Making text count: economic forecasting using newspaper text

Arthur Turrell, Bank of England

Eleni Kalamara, Nomura

George Kapetanios, King's College London

Chris Redl, International Monetary Fund

Sujit Kapadia, European Central Bank

Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia. "Making text count: economic forecasting using newspaper text." *Journal of Applied Econometrics* **37**, no. 5 (2022): 896-919.
The views expressed do not reflect those of the Bank of England, European Central Bank, or International Monetary Fund. Email: arthur.turrell@bankofengland.co.uk

Introduction

- Question: Can we use newspaper text to make (better) macro forecasts?

- Motivation: text is timely and rich in information. It could reflect new information faster than traditional data or it could cause changes in the economy via animal spirits or viral narratives. Either way, **news text could carry macroeconomic information**.

Introduction

- Question: Can we use newspaper text to make (better) macro forecasts?
- Motivation: text is timely and rich in information. It could reflect new information faster than traditional data or it could cause changes in the economy via animal spirits or viral narratives. Either way, **news text could carry macroeconomic information**.
- **Insights from this project:**
  - text significantly improves forecasts of real economy variables, with the extra performance coming during stressed times
  - simple metrics, like counting words, gives some signal – but *not* in marginal terms
  - machine learning (+ feature engineering) can increase forecast performance even in marginal terms, and is more transferable

Literature shows that text can contain signals of economic activity

- Text as data (Gentzkow, Kelly and Taddy, 2017)

- Relationship between text and activity
  - Financial markets & firms (Jegadeesh and Wu, 2013; Loughran and McDonald, 2011, 2013)
  - Uncertainty (Alexopoulos and Cohen, 2015; Baker, Bloom and Davis, 2016)
  - Sentiment (Shapiro, Sudhof and Wilson, 2018)

- Fore/Nowcasting using text
  - Financial markets and firms (Antweiler and Frank, 2004; Tetlock, 2007)
  - Sentiment using dictionaries + topics (Ardia, Bluteau and Boudt, 2019; Larsen and Thorsrud, 2019; Thorsrud, 2018)
  - Daily sentiment predicts daily stock returns more effectively during recessions (Garcia, 2013)

- Where our paper fits in: ▸ Marginal contribution

Data

	Circulation	Unique articles	% of total	⟨articles/month⟩	First article	Last article
The Guardian	138,000	288,928	54.7	828	1990-01-06	2019-01-23
The Daily Mirror	563,000	141,332	26.8	492	1995-03-01	2019-01-23
The Daily Mail	1,265,000	97,897	18.5	281	1990-01-11	2019-01-23
Total	1,966,000	528,157	100.0	1,601	-	-

Table 1: Descriptive statistics of articles from selected UK newspapers. Data from Dow Jones Factiva. First release articles covering Commodity/Financial Market News, Corporate/Industrial News, and Economic News (including editorials and commentaries/opinions). De-duplicated. Circulations from June 2018. ▸ Text cleaning details

Do simple text-based time series seem plausible as economic indicators?

We use a wide range of non-learning algorithms to turn text directly into time series – from counting single words to sentiment analysis

Positive and negative dictionary	**Boolean**	**From computer science literature**
Financial stability (Correa et al., 2017)	Economic Uncertainty (Alexopoulos and Cohen, 2009)	VADER sentiment (Gilbert, 2014)
Finance oriented (Loughran and McDonald, 2013)	Monetary policy uncertainty (Husted, Rogers and Sun, 2017)	'Opinion' sentiment (Hu et al., 2017; Hu and Liu, 2004)
Afinn sentiment (Nielsen, 2011)	**Economic Policy Uncertainty** (Baker, Bloom and Davis, 2016)	
Harvard IV (used in Tetlock (2007))		
Anxiety-excitement (Nyman et al., 2018)		
Single word counts of "uncertain" and "econom"		
tf-idf applied to "uncertain" and "econom"		

Table 2: Three broad categories of algorithm-based text metrics we use. Includes measures of both **uncertainty** and **sentiment**.

▸ Details of selected text metric types ▸ Examples of selected text metrics ▸ ADF tests of text metrics

Text sentiment vs. traditional sentiment $\Longrightarrow$ text captures some info on activity

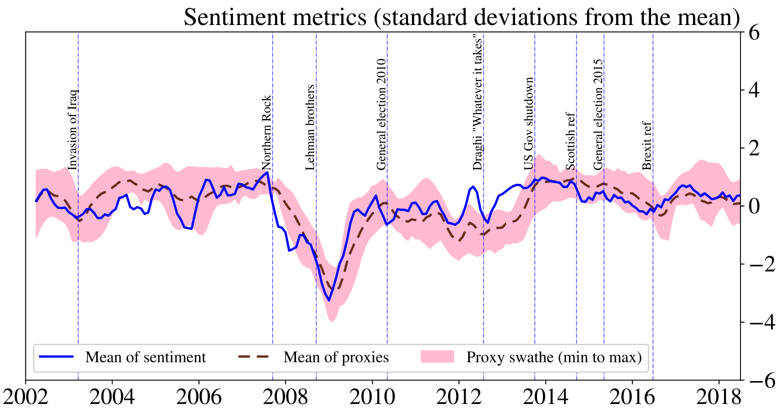


Figure 1: 3 month rolling mean of sentiment of Daily Mail text (solid line) against proxies for sentiment (broken line). Swathe is min/max of proxies. **Uncertainty much less convincing.** ‣ Proxies ‣ Proxy correlations ‣ Uncertainty

Can simple text-based time series improve forecast performance?

Forecasting with text: beating a factor model benchmark

- Simple text measures improve performance vs. an AR(1) (see Simple text in AR(1))

- But text needs to add **marginal forecast performance** – ie to improve on a model with highly statistically significant confounders

$\Longrightarrow$ We use factors $\vec{F}_t$ derived from 33 series covering output, trade, the labour market, inflation, house prices, retail sales, capacity utilisation, and business and household expectations (Redl, 2017):

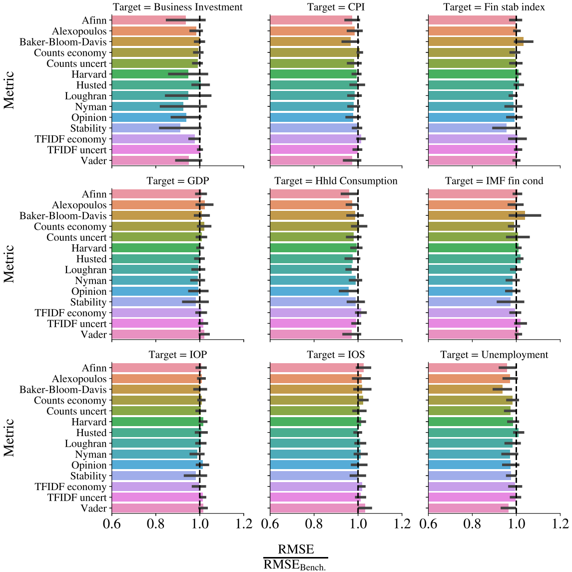
$$y_{t+h} = \alpha + \beta \cdot y_{t-1} + \overbrace{\sum_j \gamma_j F_{jt}}^{\text{Macro factors}} + \overbrace{\eta \cdot x_t}^{\text{Text term}} + \epsilon_t$$

versus

$$y_{t+h} = \alpha + \beta \cdot y_{t-1} + \sum_j \gamma_j F_{jt} + \epsilon_t$$

NB: we use $J = 2$ factors, a rolling window of 36 months, and time horizons of $h = 3, 6, 9$ months. Model supposes information on y_{t-1} only available at t.

