Comprehensive Methodology for Accessible Analysis and Forecasting of Global Agricultural Commodity Prices

Rotem Zelingher

Problem: Global Food Trade Knowledge Gap

Collect & Screen Accessible Data

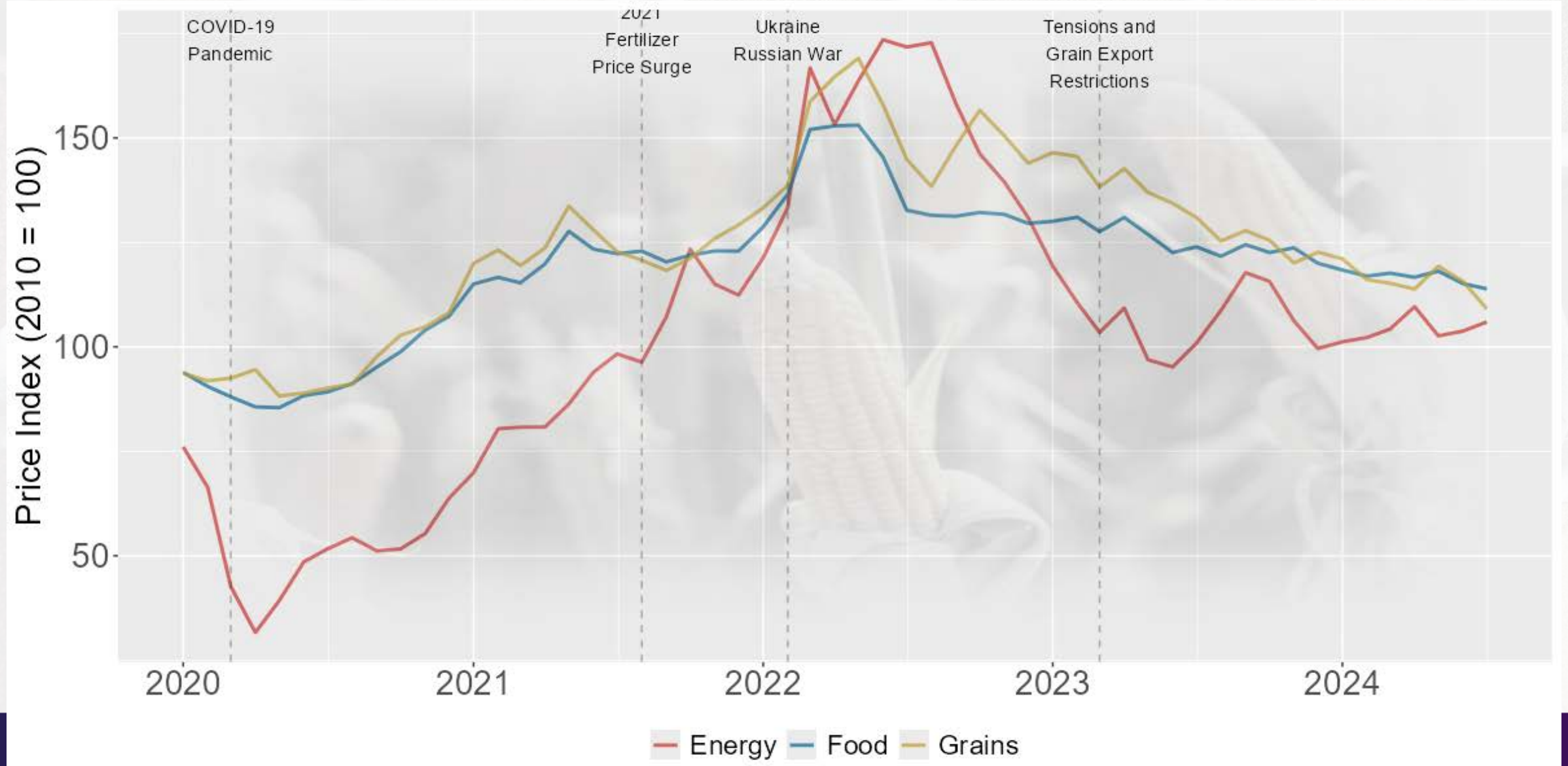
Study & Forecast Price Changes

Explain the results

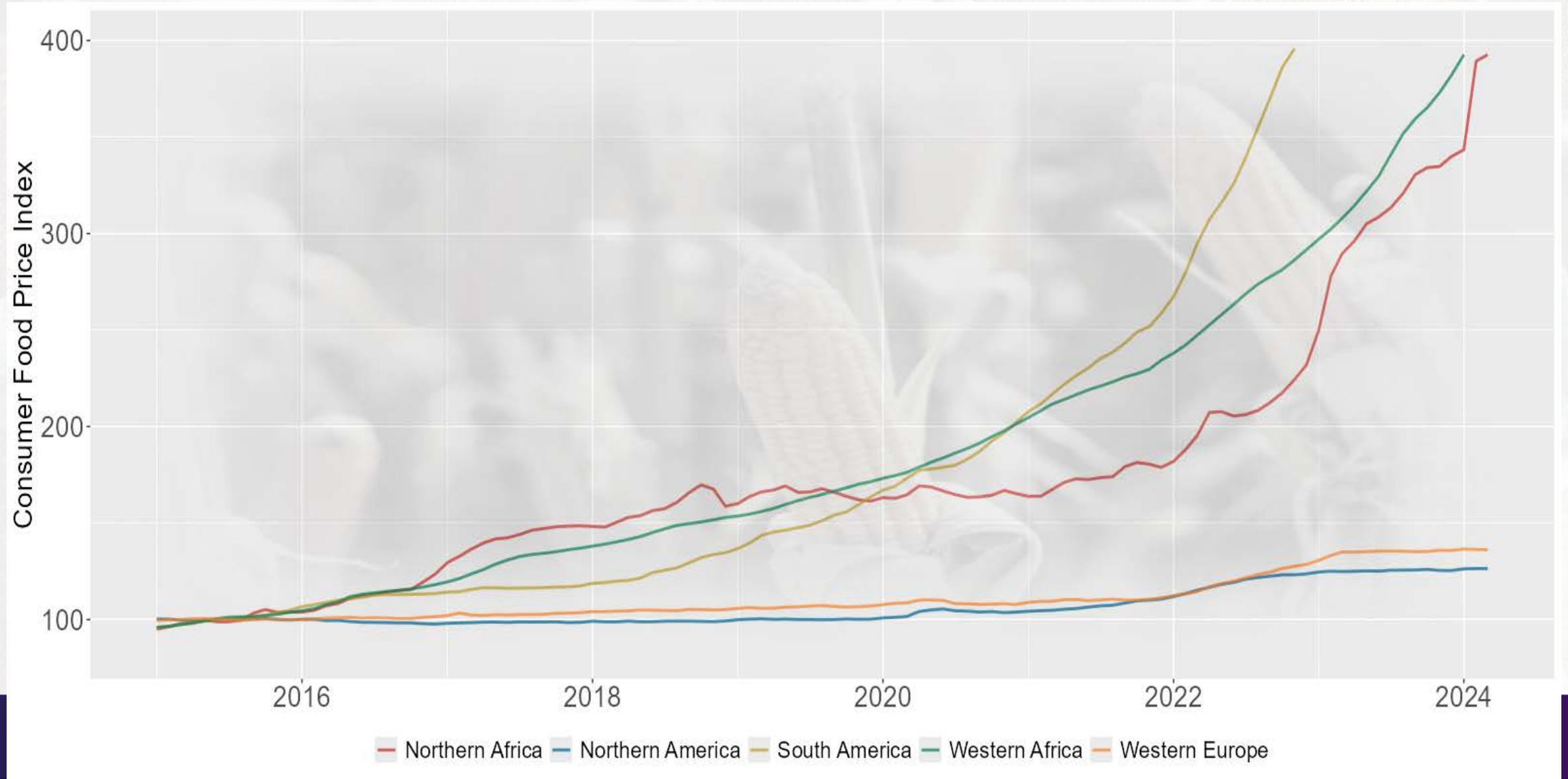
Global Food Trade for Everybody

September 3 | 2024

Commodity Price Indices – High Correlation



High Food Price Volatility in Developing Countries



Research Process – Stages I-III

I. 1st Screening: Build an initial dataset

II. Retrospective analysis & 2nd Screening

III. Forecast price changes

Global Food Trade for Everybody

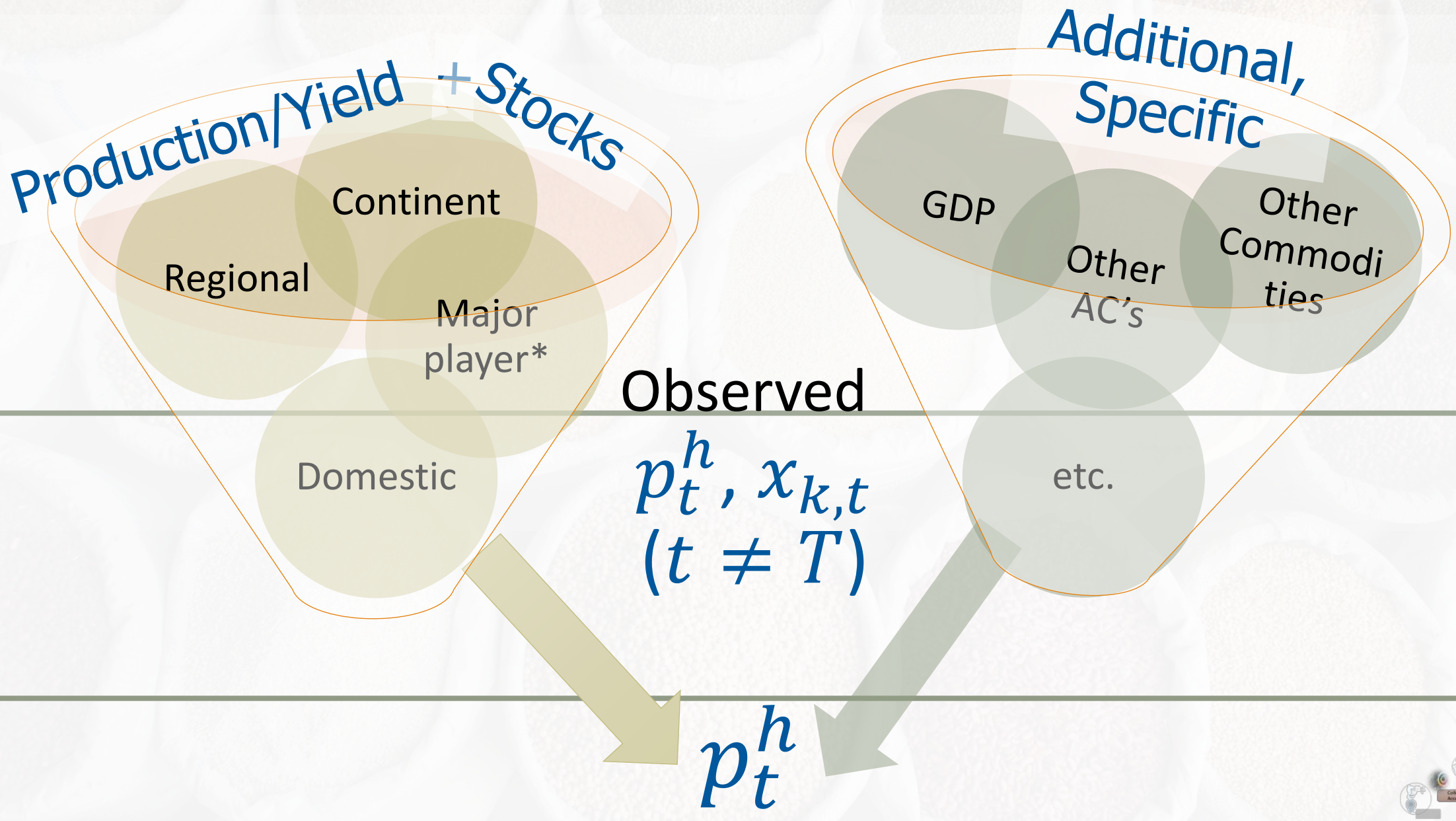
Explain the results

Study & Forecast Price Changes

Collect & Screen Accessible Data

Problem: Global Food Trade Knowledge Gap

Model Overview



Model Specification & Application

- Two steps of $p_{m,Y}^{h}$ investigation
  - **Analysis** → detect previous price changes
    detect influencive factors
  - **Forecasting** → for different forecast horizons
- Approach: Explainable & Black box models
- Output: $p_{m,Y}^{h}$
  - Monthly price, relative annual change $p_{m,Y}$
  - Forecasted 1-12 months ahead $h = 1, \dots, 12$

Algorithm	Stage			Type
	I	II	III	
ARIMA			✓	TS
CART		✓	✓	XML
GAM		✓	✓	XML
GBM		✓	✓	XML
LM		✓	✓	XML
Random Forest		✓	✓	XML
TBATS			✓	TS
VAR	✓		✓	TS
XGBoost (linear & tree)			✓	XML

