fast changing yield or inflation dynamics.

Moreover, all statistically significant coefficients in regime 1 describing the lagged factor-to-sentiment and cross-sentiment dynamics have the same sign as the corresponding coefficients from the simple VAR model depicted Table 3. Interestingly, in regime 2, we observe numerous deviations in terms of the direction of the effect compared to regime 1 for those coefficients. For example, while in regime 1 lagged unemployment sentiment negatively and statistically significantly correlates with current interest and inflation sentiment, the impact is positive and statistically significant in regime 2. In other words, not only sentiment-to-factor dynamics but also cross-sentiment dynamics show regime dependence.

Finally, looking at the transition matrix depicted in Table 4, we see that regime 1 is more persistent compared to regime 2. The probability of staying in regime 1 is 98.1%, while the probability of staying in regime 2 is 89.3%. Consistently, we observe 115 regime 1 and only 57 regime 2 observations with an average duration of 38.3 months for regime 1 and 19 months for regime 2.

Panel A	*Regime 1: Coefficients*					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_{t-1}	0.905^{***}	0.085^{*}	0.052	0.15	−0.172	0.279^{***}
	(0.043)	(0.048)	(0.119)	(0.154)	(0.154)	(0.102)
S_{t-1}	−0.025	0.972^{***}	0.058	-0.226^{**}	−0.17	0.075
	(0.031)	(0.035)	(0.094)	(0.112)	(0.122)	(0.085)
C_{t-1}	−0.014	0.047^{***}	0.945^{***}	0.123^{***}	0.115^{**}	0.086^{***}
	(0.016)	(0.014)	(0.044)	(0.044)	(0.05)	(0.031)
$S_{i,t-1}$	0.032	−0.026	−0.111	0.247^{***}	−0.07	-0.126^{**}
	(0.035)	(0.028)	(0.087)	(0.084)	(0.102)	(0.057)
$S_{\pi,t-1}$	0.025	−0.008	−0.001	0.106	0.106	-0.149^{**}
	(0.023)	(0.026)	(0.055)	(0.091)	(0.088)	(0.073)
$S_{u,t-1}$	0.049^{*}	-0.066^{**}	0.007	-0.284^{**}	-0.284^{**}	0.585^{***}
	(0.029)	(0.03)	(0.084)	(0.107)	(0.128)	(0.089)
μ	0.189^{*}	−0.132	−0.26	0.702^{**}	0.702^{**}	-0.423^{*}
	(0.095)	(0.101)	(0.277)	(0.302)	(0.318)	(0.213)

	Regime 1: Error Correlation Matrix					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_t	-	-	-	-	-	-
S_t	-0.81^{***}	-	-	-	-	-
C_t	-0.31^{***}	-0.01	-	-	-	-
$S_{i,t}$	-0.09	0.14	0.04	-	-	-
$S_{\pi,t}$	-0.08	0.07	0.08	0.12	-	-
$S_{u,t}$	0.11	-0.21^{***}	-0.02	-0.21^{***}	-0.14	-

	Estimated Transition Matrix	
	Regime 1	*Regime 2*
Regime 1	0.981	0.019
Regime 2	0.107	0.893

Table 4: Regression Output of MSVAR(1,2) Sentiment Model: Panel A
The upper table shows the coefficients of the MSVAR(1,2) model for the yield curve factors and sentiment from January 2008 to May 2022. The factors are estimated under the Dynamic Nelson-Siegel model. HC3 standard errors are shown in parentheses. The middle table reports the residual correlation matrix. Inference is made under the assumption of white-noise residuals. ***, **, and * denote significance at the 1%, 5% and 10% level, respectively. The lowest table reports the regime transition matrix.

Panel B	*Regime 2: Coefficients*					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_{t-1}	0.643***	0.083	−0.219	0.073	0.846***	−0.433*
	(0.095)	(0.064)	(0.192)	(0.204)	(0.272)	(0.252)
S_{t-1}	−0.489***	0.962***	0.478***	−0.311*	1.047***	−0.653***
	(0.09)	(0.049)	(0.165)	(0.174)	(0.241)	(0.194)
C_{t-1}	0.047*	0.058***	0.909***	−0.063	0.13*	−0.174**
	(0.025)	(0.019)	(0.053)	(0.06)	(0.071)	(0.083)
$S_{i,t-1}$	−0.177***	0.094**	0.633***	−0.136	0.396**	−0.271*
	(0.044)	(0.04)	(0.128)	(0.112)	(0.167)	(0.143)
$S_{\pi,t-1}$	0.27***	−0.242***	−0.442***	−0.156	−0.156	0.513***
	(0.056)	(0.043)	(0.092)	(0.168)	(0.162)	(0.193)
$S_{u,t-1}$	0.074	−0.037	0.233*	0.594***	0.594***	−0.106
	(0.046)	(0.041)	(0.139)	(0.127)	(0.205)	(0.173)
μ	0.032	−0.3**	1.778***	−0.33	−0.33	0.079
	(0.145)	(0.127)	(0.395)	(0.36)	(0.482)	(0.562)

