No. 12-2016

Jochen Lüdering and Peter Tillmann

Monetary Policy on Twitter and its Effect on Asset Prices: Evidence from Computational Text Analysis

Monetary Policy on Twitter and its Effect on Asset Prices: Evidence from Computational Text Analysis[*]

Jochen Lüdering[†]

Justus-Liebig-University Gießen, Germany

Peter Tillmann[‡]

Justus-Liebig-University Gießen, Germany

March 1, 2016

Abstract

In this paper we dissect the public debate about the future course of monetary policy and trace the effects of selected topics of this discourse on U.S. asset prices. We focus on the "taper tantrum" episode in 2013, a period with large revisions in expectations about Fed policy. Based on a novel data set of 90,000 Twitter messages ("tweets") covering the entire debate of Fed tapering on Twitter we use Latent Dirichlet Allocation, a computational text analysis tool to quantify the content of the discussion. Several estimated topic frequencies are then included in a VAR model to estimate the effects of topic shocks on asset prices. We find that the discussion about Fed policy on social media contains price-relevant information. Shocks to shares of "tantrum"-, "QE"- and "data"-related topics are shown to lead to significant asset price changes. We also show that the effects are mostly due to changes in the term premium of yields consistent with the portfolio balance channel of unconventional monetary policy.

Keywords: Monetary Policy, Fed, Latent Dirichlet Allocation, Text Analysis, VAR

JEL classification: E32, E44, E52

1

1 Introduction

The formation of monetary policy expectations by market participants is at the core of the monetary policy transmission process. While there is a large body of research on how central banks communicate with financial markets (Blinder et al., 2008), there is little evidence about how, given the communication of the central bank, the discourse about monetary policy by market participants shapes market expectations. This lack of research is most likely due to the lack of data about individual views on future policy.

In this paper we study the changing policy expectations of market participants and the resulting change in forward looking financial variables such as asset prices. We focus on an episode in recent U.S. monetary policy which has been characterized by a major shift in market expectations: the "taper tantrum" period in 2013. After Fed chairman Ben Bernanke mentioned an eventually exit from the Fed's asset purchase programs in May 2013, markets changed their assessment of the future course of policy resulting in a phase of unusual volatility, an increase in long-term U.S. interest rates and an appreciation of the U.S. dollar. Markets quickly coined the term "tapering" to describe the Fed's exit from QE3. Given the exaggerated market reaction, this period is referred to as the "taper tantrum".

We analyze this episode based on a data set that contains the entire traffic on Twitter.com, the social media network, on Fed tapering. The data set consists of 90,000 text messages ("tweets") between April and October 2013 and reflects the debate among market professionals during the tantrum period. Twitter data offers several advantages over alternative data sets: First, in contrast to news articles or analyst reports, which are written and read be relatively few people, Twitter allows us to exploit the views of the crowd of financial professionals. Second, while it is unclear whether news reports are actually read, Twitter messages appear as push messages on mobile phones and are actively shared, discussed, endorsed or refuted. Hence, the tweets give more reliable evidence about individuals' views than the consumption of news reports. Third, the high frequency of observation allows us to trace the public debate in real time.

Since our aim is to model the changing beliefs about future monetary policy, we need to quantify the information content of the Twitter data. In a companion paper (Meinusch and Tillmann, 2016), we construct a dictionary with keywords describing a certain policy path. The drawback of this approach is that we have to specify a list of keywords in advance, which is likely to disregard important information from expressions not on our list of keywords. Furthermore, we can focus on two alternative policy paths only, an early or a later tapering decision, and have to leave

out other dimensions of the discussion.

In this paper we employ a different tool set taken from computational linguistics. Latent Dirichlet Allocation (LDA) (Blei et al. 2003) is used to extract latent topics out of the Twitter conversation on tapering. Examples of topics are "financial conditions", "economic data", "fear", "stimulus" and many more. The resulting topic frequencies, which express the likelihood that a given tweets contains this specific topic, are then included in a vector autoregression (VAR) together with a daily series of macroeconomic fundamentals and asset prices. We show that a shock to selected topics frequencies leads to a significant change in asset prices. This finding supports the notion that the public debate about future monetary policy, which is reflected by the discourse on Twitter.com, contains information that is relevant for pricing financial assets. A shock to the likelihood of the topic describing "market worries", for example, raises bond yields and leads to an appreciation of the dollar.

This paper is closely related to the recent literature on text mining and computational text analysis, respectively, for financial and monetary policy applications. This rapidly growing literature is summarized by Loughran and McDonald (2015).[1] For the purpose of this paper, the existing research applying models of latent information in textual data is particularly interesting.

Hansen and McMahon (2015) apply LDA modeling to the entire history of policy statements issued by the Federal Open Market Committee (FOMC) of the Federal Reserve. Thus, they are able to identify at which point in time the FOMC spent time discussing a specific topic. Selected topic frequencies are included in a Factor-Augmented VAR (FAVAR) model. The authors find that forward guidance related topics and topics reflecting the current economic situation affect real and financial variables.[2] While our modeling framework is similar, we do not study central bank communication but the discourse of the market about what the Fed is likely to do.[3] Hendry (2010, 2012) uses latent semantic analysis for the communication of the Bank of Canada. He identifies "themes" of communication, which are used to explain interest rate changes. Although they do not address a financial application but instead focus on fluctuations in the business cycle, the paper by Larsen and Thorsrud (2015) is also relevant for our work. They use a data set with Norwegian newspaper articles to construct an aggregate news index by employing topic modeling. The

---

[1]Bholat et al. (2015) provide a survey of text mining applications relevant for central banks.

[2]Using the same data set, Hansen et al. (2015) model the effect of increased transparency on the policy debate in FOMC meetings.

[3]Lucca and Trebbi (2009) and Schonhardt-Baily (2013) are other recent papers on textual analysis of FOMC communication.

news indices are related to economic activity within a Bayesian regression framework. We show that topics related to the "tantrum" notion of the tapering discussion, the implications of "QE" and the debate about incoming "data" affect bond yields and exchange rates. Stock prices seem to be less affected by these topic probabilities. We also decompose yields into the expectations component and the term premium. Based on this decomposition we show that the response of bond yields is mostly due to responses of term premia. This findings lends support to the balance sheet channel as the transmission channel of shifting tapering expectations.

This paper is organized as follows: In Section 2 we introduce the Twitter data set. Section 3 gives some background on computational text analysis and presents the LDA approach used in this paper. The estimation of a VAR model that includes asset prices and selected topics frequencies is described in Section 4. The results and some robustness checks are discussed in Section 5 and Section 6 draws conclusions.

2 The data set

The data set used in this study consists of all Twitter messages containing the words "Fed" and "taper" sent between April 15 and October 30, 2013. The data has been purchased from Gnip.com. Because it is highly likely that any tweet on the Fed's exit from QE contains both filter words, we are certain to have a comprehensive data set that reflects the entire tapering debate on Twitter. After deleting a few tweets in languages other than English, we are left with 87024 tweets. Re-tweets, i.e. Twitter messages forwarded by users, are left in the data set because a forwarded tweet is likely to be a relevant tweet and, as a result, the forwarding of a tweet also contains information.

For each tweet we know the content, the sender and the time the tweet was sent. We normalize the timing of each tweet to New York time. While trading hours end at 4 pm Eastern Time, twittering continues even after markets have closed. To account for tweets sent after markets closed, tweets sent after 4 pm are attributed to the next trading day. Likewise, weekends and holidays haven been excluded due to the lack of asset price data.

Figure 1: Number of Tweets
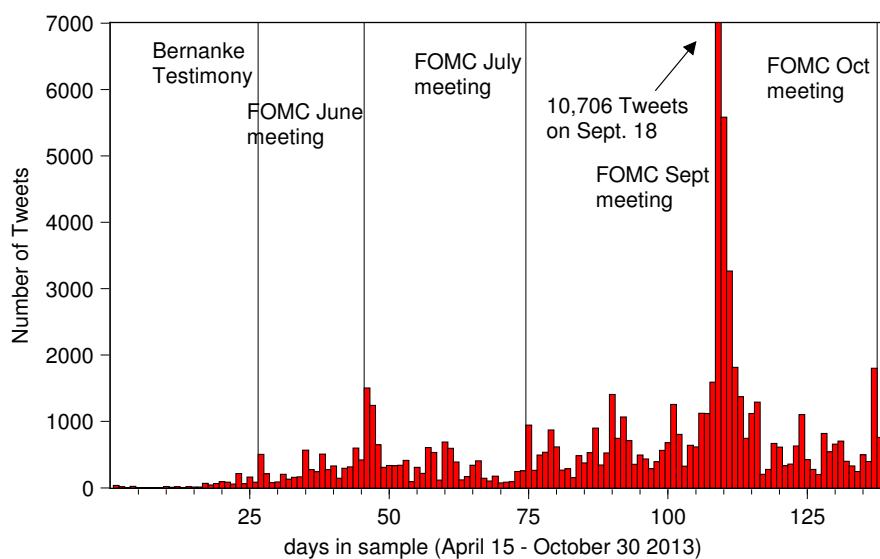
Notes: The vertical lines indicate the testimony of chairman Bernanke on May 22 and the subsequent FOMC meetings.

Figure (1) plots the daily number of tweets and, as vertical lines, the most important monetary policy events. Tweeting on the Fed's tapering decision gradually picks up before the testimony of chairman Bernanke on May 22. During this testimony, Bernanke mentioned the possibility of exiting from QE3, a statement that triggered the markets' subsequent tantrum reaction. A first peak is reached prior to the June meeting of the Federal Open Market Committee (FOMC), for which some market participants expected more detailed information about the pace of the tapering. After the FOMC postponed the tapering decision, the discourse among market participants intensified before each subsequent FOMC meeting. The peak was reached before the September FOMC meeting, for which the vast majority of Twitter users expected the decision about a reduction of monthly purchases of securities. However, market again misjudged the Fed as the FOMC again postponed the decision. The Fed eventually announced a reduction in its monthly asset purchases in the January 2014 FOMC meeting.

From Figure (1) we see that the number of tweets is systematically higher on FOMC meeting days. This is not surprising given the market's interest in monetary policy decisions. For our empirical analysis below this implies that we should control for FOMC meeting days and, in addition, for days on which FOMC minutes are published.

Since the monetary policy debate debate on Twitter captures the overall discussion among market participants well, we will now use the topic models to dissect the discussion into policy-relevant topics. We want to see which topic was most relevant on selected days and, in particular, how the information contained in these topics is reflected by asset prices.