Stage IV: Explain the Results

I. General Explanation

II. Monthly Explanation

III. Factor Explanation

IV. Explain Specific Event

Global Food Trade for Everybody

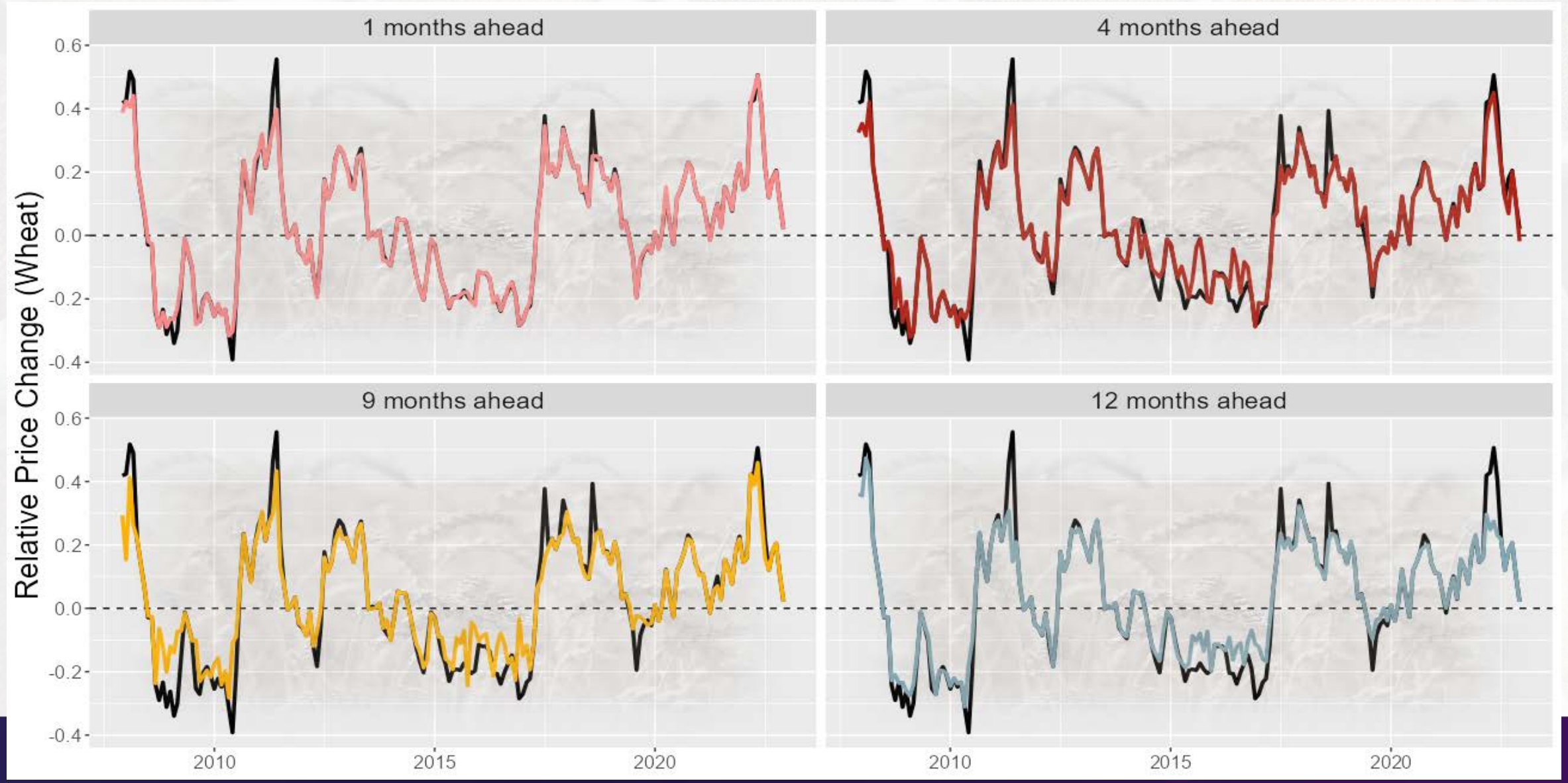
Explain the results

Study & Forecast Price Changes

