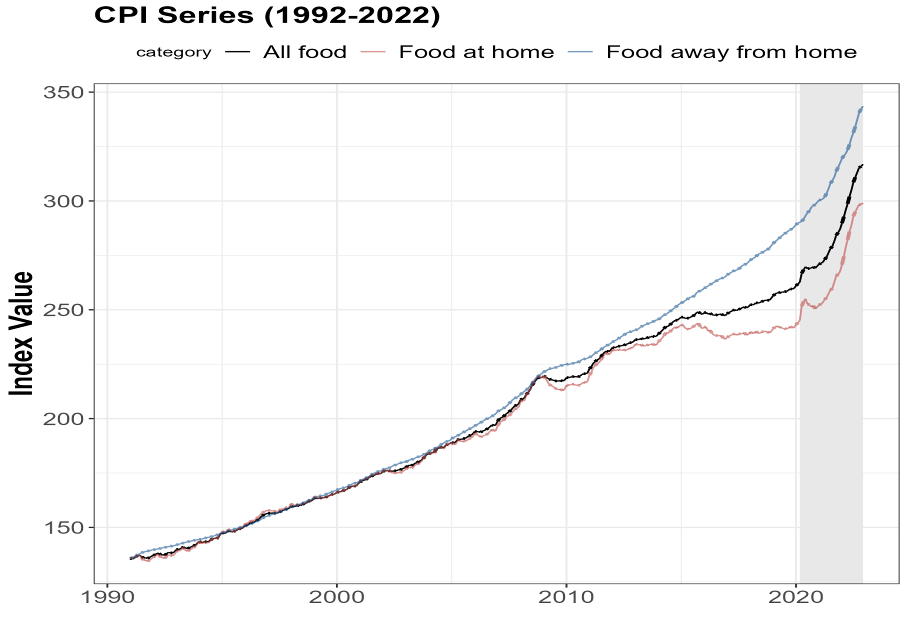
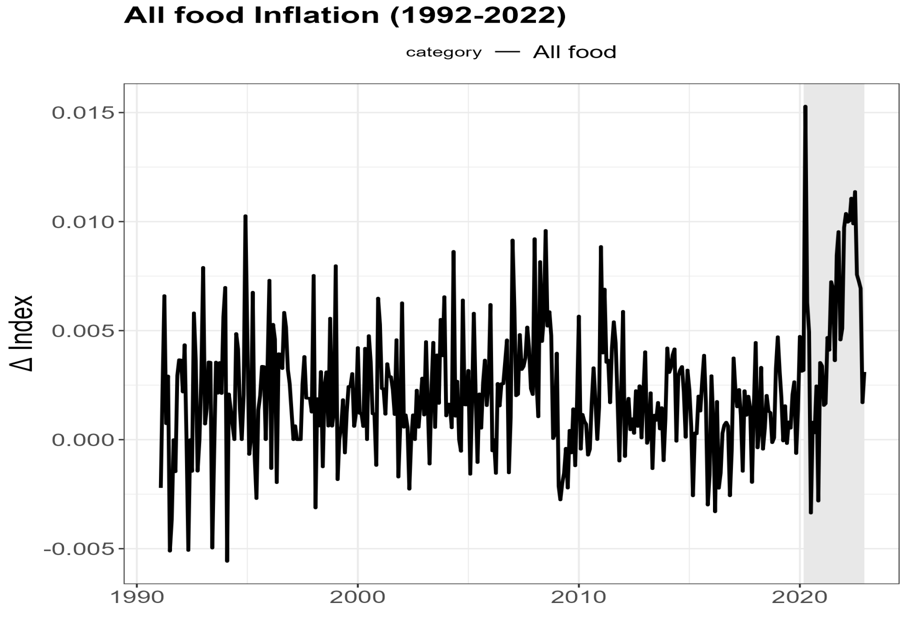
Food Price Inflation Forecasting:An Auto-Regressive Random Forest Approach

William McWilliams, Shamar L. Stewart, Olga Isengildina Massa

Virginia Tech Department of Agricultural and Applied Economics

Ag-Scholars 2024


CPI Series (1992-2022)
category
All food
Food at home
Food away from home
Index Value
350
300
250
200
150
1990
2000
2010
2020


All food Inflation (1992-2022)
category — All food
Δ Index
0.015
0.010
0.005
0.000
-0.005
1990
2000
2010
2020

Observed VS Forecasted

The Bureau of Labor Statistics (BLS)

- tracks and measures US food prices with indices
- the indices are updated monthly to reflect the prices from the previous month

Food Price Outlook

- forecasts US retail food prices since 1961
- produced by the United States Department of Agriculture (USDA) Economic Research Services (ERS)
- used for budgetary planning by government agencies, food industry participants, consumers, and media
- uses a univariate Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to forecast food related indices[1]

[1] Maclachlan et al. (2022)

Goals and Objectives

Apply an Auto-Regressive Random Forest approach for forecasting Food Price Inflation

- Maintain the recent qualitative improvements
- Allow for the option to easily include exogenous variables
- Accommodate nonlinear forecasting dynamics