The "taper tantrum" it subject to a relatively small empirical literature. Most studies, such as Eichengreen and Gupta (2013), Aizenman et al. (2014) and Mishra et al. (2014) ask whether the macroeconomic vulnerability of emerging market countries determines how strongly these countries were hit in 2013. However, these authors typically do not quantify market expectations and their revision directly. Rather, they argue that the changes in U.S. asset prices in 2013 are appropriate indicators of shifts in expectations. In Meinusch and Tillmann (2016) we use the same Twitter data set that is used in this paper. Based on a dictionary of words we build proxy variables for the beliefs of an early and a late tapering, respectively. Here we extend and broaden this line of research: we use several dimensions of the debate on Twitter, not just early or late tapering, and relate them to asset prices.

3 Applying topics models

In this section we apply topics models to dissect the discussion on Twitter regarding the unwinding of QE in its most important parts, which we then relate to asset pricing. The recent introduction of topic modeling, into the field of economics enables researchers to automatically classify texts and obtain underlying topics which constitute the document, given the assumed generative process and several predetermined parameters. It should be noted that these topics are not necessarily coherent topics in the semantic sense, but rather clusters of *terms* which repeatedly appear together over several documents. Similar to clustering methods, it is up to the researcher to make sense of the topics based on the words of which they are comprised.

For the application, we follow Grün and Hornik (2011) in using the R package tm[4] for pre-processing the data and subsequently *topicmodels*[5] for the fitting of the topic models.

3.1 Preliminary steps

Starting out with the corpus of twitter messages described in the previous section, the first step consists of cleaning the data to obtain the vocabulary V. The vocab-

---

[4]See https://cran.r-project.org/web/packages/tm/.
[5]See https://cran.r-project.org/web/packages/topicmodels/.

ulary is the set of different terms selected from the corpus, our entire set of tweets, because the terms are well suited to explain the content of individual documents. It is the goal to omit all terms that are not helpful in differentiating between topics.

As topic models are only concerned with the joint appearance of words in individual documents and are not influenced by grammar, we can remove all punctuation and redundant space characters from the tweets. A specificity of twitter messages is the appearance of hyperlinks and twitter usernames (i.e. *@username*), which we remove in their entirety. Further, all words are decapitalized and the stemming algorithm *SnowballC* is applied to create word stems. A stem is the part of a word which is common to a variety of grammatical forms. These two measures, stemming and decapitalization, lump together different grammatical forms and remove the differentiation due to a word being capitalized at the beginning of a sentence.

Words which are frequent but add little meaning to a document are called *stopwords* and are removed from the corpus and thus excluded from the vocabulary. We use the list of English language stopwords provided by the R package *tm*. In addition, the terms "Fed" and "taper" are removed from the corpus, as these words have been used to select the data set in the first place. Hence, they should be included in each tweet. Finally, all terms with a length between 4 and 20 characters are used to create the document-term matrix $\boldsymbol{F}$, which holds the frequencies $f_{i,j}$ of $|V| = 5082$ different terms in $D = 87024$ tweets and is the basis for the subsequent LDA estimation.

Figure 2: Word cloud of all tweets

Figure (2) visualizes the data set in a word cloud, which contains all words which appeared at least 100 times across all tweets, of course excluding the words "Fed" and "taper", which are included in any tweet as a result of the filtering process. As we do not see any meaningful word dominating the twitter content, we will resort to computational linguistics to disentangle different topics driving the discussion on Twitter.

In applications of topic models, it is common practice to further reduce the vocabulary by selecting only terms which are important to describe individual documents. This is usually done based on the *tf-idf* (term frequency–inverse document frequency) value (Blei and Lafferty 2009), which implies weighting the frequency in a single document against a the overall frequency. The particularities of twitter messages, e.g. the choice of words and the length of tweets limited to 140 characters, results in particularly high tf-idf values (median > 1).[6] As the importance of the individual words for explaining the different documents appears to be particularly high and the vocabulary is rather short (5082), in particular with respect to the

---

[6]Other data sets, using scientific articles, result in median tf-idf values of 0.004 (Lüdering and Winker 2016) and 0.1 (Grün and Hornik 2011).

number of documents in the corpus (87024), we do not remove any further terms.

3.2   Latent Dirichlet Allocation (LDA)

This section provides a brief overview of LDA models and the estimation method behind our analysis. For a more comprehensive introduction the reader should refer to Lüdering and Winker (2016). Before their recent arrival in economics, topic models have been used since the 1990s (Deerwester et al., 1990) to address issues in the area of information retrieval. Hofmann (1999) introduced probabilistic theory to topic models, providing a sound statistical background. His approach (probabilistic Latent Semantic Analysis) has later been extended to Latent Dirichlet Allocation (LDA) by Blei, Ng and Jordan (2003). Although LDA has subsequently been refined, e.g. time varying topics have been suggested by Wang and McCallum (2006) and the model has been extended to allow for topic correlation by Blei and Lafferty (2007), their underlying theoretical model remains the state of art in topic modeling up to today.

In LDA the creation of documents is described by an abstract generative process. It is assumed that all documents in a corpus are generated from a fixed set of K different topics. The topics consists of $|V|$ different terms w from a vocabulary V. Each term w in a document $\boldsymbol{w}$ is generated by first drawing a topic given a vector of topic probabilities $\theta_{\boldsymbol{w}}$ and afterwards drawing a term, given its probability β_k in a topic.

In order to estimate the matrices of predicted probabilities, $\hat{\theta} = K \times D$ and $\hat{\beta} = K \times |V|$, our variables of interest, an algorithm is used to reverse the generative process. Due to the complexity of the model, standard maximum likelihood procedures proof to be not suitable. Hence, a number of sophisticated methods has been developed to estimate the model nonetheless. In modern applications of LDA the original estimation algorithm (variational expectation maximization, VEM) has largely been replaced by Gibbs sampling, a Markov Chain Monte Carlo approach suggested by Griffith and Steyvers (2004). Instead of estimating the topic distribution θ and the term distribution β directly, the distribution of $\boldsymbol{z}_i$, the assignments of words to topics is estimated. In order to perform the computation, it is necessary to make the simplifying assumptions that θ and β are random draws from the Dirichlet distributions Dir(δ) and, respectively, Dir(δ). The parameters on the Dirichlet distributions are chosen according to the literature (Griffith and Steyvers, 2004) and set to $\alpha = 50/K$ and $\delta = 0.1$.

In applied work, the choice of the optimal number of topics K remains an important issue. In order to obtain an estimate for K, we apply the harmonic mean method as

suggested by Griffith and Steyvers (2004). It consists of taking a number of samples as estimates for $P(\boldsymbol{w}|K)$ from the Markov Chain, and the subsequent computation of the harmonic mean across the values. The resulting function of the relationship between K and $P(\boldsymbol{w}|K)$ is not smooth.

However, simple maximization does not necessarily lead to useful results.[7] However, we end up with an unreasonably large number of topics given our data. The obtained topics are very narrow and difficult to interpret. Hence, we follow the pragmatic approach by Hansen, McMahon and Prat (2014) of choosing a lower value in order to produce topics, which are more appealing to human judgment. By setting $K = 30$, we take into account that Twitter messages contain 140 characters as a maximum and the tweets in our sample have already been pre-selected to cover a specific area, the tapering decision of the Fed. Following Griffith and Steyvers (2004) the Markov Chain is constructed to converge to the "true" distribution of $\boldsymbol{z}_i$, which is the vector of assignments of words to topics. After 2000 iterations, from which the first 100 are discarded, the Markov Chain is assumed to have converged and approximations for document θ and topic β probabilities can be derived. The resulting matrix of term probabilities reveals the contents of the 30 different topics, which reflect the discussion on Twitter.

4   The empirical model

In this section we relate the discussion on Twitter as reflected in a selected number of identified topics on U.S. asset prices. For that purpose we use a VAR model in which we include not only information from the topics but also asset prices and macroeconomic conditions. Because we want to model a parsimonious VAR system, we cannot include all 30 topics in the VAR model jointly, which would leave too few degrees of freedom for estimation. Instead, we focus on selected topics only.

In particular, we identify three sets of topics, which are directed towards particularly important aspects of the assessment of future monetary policy. Each set includes two topics, $Topic_t^i$ and $Topic_t^j$. The following table summarizes the topics we use in the estimation.

The first topic set captures the notion of the market "tantrum" occurring in 2013. The second set reflects the debate about maintaining or unwinding the monetary stimulus. The third set includes the debate among Twitter users about data-based monetary policy decisions. We are agnostic with regard to the signs of the effects shocks to these topics have on U.S. asset prices. Figure (3) shows the frequency for

---

[7]Griffith and Steyvers (2004) circumvent this issue evaluating K at large steps.

Table 1: Selected sets of topics

	set	$Topic_t^i$	$Topic_t^j$
1	"tantrum"	T3	T13
		"market worries"	"fear"
2	"QE"	T11	T15
		"bond market"	"stimulus"
3	"data"	T17	T20
		"financial conditions"	"data"

each topics.

Figure (27) visualizes each of the selected topics by a word cloud, where the size of each word in the cloud reflects the significance (probability) of the word for a given topic. The importance of the topic over time can be assessed by the matrix of topic probabilities across tweets. In order to make meaningful interpretation possible we need to account for the fact that the number of tweets per day is not constant across the observation period. Thus, for each topic we calculate the mean topic probability over all tweets in a single day. These aggregate values can easily be compared over the course of the whole sample period.

In the empirical model each selected topic is scaled by the sum of the remaining topics. The reason for this is that all topics must add to one. An increase in one topic frequency must be associated with a decrease in the sum of all other topic frequencies. To gauge the relative importance of a topic, we have to normalize it by the importance of the remaining topics. Thus, each topic $Topic_t^n$ with $n \in (i,j)$ is included as

$$\overline{Topic}_t^n = \frac{Topic_t^n}{1 - Topic_t^n} \quad \text{for } n = i, j \text{ and } t = 1, ..., T. \tag{1}$$

We estimate a VAR model in order to derive the dynamic responses of asset price changes to the two summary indicators. The reduced-form representation of the VAR is

$$Y_t = A_0 + A(L)Y_t + D_t^{FOMC} + u_t \quad \text{for } E[u_t u_t'] = \Sigma_u, \tag{2}$$

where $A(L)$ reflects the matrix polynomial in the lag operator of order p, Y_t is a vector of endogenous variables and u_t constitutes a white noise process with variance-covariance matrix Σ_u. We also add a constant, A_0, to the model. To control for monetary policy events, we include separate dummies contained in D_t^{FOMC} for Bernanke's testimony on May 22, each FOMC meeting and each release date of FOMC minutes in our sample period. Most likely many tweets comment on FOMC

meeting outcomes or releases of FOMC minutes. We control for these days in order to see whether tweets contain information even in the inter-meeting or inter-release day period, respectively. In our discussion of the results below, we show impulse responses derived from VAR models with and without these FOMC-related dummies. The vector Y_t contains the following variables

$$Y_t = \begin{pmatrix} Y_t^{CESI} & \overline{Topic}_t^i & \overline{Topic}_t^j & A_t^{US} \end{pmatrix}'. \tag{3}$$

In this model, Y_t^{CESI} is the daily Citigroup Economic Surprise Index (CESI) for the U.S. economy. The CESI is defined as the weighted historical standard deviations of data surprises, that is, actual data releases minus the median survey expectations from the Bloomberg survey. Thus, the index captures positive and negative surprise realizations of macroeconomic data releases. It is important to control for data releases in our empirical model as market expectations reflected in Twitter messages also reflect macroeconomic news. The weights of economic indicators are derived from relative high-frequency impact on spot exchange rates.

