Forecasting UK inflation bottom up

Andreas Joseph

Bank of England*

Joint work with Eleni Kalamara (KCL & ECB), George Kapetanios (KCL) and Galina Potjagailo (BoE)

Nontraditional Data & Statistical Learning with Applications to Macroeconomics

BoI & FRB, virtual, 11. November 2020

*Disclaimer: The expressed views are my own and not necessarily those of the Bank of England (BoE) or European Central Bank (ECB). All errors are ours.

Introduction

Motivation

- Inflation forecasts have a large impact on central banks' policy decisions and communications, as well as business decisions in the wider economy
- Central banks increasingly base forecasts on large and granular sets of indicators providing a broad view on price dynamics across sectors
- Advances in data availability and computing power have promoted forecasting tools that incorporate large sets of predictors as well as non-linear features.
    - factor models with medium-sized sets of macroeconomic predictors - provide rather weak forecasting gains (Faust and Wright, 2013)
    - non-linear dynamics within unobserved component models (Stock and Watson, 2016)
    - machine learning tools (Garcia et al., 2017; Chakraborty and Joseph, 2017; Medeiros et al., 2019; Almosova and Andresen, 2019)

What we do

- We forecast UK CPI inflation (headline, core, services) combining granular disaggregated item data with a wide set of forecasting tools
  - Predictors: monthly item-level consumer price indices ($\approx$ 600 items) and macro data
  - Forecasting approaches that exploit large data set in different ways
    - dimensionality reduction techniques: DFM, PCA, PLS
    - shrinkage methods: Ridge regression, LASSO, elastic net
    - non-linear machine learning tools: random forests, SVM, artificial neural nets
- We address black box critique to ML models
  - forecasts from non-linear and non-parametric models not easily attributable to individual predictors
  - we measure the contribution of individual variables to the forecast using Shapley values (Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017)
  - we test predictive power of aggregated components (Joseph, 2019)

Key findings

- Micro item-level data often strongly improve forecasts relative to benchmark (alone or with macro data)
- Ridge regression and ML models perform best, especially at the 1-year horizon
  - ML models particularly good for service core inflation 1-year ahead
  - DFM has more difficulties to cope with disaggregated item indices, but good with macro data
- Initial Shapley value and regressions results indicate a strong role in Food & Beverages, as well as Clothing, Furnishing and Housing in predicting inflation

Where this fits in

- Vast literature on forecasting inflation: Stock and Watson (1999, 2007, 2008); Hubrich (2005); Kapetanios et al. (2008); Koop and Korobilis (2012); Koop (2013); Domit et al. (2019); Carriero et al. (2019); Martins et al. (2020)
- Dynamic factor models successful for nowcasting GDP (Giannone et al., 2008). For inflation, accounting for non-linearities important (Faust and Wright, 2013; Stock and Watson, 2016)
- Machine learning: sizeable forecast gains in forecasting US and Brazilian inflation with neural nets, random forests, and shrinkage methods (Garcia et al., 2017; Almosova and Andresen, 2019; Medeiros et al., 2019)
- Adding disaggregate price information improves forecast accuracy (Hendry and Hubrich, 2011). Use of high-frequency online price item series to forecast CPI (Aparicio and Bertolotto, 2020). Use of CPI item series for Mexico (Ibarra, 2012).

Data and forecasting setup

Setup

- Targets: Monthly UK headline, core and service core inflation
- Predictors: Representative 581 monthly item-level price indices (after cleaning; from ONS) and 46 macroeconomic and financial variables
- Sample period: 2011:M1 - 2019:M12 (no Covid yet, but easy to extend)
  - TS cross-validation & training: Until 2016:M3 (initially; expanding window)
  - Test period: 2016:M4 - 2019:M12
- Forecast horizon: 1 - 12 months
- Benchmark: AR(p) with $p \leq 12$ by BIC

