AP 2019–02

Chair of Applied Stochastics and Risk Management

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

Faculty of Economic and Social Sciences
Department of Mathematics and Statistics

Working Paper

Forecasting the Equity Premium: Mind the News!

Philipp Adämmer and Rainer Schüssler

May 21, 2019

Working Paper AP 2019–02
May 21, 2019

Forecasting the Equity Premium: Mind the News!

Philipp Adämmer

Helmut Schmidt University
Faculty of Economic and Social Sciences
Department of Mathematics and Statistics
Chair of Applied Stochastics and Risk Management
Holstenhofweg 85, D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2378
E-mail: frahm@hsu-hh.de

Rainer Schüssler

University of Rostock
Faculty of Economic and Social Sciences
Department of Economics
Chair of Applied Economic Research
Ulmenstraße 69, D-18057 Rostock, Germany

URL: www.wiwi.uni-rostock.de/en/econ/empwifo
Phone: +49 (0)381 498-4316
E-mail: rainer.schuessler@uni-rostock.de

Supervised by: Prof. Dr. Gabriel Frahm
Chair of Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

Forecasting the Equity Premium: Mind the News!

Philipp Adämmer*[1] and Rainer A. Schüssler†[2]

[1] *Helmut Schmidt University, Hamburg, Germany*

[2] *University of Rostock, Germany*

May 13, 2019

Abstract

This paper introduces a novel strategy for predicting monthly equity premia based on extracted news from more than 700,000 newspaper articles, published in *The New York Times* and *Washington Post* between 1980 and 2018. We propose a flexible data-adaptive switching approach for mapping a large set of different news-topics into forecasts of aggregate stock returns. The information embedded in our extracted news is not captured by established predictors. Compared to the prevailing historical mean between 1999 and 2018, we find large out-of-sample (OOS) gains with an R^2_{OOS} of 6.52% and sizeable utility gains for a mean-variance investor. The empirical results indicate that geopolitical news are more valuable than economic news and that gains arise in down markets.

Keywords: Topic Modeling, Big Data, Return Predictability, Text as Data

JEL classifications: G11, G12, G17, C53

*E-mail: adaemmer@hsu-hh.de; Phone: +49 (0)40 6541-2378

†Corresponding author. E-mail: rainer.schuessler@uni-rostock.de; Phone: +49 (0)381 498-4316

1 Introduction

Finding variables with strong out-of-sample predictive power for the equity premium is challenging. Empirical evidence shows that the forecasting ability of established economic predictors is, at most, short-lived and thus hard to detect in real time (see, e.g., Timmermann, 2008; Farmer et al., 2018). A crucial reason might be that investors' decisions are driven by a plethora of information that is not captured by standard economic predictors. While economic variables such as dividends, earnings and inflation can be easily included in predictive models, qualitative variables such as news are much harder to quantify and thus more difficult to exploit for equity return forecasts. In contrast to standard economic data, textual data is inherently high-dimensional. For example, a sample of 30-word Twitter messages that only uses the 1,000 most common words in the English language has roughly as many dimensions as there are atoms in the universe (Gentzkow et al., 2017). Due to this characteristic and complexity, statistical machine learning methods are natural candidates to handle qualitative data such as newspaper articles.

This paper introduces a new strategy for computing forecasts of monthly equity premia by synthesizing information from more than 700,000 newspaper articles, published in the *The New York Times* and *Washington Post* between 1980 and 2018. The information is extracted and quantified by a statistical machine learning algorithm, namely the correlated topic model (CTM) by Blei and Lafferty (2007). We mimic the information set of a real-time investor and construct continuous time series, tracking which type of news has been discussed at which point in time. The higher the media coverage of a certain topic, the higher is our assigned importance to that topic in that given point in time. We average these so-called topic proportions on a monthly basis and use them as predictors for the monthly excess returns of the S&P 500 index.

The application of text mining techniques for econometric and financial research has become a promising research field in recent years with various application possibilities. Tetlock (2007) uses the Harvard IV-4 psychosocial dictionary to score the *Wall Street Journal's* "Abreast of the Market" column on the Dow Jones index. A subsequent study is Tetlock et al. (2008). Loughran and McDonald (2011), however, show that the Harvard dictionary can lead to wrong conclusions when applied to financial reports. Building on these findings, Jiang et al. (2019) apply the financial dictionary of Loughran and McDonald (2011) to corporate financial disclosures for constructing a manager sentiment index with strong out-of-sample predictive power for the equity premium. Further studies on equity markets using dictionary-based sentiment methods are, e.g., Engelberg et al. (2012), Garcia (2013) and Glasserman and Mamaysky (2019). Turner et al. (2018) provide a broader perspective on the role of the media in financial markets.

Although empirical studies in asset pricing almost exclusively investigate the tone (positive vs negative) of textual data, theoretical and computational advances have also made it feasible to capture their content. For example, Manela and Moreira (2017) extend options implied volatility (VIX) back to the 19^{th} century by applying support vector regressions (SVR) on news coverage of the *Wall Street Journal*. Their aim is to identify which type of news affects uncertainty and stock market returns. In contrast to support vector regressions, probabilistic topic models such as the seminal latent Dirichlet allocation (LDA) algorithm by Blei et al. (2003) enable to uncover hidden themes within large collections of written documents without having to provide any prior information on the documents themselves. Interestingly, while topic models have already been used in (macro-)economics, applications in empirical asset pricing are yet missing.[1] One exception is the study by Calomiris and Mamaysky (2019) who use a variety of text mining techniques to disentangle the various effects of news on risk and returns in 51 developed as well as emerging equity markets. Among others, they use the Louvain method by Blondel et al. (2008), which is related to LDA, to decompose their text corpus.

One reason for the negligence of topic modeling in empirical asset pricing might be that the obtained topics are often manifold and prone to noise, making it difficult to extract relevant signals for future aggregate returns. Hence, an econometric approach attempting to map news-topics into equity premium forecasts should effectively deal with estimation error and thus protect against over-fitting. As a remedy, we devise a simple, yet flexible econometric strategy. In a first step, we apply univariate regressions to forecast the equity premium. Each predictive regression uses a single predictor and thus generates a one-month ahead forecast. We then aggregate those univariate forecasts by switching between model selection and model averaging in a data-adaptive manner. Model selection predicts the equity premium by relying upon only one forecast. Model averaging computes the simple mean over all forecasts. Switching between model selection and model averaging seeks to lower estimation error, while at the same time allowing to adapt to changing market environments. Based on this econometric approach, our key contribution is to investigate the benefits of topic modeling for predicting the equity premium out of sample. An important advantage of our study is that is not plagued by data snooping concerns: First, our data set has not been investigated previously for equity premium prediction neither for the US, nor for any other country. Second, although we work with many potential predictors, our pseudo real-time forecast aggregation strategy generates only *one* overall forecast at each point in time. Hence, no adjustment for multiple testing is needed.

At a broader level, this paper extends the literature on machine learning techniques for empirical asset pricing. Starting with Hutchinson et al. (1994) who use neural networks to forecast derivatives prices, statistical machine learning methods have been gaining ground, especially in recent years. For example, Rapach et al. (2013) apply regularization techniques to predict global equity returns by using

[1]In (macro-)economics, Larsen and Thorsrud (2019) use LDA to measure the effect of economic news on key economic variables. Thorsrud (2018) uses LDA to construct a daily business cycle index based on a Norwegian business newspaper. He shows that the index has high out-of-sample predictive power for the business cycle. Hansen and McMahon (2016) and Hansen et al. (2017) investigate the mechanisms and the economic effects of central bank communication. Dyer et al. (2017) quantify trends in 10-K disclosures with LDA.

lagged returns of 11 industrialized countries. Machine learning methods have also become popular to explore the cross section of stock returns (see, e.g., Harvey et al., 2016; Giglio and Xiu, 2017; Kelly et al., 2017; Freyberger et al., 2017; Kozak et al., 2017; Gu et al., 2018).

Our empirical results document large economic and statistical out-of-sample (OOS) gains. More precisely, we obtain an R^2_{OOS} of 6.52% and sizeable utility gains for a mean-variance investor. We show that the predictive content of our news-based forecast is not captured by established predictors. On closer inspection, the empirical results reveal that the gains are achieved in down markets and stem both from the discount rate as well as the cash flow channel. In addition, the empirical results imply that geopolitical rather than economic news are more valuable for predicting the equity premium out of sample.

The rest of this paper is organized as follows. Section 2 outlines our methodology. Section 3 reports and discusses the empirical results. Section 4 concludes.

2 Methodology

In this section we outline the process of extracting our news-based forecasts of the equity premium. We first delineate the CTM, then proceed to explain how the news-topics are estimated and describe our strategy of aggregating the univariate forecasts into an overall point forecast.

2.1 Correlated topic model

Probabilistic topic models uncover hidden themes within a large collection of written documents without providing any information other than the texts themselves. Although word order and grammar are important to comprehend a written text, such features can be neglected when the goal is to uncover general themes within document collections. Topic models treat words within a document as *exchangeable*. The assumption that texts are merely "bag of words" allows to represent documents as vectors of word counts. A tool for this representation is the document-term matrix (*dtm*), where each row denotes a document and each column a word (or vice versa). Values in the matrix-cells (ij) count how often the word j occurs in document i (Blei et al., 2003).

Latent Dirichlet allocation (LDA) by Blei et al. (2003) is the most prominent and widely applied topic model ($\sim$ 26,000 citations). It assumes that documents are written by a stochastic process where each text is a mixture of K latent *topics*. Each topic is a discrete probability distribution over the same vocabulary, differing only in the probabilities given to each word. For example, a topic about "unemployment" assigns high probabilities to words such as *job* and *unemployment*, whereas a topic about "central banking" assigns high probabilities to the words *rate* and *inflation*. The topics are shared by all documents, but each document has a different mixture over those topics. LDA is a Bayesian mixture model and owes its name to the fact that the topics and the topic proportions are assumed to be drawn from a Dirichlet distribution.

Using a Dirichlet distribution is computationally convenient as it is conjugate to the multinomial distribution.[2] The drawback, however, is that LDA cannot account for correlations between topics as the values of random Dirichlet-vectors are nearly independent. It is yet more realistic to assume that certain topics co-occur together. For example, an article about central banking is more likely to contain words associated with unemployment than with cooking. To overcome this limitation, we estimate a correlated topic model (CTM), proposed by Blei and Lafferty (2007). Aside from being more realistic, the CTM also tends to perform better in predicting unseen documents (Blei and Lafferty, 2007). This fact is important since we aim to predict the equity premium out of sample. In contrast to LDA, the CTM assumes that the topic proportions are drawn from a logistic normal distribution, allowing to include a covariance matrix for the topic correlations. The CTM has been less frequently used than LDA and with much fewer documents due to estimation difficulties (He et al., 2017).