Forecasting with simple text metrics with factor model benchmark



Using machine learning to get more out of text

Turning text into time series – feature engineering + machine learning

Let model decide what terms to include. Method:

- Choose dictionary of terms that is **independent** of source text to avoid information leakage. This is very important for forecasts ‣ Information leakage and other text-based forecasting pitfalls
- We use union of all terms in dictionary methods + terms from a dictionary of economic terms up to 3-grams[1]: 9660 distinct terms (some of which never appear)
- Each article represented as a vector of terms:

$$\overrightarrow{\mathrm{tf}(a)} = \left(\mathrm{tf}(a)_{w_1}, \mathrm{tf}(a)_{w_2}, \cdots\right)$$

- Take mean vector at whatever required frequency
- Use as features for machine learning model, i.e. predict $\hat{y} = f_{\mathrm{ML}}\left(\dots, \overrightarrow{\mathrm{tf}}\right)$

[1] An N-gram is a term N words long, up to 3-grams includes 1-, 2-, and 3-grams.

Forecasting with ML – beating an OLS factor model benchmark

Model:

$$y_{t+h} = \overbrace{f_{\text{ML}}}^{\text{Model}} \left(y_{t-1}, \overbrace{\overrightarrow{F_t}}^{\text{Macro vector}}, \overbrace{\overrightarrow{\text{tf}_t}}^{\text{Text vector}} \right) + \epsilon_t$$

versus

$$y_{t+h} = \alpha + \beta \cdot y_{t-1} + \sum_j \gamma_j F_{jt} + \epsilon_t$$

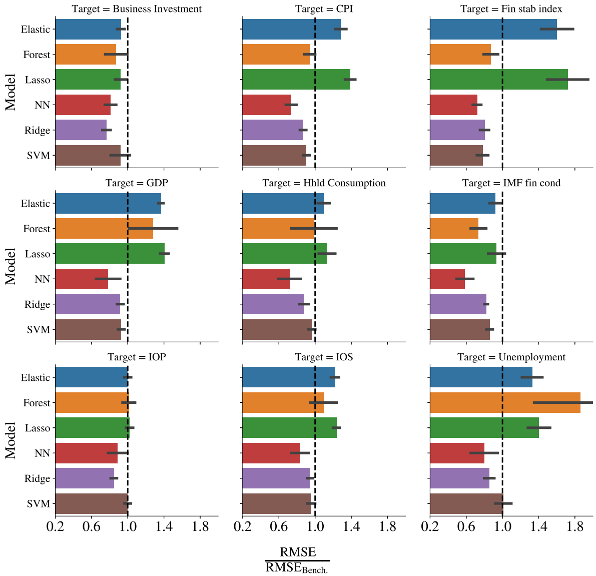
NB: we use $J = 2$ factors, a rolling window of 36 months, and time horizons of $h = 3, 6, 9$ months. Model supposes information on y_{t-1} only available at t. ‣ Example neural network forecast of GDP

‣ Diebold-Mariano tests for ML-OLS factor model ‣ Alternative specification versus ML factor model

Results hold against simpler AR(1) benchmark:

‣ AR(1) ML-OLS model ‣ Diebold-Mariano test for ML-OLS AR(1) model ‣ AR(1) ML-ML model

Forecasting with ML: factor model benchmark results



Forecast improvements are during stressed times

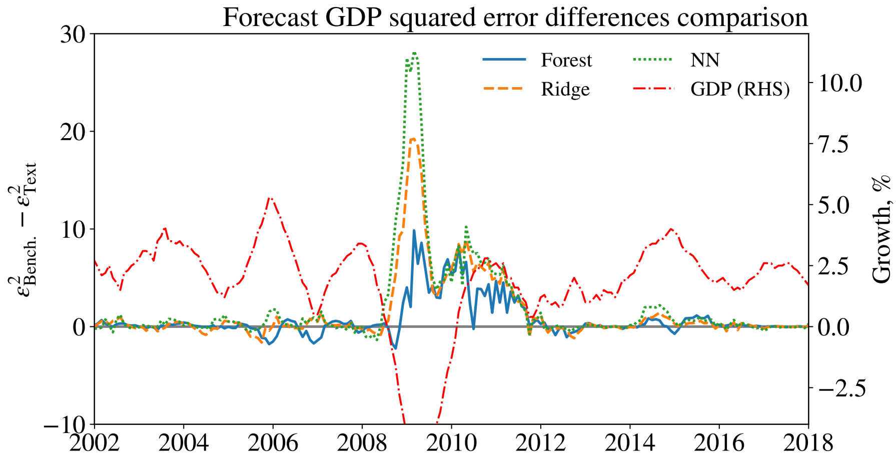


Figure 2: Above $y = 0$ means out-of-sample improvement versus benchmark.

Forecasting with machine learning

Results suggest:

- Machine learning approach gives clear RMSE improvements for every target variable (though not for every model – neural network performs best here)
- ML can improve forecasts of the three time series in the BOE *Inflation Report* – GDP, unemployment, and CPI – relative to our benchmark
- Results persist across different specifications
- “If it bleeds, it leads”: text adds value to forecasts when it matters most, during stressed times

Conclusion and future possibilities

Conclusion and future possibilities

- Text can be forward looking economic indicator and forecast input (robust to horizon and newspaper)

Conclusion and future possibilities

- Text can be forward looking economic indicator and forecast input (robust to horizon and newspaper)
- Methods to use text in real-time macroeconomic forecasts:
    - Simple text metrics do well in absolute terms, but not in marginal terms

Conclusion and future possibilities

- Text can be forward looking economic indicator and forecast input (robust to horizon and newspaper)
- Methods to use text in real-time macroeconomic forecasts:
  - Simple text metrics do well in absolute terms, but not in marginal terms
  - Machine learning (esp. neural network) + term frequency vector does well even in marginal terms and method is transferable to other problems

Conclusion and future possibilities

- Text can be forward looking economic indicator and forecast input (robust to horizon and newspaper)
- Methods to use text in real-time macroeconomic forecasts:
  - Simple text metrics do well in absolute terms, but not in marginal terms
  - Machine learning (esp. neural network) + term frequency vector does well even in marginal terms and method is transferable to other problems
- Future work: can large language models do better? (Though be careful of training data!)

Thanks for listening

“A good newspaper, I suppose, is a nation talking to itself.”

