Forecasting UK inflation bottom up

Andreas Joseph
Bank of England

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

New Approaches to macroeconomic monitoring, nowcasting and forecasting
OECD & Banque de France
8. June 2021

*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 (≈ 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 (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 test predictive power of aggregated components (Joseph, 2019)
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, 1-12 months ahead
- **Predictors:**
  - 581 monthly item-level price indices (from ONS), chain-linked
  - 43 macroeconomic and financial variables
  - all series transformed to y-o-y log differences, mean-variance standardised
- **Sample period:** 2011:M1 - 2019:M12 (no Covid yet)
  - TS cross-validation & training sample: Until 2015:M3
  - Test period: 2015:M4 - 2019:M12
- **Extended sample period** with macro data only
  - 2000M1:2019M12, test period starts in 2005M4
- **Benchmarks:**
  - AR(\(p\)) with \(p \leq 12\) by BIC
  - DFM and PCA with 5 factors / principal components
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
  - 581 item indices without missing values, 2011M1-2019M12
  - Items cover 84% of total CPI, similar coverage across broad CPI categories
Classification of Individual Consumption According to Purpose (COICOP)

- All Items index (COICOP1)
  - CPIH/CPI

- Divisions (COICOP2)
  - (01) Food and non-alcoholic beverages
  - (02) Alcohol and tobacco
  - (n) Other COICOP divisions

- Groups (COICOP3)
  - (01.1) Food
  - (01.2) Non-alcoholic beverages

- Classes (COICOP4)
  - (01.1.6) Fruit
  - (01.1.7) Vegetables
  - (01.1.n) Other food classes

- Subclasses (COICOP5)
  - (01.1.6.1) Fresh or chilled fruit
  - (01.1.6.3) Dried fruit and nuts
  - (01.1.6.n) Other fruit subclasses

- Items
  - Apples (dessert)
  - Bananas
  - Other fruit

- Shop type/region
  - Apples, in multiples, in Wales
  - Apples, in independents in Wales
  - Multiples/Independents in other locations

- Product
  - Royal Gala, Cardiff, multiple store 1
  - Pink Lady, Cardiff, multiple store 2
  - Other apples, other locations in Wales

Source: ONS.
Classification of Individual Consumption According to Purpose (COICOP)

- All items index (COICOP1)
  - CPIH/CPI

- Divisions (COICOP2)
  - (01) Food and non-alcoholic beverages
  - (02) Alcohol and tobacco
  - (n) Other COICOP divisions

- Groups (COICOP3)
  - (01.1) Food
  - (01.2) Non-alcoholic beverages

- Classes (COICOP4)
  - (01.1.6) Fruit
  - (01.1.7) Vegetables
  - (01.1.n) Other food classes

- Subclasses (COICOP5)
  - (01.1.6.1) Fresh or chilled fruit
  - (01.1.6.3) Dried fruit and nuts
  - (01.1.6.n) Other fruit subclasses

- Items
  - Apples (dessert)
  - Bananas
  - Other fruit

- Shop type/region
  - Apples, in multiples, in Wales
  - Apples, in independents in Wales
  - Multiples/Independents in other locations

- Product
  - Royal Gala, Cardiff, multiple store 1
  - Pink Lady, Cardiff, multiple store 2
  - Other apples, other locations in Wales