Figure 1 shows a graphical model, depicting the stochastic process assumed by the CTM. Shaded nodes are observable and unshaded nodes are unobservable random variables. Edges indicate dependence. Plates are replicated variables, namely topics (K), documents (D) and the words within each document (N). The topics $\beta_{1:K}$ are shared by all documents and they are assumed to be drawn from a Dirichlet distribution with prior γ. The topic proportions θ_d are drawn for each document d from a logistic normal distribution with mean vector μ and covariance matrix Σ. $z_{d,n}$ is the so-called *topic assignment*, which is drawn from a multinomial distribution with probability vector θ_d. Each document's word $w_{d,n}$ is thus drawn from topic $z_{d,n}$ whose probabilities are given in $\beta_{z_{d,n}}$. The goal of the algorithm is to uncover the hidden structure that has most likely generated the observed words within each document. As such, it estimates the topic distributions $\beta_{1:K}$ and the topic mixtures θ_d.[3]

2.2 Quantifying news

Our source for newspaper articles is the *LexisNexis Group* who provide access to a variety of legal and journalistic articles. We downloaded each *New York Times* and *Washington Post* article beginning in June 1980 that either contains the character string *econom* in its text or a proportion of economic relevance greater than zero. Economic relevance is given by the search engine of the *LexisNexis Group*.

[2]Conjugacy means that the posterior distribution of the quantities of interest is from the same family as their prior distribution, that is, in this case, the Dirichlet distribution.

[3]The nonconjugacy of the logistic normal to the multinomial distribution impedes the use Markov Chain Monte Carlo (MCMC) methods such as Gibbs-sampling that have been developed for Dirichlet-based mixed membership models (Blei and Lafferty, 2007). To estimate the CTM, Blei and Lafferty (2007) propose a variational Expectation Maximization (EM) algorithm. We, however, use the more efficient and recently proposed *partially collapsed variational EM algorithm* by Roberts et al. (2016), implemented in the stm-package in *R*. Roberts et al. (2018) compare the efficiency of both algorithms and show the superior performance of theirs. We further follow Roberts et al. (2018) by using a "spectral initialization" method which is deterministic – thus making the results reproducible – and globally consistent.

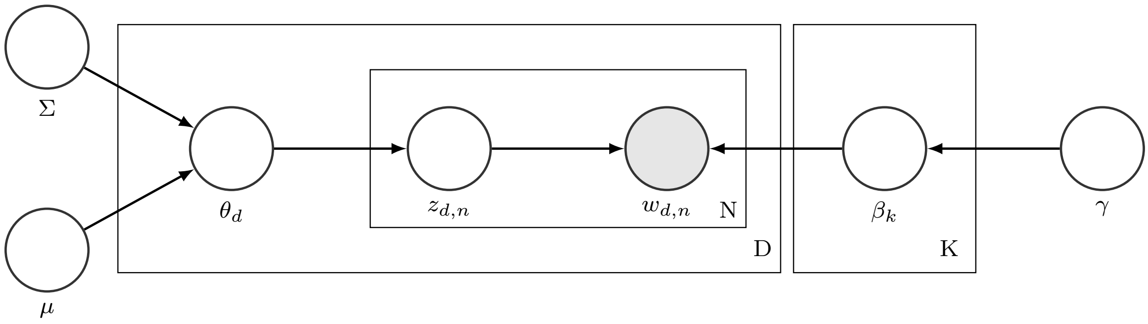

Figure 1: Graphical model of the CTM based on Blei and Lafferty (2007).

We pre-selected articles in this manner to reduce the number of economically irrelevant documents but yet include articles that only marginally touches upon economic issues. In total, we obtained 732,354 articles. As we aim to make out-of-sample forecasts of the equity premium, we have to divide the corpus into a training and a test sample. The training sample spans from 1980:06 to 1995:12 and the test sample from 1996:01 to 2018:12.

We pre-processed all documents according to standard routines, such as removing stopwords (e.g. *the, but, for*), hyphens, apostrophes, numbers, and words that only contain two characters. We also deleted words that occur fewer than 70 times in the corpus, and reduce words to their stems such that, for instance, *economics* becomes *economi*. Documents whose number of words are shorter (longer) than the lower (upper) 2.5% quantiles of all documents were removed to delete very short articles (such as corrections) as well as very long ones (such as dossiers). Some articles were in Spanish which we removed as well. The pre-processing is done for the training sample. The final corpus consists of 694,506 documents, separated into 274,103 documents for the training sample and 420,403 documents for the test sample.

Estimating a CTM requires to choose the number of topics K. It is important to note that there is no "right" answer to the question how many topics to use (Grimmer and Stewart, 2013; Roberts et al., 2018). Following Hoffman et al. (2013); Thorsrud (2018) and Larsen and Thorsrud (2019), we choose K = 100, which strikes a good balance between two aspects: First, as we have to fit unseen documents, the number of topics has to be sufficiently large. Second, given that estimation time should be kept at a reasonable level, the number of topics should not be too large. We report empirical results based on 100 topics. As a robustness check, we replicated all empirical exercises for models with 75, 125 and 150 topics. Those results are similar and omitted for the sake of brevity, but available upon request.

Table 1 shows the five most probable words of eight selected topics, estimated by the CTM with $K = 100$. The results are based on the training sample (1980:06 to 1995:12). The topic numbers themselves convey no meaning and the content of each topic has to be inferred by the researcher. For example, if a newspaper article has a high proportion of topic 11, which includes words such as *oil, price* and *barrel*, we most certainly know that it is about the oil market. Based on the results of the training sample, we estimate and compute monthly averages of topic proportions, beginning with out-of-sample documents in 1996:01. We stress at this point that we mimic the information set of a real-time investor, thus only using topic proportions out-of-sample for the equity premium forecasts. Eight selected monthly topic proportions are shown in Figure 2. The titles for each topic are our choices and based on the word distributions in each topic. We use all monthly topic proportions as predictors for univariate regressions to which we turn now.[4]

Table 1: Word distributions for eight selected topics

Topic 53	**Topic 48**	**Topic 93**	**Topic 33**	**Topic 11**	**Topic 62**	**Topic 68**	**Topic 20**
price	budget	compani	job	oil	iraq	israel	west
futur	cut	busi	work	price	iran	isra	germani
cent	spend	corpor	worker	gas	gulf	palestinian	german
market	deficit	firm	employ	energi	war	arab	east
commod	propos	execut	employe	barrel	iraqi	peace	western

The table shows the five most probable words in descending order for eight selected topics, estimated by the CTM. The last word in topic 53 (*commod*) is originally on position six. We switched the order for illustration purposes.

[4] A potential concern is that language changes over time. One might be inclined to argue that words from newspaper articles between 1980 and 1995 (our training sample) might have been different than the words used between 1995 and 2018 (our test sample). Yet topic models such as the dynamic topic model by Blei and Lafferty (2006) have been developed to analyze text corpora over centuries and not decades. Hence, we do not consider the CTM approach as problematic.

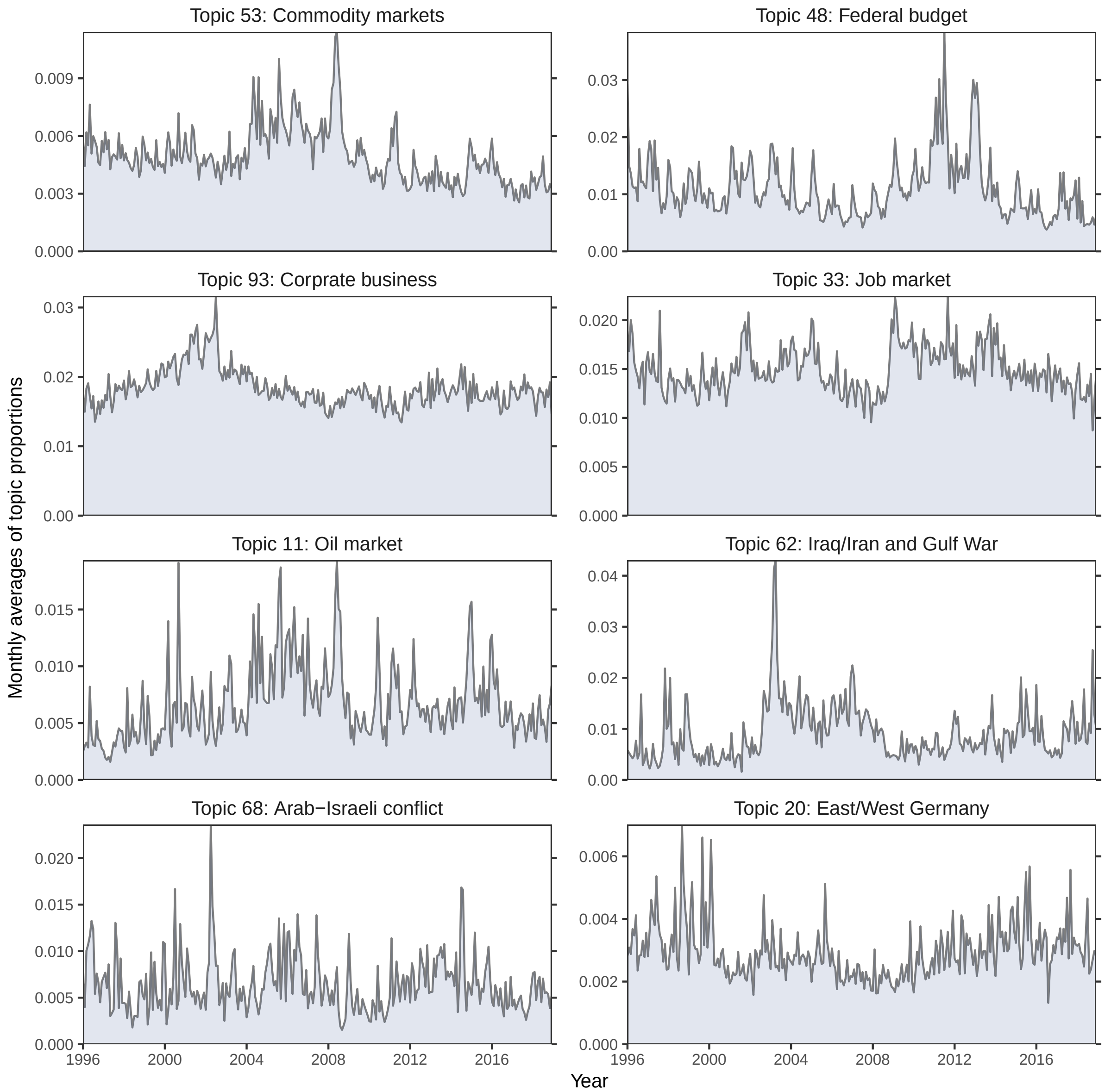

Figure 2: Monthly averages of topic proportions for eight selected topics, estimated by the CTM. The headers are based on the word distribution within each topic.

2.3 Forecasting with topic proportions

Equipped with the estimated monthly topic proportions, the question arises of how to optimally exploit them for the equity premium forecasts.[5] Our goal is to map the 100 estimated predictors into an overall point forecast of the equity premium at a monthly frequency. Our choice of the econometric technique is driven by the following considerations: First, given that we have 100 predictors of which the majority is most likely very noisy or irrelevant, an appropriate econometric method has to effectively limit estimation error. Pre-selecting certain topics on the basis of their alleged relevance based on human judgement might be somewhat arbitrary and might

[5] Henceforth, we use the terms *topic proportions* and *predictors* synonymously.

remove relevant topics that do not appear so at first sight.[6] Second, topics with strong predictive power should be attached higher weights for constructing the aggregate forecast than topics with lower predictive power. Third, the forecasting power of the predictors might vary through time. We thus need an econometric method that quickly adapts to a changing market environment. Fourth, we want a method that is transparent and ensures interpretability of the results. We opt for an econometric technique that simultaneously accommodates all four aspects in a data-adaptive manner.