CPI item series

- UK monthly CPI is constructed from about 700 representative item indices by the Office for National Statistics (ONS), publically available
  - Constructed by the ONS from single product prices collected in shops (price quotes)
  - Item indices are weighted and aggregated into classes, groups, divisions, and finally the CPI based on COICOP classification
- Our CPI item data set
  - 581 item indices without missing values, 2011M1-2019M12
  - items cover 84% of total CPI, similar coverage across broad CPI categories
  - we use item indices in levels - ONS rebases and chain-links them to the January value annually, which already removes trends from the series ◂ data
  - indices are mean-variance standardised
  - indices are seasonally adjusted - though a sub-set show additional spikes once a year due to changes in weights. Could ML methods also deal with seasonality?

Forecasting models

We are interested forecasting inflation y_t in period $t+h$, based on the past dynamics of y_t and the set of predictors $x_t = (x_{1t}, \ldots, x_{Nt})\prime$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$

$$\hat{y}_{t+h} = \hat{\alpha} + \Sigma_{i=1}^{N} \hat{\beta} x_t + \Sigma_{j=1}^{p} \hat{\gamma}_j y_{t-j+1}. \tag{1}$$

A wide range of forecasting methods that can deal with large data sets

- Dimensionality reduction techniques: DFM, PCA, PLS
- Shrinkage methods: Ridge Regression, LASSO, Elastic Net
- Non-Linear Machine Learning Models: Support Vector Machines (SVM), 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)

Results

Forecast comparison (I): Headline inflation - CPI item predictors

	Target: headline inflation				
horizon	**1**	**3**	**6**	**9**	**12**
State Space Model (DFM)	0.97	0.92	0.88	0.85	0.93
Principal Components (PCR)	**0.61*****	**0.75***	1.13	1.28**	1.35***
Partial Least Squares (PLS)	1.72***	0.91	1.00	0.90	0.95
Ridge Regression	1.09	0.80	**0.69****	**0.57*****	**0.48*****
LASSO	1.14*	0.97	0.94	0.95	1.03
Elastic Net	0.91	**0.83****	0.85	**0.61*****	**0.67*****
Support Vector Machines (SVM)	1.05	**0.74*****	**0.6*****	**0.61*****	**0.55*****
Random Forest	1.14	1.00	0.79	0.8	**0.69*****
Artificial Neural Net (ANN)	1.09	**0.77****	**0.65****	**0.64****	**0.53*****

Notes: Forecasts using CPI item series as predictors. Root mean squared errors, relative to AR(p) model. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. $***\backslash**\backslash*$ indicates significance at $10\%, 5\%, and 1\%$, respectively. Relative RMSE that are significant at a level of 10% or lower and taking values below 1 are marked in bold. Source: ONS and authors' calculation.

Forecast comparison (II): Core inflation - CPI item predictors

			Target: core inflation		
horizon	**1**	**3**	**6**	**9**	**12**
DFM	0.99	1.03	0.92	0.81	**0.73***
PCR	0.85	0.85	1.01	1.05	0.94
PLS	1.42***	1.13	0.91	**0.62****	**0.81****
Ridge	1.03	0.80	**0.75***	**0.55*****	**0.47*****
LASSO	1.14*	1.01	0.97	0.96	0.96
Elastic	1.02	0.90	**0.77****	**0.68*****	**0.79****
SVM	1	**0.82***	**0.65*****	**0.55*****	**0.45*****
Forest	1.16	0.96	**0.72***	**0.66****	**0.57*****
ANN	1.43*	1.24	1.16	0.95	**0.74****

