



# WHEAT PRICE PREDICTION

---

Business, Economic and Financial Data

Supervisor: Prof. Mariangela Guidolin

Karakus Isikay  
Mahfouz Auriane



# OUTLINE

---

1

INTRODUCTION

2

EXPLORATORY DATA  
ANALYSIS

3

DATA MODELLING

4

MODELS COMPARISON &  
FORECASTING

# 1

## INTRODUCTION

---

---

A vital food source for all living beings  
aka “edible gold”





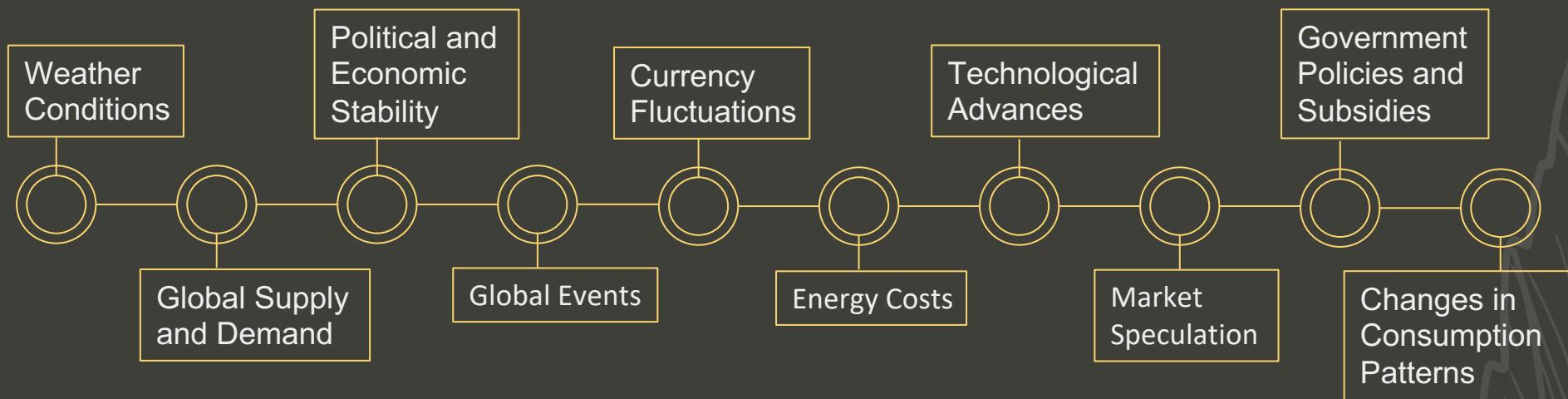
# PROJECT SPECIFICATIONS

---

- Our aim is to develop a reliable model for predicting and forecasting wheat prices for the next 5 years.

# FACTORS THAT INFLUENCE THE WHEAT PRICE

---



# DATA COLLECTION & DATASET INTRODUCTION



Wheat price (1990-2022) [Main Dataset]  
Pasta price (1984-2023)  
Flour price (1980-2023)



Food and Agriculture Organization  
of the United Nations

Export Quantity (1990-2022)  
Export Value (1990-2022)  
Import Quantity (1990-2022)  
Import Value (1990-2022)



National Centers for  
Environmental Information  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Global Land and Ocean  
January-December Temperature Anomalies  
(1990-2023)



Crop Production (1990-2023)



Economic Research Service  
U.S. DEPARTMENT OF AGRICULTURE

Supply and Disappearances (2003-2023)