In a first step, we compute univariate predictive regressions of the equity premium based on one of the 100 predictors. Building on the 100 univariate forecasts, our proposed forecast aggregation method can switch between model averaging and model selection at each point in time in a data-based manner. The rationale for our switching strategy is the purpose to merge the flexibility of model selection with the robustness of model averaging. Although our switching strategy is simple, this paper is, to our knowledge, the first to consider data-adaptive switching between model averaging and model selection.

2.3.1 Predictive regression models

For each of the 100 predictors, we compute univariate regressions for the monthly equity premium:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} denotes the discrete return of the S&P 500 index in excess of the Treasury bill rate, $x_{i,t}$ is the estimated topic proportion, and ε_{t+1} is an error term. We generate out-of-sample forecasts of the equity premium on an expanding window. We divide the entire sample of T observations for r_t and $x_{i,t}$ into an in-sample portion comprising the first m observations and an out-of-sample portion comprising the last q observations. The initial out-of-sample forecast of the equity premium based on predictor $x_{i,t}$ is obtained as

$$\hat{r}_{i,m+1} = \hat{\alpha}_{i,m} + \hat{\beta}_{i,m} x_{i,m}, \tag{2}$$

where $\hat{\alpha}_{i,m}$ and $\hat{\beta}_{i,m}$ are the ordinary least squares (OLS) estimates of α_i and β_i, respectively. These estimates are obtained by regressing $\{r_t\}_{t=2}^{m}$ on a constant and $\{x_{i,t}\}_{t=1}^{m-1}$. The second out-of-sample forecast is obtained by the regression

$$\hat{r}_{i,m+2} = \hat{\alpha}_{i,m+1} + \hat{\beta}_{i,m+1} x_{i,m+1}, \tag{3}$$

where $\hat{\alpha}_{i,m+1}$ and $\hat{\beta}_{i,m+1}$ are obtained by regressing $\{r_t\}_{t=2}^{m+1}$ on a constant and $\{x_{i,t}\}_{t=1}^{m}$. We proceed in this manner until the end of the out-of-sample period, which leaves us with q out-of-sample forecasts. The out-of-sample forecasts mimick the situation of a real-time investor since we only use information that would have been available to the forecaster at any given point in time.

[6] Our aim is to avoid arbitrary and/or discretionary decisions as far as possible in our real-time forecast aggregation strategy. We therefore treated all topics equally for being useful. Alternatively, one could, for instance, attach different prior probabilities to the topics according to their assumed relevance in a Bayesian setup.

2.3.2 Forecast aggregation

We have computed out-of-sample forecasts of the equity premium at each point in time based on the 100 univariate predictive regressions, ($i = 1, ..., N$ and $N = 100$), see Equation 1. We first outline the model averaging/selection strategy and then describe the mechanism for switching between them.

Model averaging

Following, among others, Rapach et al. (2010), we consider a simple version of model averaging. The combination forecast of r_{t+1} made at time t takes the average of all individual forecasts implied by Equation 1:

$$\hat{r}_{t+1}^{CTM^{Avg}} = \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{i,t+1}. \tag{4}$$

We label this aggregation strategy CTM^{Avg}. Each univariate forecasting model has a high shrinkage intensity because all but one predictor are implicitly zero. When taking the average over all univariate predictions, this approach is robust to estimation error.

Model selection

The model selection strategy attaches the entire weight to the best univariate forecast. The *best* individual forecast is the one with the lowest out-of-sample mean squared prediction error (MSPE). To apply this strategy, we first calculate the realized out-of-sample MSPE for each individual forecasting model $\text{MSPE}_{i,t}$ at each period t, where $t > m + 1$. We do so in a recursive manner based on the forecast errors from period $m + 1$ to $t - 1$. The 100 models are then ranked in ascending order according to their realized out-of-sample MSPE. Rank 1 is thus assigned to the forecasting model with the lowest MSPE, and rank 100 to the model with the highest out-of-sample MSPE. As we constantly re-calculate the MSPE, different univariate models may emerge as the *best* model at different points in time. We label this aggregation strategy CTM^{Sel}.

Model switching

Both model averaging and model selection have their own advantages and disadvantages. The model averaging strategy provides a high degree of shrinkage and offers protection against over-fitting since it averages over a large number of signals. The adaptation to changing market environments, however, might be slow. Conversely, the model selection strategy might be faster in adapting to a changing market environment since it relies upon only one single forecast at a given point in time and discards the remaining 99. Yet the increased flexibility makes it more prone to estimation error. Contemplating the pros and cons of both aggregation strategies, we attempt to exploit the advantages of both aggregation strategies while minimizing their drawbacks. To do so, we consider a third aggregation strategy, labeled CTM^{Sw}, which involves switching between model averaging (CTM^{Avg}) and model selection (CTM^{Sel}). The model selection strategy is preferred over the model averaging strategy if a particular predictor shows temporarily stronger forecasting power compared to model averaging. If no single predictor emerges as more powerful, the model averaging strategy is chosen. We will extend our switching strategy to the case that allows anything between model selection and model averaging over all predictors in Section 3.3.1. To formalize the switching strategy, we have to define a criterion that measures which strategy $-$ CTM^{Sel} or CTM^{Avg} $-$ is superior at any given point in time.

As we strive to allow for rapid adaptation to a changing market environment (if empirically warranted), we use the discounted mean squared prediction error (DMSPE) as a criterion to decide whether we pick model selection or model averaging. We compute the DMSPE^c separately for the case in which the aggregate forecast $\hat{r}^c_{s+1}$ is based on model averaging and the case in which the aggregate forecast is based on model selection, indicated by the superscript $c = \{CTM^{Avg}, CTM^{Sel}\}$. For a typical period t, we calculate the discounted mean squared prediction error as

$$DMSPE^c_t|\delta = \sum_{s=m}^{t-1} \delta^{t-1-s}(r_{s+1} - \hat{r}^c_{s+1})^2, \tag{5}$$

where $0 < \delta \leq 1$ denotes an exponential discount factor. When $\delta = 1$, there is no discounting and equal weights are attached to forecast errors of the recent as well as to the more distant past. The lower the value of the discount factor δ, the more emphasis is put on the recent forecast errors compared to the forecast errors of the more distant past. If the value of the discount factor δ was known, the DMSPE could be easily computed. However, as we wish to avoid arbitrary choices of δ and further wish to allow for different values at different points in time, we determine the value of the parameter sequentially by using a grid search from 0.40 to 1.00 with step length 0.01. This wide range of possible values allows to identify appropriate values of the discount factor in an automatic data-driven manner. Notably, we allow for discounting rather than imposing it by also including 1.00 (no discounting) as a possible choice within the grid. The economic intuition for changing values of the discount factor is that different economic regimes might require different degrees of adaptation. We pick the value of the discount factor $\delta \in \{0.40{:}0.01{:}1.00\}$ that would have generated the lowest out-of-sample MSPE until the given point in time t:

$$\underset{\{\delta\}}{min}\colon \; \frac{1}{t-s} \sum_{s=m}^{t-1} (\hat{r}_{s+1} - \hat{r}^{CTM^{Sw}}_{s+1}(\delta))^2, \tag{6}$$

where $\hat{r}^{CTM^{Sw}}_{s+1}(\delta)$ indicates that the forecast for period $s+1$ depends on the value of the discount factor δ. That is, the value of the discount factor δ determines whether $\hat{r}^{CTM^{Sw}}_{s+1}$ represents the forecast implied by CTM^{Avg}_{s+1} or CTM^{Sel}_{s+1}, depending on which aggregation strategy has achieved the lowest DMSPE until the given point in time according to Equation 5.

2.3.3 Timeline

The corpus of newspaper articles covers the period from 1980:06 to 2018:12. We use the sample from 1980:06 to 1995:12 as our training sample. Hence, our entire sample for exploring the out-of-sample predictive power of the topic proportions comprises $T = 276$ monthly observations from 1996:01 to 2018:12. We use the first $m = 18$ observations as in-sample portion and run the first 100 univariate predictive regressions in 1997:06, using the estimated topic proportions as predictors for the equity premium in 1997:07 (see Equation 1). Our out-of-sample period thus comprises $q = 258$ observations, from which we use another $q_0 = 18$ observations (from 1997:07 to

1998:12) to initialize the discount factor δ. Our post-holdout out-of-sample period thus spans exactly 20 years, from 1999:01 to 2018:12. All empirical results are reported for the post-holdout out-of-sample period.

3 Empirical results

In this section we report our empirical results. We first evaluate the forecasts using both statistical and economic measures (Section 3.1 and 3.2). To put our findings into perspective, we then look at the baseline results from different angles (Section 3.3).

3.1 Statistical evaluation

3.1.1 Forecasting accuracy

Welch and Goyal (2008) demonstrate that the simple historical mean model (HM) provides a stringent out-of-sample benchmark in a mean squared error sense. Hence, we take the prevailing historical mean, $\bar{r}_{t+1}$, as a benchmark to evaluate our proposed strategy with respect to point forecasting accuracy. We use the out-of-sample R^2 by Campbell and Thompson (2008), R^2_{OOS}, to compare the strategy forecast $\hat{r}_{t+1}$, to the forecast of the historical mean model. The R^2_{OOS}-statistic is computed as

$$R^2_{OOS} = 1 - \frac{\sum_{s=q_0+1}^{q}(r_{m+s} - \hat{r}_{m+s})^2}{\sum_{s=q_0+1}^{q}(r_{m+s} - \bar{r}_{m+s})^2}. \tag{7}$$

The R^2_{OOS}-statistic measures the reduction in the mean squared prediction error of a forecasting strategy relative to the historical mean forecast. If $R^2_{OOS} > 0$, the strategy forecast outperforms the historical average forecast in a mean squared error sense. To assess whether the proposed forecasting strategy has a significantly lower mean squared prediction error than the historical mean benchmark, we test the null hypothesis $R_{OOS} \leq 0$ against the alternative $R_{OOS} > 0$, using the test by Clark and West (2007).

Calculated over the entire post-holdout out-of-sample period from 1999:01 to 2018:12 (240 months), R^2_{OOS} is 6.52% with an associated p-value of 0.003, based on the test by Clark and West (2007). Interestingly, this R^2_{OOS}-statistic is much greater than the predictive power of established macroeconomic and financial variables (see, e.g., Welch and Goyal, 2008). While the substantially and significantly positive R^2_{OOS}-statistic indicates superior performance of CTM^{Sw} relative to the simple historical mean model, it is of interest how the forecasting gains have accrued over time. To gain insight on this question, we compute the cumulative squared prediction errors of the prevailing mean minus the cumulative squared prediction error of the forecast aggregation strategies implied by CTM^{Sw}, CTM^{Avg} and CTM^{Sel}. Figure 3 shows these cumulative differences in squared prediction errors. Four important findings emerge. First, most gains in terms of forecast accuracy were achieved in periods when model selection was picked. Second, switching between model selection and model averaging occurs quite frequently. This happens as there is data support for emphasizing the recent forecasting performance with low values of the discount factor δ, fluctuating between 0.77 and 0.86 over the post-holdout out-of-sample period (see Figure 4). Third, out-performance against the historical mean benchmark was particularly strong during

NBER-dated recessions, although, with declines towards their ends. Fourth, CTM^{Sw} picked up strong periods of CTM^{Sel} and switched to CTM^{Avg} in periods when CTM^{Sel} did worse than the historical mean model. In those periods, CTM^{Sw} benefited from the more stable forecasts implied by CTM^{Avg}.

