North American Journal of Economics and Finance 51 (2020) 100875

Contents lists available at ScienceDirect

North American Journal of Economics and Finance

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

Monetary policy on twitter and asset prices: Evidence from computational text analysis

Jochen Lüdering[1], Peter Tillmann*,[2]

Justus-Liebig-University Gießen, Germany

A R T I C L E I N F O

Keywords:
Monetary policy
Fed
Latent Dirichlet Allocation
Text analysis
VAR

JEL classification:
E32
E44
E52

A B S T R A C T

In this paper, we dissect the Twitter 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 future Fed policy. Based on a novel data set of 90,000 Twitter messages ("tweets") covering the 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 the discussion about the timing of the tapering, about the broader economic policy context and worrying investors 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.

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, Ehrmann, Fratzscher, de Haan, & Jansen, 2008), there is little evidence about how, given the communication of the central bank, the discourse about monetary policy among 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 eventual 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

We thank Matthias Neuenkirch, Peter Winker, seminar participants at the Bundesbank Big Data group and the COMPSTAT, CMStatistics and Money, Macro and Finance Group conferences for very helpful comments. A referee of this journal provided important feedback. We are grateful to Annette Meinusch, Lucas Hafemann and Florian Viereck for preparing the data set.

* Corresponding author.

E-mail addresses: research@jochenluedering.de (J. Lüdering), peter.tillmann@wirtschaft.uni-giessen.de (P. Tillmann).

[1] Justus-Liebig-University Gießen, Center for International Development and Environmental Research, Senckenbergstrasse 3, 35390 Giessen, Germany.

[2] Justus-Liebig-University Gießen, Department of Economics, Licher Str. 66, 35394 Giessen, Germany.

https://doi.org/10.1016/j.najef.2018.11.004
Received 30 April 2018; Received in revised form 26 October 2018; Accepted 13 November 2018
Available online 15 November 2018
1062-9408/

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 traffic on Twitter.com, the social media network, on Fed tapering. The data set consists of 90,000 text messages ("tweets"), sent 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 by 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, or endorsed. 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 & Tillmann, 2016), we construct a dictionary with keywords describing a certain expected 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 ignore other dimensions of the discussion. In this paper we employ a tool set taken from computational linguistics. Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) is used to extract latent topics out of the Twitter conversation on tapering. LDA dissects a text document into different topics based on Markov Chain Monte Carlo estimation. It estimates a certain number of different topics which, based on each topic's most frequent words, will be labeled manually.

We extend the results of Meinusch and Tillmann (2016) in two dimensions. First, we are able to deconstruct the discussion on Twitter into topics. Some of these topics contain price-relevant information, others do not. Hence, we can separate the news from the noise in the data. Second, we let the data speak freely and do not impose a pre-specified list of words that carry a certain meaning.

Examples of topics are "September", "next year", "market worries", "dollar" and many more. The resulting topic frequencies, which express the likelihood that a given tweet contains words belonging to 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 topic 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 "September", for example, raises bond yields, leads to an appreciation of the dollar and a fall in stock prices as it reflects the heightened discussion about an early tapering decision.

We also use a decomposition of 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 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 (2016).[3] For the purpose of this paper, the existing research applying models of latent information in textual data is particularly interesting.

Hansen and McMahon (2016) apply LDA modeling to the entire history of policy transcripts 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.[4] 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.[5] Hendry and Madeley (2010) and Hendry (2012) use latent semantic analysis for the communication of the Bank of Canada. They identify "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 news indices are related to economic activity within a Bayesian regression framework. Using a similar methodology but covering a time horizon of 61 years, Lüdering and Winker (2016) examine whether economists anticipate changes in the state of the economy or merely discuss these events *ex-post* by looking at the relationship between economic publications and real-world time-series. They find a significant link between the economic discussion and the time-series. In two out of five cases, the economic discussion precedes the economic developments, supporting the argument that economists are – to a certain degree – able to anticipate economic developments.

Two recent papers also use Twitter data for studying monetary policy. Azar and Lo (2016) use Twitter to predict the evolution of the stock market around meetings of the Fed's FOMC. Luik and Wesselbaum (2016) employ tweets from the Fed's own Twitter account to identify changes to the Fed's monetary policy communication.

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.

[3] Bholat, Hansen, Santos, and Schonhardt-Bailey (2015) provide a survey of text mining applications relevant for central banks.

[4] Using the same data set, Hansen, McMahon, and Prat (2015) model the effect of increased transparency on the policy debate in FOMC meetings.

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

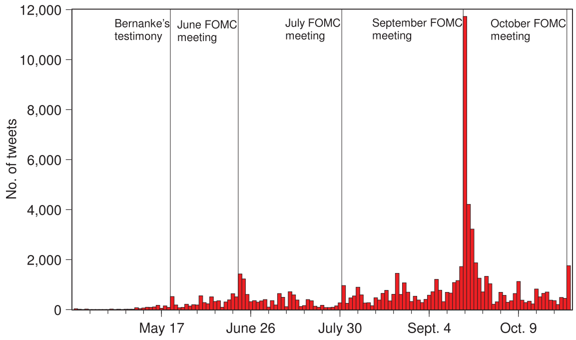


Fig. 1. Number of Tweets. Notes: The vertical lines indicate the testimony of chairman Bernanke on May 22 and the subsequent FOMC meetings.

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 31, 2013. The data has been purchased from Gnip.com, which in the meantime has been acquired by Twitter. 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 most of the tapering debate on Twitter. After deleting a few tweets in languages other than English, we are left with 87038 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 have been excluded due to the lack of asset price data. Finally, tweets are aggregated to daily frequency. Fig. 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 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, markets 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.

We believe most of the Twitter accounts represented in this data set belong to financial professionals. This assessment is based on three pieces of information: First, we filter the tweets by the word ”taper”, which together with ”Fed” is the jargon used only by journalists, Fed watchers and researchers. Second, most tweets are geo-tagged and suggest most users are located either on the east coast of the U.S. or in the city of London. There is only a much smaller fraction of users from other parts of the world. Finally, we have lots of tweets from well-known market economists in the sample.

From Fig. 1 we see that the number of tweets is systematically higher on FOMC meeting days.[6] 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 topic models to dissect the discussion into policy-relevant topics. We are interested in 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 (2015), Aizenman, Binici, and Hutchison (2016) and Mishra, Moriyama, N’Diaye, and Nguyen (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. Meinusch and Tillmann (2016) use the same Twitter data set that is used in this paper. Based on a dictionary of words they 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.

[6] Due to the convention of the software used, the vertical lines in Fig. 1 are drawn at the beginning of each of each day, while the bars are plotted from the beginning to the end of a given day.

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*[7] for pre-processing the data and subsequently *topicmodels*[8] in order to fit 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 vocabulary is a subset of all the terms in the corpus, our entire set of tweets, which are well suited to differentiate the individual documents. Conversely, all terms which have little explanatory power are eliminated from the vocabulary in order to ease the estimation and produce meaningful topics. This is achieved by removing so called stopwords, which include pronouns, conjunctions and auxiliary verbs with little meaning on their own. We use the list of English language stopwords provided by the R package *tm* as a basis. In addition, the terms "Fed" and "taper", present in each tweet, are removed from the corpus, as these words have been used to filter the tweets in the first place.

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 turned to lowercase. By applying the stemming algorithm *SnowballC* wordstems are obtained. 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 differences in capitalization.

Due to the short length of a tweet, we only impose the restriction that a term has to appear in at least five documents and should have a length, after stemming, between 4 and 20 characters. We refrain from any further filtering based on a word's importance. It would be common practice to discard terms with a tf-idf score below the median (Blei & Lafferty, 2007), implying that these terms are not very characteristic for individual documents but are evenly distributed over the whole corpus. However, tweets appear to contain very few of these words, probably due to their length limited in 2013 to 140 characters, leading to a median tf-idf score larger than one (compared to 0.1 in Grün and Hornik). Compared to other applications of LDA, our dataset differs in its ratio of different terms (5076) to the number of documents in the corpus (87038). Usually, the number of different terms is much larger than the number of documents. This data is used to create a document-by-term matrix $\boldsymbol{F}$, which holds the frequencies $f_{i,j}$ of $|V| = 5076$ different terms in $D = 87038$ tweets, serving as the basis for the subsequent LDA estimation.