	Regime 2: Error Correlation Matrix					
	L_t	S_t	C_t	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$
L_t	-	-	-	-	-	-
S_t	-0.76***	-	-	-	-	-
C_t	-0.35***	0.13	-	-	-	-
$S_{i,t}$	-0.36***	0.29***	0.13	-	-	-
$S_{\pi,t}$	-0.35***	0.12	0.13	0.20***	-	-
$S_{u,t}$	0.18**	-0.25***	-0.06	-0.21***	-0.11	-

	Regime Duration Information					
		Regime 1		*Regime 2*		
# Observations		115		57		
Duration Average		38.3		19		
Observations	172	172	172	172	172	172
Adjusted R^2	0.946	0.956	0.928	0.261	0.168	0.653

Table 5: Regression Output of MSVAR Sentiment Model: Panel A
The upper table shows the coefficients of the MSVAR(2,1) model for the yield curve factors and sentiment from January 2008 to May 2022. The factors are estimated under the Dynamic Nelson-Siegel model. HC3 standard errors are shown in parentheses. The middle table reports the residual correlation matrix. Inference is made under the assumption of white-noise residuals. ***, **, and * denote significance at the 1%, 5% and 10% level, respectively. The lowest table depicts regime duration information. Regime prevalence is assumed if its filtered probability is higher than 50%. The adjusted R squared describes the model fit for the respective dependent variable over the entire sample, using regime 1 and regime 2.

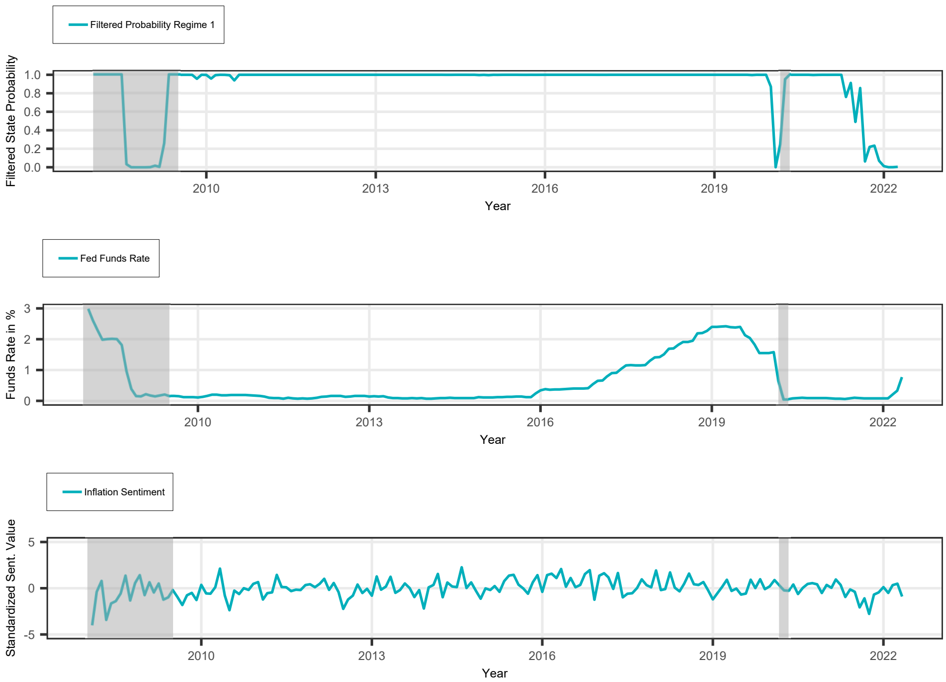

Figure 3: Filtered State Probabilities Regime 1

This figure shows the filtered state probabilities of regime 1. They cover the period from January 2008 to May 2022. In dark gray shading are the NBER recession periods.