Asset prices in the U.S. are reflected by A_t^{US}. We use four alternative asset prices: the 10-year and 5-year Treasury yields, the S&P 500 stock price index, and the nominal USD-EUR exchange rate. The latter two asset prices are taken from the St. Louis Fed's FRED database. The yield series are fitted yields from the Adrian et al. (2013) term structure model. Note that an increase in the USD-EUR exchange rate implies a depreciation of the dollar. The estimated VAR model includes four lags of the endogenous variables and is estimated on daily data for $T = 139$.

Estimating a VAR model in order to derive impulse response functions necessitates the identification of structural shocks from the estimated reduced form residuals. Here we impose a Cholesky identification on our VAR system which implies that on a given day each variable affects only those variables ordered behind it in the VAR model. The ordering of the variables corresponds to the order given in the description of the model, i.e. $Y_t^{CESI}, \overline{Topic}_t^i, \overline{Topic}_t^j, A_t^{US}$. This implies that a change in $\overline{Topic}_t^i$, for example, has an effect on asset prices on the same day, while a change in asset prices can affect the Twitter topics only on the following day. We believe it is important in our context to allow Twitter messages to have a contemporaneous effect on asset prices, while it is less convincing that Twitter users respond to, say, yield changes on the very same day. This is because responding to asset price change on a specific day requires continuous monitoring of intraday information about asset price changes. In the robustness section presented below, we will also present results based on alternative identification schemes.

5 Results

In the following subsections, we present the resulting impulse response functions describing the responses to topic shocks. In each impulse response graph, we show bootstrapped confidence bands reflecting the 16th and the 84th percentiles of the draws. We also show the responses from a model in which we include dummies for major FOMC events (red line) and a model without FOMC dummies (green, dotted line). We focus on the responses of asset prices, which are the focus of this paper.

5.1 Baseline results

Figures (4) to (7) report the responses of the three selected U.S. asset prices to an unexpected increase in the frequencies of the two "tantrum" topics. An increase in the importance of topic 3 ("market worries") leads to a strong increase in bond yields. A one standard deviation increase raises yields by two basis points. The results are very similar for 10-year and 5-year bond yields. An increase in the frequency of topic 13 ("fear"), however, lowers Treasury yields by two basis points. For the shock to topic 3 both impulse responses, the one in red based on the baseline model and the one in green coming from a model without FOMC dummies, exhibit virtually identical dynamics. For shocks to topic 13, however, controlling for FOMC events is particularly important as the green line indicated a weaker response to the shock. The US dollar appreciates following an increase in the frequency of topic 3 ("market worries") and depreciates after a shock to topic 13 ("fear") as shown in Figure (6). Stock prices, see Figure (7) appear insensitive to tantrum-related discussions.

These results show that a heightened discourse of the Fed's tapering decision, which leads to larger shares on topics 3 and 13 in the discussion, leads to high interest rate volatility. It is important to note that these estimated effects of the discussion on Twitter on asset prices do not stem from changes in the macroeconomic environment that could make a policy tightening or easing more likely. Since we control for the business cycle by including the CESI measure, the effects are driven by views of Twitter users alone. Likewise, we dummy out the days with FOMC meeting and releases of FOMC minutes. Thus, the results are not driven by tweets that simply reflect the information contained in official Fed communication on these policy days. The asset prices responses to "QE"-related topic frequencies are shown in Figures (8) to (11). An increase in the share of topic 11 ("bond market") or topic 15 ("stimulus") raises bond yields. Hence, the discussion about both topics reflects the market assessment that the Fed is concerned about the overheated bond market and

will likely reduce its monetary stimulus. The exchange rate response is consistent with this interpretation: the dollar appreciates following a surprise increase in topic 11 and and topic 15, respectively. Stock prices do not exhibit a significant response to both "QE"-topics, see Figure (11).

The responses to "data"-related topic shocks, see Figures (12) to (15), give rise to a consistent interpretation: as the Twitter discussion of topic 17 ("financial conditions") intensifies, Treasury yields increase, the dollar appreciates and stock prices fall. Thus, this topic reflects the public's understanding of the Fed's concern about overheating financial conditions. An increase in the likelihood of topic 20 ("economic data") depresses yields, leads to a depreciation of the dollar and to stock market gains. The more the Twitter discussion centers around the macroeconomic environment, the less likely an early tapering decision seems to be.

5.2   The contribution of shocks

A historical decomposition of the baseline model for 10-year bond yields shows the contribution of each topic to evolution bond yields over time. Figures (24) to (26). An increase in topic 3 ("market worries") explains a large fraction of the yield increase before the September 2013 FOMC meeting. Likewise, the steep increase in bond yield prior to the June FOMC meeting is driven by topic 13 ("fear"). Another interesting finding is that topic 17 ("financial conditions") is responsible for the drop in bond yields before and after the September 2013 FOMC meeting. A negative contribution means that the increase in topic 17 is smaller than expected, leading to a negative surprise component. Hence, compared to the tightening of market conditions in the run-up to the September FOMC meeting, the debate about financial conditions remains subdued.

5.3   Decomposing transmission channels

The previous result shed light on the overall effects of topic shocks. The model is not able to disentangle different transmission channels. At the same time, unconventional monetary policy such as asset purchases is often believed to work through two main channels: first, to the extent different asset classes are imperfect substitutes, asset purchases by the central bank raise bond prices and, through portfolio readjustments of investors, also other asset prices. This channel is referred to as the *portfolio balance channel*. Second, by purchasing assets the central bank conveys information about persistently low policy rates in the future. This affects market expectations and, as a result, asset prices. The latter effect is known as the *signaling*

14

channel.

Based on an estimated term structure mode, any change in Treasury yields can be decomposed into changes in expected short rates and changes in the term premium. In the context of quantitative easing this is particularly important as changes in the term premium are often associated with the *portfolio balance channel* and changes in the expectation component are reflecting the *signaling channel* (See, among others, Thornton, 2012; Bauer and Rudebusch, 2014; Wu, 2014).

To shed light on the two main transmission channels of asset purchases and, as a consequence, tapering, we substitute the Treasury yield used before by the estimated expectation component and, as a a separate variable, the estimated term premium. All three variables are taken from the model of Adrian et al. (2013).[8] The results for decomposed 10-year yields are shown in Figures (16) to (18). In each figure the red and green lines are depicting the impulse responses of the term premium. The black dotted line is the response of the expectations component of 10-year yields to a topics shock.

We find that the effects shown before were driven by the response of the term premium, which exhibits a significant response to the topics shocks. The expectations component shows a very weak response only. This difference between the responses of the two components of bond yields is most clearly visible for the "tantrum" and the "QE" topics. This finding is intuitive as the tapering decision of the Fed, which is discussed in our Twitter data, pertains to the exit from asset purchases under its QE3 program and not the eventual "lift-off" of short term interest rates. As a matter 0f fact, unwinding QE is a tightening policy action that make an eventual "lift-off" more likely, but the question of ending asset purchases clearly dominates the tapering discussion. Thus, our results point to a reversed *portfolio balance channel* during the taper tantrum episode.

5.4 Including the number of tweets

In the VAR models used so far, we derived impulse responses following a shock to each topic frequency. This shock reflects an increase in the share of a given topic on a particular day. The model does not include, however, a measure of how much Twitter activity there is on this particular day. It should matter whether we see a shift in topic probabilities in a day with only 100 tweets or on a day with 10.000 tweets. The number of tweets contains information about the attention the debate receives on Twitter. To control for the number of tweets, we include the log-level

---

[8]The fitted yields, the estimated expectation component the term premium are available at https://www.newyorkfed.org/research/data_indicators/term_premia.html.

of tweets in the vector Y_t. This variable is ordered fourth, i.e. after the topic frequencies and before the asset price. The results for 10-year yields are shown in Figures (19) to (21). When controlling for the number of tweets all results remain unchanged. We find results which are broadly similar to our baseline findings. Thus, our findings are independent from the overall number of tweets sent on a given day.

5.5 Changing the ordering in the VAR model

Our identification of shocks to topic frequencies is based on the assumption that on a given day the topic frequencies affect asset prices but not vice versa. This identifying assumption is reflected in the ordering of the endogenous variables in (3). To corroborate the robustness of our results, we reverse this ordering, that is, we order asset prices second, i.e. after the macroeconomic conditions, and the topics third and fourth. If the impulse responses do not change, our baseline results are robust with regard to the identification assumption.

Figure (22) reports the impulse response of 10-year yields to shocks in "tantrum"-related topics. As in the baseline model, bond yields rise after a shock to topic 3 ("market worries") and fall after a shock to topic 13 ("fear"). We do not show the responses to the other shocks to save space. All other results remain also unchanged. We can conclude that using a specific ordering of the variables in our Choleski identification scheme does not drive our findings.

5.6 Exogenous fundamentals

In our baseline model the daily proxy for macroeconomic fundamentals is ordered first, that is, changes in fundamentals have a contemporaneous effect on all other variables. However, with a delay of one day, changes in topics or in financial conditions are also allowed to affect fundamentals. To shut down this latter feedback effect, we estimate a model, in which fundamentals enter as a purely exogenous variable. They still affect all other variables contemporaneously, but the feedback from financial conditions on fundamentals is absent. Thus, the VAR model becomes a VARX model

$$Y_t = A_0 + A(L)Y_t + D_t^{FOMC} + A_Y Y_t^{CESI} + u_t \ \text{ for } E[u_t u_t'] = \Sigma_u, (4)$$

where Y_t^{CESI} enters with a coefficient vector A_Y. The results, see Figure (23) suggest that our results are also robust with respect to the treatment of the macroeconomic fundamentals.

16

6 Conclusions

The expectations about future monetary policy matter for asset prices. The process in which expectations are formed, however, is opaque. A key contribution to expectations formation is the public debate about future monetary policy among households and investors. This paper dissects the debate about monetary policy for a period with large swings in policy expectations - the "taper tantrum" episode in 2013. Based on a large data set containing all Twitter messages on the Fed's unwinding of asset purchases ("tapering") we use computational linguistic methods (LDA) to slice the debate into different topics. The frequencies of selected topics are then modeled in a VAR framework. We show that shock to selected topic frequencies have significant effects on U.S. bond yields, exchange rates and stock prices.