3.2. Latent Dirichlet Allocation

This section provides a brief overview of LDA models and the estimation method behind our analysis. Before their recent arrival in economics, topic models have been used since the 1990s (Deerwester, Dumais, Landauer, Furnas, & Harshman, 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 et al. (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 D are generated from a fixed set of K different topics. The topics consists of a set of terms from a vocabulary V. Each term w in a document $\boldsymbol{w}$ is generated by drawing a topic k, given a vector of topic probabilities $\theta_{\boldsymbol{w}}$, and subsequently drawing a term from the topic, given its probability β_k conditional on the topic. Hence, the probability of word w_i is given by

$$P(w_i) = \sum_{j=1}^{K} P(w_i|k_j)P(k_j). \tag{1}$$

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 do not proof suitable. Hence, a number of sophisticated methods have 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,

[7] Seehttps://cran.r-project.org/web/packages/tm/.
[8] Seehttps://cran.r-project.org/web/packages/topicmodels/.

a Markov Chain Monte Carlo approach suggested by Griffiths and Steyvers (2004). Instead of estimating the topic distribution θ and the term distribution β directly, the vector $\boldsymbol{z}_i$, the assignments of words to topics is estimated. Based on these assignments of words to topics, approximations of θ and β can be computed. 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, i.e. Dir(α) and Dir(δ). The parameters on the Dirichlet distributions are chosen according to the literature (Griffiths & 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. It represents a trade-off between broader topics (smaller K), aiding the interpretation, and a better decomposition of the data (larger K). In order to obtain an estimate for K, we apply the harmonic mean method as suggested by Griffiths and Steyvers (2004).[9] 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. Thus, simple maximization does not necessarily lead to useful results.[10] We end up with an unreasonably large number of topics given our data. As a consequence the obtained topics are very narrow and difficult to interpret. Hence, we follow the pragmatic approach by Hansen et al. (2015) 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 Griffiths 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 of sequential updating, the Markov Chain is assumed to have converged. The approximation for document probabilities θ is calculated based on the number of times the MCMC algorithm has assigned document $\boldsymbol{w}$ to topic k, measured by the count variable $\eta_k^{\boldsymbol{w}}$, relative to the sum of all assignment of document $\boldsymbol{w}$ to any topic

$$\hat{\theta}_k^{\boldsymbol{w}} = \frac{\eta_k^{\boldsymbol{w}} + \alpha}{\sum_k \eta_k^{\boldsymbol{w}} + K\alpha}. \tag{2}$$

Analogously the vector of topic probabilities β is approximated from the number of times that term w_i has been associated with topic k as indicated by count variable $\eta_k^{w_i}$, relative to the sum of all associations of any word to topic k

$$\hat{\beta}_k^{w_i} = \frac{\eta_k^{w_i} + \delta}{\sum_{w_i} \eta_k^{w_i} + |V|\delta}. \tag{3}$$

The resulting matrix of term probabilities $\hat{\beta}$ reveals the contents of the 30 different topics, which reflect the discussion on Twitter. Table 2 lists the five most frequent words of each topic.

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.

4.1. The information content of topic frequencies

Because we want to model a parsimonious VAR system, we cannot include all 30 topic frequencies in the VAR model jointly, which would leave too few degrees of freedom for estimation. Moreover, including all topics would make identification even more problematic. Instead, we focus on selected topics only. Hence, we need to select a few topics from the overall set of 30 topics.

To obtain an overview over the information content in all topics, we start by estimating a VAR model that includes one topic frequency at a time. In particular, we use the model to derive the dynamic responses of asset prices to shocks in each topic frequency. The reduced-form representation of the VAR is

$$Y_t = A_0 + A(L)Y_t + s_t^{FOMC} + D_t^{Minutes} + x_t^{ADS} + u_t \text{for} E[u_t u_t'] = \Sigma_u, \tag{4}$$

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. The vector Y_t contains topic i for $i \in (1, ..., 30)$ and an U.S. asset price A_t^{US}

$$Y_t = (Topic_t^i A_t^{US})'. \tag{5}$$

In the VAR model, three variables enter exogenously. First, we include the Aruoba, Diebold, and Scotti (2009) index of daily macroeconomics news announcements, x_t^{ADS}. It is important to control for data releases in our empirical model as market expectations reflected in Twitter messages also reflect macroeconomic news. Second, s_t^{FOMC} is the series of monetary policy shocks

[9] For an alternative procedure that determines the number of topics as a result of a trade-off between the salience of topics and the load on an given topic see Goldsmith-Pinkham, Hirtle, and Lucca (2016).

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

identified by Nakamura and Steinsson (2017). These authors identify surprise changes in interest rates around FOMC meeting days based on the high-frequency change federal funds futures in a narrow window around the FOMC announcement. By including this series we control for the monetary policy news on FOMC days. Third, we include a set of three dummies, $D_t^{Minutes}$, each of which is equal to one at a particular release date of FOMC minutes and zero otherwise. 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 the dummies for the releases of minutes. All three variables are not affected by shocks to Twitter beliefs, at least not on a daily basis over a relatively small sample, and hence enter as exogenous variables.

Asset prices in the U.S. are reflected by A_t^{US}. We use three alternative asset prices: the 10-year yield, the (log of the) nominal USD/EUR exchange rate and the (log) S&P500 stock price index. All asset prices are taken from the St. Louis Fed's FRED database. The yield data are fitted yields from the Adrian, Crump, and Moench (2013) term structure model. We use fitted yields which allow us to decompose yields into the expected short rates and the term premium which we will use below. 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 $T = 140$.

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 the variable ordered behind it in the VAR model. Here the ordering implies that a change in $Topic_t^i$, for example, has an effect on asset prices on the same day, while a change in asset prices can affect the Twitter topic 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. The attractiveness from Twitter as a medium of exchange stems from the speed by which users can respond to news and interact with others. Any information contained in the Twitter exchange should be allowed to move markets instantaneously. At the same time, we accept the restriction that users cannot contemporaneously respond to asset price developments. In fact, this restrictions helps us controlling our results for those tweets that simply comments on the ups and downs in asset markets and focusing on tweets that contain views about future policy. To assess whether this ordering assumption is innocuous, we will present results from an alternative ordering below.[11]

Since we cannot present the full impulse responses for each of the 30 estimated models, we report the cumulative response of the asset price variable only. Fig. 2 shows the cumulative response over five day following the shock for the baseline model with 10-year bond yields, the extended baseline model that includes the MOVE index of bond market volatility, which is explained below, and the baseline model for the USD/EUR exchange rate. As a matter of fact, the cumulative response is just one out of many ways to summarize an impulse response by one number. The caveat is that this measure hides information about the shape of the response. A response that is significantly different from zero on impact and then quickly converges back to zero could turn out to be insignificantly different from zero when accumulated over five days. Therefore, the confidence band around the cumulative responses has to be interpreted with some caution. We find that many topics contain information which is relevant for asset prices. Bond yields and the exchange rate respond significantly to shocks to several topic frequencies.