6.2 Discussion trading results

The trading results from betting on changes of the slope factor (curvature factor) are depicted in Table 6 (7). The results are split into different lengths of trading periods. Panel A describes the results for the entire trading period from January 2016 to May 2022, while Panel B describes the time period before the outbreak of the Corona pandemic ending in February 2020, and Panel C includes the trading results for the Corona time period starting in March 2020. There are 50 trading months before the Corona pandemic and 27 trading months during Corona. In each panel, we show the top six and bottom six models out of 30 models with respect to cumulative return, separated by the dashed line in each panel.[9] The complete list of models is depicted in Appendix A.3. In order to benchmark the cumulative returns of our trading strategy, we build an equally weighted, long only benchmark portfolio of 2-year-, 5-year-, and 10-year U.S. Treasury notes. The performance is given in Table 8.

The results from betting on changes in the slope are depicted in Table 6. A total of

[9]The numerical results of the other models are available from the authors upon request.

9 models beat the benchmark in terms of cumulative return before Corona, while all butterfly strategies beat the benchmark during Corona. The gains from the butterfly strategy during the Corona period lift the results for the entire trading period, such that all 30 models beat the benchmark portfolio concerning the entire trading period. The top six models all have Sharpe ratios higher than 1 in all trading periods shown in panels A, B, C. Furthermore, the p-values of the T-tests of the bottom six models indicate that at the 5% level, we can only statistically significantly differentiate two (six, two) models from the best performing model in terms of their mean return during the respective trading period depicted in Panel A (B, C). Similarly, the bootstrapped p-values of the Sharpe ratios of the bottom six models show that we can statistically significantly differentiate five (four, one) from the best performing model in terms of the Sharpe ratio, as illustrated in Panel A (B, C). With respect to minimum cumulative return, there are some models that never lost money over the entire trading period, which does imply that those models are the best in terms of cumulative return. In terms of maximum drawdown, the models do not show any notable differences for all three trading periods. The answer to the question of whether macro sentiment data-augmented models outperform other models in terms of realized cumulative return is no. While there are five (three) sentiment-augmented models within the top six in Panel A and B (C), even the best performing model in each category does not statistically significantly outperform the raw models, in which we only use the DNS factors to predict changes in the slope of the yield curve. In Table A.4, we can see that the model *MSVAR DNS Only (VAR DNS Only)* ranks on place 8 (18), 6 (22), 25 (4) for the entire, before Corona and during Corona periods, respectively, with T-test p-values never below 0.06. Furthermore, we observe that the models are unstable in that well performing models before Corona had much worse results during Corona and vice versa. The benchmark models augmented with monetary policy indicators do not perform significantly better or worse compared to the sentiment-augmented models. To summarize, for slope bets, we do not find convincing evidence that sentiment-augmented models yield better results compared to non-sentiment-augmented models. However, we find many models that yield butterfly trading results that perform consistently better than the benchmark.

The results from the bets on the curvature factor are shown in Table 7. For the entire trading period, the top 9 models beat the benchmark in terms of cumulative return. Before Corona, only two models are better than the benchmark and during Corona 15 models beat the benchmark. Overall, the returns from the curvature bets are much more volatile, which results in significantly lower Sharpe ratios in all trading periods compared to the slope bets in Table 6. For all models, the Sharpe ratios are lower than 1, except for the model *MSVAR Percentage* during Corona. Again, from a statistical point of view, sentiment-augmented models do not perform significantly better than the raw DNS factor models. In Table A.4 we show that only during Corona does the best performing model, *MSVAR Percentage*, outperform the raw models at the five percent level. However, this model performs very badly for the 50 trading months before Corona, ranking at place 27. No sentiment-augmented model outperforms the raw models from a statistical point of view. However, in terms of model rank consistency, we find that the model *MSVAR Inflation Sentiment (Ravenpack)* performs consistently very well. For the trading period before Corona, the model ranks at place 5, and during Corona at place 6. For the entire trading period, the model is in third place. While we cannot find statistical evidence that the model performs better than the raw models, economically it delivers significantly better results, beating the benchmark portfolio especially during Corona but not before Corona. Interestingly, for every trading period, there are at least two inflation sentiment-augmented models within the top six models. This delivers some evidence that inflation sentiment can be helpful in predicting changes in the curvature of the yield curve, which can be translated into real economic returns. In particular, the cumulative returns are high for the time period during Corona. In fact, inflation has been sharply rising since the onset of the Corona pandemic and the yield curve has experienced decreasing levels of curvature. This is consistent with the highly negative and statistically and economically significant coefficient describing the relationship between lagged inflation sentiment and the curvature factor in regime 2, which dominates during the Corona trading period.