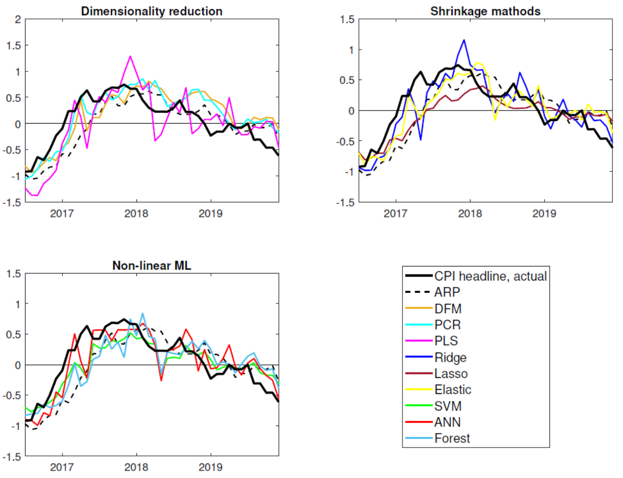
Forecasts using CPI item series as predictors. Root mean squared errors, relative to AR(p) model. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. $***\backslash**\backslash*$ indicates significance at $10\%, 5\%, and 1\%$, respectively. Relative RMSE that are significant at a level of 10% or lower and taking values below 1 are marked in bold. Source: ONS and authors' calculation.

Forecast comparison (III): Service core inflation - CPI item predictors

			Target: services core inflation		
horizon	**1**	**3**	**6**	**9**	**12**
DFM	0.98	1.00	1.06	0.94	0.87
PCR	0.92	1.04	1.14	0.93	0.95
PLS	1.32	1.32**	1.20	0.87	0.90
Ridge	1.18	1.25	1.17	**0.78*** **	0.90
LASSO	0.96	1.05	1.02	0.93	0.91
Elastic	0.97	1.07	1.04	**0.77*** **	**0.76* **
SVM	0.93	1.04	0.93	**0.77*** **	**0.73** **
Forest	0.99	1.07	0.95	**0.80** **	**0.73** **
ANN	1.34*	1.33**	1.26	0.86	0.87

Forecasts using CPI item series as predictors. Root mean squared errors, relative to AR(p) model. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. $* * *\backslash ** \backslash *$ indicates significance at $10\%, 5\%, and 1\%$, respectively. Relative RMSE that are significant at a level of 10% or lower and taking values below 1 are marked in bold. Source: ONS and authors' calculation.

Forecast comparison (IV): Point forecasts, headline inflation (3m)



Forecasts of CPI headline inflation 3-months ahead from AR benchmark (dashed black line) and from various forecasting models using CPI items predictors (coloured lines), compared to the actual outcome (solid balck lines), lagged by 3 months. Source: ONS and authors' calculation.

Opening the black box of machine learning [work in progress]

Opening black box using Shapley additive explanations framework

We propose a model-agnostic approach aimed at informing interpretations of results: how much do disaggregated CPI components contribute to predicting aggregate CPI?

1. Model decomposition: Shapley values (Joseph, 2019; Lundberg and Lee, 2017)
   - compute predictions from many sampled (or all possible) combinations of other predictors and get the marginal contribution (or "payoff") of including predictor k
   - gives exact decomposition of each prediction into marginal contributions from each of the n predictors, beyond mean predicted value from training set c
2. Context-specific partial re-aggregation
   - aggregate the Shapley components of items according to CPI sectors for each model
   - reduces dimension, helps inference and tractability
3. Statistical testing: Shapley regression
   - is the component actually a useful indicator of future inflation? Model accuracy needs to be taken into account – *now* we can run a linear regression!
   - addressed by regressing the target variables (CPI inflation) on Shapley components

Statistical analysis using Shapley values (Joseph, 2019)

n - set of predictors in the model (e.g. lagged CPI inflation and 581 CPI items)

x_t - set of observations for which we want to explain / decompose the predictive value

c - mean predicted value based on training set

$$
\begin{aligned}
f(x_t) &= \sum_{k=1}^{n} \phi_k^S(x_t) + c \equiv \Phi_t^S(x_t) + c\,, \quad \text{(1. model decomposition)} \\
\phi_k^S(f, x_t) &= \sum_{S \subseteq \mathcal{C} \setminus \{k\}} \frac{|S|!(n - |S| - 1)!}{n!} \left[f(x_t | S \cup \{k\}) - f(x_t | S) \right] \quad \text{(Shapley value)} \\
\mathcal{F}\left(\Phi_t^S(x_t)\right) &= \sum_{j=1}^{p} \psi_{j,t}^S(x_t) \equiv \Psi_t^S(x_t) \quad \text{(2. meso-aggregation)} \\
y_{t+h} &= \alpha_t' + \sum_{j=1}^{p} \beta_j^S \, \psi_j^S(x_t) + \epsilon_t' \quad \text{(3. component alignment test)}
\end{aligned}
$$