– Arthur Miller, *The Observer*, 1961

Appendix

This paper ▸ Back

- Many different methods compared on same text sources
- Newspaper text is about consumer (rather than investor) news; targets are macro variables
- Text sources broadly cover political spectrum
- Going beyond simple counting and dictionaries ($=$ algorithms), and one-off estimation of topic models
- Comprehensive forecasting exercises - direct h-step ahead forecasts from rolling window rather than single train/test (in/out of sample) forecast

Text cleaning details

We use the following steps to pre-process newspaper text:

1. remove punctuation, hyperlinks, hyper text markup language (HTML) tags, special characters, leading or trailing white space characters, and digits;
2. set all characters in lower case; and
3. drop words in our list of stop words.

Stop words include ‘and’, ‘is’, ‘in’, and so on – see Nothman, Qin and Yurchak (2018) for a discussion. We drop words from the union of two popular lists of stop words: the NLTK word list (Bird and Loper, 2004) and the list proposed by Puurula (2013). ▸ Back

Punctuation economy text metric

Metric based on measuring sentiment within individual sentences, discarding the information contained in the remainder of the article. ▸ Back

- Given a specific term – here ‘econom’ – the metric returns the sentiment of the words of the surrounding sentence fragment.
- Retains and processes surrounding snippets of text up to the closest punctuation characters.
- Ignores punctuation that does not indicate sentence fragment , eg ‘Mr.’, ‘Mrs.’, ‘Dr.’, ‘etc.’, etc.
- Coreference resolution[2] performed before text surrounding terms is obtained (Clark and Manning, 2016; Elango, 2005).
- Sentiment analysis uses union of dictionaries of Nyman et al. (2018), Nielsen (2011), Correa et al. (2017), and the Harvard IV psychological dictionary.

[2] Coreference resolution allows for any linguistic expressions that refer to the same real-world entity indirectly to be replaced by explicit references to that real world entry. An example would be “The cat is on the mat. It looks hungry.”, which would be converted to “The cat is on the mat. The cat looks hungry.”

Proxies for sentiment and uncertainty

Name	Description	Proxy for	Type
Lloyds Bus Conf	Lloyds Business Barometer – confidence	Sentiment	Survey
Lloyds Bus Activity	Lloyds Business Barometer – activity over next 12 months	Sentiment	Survey
OECD Bus Conf	OECD UK business confidence	Sentiment	Survey
Composite PMI	Composite measure of PMI	Sentiment	Survey
GfK Consumer Conf	GfK Consumer Confidence	Sentiment	Survey
IG Corp Bond spread	Investment Grade Corporate Bond spread	Uncertainty, sentiment	High-frequency market-based
Jurardo Fin Uncert	UK version of Jurado, Ludvigson and Ng (2015) from Redl (2018); financial uncertainty, $h = 3$	Uncertainty	Forecast error
Jurardo Macro Uncert	UK version of Jurado, Ludvigson and Ng (2015) from Redl (2018); macroeconomic uncertainty, $h = 3$	Uncertainty	Forecast error
BoE agg credit spread	Bank of England measure of aggregate credit spread	Uncertainty	Market-based
VIX	CBOE volatility index	Uncertainty	High-frequency market-based
VFTSEIX	FTSE volatility	Uncertainty	High-frequency market-based
GDP forecast std dev	UK Treasury collected standard deviation of professional forecasts of GDP, 3 months ahead	Uncertainty	Low-frequency forecast spread
BoE Uncert	Bank of England uncertainty index	Uncertainty	Composite
ERI volatility	GBP Exchange Rate Index volatility	Uncertainty	High-frequency market-based

Table 3: Descriptions of the proxy time series.

Back

Turning text into time series – details of common methods (applied at article level)

Back

- Boolean: two sets of terms, E and U, and w a term in article a. a is counted as a '1' iff

$$(w \in E) \wedge (w' \in U) \quad \forall \quad w, w' \in a$$

- Dictionary: D split into positive, D^+, and negative, D^- term sets and defines a mapping $D : W \rightarrow C$ such that $w \in W$ has an associated score $c \in \mathbb{R}$. Score for an article a with terms w is given by

$$S = \frac{1}{|w|}\left(\sum_w D^+(w) - \sum_w D^-(w)\right)$$

- Term frequency – inverse document frequency (tf-idf): define number articles per day (N) and number of articles with term as ($n_w \leq N$).

$$\text{tf-idf}(a)_w = \frac{\ln\left(1 + \text{tf}(a)_w\right)}{\ln\left(1 + N/n_w\right)}$$

Turning text into time series – examples of algorithms

Back

Text	TFIDF economy	Vader	Counts economy	Alexopoulos	Stability
Global GDP growth picked up during 2016 and has been strong over the past year (Section 1.1). Weighted by countries' shares of UK exports, global growth is estimated to have remained at 0.8% in 2017 Q4. That pace of growth is expected to persist in the near term, above expectations in November. Survey indicators of output (Chart 1.1) and new orders remain robust, particularly in the euro area and United States. Measures of business and consumer confidence are also healthy...	-0.00	0.97	0	0	0.03
The economy has struggled and is in a bad state with disappointing performance, unhappy consumers, low confidence with high uncertainty. Policy faces a number of risks which could transmit to the real economy, and pundits are increasingly concerned about a crash.	-0.15	-0.93	-2	1	-0.11
The current direction of policy is very bad.	-0.00	-0.54	0	0	-0.25
The current direction of policy is very good.	-0.00	0.44	0	0	0.25

Table 4: Selected algorithm-based methods applied to example text.

Augmented Dickey-Fuller test

	The Daily Mirror	No. obs.	The Daily Mail	No. obs.	The Guardian	No. obs.
TFIDF uncert	-03.37**	254	-07.66***	272	-04.22***	318
Counts uncert	-01.28	242	-04.05***	268	-01.79	308
Alexopoulos	-01.87	241	-04.14***	267	-01.70	309
Baker-Bloom-Davis	-05.81***	255	-05.29***	271	-00.99	309
Husted	-08.70***	256	-08.97***	272	-04.03***	319
Opinion	-04.35***	254	-04.56***	269	-03.09**	320
Harvard	-07.67***	255	-03.08**	264	-04.01***	319
Loughran	-04.61***	255	-04.43***	268	-02.16	320
Vader	-02.78*	251	-02.95**	267	-03.13**	320
Afinn	-02.63*	251	-03.27**	268	-02.99**	320
Counts economy	-01.96	249	-03.46***	270	-03.31**	320
Stability	-04.47***	255	-04.90***	270	-04.47***	321
TFIDF economy	-02.87*	255	-02.69*	264	-03.36**	311
Nyman	-05.60***	255	-04.54***	271	-03.45***	311

Table 5: Results of an Augmented Dickey-Fuller test on all text metrics. The number of observations differ as the number of lags to include is chosen using the AIC information criterion. Asterisks denote p-values; 1%: ***, 5%: **, 10%: *. Back