	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.B.C.	R.D.C.
Panel A: Entire Out-of-Sample Period										
VAR Michigan Cons. Sentiment	10.84	0.13	1	1.74	1	1.04	-1.04	10.99	4	10
VAR All Sent. (Ravenpack)	10.7	0.13	0.97	1.72	0.95	0.9	-0.9	10.85	3	13
MSVAR Michigan Cons. Sentiment	10.03	0.12	0.83	1.61	0.83	0.74	0	10.51	1	19
MSVAR Unemp. Sent. (Twitter)	9.84	0.12	0.79	1.58	0.74	0.57	0	10.34	2	21
VAR Unemp. Sent. (Ravenpack)	9.28	0.12	0.67	1.48	0.25	0.9	-0.9	9.43	7	15
VAR Fundsrate	8.34	0.1	0.5	1.33	0.31	1.17	0	8.49	9	12
MSVAR All Sent. (Twitter)	4.38	0.06	0.08	0.69	0.1	2.25	-0.73	6.57	11	30
VAR Percentage	4.34	0.06	0.07	0.71	0.02	1.04	-1.04	4.48	30	2
MSVAR Maturity	4.24	0.05	0.07	0.67	0.04	1.88	-1.88	5.04	21	23
VAR Inf. Sent. (Ravenpack)	4.07	0.05	0.06	0.66	0	1.21	-1.04	4.22	29	14
MSVAR Log Assets	3.7	0.05	0.05	0.59	0.03	2.51	-2.51	3.99	28	18
VAR Inf. Sent. (Twitter)	3.22	0.04	0.04	0.52	0	1.15	-1.04	3.36	27	24
Panel B: Before Corona										
	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.A.P.	R.D.C.
MSVAR Michigan Cons. Sentiment	7.61	0.15	1	1.92	1	0.65	0	7.89	3	19
MSVAR Unemp. Sent. (Twitter)	7.59	0.15	0.99	1.92	1	0.57	0	7.87	4	21
VAR All Sent. (Ravenpack)	7.28	0.14	0.91	1.8	0.87	0.9	-0.9	7.55	2	13
VAR Michigan Cons. Sentiment	7.08	0.14	0.85	1.72	0.8	1.04	-1.04	7.35	1	10
MSVAR Inf. Sent. (Twitter)	6.98	0.14	0.83	1.72	0.79	0.73	-0.73	7.25	9	27
MSVAR DNS Only	6.39	0.12	0.67	1.58	0.54	0.57	0	6.67	8	25
VAR Log Divisia	2.04	0.04	0.05	0.5	0.05	1.07	-1.04	2.3	16	1
VAR Int. Sent. (Ravenpack)	1.97	0.04	0.05	0.5	0.09	0.92	0	2.23	21	11
VAR Inf. Sent. (Twitter)	1.29	0.03	0.03	0.33	0.02	1.15	-1.04	1.65	30	24
MSVAR Log Assets	1.08	0.02	0.02	0.27	0.06	2.51	-2.51	1.34	29	18
VAR Inf. Sent. (Ravenpack)	0.86	0.02	0.02	0.22	0.02	1.21	-1.04	1.12	28	14
VAR Percentage	0.49	0.01	0.01	0.13	0.02	1.04	-1.04	0.93	26	2
Panel C: During Corona										
	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.A.P.	R.B.C.
VAR Log Divisia	3.87	0.14	1	1.96	1	0.35	0	4.01	16	25
VAR Percentage	3.83	0.14	0.98	1.94	0.98	0.31	0	3.97	26	30
MSVAR Unemp. Sent. (Ravenpack)	3.53	0.13	0.86	1.73	0.5	0.38	0	3.53	11	17
VAR DNS Only	3.51	0.13	0.85	1.74	0.31	0.38	0	3.65	18	22
VAR All Sent. (Twitter)	3.51	0.13	0.85	1.74	0.31	0.38	0	3.65	7	13
VAR Int. Sent. (Twitter)	3.51	0.13	0.85	1.74	0.31	0.38	0	3.65	17	20
MSVAR DNS Only	1.78	0.07	0.29	0.83	0.1	0.78	0	2.56	8	6
MSVAR Inf. Sent. (Ravenpack)	1.07	0.04	0.16	0.5	0.1	0.93	-0.05	1.94	24	18
MSVAR Inf. Sent. (Twitter)	0.98	0.04	0.15	0.46	0.09	0.86	0	1.84	9	5
MSVAR Int. Sent. (Ravenpack)	0.94	0.04	0.15	0.44	0.19	1.09	-0.67	1.72	22	14
MSVAR Percentage	-0.34	-0.01	0.04	-0.15	0.1	1.05	-0.62	0.43	23	8
MSVAR All Sent. (Twitter)	-0.36	-0.01	0.04	-0.16	0.03	2.14	-0.41	1.73	25	11

