Forecasting UK inflation bottom up

Andreas Joseph

Bank of England

Joint work with Chiranjit Chakraborty (BoE), Galina Potjagailo (BoE), Eleni Kalamara (KCL) and George Kapetanios (KCL)

Challenges and recent advances in modelling and forecasting inflation

SURF workshop in cooperation with ECB, Bank of Finland, Banca d’Italia and OeNB

28. November 2022, virtual

*Disclaimer: The expressed views are my own and not necessarily those of the Bank of England (BoE). All errors are ours.
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 ($\approx 700$ 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 (and high-dim settings)
  - 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 investigate the relative importance of aggregated components (Buckmann and Joseph, 2022)
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).
Forecasting setup & data
Setup

- **Targets:** Monthly UK headline, core and service core inflation, 1-12 months ahead
- **Predictors:**
  - 700 monthly chain-linked item-level price indices (from ONS)
  - 46 macroeconomic and financial variables
  - all series transformed to y-o-y log differences, mean-variance standardised
- **Sample period:** 2001:M2 - 2021:M11
  - initial TS cross-validation & training sample: 2002M2-2009M1 (incl. changes)
  - Test period: 2009:M2 - 2021:M11
  - 7-year rolling window (data trade-off, 69% coverage in baseline)
- **Benchmarks:**
  - AR(2) by BIC
- **Two-step approach** to separate out effect of micro data
  - Step 1: AR(2) benchmark
  - Step 2: Use micro data (w/wo macro indicators) & model zoo on Step 1 residuals
Classification of Individual Consumption According to Purpose (COICOP)

Source: ONS.
Micro inflation has leptokurtic distribution

Figure source: Authors’ calculation using ONS data.

In line with findings for US (Klenow and Kryvtsov, 2008) and Turkey (Özmen and Sevinc, 2016).
Micro inflation time series

Median item index changes in line with aggregate inflation, but wide dispersion.

Figure source: Authors’ calculation using ONS data.
Results
### Forecast comparison (I): Headline inflation - CPI item predictors

**Source:** Authors’ calculation using ONS data.

#### Predictors: CPI item indices

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Target: headline CPI</th>
<th>Target: Core CPI</th>
<th>Target: Service CPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>horizon 3</td>
<td>horizon 6</td>
<td>horizon 12</td>
</tr>
<tr>
<td>PCA</td>
<td>1.0</td>
<td>0.98</td>
<td>1.04</td>
</tr>
<tr>
<td>PLS</td>
<td>1.09</td>
<td>1.19</td>
<td>1.29</td>
</tr>
<tr>
<td>Ridge</td>
<td>1.0</td>
<td>0.84*</td>
<td>0.96</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.98</td>
<td>1.01</td>
<td>0.96</td>
</tr>
<tr>
<td>Elastic</td>
<td>1.0</td>
<td>1.0</td>
<td>0.98</td>
</tr>
<tr>
<td>SVM</td>
<td>1.0</td>
<td>1.02</td>
<td>1.14</td>
</tr>
<tr>
<td>Forest</td>
<td>0.99</td>
<td>1.07</td>
<td>1.21</td>
</tr>
<tr>
<td>NN</td>
<td>1.19**</td>
<td>1.24*</td>
<td>1.16</td>
</tr>
</tbody>
</table>

**Notes:** Root mean squared errors, relative to AR(2) model. Forecasts of headline CPI inflation (left panel), Core inflation (middle) and Service inflation (right) using CPI item series as predictors. Rolling window samples over sample period 2002-2021 with 7 years of training sample and out-of-sample forecasts at horizons of 3, 6, and 12 months. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey’s adjustment. Significance levels: ***:1%, **:5%, *:10%. Relative RMSE for forecasts at the 1-month horizon were not significant and are not presented for space constraints.
Forecast comparison (II): Headline inflation - CPI item and macro variables

<table>
<thead>
<tr>
<th>Predictors:</th>
<th>CPI items &amp; Macro ind.</th>
<th>Predictors: Macro ind. only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target: headline CPI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>horizon 3 6 12</td>
<td>horizon 3 6 12</td>
</tr>
<tr>
<td>PCA</td>
<td>1.04 1.01 1</td>
<td>PCA 0.98 0.99 1.1</td>
</tr>
<tr>
<td>PLS</td>
<td>1.3 1.2* 1.09</td>
<td>PLS 0.98 1.01 1.81</td>
</tr>
<tr>
<td>Ridge</td>
<td>0.97 0.9 0.99</td>
<td>Ridge 0.99 1.22** 1.72**</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.94 1.05 0.98</td>
<td>Lasso 0.92** 0.97 0.99*</td>
</tr>
<tr>
<td>Elastic</td>
<td>0.98 0.95 1</td>
<td>Elastic 1 0.98 1.02</td>
</tr>
<tr>
<td>SVM</td>
<td>1.14 1.01 0.99</td>
<td>SVM 0.96 1.18 1.52**</td>
</tr>
<tr>
<td>Forest</td>
<td>1.2 1.07 0.96</td>
<td>Forest 0.97 0.99 1.18</td>
</tr>
<tr>
<td>NN</td>
<td>1.84** 1.25* 1.28**</td>
<td>NN 0.95 1.16* 1.44**</td>
</tr>
</tbody>
</table>

Source: Authors’ calculation using ONS data.

