

# Forecasting UK inflation bottom up

---

**Andreas Joseph**

Bank of England\*

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

*ESCoE Economic Measurement '21*

11. May 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** ( $\approx 600$  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 (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**

---

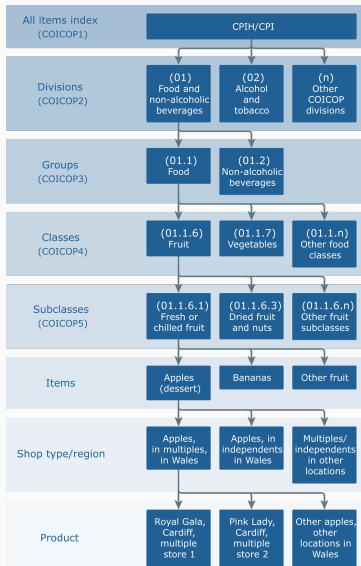
# Setup

- **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

- 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

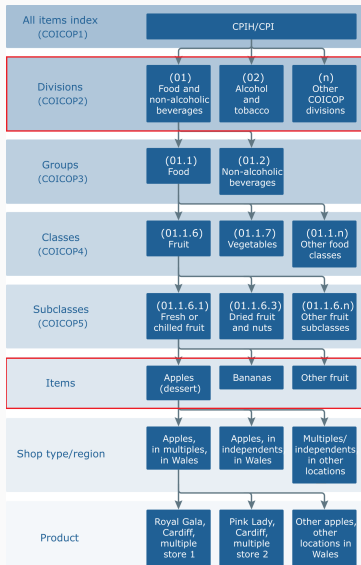


# Classification of Individual Consumption According to Purpose (COICOP)



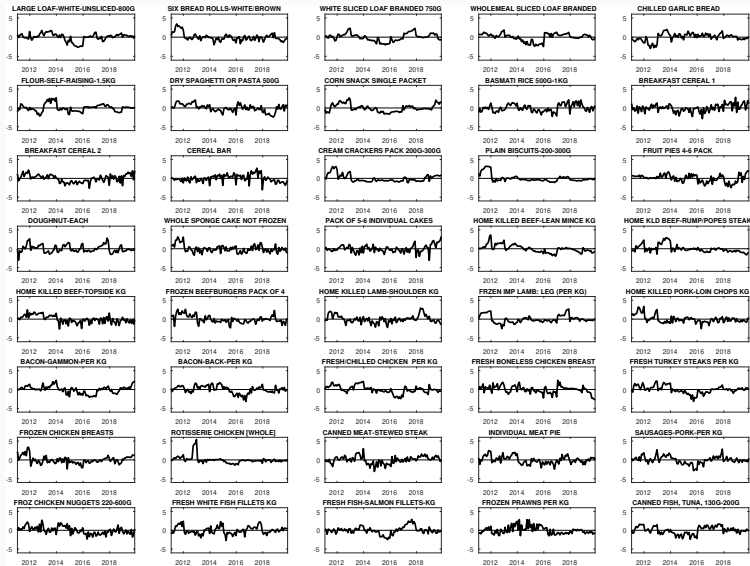
Source: ONS.

# Classification of Individual Consumption According to Purpose (COICOP)



Source: ONS.

# Some items: mostly stationary, but heterogeneous with jumps



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

## Data coverage

	division	weight (%)	# total	# in- cluded	coverage	mean	SD
1	Food & non-alc. bev.	10	159.8	129	81	0.66	6.98
2	Alc. bev. & tobacco	4	27	20	74	1.13	4.68
3	Clothing & footwear	7	76.7	71	93	1.68	5.18
4	Housing & fuels	12	30.2	30	83	1.57	5.20
5	Furnishing & house maint.	7	70.8	55	78	1.29	4.53
6	Health	3	18.6	19	92	1.84	3.92
7	Transport	15	41.3	36	87	2.12	6.03
8	Communication	2	10.9	9	83	2.12	13.05
9	Recreation & culture	16	115.4	92	80	1.51	6.80
10	Education	1	3	3	100	10.48	8.18
11	Restaurants & hotels	12	51.3	44	86	2.49	1.80
12	Misc. goods & services	10	79	73	92	0.91	5.31
–	Total	100	684	581	84	2.32	5.97

