Forecasting UK GDP growth with large survey panels

Niki Anesti[1], **Eleni Kalamara**[2] **and George Kapetanios**[2]

Bank of England[1], King's College London and QCGBF[2]

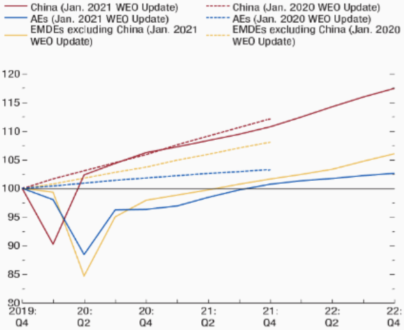
Second joint European Central Bank-Federal Reserve Bank of New York conference on expectations surveys: a tool for research and monetary policy.

Virtual, 11. November 2021

[*] Disclaimer: The expressed views are our own and not necessarily those of the Bank of England or its policy committees.

Motivation

IMF: World Economic Outlook Update (Jan 21')



Source: IMF staff estimates.

Notes: AE:Advanced Economies,EMDES: Emerging Market and Developing Economies, WEO, World Economic Outlook.

- Central banks face difficulties in monitoring movements of the macro-economy. Current pandemic shock is a prominent example.
- Critical juncture of two mega-trends.
    - Rapidly increasing amount of granular data, often referred to as "alternative" (e.g **surveys**, **textual data**, financial transactions, internet searches).
    - Large advances in data-driven modelling techniques, e.g Machine learning (ML, henceforth).

Existing evidence towards this direction

- Economic forecasting can benefit from the availability of a wider information set and new techniques drawn from the ML literature (see e.g. Chakraborty and Joseph (2017); Larsen (2017)).
- Even though ML methods have been used in the past (Swanson and White, 1997; Nakamura, 2005), it's only recently that this trend re-emerged, partly due to data explosion (Smeekes and Wijler, 2018; Coulombe et al., 2020; Medeiros et al., 2021).
- Long-lasting literature on macroeconomic forecasting using surveys, see e.g. Giannone et al. (2008); Bańbura et al. (2013); Anesti et al. (2017)
- A number of recent studies show that textual data can answer economic related questions (Bholat et al., 2015) and can be particularly helpful in monitoring economic conditions (Thorsrud, 2016; Kalamara et al., 2020, 2021).

What we do

What we do

- We explore the informational content of an extensive dataset of disaggregate business & consumer survey balances.
- We compare the predictive content of surveys with text based indicators from newspaper articles.
- But also with a more traditional macroeconomic dataset in the spirit of Stock and Watson (2002a).
- Use a battery of linear and non-linear modelling approaches to forecast the monthly GDP growth estimate (official statistic) for the UK.

What we do

Forecastig Methods used: a wide range of models that can deal with large and alternative data sets:

- Dimensionality reduction techniques: PCR (as benchmark here), PLS
- Shrinkage methods: Ridge Regression, LASSO, Elastic Net
- Non-Linear Machine Learning Models: Support Vector Regressions (SVR), Random Forests, Artificial Neural Networks (ANN) SVM RF ANN

Hyperparameter tuning via K-fold cross-validation.

- in-sample data devided into $k = 5$ folds, training based on 4 folds, testing on 5th (avoids correlation between training and testing instances)

Key Findings

What we find

- Largest forecasting gains are overwhelmingly concentrated at the shorter horizons for the majority of models and datasets.
- Text indicators offer fairly similar predictive content to this of survey balances.
- Macroeconomic time series do not appear to add much more predictive power.
- Ridge and PLS models report the largest gains consistently for most of the forecasting horizons.
- Non-linear machine learning models appear to be more useful during the Great Recession.

Data

Surveys

- Monthly disaggregate survey balances from businesses and consumers.
- Questions related to the *current and future state of the economy*, *employment prospects*, *level of sales*, *volume of exports*, *current and future household financial conditions*,
- Surveys from CBI, IHS Markit/CIPS, Lloyds, EC Consumer Confidence at monthly frequency,
- 96 time series in total starting in Jan 2000 until Aug 2018,

Text Data

- We use 15 text-based indicators introduced by Kalamara et al. (2020) that aim to capture uncertainty and sentiment in the UK economy.
- Raw data: news articles from the “the Guardian” , “ Daily Mail” and “Daily Mirror ” .
- Total of 5m articles covering a period from Jan 2000 to Aug 2018.
- Include only economic, financial and corporate articles to reduce noise.
- Text methods: Dictionary - based approaches.