So far, results quite sobering. But what about outcome distributions, like tail events and turning points?
We consider sub-periods during which our target CPI indicators are

- **high**: above 3% (upper quartile since the year 2008)
- **low**: below 1.5% (lower quartile since the year 2008)
- **IQR**: between 1.5% and 3%
- **rising**: 3-m moving average %-change *positive* for at least 5 consecutive months
- **falling**: 3-m moving average %-change *negative* for at least 5 consecutive months
- **stable**: neither falling nor rising
Forecast comparison (III): Headline inflation - by inflation regime

<table>
<thead>
<tr>
<th>Horizon</th>
<th>CPI inflation falling</th>
<th>CPI inflation rising</th>
<th>CPI inflation stable</th>
<th>CPI inflation low</th>
<th>CPI inflation high</th>
<th>CPI inflation in IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PCA</td>
<td>Elastic</td>
<td>Ridge</td>
<td>Forest</td>
<td>SVM</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>0.93**</td>
<td>0.9**</td>
<td>0.78***</td>
<td>0.81**</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>Elastic</td>
<td>Ridge</td>
<td>-</td>
<td>SVM</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>0.82***</td>
<td>-</td>
<td>Ridge</td>
<td>SVM</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>Elastic</td>
<td>-</td>
<td>-</td>
<td>Elastic</td>
</tr>
</tbody>
</table>

Source: Authors’ calculation using ONS data.

- Micro data particularly useful around **turning points and tail outcomes**
- **Penalised regressions** produce the most robust forecast improvements
- Results on **long horizons** somewhat stronger
Forecast comparison (IV): Core & Service inflation

<table>
<thead>
<tr>
<th>Core CPI forecast - CPI items &amp; Macro indicators</th>
<th>Horizon</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core inflation falling</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>PCA 0.89***</td>
</tr>
<tr>
<td>Core inflation rising</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Lasso 0.88***</td>
</tr>
<tr>
<td>Core inflation high</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>SVM 0.94***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Service CPI forecast - CPI items &amp; Macro indicators</th>
<th>Horizon</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service inflation falling</td>
<td>-</td>
<td>Ridge 0.68***</td>
<td>Ridge 0.68**</td>
<td>Lasso 0.87***</td>
<td></td>
</tr>
<tr>
<td>Service inflation stable</td>
<td>-</td>
<td>Elastic 0.59**</td>
<td>PCA 0.83**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Service inflation low</td>
<td>-</td>
<td>Ridge 0.75**</td>
<td>-</td>
<td>Ridge 0.88***</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ calculation using ONS data.

- Core inflation more difficult to forecast
- Results for service inflation comparable but somewhat weaker than headline
Opening the black box
Model-agnostic approach (Buckmann and Joseph, 2022) aimed at informing interpretations of results: how much do CPI subcomponents contribute to predicting aggregate CPI?

1. Model decomposition: Shapley values

2. Context-specific partial re-aggregation (core goods/services, volatile goods/service and energy)
Model decomposition using Shapley values

\( n \) - set of predictors in the model (e.g. macro indicators and 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

\[
f(x_t) = \sum_{k=1}^{n} \phi_k^S(x_t) + c, \quad (1. \text{ model decomposition})
\]

\[
\phi_k^S(f, x_t) = \sum_{S \subseteq C \setminus \{k\}} \frac{|S|!(n-|S|-1)!}{n!} \left[ f(x_t|S \cup \{k\}) - f(x_t|S) \right] \quad (\text{Shapley value})
\]

\[
\mathcal{F}(\Phi_t^S(x_t)) = \sum_{j=1}^{p} \psi_j^S(x_t) \quad (2. \text{ partial re-aggregation})
\]

We consider the Ridge regression and the Random Forest in the following (no Shapley value approximation needed).
Shapley decomposition (I): Ridge vs Random Forest

Absolute Shapley values shares by item groups. Source: Authors’ calculation using ONS data.

- Ridge group weights roughly follow input proportions
- Random forest allocation more varied across horizons
Shapley decomposition (II): Falling / rising headline inflation

Shapley values shares by item groups relative to input share. Source: Authors’ calculation using ONS data.

- Models tend to over-proportionally draw on volatile items
- less so on core services and macro data
- Effects particularly strong for Random Forest
Take-away messages

- Using Micro item-level data can lead to sizeable forecast improvements around turning points and in the tails of aggregate inflation measures.

- However, hard to beat the AR benchmark overall (high persistence in aggregate inflation), especially during ‘normal times’.