Forecast comparison (V): Shapley regression (selected components only)

	Ridge Regression			Random Forest		
division	*headline*	*core*	*service core*	*headline*	*core*	*service core*
LAG	**0.19****	**0.17****	**0.27*****	**0.58*****	**0.30*****	**0.33*****
	0.02	0.03	0.03	0.27	0.14	0.26
Food & non-alc. bev.	**0.37*****	–	–	**0.24*****	–	–
	0.19			0.28		
Acl. bev. & tobacco	**0.15*****	**0.16****	-0.10	0.05	-0.00	0.04
	0.05	0.08	0.07	0.01	0.02	0.01
Clothing & footwear	**0.31*****	0.09	**0.16****	0.01	**0.13****	0.01
	0.09	0.10	0.12	0.03	0.04	0.06
Housing & fuels	**0.31*****	**0.19****	**0.15****	0.09	**0.23*****	-0.01
	0.05	0.05	0.06	0.03	0.11	0.10
Furnishing & house maint.	**0.22*****	0.06	0.10	**0.21*****	0.06	**0.15****

Summary of Shapley regression (*) for Ridge Regression (LHS) and Random Forest (RHS) models for headline, core and service core inflation. CPI models components for different horizons are grouped. The share of each aggregate component is given below each coefficient. Core and service core targets do not contain item components from food and non-alcoholic beverages. Significance levels: ***:1%, **:5%, *:10%. Panel-HAC standard errors grouped by forecast horizon have been used. Source: ONS and authors calculations.

Take-away messages

- Micro-data are helpful to forecast macro outcomes
- Machine learning models deal well with high-dimensional datasets
- Ridge regression, SVMs and Random Forest show strong forecasting performance 9-12 months ahead
- Shapley values and regressions help to overcome the black box critique and allows for standardised communication of results

Thanks for listening

contact: andreas.joseph@bankofengland.co.uk

References i

Almosova, A. and Andresen, N. (2019). Nonlinear inflation forecasting with recurrent neural networks. Unpublished manuscript.

Aparicio, D. and Bertolotto, M. I. (2020). Forecasting inflation with online prices. *International Journal of Forecasting*, 36(2):232–247.

Blake, A. P. and Kapetanios, G. (2000). A radial basis function artificial neural network test for arch. *Economics Letters*, 69(1):15–23.

Blake, A. P. and Kapetanios, G. (2010). Tests of the martingale difference hypothesis using boosting and rbf neural network approximations. *Econometric Theory*, 26(5):1363–1397.

Carriero, A., Galvao, A. B., and Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 35(4):1226–1239.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

Domit, S., Monti, F., and Sokol, A. (2019). Forecasting the UK economy with a medium-scale Bayesian VAR. *International Journal of Forecasting*, 35(4):1669–1678.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In *Handbook of economic forecasting*, volume 2, pages 2–56. Elsevier.

Garcia, M. G., Medeiros, M. C., and Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, 33(3):679–693.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Hendry, D. F. and Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of business & economic statistics*, 29(2):216–227.

References ii

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by hicp component improve forecast accuracy? *International Journal of Forecasting*, 21(1):119–136.

Ibarra, R. (2012). Do disaggregated cpi data improve the accuracy of inflation forecasts? *Economic Modelling*, 29(4):1305–1313.

Joseph, A. (2019). Parametric inference with universal function approximators. *Bank of England Staff Working Paper Series*, (784).