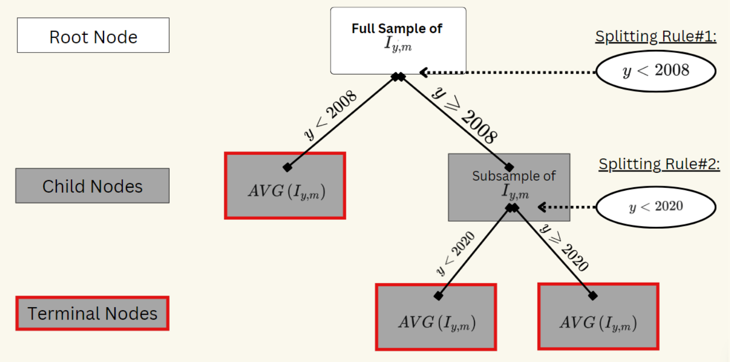
Roadmap

USDA FPO Forecast Cycles from Years 2024 to 2025

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Year = 2024	Step = 7	8	9	10	11	12	13	14	15	16	17	18
							1	2	3	4	5	6
2025	7	8	9	10	11	12	13	14	15	16	17	18

$$\hat{\Pi}_{y,m|s} = \begin{cases} 100 \cdot \dfrac{\overbrace{\sum_{m=1}^{12} \widehat{\mathcal{I}}_{y,m}}^{Target \quad Year} - \overbrace{\sum_{m=1}^{12} \widehat{\mathcal{I}}_{y-1,m} + \sum_{m=1}^{12} \mathcal{I}_{y-1,m}}^{Base \quad Year}}{\sum_{m=1}^{12} \widehat{\mathcal{I}}_{y-1,m} + \sum_{m=1}^{12} \mathcal{I}_{y-1,m}} & \textit{for } s = 1 \ldots 6 \\ 100 \cdot \dfrac{\sum_{m=1}^{12} \widehat{\mathcal{I}}^{j}_{y,m} - \sum_{m=1}^{12} \mathcal{I}^{j}_{y-1,m}}{\sum_{m=1}^{12} \mathcal{I}^{j}_{y-1,m}} & \textit{for } s = 7 \\ 100 \cdot \dfrac{\sum_{m=1}^{12} \widehat{\mathcal{I}}_{y,m} + \sum_{m=1}^{12} \mathcal{I}_{y,m} - \sum_{m=1}^{12} \mathcal{I}_{y-1,m}}{\sum_{m=1}^{12} \mathcal{I}_{y-1,m}} & \textit{for } s = 8 \ldots 18 \end{cases} \tag{1}$$

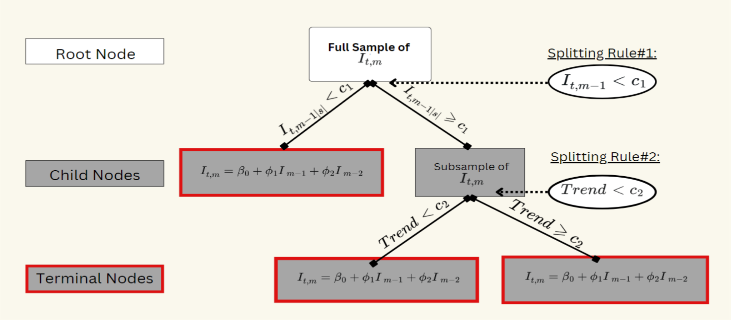
Regression Trees



Note: Hypothetical example to illustrate a forecast

$$\min_{j \in \mathcal{J}^-,\, c \in \mathbb{R}} \Big[\sum_{\{I_{y,m} \in l_1 | j < c\}} (I_{y,m} - \hat{I}_{y,m})^2 + \sum_{\{I_{y,m} \in l_2 | j \geq c\}} (I_{y,m} - \hat{I}_{y,m})^2 \Big] \tag{2}$$

Auto Regressive Trees



Note: Hypothetical example to illustrate a forecast

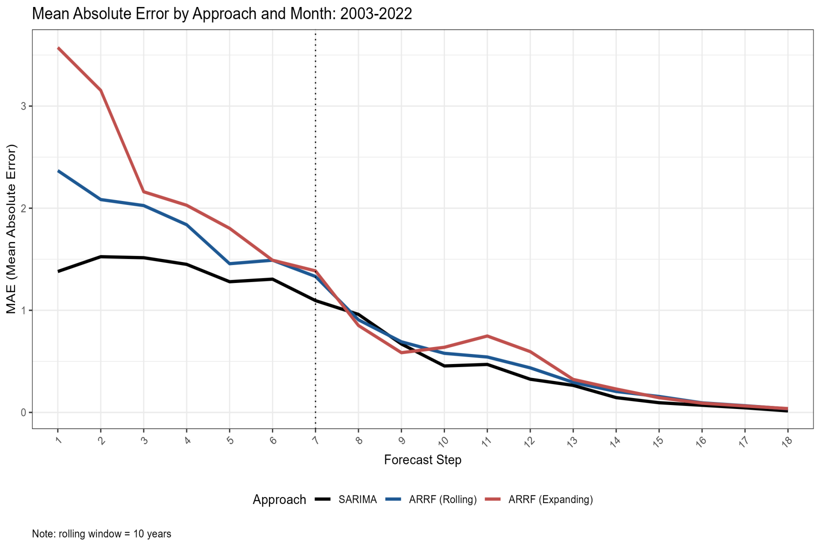
$$\begin{aligned}\min_{j\in\mathcal{J}^-,\, c\in\mathbb{R}} \Big[\min_{\Phi_1} \sum_{\{I_{y,m}\in l_1 | j<c\}} (I_{m,t} - \mathbf{X\Phi_1})^2 + \lambda\|\Phi_1\|_2 + \\ \min_{\Phi_2} \sum_{\{I_{y,m}\in l_2 | j\geq c\}} (I_{m,t} - \mathbf{X\Phi_2})^2 + \lambda\|\Phi_2\|_2 \Big]\end{aligned} \tag{3}$$