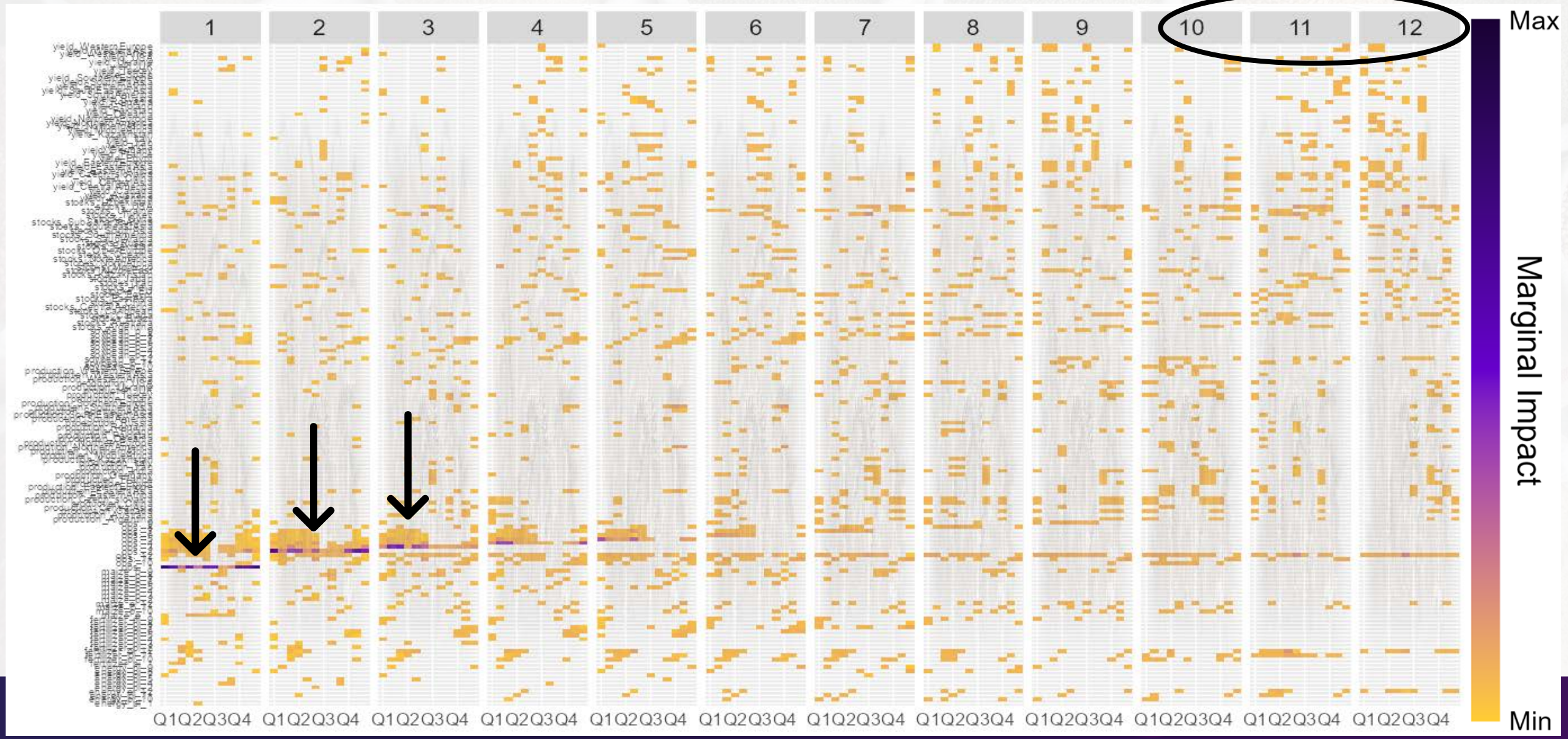
Collect & Screen Accessible Data

Problem: Global Food Trade Knowledge Gap

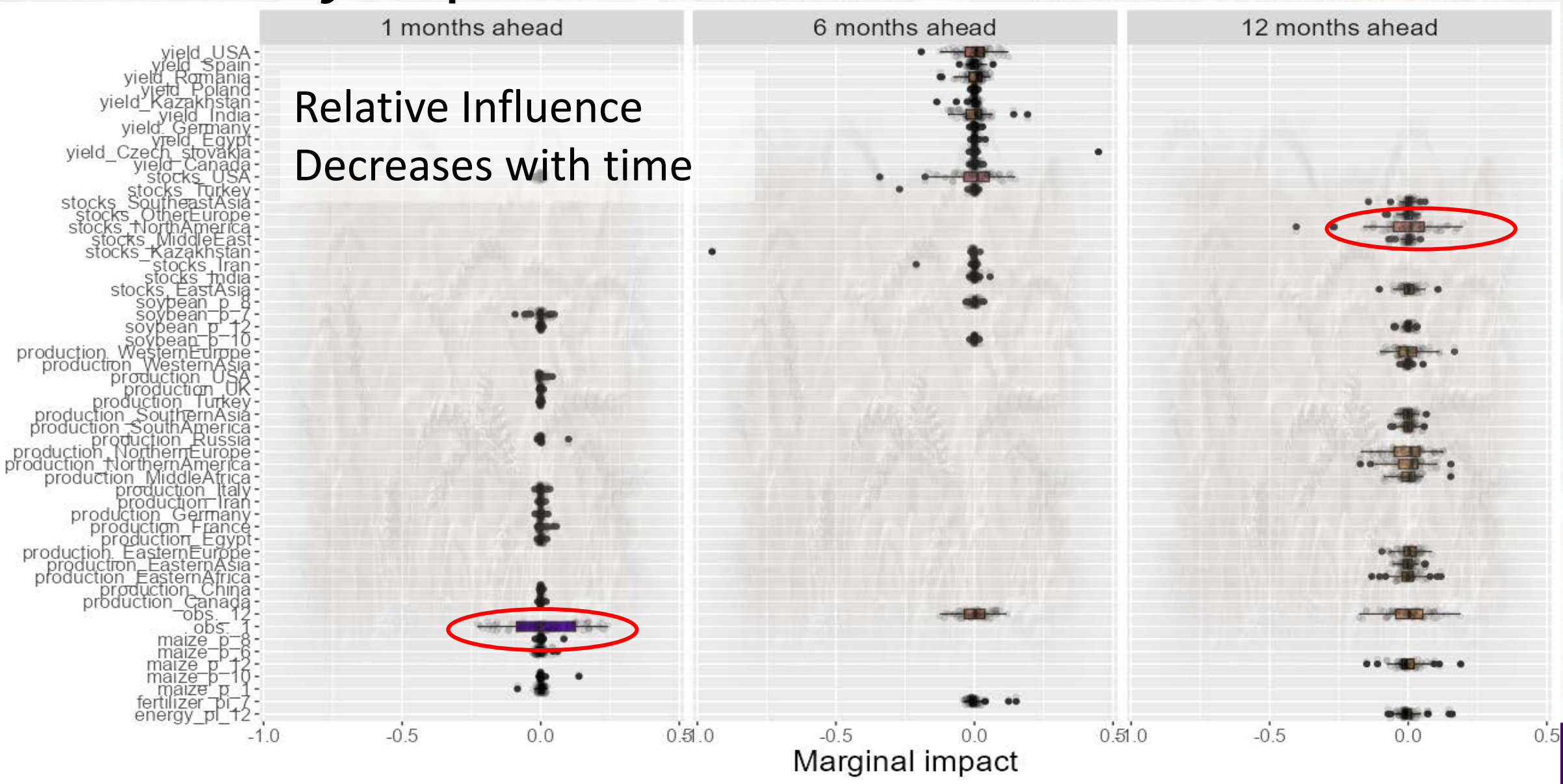
Price: Observed Vs. Forecasted, Wheat



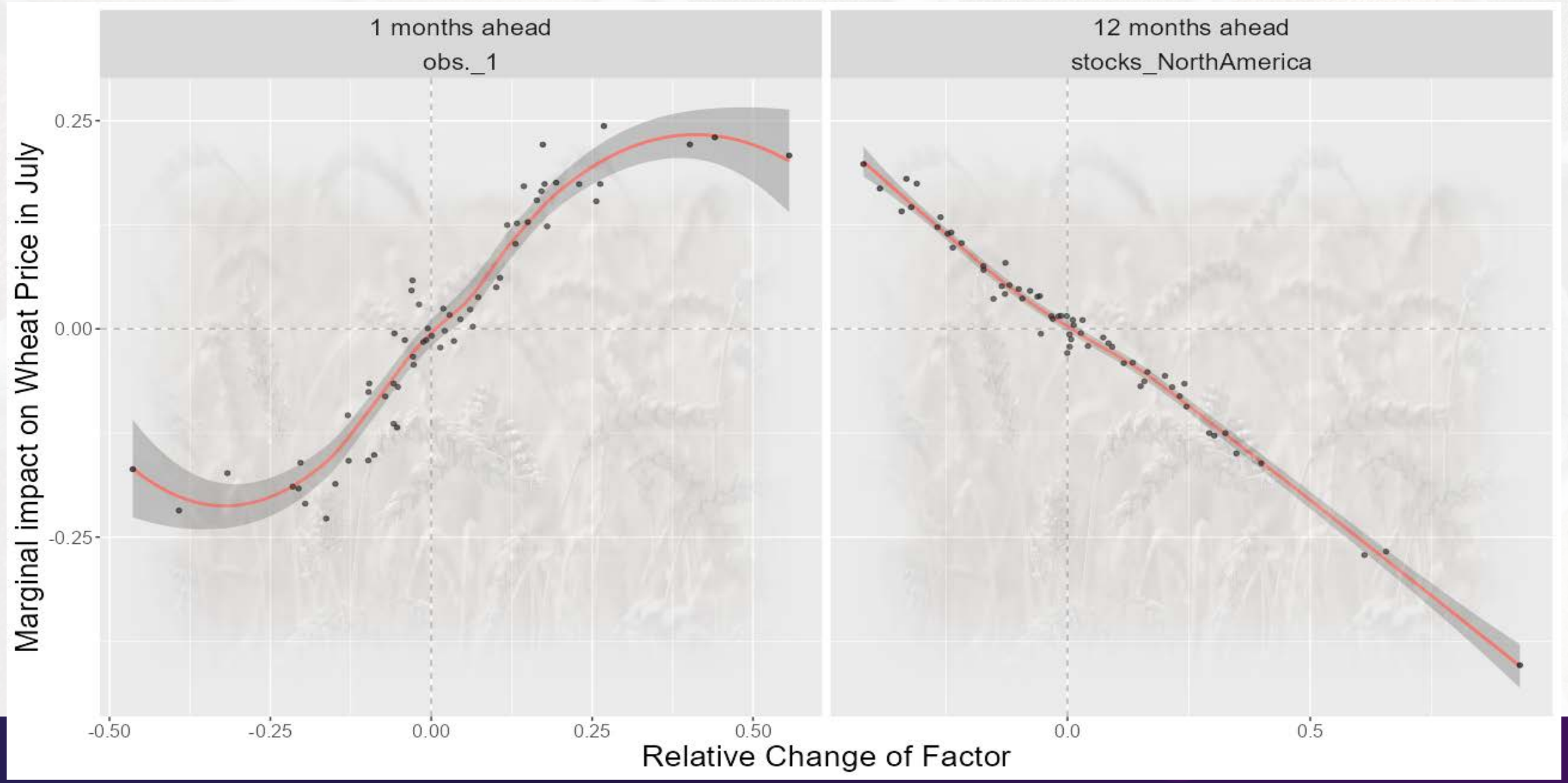
Global Explanation, All Months & Horizons