The results are robust to the specification of the VAR model and suggest that the discourse about policy in social media matters for asset prices. With the help of social media we can shed light on the black box of expectations formation, that is, how people share and comment on information and how an aggregate market view evolves. For applications for which expectations play an important role, such as monetary policy, asset pricing and central bank communication, social media offers interesting opportunities. In particular, questions related to central bank communication and expectations management, respectively, could be addressed by using high-frequency social media data.

In future research the cross-section or network dimension of the data can be used. The present paper employs daily aggregates of the Twitter exchange. It might also be fruitful to exploit the high-frequency flow of information in the network of Twitter users and the resulting formation of expectations.

References

[1] Adrian, T., R. K. Crump and E. Moench (2013): "Pricing the term structure with linear regressions", *Journal of Financial Economics* 110, 110-138.

[2] Aizenman, J., M. Binici and M. H. Hutchison (2014): "The transmission of Federal Reserve tapering news to emerging financial markets", *NBER Working Paper* No. 19980, National Bureau of Economic Research.

[3] Bauer, M. D. and G. D. Rudebusch (2014): "The signaling channel of Federal Reserve bond purchases", *International Journal of Central Banking* 10, 233-289.

[4] Bholat, D., S. Hansen, P. Santos and C. Schonhardt-Bailey (2015): "Text mining for central banks", *CCBS Handbook* No. 33, Centre for Central Banking Studies, Bank of England.

[5] Blei, D. M., Ng, A. Y. and M. I. Jordan (2003): "Latent Dirichlet Allocation", *Journal of Machine Learning Research* 3, 993-1022.

[6] Blei, D. M. and J. D. Lafferty (2007): "A correlated topic model of science", *Annals of Applied Statistics* 1, 17-35.

[7] Blinder, A. S., M. Ehrmann, M. Fratzscher, J. de Haan, and D.-J. Jansen (2008): "Central bank communication and monetary policy: A survey of theory and evidence", *Journal of Economic Literature* 46, 910-945.

[8] Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W. and R. A. Harshman (1990): "Indexing by latent semantic analysis", *Journal of the American Society for Information Science* 41, 391-407.

[9] Eichengreen, B. and P. Gupta (2013): "Tapering talk: the impact of expectations of Reduced Federal Reserve security purchases on emerging markets", *unpublished*, University of Berkeley.

[10] Griffiths, T. L. and M. Steyvers (2004): "Finding scientific topics", *Proceedings of the National Academy of Sciences of the United States of America* 101, 228-5235.

[11] Grün, B. and K. Hornik (2011): "`topicmodels`: An `R` package for fitting topic models", *Journal of Statistical Software* 40, 1-30.

[12] Hansen, S. and M. McMahon (2015): "Shocking language: understanding the macroeconomic effects of central bank communication", *unpublished*, University of Warwick.

[13] Hansen, S., M. McMahon and A. Prat (2015): "Transparency and deliberation within the FOMC: a computational linguistics approach", *unpublished*, University of Warwick.

[14] Hendry, S. (2010): "Text mining and the information content of Bank of Canada communication", *Bank of Canada Working Paper* 2010-31, Bank of Canada.

[15] Hendry, S. (2012): "Central bank communication or the media's interpretation: what moves markets?", *Bank of Canada Working Paper* 2012-9, Bank of Canada.

[16] Hofmann, T. (1999): "Probabilistic latent semantic indexing", *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval* 50-57.

[17] Larsen, V. H. and L. A. Thorsrud (2015): "The values of news", *unpublished*, BI Norwegian Business School.

[18] Loughran, T. and B. McDonald (2015): "Textual analysis in accounting and finance: a survey", *unpublished*, University of Notre Dame.

[19] Lucca, D. O. and F. Trebbi (2009): "Measuring central bank communication: an automated approach with application to FOMC statements", *NBER Working Paper* No. 15367, National Bureau of Economic Research.

[20] Lüdering, J. and P. Winker (2016): "Forward or Backward Looking? The Economic Discourse and the Observed Reality", MAGKS Joint Discussion Papers in Economics 2016 Nr. 07-2016, Marburg

[21] Mishra, P. K. Moriyama, P. N'Diaye and L. Nguyen (2014): "Impact of Fed tapering announcements on emerging markets", *IMF Working Paper* No. 14/109, International Monetary Fund.

[22] Meinusch, A. and P. Tillmann (2016): "Quantitative Easing and tapering uncertainty: evidence from Twitter", *unpublished*, Giessen University.

[23] Schonhardt-Baily, C. (2013): *Deliberating Monetary Policy*, MIT Press, Cambridge.

[24] Thornton, D. L. (2012): ""Evidence on the portfolio balance channel of Quantitative Easing", *Working Paper* No. 2012-015A, Federal Reserve Bank of St. Louis.

[25] Wang, X. McCallum, A. (2006): "Topics over time: a non-Markov continuous-time model of topical trends", *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* 424-433.

[26] Wu, T. (2014): "Unconventional monetary policy and long-term interest rates", *IMF Working Paper* No. 14/189, International Monetary Fund.

Figure 3: Frequencies of selected topics
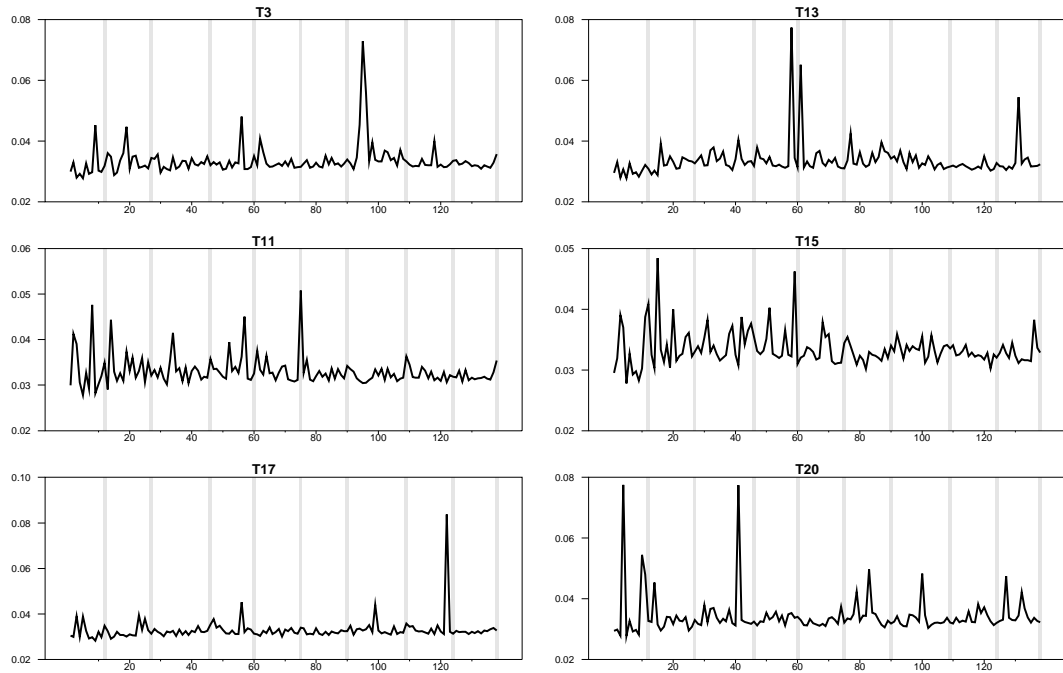
Notes: The shaded areas indicate FOMC meeting dates and dates of FOMC minutes releases. The horizontal axis reflects our sample period, i.e. April 15 to October 30 2013.

21

Figure 4: Response of 10-year yields
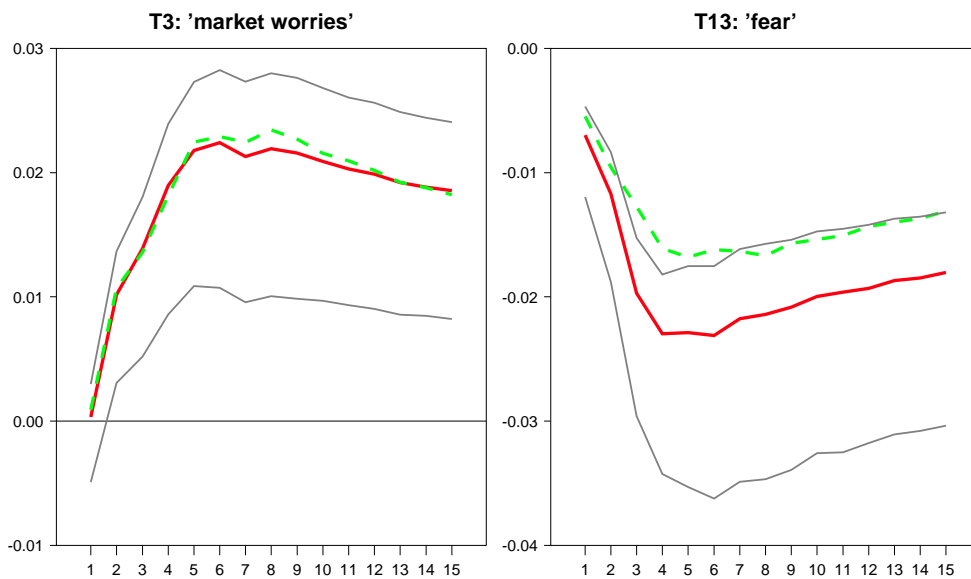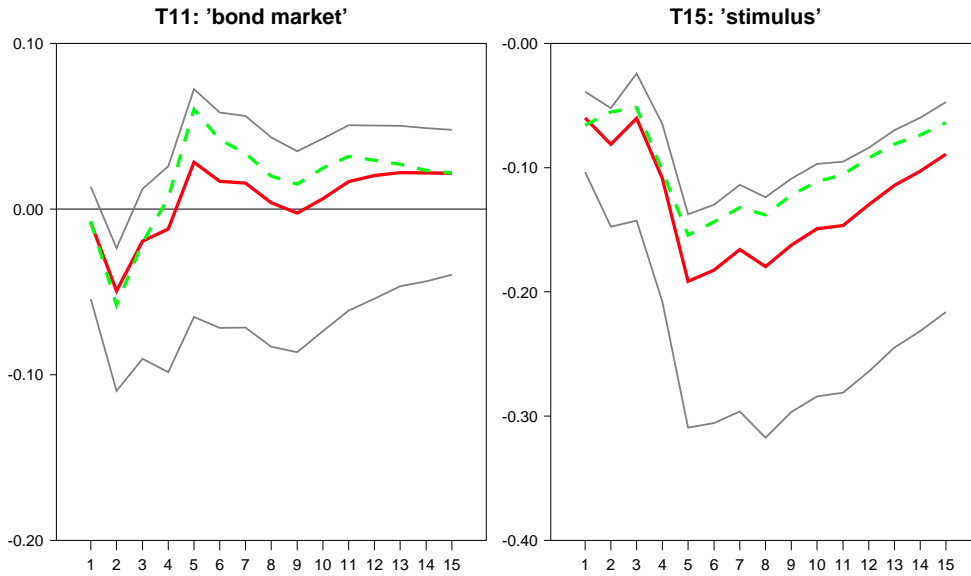