4.2. Selected topics

In order to take a closer look on the responses, we select eight topics out of the 30. We do not pick the topics that exhibit the strongest impact on asset prices as shown in Fig. 2, but those for which we believe the content is most relevant for the tapering debate in 2013. In particular, we identify four pairs of topics, which are directed towards particularly important aspects of the assessment of future monetary policy and the tapering decision, receptively. We allocate names to each selected topic which summarize each topic's content as described by the five most frequent words in Table 2. Each pair comprises $Topic_t^i$ and $Topic_t^j$. Fig. 3 presents the content of each selected topic as a word cloud, where the size of each word in the cloud reflects the significance (probability) of the word for a given topic. While the word list in Table 2 gives a broad assessment of each topic, the word clouds help interpreting the content of each topic in light of the debate about monetary policy. Table 1 summarizes the topics we selected for the following estimation:

The selected topics capture the uncertainty about the timing of the exit from QE, topic "September" and topic "next year", the policy mix in the U.S. and the business cycle, topic "policy" and topic "business cycle", the policy rate and the dollar, topic "interest rate" and topic "dollar" and the communication and perception of policy, respectively, topic "market worries" and topic "talk". We are agnostic with regard to the signs of the effects shocks to these topics have on U.S. asset prices and rely on the VAR model to highlight the market impact of each topic.

We estimate a VAR model for each pair of topics. The VAR model presented before is extended for a second topic, $Topic_t^j$,

$$Y_t = (Topic_t^i \, Topic_t^j \, A_t^{US})'. \tag{6}$$

All other parts of the VAR model, including the exogenous variables and the identification of shocks, remain unchanged.

[11] The problem with alternative identification schemes, e.g. sign restrictions or heteroscedasticity-based approaches, is that we lack the theory-based restrictions needed to generate reliable impulse responses.

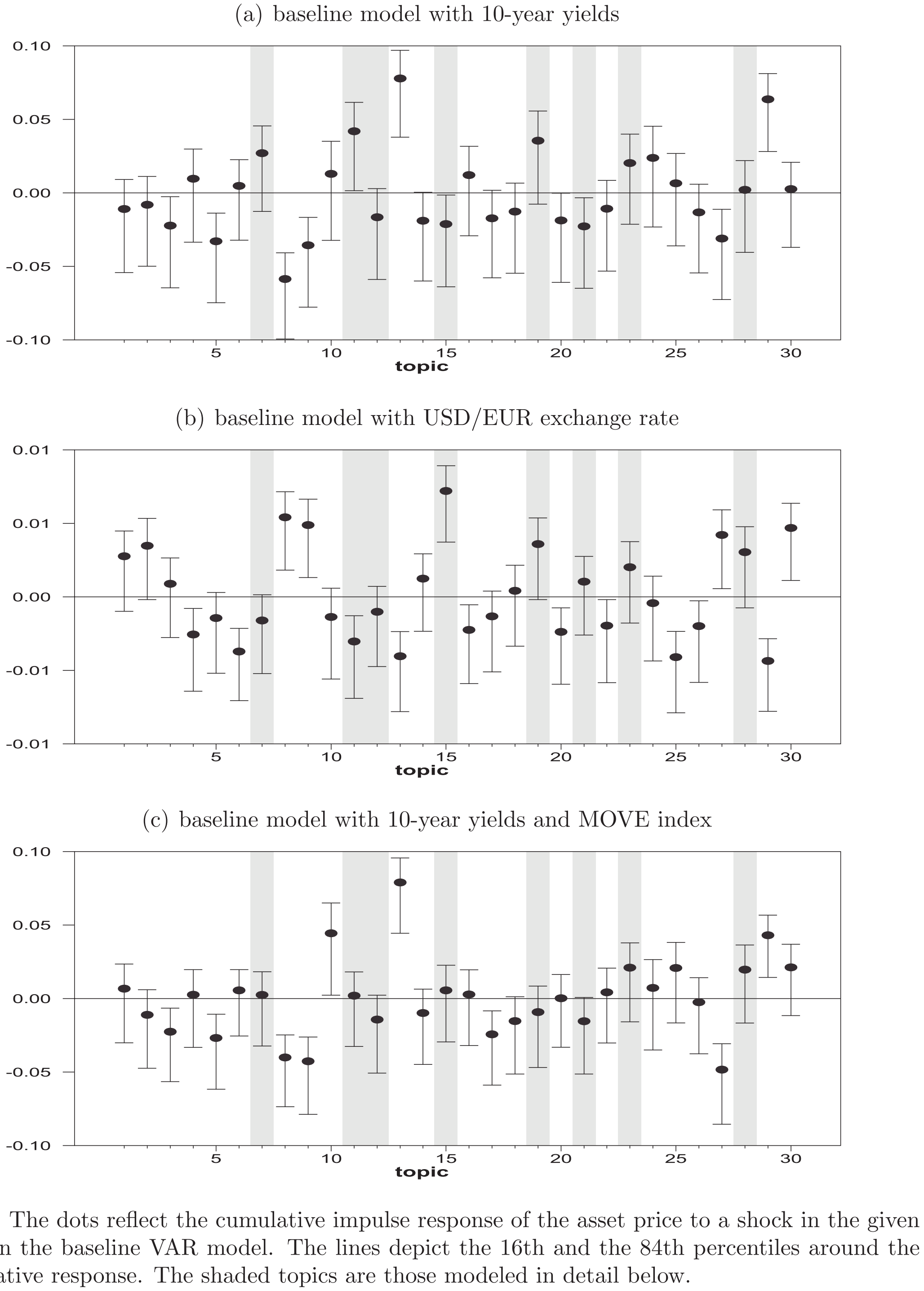


Fig. 2. Cumulative impulse responses. *Notes:* The dots reflect the cumulative impulse response of the asset price to a shock in the given topic in the baseline VAR model. The lines depict the 16th and the 84th percentiles around the cumulative response. The shaded topics are those modeled in detail below..

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 show the responses from a model in which we include the exogenous variables reflecting monetary policy shocks and releases of FOMC