Source: ONS.
Some items: mostly stationary, but heterogeneous with jumps

Data in levels, standardised. Item identifiers No. 210102 to No. 211207. Source: ONS.
<table>
<thead>
<tr>
<th>division</th>
<th>weight (%)</th>
<th># total</th>
<th># included</th>
<th>coverage</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food &amp; non-alc. bev.</td>
<td>10</td>
<td>159.8</td>
<td>129</td>
<td>81</td>
<td>0.66</td>
<td>6.98</td>
</tr>
<tr>
<td>Acl. bev. &amp; tobacco</td>
<td>4</td>
<td>27</td>
<td>20</td>
<td>74</td>
<td>1.13</td>
<td>4.68</td>
</tr>
<tr>
<td>Clothing &amp; footwear</td>
<td>7</td>
<td>76.7</td>
<td>71</td>
<td>93</td>
<td>1.68</td>
<td>5.18</td>
</tr>
<tr>
<td>Housing &amp; fuels</td>
<td>12</td>
<td>30.2</td>
<td>30</td>
<td>83</td>
<td>1.57</td>
<td>5.20</td>
</tr>
<tr>
<td>Furnishing &amp; house maint.</td>
<td>7</td>
<td>70.8</td>
<td>55</td>
<td>78</td>
<td>1.29</td>
<td>4.53</td>
</tr>
<tr>
<td>Health</td>
<td>3</td>
<td>18.6</td>
<td>19</td>
<td>92</td>
<td>1.84</td>
<td>3.92</td>
</tr>
<tr>
<td>Transport</td>
<td>15</td>
<td>41.3</td>
<td>36</td>
<td>87</td>
<td>2.12</td>
<td>6.03</td>
</tr>
<tr>
<td>Communication</td>
<td>2</td>
<td>10.9</td>
<td>9</td>
<td>83</td>
<td>2.12</td>
<td>13.05</td>
</tr>
<tr>
<td>Recreation &amp; culture</td>
<td>16</td>
<td>115.4</td>
<td>92</td>
<td>80</td>
<td>1.51</td>
<td>6.80</td>
</tr>
<tr>
<td>Education</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>100</td>
<td>10.48</td>
<td>8.18</td>
</tr>
<tr>
<td>Restaurants &amp; hotels</td>
<td>12</td>
<td>51.3</td>
<td>44</td>
<td>86</td>
<td>2.49</td>
<td>1.80</td>
</tr>
<tr>
<td>Misc. goods &amp; services</td>
<td>10</td>
<td>79</td>
<td>73</td>
<td>92</td>
<td>0.91</td>
<td>5.31</td>
</tr>
<tr>
<td>–</td>
<td>Total</td>
<td>100</td>
<td>684</td>
<td>581</td>
<td>84</td>
<td>2.32</td>
</tr>
</tbody>
</table>

Source: Authors’ calculation using ONS data.
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).
Average index changes in line with aggregate inflation.

Figure source: Authors’ calculation using ONS data.
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 $x_t = (x_{1t}, \ldots, x_{Nt})'$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$

$$\hat{y}_{t+h} = \hat{\alpha} + \sum_{i=1}^{N} \hat{\beta}_i x_t + \sum_{j=1}^{P} \hat{\gamma}_j y_{t-j+1}. \quad (1)$$

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)

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Horizon</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>0.91</td>
<td>0.73**</td>
<td>0.56***</td>
<td>0.44***</td>
<td>0.41***</td>
<td></td>
</tr>
<tr>
<td>DFM</td>
<td>1.04</td>
<td>0.97</td>
<td>0.85</td>
<td>0.72**</td>
<td>0.8*</td>
<td></td>
</tr>
<tr>
<td>PLS</td>
<td><strong>0.76</strong></td>
<td>0.57***</td>
<td>0.46***</td>
<td>0.41***</td>
<td>0.37***</td>
<td></td>
</tr>
<tr>
<td>Ridge</td>
<td>0.77</td>
<td>0.52***</td>
<td>0.37***</td>
<td>0.29***</td>
<td>0.29***</td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>0.77*</td>
<td><strong>0.59</strong>*</td>
<td>0.52***</td>
<td>0.42***</td>
<td>0.37***</td>
<td></td>
</tr>
<tr>
<td>Elastic</td>
<td>1.2</td>
<td>0.82</td>
<td>0.61**</td>
<td>0.49***</td>
<td>0.44***</td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>1.96***</td>
<td>1.22**</td>
<td>0.85</td>
<td>0.71***</td>
<td>0.69***</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>1.18**</td>
<td><strong>0.83</strong></td>
<td>0.57***</td>
<td>0.58***</td>
<td>0.57***</td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>1.26</td>
<td>0.71*</td>
<td>0.51***</td>
<td>0.31***</td>
<td>0.39***</td>
<td></td>
</tr>
</tbody>
</table>

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. ***, **, * 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: Authors’ calculation using ONS data.
Forecast comparison (I): Headline inflation - CPI item predictors