Table 6: Trading Results Slope Bets With Transaction Costs

This table shows the top six and bottom six models in terms of realized cumulative return over the respective trading period. The top and bottom models are separated with the dashed line in each trading period. Panel A depicts the results covering the period from January 2016- May 2022, Panel B shows the results for the trading period between January 2016 and February 2020, while Panel C includes the results for the period between March 2020 and May 2022. The column labelled T-Test p-value describes the p-value of the t-test of the mean return of the best performing model and the respective row model. The column SR p-value describes the p-value testing equality of the Sharpe ratio of the best performing model and the respective row model (Ledoit & Wolf, 2008). The column maximum drawdown describes the maximum decline in percentage points from a historical peak in terms of cumulative return. The minimum (maximum) cumulative return describes the lowest (highest) cumulative return achieved over the entire trading period. The last two columns describe the rank of the respective model in different trading periods. R.B.C. stands for rank before Corona, and R.D.C. refers to rank during corona, while R.A.P. is the abbreviation for rank in all periods.

	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.B.C.	R.D.C.
Panel A: Entire Out-of-Sample Period										
MSVAR Percentage	6.93	0.09	1	0.29	1	7.43	-7.43	8.67	27	1
VAR Percentage	6.75	0.09	0.99	0.29	0.99	9.1	-5.33	8.49	12	2
MSVAR Inf. Sent. (Ravenpack)	4.97	0.07	0.89	0.22	0.87	13.3	-3.94	9.36	5	6
MSVAR Inf. Sent. (Twitter)	2.65	0.04	0.76	0.13	0.72	15.06	-9.91	5.15	20	3
VAR Unemp. Sent. (Ravenpack)	0.1	0.01	0.63	0.02	0.6	8.96	-3.9	5.06	16	5
VAR Michigan Cons. Sentiment	-0.07	0	0.62	0.02	0.6	11.8	-0.25	11.73	10	8
VAR Int. Sent. (Ravenpack)	-15.2	-0.21	0.09	-0.67	0.15	15.2	-15.2	0	25	17
MSVAR All Sent. (Ravenpack)	-15.64	-0.21	0.08	-0.69	0.18	16.86	-15.64	1.22	13	28
MSVAR Log Divisia	-15.82	-0.22	0.08	-0.71	0.03	18.35	-18.35	0	30	9
MSVAR Int. Sent. (Ravenpack)	-17.49	-0.24	0.05	-0.79	0.1	18.79	-18.79	0	26	23
VAR All Sent. (Ravenpack)	-19.45	-0.27	0.03	-0.91	0.11	21.95	-19.45	2.5	22	29
VAR Maturity	-22.72	-0.33	0.02	-1.1	0.07	25.61	-22.72	2.89	21	30

Panel B: Before Corona	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.A.P.	R.D.C.
VAR Inf. Sent. (Ravenpack)	9.29	0.18	1	0.76	1	2.91	0	11.43	9	22
MSVAR Unemp. Sent. (Twitter)	5.53	0.11	0.68	0.46	0.39	4.28	-0.13	8.15	12	19
VAR Log Divisia	5.15	0.1	0.64	0.43	0.36	4.89	-0.64	9.31	10	16
MSVAR All Sent. (Twitter)	3.94	0.08	0.55	0.33	0.5	4.46	0	6.5	8	11
MSVAR Inf. Sent. (Ravenpack)	3.91	0.08	0.55	0.33	0.44	5.52	0	9.36	3	6
VAR Int. Sent. (Twitter)	3.38	0.07	0.51	0.29	0.24	4.81	-0.47	7.46	13	18
VAR Int. Sent. (Ravenpack)	-5.25	-0.1	0.09	-0.44	0.13	6.56	-6.56	0	25	17
MSVAR Int. Sent. (Ravenpack)	-5.74	-0.11	0.08	-0.48	0.16	5.8	-5.8	0	28	23
MSVAR Percentage	-5.93	-0.12	0.07	-0.5	0.11	7.43	-7.43	0	1	1
MSVAR Unemp. Sent. (Ravenpack)	-8.3	-0.17	0.03	-0.73	0.08	11.32	-11.32	0	15	7
MSVAR Int. Sent. (Twitter)	-8.43	-0.17	0.03	-0.74	0.05	9.49	-9.49	0	20	12
MSVAR Log Divisia	-12.34	-0.26	0.01	-1.16	0.03	12.95	-12.95	0	27	9