Monthly Explanation



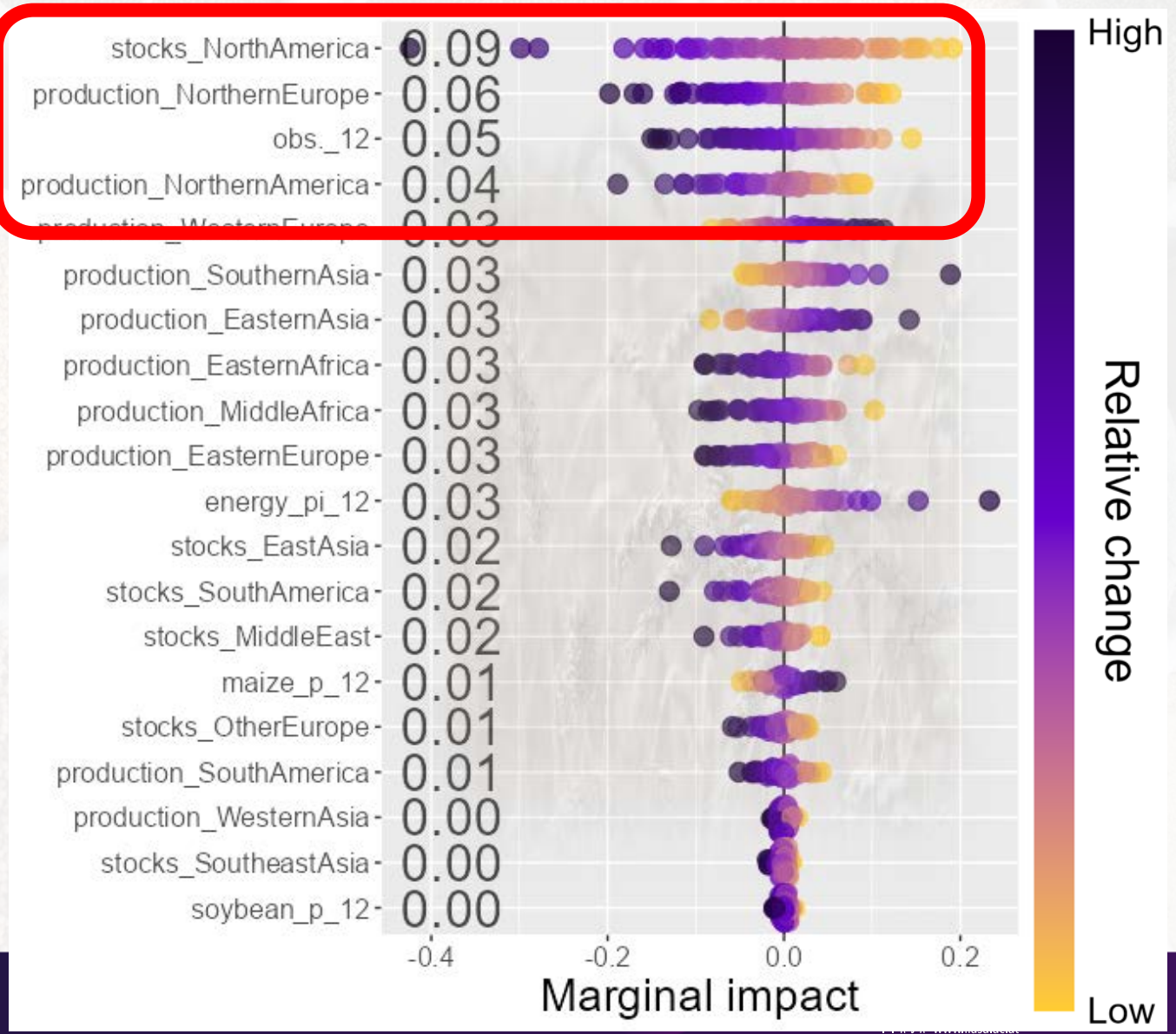
Monthly Explanation, Chosen Factor



Specific Event – Wheat Price, July 2022

h	$p^h_{m,Y}$
1	21.7%
2	21.9%
3	21.5%
4	21.3%
5	25.2%
6	19.7%
7	22.0%
8	17.2%
9	15.8%
10	20.6%
11	19.8%
12	**21.8%**

Observed price change
21.7%



AGRICAF can make price forecasting and analysing [Social Good]

Because Everybody Deserves Food Security!

Global Food Trade for Everybody

Explain the results

Study & Forecast Price Changes

Collect & Screen Accessible Data

Problem: Global Food Trade Knowledge Gap

International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Thank you!

Rotem ZELINGHER

רתם זלינגר

zelingher@iiasa.ac.at

AGRICAF Primary Goal

- Reliable and accessible forecasts
- Empower stakeholders for informed decision-making
- Enhance Food Justice & Social Equity

Why AGRICAF for Food Justice?

Guidelines

- **Accessibility**: Utilizes publicly accessible, regularly updated data.
- **Comprehensiveness**: Integrates XML and econometric methods.
- **Accuracy**: Price forecasts for 1-12 months ahead.
- **Interpretability**: Clear visual explanations.
- **Practicality**: A valuable tool for wide audience

Stage I: 1st Screening, Build an initial dataset

Are the variables potential regressors? › How do they behave together? › New dataset

- Test for stationarity (ADF)
- Associate data relative to forecast horizon

- Correlation between features
- Removal of multicollinearity*
- Importance ranking
- Remove variables with low impact

Part II – 2nd Screening: Retrospective analysis

1. **Split**: Train (i years), Test (1 year) sets
2. **Train** an algorithm using the training set
3. **Identify** prices using the Train set to

$$p_{m,Y} = f\left(x^d_{1,y}, \dots, x^d_{k,y}, x^a_{k,m,y}\right)$$

4. **Test** the algorithms based on test dataset
5. **Assess*** detection accuracy using LOOCV
6. **Rank** features by their contribution level
7. **Filter** – leave most contributing features

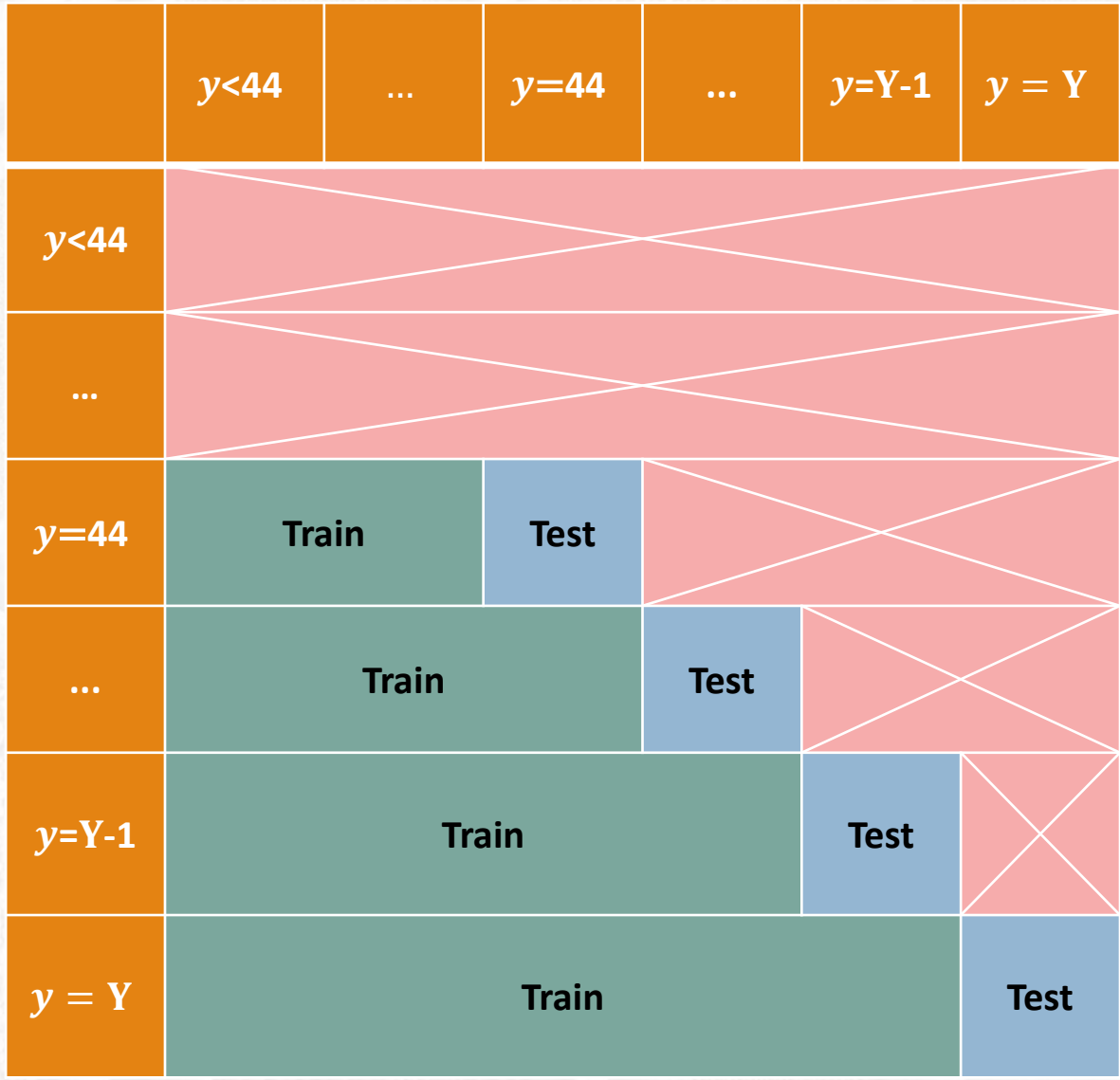
	y=1	y=2	…	…	y=Y-1	$y = Y$
y=1	Test	Train				
y=2	Train	Test	Train			
…	Train		Test	Train		
…	Train			Test	Train	
y=Y-1	Train				Test	Train
$y = Y$	Train					Test

Part III – Price forecast

1. **Split** data: Train (45≤ i years),
   Test (1 year) sets
2. **Train** an algorithm using the training set
3. **Forecast** using the Train set to

$$p_{m,Y} = f\left(x^d_{1,y}, \dots, x^d_{k,y}, x^a_{k,m,y}\right); Y = y_{max} + 1$$

4. **Test** the algorithms based on test dataset
5. **Assess** forecasting accuracy using Rolling CV
6. **Rank** features by their contribution level



Model Assessement & Application

- Two types of $p_{m,Y}$ investigation
  - **Analysis** → detect previous price
  - **Forecast** → for different time zones
- Approach: Explainable Vs. Black box models
- Model accuracy assessment**