Indicators of uncertainty versus proxies for uncertainty

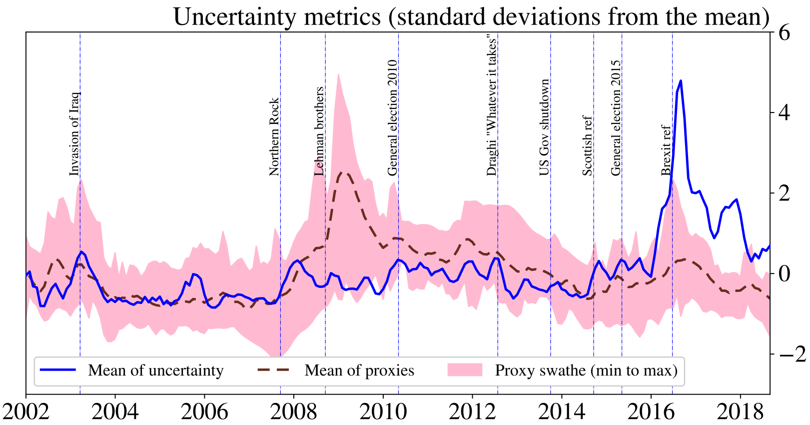


Figure 3: Three month rolling mean of uncertainty from text using The Guardian (solid line) against proxies for uncertainty (broken line). Swathe is min/max of proxies. ▸ Proxies ▸ Proxy correlations ▸ Back

Correlations of sentiment metrics with proxies for sentiment

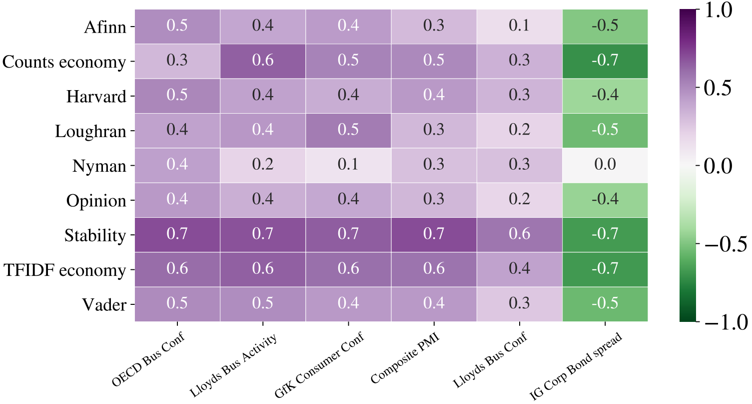


Figure 4: Heatmap of correlations between text metrics, averaged over newspapers, and proxies for macroeconomic sentiment at a three month horizon. Back

Correlations of uncertainty metrics with proxies for uncertainty

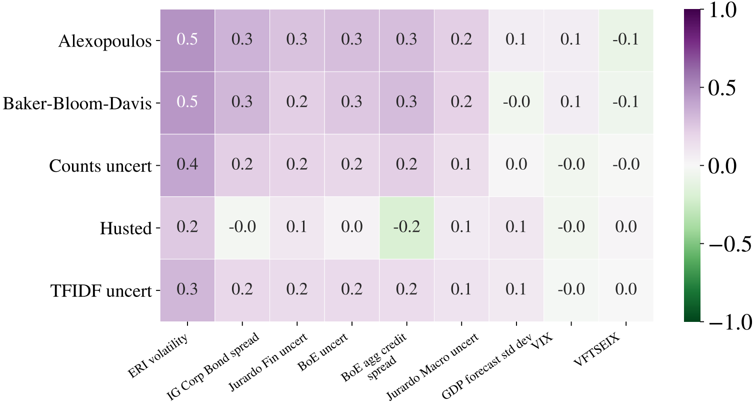


Figure 5: Heatmap of correlations between text metrics, averaged over newspapers, and proxies for financial sentiment at a three month horizon.

▸ Back to uncertainty time series ▸ Back to summary of text metric results

Information leakage and text forecasting pitfalls (aka this is hard, be careful)

Back

- Terms used from entire corpus based on, eg, a frequency threshold (5–95% quantiles). Problem: terms that suddenly appear at one point in time correlated with big macro developments.
    - Example: ‘sub-prime’

Information leakage and text forecasting pitfalls (aka this is hard, be careful) ▸ Back

- Terms used from entire corpus based on, eg, a frequency threshold (5–95% quantiles). Problem: terms that suddenly appear at one point in time correlated with big macro developments.
  - Example: 'sub-prime'
- Topic model trained on entire corpus, topics deployed to forecast text earlier in time. Problem: topics that 'know' about future events.
  - Example: topics that put weight on words unique to one crisis or time period

Information leakage and text forecasting pitfalls (aka this is hard, be careful)

Back

- Terms used from entire corpus based on, eg, a frequency threshold (5–95% quantiles). Problem: terms that suddenly appear at one point in time correlated with big macro developments.
  - Example: ‘sub-prime’
- Topic model trained on entire corpus, topics deployed to forecast text earlier in time. Problem: topics that ‘know’ about future events.
  - Example: topics that put weight on words unique to one crisis or time period
- Methods restricted to employ single in-sample (train) and out-of-sample (test) period. Problem: not feasible to use for forecasts in practice.
  - Example: topics cannot be joined together over time when model re-estimated.

Information leakage and text forecasting pitfalls (aka this is hard, be careful) ▸ Back

- Terms used from entire corpus based on, eg, a frequency threshold (5–95% quantiles). Problem: terms that suddenly appear at one point in time correlated with big macro developments.
  - Example: 'sub-prime'
- Topic model trained on entire corpus, topics deployed to forecast text earlier in time. Problem: topics that 'know' about future events.
  - Example: topics that put weight on words unique to one crisis or time period
- Methods restricted to employ single in-sample (train) and out-of-sample (test) period. Problem: not feasible to use for forecasts in practice.
  - Example: topics cannot be joined together over time when model re-estimated.
- Dictionary or boolean method created with benefit of hindsight. Problem: great for that particular event/crisis, not so good for future ones.
  - Example: algorithmic indicator counts "dodd-frank", "bank stress test"

Information leakage and text forecasting pitfalls (aka this is hard, be careful) ▸ Back

- Terms used from entire corpus based on, eg, a frequency threshold (5–95% quantiles). Problem: terms that suddenly appear at one point in time correlated with big macro developments.
    - Example: ‘sub-prime’
- Topic model trained on entire corpus, topics deployed to forecast text earlier in time. Problem: topics that ‘know’ about future events.
    - Example: topics that put weight on words unique to one crisis or time period
- Methods restricted to employ single in-sample (train) and out-of-sample (test) period. Problem: not feasible to use for forecasts in practice.
    - Example: topics cannot be joined together over time when model re-estimated.
- Dictionary or boolean method created with benefit of hindsight. Problem: great for that particular event/crisis, not so good for future ones.
    - Example: algorithmic indicator counts “dodd-frank”, “bank stress test”
- Text is processed with global transforms. Problem: future mean/std dev/frequency different.
    - Example: Computing tf-idf with idf based on entire corpus frequency of appearance.