To quantify the added value of the flexibility of the proposed CTM^{Sw} strategy, we compute the R^2_{OOS}-statistic separately for each of the forecast aggregation strategies and decompose the MSPE into the squared bias and forecast variance to elucidate where the forecasting gains stem from and to explore potential bias-variance tradeoffs. Table 2 summarizes a decomposition into squared bias and variance:

$$MSPE = (\bar{\hat{\varepsilon}})^2 + Var(\hat{\varepsilon}), \tag{8}$$

where $\hat{\varepsilon}$ denotes the forecast error, $(\bar{\hat{\varepsilon}})^2$ is the squared bias and $Var(\hat{\varepsilon})$ is the forecast error variance. CTM^{Sw} is clearly ahead in terms of R^2_{OOS} with 6.52% compared to CTM^{Sel} and CTM^{Avg} with an R^2_{OOS} of 2.42% and 0.79%, respectively. In comparison to the prevailing historical mean model (HM), all considered forecast aggregation strategies have a considerably lower squared bias and also the forecast variances of the aggregation strategies are slightly lower than in the historical mean model. CTM^{Sw} exhibits the lowest forecast variance, while CTM^{Sel} has the lowest squared bias.

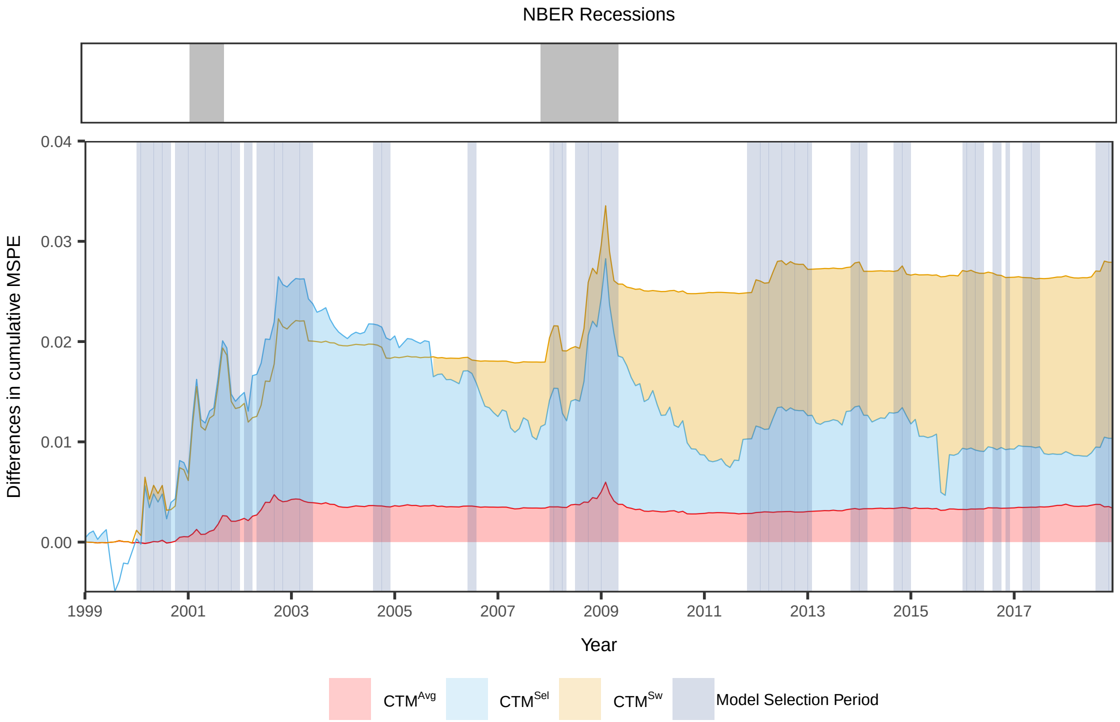

Figure 3: The figure shows the cumulative differences in MSPE between the simple historical mean forecast and CTM^{Avg}, CTM^{Sel} and CTM^{Sw}, respectively. The blue shaded areas correspond to periods where model selection (CTM^{Sel}) was selected. The gray shaded areas in the top panel correspond to NBER-dated recessions.

Table 2: Forecasting strategies and bias-variance decomposition

Strategy	$(\bar{\hat{\varepsilon}})^2$	$Var\,(\hat{\varepsilon})$	R^2_{OOS}
CTM^{Sw}	$2.4131e^{-7}$	$1.6727e^{-3}$	6.52^{***}
CTM^{Avg}	$2.4265e^{-6}$	$1.7731e^{-3}$	0.79^{*}
CTM^{Sel}	$2.0300e^{-6}$	$1.7444e^{-3}$	2.42^{***}
HM	$4.8673e^{-6}$	$1.7848e^{-3}$	0

The table reports the R^2_{OOS}-statistics for the different forecasting strategies and the decomposition of the mean squared prediction error into the squared bias, $(\bar{\hat{\varepsilon}})^2$, and the forecast error variance, $Var\,(\hat{\varepsilon})$. *** indicates that the result is significant at the 1% level, * at the 10% level.

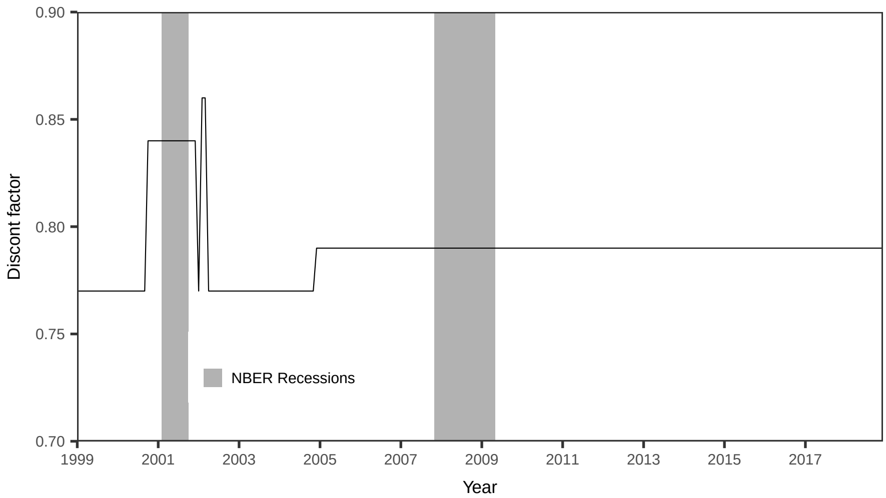

Figure 4: The plot shows the estimated value of the discount factor δ at each point in time.

To elucidate how sensitive the R^2_{OOS}-statistic is with respect to the choice of δ, we show the R^2_{OOS}-statistic as a function of the (fixed) value of the discount factor δ. We choose the range from 0.40 to 1.00 for values of δ. Figure 5 reveals that certain fixed values of δ would have led to even higher R^2_{OOS}-statistics than the data-adaptive choice of δ in our out-of sample setting. The highest R^2_{OOS}-statistics are achieved for values of δ between 0.6 and 0.85. Values near one lead to negative R^2_{OOS}-statistics, pointing to the importance of fast model switching. This finding lines up with Beckmann et al. (2018), who also find strong evidence for fast model switching in their forecasting strategy for exchange rate returns.

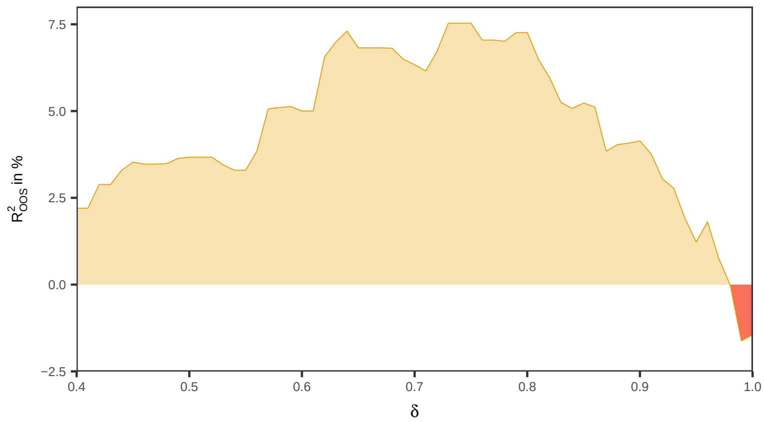

Figure 5: The figure depicts the sensitivity of the R^2_{OOS} as a function of the fixed value of the discount parameter δ.

3.1.2 Forecast encompassing test

We conduct forecast encompassing tests to compare the predictive information embedded in CTM^{Sw} to the individual forecasts based on one of 15 established predictors. We attempt to find a weight λ that yields an optimal convex combination between the individual forecast and the forecast implied by CTM^{Sw},

$$\hat{r}^{*}_{t+1} = (1-\lambda)\,\hat{r}^{i}_{t+1} + \lambda \hat{r}^{CTM^{Sw}}_{t+1}, \tag{9}$$

where $\hat{r}^{i}_{t+1}$ is the predictive regression forecast based on one of the established predictors and $\hat{r}^{CTM^{Sw}}_{t+1}$ is the predictive regression forecast implied by CTM^{Sw}, and $0 \leq \lambda \leq 1$. If the optimal forecast combination given by Equation 9 sets $\lambda = 0$, the forecast based on CTM^{Sw} is excluded. In this case CTM^{Sw} contains no additional information for predicting aggregate stock returns beyond the information already contained in the considered established predictor. Otherwise, if $\lambda > 0$, the predictive regression forecast based on the considered established predictor does not encompass the forecast implied by CTM^{Sw}. In this case there is additional information contained in the forecast implied by CTM^{Sw} beyond the forecast based on the established predictor.

The established predictors we use are those considered in Welch and Goyal (2008): dividend yield (DY), earnings-price-ratio (EP), dividend payout ratio (dpayr), stock variance (SVAR), book-to-market ratio (BMR), net equity expansion (NTIS), treasury bill rate (TBL), long-term government bond yield (LTY), return on long-term government (LTR), default return spread (DFR), default yield spread (DFY), inflation (INF)[7], term spread (TMS) and the dividend-price ratio (DP).[8] In addition, we consider the short interest rate (SII) by Rapach et al. (2016). Table 3 reports the estimates of λ in Equation 9, corresponding to each of the established predictors. We apply the test by Harvey et al. (1998) to indicate whether the estimate is significantly different from 0. The message from Table 3 is unambiguous: with all $\hat{\lambda}$ estimates being large and significantly different from 0, none of the

[7] Inflation is lagged by one additional month.

[8] All variables are collected from Amit Goyal's webpage: `http://www.hec.unil.ch/agoyal/` and constructed as in Welch and Goyal (2008). We initialize all predictive regressions at the earliest possible date (in 1927:01 (or when the predictive variable is available for the first time) and calculate the forecasts recursively until 2017:12.

forecasts based on the popular predictors encompasses the forecast implied by CTM^{Sw}. Except for DY, DP and SII, the entire weight is attached to the forecast implied by CTM^{Sw} in the combination forecast. The results of the encompassing tests clearly highlight the superior forecasting ability of CTM^{Sw} compared to the large set of established predictors.