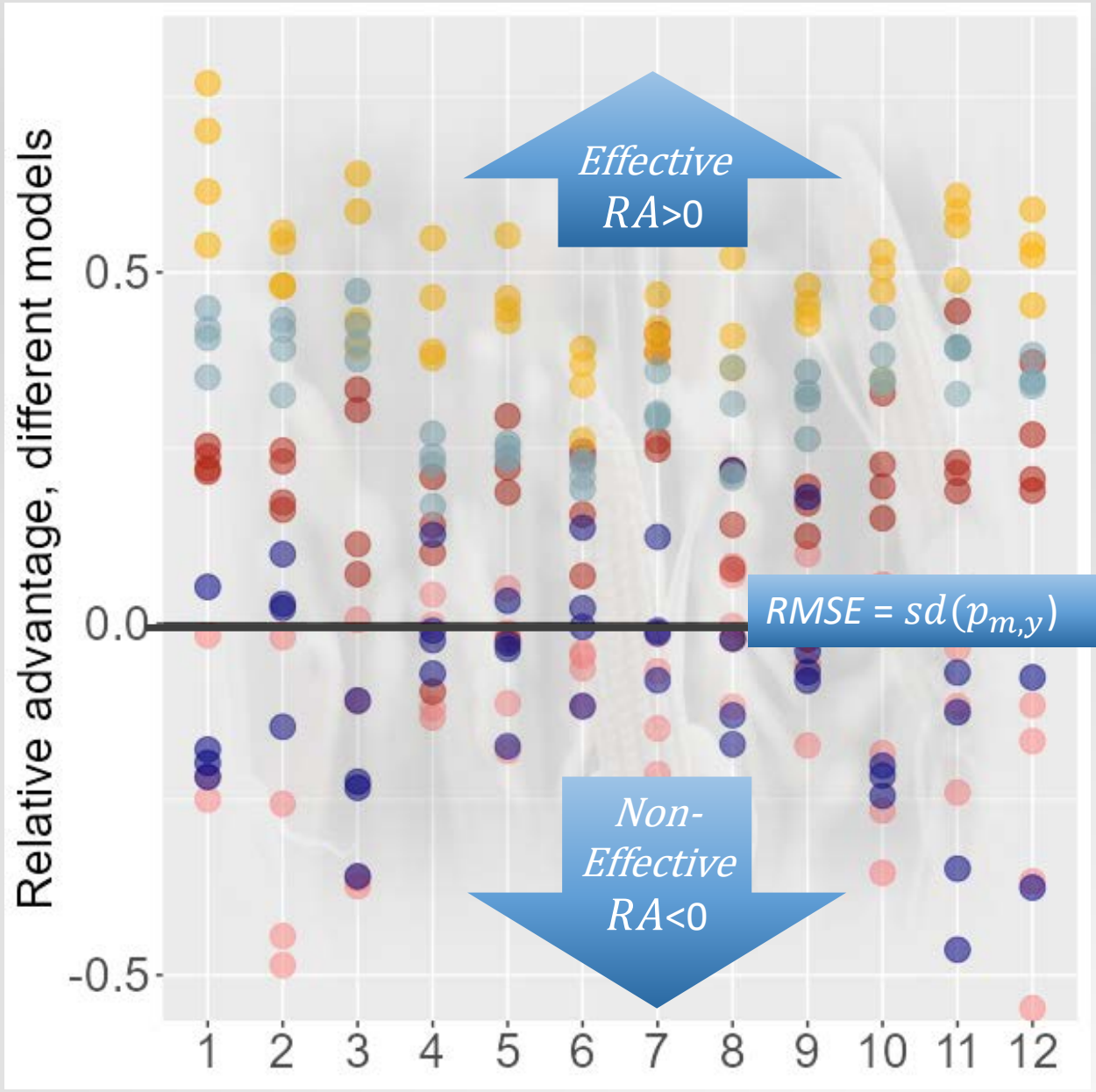
Metric	Formula	Stage
CP	$\frac{1}{T}\sum_{t=1}^{T}(step)\frac{p_t-\hat{p}_t}{error_t}$	2,3
MAE	$\frac{1}{T}\sum_{t=1}^{T}\lvert p_t-\hat{p}_t\rvert$	2,3,4
MAPE	$\frac{1}{T}\sum_{t=1}^{T}\lvert\frac{p_t-\hat{p}_t}{error_t}\rvert$	2,3
R^2	$1-\frac{\sum_{t=1}^{T}(p_t-\hat{p}_t)^2}{\sum_{t=1}^{T}(p_t-\bar{p}_t)^2}$	2,3
RMSE	$\frac{1}{T}\sum_{t=1}^{T}(p_t-\hat{p}_t)^2$	2,3
RA	$1-\frac{RMSE}{sd(p_t)}$	2,3,4

Relative advantage

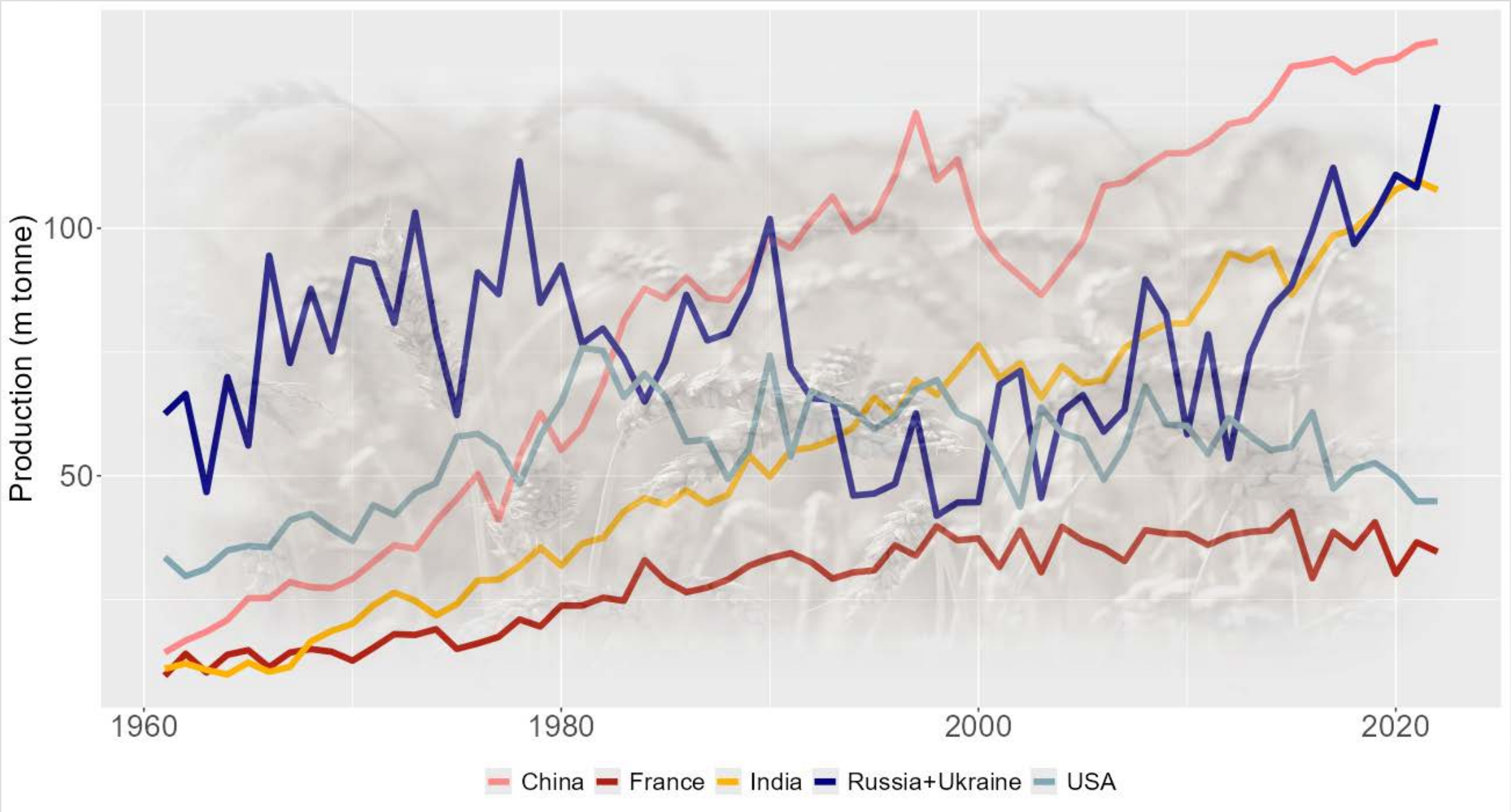
$$\mathrm{RA} = 1 - \frac{RMSE}{sd(p_{m,y})}$$