Forecasting environment

Models are re-estimated at every step, and indexed by μ, with $\mu = 1, \ldots, T - \alpha - 1$ and $t = 0, \ldots, T$. Data transforms avoid information leakage - i.e. they are performed with in-sample data only. μth in-sample I and out-of-sample O data given by:

$$I_\mu(\vec{z}) = \{z_t\}_{t=\mu-1}^{t=\mu+\alpha-1}$$

$$O_\mu(\vec{z}) = \{z_t\}_{t=\mu+\alpha}^{t=T}$$

Shown results are union of last period of rolling window in-sample forecasts (in-sample), and union of first period out-of-sample rolling window forecasts (out-of-sample) given by:

$$\mathcal{I} = \bigcup_\mu \left\{ f_\mu \left(I_\mu(X) \right) \right\}_{t=\mu+\alpha-1}$$

$$\mathcal{O} = \bigcup_\mu \left\{ f_\mu \left(O_\mu(X) \right) \right\}_{t=\mu+\alpha}$$

Time index on matrix of features X is implicit. Back

Baseline specification for forecasts with algorithms

Evaluate forecasting power of each text metric, x_t, in turn using the model

$$y_{t+h} = \gamma + \beta \cdot y_{t-1} + \eta \cdot x_t + \epsilon_t$$

versus

$$y_{t+h} = \gamma + \beta \cdot y_{t-1} + \epsilon_t$$

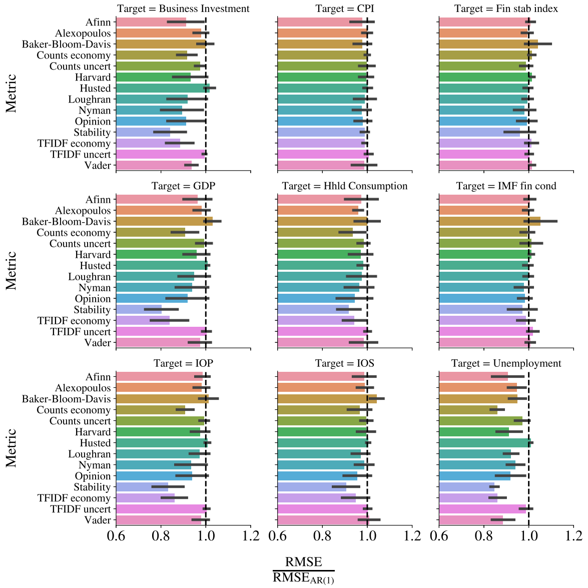
Targets are GDP, unemployment rate, business investment (quarterly), household consumption (quarterly), consumer price inflation (CPI), index of production (IOP), the index of services (IOS), the financial stress index of Chatterjee et al. (2017), and the IMF financial conditions index. We use a rolling window of 36 months for fitting, and time horizons of $h = 3, 6, 9$.

▸ Details of forecast environment ▸ Example plot with unemployment

General idea: information on y_t not available at t, but text and y_{t-1} is.

Forecasting with text metrics - AR(1) benchmark (ratio of RMSEs)

Back



Forecasting with simple text metrics

Results suggest:

- can beat simple benchmark on wide range of variables but do best on GDP (and components)
- simple and transformed counts (e.g. tf-idf on economy) do surprisingly well
- Stability dictionary method from Correa et al. (2017) strongest overall method
- uncertainty measures not effective inputs into forecasts, not so convincing as indicators either ▸ Uncertainty index correlations

Forecasting with simple text metrics

Results suggest:

- can beat simple benchmark on wide range of variables but do best on GDP (and components)
- simple and transformed counts (e.g. tf-idf on economy) do surprisingly well
- Stability dictionary method from Correa et al. (2017) strongest overall method
- uncertainty measures not effective inputs into forecasts, not so convincing as indicators either ▸ Uncertainty index correlations
- ...but results do **not** persist with richer model that incorporates more information

Out-of-sample forecasting with text metrics - unemployment example

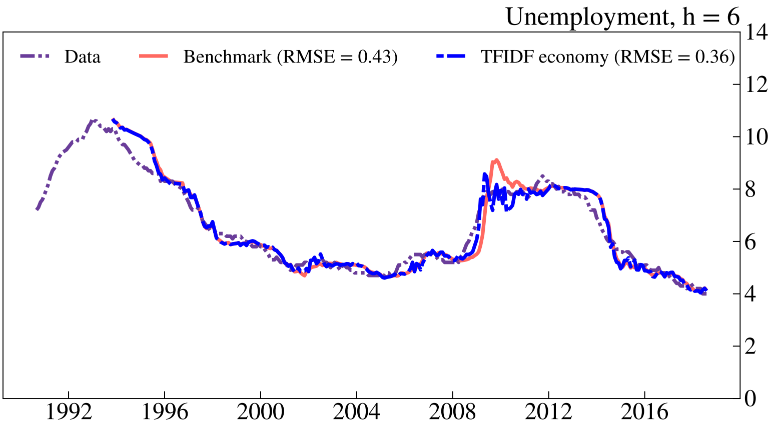


Figure 6: OLS based out-of-sample forecasts for unemployment 6 months ahead using the text metric TFIDF economy + one lag versus one lag alone. Text metric is mean across papers. Rolling window. Back.

Forecasting with text and an artificial neural network

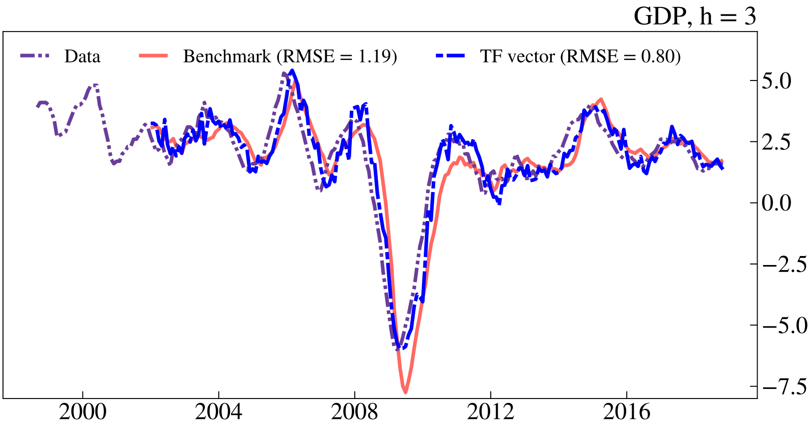


Figure 7: Artificial neural network based out-of-sample forecasts for GDP 3 months ahead using text plus one lag of the target versus one lag of the target alone in an OLS model. Daily Mail.

▸ Back

Baseline specification for out-of-sample forecasts with machine learning

Evaluate forecasting power of each model f_{ML} in turn using the vector of term-frequencies $\overrightarrow{\text{tf}}_t$

$$y_{t+h} = f_{\text{ML}}(y_{t-1}, \overrightarrow{\text{tf}}_t) + \epsilon_t$$

versus

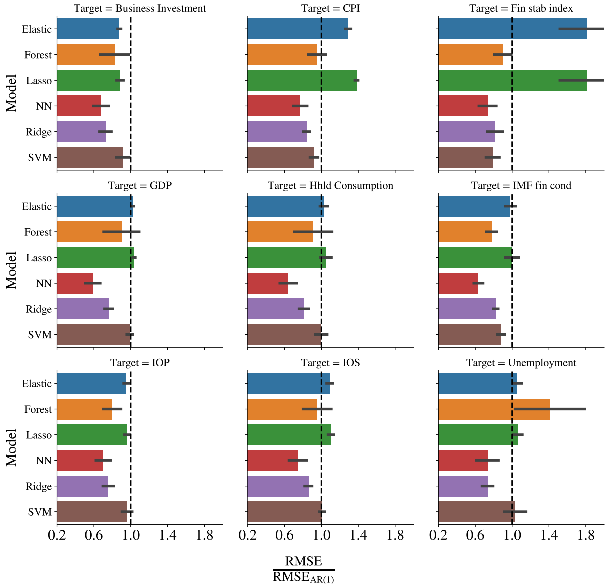
$$y_{t+h} = \gamma + \beta \cdot y_{t-1} + \epsilon_t$$

As before, a rolling window of $\alpha = 36$ months for fitting, and time horizons of $h = 3, 6, 9$.