See all approaches

Macro Data

- Test a dataset with "hard" indicators.
- 49 macroeconomic series selected in the spirit of (Stock and Watson, 2002a,b)
- Mostly data on production, services, prices and labour market statistics, but also some financial market data taken as monthly averages.
- Target variable: the monthly GDP estimate as published by the Office of National Statistics, transformed to three month-on-three month growth rate. All series are seasonally adjusted.

Forecasting environment

Evaluation design

- Estimation: Use data from Jan 2000 until Aug 2006,
- Evaluation: Direct out-of-sample forecasts for $h = 1, 3, 6, 9, 12, 24$ at every month starting at Sep 2006 using an expanding window.
- Benchmarks: PCR and AR(1)
- Baseline Forecasts:

$$\hat{y_{t+h}} = g\left(x_t, \theta, \eta\right),$$

where $\hat{y}_{t+h}$ is the forecast of GDP growth at h horizon; x_t is a vector of predictors; θ is a vector of estimated parameters; and η is a vector of hyperparameters used to train the model. The function $g(\cdot)$ varies with the model at hand.

Results

A first-pass using survey data

PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.991	0.977	0.961	0.940	0.950	0.990
Ridge	0.726^{***}	0.824^{*}	0.846	0.891	0.935	1.104
Elastic	0.817^{***}	0.843^{*}	0.843	0.868	0.908	1.065
PLS	0.826^{***}	0.841^{*}	0.840	0.861	0.899	1.068
Random Forest	0.879^{***}	0.901	0.856	0.885	0.914	1.045
SVM	0.722^{***}	0.859^{*}	0.872	0.889	0.920	1.074
NN	0.768	0.874	0.869	0.900	0.937	1.103
AR(1)						
Lasso	0.720^{***}	0.826^{*}	0.846	0.896	0.937	1.109
Ridge	0.702^{***}	0.820	0.890	0.969	1.011^{*}	0.975
Elastic	0.726^{***}	0.716^{*}	0.758	0.907	0.993	0.965
PLS	0.792^{***}	0.789	0.830	0.939	0.998	0.976
Random Forest	0.751^{***}	0.770	0.835	0.877	0.972	0.972
SVR	0.625^{***}	0.850	0.891	0.970	1.010	0.985
NN	0.687^{***}	0.895	0.923	0.986	1.011	0.968

Notes:
PCR model is a harder benchmark and thus it is used as the main competitor.
Significant forecast gains are mostly concentrated at $h = 1$.
For both panels gains are getting smaller after the first two quarters.
On average, linear models consistently outperform the non-linear models.

Testing Text and Macro data

Text Data	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.902*	0.914	0.896	0.898	0.918	1.033
Ridge	0.904*	0.914	0.897	0.898	0.918	1.033
Elastic	0.986	0.975	0.959	0.940	0.950	0.990
PLS	0.937***	0.932	0.924	0.916	0.932	1.012
Random Forest	0.930***	0.941	0.936	0.927	0.940	1.014
SVM	0.823***	0.882*	0.904	0.912	0.939	1.040
NN	0.879***	0.904	0.950	0.903	0.924	1.056
Macro Dataset						
Lasso	0.959***	0.956	0.940	0.931	0.945	0.997
Ridge	0.855***	0.884	0.863	0.882	0.896	1.065
Elastic	0.946***	0.947	0.932	0.925	0.940	1.003
PLS	0.885***	0.901	0.876	0.888	0.899	1.045
Random Forest	0.933***	0.947	0.920	0.921	0.938	1.020
SVR	0.722***	0.886	0.872	0.896	0.929	1.118
NN	0.802***	0.895*	0.928	0.897	0.934	1.125

Notes:
Predictive content is fairly similar to this of survey balances...
but text maintains some of its gains for longer horizons up to one-year ahead.
However, both of this type of data favour non-linear ML models at shorter horizons.
On average, the macro-only forecasts do not outperform the forecasts of text data.

Does combined information from surveys and text provide better forecasts?

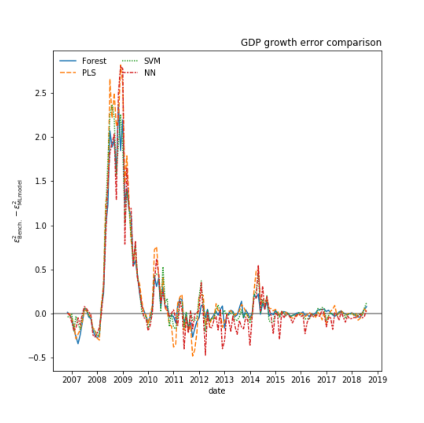
Does combined information from surveys and text provide better forecasts?