# Micro inflation has leptokurtic distribution

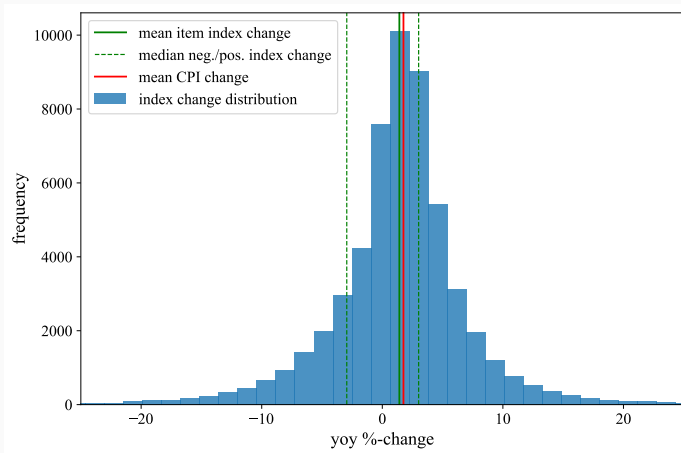
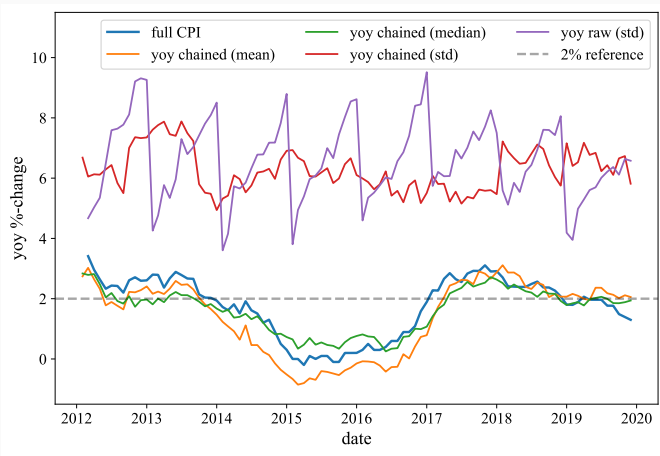


Figure source: Authors' calculation using ONS data. In line with findings for US (Klenow and Kryvtsov, 2008) and Turkey (Özmen and Sevinc, 2016).

# Micro inflation time series



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  $x_t = (x_{1t}, \dots, x_{Nt})'$ ,  $i = 1, \dots, N$  and  $t = 1, \dots, T$

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

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) ◀ SVM ◀ RF ◀ 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

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

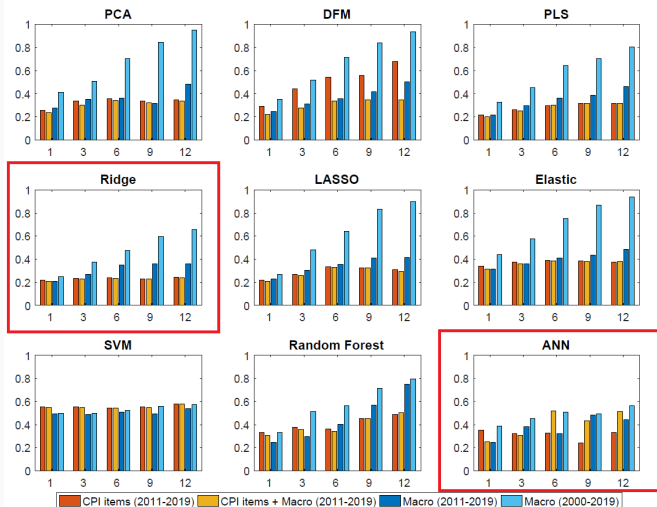
Notes: Forecasts using CPI item series as predictors. Root mean squared errors, relative to AR(p) model. Significance of forecast accuracy is assessed via Diebold and Mariano (1995) test statistics with Harvey's adjustment. \*\*\*\\*\*\\* indicates significance at 10%, 5%, and 1%, respectively. Relative 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

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

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 (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

---

# Opening black box using Shapley additive explanations framework

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

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

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

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

$$y_{t+h} = \alpha'_t + \sum_{j=1}^p \beta_j^S \psi_j^S(x_t) + \epsilon'_t \quad (3. \text{ component alignment test})$$