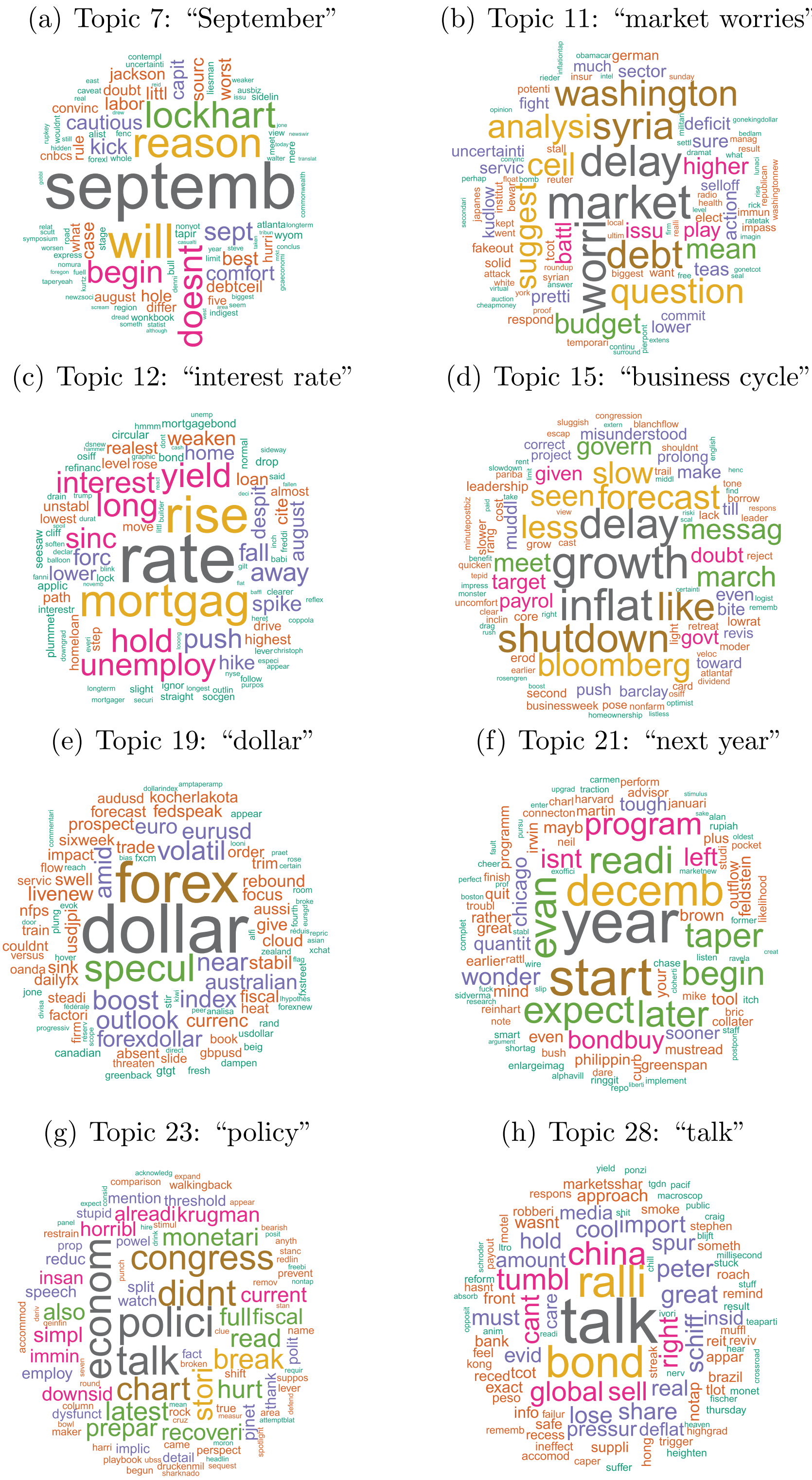


Fig. 3. Selected topics.

Table 1
Selected pairs of topics.

Pair	$Topic_t^i$	$Topic_t^j$
1	Topic 7 "September"	Topic 11 "market worries"
2	Topic 23 "policy"	Topic 15 "business cycle"
3	Topic 12 "interest rate"	Topic 19 "dollar"
4	Topic 21 "next year"	Topic 28 "talk"

Table 2
The five most frequent words in each topic.

	1	2	3	4	5
Topic 1	news	stimulus	busi	financ	market
Topic 2	decis	surpris	risk	market	never
Topic 3	feder	offici	reserv	reuter	delay
Topic 4	will	gross	start	report	elerian
Topic 5	come	think	make	move	first
Topic 6	time	bullard	octob	possibl	reuter
Topic 7	septemb	will	reason	lockhart	doesnt
Topic 8	like	need	soon	look	taper
Topic 9	expect	next	will	announc	just
Topic 10	data	fear	fall	share	asian
Topic 11	market	delay	worri	debt	syria
Topic 12	rate	rise	mortgag	yield	long
Topic 13	bernank	bank	dont	goldman	yellen
Topic 14	back	just	shock	thing	nontap
Topic 15	growth	inflat	delay	shutdown	like
Topic 16	month	purchas	treasuri	asset	billion
Topic 17	gold	price	drop	jump	concern
Topic 18	week	market	wall	street	ahead
Topic 19	dollar	forex	specul	volatil	boost
Topic 20	economi	dudley	still	want	weak
Topic 21	year	start	decemb	evan	readi
Topic 22	investor	chang	increas	post	live
Topic 23	polici	talk	econom	didnt	congress
Topic 24	stock	high	sampp	record	amid
Topic 25	wont	financi	trade	last	asia
Topic 26	bond	keep	plan	money	point
Topic 27	market	talk	gain	earli	world
Topic 28	talk	ralli	bond	china	tumbl
Topic 29	fomc	minut	meet	still	report
Topic 30	decis	good	call	emerg	close

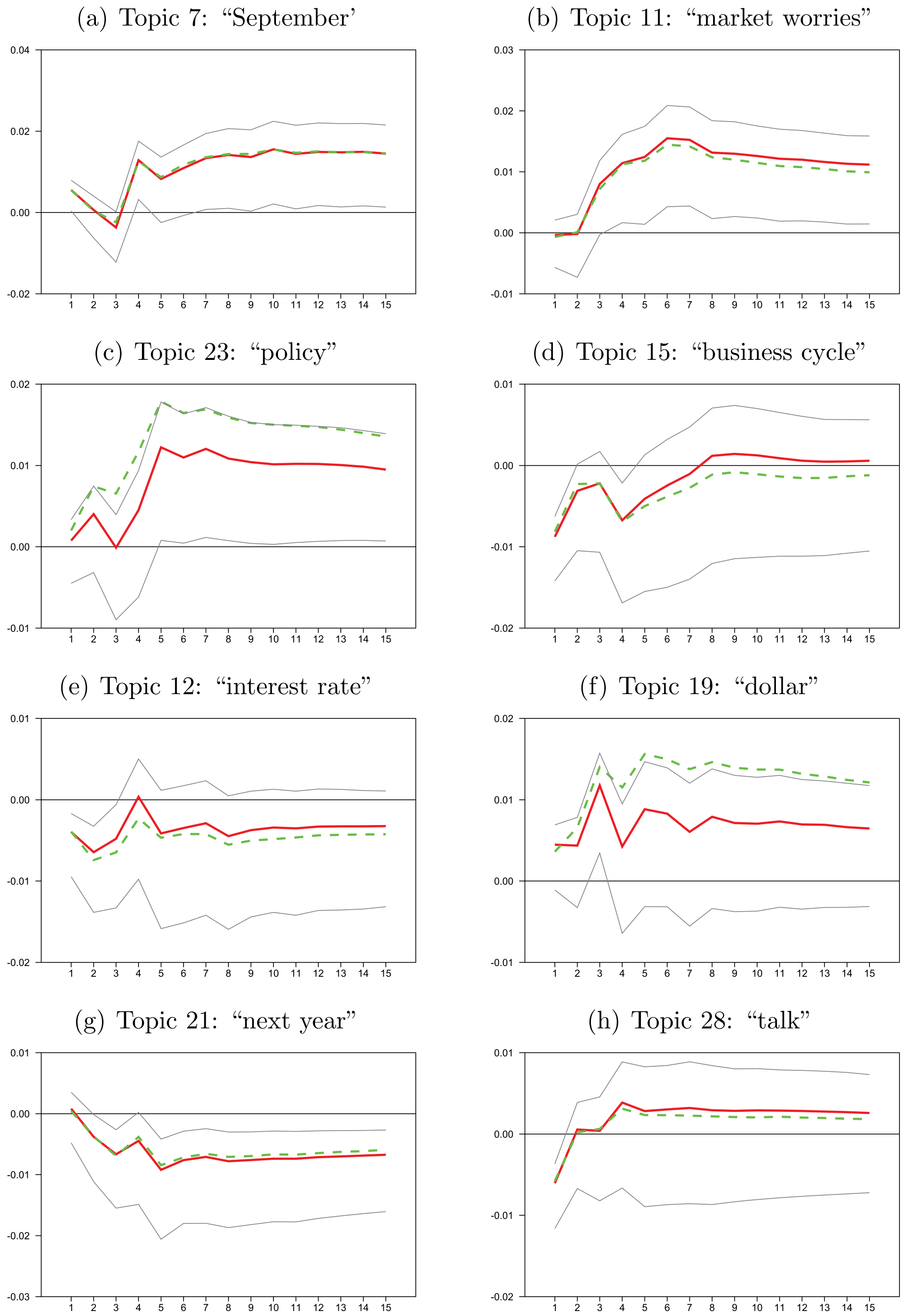
minutes (red line) and a model without this information (green, dotted line). We focus on the responses of asset prices, which are the focus of this paper. In all cases the size of the shock is one standard deviation.

5.1. Baseline results

Fig. 4 reports the responses of U.S. bond yields to shocks to each selected topic stemming from the estimation of the VAR model with each of the topic pairs.

An increase in the importance of topic "September" and topic "policy" leads to a significant increase in bond yields. An increase in the frequency of the topic "September" by one standard deviation raises bond yields by 2 basis points. This is consistent with the notion that a heightened discussion about a tapering decision in September 2013 leading to higher bond yields. A shock to topic "next year", in contrast, reduces bond yields as this reflects the belief that the tapering decision is postponed until 2014. Likewise, more talk on Twitter about the state of the business cycle, see topic 15, weakens the belief about an early tapering and, as a result, leads to a fall in bond yields. 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 ADS measure, the effects are driven by views of Twitter users alone. Likewise, we control for FOMC-related policy shocks and releases of minutes. Thus, the results are not driven by tweets that simply reflect the information contained in official Fed communication on these policy days.