Table 3: Encompassing tests

Predictor	$\hat{\lambda}$
DY	0.85^{**}
EP	1.00^{***}
DPAYR	1.00^{***}
SVAR	1.00^{***}
BMR	1.00^{***}
NTIS	1.00^{***}
TBL	1.00^{***}
LTY	1.00^{***}
LTR	1.00^{***}
DFR	1.00^{***}
DFY	1.00^{***}
INF	1.00^{***}
TMS	1.00^{***}
DP	0.89^{***}
SII	0.88^{***}

The table reports the estimates of the encompassing tests with all considered established predictors. *** indicates that the result is significant at the 1% level, ** at the 5% level.

3.2 Economic evaluation

3.2.1 Economic utility measures

We calculate economic utility gains from the perspective of a real-time mean-variance investor who allocates her wealth monthly between stocks and the Treasury bill rate. At the end of each month, the investor computes the optimal share to allocate to equities:

$$w_t^* = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2}\right), \tag{10}$$

where $\hat{r}_{t+1}$ and $\hat{\sigma}_{t+1}^2$ denote the expected return and the expected variance. Parameter γ denotes the relative risk aversion. We restrict the portfolio weights to the range from -0.5 and 1.5, thus allowing for short sales and a moderate degree of leverage. We report our empirical results for $\gamma = 3$. The certainty-equivalent return (CEQ) based on the allocation of Equation 10 is

$$\widehat{CEQ} = \hat{\mu} - \gamma\hat{\sigma}^2, \tag{11}$$

where $\widehat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of the excess return over the post-holdout out-of-sample period. We multiply $\widehat{CEQ}$ by $1,200$ to express it in average annualized percentage return. Our second measure of economic utility is the out-of-sample Sharpe ratio (SR), computed as

$$\widehat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}. \tag{12}$$

We multiply this quantity by $\sqrt{12}$ to obtain the annualized Sharpe ratio. Following Johannes et al. (2014) as well as Wachter and Warusawitharana (2009), we assess the statistical significance of $\widehat{CEQ}$ and $\widehat{SR}$ by simulated distributions under the null of no predictability. In the null model, there is no predictability, meaning that we have a constant mean and variance of the equity premium. To generate the null distributions, we simply shuffle the realizations of the equity premium forecasts, thereby ensuring that the simulated returns match the mean and the variance of the actual returns over the post-holdout out-of-sample period. Subsequently, given the returns simulated from the null model, we sequentially estimate each of our considered forecasting strategies, using the same estimation rules which we used on the real data. We repeat this exercise 1,000 times for each forecasting strategy to obtain a distribution of $\widehat{CEQ}$ and $\widehat{SR}$. Equipped with the null distributions, we can evaluate whether the statistics observed from the real-world data are significantly higher than those generated from the null model. We report economic results net of transaction costs ($\widehat{CEQ}^{TC}$ and $\widehat{SR}^{TC}$), assuming proportional transaction costs equal to 50 basis points per transaction (see, e.g., Balduzzi and Lynch, 1999; Neely et al., 2014; Li and Tsiakas, 2017).

We evaluate our proposed CTM^{Sw} strategy to compute the forecasts of the excess returns and the prevailing historical mean (HM) as a benchmark strategy. As strategies for computing the expected variance, we calculate the variance on an expanding window (*Constant*) and also use a time-varying version based on the realized variance (RV). For the latter, we simply sum over the squared daily returns in the last 21 trading sessions. Beside strategies that differ with regard to the computation of expected returns and the expected variance, we also consider a "buy-and-hold" investor who is 100% long in equities and who does not rely upon any forecasts.

3.2.2 Realized economic utility

Table 4 reports our results on economic utility. The most important finding is that economic utility is highest when expected returns are based on CTM^{Sw}. CTM^{Sw}-*RV* achieves an annualized certainty equivalent return of 9.02% (6.50%) before (after) transaction costs and an annualized Sharpe ratio of 0.94 (0.73) before (after) transaction costs. Both certainty equivalent returns and Sharpe ratios are significantly higher compared to the portfolio results based on the *HM-Constant* model. A risk-averse mean-variance investor with risk aversion $\gamma = 3$ would thus be willing to pay an annualized performance fee of 391 basis points before transaction costs and 297 basis points after transaction costs to switch from the historical mean model with realized volatility *HM-RV* to the CTM^{Sw}-*RV* strategy. Economic gains are similar for both versions of expected variance in case of CTM^{Sw}. Yet taking into account time variation in volatility pays off in case of the prevailing mean. The reason is that CTM^{Sw} relies upon precise forecasts in episodes of negative excess returns (see also Section 3.3.5), while *RV*-based forecasts of the variance are generally higher in down markets than the expected variance computed over an expanding window. Higher volatility forecasts decrease the share invested in equities. Given, however, the precise forecasts of expected excess returns in CTM^{Sw} for down markets, a decreased share invested in equities is not necessarily beneficial. In contrast, the *HM* model strongly benefits from relying upon time-varying variance forecasts since forecasts of excess returns are imprecise for down markets and a lower share in equities is thus advantageous. Figure 6 compares the evolution of the wealth of two investors after transaction costs, where one investor relies upon the CTM^{Sw}-*RV* model and the other upon the *HM-RV* model. The initial endowment of both investors is one dollar each.

Table 4: Evaluation of economic utility

$\hat{r}_t$	$\hat{\sigma}_t^2$	Weight restriction	$\widehat{CEQ}$	$\widehat{CEQ}^{TC}$	$\widehat{SR}$	$\widehat{SR}^{TC}$
CTM^{Sw}	RV	$[-0.5; 1.5]$	9.02^{***}	6.50^{***}	0.94^{***}	0.73^{***}
CTM^{Sw}	Constant	$[-0.5; 1.5]$	8.36^{***}	6.24^{***}	0.91^{***}	0.72^{***}
HM	RV	$[-0.5; 1.5]$	5.11	3.53	0.59	0.47
HM	Constant	$[-0.5; 1.5]$	1.45	1.22	0.31	0.29
"Buy-and-hold"		1	3.47	3.45	0.46	0.46

The table summarizes the results with respect to economic utility gains.*** indicates that the result is significant at the 1% level based on $1,000$ simulated data sets with constant mean and volatility.

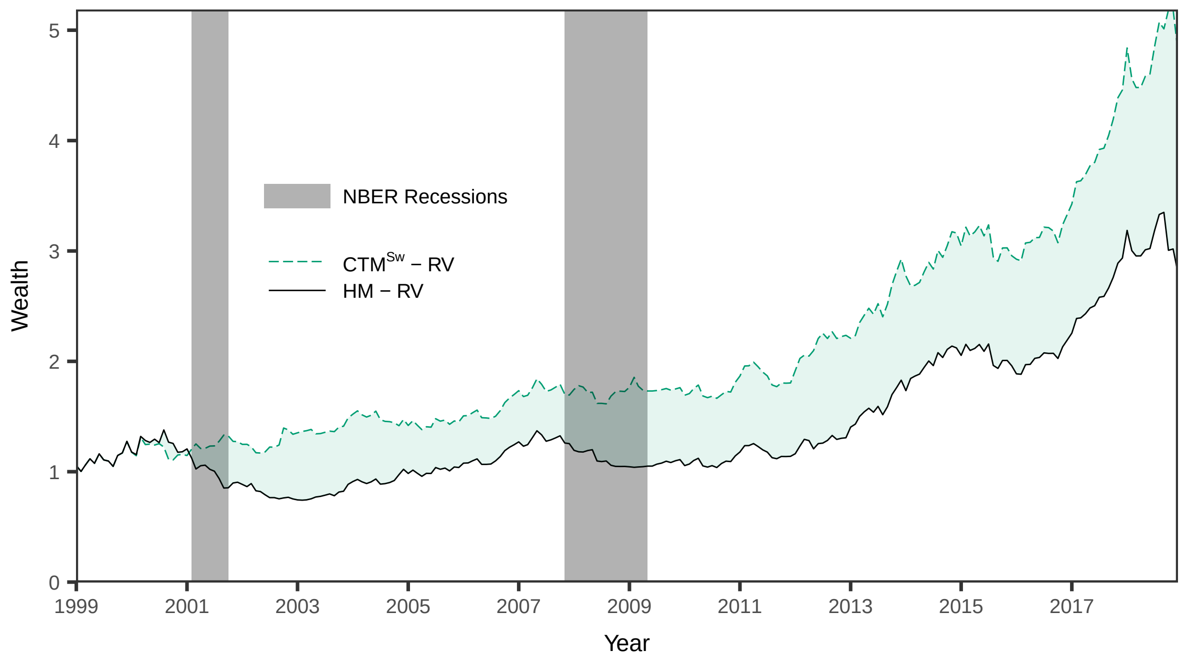

Figure 6: The figure compares the evolution of wealth when asset allocation decisions are based on CTM^{Sw}-*RV* vs *HM-RV*, after deducting transaction costs.

3.3 Additional results and interpretation

3.3.1 Generalized forecast aggregation

In our baseline strategy (CTM^{Sw}), we have allowed for switching between an average over all 100 individual forecasts (CTM^{Avg}) and selecting the univariate forecasting model with the lowest MSPE until a given point in time (CTM^{Sel}). Here, we explore a generalized version of the switching strategy, allowing for an average over any subset of forecast models. That is, rather than the extremes of using only one forecast or averaging over all models, we allow for any value in between. At certain points in time, the switching mechanism may opt for an aggregation strategy that computes the simple average of the, say 20, best forecasting models in terms of MSPE. The ranking procedure and the discounted mean squared prediction error remain the same as in our baseline setting.

Figure 7 shows how many individual forecasts were averaged to compute the aggregate forecast at each point in time. We observe many periods where only the single best model was selected, although the optimal number of included individual forecasts differs over time and changes quite rapidly.

Beside the number of included individual forecasts in the aggregate forecast, it is of interest *which* predictors were included. Figure 8 visualizes the inclusion frequencies for all individual topics. Each topic can be included up to 240 times, i.e., the number of observations in our post-holdout out-of-sample period. The forecasts based on the topic proportions of topic 20 ("East/West Germany") were selected in each period. We have the same situation in our baseline CTM^{Sw} strategy: topic 20 was included in each period. Individual forecasts that were frequently part of the aggregate forecast are based on topic 11 ("Oil market") and topic 62 ("Iraq/Iran & Gulf War").