Random Forests

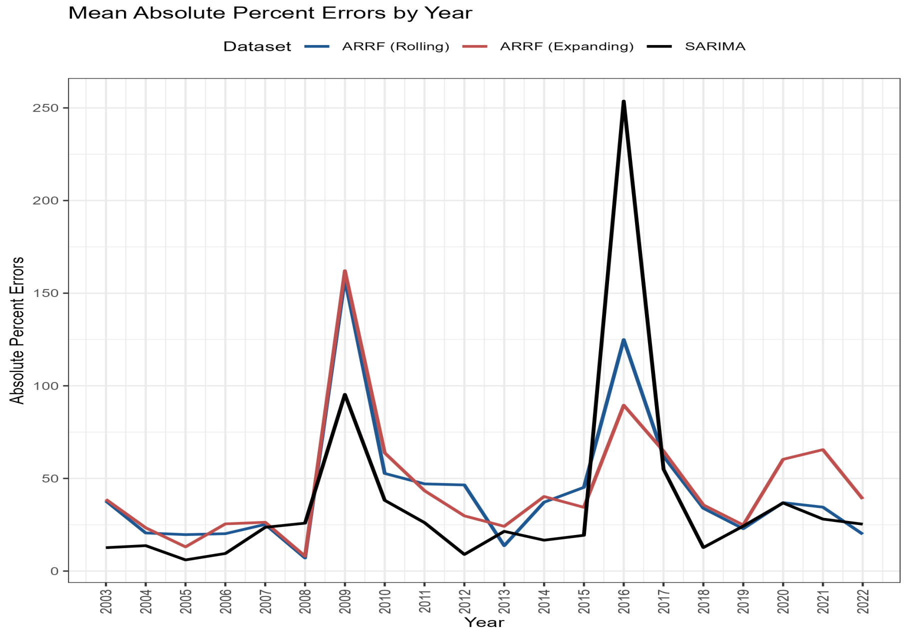
1. Select random samples from the data and create a decision tree using each of them
2. At each split only consider a subset of the available splitting variables
3. Repeat until you have a forest consisting of 1000 trees

Average across all 1000 predictions for a final estimate

Results



Mean Absolute Error by Approach and Month: 2003-2022
MAE (Mean Absolute Error)
0
1
2
3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Forecast Step
Approach
SARIMA
ARRF (Rolling)
ARRF (Expanding)
Note: rolling window = 10 years

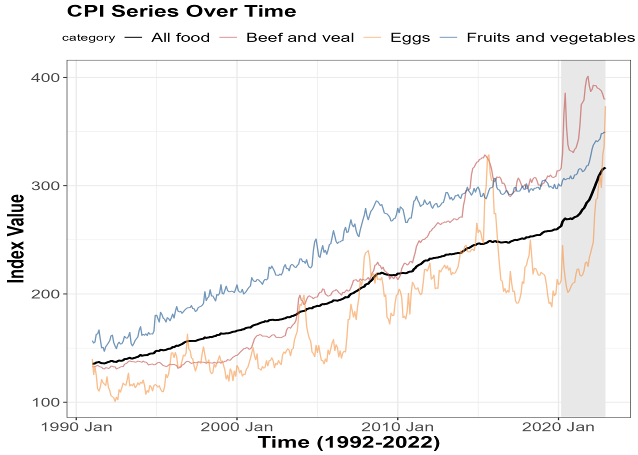

Mean Absolute Percent Errors by Year
Dataset
ARRF (Rolling)
ARRF (Expanding)
SARIMA
Absolute Percent Errors
0
50
100
150
200
250
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Year

Conclusions

- The ARRF Approach was comparable to the SARIMA over steps 6 through 18
- The ARRF approach proved more accurate over certain years

Next Steps

- Explore ARRF when forecasting more variable food price indices (fruits and veggies, eggs, etc.)
- Include Exogenous variables into the ARRF



Thank you!

Farm Foundation, USDA Economic Research Service, and Megan Sweitzer

Presenter: William McWilliams

wnm007@vt.edu