## 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.**

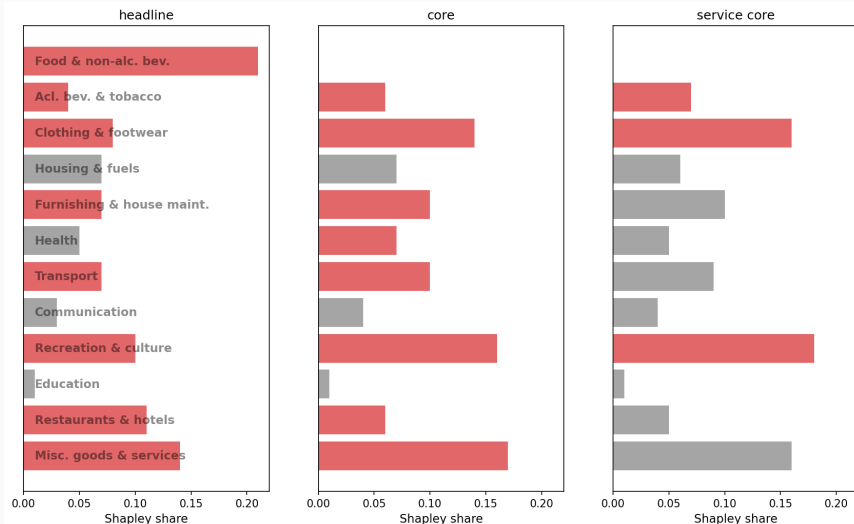
# Shapley value regression results

	Ridge Regression			Random Forest		
	<i>headline</i>	<i>core</i>	<i>service core</i>	<i>headline</i>	<i>core</i>	<i>service core</i>
7 – 12 months horizon						
Food & non-alc. bev.	<b>0.23***</b> 0.21	–	–	<b>0.08**</b> 0.18	–	–
Acl. bev. & tobacco	<b>0.06***</b> 0.04	<b>0.09***</b> 0.06	<b>0.17***</b> 0.07	-0.03 0.01	0.0 0.01	0.10 0.05
Clothing & footwear	<b>0.07***</b> 0.08	<b>0.11***</b> 0.14	<b>0.17***</b> 0.16	<b>0.10***</b> 0.02	<b>0.06*</b> 0.06	<b>0.14**</b> 0.04
Recreation & culture	<b>0.13***</b> 0.10	<b>0.29***</b> 0.16	<b>0.29***</b> 0.18	<b>0.12***</b> 0.06	<b>0.23***</b> 0.15	<b>0.18***</b> 0.31

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.

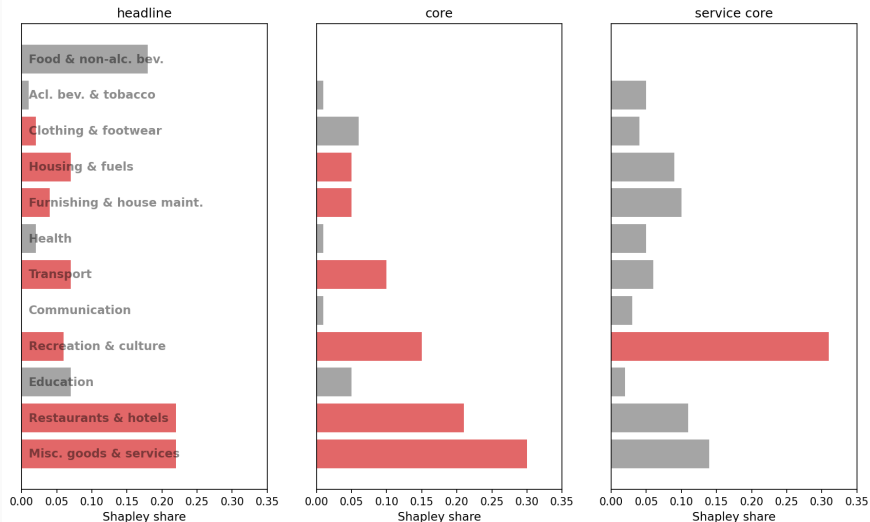


# Shapley results for Ridge regression



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 forest 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

# 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.
- Klenow, P. J. and Kryvtsov, O. (2008). State-Dependent or Time-Dependent Pricing: Does it Matter for Recent U.S. Inflation?\*. *The Quarterly Journal of Economics*, 123(3):863–904.
- 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.
- Özmen, M. U. and Sevinc, O. (2016). Price Rigidity in Turkey: Evidence from Micro Data. *Emerging Markets Finance and Trade*, 52(4):1029–1045.
- 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.
- 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.

# 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)

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

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

- Gaussian kernel  $\mathcal{K}(\cdot, \cdot)$  (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

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

- 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  $y_t$ .

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

- 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](#)