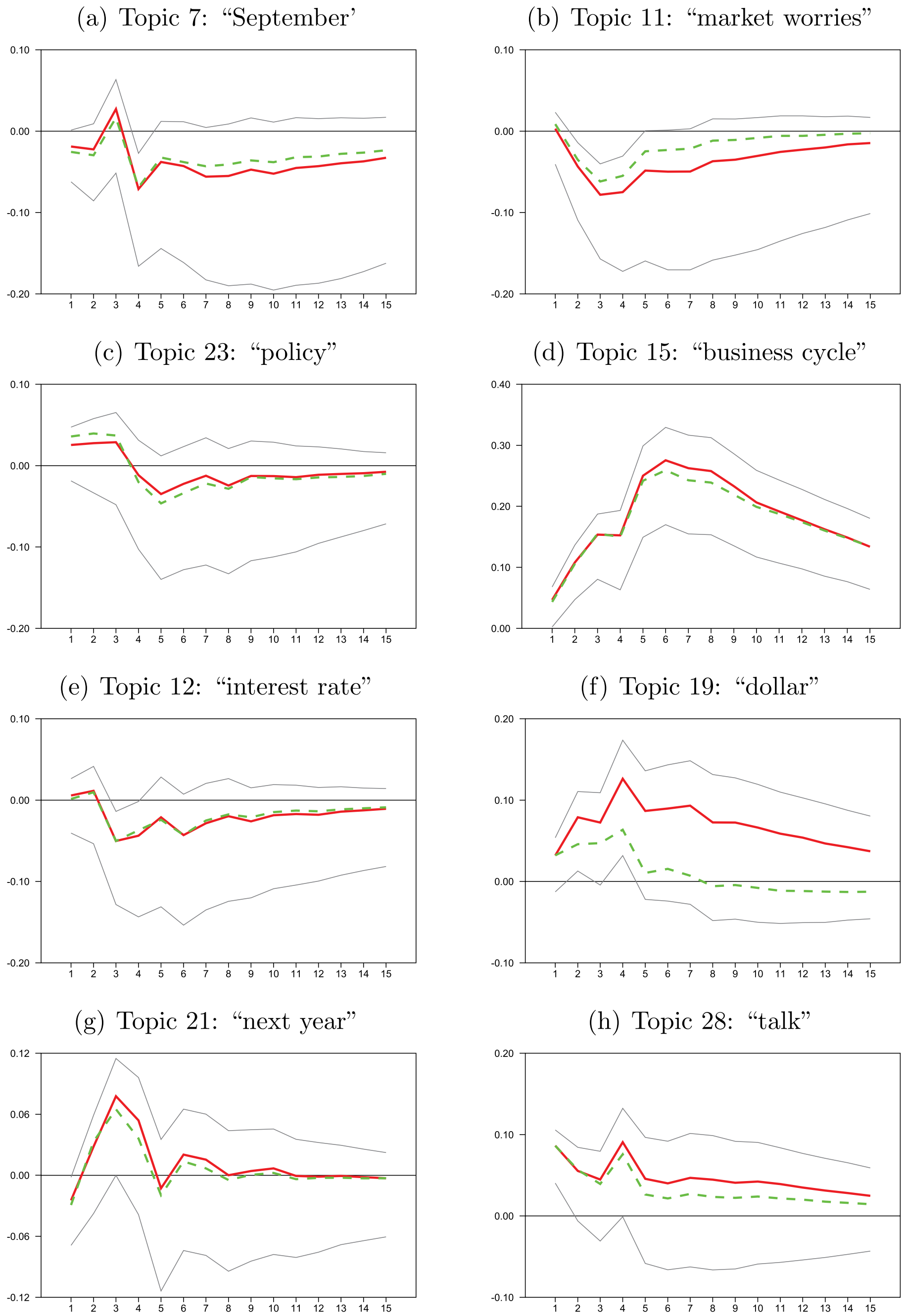
Fig. 5 presents the response of the USD/EUR exchange rate. The dollar appreciates following a shock to topics "September" and "market worries". This is consistent with the positive effect both topics have on bond yields. The more the market discusses the (still weak) business cycle and the rising dollar as reflected in topics 15 and 19, respectively, the more the dollar looses value as an early tapering decision appears less likely.



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

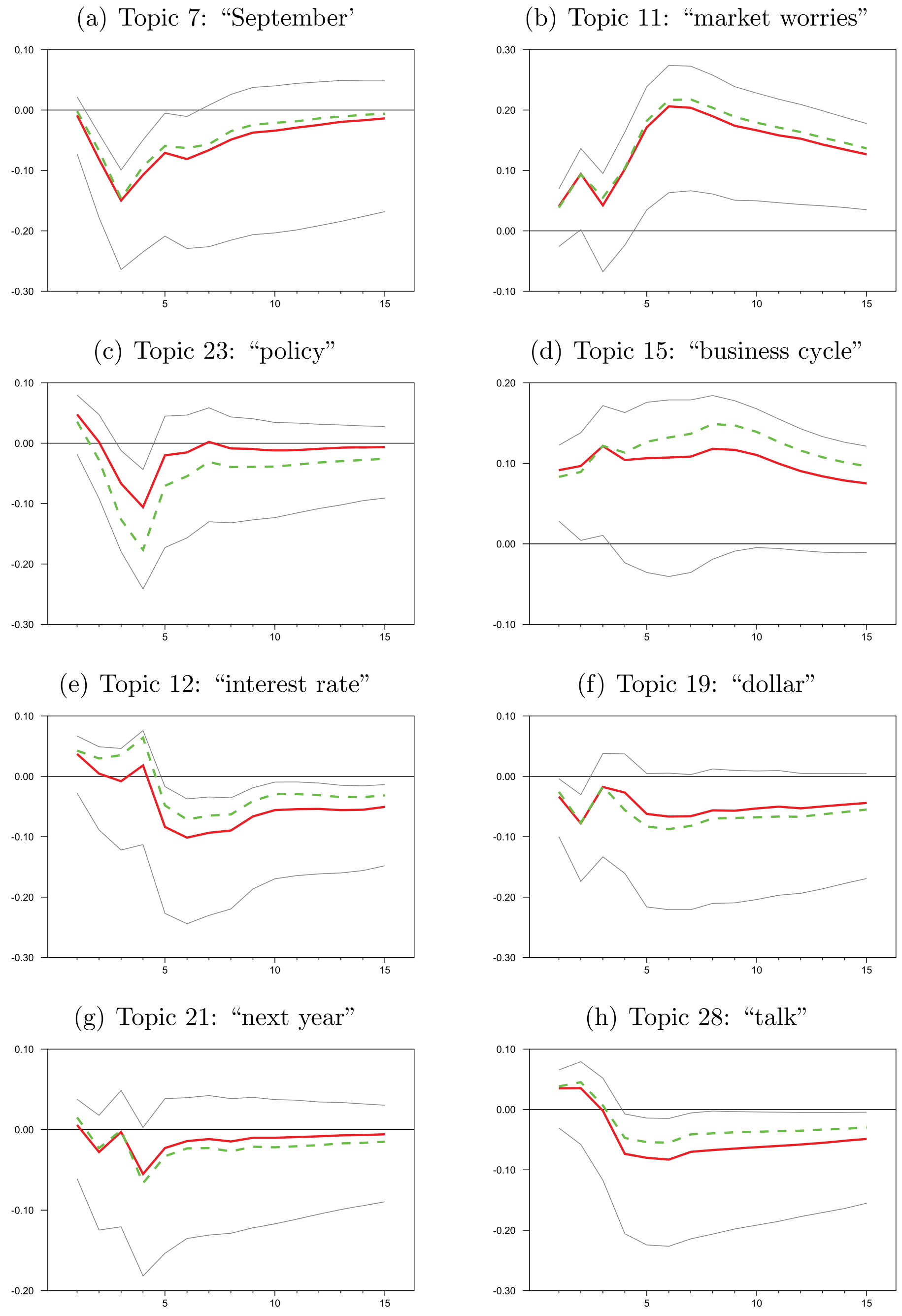
Fig. 4. Impulse responses for baseline model with 10-year bond yield. *Notes:* Yields are ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

The results for the equity market response are reported in Fig. 6. We find that a shock to the topics "September", "policy" and others leads to a significant drop in stock prices. This pattern is consistent with the interest rate responses discussed before. Only a shock to topic 11 "market worries" is puzzling as equity valuations increase.