Panel C: During Corona	Cumulative Return (in %)	Average Monthly Return (in %)	T-Test p-value	Sharpe Ratio	SR p-value	Max. Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)	R.A.P.	R.B.C.
MSVAR Percentage	13.67	0.49	1	1.2	1	2.17	-1.55	15.53	1	27
VAR Percentage	7.42	0.28	0.59	0.66	0.42	4.78	-4.74	9.17	2	12
MSVAR Inf. Sent. (Twitter)	5.79	0.22	0.5	0.52	0.35	7.16	-7.16	7.51	4	20
MSVAR Michigan Cons. Sentiment	3.91	0.15	0.4	0.36	0.23	5.01	-1.58	5.73	7	24
VAR Unemp. Sent. (Ravenpack)	1.72	0.07	0.3	0.17	0.27	5.6	-2.34	6.69	5	16
MSVAR Inf. Sent. (Ravenpack)	1.01	0.05	0.27	0.11	0.13	7.59	-7.56	2.66	3	5
VAR Unemp. Sent. (Twitter)	-13.05	-0.51	0.01	-1.27	0.09	15.5	-13.05	2.46	23	18
VAR All Sent. (Twitter)	-13.29	-0.52	0.01	-1.31	0.09	14.63	-13.29	1.34	24	17
VAR Inf. Sent. (Twitter)	-14.52	-0.57	0.01	-1.46	0.05	15.99	-15.95	0.03	21	8
MSVAR All Sent. (Ravenpack)	-15.05	-0.59	0.01	-1.53	0.07	15.82	-15.05	0.76	26	13
VAR All Sent. (Ravenpack)	-15.9	-0.63	0	-1.65	0.07	15.9	-15.9	0	29	22
VAR Maturity	-20.11	-0.82	0	-2.35	0.03	20.14	-20.11	0.03	30	21

Table 7: Trading Results Curvature Bets With Transaction Costs

This table shows the top six and bottom six models in terms of realized cumulative return over the respective trading period. The top and bottom models are separated with the dashed line in each trading period. Panel A depicts the results covering the period from January 2016- May 2022, Panel B shows the results for the trading period between January 2016 and February 2020, while Panel C includes the results for the period between March 2020 and May 2022. The column labelled T-Test p-value describes the p-value of the t-test of the mean return of the best performing model and the respective row model. The column SR p-value describes the p-value testing equality of the Sharpe ratio of the best performing model and the respective row model (Ledoit & Wolf, 2008). The column maximum drawdown describes the maximum decline in percentage points from a historical peak in terms of cumulative return. The minimum (maximum) cumulative return describes the lowest (highest) cumulative return achieved over the entire trading period. The last two columns describe the rank of the respective model in different trading periods. R.B.C. stands for rank before Corona, and R.D.C. refers to rank during corona, while R.A.P. is the abbreviation for rank in all periods.

	Benchmark Portfolio					
	Cumulative Return (in %)	Average Monthly Return (in %)	Sharpe Ratio	Max Drawdown (in %)	Min. Cum. Return (in %)	Max. Cum. Return (in %)
Entire Out-of-Sample Period	-4.88	-0.06	-0.451	14.44	-5.63	9.21
Before Corona	5.37	0.11	-0.05	10.04	-5.63	5.37
During Corona	-9.73	-0.37	-1.16	13.7	-10.06	3.64

Table 8: Benchmark Statistics
The table shows the performance of our benchmark portfolio. The full sample ranges from Janaury 2016 to May 2022. The period before Corona dates from January 2016 until February 2020, while the Corona period starts in March 2020 and ends in May 2022. The benchmark portfolio consists of equally weighted U.S. 2-year-, 5-year-, and 10-year Treasury notes.

7 Conclusion

While there exists substantial evidence that sentiment-augmented models predict the yield curve better than non-sentiment-augmented models, we fail to construct trading strategies based upon sentiment data that perform consistently and statistically significantly better than their non-sentiment-augmented counterparts. However, we find mild evidence that inflation sentiment-augmented models perform consistently economically better when betting on changes in the curvature of the yield curve compared to other sentiment-augmented and non-sentiment-augmented models. In particular, the Markov-switching VAR model augmented with inflation sentiment data from RavenPack performs consistently well in all trading periods before as well as during the Corona pandemic. Overall, however, the economic returns from betting on changes in the slope of the yield curve, as opposed to the curvature, are better. Considering the entire trading period, slope bets based on all models applied achieve cumulative returns that are higher than the long-only Treasury note benchmark portfolio. This is due to the better economic returns of the butterfly trading strategy for all underlying models during the Corona pandemic compared to the benchmark. Considering the economic returns of betting on changes in the curvature of the yield curve, the returns are much more volatile, resulting in lower Sharpe ratios and higher drawdowns. Over the entire trading period, only around a third of the models beat the long-only benchmark portfolio. Our models are all Nelson-Siegel type models, which differ from arbitrage-free term structure models. Testing whether trading strategies based upon sentiment-augmented, arbitrage-free term structure models yield different results presents an interesting future research opportunity.