Back

Forecasting with ML: Ratio of RMSEs

Back



Alternative ML specification I

Evaluate forecasting power of each model f_{ML} in turn using the vector of term-frequencies $\overrightarrow{\text{tf}}_t$

$$y_{t+h} = f_{\text{ML}}(y_{t-1}, \overrightarrow{\text{tf}}_t) + \epsilon_t$$

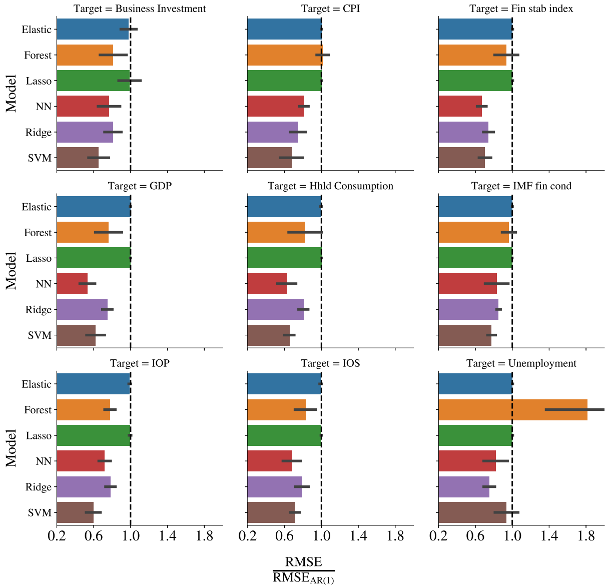
versus

$$y_{t+h} = f_{\text{ML}}(y_{t-1}) + \epsilon_t$$

As before, a rolling window of $\alpha = 36$ months for fitting, and time horizons of $h = 3, 6, 9$.

▸ Back

Alternative ML specification I: Ratio of RMSEs



Alternative ML specification I: Diebold-Mariano test

Paper	Model	Target Horizon	Business Investment	CPI	Fin stab index	GDP	Hhld Consumption	IMF fin cond	IOP	IOS	Unemployment
The Daily Mail	Forest	6	-1.68*							-2.17**	
		9	-2.21**			-1.67*	-2.24**				
	Lasso	9									-2.32**
	NN	3	-2.05**	-1.75*	-2.00**	-2.18**	-2.04**		-1.68*	-1.97*	
		6	-1.91*	-2.18**	-1.89*		-1.90*				
		9		-2.22**	-1.72*	-1.78*	-2.09**				
	Ridge	3		-1.91*	-1.91*	-2.18**	-2.39**	-1.77*	-2.12**	-3.38***	-1.89*
		6	-1.94*	-2.19**	-1.77*	-2.48**	-1.96*	-2.09**	-1.74*	-3.57***	-2.27**
		9	-2.15**	-2.29**		-3.15***	-2.08**	-1.69*		-4.17***	-2.50**
	SVM	3	-2.01**			-2.21**	-2.46**	-1.71*	-3.54***	-2.85***	
		6	-3.79***	-1.85*					-3.19***	-3.48***	
		9	-3.95***				-1.68*	-2.34**	-2.30**	-2.84***	
The Daily Mirror	Forest	6								-2.20**	
		9	-2.02**				-2.28**		-1.69*	-1.67*	
	NN	3	-1.99**		-2.45**	-1.79*	-1.76*		-1.71*	-2.04**	
		6		-2.20**	-1.83*		-1.87*			-1.69*	
		9	-2.11**	-2.14**	-1.86*	-1.76*	-1.95*			-1.76*	
	Ridge	3		-1.70*	-1.74*	-1.67*			-1.79*	-2.17**	-2.41**
		6	-1.75*	-2.32**	-1.72*	-2.42**	-1.97*			-2.95***	-2.52**
		9	-2.21**	-2.15**		-4.48***	-2.10**			-3.18***	-2.58**
	SVM	3	-5.67***			-2.19**	-2.23**		-3.64***	-2.86***	
		6	-3.78***	-1.87*					-2.78***	-3.45***	
		9	-4.08***					-2.44**	-2.36**	-2.52**	
The Guardian	Elastic	9					-1.95*				
	Forest	6								-2.14**	
		9				-1.72*	-2.36**		-1.90*	-1.79*	
	NN	3			-2.31**	-1.97**	-1.85*			-2.11**	-2.16**
		6		-1.80*	-2.05**	-1.72*	-1.84*			-1.76*	
		9	-2.19**	-2.20**	-1.78*	-1.83*	-2.13**		-1.98**	-1.68*	
	Ridge	3		-2.31**		-1.99**	-1.94*	-1.86*	-1.90*	-2.88***	-3.11***
		6		-2.28**		-2.60***	-1.87*	-1.74*	-1.72*	-3.79***	-2.80***
		9		-2.16**		-3.26***	-2.10**			-4.04***	-2.74***
	SVM	3				-2.24**	-2.73***		-3.54***	-2.58**	-1.72*
		6	-2.11**	-1.79*			-1.83*		-3.23***	-3.72***	
		9		-1.91*				-3.11***	-2.35**	-3.06***	-1.93*

Table 6: Statistically significant differences in RMSE from a Diebold-Mariano test (TF-IDF, lag, and ML). *, **, *** denote rejection of the null at the 10%, 5%, 1% levels respectively. Only targets for which at least one of the models had a p-value of less than 10% are shown.