T3: 'market worries'　　　　**T13: 'fear'**

Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.
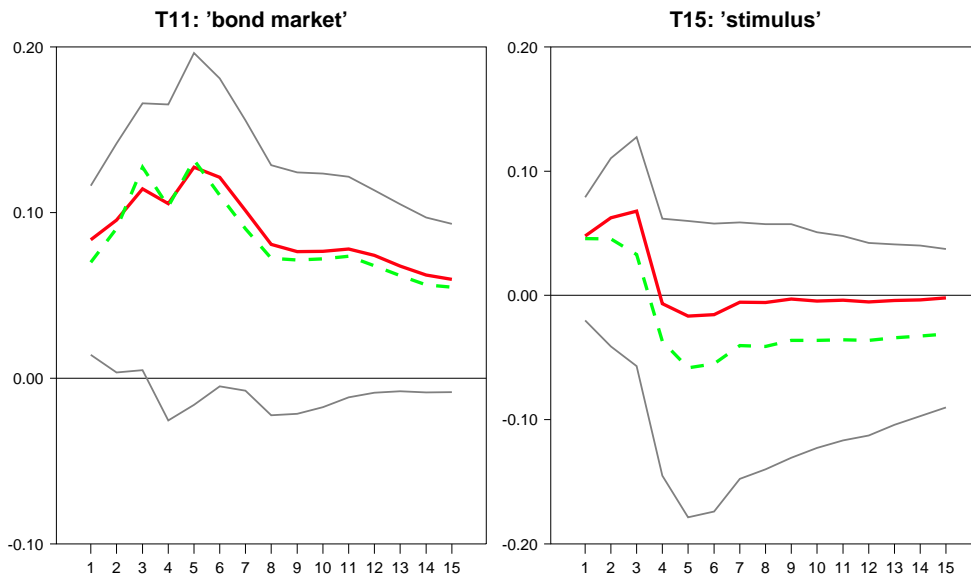
Figure 5: Response of 5-year yields
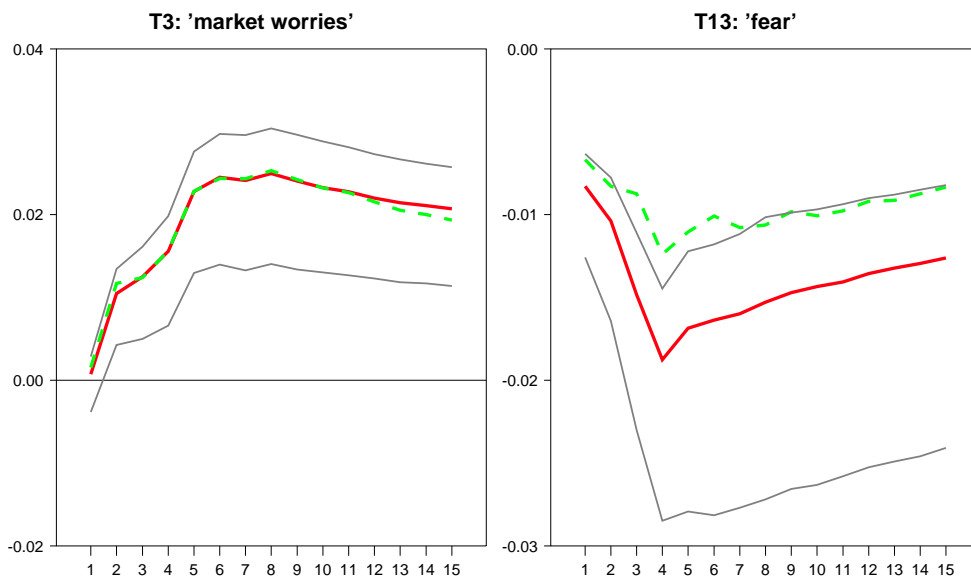
T3: 'market worries'　　　　**T13: 'fear'**

Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.
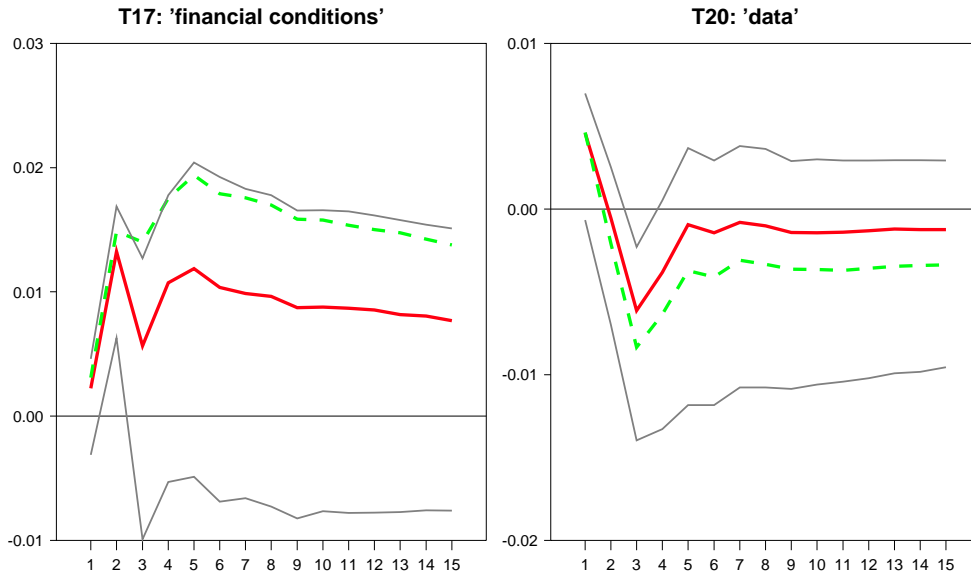
Figure 6: Response of USD-EUR exchange rate
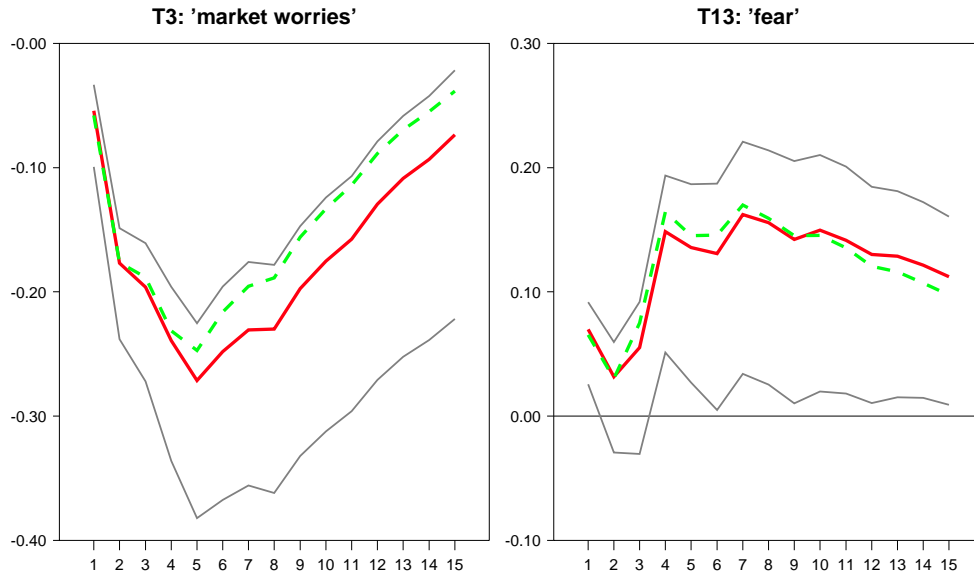
Notes: The exchange rate is ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

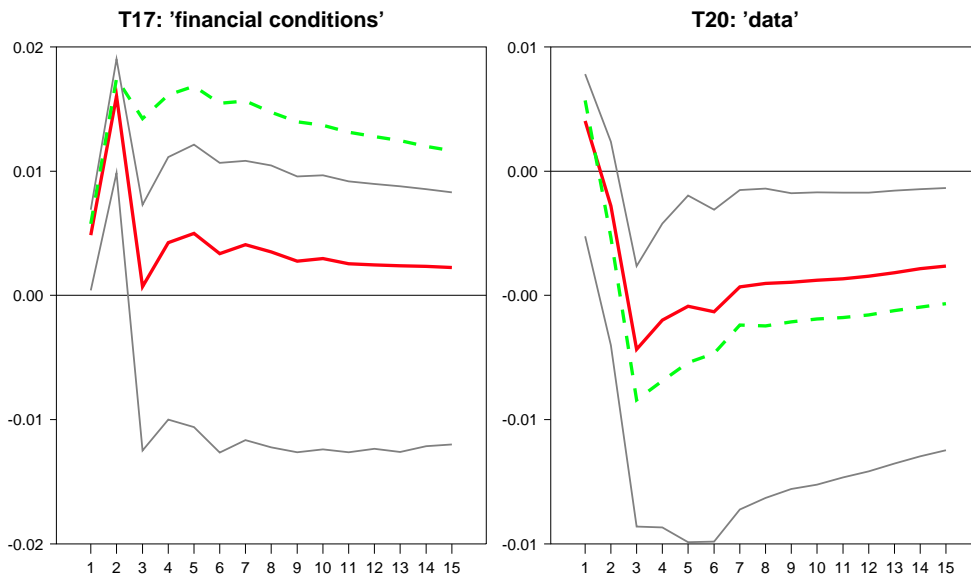Figure 7: Response of stock prices
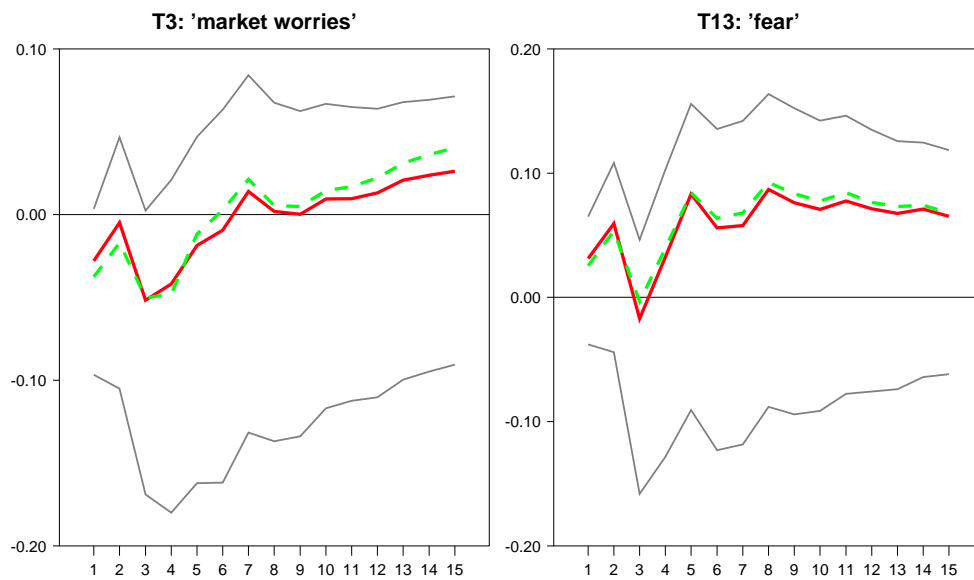
Notes: Stock prices are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