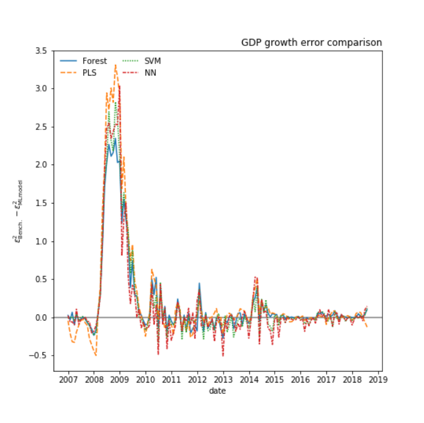
RELATIVE TO PCR	(1)	(3)	(6)	(9)	(12)	(24)
Lasso	0.992	0.977	0.960	0.940	0.950	0.990
Ridge	0.698***	0.831*	0.855	0.901	0.946	1.097
Elastic	0.805***	0.837*	0.844	0.873	0.913	1.062
PLS	0.818***	0.834*	0.840	0.867	0.909	1.057
Random Forest	0.794***	0.882	0.873	0.882	0.916	1.044
SVR	0.663	0.874	0.865	0.895	0.936	1.125
NN	0.667	0.851	0.878	0.903	0.952	1.148

Notes: The combination of survey balances with text indicators does not add much predictive content. The general pattern of the relative RMSFEs is closer to that of the forecasting exercise when using just the survey balances.

Are machine learning models more useful around recessionary episodes?

Are machine learning models more useful around recessionary episodes?





Average mean squared error differences between the PCR benchmark and other models over h-month ahead out-of-sample forecasts, where h = 1,3,6,9,12,24. A line above zero means that the relative model produces smaller errors than the benchmark model.
Right figure: results with Survey data while left figure: results with a combination of text, survey and macro series.
Largest forecast errors both in absolute and in relative terms during the Great Recession.
ML models are more appropriate to capture large non-linearities in the data compared to their linear counterparts.
Generalised finding for all the datasets and their combinations.

Conclusions

Take away message

- Survey and text-based forecasts perform similarly, however survey-only models are marginally better at shorter horizons and the text-only models can maintain some of their gains for longer horizons up to one-year ahead.
- Macroeconomic time series do not appear to add much more predictive power.
- Among the linear models, the Ridge and the PLS models report the largest gains consistently for most of the forecasting horizons.
- For the non linear ML models, the SVR performs better at shorter horizons compared to the NN and Random forest. Most improvements up to one year ahead.
- Non-linear ML models appear to be more useful during the Great Recession.

Thanks for listening

Thanks for listening!

In more depth: Machine learning methods - Random forests

- tree models consecutively split the training dataset until an assignment criterion with respect to the target variable into a "data bucket" (leaf) is reached
  - algorithm minimises objective function within "buckets", conditioned on input x_t
  - sparse models: only variables which actually improve the fit are chosen

The regression function is

$$y_{t+h} = \sum_{m=1}^{M} \beta_m I(x_t \in P_m) + \varepsilon_t, \quad \text{with} \quad \beta_m = 1/|P_m| \sum_{y^{tr} \in P_m} y^{tr}, \; m \in \{1, \ldots, M\}. \quad (1)$$

- A random forest contains a set of *uncorrelated trees* which are estimated separately
  - this overcomes overfitting of standard tree models
  - but also harder to interpret due to the built-in randomness back

In more depth: Machine learning methods - Artificial Neural Networks (ANN)

- Standard architecture: multilayer perceptrons (MLP), i.e. a feed-forward network
  - can be viewed as alternative statistical approach to solving the least squares problem, but a hidden layer is added
  - predictors x_t in the input layer are multiplied by weight matrices, then transformed by an activation function in the first hidden layer and passed on to the next hidden or the output layer resulting a prediction y_t.

$$y_{t+H} = G(x_t, \beta) + \varepsilon = g_L(g_{L-1}(g_{L-2}(\dots g_1(x_t, \beta_0), \dots, \beta_{L-2}), \beta_{L-1}), \beta_L) + \varepsilon \quad (2)$$

  - activation function $g(\cdot)$ introduces non-linearity into the model. We use rectified linear unit functions (ReLU)

- Number of layers L, the number of neurons in each layer and appropriate weight penalisation are determined by cross-validation. Deeper networks being generally more accurate but also needing more data to train them. back