Alternative ML specification II

Richer model with highly statistically significant confounders. Macroeconomic factors $\vec{F}$ derived from 33 series covering real output, international trade, the labour market, inflation, house prices, retail sales, capacity utilisation, and business and household expectations (Redl, 2017):

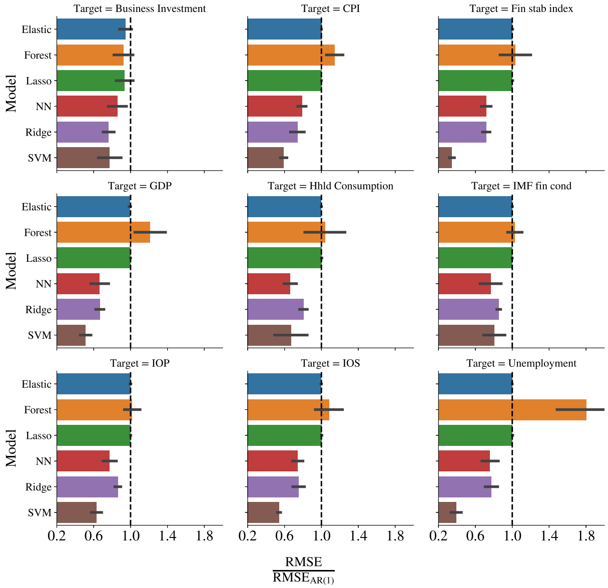
$$y_{t+h} = f_{\mathrm{ML}}(y_{t-1}, \vec{F}_t, \overrightarrow{\mathrm{tf}}_t) + \epsilon_t$$

versus

$$y_{t+h} = f_{\mathrm{ML}}(y_{t-1}, \vec{F}_t) + \epsilon_t$$

$J = 2$ factors. As before, a rolling window of $\alpha = 36$ months for fitting, and time horizons of $h = 3, 6, 9$. ▸ Back

Alternative ML specification II: Ratio of RMSEs



Alternative ML specification II: Diebold-Mariano test

Paper	Model	Target Horizon	Business Investment	CPI	Fin stab index	GDP	Hhld Consumption	IMF fin cond	IOP	IOS	Unemployment
The Daily Mail	NN	3		-2.18**	-2.27**	-2.21**	-2.67***	-1.91*	-2.35**	-3.33***	-2.82***
		6		-3.06***	-1.68*	-2.81***	-3.22***		-2.00**	-3.06***	-2.63***
		9	-1.93*			-1.93*	-2.45**	-1.90*	-2.04**	-2.59**	-2.43**
	Ridge	3	-2.39**	-2.01**	-1.85*	-3.18***	-2.37**	-1.94*	-2.17**	-3.51***	-2.27**
		6	-2.15**	-2.21**		-3.11***	-2.64***	-1.85*		-2.99***	-3.77***
		9	-2.29**	-2.25**		-3.18***	-2.85***	-1.76*		-3.47***	-3.83***
	SVM	3	-3.52***	-2.91***	-3.95***	-3.13***	-4.99***	-1.76*	-3.81***	-3.42***	-5.00***
		6	-2.11**	-2.27**	-2.28**	-3.54***	-3.85***		-4.08***	-3.66***	-2.83***
		9		-3.53***	-2.26**	-3.65***			-2.56**	-2.51**	-2.22**
The Daily Mirror	Elastic	3	-1.83*	-1.70*							
	Forest	9					-2.07**				
	Lasso	6									-3.95***
	NN	3		-2.13**	-2.34**	-2.63***	-2.16**	-2.07**	-2.18**	-1.79*	-3.85***
		6	-2.26**	-2.27**	-1.90*	-2.45**	-2.26**		-2.50**	-2.25**	-2.51**
		9	-3.26***	-1.67*		-1.88*	-2.05**	-1.97*	-2.51**	-2.15**	-2.33**
	Ridge	3	-2.74***	-1.80*		-2.93***	-1.66*	-1.95*	-1.75*	-3.10***	-2.32**
		6	-2.15**	-2.47**		-2.82***	-2.14**			-3.27***	-3.30***
		9	-1.88*	-2.13**		-2.88***	-2.44**			-3.45***	-2.81***
	SVM	3	-2.55**	-2.91***	-3.90***	-3.10***	-5.54***		-3.62***	-3.06***	-4.87***
		6	-1.81*	-1.80*	-2.40**	-3.59***	-3.71***		-3.98***	-3.59***	-2.68***
		9		-3.47***	-2.31**	-4.18***			-2.29**	-2.22**	-2.08**
The Guardian	Elastic	3				-2.00**				-1.83*	
		6								-1.66*	
		9				-1.73*				-2.03**	
	NN	3	-2.09**	-1.75*	-2.23**	-1.98**	-2.69***			-2.17**	-2.35**
		6		-2.80***		-2.62***	-2.77***		-1.87*	-2.43**	-2.54**
		9		-2.31**		-2.30**	-2.76***		-1.94*	-2.17**	-2.48**
	Ridge	3		-1.89*		-3.51***	-2.10**	-2.04**	-1.84*	-3.81***	-2.83***
		6		-2.27**		-3.17***	-2.66***			-3.28***	-3.52***
		9		-2.12**		-3.06***	-2.78***			-3.23***	-3.50***
	SVM	3		-3.83***	-4.00***	-3.19***	-3.71***	-1.83*	-3.74***	-3.73***	-6.12***
		6		-2.53**	-2.27**	-3.74***	-4.03***		-4.21***	-3.95***	-3.23***
		9		-3.87***	-2.61***	-4.08***			-2.49**	-2.59**	-2.47**

Table 7: Statistically significant differences in RMSE from a Diebold-Mariano test (TF-IDF, lag, and factors). *, **, *** denote rejection of the null at the 10%, 5%, 1% levels respectively. Only targets for which at least one of the models had a p-value of less than 10% are shown.

Diebold-Mariano tests for ML-OLS AR(1) model benchmark

Paper	Model	Target Horizon	Business Investment	CPI	Fin stab index	GDP	Hhld Consumption	IOP	IOS	Unemployment
The Daily Mail	Elastic	3	-2.65***							
		6	-2.25**							
		9	-2.22**							
	Forest	6	-1.73*			-1.97*			-1.74*	
		9	-2.46**			-1.66*	-1.98**	-1.85*		
	Lasso	6	-1.97*							
		9	-1.92*							
	NN	3	-2.98***	-2.39**		-2.65***	-2.23**		-1.66*	
		6	-2.48**	-2.88***		-1.86*	-1.88*			-2.20**
		9	-2.32**	-2.57**		-1.72*	-1.81*			-2.09**
	Ridge	3	-2.96***	-1.74*		-2.50**	-1.94*	-2.20**	-2.09**	-1.91*
		6	-2.51**	-2.14**		-1.83*	-1.67*	-1.78*		-2.25**
		9	-2.52**	-1.85*			-1.81*	-1.70*		-2.05**
	SVM	3	-2.34**	-2.18**						
		6		-2.56**						
		9		-2.43**						
The Daily Mirror	Elastic	3	-2.03**							
		6	-1.84*							
		9	-1.81*							
	Forest	6	-2.14**			-1.99**				
		9	-2.23**	-1.72*			-1.72*	-1.71*		
	Lasso	3	-2.05**							
		6	-1.78*							
		9	-1.80*							
	NN	3	-2.59**	-1.73*		-2.28**	-1.83*	-1.71*		
		6	-2.47**	-3.02***	-1.71*	-1.85*	-2.01**		-1.65*	-2.53**
		9	-2.82***	-2.79***		-1.73*				-1.80*
	Ridge	3	-2.71***	-2.20**		-2.11**		-1.83*	-1.77*	
		6	-2.63***	-1.92*		-1.67*	-1.84*			-1.90*
		9	-2.40**				-1.68*			-1.69*
	SVM	6		-1.86*						
		9		-1.94*						
The Guardian	Elastic	3	-1.94*							
		6	-1.86*							
		9	-1.72*							
	Forest	9					-1.95*			
	Lasso	3	-1.80*							
	NN	3	-3.19***	-1.74*			-1.84*			-2.14**
		6		-2.46**	-1.76*		-2.01**			-2.19**
		9		-2.78***	-1.76*		-1.82*			-2.03**
	Ridge	3	-2.05**	-2.35**						-2.49**
		6	-1.85*	-2.18**			-1.76*			-2.05**
		9		-2.20**			-1.92*			-1.88*
	SVM	6		-1.93*	-1.74*					
		9		-2.66***						

Table 8: Diebold-Mariano tests on forecasts using term frequency vectors with a lag of the target variable versus a lag alone within an OLS model. Statistically significant differences in RMSE are shown. *, **, *** denote rejection of the null at the 10%, 5%, 1% levels respectively.