23

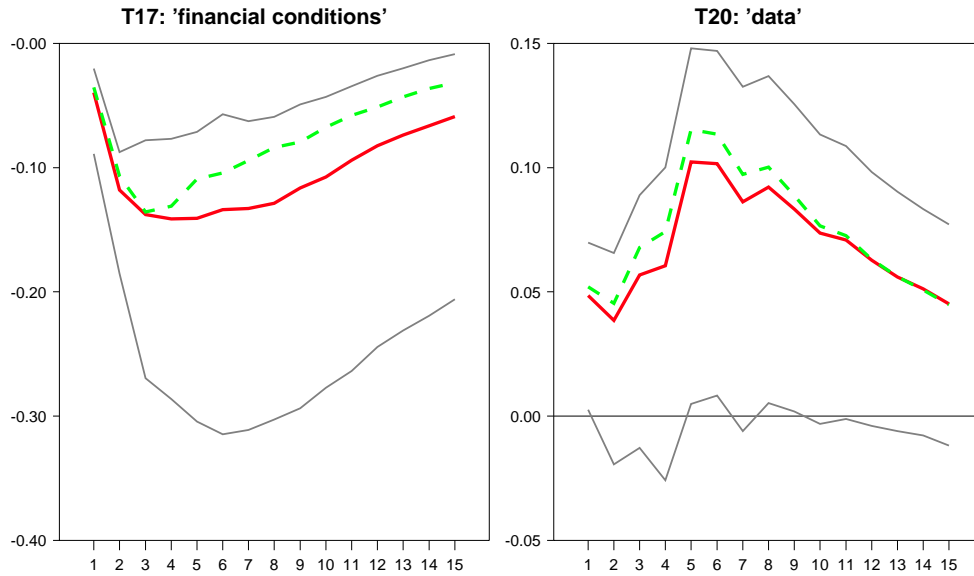Figure 8: Response of 10-year yields
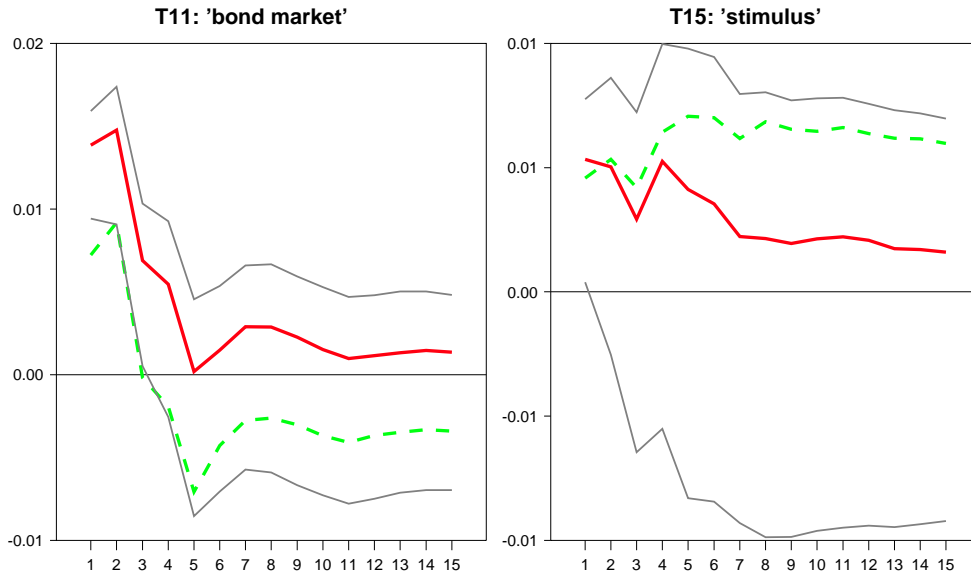
Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.
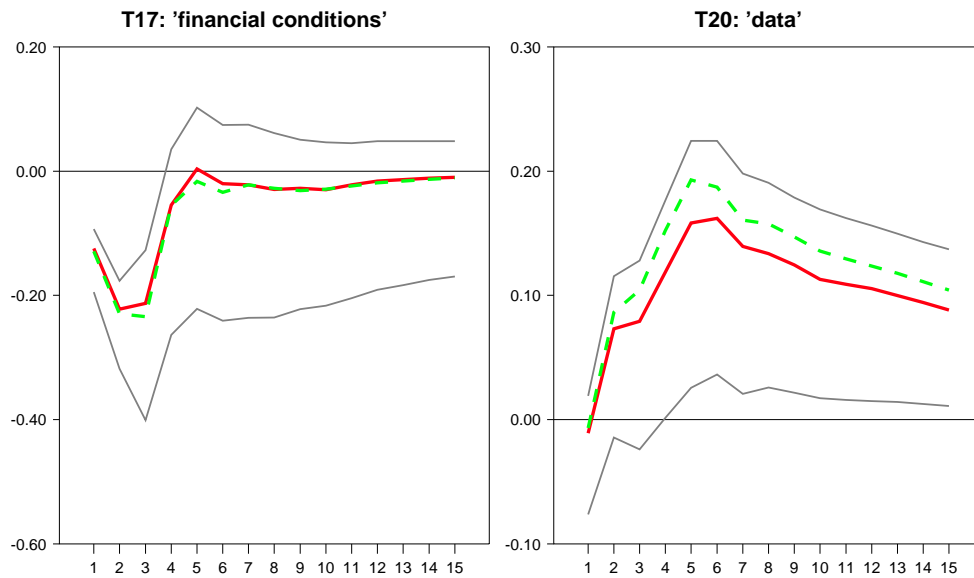
Figure 9: Response of 5-year yields
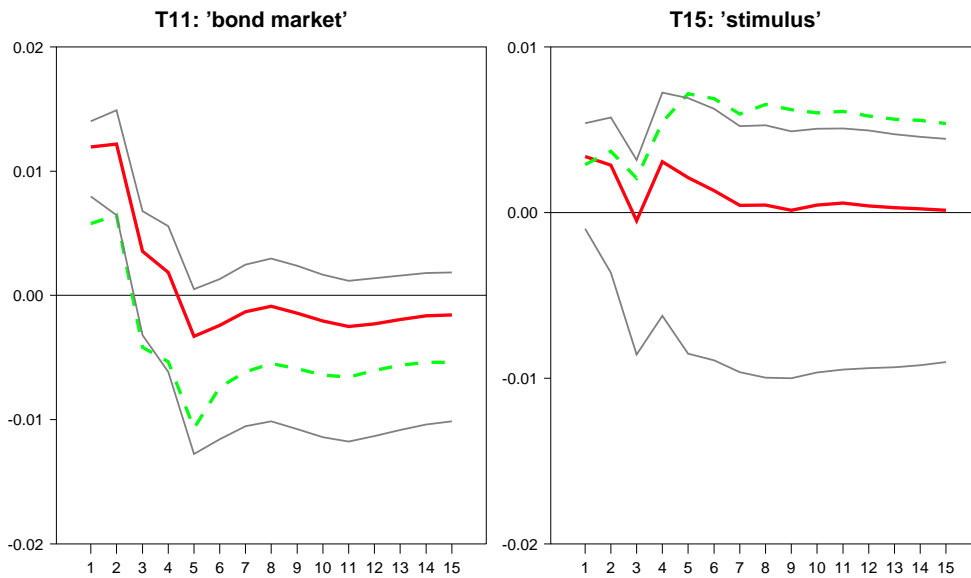
Notes: Yields ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 10: Response of USD-EUR exchange rate

Notes: The exchange rate is ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 11: Response of stock prices

Notes: Stock prices are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 12: Response of 10-year yields

Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 13: Response of 5-year yields

Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 14: Response of USD-EUR exchange rate

Notes: The exchange rate is ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.
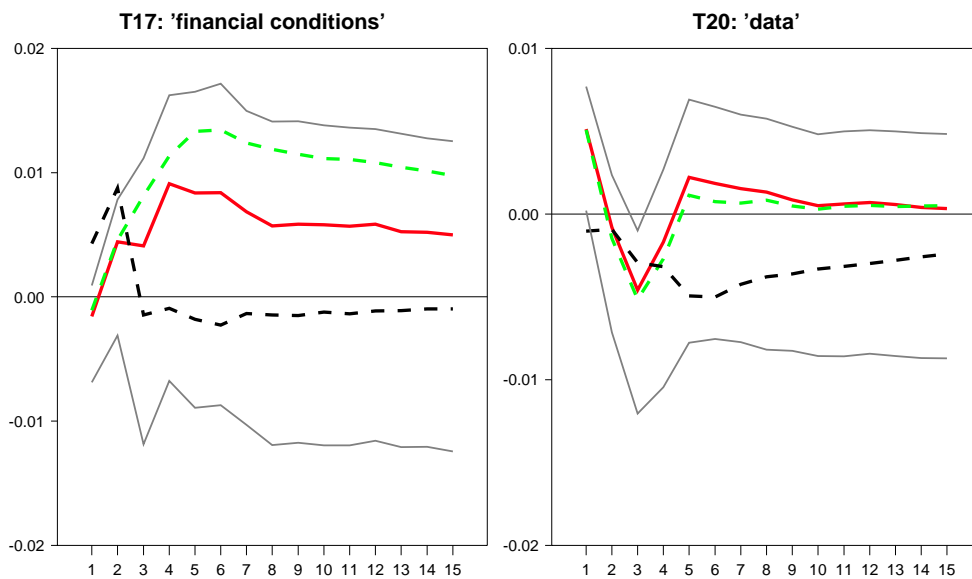
Figure 15: Response of stock prices