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

Fig. 5. Impulse responses for baseline model with USD/EUR exchange rate. *Notes:* The USD/EUR exchange rate is ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).



Notes: The S&P500 inidex is ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles.

Fig. 6. Impulse responses for baseline model with stock prices. *Notes:* The S&P500 inidex is ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

5.2. 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* of asset purchases. 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 channel* of unconventional monetary policy.

Based on an estimated term structure model, 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 expectations component are reflecting the *signaling channel* (see, among others, Thornton, 2012; Bauer & 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 expectations component and, as a separate variable, the estimated term premium. All three variables are taken from the model of Adrian et al. (2013).[12] The results for decomposed 10-year yields are shown in Fig. 7. In each figure, the red and green lines are depicting the impulse responses of the term premium in the model with and without FOMC information. The black dotted line is the response of the expectations component of 10-year yields to a topics shock in the model with FOMC information.

We find that the effects shown before were largely driven by the response of the term premium, which exhibits a significant response to the topics shocks. Often the response of the term premium is much stronger that the response of the expectations component. For example, the yield increases as a response to a shock to topic “September” is exclusively due to the increase in the term premium. The same is true for topic “next year”. Hence, the discussion of the timing of the exit from QE works through a reversed *portfolio balance channel* during the taper tantrum episode.

5.3. Controlling for bond market volatility

The tapering period in 2013 has been characterized by high bond market volatility. In order to control for shifts in volatility, we include the Merrill Lynch Option Volatility Estimate (MOVE) Index, which reflects the implied volatility of one-months Treasury options three months ahead. We order this variable, $MOVE_t$, first in the VAR system, i.e. the vector of endogenous variables becomes

$$Y_t = (MOVE_t\, Topic_t^i\, Topic_t^j\, A_t^{US})'. \tag{7}$$

This implies that on a given day innovations to the MOVE index can affect all other variables in the system. Fig. 8 presents the resulting impulse response functions. Most of the dynamic responses remain unchanged. Only two results change: the response to topic “market worries” is no longer significant. It is plausible to assume that changes in market volatility as reflected in the MOVE index already cover some of the information contained in this topic. The other response is that to topic “business cycle”, which is no longer significant.

5.4. 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 the VAR model. To corroborate the robustness of our results, we reverse this ordering, that is, we order asset prices first, i.e. before the topic frequencies, and the topics second and third. If the impulse responses do not change, our baseline results are robust with regard to the identification assumption.

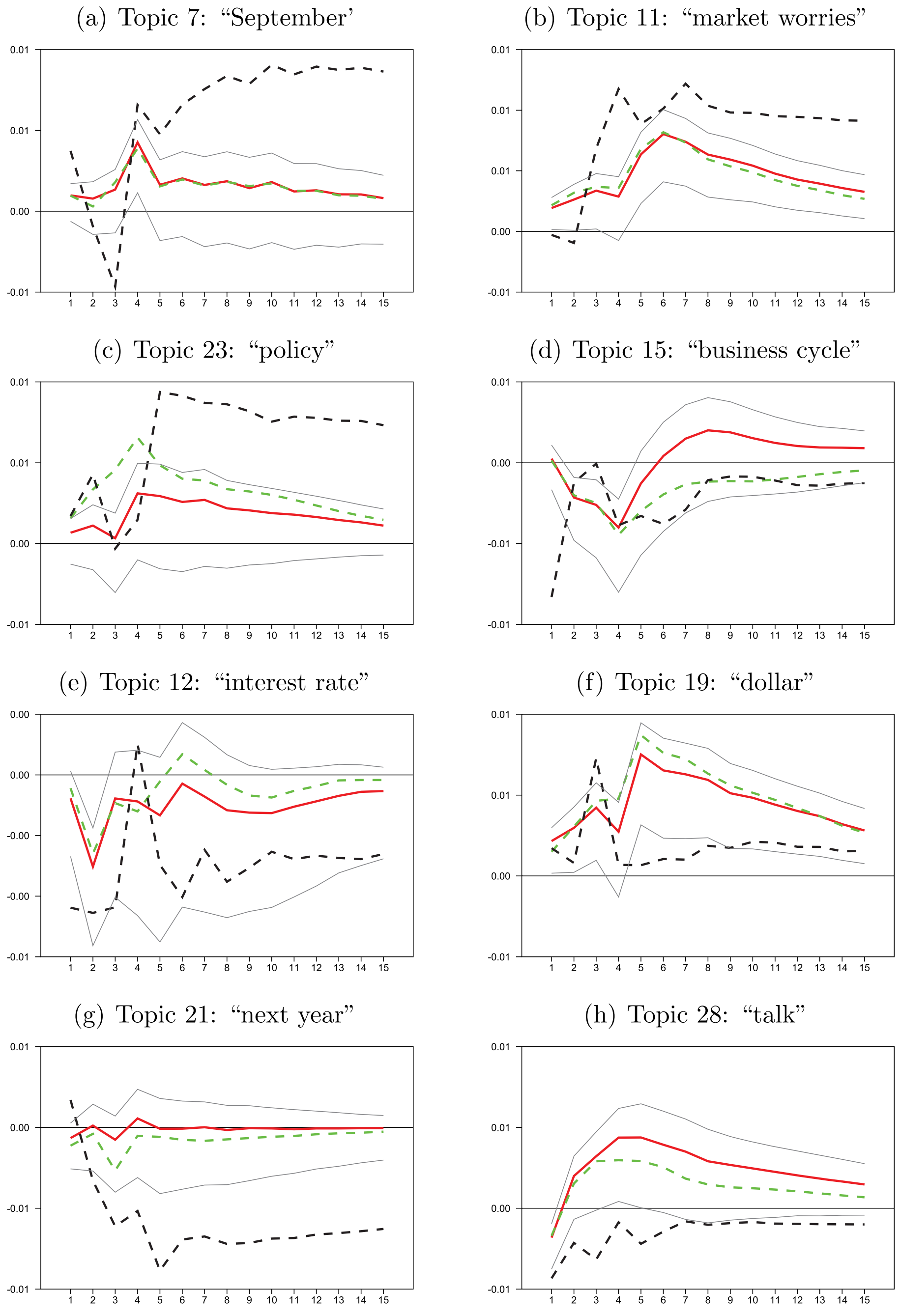
Fig. 9 reports the impulse response of 10-year yields to a shock in each of the selected topics. Overall, our results appear relatively insensitive with regard to the ordering of the variables. The only exceptions are the response to topic “business cycle” and to topic “talk”. While the former is no longer significant, the latter becomes significant once the ordering changes. Overall, we can conclude that using a specific ordering of the variables in our Choleski identification scheme does not drive our findings.

We also run a robustness check of the VAR model which includes a measure of bond market volatility. In contrast to the model used before, we order bond market volatility fourth. Hence, shocks to topic frequencies are allowed to contemporaneously affect bond yields and market volatility. Fig. 10 shows the resulting impulse responses of bond yields to topic-related shocks. Most responses remain unchanged compared to the model used before with volatility ordered first. Again, however, the response to shocks to “market worries” is affected by the change in the model specification. Bond yields increase after a shock to this topic, which is consistent with economic intuition.