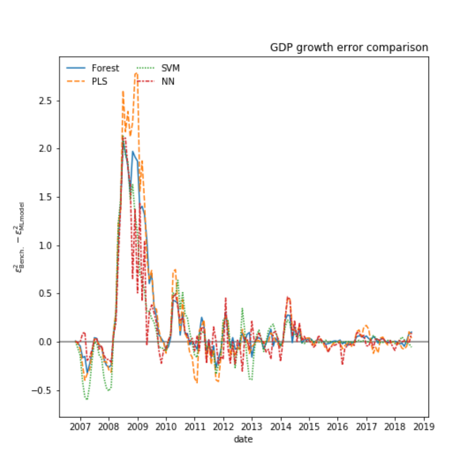
Text analysis methods

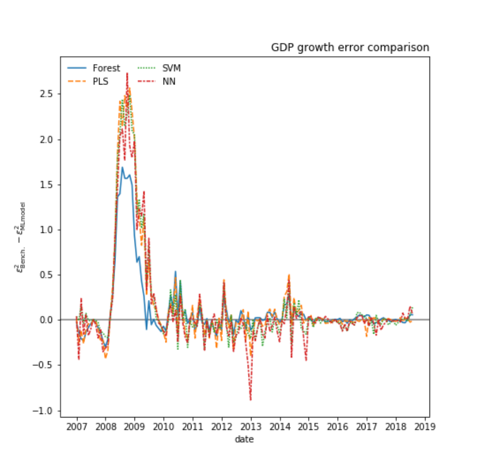
Positive and negative dictionary	Boolean	Computer science-based
Financial stability (Correa et al., 2017)	Economic Uncertainty (Alexopoulos et al., 2009)	VADER sentiment (Gilbert, 2014)
Finance oriented (Loughran and McDonald, 2013)	Monetary policy uncertainty (Husted et al., 2017)	'Opinion' sentiment (Hu and Liu, 2004; Hu et al., 2017)
Afinn sentiment (Nielsen, 2011)	Economic Policy Uncertainty (Baker et al., 2016)	Punctuation economy (this paper)
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 1: The three broad categories of algorithm-based text metrics used.

▸ Go back

Surveys and Text combined and Macro





j

References i

Alexopoulos, M., Cohen, J., et al. (2009). Uncertain times, uncertain measures. *University of Toronto Department of Economics Working Paper*, 352.

Anesti, N., Hayes, S., Moreira, A., and Tasker, J. (2017). Peering into the present: the Bank's approach to GDP nowcasting. *Bank of England Quarterly Bulletin*, 57(2):122–133.

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

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of Economic Forecasting, volume 2A*, chapter 4, pages 195–237. Elsevier.

Bholat, D. M., Hansen, S., Santos, P. M., and Schonhardt-Bailey, C. (2015). Text mining for central banks.

Chakraborty, C. and Joseph, A. (2017). Machine learning at central banks.

Correa, R., Garud, K., Londono, J. M., Mislang, N., et al. (2017). Constructing a dictionary for financial stability. Technical report, Board of Governors of the Federal Reserve System (US).

Coulombe, P. G., Leroux, M., Stevanovic, D., and Surprenant, S. (2020). How is machine learning useful for macroeconomic forecasting?

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting gdp and inflation: The real-time informational content of macroeconomic data releases. *Journal of Monetary Economics*, 55:665–676.

Gilbert, C. H. E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Hu, G., Bhargava, P., Fuhrmann, S., Ellinger, S., and Spasojevic, N. (2017). Analyzing users' sentiment towards popular consumer industries and brands on twitter.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ?04, page 168?177, New York, NY, USA. Association for Computing Machinery.

Husted, L. F., Rogers, J. H., and Sun, B. (2017). Monetary policy uncertainty.

Kalamara, E., Ashwin, J., and Saiz, L. (2021). Nowcasting euro area gdp with news sentiment: A tale of two crises. *Unpublished manuscript*.

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., and Kapadia, S. (2020). Making text count: economic forecasting using newspaper text. *Bank of England Staff Working Papers*, 865.

Larsen, V. H. (2017). Components of uncertainty.

Loughran, T. and McDonald, B. (2013). Ipo first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics*, 109(2):307–326.

Medeiros, M. C., Vasconcelos, G. F. R., Ãlvaro Veiga, and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, 86(3):373–378.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Nyman, R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P., and Smith, R. (2018). News and narratives in financial systems: exploiting big data for systemic risk assessment. *Bank of England Staff Working Papers*, 704.

Smeekes, S. and Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, 34(3):408–430.

Stock, J. and Watson, M. (2002a). Forecasting using principal components from large numbers of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

References v

Stock, J. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Swanson, N. R. and White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4):540–550.

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

Thorsrud, L. A. (2016). Nowcasting using news topics. big data versus big bank.