Notes: Stock prices are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 16: Response of 10-year yields with different transmission channels
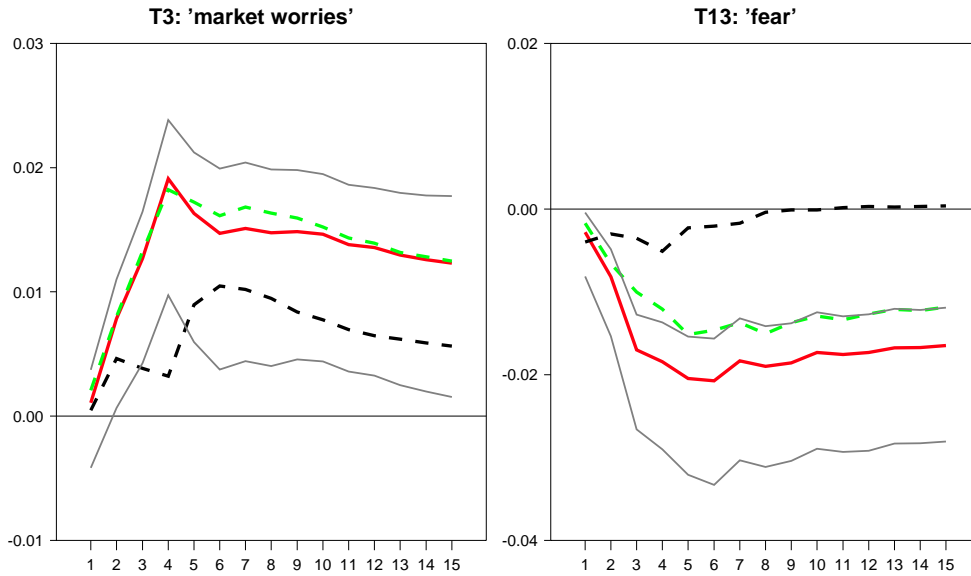
T3: 'market worries'

T13: 'fear'

Notes: The expectations component and the term premium are ordered second last and last, respectively. The red line plots the response of the <u>term premium</u> for a VAR with dummies for FOMC meetings and minutes releases, the green line depicts the response of the term premium for a model without dummies and the black line is the response of the <u>expectations component</u> of yields.

Figure 17: Response of 10-year yields with different transmission channels
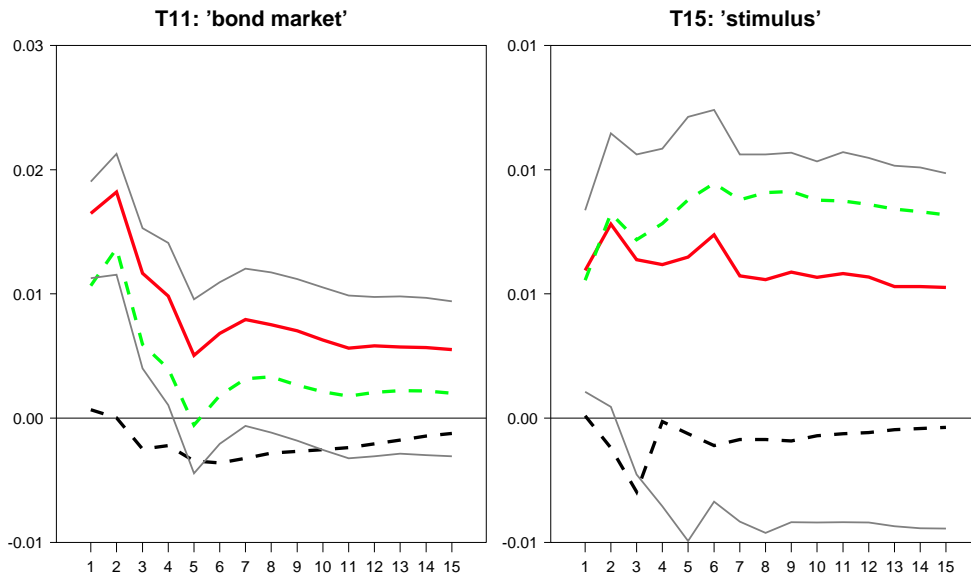

T11: 'bond market'

T15: 'stimulus'

Notes: The expectations component and the term premium are ordered second last and last, respectively. The red line plots the response of the <u>term premium</u> for a VAR with dummies for FOMC meetings and minutes releases, the green line depicts the response of the term premium for a model without dummies and the black line is the response of the <u>expectations component</u> of yields.

Figure 18: Response of 10-year yields with different transmission channels

Notes: The expectations component and the term premium are ordered second last and last, respectively. The red line plots the response of the term premium for a VAR with dummies for FOMC meetings and minutes releases, the green line depicts the response of the term premium for a model without dummies and the black line is the response of the expectations component of yields.

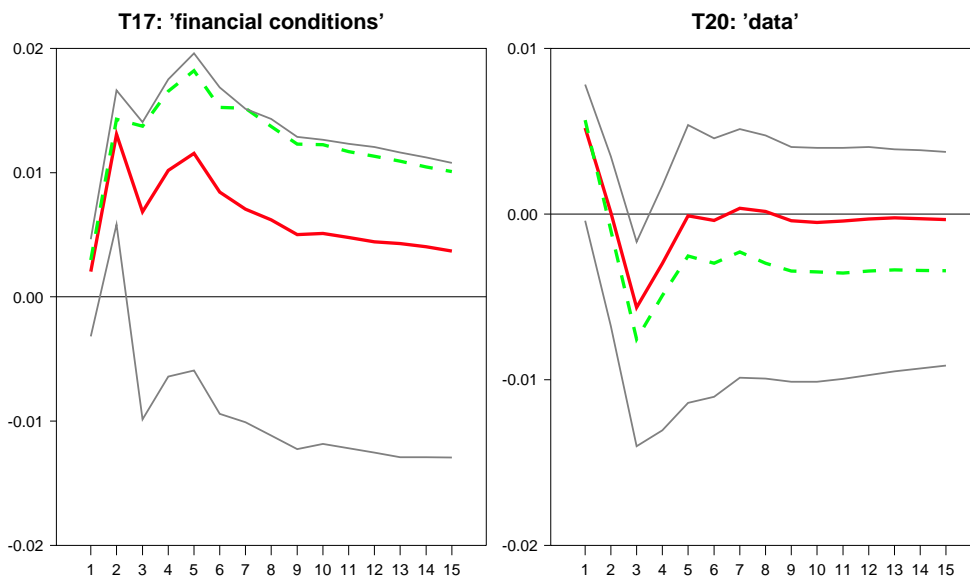Figure 19: Response of 10-year yields with number of tweets
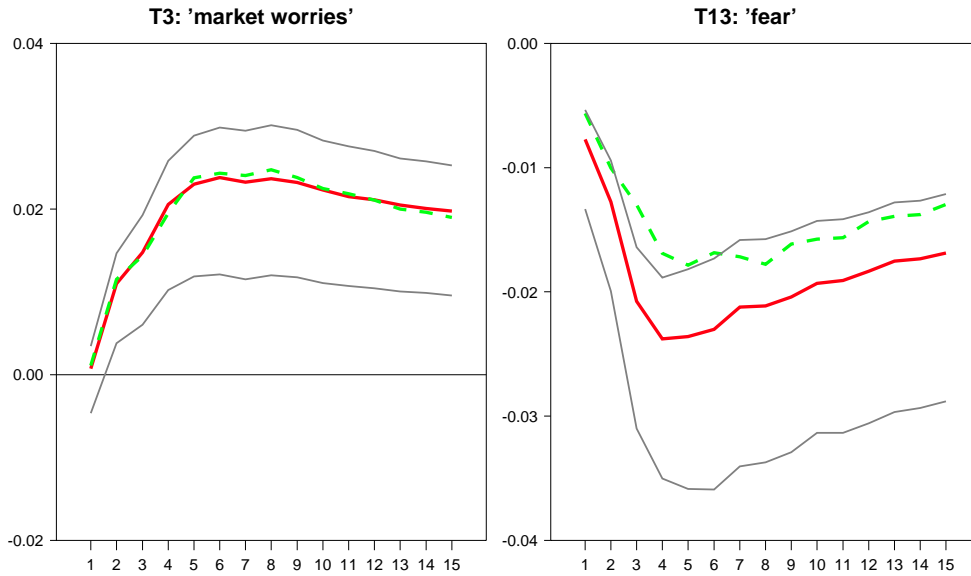
Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. The model includes the log of daily <u>number of tweets</u> which are ordered fourth.

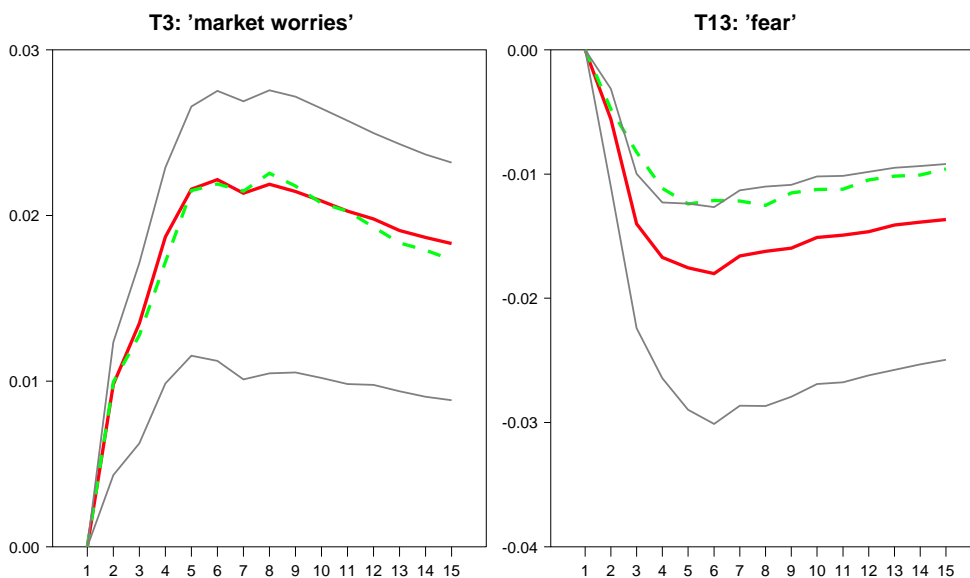Figure 20: Response of 10-year yields with number of tweets
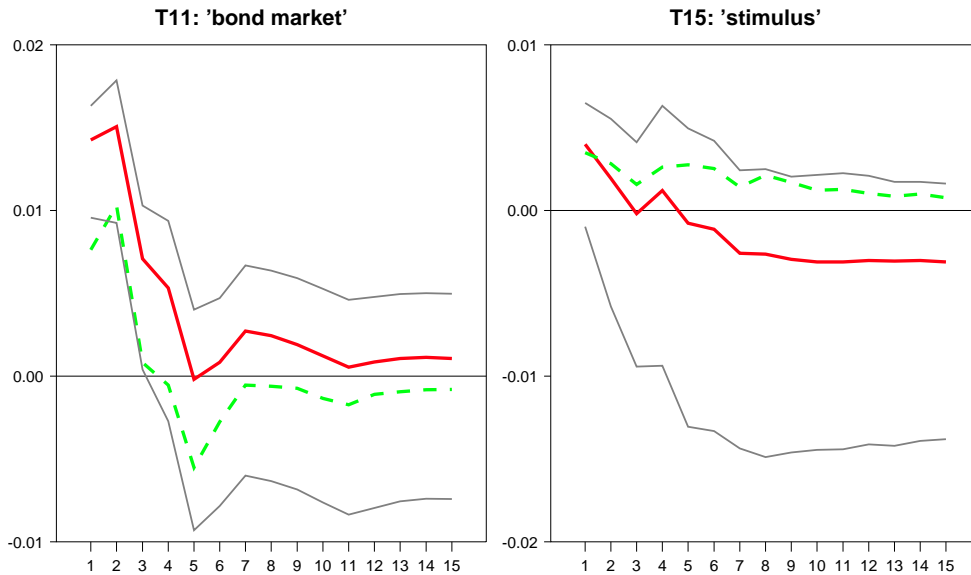
Notes: Yields are ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. The model includes the log of daily <u>number of tweets</u> which are ordered fourth.