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 market participants. This paper dissects the debate about monetary policy for a period with large swings in policy expectations - the “taper tantrum”

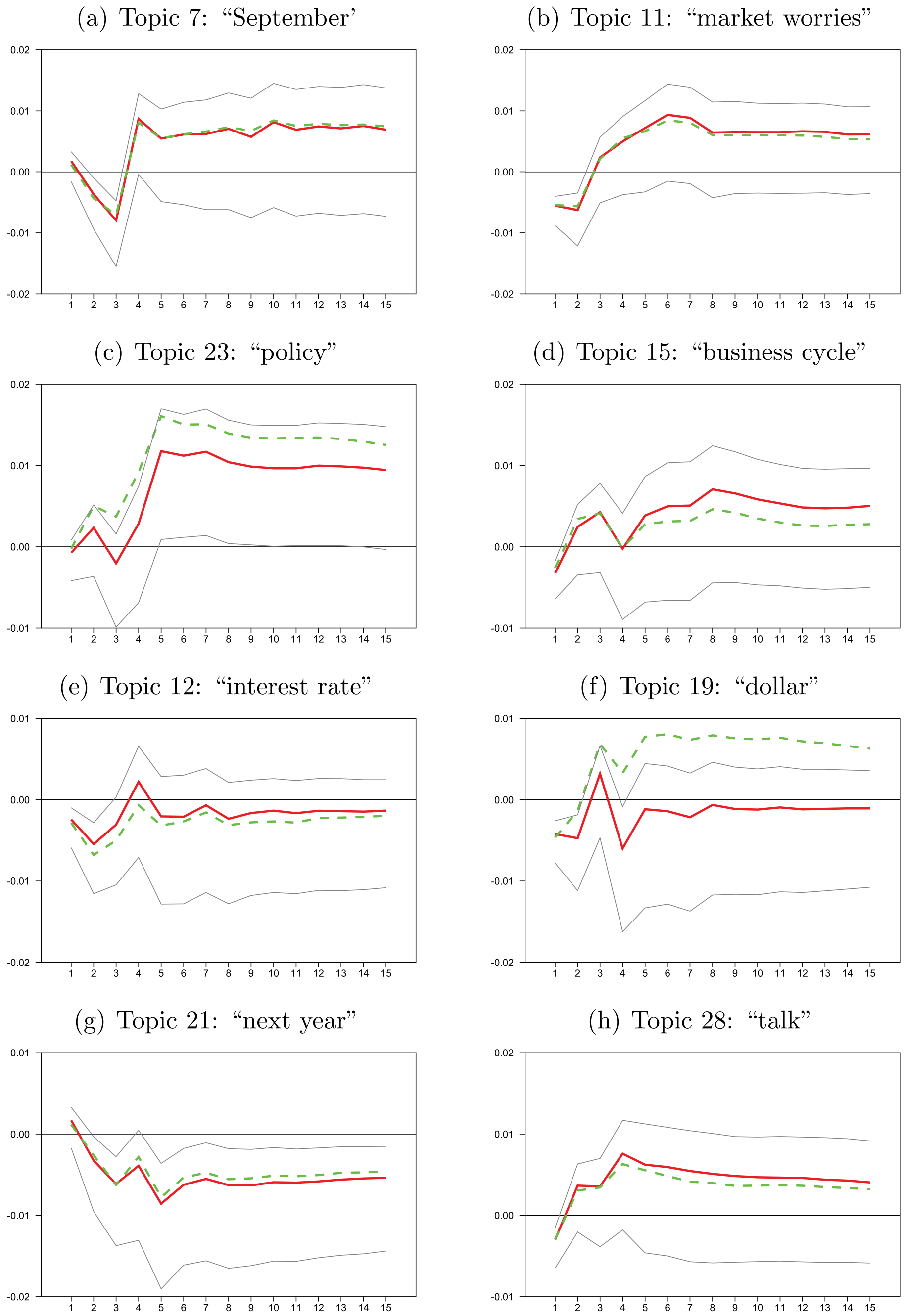
[12] The fitted yields, the estimated expectations component the term premium are available athttps://www.newyorkfed.org/research/data_indicators/term_premia.html.



Notes: Yields are ordered last. The red (solid) and green (dotted) lines are depicting the impulse responses of the term premium in the model with and without FOMC information. The black dotted line is the response of the expectations component of 10-year yields to a topics shock in the model with FOMC information.

Fig. 7. Impulse responses for baseline model with yield decomposition. *Notes:* Yields are ordered last. The red (solid) and green (dotted) lines are depicting the impulse responses of the term premium in the model with and without FOMC information. The black dotted line is the response of the expectations component of 10-year yields to a topics shock in the model with FOMC information. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

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 shocks to selected topic frequencies have significant effects on U.S. bond yields,



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

Fig. 8. Impulse responses for baseline model with 10-year yields and MOVE index. *Notes:* Yields are ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