- Model comparison
  - Penalised regressions see the most improvements.
  - Machine learning models deliver promising results on shorter forecast horizons.

- Shapley values allow to view micro-based modelling results based on broad product groups used in policy discussion.

- Challenge to clearly connect model performance to general micro drivers. Likely idiosyncratic analysis needed for which Shapley decomposition offers basis.
Thanks for listening

Q & A


CPI item series

- UK monthly CPI is constructed from about 700 representative item indices by the Office for National Statistics (ONS), publicly 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
  - Locally and centrally collected prices
  - About 700 item indices without missing values in 7-window approach
  - Items cover 69% of total CPI, similar coverage across broad CPI categories
  - Increase of coverage across sample, 250 (2002) - 600 (2019), about 400 mean.
Some items: mostly stationary, but heterogeneous with jumps
## Data coverage

<table>
<thead>
<tr>
<th>description</th>
<th>weight</th>
<th>#items, raw</th>
<th>#items, balanced</th>
<th>% covered</th>
<th>median</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Food &amp; non-alc. bev.</td>
<td>12</td>
<td>155.41</td>
<td>111.6</td>
<td>72</td>
<td>1.74</td>
<td>7.31</td>
</tr>
<tr>
<td>2 Alcohol &amp; tobacco</td>
<td>5</td>
<td>26.57</td>
<td>15.2</td>
<td>56</td>
<td>2.02</td>
<td>4.23</td>
</tr>
<tr>
<td>3 Clothing &amp; footwear</td>
<td>7</td>
<td>77.63</td>
<td>54.6</td>
<td>70</td>
<td>-0.71</td>
<td>5.59</td>
</tr>
<tr>
<td>4 Housing &amp; fuels</td>
<td>13</td>
<td>37.04</td>
<td>30.3</td>
<td>82</td>
<td>2.67</td>
<td>6.42</td>
</tr>
<tr>
<td>5 Furnishing &amp; house maint.</td>
<td>6</td>
<td>72.85</td>
<td>53.1</td>
<td>73</td>
<td>1.34</td>
<td>5.02</td>
</tr>
<tr>
<td>6 Health</td>
<td>3</td>
<td>20.09</td>
<td>14.7</td>
<td>73</td>
<td>1.77</td>
<td>4.28</td>
</tr>
<tr>
<td>7 Transport</td>
<td>14</td>
<td>43.58</td>
<td>31.5</td>
<td>72</td>
<td>2.46</td>
<td>7.27</td>
</tr>
<tr>
<td>8 Communication</td>
<td>3</td>
<td>9.18</td>
<td>5.6</td>
<td>61</td>
<td>1.91</td>
<td>10.69</td>
</tr>
<tr>
<td>9 Recreation &amp; culture</td>
<td>15</td>
<td>112.35</td>
<td>64.6</td>
<td>56</td>
<td>1.41</td>
<td>8.07</td>
</tr>
<tr>
<td>10 Education</td>
<td>2</td>
<td>3.05</td>
<td>2.1</td>
<td>69</td>
<td>6.71</td>
<td>5.44</td>
</tr>
<tr>
<td>11 Restaurants &amp; hotels</td>
<td>9</td>
<td>48.36</td>
<td>31.8</td>
<td>66</td>
<td>2.85</td>
<td>1.88</td>
</tr>
<tr>
<td>12 Misc. goods &amp; services</td>
<td>11</td>
<td>76.13</td>
<td>52.4</td>
<td>69</td>
<td>1.65</td>
<td>8.42</td>
</tr>
<tr>
<td>13 Total</td>
<td>100</td>
<td>682.2</td>
<td>467.5</td>
<td>69</td>
<td>2.15</td>
<td>6.22</td>
</tr>
</tbody>
</table>

Source: Authors’ calculation using ONS data.
We are interested forecasting inflation $y_t$ in period $t + h$, based on the past dynamics of $y_{t-j}$ and the set of predictors $x_t = (x_{1t}, \ldots, x_{Nt})'$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$

$$\hat{y}_{t+h} = y_{t+h} - \sum_{j=1}^{2} \hat{\gamma}_j y_{t-j+1} = f(x_t) + \tilde{\epsilon}_{t+h} \quad \text{(Step 2)}.$$

We consider a wide range of forecasting methods $f$ to deal with high-dim data

- **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)

Hyperparameter tuning via K-fold cross-validation

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

\[ y_{t+h} = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i \mathcal{K}(x_{tr}^i, x) + \varepsilon, \]  

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
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

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

- 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
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 $y_t$.

$$y_{t+H} = G(x_t, \beta) + \varepsilon = g_L(g_{L-1}(g_{L-2}(\ldots g_1(x_t, \beta_0), \ldots, \beta_{L-2}), \beta_{L-1}), \beta_L) + \varepsilon$$ (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.
Input sizes by window length and time period (headline inflation)

Source: Authors’ calculation using ONS data.
All models forecast lines (headline inflation)

Source: Authors’ calculation using ONS data.