Higher = Better performance

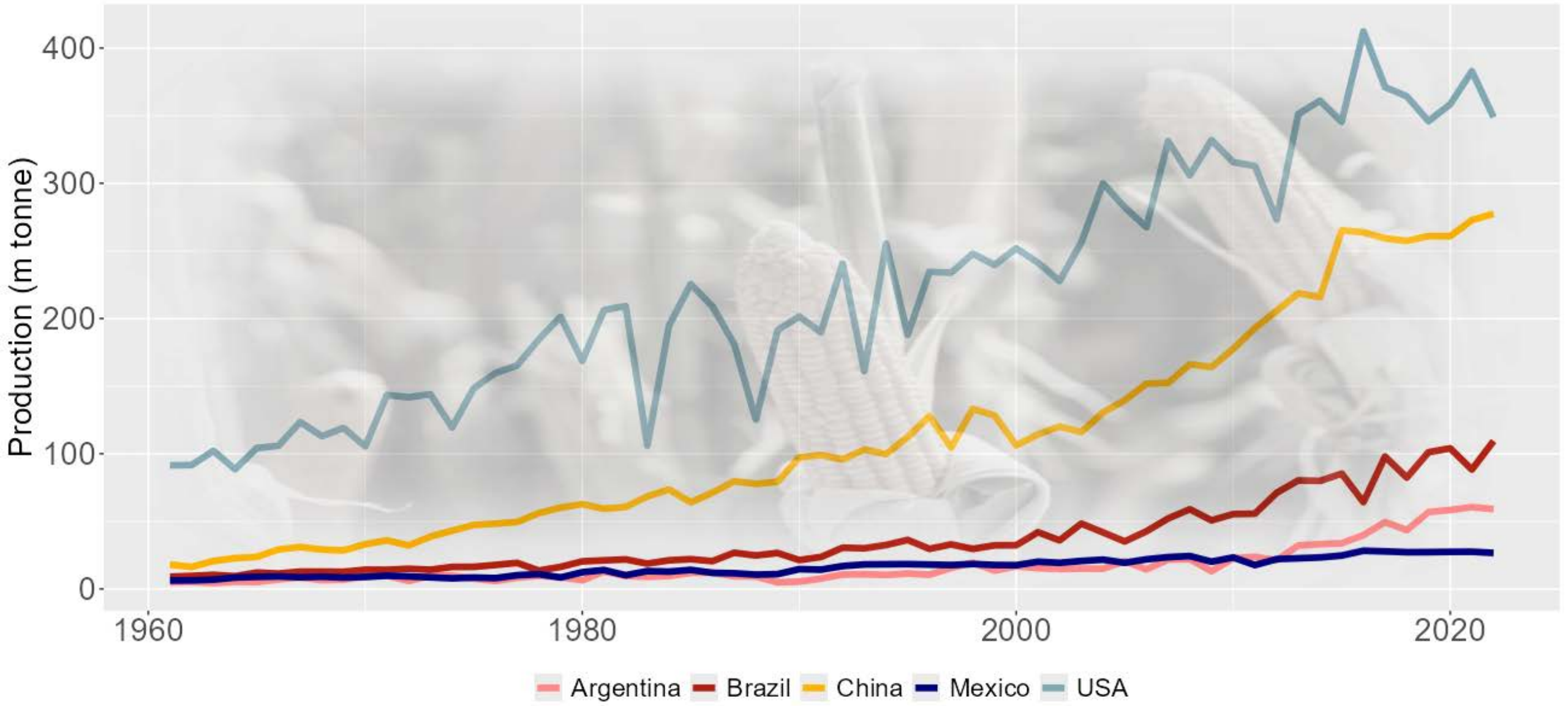
Lower = Worse performance



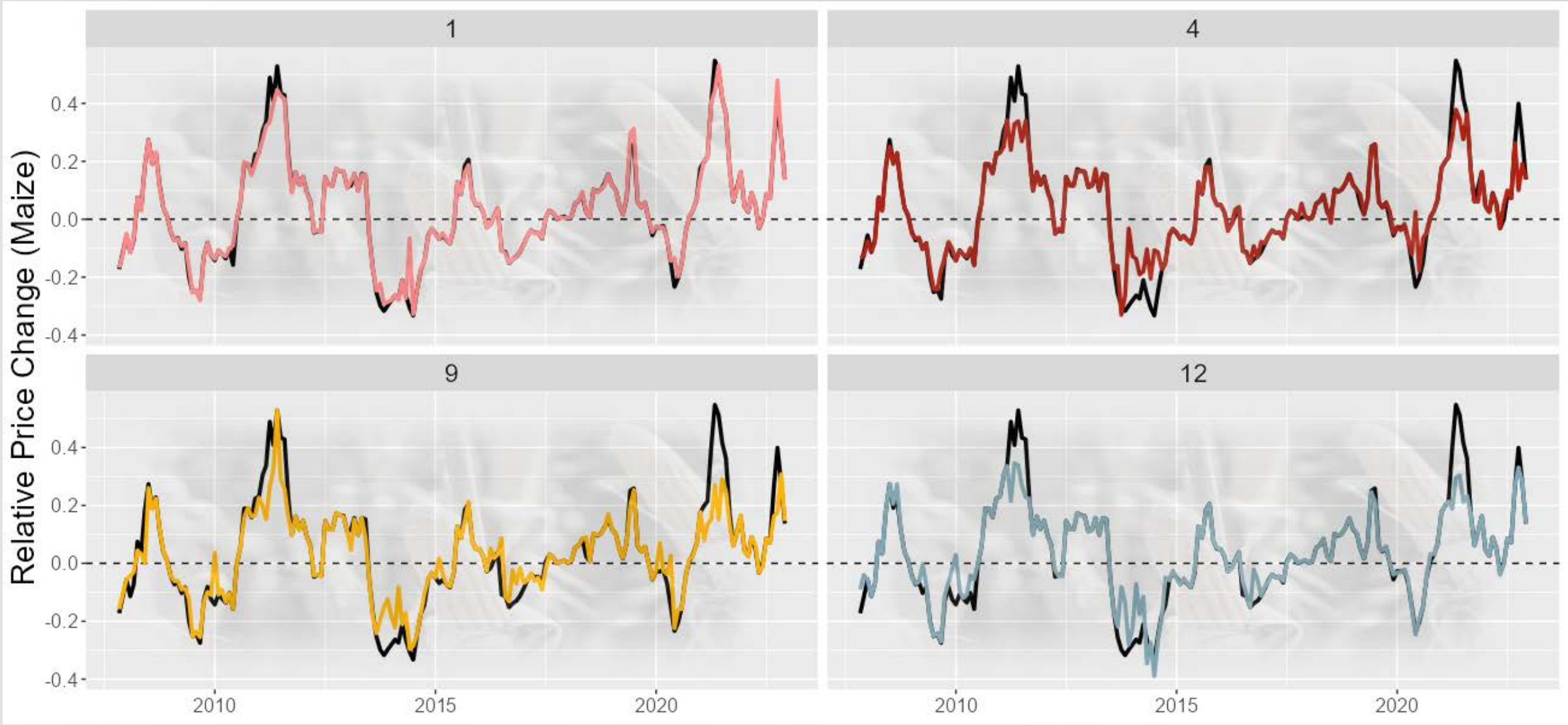
Wheat Top Producers – Concentrated Market