<table>
<thead>
<tr>
<th>horizon</th>
<th>1</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>0.91</td>
<td>0.73**</td>
<td>0.56***</td>
<td>0.44***</td>
<td>0.41***</td>
</tr>
<tr>
<td>DFM</td>
<td>1.04</td>
<td>0.97</td>
<td>0.85</td>
<td>0.72**</td>
<td>0.8*</td>
</tr>
<tr>
<td>PLS</td>
<td>0.76**</td>
<td>0.57***</td>
<td>0.46***</td>
<td>0.41***</td>
<td>0.37***</td>
</tr>
<tr>
<td>Ridge</td>
<td>0.77</td>
<td>0.52***</td>
<td>0.37***</td>
<td>0.29***</td>
<td>0.29***</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.77*</td>
<td>0.59***</td>
<td>0.52***</td>
<td>0.42***</td>
<td>0.37***</td>
</tr>
<tr>
<td>Elastic</td>
<td>1.2</td>
<td>0.82</td>
<td>0.61**</td>
<td>0.49***</td>
<td>0.44***</td>
</tr>
<tr>
<td>SVM</td>
<td>1.96***</td>
<td>1.22**</td>
<td>0.85</td>
<td>0.71***</td>
<td>0.69***</td>
</tr>
<tr>
<td>Forest</td>
<td>1.18**</td>
<td>0.83*</td>
<td>0.57***</td>
<td>0.58***</td>
<td>0.57***</td>
</tr>
<tr>
<td>ANN</td>
<td>1.26</td>
<td>0.71*</td>
<td>0.51***</td>
<td>0.31***</td>
<td>0.39***</td>
</tr>
</tbody>
</table>

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. ***, **, and * 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: Authors’ calculation using ONS data.
Forecast comparison (II): All cases

Average RMSE of CPI headline inflation forecasts, different forecasting models (sub-plots) and horizons (x-axis) and specifications (colours). Source: Authors' calculation using ONS data.
Opening the black box of machine learning
We propose a model-agnostic approach 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

3. Statistical testing: Shapley regression
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

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

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

\[
\mathcal{F} \left( \Phi_t^S(x_t) \right) = \sum_{j=1}^{p} \psi_{j,t}^S(x_t) \quad \text{(2. meso-aggregation)}
\]

\[
y_{t+h} = \alpha_t' + \sum_{j=1}^{p} \beta_j^S \psi_j^S(x_t) + \epsilon_t' \quad \text{(3. component alignment test)}
\]
Differences in inference between models

- Inference results depend on model convergence (learning progress)
- Usually different for different models, especially for small and high-dim samples
- Explains different performance based on different signals

⇒ Model inference important and insightful.
<table>
<thead>
<tr>
<th></th>
<th>Ridge Regression</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>headline</td>
<td>core</td>
</tr>
<tr>
<td><strong>Food &amp; non-alc. bev.</strong></td>
<td>0.23***</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>-</td>
</tr>
<tr>
<td><strong>Acl. bev. &amp; tobacco</strong></td>
<td>0.06***</td>
<td>0.09***</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Clothing &amp; footwear</strong></td>
<td>0.07***</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>Recreation &amp; culture</strong></td>
<td>0.13***</td>
<td>0.29***</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Comparison of Shapley-value-based model inference using only item indices for Ridge regression (LHS) and Random Forest (RHS) for headline, core and service core inflation and 4/12 divisions. 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: Authors calculations using ONS data.
Shapley results for Ridge regression

Shapley components of CPI divisions for Ridge regression forecasts, averaged over horizons of 7-12 months. Red bars indicate significance at 1% confidence level of Shapley regression coefficients.
Shapley results for Random Forest

Shapley components of CPI divisions for Random forecast forecasts, averaged over horizons of 7-12 months. Red bars indicate significance at 1% confidence level of Shapley regression coefficients.
Take-away messages

- Micro item-level data often **strongly improve forecasts** relative to benchmark (beyond improvement from macro data)

- Model comparison
  - Ridge regression performs best across horizons and specifications, but also Lasso, PCA, PLS and ANN perform well
  - DFM less useful with disaggregated item indices due to lack of dynamic co-movement
  - ML models strong in capturing **turning points / economic cycle** over longer sample (with macro data) [not for today, but some promising results in ongoing work]

- Provide model-agnostic method based on Shapley values and regressions to **communicate high-dimensional and non-linear modelling results**
Thanks for listening

Q & A


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)

\[ y_{t+h} = \hat{\alpha}_0 + \sum_{i=1}^{m} \hat{\alpha}_i K(x_{i}^{tr}, x) + \varepsilon, \quad (2) \]

weights \( \hat{\alpha}_i \geq 0 \) mark the support vectors, \( m \) is size of training vector.

- Gaussian kernel \( K(\cdot, \cdot) \) (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

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

(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 $y_t$.

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

- 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.