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,

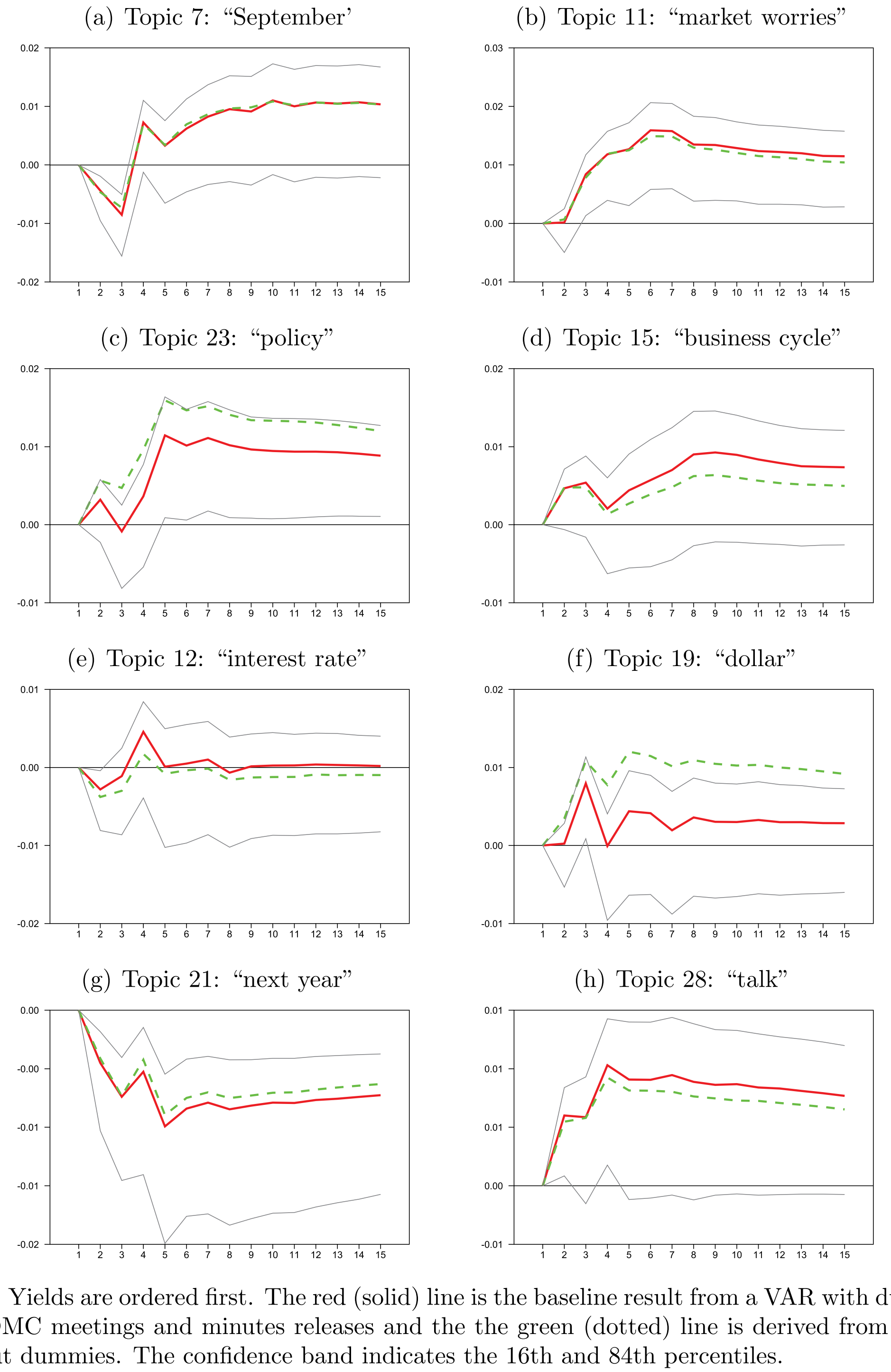
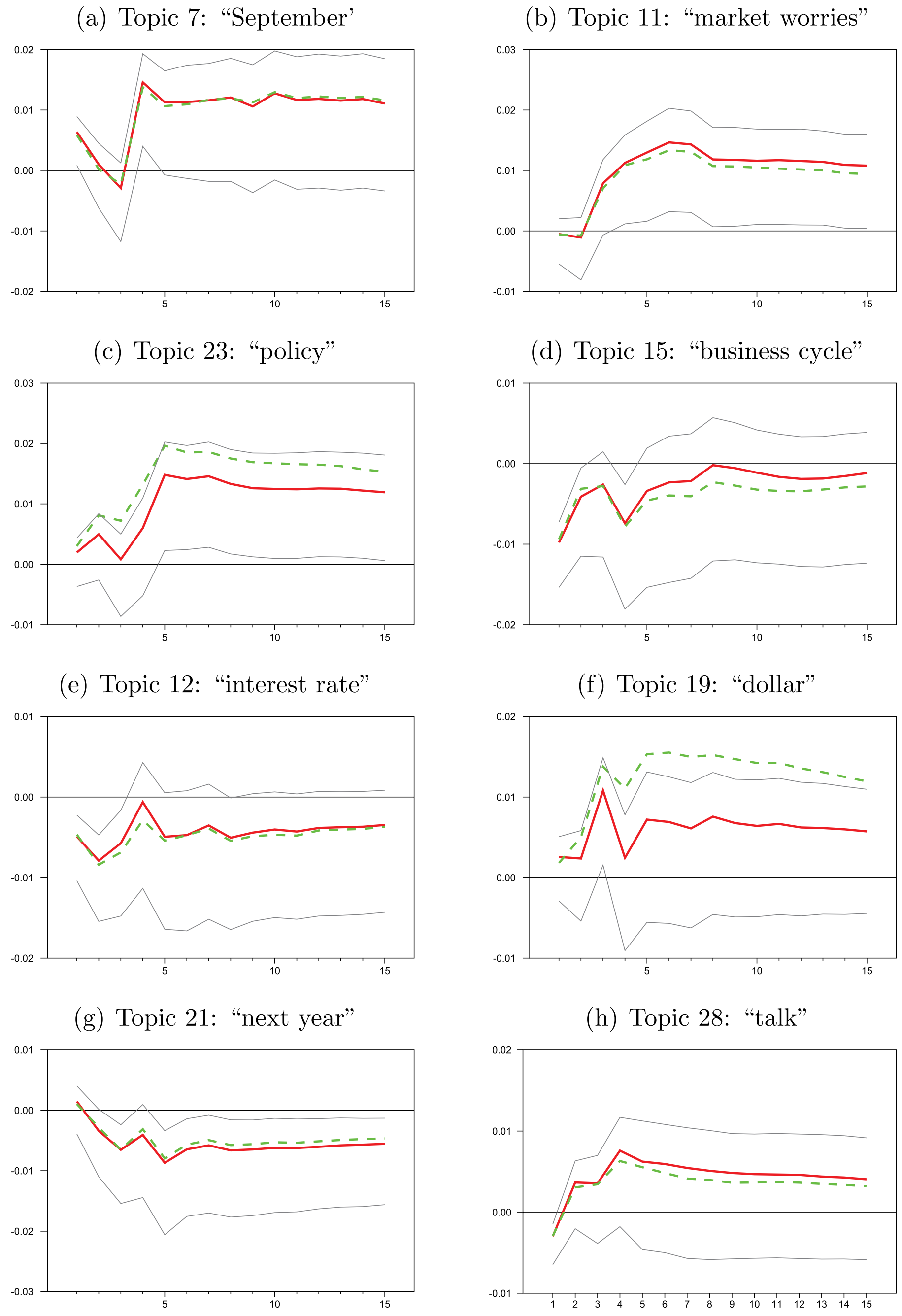


Fig. 9. Impulse responses for baseline model with 10-year bond yield and alternative ordering. *Notes:* Yields are ordered first. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

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.

This analysis leaves several other issues for future research. For example, the cross-section or network dimension of the data can



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

Fig. 10. Impulse responses for baseline model with 10-year yields, MOVE index and an alternative ordering. *Notes:* Volatility are ordered last. The red (solid) line is the baseline result from a VAR with dummies for FOMC meetings and minutes releases and the the green (dotted) line is derived from a VAR without dummies. The confidence band indicates the 16th and 84th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

be used, which provides information about the information flow among different types of investors. The present paper employs daily aggregates of the Twitter exchange. It might also be fruitful to exploit the high-frequency intraday flow of information in the network of Twitter users and the resulting formation of expectations.

References

Adrian, T., Crump, R. K., & Moench, E. (2013). Pricing the term structure with linear regressions. *Journal of Financial Economics, 110*, 110–138.

Aizenman, J., Binici, M., & Hutchison, M. H. (2016). The transmission of Federal Reserve tapering news to emerging financial markets. *International Journal of Central Banking,* 317–356.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business and Economic Statistics, 27*, 417–427.

Azar, P. D., & Lo, A. W. (2016). The wisdom of twitter crowds: Predicting stock market reactions to FOMC meetings via twitter feeds. *The Journal of Portfolio Management, 42*, 123–134.

Bauer, M. D., & Rudebusch, G. D. (2014). The signaling channel of Federal Reserve bond purchases. *International Journal of Central Banking, 10*, 233–289.

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

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *Annals of Applied Statistics, 1*, 17–35.

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

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*, 391–407.

Eichengreen, B., & Gupta, P. (2015). Tapering talk: The impact of expectations of Reduced Federal Reserve security purchases on emerging markets. *Emerging Markets Review, 25*, 1–15.

Goldsmith-Pinkham, P., Hirtle, B., Lucca, D. (2016). Parsing the content of bank supervision. Staff Report No. 770, Federal Reserve Bank of New York.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 228–5235.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software, 40*, 1–30.

Hansen, S., & McMahon, M. (2016). Shocking language: understanding the macroeconomic effects of central bank communication. *Journal of International Economics, 99*(Supplement 1) S114–S113.

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

Hendry, S., Madeley, A. (2010). Text mining and the information content of Bank of Canada communication. *Bank of Canada Working Paper* 2010-31, Bank of Canada.

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

Hofmann, T. (1999). Probabilistic latent semantic indexing. In: Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval, pp. 50–57.

Larsen, V. H., Thorsrud, L. A. (2015). The values of news. *unpublished*, BI Norwegian Business School.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research, 54*, 1187–1230.

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

Lüdering, J., & Winker, P. (2016). Forward or backward looking? The economic discourse and the observed reality. *Jahrbücher für Nationalökonomie und Statistik, 236*(4), 483–515.

Luik, M. -A. Wesselbaum, D. (2016). Central Bank Communication and Social Media #FED. University of Otago (unpublished).

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

Meinusch, A., & Tillmann, P. (2016). Quantitative Easing and tapering uncertainty: Evidence from Twitter. *International Journal of Central Banking* forthcoming.

Nakamura, E., Steinsson J. (2017). High Frequency Identification of Monetary Non-Neutrality: The Information Effect. Quarterly Journal of Economics (forthcoming).

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

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

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

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