Back

Diebold-Mariano tests for ML-OLS factor model benchmark

Paper	Model	Target Horizon	Business Investment	CPI	Fin stab index	GDP	Hhld Consumption	IMF fin cond	IOP	IOS	Unemployment
The Daily Mail	Forest	9	-2.26**				-2.27**				
	NN	3		-2.07**			-2.09**	-1.80*			
		6	-1.77*	-1.99**	-1.76*	-2.36**	-2.07**				-1.93*
		9	-1.85*	-2.06**		-2.24**	-2.19**			-2.17**	-2.11**
	Ridge	3	-2.51**				-2.21**	-1.85*	-2.04**		
		6	-2.20**	-1.98**	-1.77*		-2.06**				
		9	-2.79***				-2.64***			-1.71*	
	SVM	3	-1.96*		-1.79*		-1.88*	-2.10**			
		6			-1.90*						
		9		-2.16**							-2.03**
The Daily Mirror	Forest	6			-2.01**						
		9					-2.13**				
	NN	3		-1.84*				-1.68*			
		6		-1.87*	-1.66*	-2.70***	-2.04**			-2.08**	-2.19**
		9	-1.87*	-2.14**		-1.96*	-1.86*			-1.90*	-1.88*
	Ridge	3	-1.80*					-1.79*	-1.80*		
		6	-3.28***				-1.76*				
		9	-2.31**				-2.05**				
	SVM	9		-1.81*							
The Guardian	Elastic	3	-1.86*								
	Forest	9					-2.07**				
	Lasso	3	-1.81*								
	NN	3	-1.89*								
		6		-2.21**	-1.82*		-1.92*				-1.76*
		9		-2.12**			-2.21**				-2.66***
	Ridge	3	-2.17**	-1.89*				-1.91*			-2.28**
		6	-1.76*	-1.81*	-1.79*		-2.15**				
		9	-1.65*	-1.89*			-2.73***				
	SVM	3	-2.37**					-1.77*			
		9		-2.65***							-2.18**

Table 9: Statistically significant differences in RMSE from a Diebold-Mariano test (TF-IDF, ML, factors, and lags). *, **, *** denote rejection of the null at the 10%, 5%, 1% levels respectively. Only those targets for which at least one model-newspaper pair had a p-value of less than 10% are included.

▸ Back

References

Alexopoulos, Michelle, and Jon Cohen. 2009. "Uncertain times, uncertain measures." *University of Toronto Department of Economics Working Paper*, 352.

Alexopoulos, Michelle, and Jon Cohen. 2015. "The power of print: Uncertainty shocks, markets, and the economy." *International Review of Economics & Finance*, 40: 8–28.

Antweiler, W., and M. Z. Frank. 2004. "Is all that talk just noise? The information content of internet stock message boards." *Journal of Finance*, 59(3): 1259–1294.

Ardia, David, Keven Bluteau, and Kris Boudt. 2019. "Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values." *International Journal of Forecasting*.

References ii

Baker, Scott R, Nicholas Bloom, and Steven J Davis. 2016. “Measuring economic policy uncertainty.” *The Quarterly Journal of Economics*, 131(4): 1593–1636.

Bird, Steven, and Edward Loper. 2004. “NLTK: the natural language toolkit.” 31, Association for Computational Linguistics.

Chatterjee, Somnath, Ching-Wai (Jeremy) Chiu, Sinem Hacioglu-Hoke, and Thibaut Duprey. 2017. “A financial stress index for the United Kingdom.” Bank of England Staff Working Paper 697.

Clark, Kevin, and Christopher D Manning. 2016. “Deep reinforcement learning for mention-ranking coreference models.” *arXiv preprint arXiv:1609.08667*.

Correa, Ricardo, Keshav Garud, Juan M Londono, Nathan Mislang, et al. 2017. “Constructing a Dictionary for Financial Stability.” Board of Governors of the Federal Reserve System (US).

Elango, Pradheep. 2005. “Coreference resolution: A survey.” *University of Wisconsin, Madison, WI*.

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

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2017. “Text as data.” National Bureau of Economic Research.

Gilbert, CJ Hutto Eric. 2014. “Vader: A parsimonious rule-based model for sentiment analysis of social media text.”

Hu, Guoning, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic. 2017. “Analyzing Users’ Sentiment Towards Popular Consumer Industries and Brands on Twitter.” 381–388, IEEE.

Hu, Minqing, and Bing Liu. 2004. "Mining and summarizing customer reviews." 168–177, ACM.

Husted, Lucas F., John Rogers, and Bo Sun. 2017. "Monetary Policy Uncertainty." Board of Governors of the Federal Reserve System (U.S.) International Finance Discussion Papers 1215.

Jegadeesh, Narasimhan, and Di Wu. 2013. "Word power: A new approach for content analysis." *Journal of Financial Economics*, 110(3): 712–729.

Jurado, Kyle, Sydney C Ludvigson, and Serena Ng. 2015. "Measuring uncertainty." *The American Economic Review*, 105(3): 1177–1216.

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

Loughran, Tim, and Bill McDonald. 2011. “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks.” *The Journal of Finance*, 66(1): 35–65.

Loughran, Tim, and Bill McDonald. 2013. “IPO first-day returns, offer price revisions, volatility, and form S-1 language.” *Journal of Financial Economics*, 109(2): 307–326.

Nielsen, Finn Årup. 2011. “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.” *arXiv preprint arXiv:1103.2903*.

Nothman, Joel, Hanmin Qin, and Roman Yurchak. 2018. “Stop Word Lists in Free Open-source Software Packages.” 7–12.

Nyman, Rickard, Sujit Kapadia, David Tuckett, David Gregory, Paul Ormerod, and Robert Smith. 2018. “News and narratives in financial systems: exploiting big data for systemic risk assessment.” *Bank of England Staff Working Papers*, 704.

Puurula, Antti. 2013. "Cumulative progress in language models for information retrieval." 96–100.

Redl, Chris. 2017. "The impact of uncertainty shocks in the United Kingdom." Bank of England Bank of England Staff Working Papers.

Redl, Chris. 2018. "Uncertainty matters: evidence from close elections." Bank of England Bank of England Staff Working Papers.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel Wilson. 2018. "Measuring news sentiment." Federal Reserve Bank of San Francisco.

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

Thorsrud, Leif Anders. 2018. "Words are the new numbers: A newsy coincident index of the business cycle." *Journal of Business & Economic Statistics*, 1–17.