References

Audrino, F. (2006). Tree-structured multiple regimes in interest rates. *Journal of Business & Economic Statistics*, *24* (3), 338–353.

Audrino, F., & Medeiros, M. C. (2010). Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, *26* (6), 999–1022.

Audrino, F., & Offner, E. (2022). The impact of macroeconomic news sentiment on interest rates. *SSRN*, 72. (Available at SSRN: https://ssrn.com/abstract=4170954)

Ballinari, D., & Behrendt, S. (2021). How to gauge investor behavior? a comparison of online investor sentiment measures. *Digital Finance*, *3* (2), 169–204.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130* (2), 337–364.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, *131* (1-2), 309–338.

Duffee, G. R. (2011). Information in (and not in) the term structure. *Review of Financial Studies*, *24* (9), 2895–2934.

Edison, H. J. (1997). The reaction of exchange rates and interest rates to news releases. *International Journal of Finance & Economics*, *2* (2), 87–100.

Erlwein-Sayer, C. (2017). *Forecasting sovereign bond spreads with macroeconomic news sentiment* (Tech. Rep.). OptiRisk Systems.

Erlwein-Sayer, C. (2018). Macroeconomic news sentiment: Enhanced risk assessment for sovereign bonds. *Risks*, *6* (4), 141.

Goldberg, L., & Grisse, C. (2013). *Time variation in asset price responses to macro announcements* (Tech. Rep.).

Goldberg, L., & Leonard, D. (2005). What moves sovereign bond markets? the effects of economic news on U.S. and german yields. *Federal Reserve Bank of New York*, 1-7. (Available at SSRN: https://ssrn.com/abstract=683269)

Gotthelf, N., & Uhl, M. W. (2018). News sentiment: A new yield curve factor. *Journal of Behavioral Finance*, *20* (1), 31–41.

Guidolin, M., & Pedio, M. (2019). Forecasting and trading monetary policy effects on the riskless yield curve with regime switching nelson–siegel models. *Journal of*

Economic Dynamics and Control, *107*, 103723.

Guidolin, M., & Timmermann, A. (2009). Forecasts of us short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, *150*(2), 297–311.

Hamilton. (1994). *Time series analysis.* Princeton: Princeton University Press.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65.

Lucca, D., & Trebbi, F. (2009). *Measuring central bank communication: An automated approach with application to FOMC statements* (Tech. Rep.). doi: 10.3386/w15367

Nelson, C., & Siegel, A. (1985). *Parsimoneous modeling of yield curves for U.S. treasury bills* (Tech. Rep.).

Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. stock market. *Journal of Banking & Finance*, *84*, 25–40.

Renault, T. (2019). Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. *Digital Finance*, *2*(1-2), 1–13.

Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, *101*(4), 1514–1534.

Xiang, J., & Zhu, X. (2013). A regime-switching nelson-siegel term structure model and interest rate forecasts. *Journal of Financial Econometrics*, *11*(3), 522–555.

Yang, S., Liu, Z., & Wang, X. (2020). News sentiment, credit spreads, and information asymmetry. *The North American Journal of Economics and Finance*, *52*, 1–19.

Yu, W.-C., & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields using dynamic nelson-siegel models. *International Journal of Forecasting*, *27*(2), 579–591.

Appendix

A.1 Attributes and Summary Statistics for Sentiment Variables

	Interest Rates	**Inflation**	**Unemployment Rate**
Keywords	interest-rate, treasury-bill (-note, -bond)	inflation, deflation	unemployment, employment
	sovereign-debt, government-budget	consumer-price-index	jobless-claims, non-farm-payrolls
Number of News	22,548	7,821	8,281
Mean	49.97	43.94	65.52
Std. Deviation	22.58	22.50	25.48
Skewness	-0.41	0.13	-0.67
Kurtosis	-0.38	-0.41	-0.51
ACF(1)	0.71	0.44	0.46
ACF(2)	0.55	0.20	0.35

Table A.1: Attributes and Summary Statistics for Sentiment Variables of RavenPack

The table shows the attributes and summary statistics for the selected sentiment variables from January 2008 - May 2022. Each column represents the sentiment on the underlying macroeconomic variables, i.e. interest rates, inflation rate, and unemployment rate. The first row shows the keywords used to extract the sentiment variables.