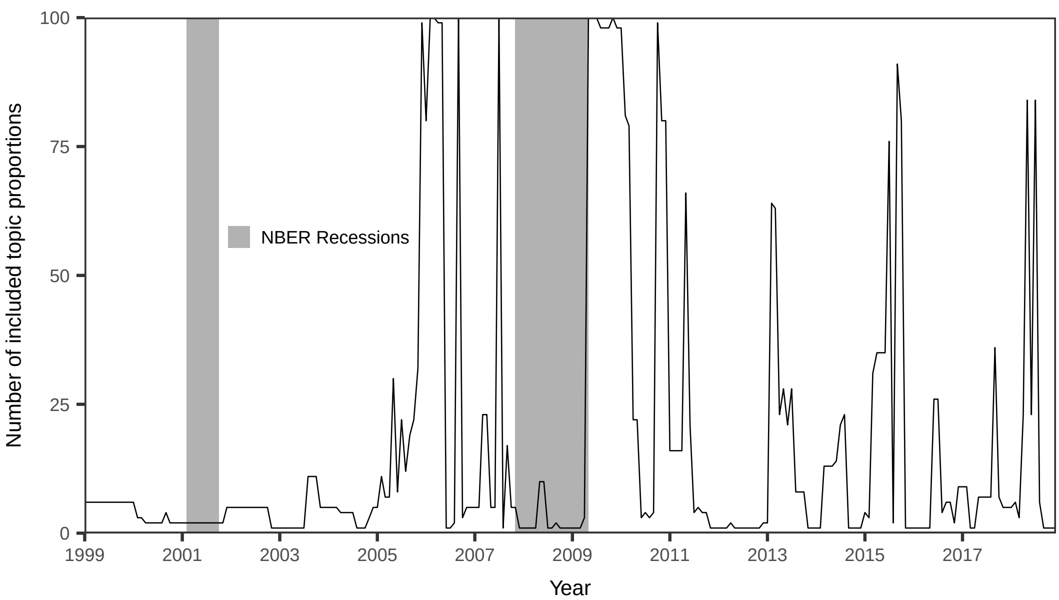

Figure 7: The figure shows which topics are included in the aggregate forecast at each point in time.

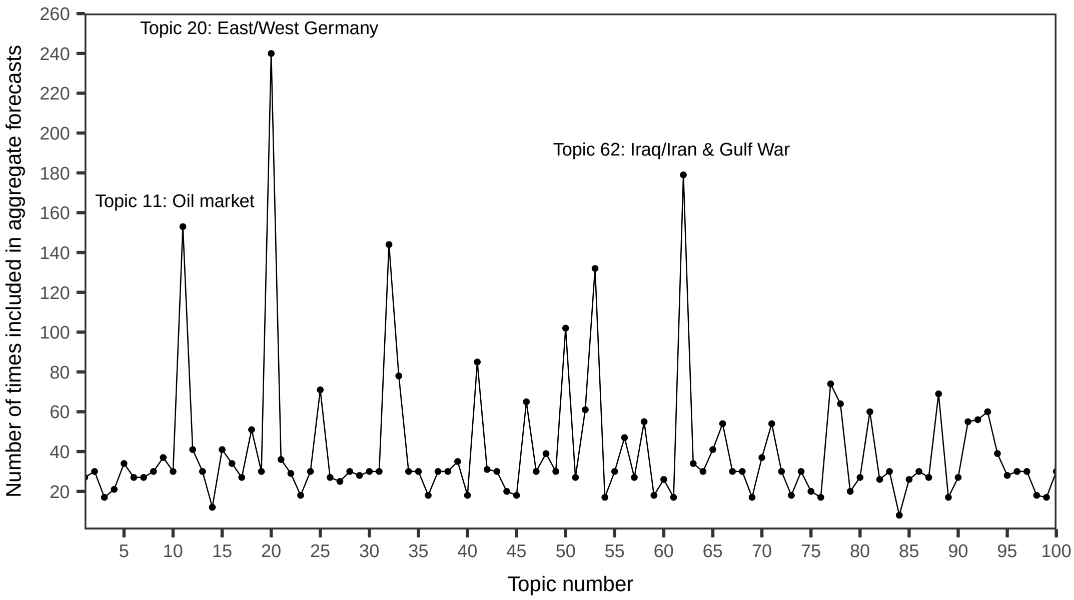

Figure 8: The figure shows how often each topic (proportion) is included in the aggregate forecast.

The performance of the generalized aggregation strategy is similar to the baseline strategy (CTM^{Sw}). In terms of statistical forecasting accuracy, the R^2_{OOS}-statistic is 5.08 (with a p-value of 0.0034) and, hence, slightly lower than our baseline result. On the contrary, economic gains of the generalized aggregation strategy are slightly higher than our baseline result: the annualized certainty equivalent return is 10.01% (7.09%) before (after) transaction costs, the annualized Sharpe ratio is 1.07 (0.80) before (after) transaction costs.

3.3.2 Topic analysis during recessions

The empirical results show that topic 20 is always included as the single predictor when the algorithm switches to the model selection approach. In addition, topic 20 is always included in our generalized forecast aggregation strategy. Given that topic 20 assigns high probabilities to words such as *west, german* and *east*, one might be inclined to argue that political and economic events in Germany contain the most valuable information to forecast the equity premium out of sample. However, even though Germany is one of the most developed global economies, it is more likely to assume that topic 20 captures economic and political events that are not directly related to Germany. We remind the reader at this point that each topic contains *all* unique words of the training sample, differing only in the order. It is thus possible that the CTM predicts a high probability for topic 20 in documents that do not necessarily contain the highest probable words of topic 20, namely *west, german* and *east*. To investigate which events, apart from Germany, might be captured by topic 20, we estimate two new and separate topic models with those articles during the NBER-dated recessions. We focus on these time spans because CTM^{Sw} almost exclusively picked model selection (i.e., topic 20) in those periods (see Figure 3). Given that we use much fewer documents compared to the entire training sample, we set the topic number to K = 40 for both periods.

To identify which new (recession) topics are similar to topic 20, we calculate cosine similarities between the topics by treating the topics as vectors.[9] More precisely, we use the 100 most probable words of each topic and represent all topics in a document term matrix (*dtm*), where each row corresponds to a topic and each column to a word. Each cell of the *dtm* then contains a 0 or 1, depending on whether the word is included in the topic or not. A topic can thus be represented as a vector in a high-dimensional space. The angles between the topic vectors indicate which topics are similar or not. Figure 9 visualizes this approach in a two-dimensional space. Assume that the x-axis denotes *word 1* and the y-axis *word 2*. In this example, topic 1 includes both words, whereas topic 2 only includes the first but not the second word. The smaller the angle between the vectors, the more similar are the topics. The highest angle between those vectors is 90°and the lowest 0°. Calculating the cosine of the angle thus leads to a bounded number between [0;1], where 0 equals no similarity and 1 complete similarity (Huang, 2008).

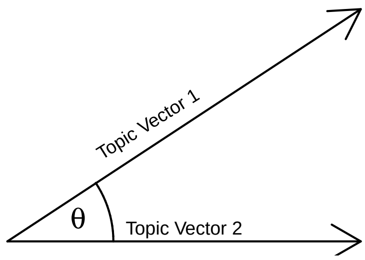

Figure 9: The figure shows two vectors of topics that can only contain two words (x and y). The angle θ hints at the similarity of the two topics.

The cosine similarity can be computed as

$$\cos(\theta) = \frac{\mathbf{t}_{20} \cdot \mathbf{t}_{r,i}}{||\mathbf{t}_{20}||\ ||\mathbf{t}_{r,i}||} = \frac{\sum_{w=1}^{100} t_{20,w} \cdot t_{r,i,w}}{\sqrt{\sum_{w=1}^{100} t_{20,w}^2} \cdot \sqrt{\sum_{w=1}^{100} t_{r,i,w}^2}}, \tag{13}$$

where $\mathbf{t}_{20}$ is the vector of topic 20. $r = 1, 2$ denotes the two NBER-identified recessions and $i = 1, ..., 40$ corresponds to the new (recession) topics. Parameter w denotes the word in each topic. The numerator of Equation 13 is the dot product between two topic vectors and the denominator is the product of the vectors' Euclidean norms.

Table 5 shows the five most probable words for three new topics of the recession periods that have the highest cosine similarity with topic 20. Given that a topic number does not convey any meaning, we simply renumbered the topics from 1 to 3. The word distributions of each topic show that five out of six topics are related to geopolitical events, especially to Russia, China and Israel. The only exception is topic 2 during the dotcom bubble, having most probable words such as *bank, european* and *world*.

Tables 8 and 9 in the Appendix show two newspaper excerpts for each recession period whose documents have the highest topic proportion of the new estimated (recession) topics. For example, during the subprime crisis,

[9]For example, Calomiris and Mamaysky (2019) use the technique of cosine similarity to find those words in their corpus that most often co-occur with their economic word list.

two events are related to regional conflicts, namely the Russian-Georgian war in August 2008 and the Gaza war between Israel and the Palestinians at the end of 2008.

Table 5: Word distributions for recession topics and cosine similarities with topic 20

NBER-recession: 2001-2002			NBER-recession: 2007-2009		
Topic 1* (0.27)	**Topic 2*** (0.25)	**Topic 3*** (0.23)	**Topic 1**** (0.3)	**Topic 2**** (0.23)	**Topic 3**** (0.23)
russia	bank	china	european	iran	china
bush	european	chines	russia	israel	chines
russian	europ	state	russian	palestinian	govern
state	world	taiwan	europ	isra	beji
missil	countri	prison	union	gaza	offici

The table shows the five most probable words in descending order for three out of 40 topics that have the highest cosine similarity with topic 20. The values of the cosine similarity are shown in parentheses. The new topics are separately estimated based on those documents during the NBER-dated recessions.
* The original topic numbers are 25, 3 and 31.
** The original topic numbers are 3, 38 and 35.

3.3.3 Topic analysis with news implied volatility (NVIX)

Our findings that geopolitical events have an impact on aggregate equity returns line up with evidence by Pastor and Veronesi (2012) and Baker et al. (2016). In particular, our result that news about armed conflicts might have predictive power for the equity premium corroborates the finding by Manela and Moreira (2017), who decompose their news-based implied volatility index (NVIX) into sub-categories. They show that the majority of the index' variance is driven by word categories related to (i) Government (Tax Policy), (ii) Financial Intermediation, (iii) Stock Market, (iv) War and (v) Remaining. Figure 10 shows the original NVIX (upper left panel) as well as its constituent parts.[10]

Based on our analysis in the previous section, we conjecture that topic 20 captures relevant information regarding geopolitical news. Remember that topic 20 is always the univariate predictor whenever CTM^{Sw} chooses the model selection strategy. Hence, we investigate whether CTM^{Sw} tends to pick the model selection strategy in periods of high (news-based) uncertainty. To do so, we run the following logit regressions,

$$X_t = \alpha + \beta NVIX_t^i + \varepsilon_t, \tag{14}$$

where X_t is a binary variable, taking the value of 0 if our CTM^{Sw} strategy picks model selection in period t and 1 in case model averaging is selected. $NVIX_t^i$ denotes the value of the NVIX index, whereby the superscript i refers to the sub-index of the NVIX, that is: Original, Government, Financial Intermediation, Stock Market, War or Remaining. The empirical results of Equation 14 are summarized in Table 6. The results reveal that the estimated coefficients of all sub-indices are negative. While model selection is significantly associated with higher news-based uncertainty in the sub-indices Stock Market and War, no significant relationship can be

[10]The values of the sub-categories can be negative since the values are standardized.

established between model selection and news-based uncertainty in the sub-indices Government and Financial Intermediation.[11]

Table 6: Regression results of news-based uncertainty

Index-Category	$\hat{\beta}$
Original	-0.14^{***}
Government	-0.25
Financial Intermediation	-0.14
Stock Market	-1.05^{***}
War	-4.30^{***}
Remaining	-0.19^{***}

The table summarizes the empirical estimates of the slope coefficients from Equation 14. *** indicates that the result is significant at the 1% level. The regressions are based on the period from 1999:01 to 2016:03 (due to availability of the NVIX data until 2016:03.)