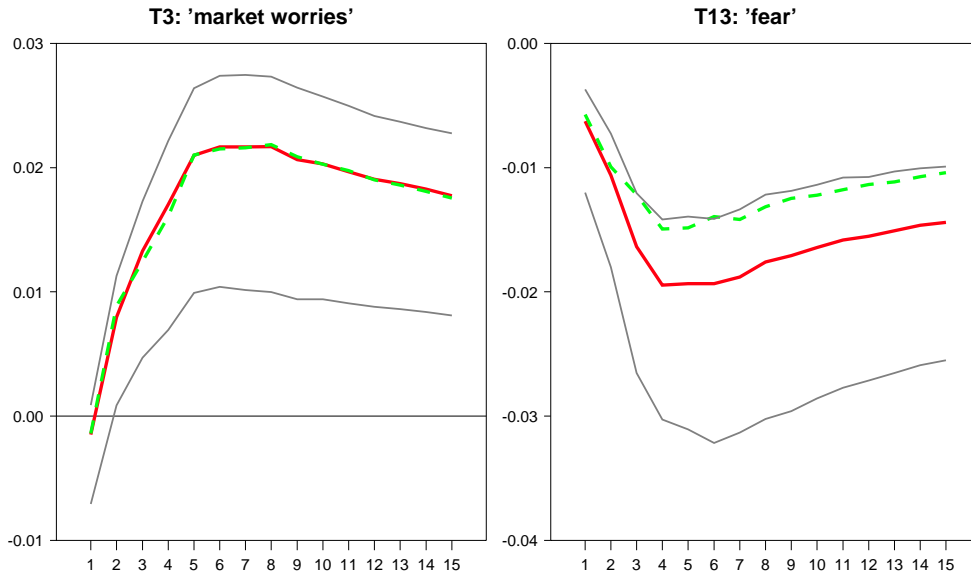Figure 21: Response of 10-year yields with number of tweets

Notes: The exchange rate is ordered last. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. The model includes the log of daily <u>number of tweets</u> which are ordered fourth.

Figure 22: Response of 10-year yields for alternative VAR ordering

Notes: <u>Yields are ordered second</u>, i.e. before topics. The red line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Figure 23: Response of 10-year yields for exogenous fundamentals

T3: 'market worries'

T13: 'fear'

Notes: Yields are ordered first, i.e. before topics. The red line is the baseline result from a VARX with dummies for FOMC meetings and minutes releases and the the green line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

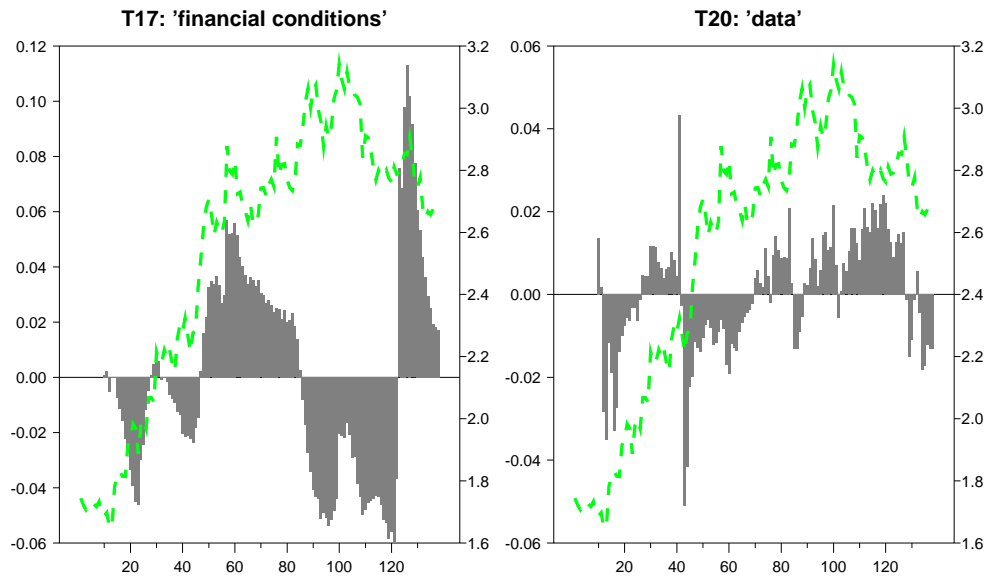Figure 24: Contribution to 10-year yields
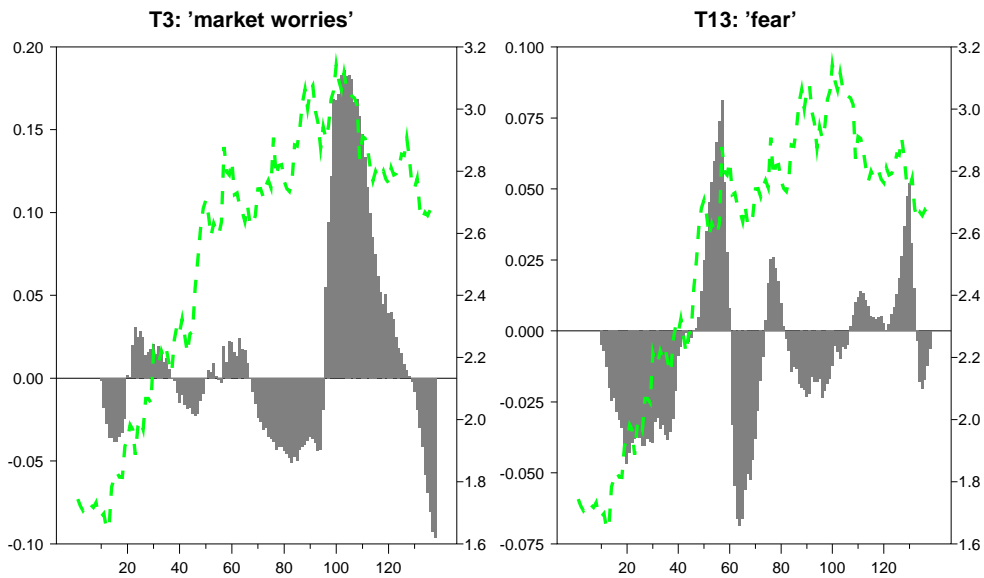
T3: 'market worries'

T13: 'fear'

Notes: The gray bars indicate the contribution of the topic to the evolution of bond yields. The green (dotted) line (right scale) gives the 10-year bond yield.

Figure 25: Contribution to 10-year yields
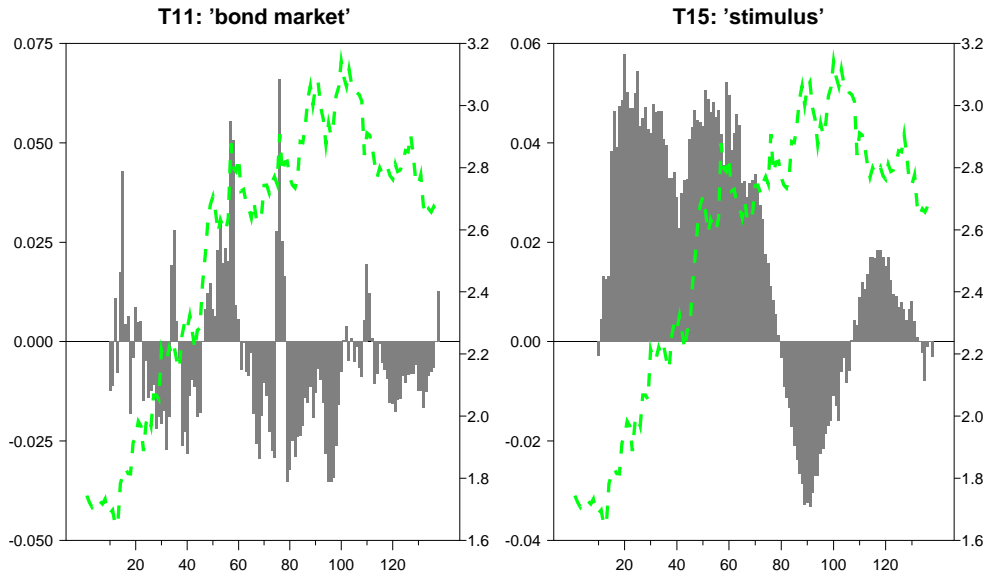
T11: 'bond market'   **T15: 'stimulus'**

Notes: The gray bars indicate the contribution of the topic to the evolution of bond yields. The green (dotted) line (right scale) gives the 10-year bond yield.

Figure 26: Contribution to 10-year yields

T17: 'financial conditions'   **T20: 'data'**

Notes: The gray bars indicate the contribution of the topic to the evolution of bond yields. The green (dotted) line (right scale) gives the 10-year bond yield.

(a) Topic 3: "market worries"

(b) Topic 11: "bond market"

(c) Topic 13: "fear"

(d) Topic 15: "stimulus"

(e) Topic 17: "financial conditions"

(f) Topic 20: "data"

Figure 27: Selected topics