Kapetanios, G., Labhard, V., and Price, S. (2008). Forecasting using bayesian and information-theoretic model averaging: an application to uk inflation. *Journal of Business & Economic Statistics*, 26(1):33–41.

Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886.

Koop, G. M. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, 28(2):177–203.

Lundberg, S. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, pages 4765–4774.

Martins, M. M., Verona, F., et al. (2020). Forecasting inflation with the new keynesian phillips curve: Frequency matters. Bank of Finland Research Discussion Papers 4, 2020.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2019). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, pages 1–22.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Stock, J. H. and Watson, M. W. (2008). Phillips curve inflation forecasts.

Stock, J. H. and Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4):770–784.

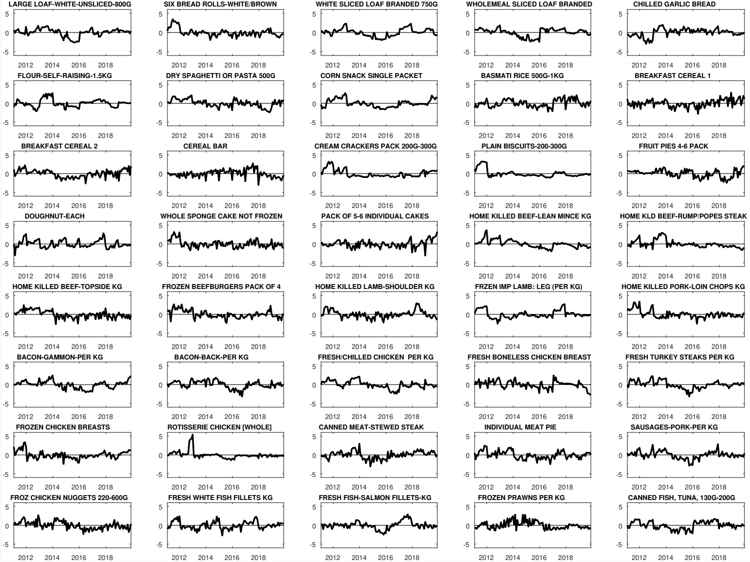
References iii

Strumbelj, E. and Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research*, 11:1–18.

Vapnik, V. (1998). *Statistical learning theory.* John Wiley&Sons Inc., New York.

Wang, Y., Wang, B., and Zhang, X. (2012). A new application of the support vector regression on the construction of financial conditions index to cpi prediction. *Procedia Computer Science*, 9:1263–1272.

Looking at some items (Source: ONS)



Data in levels, standardised. Item identifiers No. 210102 to No. 211207. Source: ONS.

back

Machine learning methods - Support Vector Machines (SVM)

- Support vectors represent class boundaries in classification problems (Vapnik, 1998), similar to logistic regressions, but SVMs also capture non-linearities through kernel function (Wang et al., 2012)

$$y_{t+h} = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i \mathcal{K}\left(x_i^{tr}, x\right) + \varepsilon \,, \tag{2}$$

  weights $\hat{\alpha}_i \geq 0$ mark the support vectors, m is size of training vector.
- Gaussian kernel $\mathcal{K}(\cdot, \cdot)$ (radial basis function)
- penalisation through restrictions on $\hat{\alpha}_i$, returning a dense model with local sparsity around support vectors back

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

$$y_{t+h} = \sum_{m=1}^{M} \beta_m I(x_t \in P_m) + \varepsilon_t, \quad \text{with} \quad \beta_m = 1/|P_m| \sum_{y^{tr} \in P_m} y^{tr}, \; m \in \{1, \ldots, M\}. \tag{3}$$

- 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

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.

$$y_{t+H} = G(x_t, \beta) + \varepsilon = g_L(g_{L-1}(g_{L-2}(\dots g_1(x_t, \beta_0), \dots, \beta_{L-2}), \beta_{L-1}), \beta_L) + \varepsilon \quad (4)$$

  - activation function $g(\cdot)$ introduces non-linearity into the model. We use rectified linear unit functions (ReLU) (Blake and Kapetanios, 2000, 2010)
- 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