Interestingly, Manela and Moreira (2017) find that a substantial fraction of the variation in risk premia stems from concerns regarding tax changes (captured by the Government sub-index) and Wars. Although we identified a tax topic as well, it does not play any role for predicting the equity premium out of sample. Topic 20, which evidently captures political events, does.

[11]We omit the sub-index Natural Disaster since it makes up only a very small part of the aggregate NVIX index.

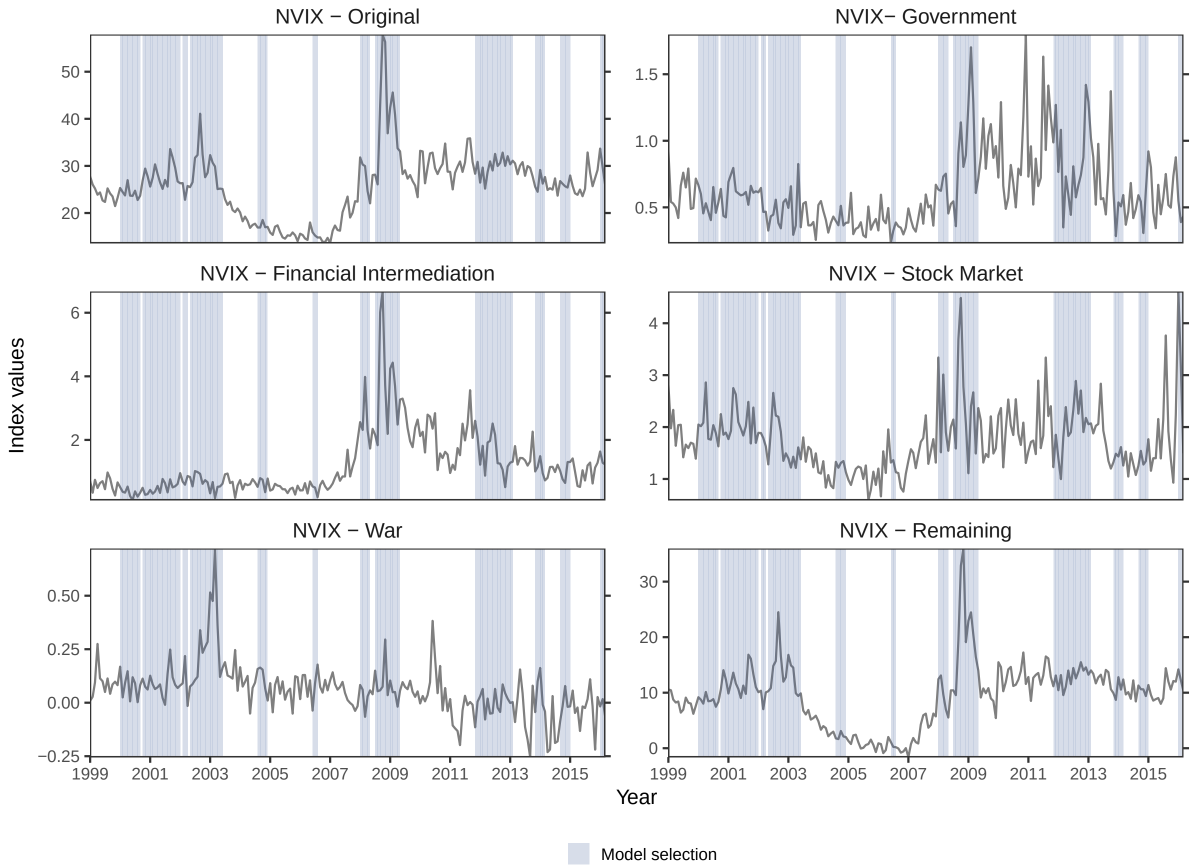

Figure 10: The figure shows the NVIX index (upper left) by Manela and Moreira (2017) as well as five components explaining the index, namely words related to Government, Stock Market, Financial Intermediation and War.

3.3.4 Geopolitical vs economic news

The previous chapters point to the general importance of geopolitical news for predicting equity premia. To investigate this hypothesis further, we compare the predictive power of geopolitical to economic news. For this purpose, we focus on three geopolitical and three economic topics that were picked most frequently within the generalized forecast aggregation strategy in Section 3.3.1. Those were topics 11, 20 and 62 (geopolitical news) and topics 33, 53, 93 (economic news). For the word distributions of these topics, see Table 1. We omit all remaining topics and carry out the forecasting analysis for (i) three geopolitical topics, (ii) three economic topics and (iii) a combination of them.

Table 7 reports the empirical results. Geopolitical topics exhibit stronger predictive power in terms of point prediction accuracy with an R^2_{OOS}-statistic of 6.71% compared to economic topics with an R^2_{OOS}-statistic of 1.87%. With respect to the certainty equivalent return and the Sharpe ratio, the geopolitical topics also prevail, albeit by a smaller margin. Combining geopolitical and economic topics leads to a boost in performance, both in statistical and economic terms: the R^2_{OOS}-statistic is 9.35%, the annualized Sharpe ratio is 0.90 after transaction costs and the certainty equivalent return of 8.57% after transaction costs. Of course, the topics in this chapter were selected with hindsight and the results are thus not directly comparable to our real-time results.

Table 7: Results for geopolitical and economic topics

Included topics	R^2_{OOS}	$\widehat{SR}^{TC}$	$\widehat{CEQ}^{TC}$
Geopolitical $_{\text{topics:}\{11,\ 20,\ 62\}}$	6.71%***	0.74***	6.65%***
Economic $_{\text{topics:}\{33,\ 53,\ 93\}}$	1.87%**	0.67**	6.30%**
Combined $_{\text{topics:}\{11,\ 20,\ 62,\ 33,\ 53,\ 93\}}$	9.35%***	0.90***	8.57%***

The table reports the R^2_{OOS}-statistics, $\widehat{SR}^{TC}$ and $\widehat{CEQ}^{TC}$ when we include only the most frequently selected geopolitical and economic topics or a combination of them as predictors in CTM^{Sw}. The topic numbers are written in subscripts. The word distribution of those topics can be seen in Table 1. *** indicates that the result is significant at the 1% level, ** at the 5% level. $\widehat{SR}^{TC}$ and $\widehat{CEQ}^{TC}$ were generated based on realized variance (RV) and the same weight restriction as in Section 3.2.1.

We also estimated a topic model based on LDA. The posterior was approximated by Gibbs sampling and the priors were chosen according to Griffiths and Steyvers (2004).[12] We estimated (out-of-sample) monthly topic proportions for the LDA results and chose three geopolitical and economic LDA-topics that had the highest correlations with the geopolitical and economic CTM-topics. The word distributions of the LDA-topics closely matched those of the CTM-topics. We then conducted the same analysis as outlined above, namely comparing the predictive power of geopolitical to economic news. Although being lower compared to the CTM, the result of the LDA analysis is unambiguous: geopolitical news contain more valuable information to predict the equity premium out of sample than economic news.[13]

3.3.5 Forecasting gains in up and down markets

To assess the state-dependent behavior of CTM^{Sw}, we first compute the R^2_{OOS}-statistics for NBER-dated expansions and recessions separately. The R^2_{OOS} in expansions is 6.55% and 6.35% in recessions. As we only observe 28 NBER-dated recession periods (out of 240), we also use a different categorization of periods to analyze state-dependent behavior. To this end, we follow Baltas and Karyampas (2018) by calculating the cumulative sum of differences in the MSPE for months with non-negative realizations of the equity premium and months with negative realizations. Our post-holdout out-of-sample period contains 148 periods with non-negative excess returns and 92 periods with negative excess returns. The cumulative differences in MSPE ($DCMSPE$) are calculated in three different versions: over all months ($DCMSPE$)[14], over months with non-negative realizations of the equity premium ($DCMSPE^+$) and over months with negative realizations of the equity premium ($DCMSPE^-$):

[12] The approach is implemented in the R-package by Hornik and Grün (2011).

[13] The R^2_{OOS}-statistic based on geopolitical topics is 3.98% and 0.57% for economic topics. Both results are statistically significant at the one and five percent level, respectively.

[14] Note that this quantity has already been reported in Figure 3 and is repeated here for convenience.

$$DCMSPE = \sum_{s=q_0+1}^{q} (r_{m+s} - \bar{r}_{m+s})^2 - \sum_{s=q_0+1}^{q} (r_{m+s} - \hat{r}_{m+s})^2 \tag{15}$$

$$DCMSPE^{+} = \sum_{s=q_0+1}^{q} (r_{m+s} - \bar{r}_{m+s})^2 \cdot \mathbb{I}_{r_{m+s} \geq 0} - \sum_{s=q_0+1}^{q} (r_{m+s} - \hat{r}_{m+s})^2 \cdot \mathbb{I}_{r_{m+s} \geq 0} \tag{16}$$

$$DCMSPE^{-} = \sum_{s=q_0+1}^{q} (r_{m+s} - \bar{r}_{m+s})^2 \cdot \mathbb{I}_{r_{m+s} < 0} - \sum_{s=q_0+1}^{q} (r_{m+s} - \hat{r}_{m+s})^2 \cdot \mathbb{I}_{r_{m+s} < 0}, \tag{17}$$

where $\mathbb{I}_{r_{m+s} \geq 0}$ and $\mathbb{I}_{r_{m+s} < 0}$ denote indicator functions for non-negative and negative realizations of the equity premium. Figure 11 presents the evolution of the differences in the cumulative sums of MSPE. The unambiguous message from the figure is that the gains in forecasting accuracy are accrued in months with negative realizations of the equity premium.

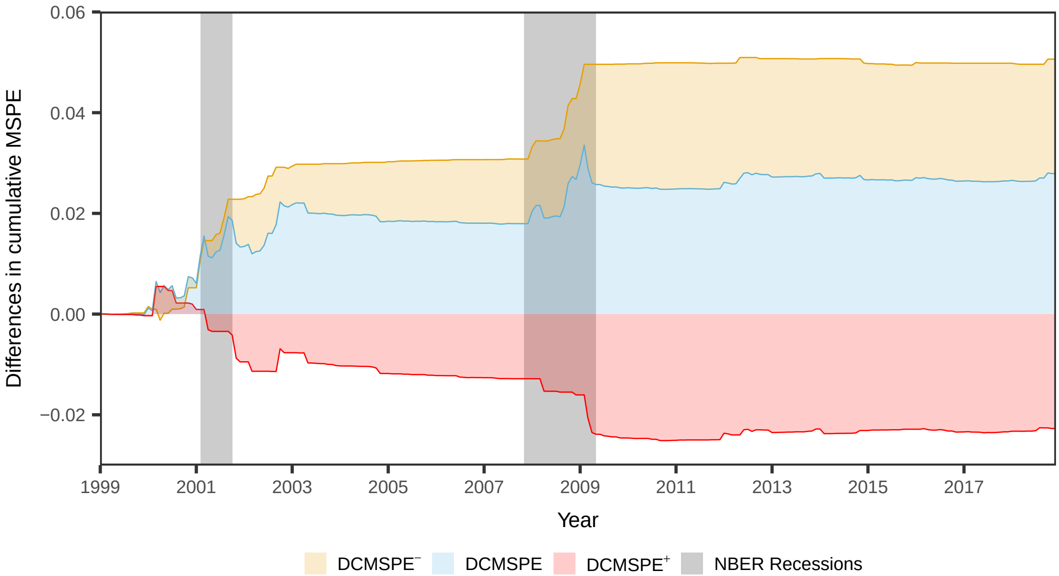

Figure 11: The figure presents the difference in the cumulative sum of MSPE for the simple historical mean forecast and CTM^{Sw}. The cumulative differences in MSPEs are calculated in three different versions: over all months ($DCMSPE$), over months with non-negative realizations of the equity premium ($DCMSPE^{+}$) and over months with negative realizations of the equity premium ($DCMSPE^{-}$).