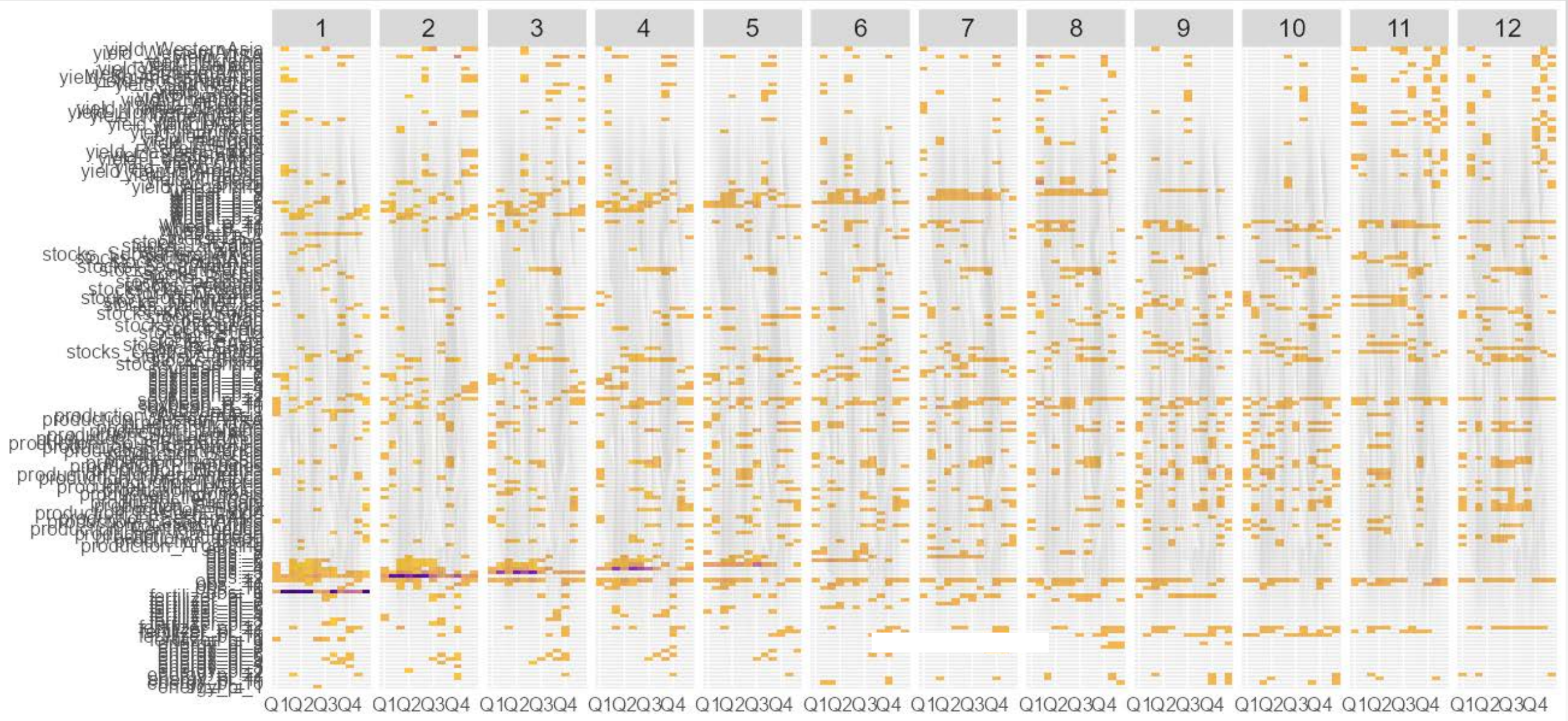
Maize Top Producers – Concentrated Market



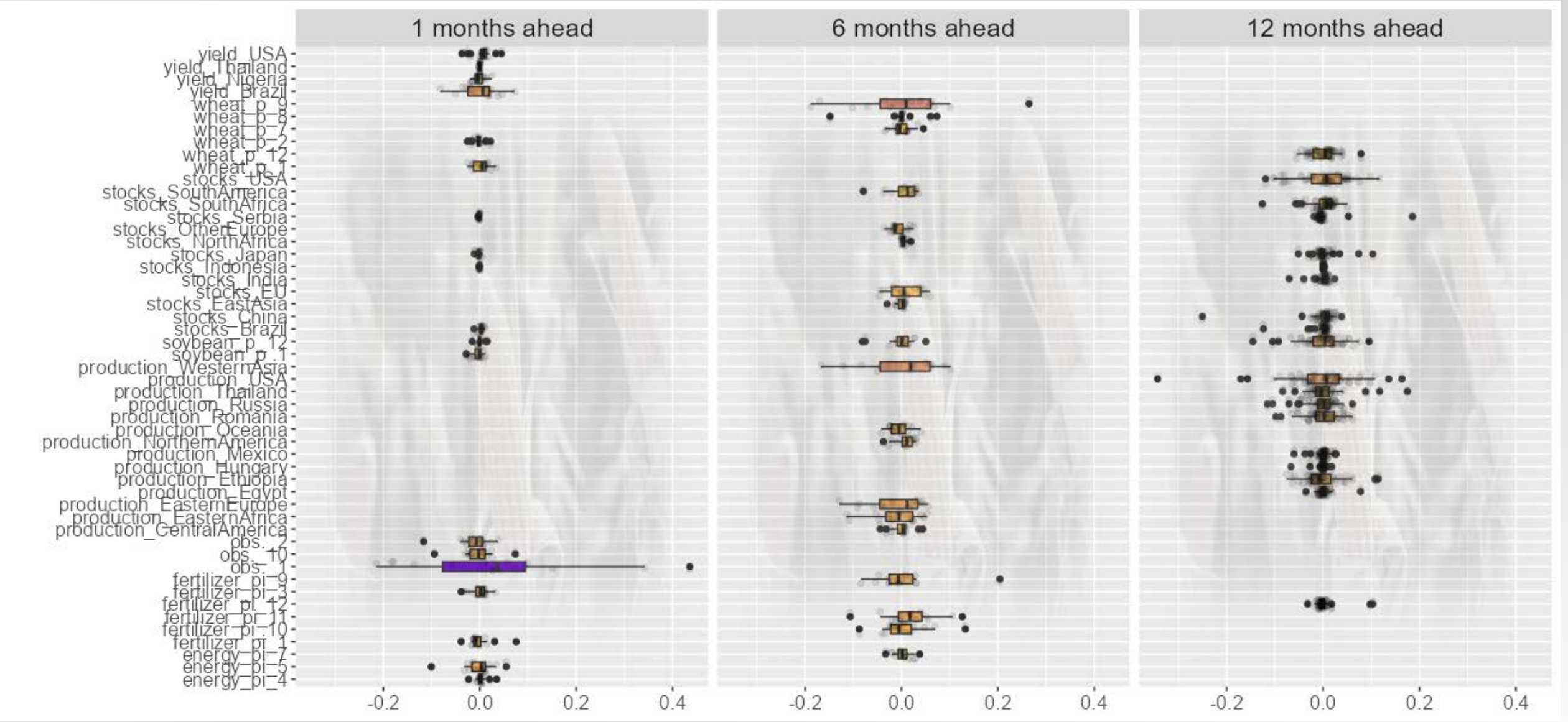
Price: Observed Vs. Forecasted, Maize



Global Agnostic, Maize: General Relative Importance



Marginal Impact on Maize Price, September



Partial Dependence

- Average response of $p_{m,y}$ to variations of USA's maize production
- PDP shows negative correlation

$$x_{k,y} \downarrow \Rightarrow \quad p_{m,y} \uparrow$$