	Interest Rates	**Inflation**	**Unemployment Rate**
Keywords	interest rate, treasury bill	inflation	unemployment, labor market
Number of Tweets	1,019,957	1,710,748	1,632,518
Mean	-0.02	-0.02	-0.05
Std. Deviation	0.04	0.05	0.05
Skewness	-0.64	-0.94	-0.46
Kurtosis	1.48	3.45	1.03
ACF(1)	0.46	0.35	0.48
ACF(2)	0.43	0.33	0.46

Table A.2: Attributes and Summary Statistics for Sentiment Variables of Twitter

The table shows the attributes and summary statistics for the selected sentiment variables from January 2008 - May 2022. Each column represents the sentiment on the underlying macroeconomic variables, i.e. interest rates, inflation rate, and unemployment rate. The first row shows the keywords used to extract the sentiment variables. In order to guarantee that the extracted tweets are about the economy of the United States of America, all tweets had to contain at least one of the following keywords: US, USA, United States, FOMC, FED, America, Powell, Yellen, Bernanke or Greenspan.

A.2 RavenPack Sentiment vs. Macroeconomic Variables

Figure A.1: Sentiment vs. Macroeconomic Variables (RavenPack)

This figure shows different macroeconomic variables in blue against their sentiment counterpart in yellow. All time series are standardized. They cover the period from January 2008 to May 2022.

	$S_{i,t}$	$S_{\pi,t}$	$S_{u,t}$	$S_{i,t+1}$	$S_{\pi,t+1}$	$S_{u,t+1}$	$S_{i,t-1}$	$S_{\pi,t-1}$	$S_{u,t-1}$
Δi_t	**-0.42**	-0.2	0.05	**-0.23**	-0.05	**0.51**	**-0.35**	0.07	-0.09
π_t	**0.22**	**-0.16**	**0.13**	**0.22**	-0.13	0.09	**0.2**	**-0.21**	0.13
u_t	0.15	-0.01	-0.2	0.12	-0.07	-0.07	**0.22**	0.09	**-0.18**

Table A.3: Correlation Analysis Sentiment vs. Macroeconomic Variables
The table depicts the correlation between the RavenPack sentiment variables (contemporaneously, lead one period and lagged one period) and their macroeconomic counterparts. Δi_t, π_t and u_t represent the first difference in the short rate, inflation and unemployment growth, respectively. S_i, S_π, S_u denote the sentiment on each macroeconomic variable. Bold font indicates a p-value lower than 0.05, testing whether the true correlation coefficient is equal to zero. Due to the relatively small number of posts during the initial years of Twitter, the time period ranges from January 2011 to May 2022.

A.3 Complete List of Models

- VAR DNS Only
- VAR Log Divisia
- VAR Percentage
- VAR Maturity
- VAR Log Assets
- VAR Michigan Cons. Sentiment
- VAR Fundsrate
- VAR All Sent. (Twitter)
- VAR Inf. Sent. (Twitter)
- VAR Int. Sent. (Twitter)
- VAR Unemp. Sent. (Twitter)
- VAR Int. Sent. (Ravenpack)
- VAR All Sent. (Ravenpack)
- VAR Inf. Sent. (Ravenpack)
- VAR Unemp. Sent. (Ravenpack)

- MSVAR DNS Only
- MSVAR Log Divisia
- MSVAR Percentage
- MSVAR Maturity
- MSVAR Log Assets
- MSVAR Michigan Cons. Sentiment
- MSVAR Fundsrate
- MSVAR All Sent. (Twitter)
- MSVAR Inf. Sent. (Twitter)
- MSVAR Int. Sent. (Twitter)
- MSVAR Unemp. Sent. (Twitter)
- MSVAR All Sent. (Ravenpack)
- MSVAR Inf. Sent. (Ravenpack)
- MSVAR Int. Sent. (Ravenpack)
- MSVAR Unemp. Sent. (Ravenpack)

A.4 Performance of Raw Models

		Entire Trading Period		*Before Corona*		*During Corona*	
		VAR DNS Only	MSVAR DNS Only	VAR DNS Only	MSVAR DNS Only	VAR DNS Only	MSVAR DNS Only
Slope Bets	*Rank*	18	8	22	6	4	25
	T-Test p-value	0.15	0.49	0.06	0.67	0.85	0.29
Curvature Bets	*Rank*	18	14	7	11	24	14
	T-Test p-value	0.2	0.27	0.47	0.27	0.01	0.05

Table A.4: Raw Model Performance
The table shows the performance of the two raw models *VAR DNS Only* and *MSVAR DNS Only* in terms of their ranks and the p-value of the test comparing the mean of the returns and the best-performing model in the category.