3.3.6 Forecasting channel

We shed some light on whether the predictability either stems from the discount rate channel, the cash flow channel or from both. Following Cochrane (2011), we use the dividend-price ratio (DP) as a proxy of discount rates and dividend growth (DG) as a proxy of cash flows. Given that CTM^{Sw} successfully predicts aggregate stock returns, it must predict either the discount rate (proxied by DP) or the cash flow (proxied by DG), or both. To explore these channels, we regress DP and DG on the forecast implied by CTM^{Sw} strategy and its lagged value, respectively. We obtain the following results with robust standard errors (in parentheses) by Newey and West (1987):

$$\widehat{DP}_{t+1} = \underset{(0.0844)}{-0.1150} - \underset{(0.3172)}{0.6981}\hat{r}_{t+1}^{CTM^{Sw}} + \underset{(0.0205)}{0.9704}DP_t \tag{18}$$

$$\widehat{DG}_{t+1} = \underset{(0.0034)}{0.0013} + \underset{(0.2854)}{0.7973}\hat{r}_{t+1}^{CTM^{Sw}} + \underset{(0.0805)}{0.0329}DG_t. \tag{19}$$

The results from Equations 18 and 19 reveal that the forecasts of CTM^{Sw} are associated with positive dividend growth (t-statistic of 2.7935) and lower dividend-price ratios (t-statistic of −2.2010). These findings indicate that the forecasting ability of our forecasting strategy stems from both the discount rate channel and the cash flow channel. The results are robust to different lag lengths of DP and DG.

4 Conclusion

We have introduced a new strategy for predicting the equity premium out of sample by synthesizing news from a vast collection of newspaper articles. Based on a correlated topic model, we identify various themes and their media coverage over time. We exploit the estimated topic proportions as predictors for the equity premium in univariate predictive regressions. To obtain an overall forecast of the equity premium, we propose a flexible data-adaptive switching strategy that merges the advantages of model averaging and model selection.

Our empirical results document strong out-of-sample predictive power based on our proposed forecasting strategy. The news-based aggregate forecast embeds predictive information that is not captured by established predictors of the equity premium. Gains over the historical mean model are achieved when they are most needed from an investor's perspective, namely in down markets. Switching between model averaging and model selection substantially enhances the forecasting gains. In addition, our findings support the notion that geopolitical rather than economic news are more important for predicting the equity premium out of sample.

The results in this paper, along with findings by, e.g., Jiang et al. (2019) and Calomiris and Mamaysky (2019), demonstrate the huge potential of text mining for analyzing and forecasting stock returns in various facets. While sentiment analysis has arguably been the most prominent tool for investigating the impact of news on stock returns, our study suggests that text analysis based on topic modeling is worthwhile for predicting the equity premium. For example, future analysis could continue deepening the analysis on the kind of topics that encode predictive information, use alternative topic modeling approaches or different strategies for mapping the estimated topic proportions into equity premium forecasts.

References

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Balduzzi, P. and Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78.

Baltas, N. and Karyampas, D. (2018). Forecasting the equity risk premium: The importance of regime-dependent evaluation. *Journal of Financial Markets*, 38:83–102.

Beckmann, J., Koop, G., Korobilis, D., and Schüssler, R. (2018). Exchange rate predictability and dynamic bayesian learning. *Available at SSRN*.

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM.

Blei, D. M. and Lafferty, J. D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 1(1):17–35.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Calomiris, C. W. and Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, Forthcoming.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.

Dyer, T., Lang, M., and Stice-Lawrence, L. (2017). The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation. *Journal of Accounting and Economics*, 64(2-3):221–245.

Engelberg, J. E., Reed, A. V., and Ringgenberg, M. C. (2012). How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics*, 105(2):260–278.

Farmer, L., Schmidt, L., and Timmermann, A. (2018). Pockets of predictability. Unpublished working paper.

Freyberger, J., Neuhierl, A., and Weber, M. (2017). Dissecting characteristics nonparametrically. Technical report, National Bureau of Economic Research.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Gentzkow, M., Kelly, B. T., and Taddy, M. (2017). Text as data. *Journal of Economic Literature*, Forthcoming.

Giglio, S. and Xiu, D. (2017). Inference on risk premia in the presence of omitted factors. Technical report, National Bureau of Economic Research.

Glasserman, P. W. and Mamaysky, H. (2019). Does unusual news forecast market stress? *Journal of Financial and Quantitative Analysis*, Forthcoming.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1):5228–5235.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):267–297.

Gu, S., Kelly, B., and Xiu, D. (2018). Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research.

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99, Supplement 1:S114 – S133.

Hansen, S., McMahon, M., and Prat, A. (2017). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259.

He, J., Hu, Z., Berg-Kirkpatrick, T., Huang, Y., and Xing, E. P. (2017). Efficient correlated topic modeling with topic embedding. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 225–233. ACM.

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347.

Hornik, K. and Grün, B. (2011). topicmodels: An r package for fitting topic models. *Journal of Statistical Software*, 40(13):1–30.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, volume 4, pages 9–56.

Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, 49(3):851–889.

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1):126–149.

Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644.

Kelly, B. T., Pruitt, S., and Su, Y. (2017). Some characteristics are risk exposures, and the rest are irrelevant. *Unpublished working paper, University of Chicago.*

Kozak, S., Nagel, S., and Santosh, S. (2017). Shrinking the cross section. Technical report, National Bureau of Economic Research.

Larsen, V. H. and Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1):203 – 218.

Li, J. and Tsiakas, I. (2017). Equity premium prediction: The role of economic and statistical constraints. *Journal of Financial Markets*, 36:56–75.

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

Manela, A. and Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1):137–162.

Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, 60(7):1772–1791.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pastor, L. and Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, 67(4):1219–1264.

Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1):46–65.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2013). International stock return predictability: what is the role of the united states? *The Journal of Finance*, 68(4):1633–1662.

Roberts, M., Stewart, B., and Tingley, D. (2018). stm: R package for structural topic models. r package version 1.3.3. *Journal of Statistical Software*, Forthcoming.

Roberts, M. E., Stewart, B. M., and Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, pages 1–17.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18.

Turner, J. D., Ye, Q., and Walker, C. B. (2018). Media coverage and stock returns on the london stock exchange, 1825–70. *Review of Finance*, 22(4):1605–1629.

Wachter, J. A. and Warusawitharana, M. (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148(2):162–178.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

A Excerpts from newspaper articles

Table 8: Excerpts from newspaper articles between March 2001 and November 2001

	Newspaper	Date	Topic
Meta	Washington Post	2001-11-11	1
Headline	U.S., Russia Likely To Agree on Arms; Summit Could Lead to Historic Cuts		
Excerpt	*"The United States and Russia are working to establish an unprecedented arms control agreement that calls for deep but unequal reductions in strategic nuclear weapons over the next decade. [...]"*		
Meta	Washington Post	2001-08-11	1
Headline	Rumsfeld to Discuss ABM Treaty in Moscow		
Excerpt	*"Defense Secretary Donald H. Rumsfeld leaves tonight for talks in Moscow on missile defense and possible cuts in offensive nuclear forces, hoping to convince his Russian counterpart that mutual withdrawal from the Anti-Ballistic Missile Treaty is in the best interests of both nations. [...] "*		
Meta	New York Times	2001-07-10	2
Headline	THE MARKETS: CURRENCIES		
Excerpt	*"Brazil's real strengthened after the central bank sold an undisclosed sum of dollars to help prop it up. [...] "*		
Meta	New York Times	2001-03-14	2
Headline	THE MARKETS: CURRENCIES		
Excerpt	*"Euro drops. The euro fell as accelerating French and German inflation raised concern the European Central Bank will not cut interest rates soon. [...] "*		
Meta	New York Times	2001-04-04	3
Headline	COLLISION IN CHINA: THE OVERVIEW; CHINA FAULTS U.S. IN INCIDENT; SUGGESTS RELEASE OF CREW HINGES ON OFFICIAL APOLOGY		
Excerpt	*"The Chinese government blamed the United States today for Sunday's midair collision of a spy plane and a trailing Chinese fighter jet and suggested that the release of the 24 American crew members hinged on Washington's willingness to apologize. [...] "*		
Meta	Washington Post	2001-09-29	3
Headline	China Frees U.S. Writer Held on Spy Charges		
Excerpt	*"China released an American writer jailed since April on charges of spying for Taiwan and put him on a plane to the United States tonight, U.S. officials and Chinese state media reported. [...] "*		

The table shows excerpts from newspaper articles between March 2001 and November 2001. The excerpts are from those articles whose topics had the highest cosine similarity with topic 20 from the training sample.

Table 9: Excerpts from newspaper articles between December 2007 and June 2009

	Newspaper	Date	Topic
Meta	New York Times	2001-11-11	1
Headline	European Union to Resume Russian Partnership Talks		
Excerpt	*"The European Union said Monday that it would resume negotiations with Russia that it had halted following Russia's invasion of Georgia, in a significant step toward normalizing ties with Moscow. [...]"*		
Meta	New York Times	2008-09-16	1
Headline	U.S. Envoy Says Conflicts With Enclaves Shouldn't Keep Georgia Out of NATO		
Excerpt	*"The West should not use Georgia's conflicts with the Russian-backed breakaway enclaves of South Ossetia and Abkhazia as an excuse to keep Georgia out of NATO, the United States ambassador to the alliance said Monday. [...] "*		
Meta	New York Times	2008-11-05	2
Headline	Israeli Strike Is First in Gaza Since Start Of Cease-Fire		
Excerpt	*"Israel carried out an airstrike on Gaza on Tuesday night after its troops clashed with Hamas gunmen along the border in the first such confrontation since a cease-fire took effect in June. [...] "*		
Meta	New York Times	2001-03-14	2
Headline	Israel and Hamas Near Truce on Gaza		
Excerpt	*"The prime minister and defense minister of Israel have agreed to an Egyptian-brokered cease-fire with Hamas for the Gaza area starting Thursday, Israel Radio reported on Wednesday morning. [...] "*		
Meta	New York Times	2008-06-30	3
Headline	China to Resume Talks With Dalai Lama		
Excerpt	*"China said Sunday that it would soon resume talks with representatives of the Dalai Lama, weeks before the start of the Olympic Games. [...] "*		
Meta	Washington Post	2008-05-03	3
Headline	Dalai Lama's Envoys Heading to China; Informal Talks to Focus on Unrest in Tibetan Areas		
Excerpt	*"Representatives of the Dalai Lama are scheduled to arrive in China on Saturday to begin informal talks with their Chinese counterparts on the unrest in Tibet. [...] "*		

The table shows excerpts from newspaper articles between December 2007 and June 2009. The excerpts are from those articles whose topics had the highest cosine similarity with topic 20 from the training sample.