# 2

## EXPLORATORY DATA ANALYSIS

---

---

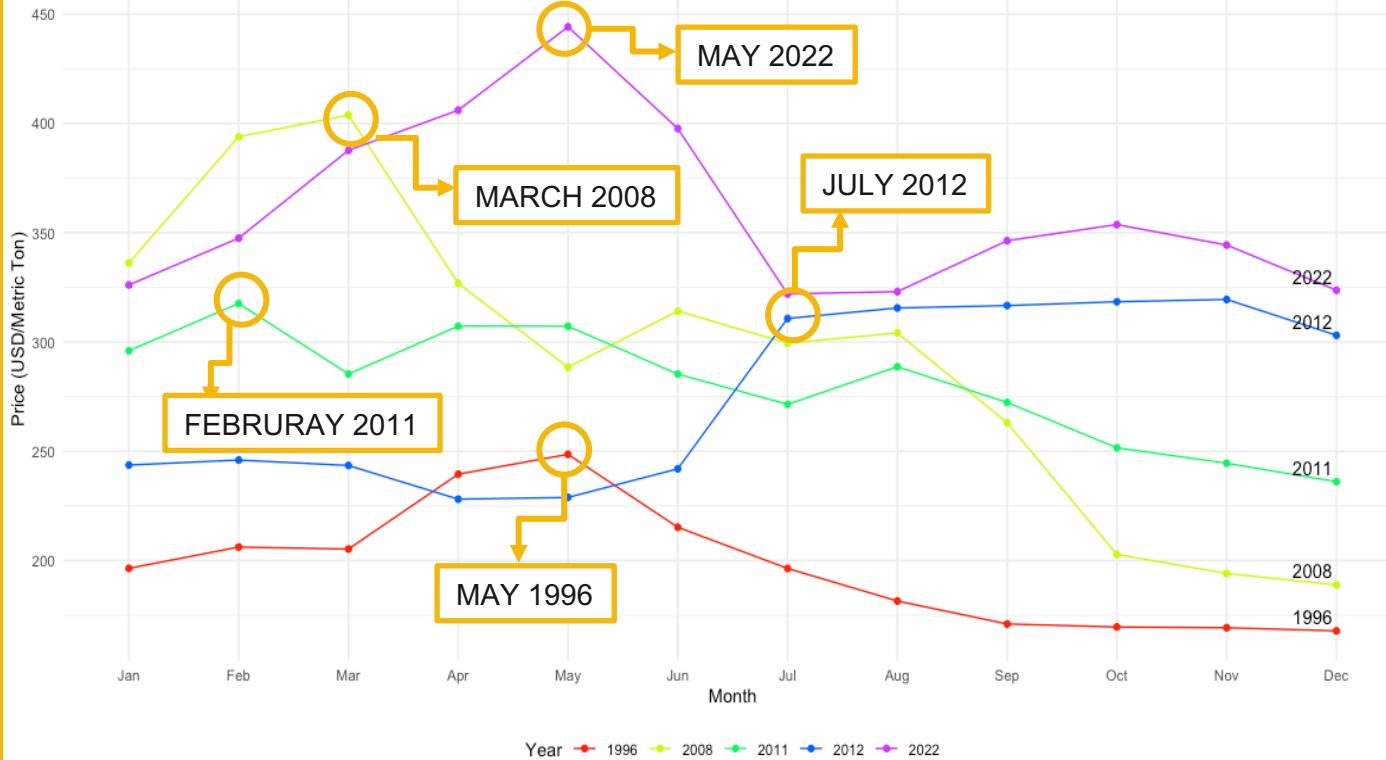


# WHEAT PRICE

Average Wheat Prices Over Years



Global Price of Wheat for Selected Years

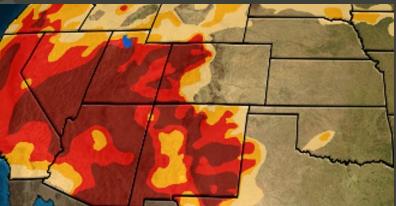


WHAT HAPPENED EXACTLY?

# WHAT HAPPENED IN THESE YEARS & HOW DID IT AFFECT THE WHEAT PRICE?



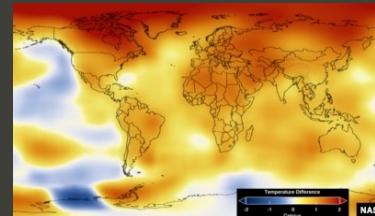
Drought in  
the Midwest and West  
**1996**



**2008**  
Global Food Crisis

NASA: 2012 Was 9th Hottest  
Year on Record for the 36th  
years in a row

**2011-2012**



Floods In Argentina And Australia In 2011  
Export ban of Russia

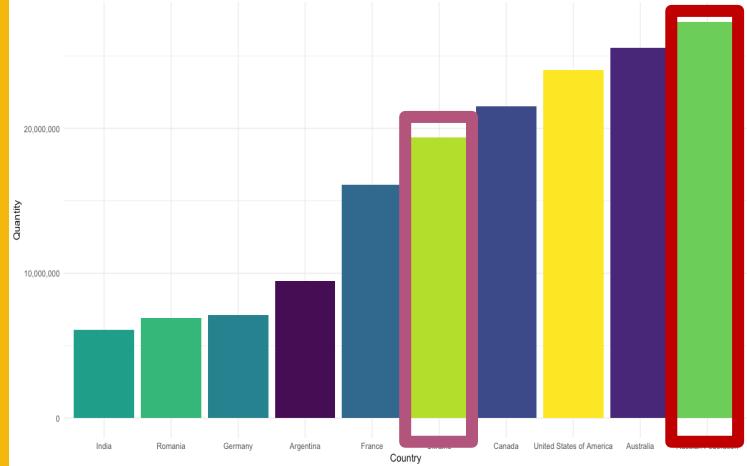


**2022**  
Russian invasion of  
Ukraine

# TOP 10 WHEAT EXPORTERS

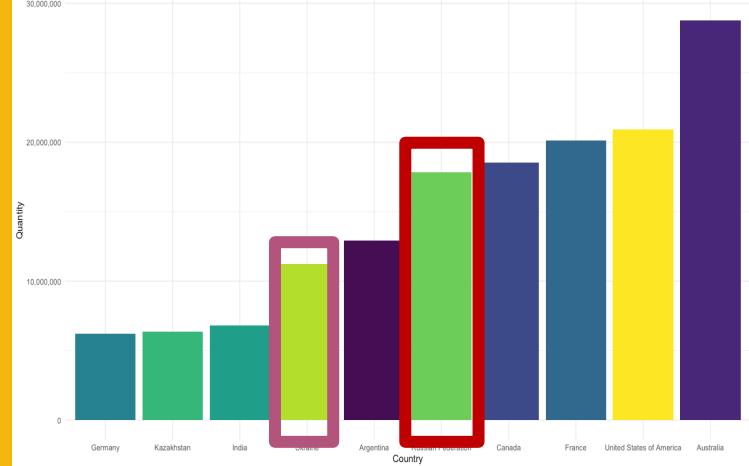


Top 10 Wheat Exporters by Quantity in 2021



2021

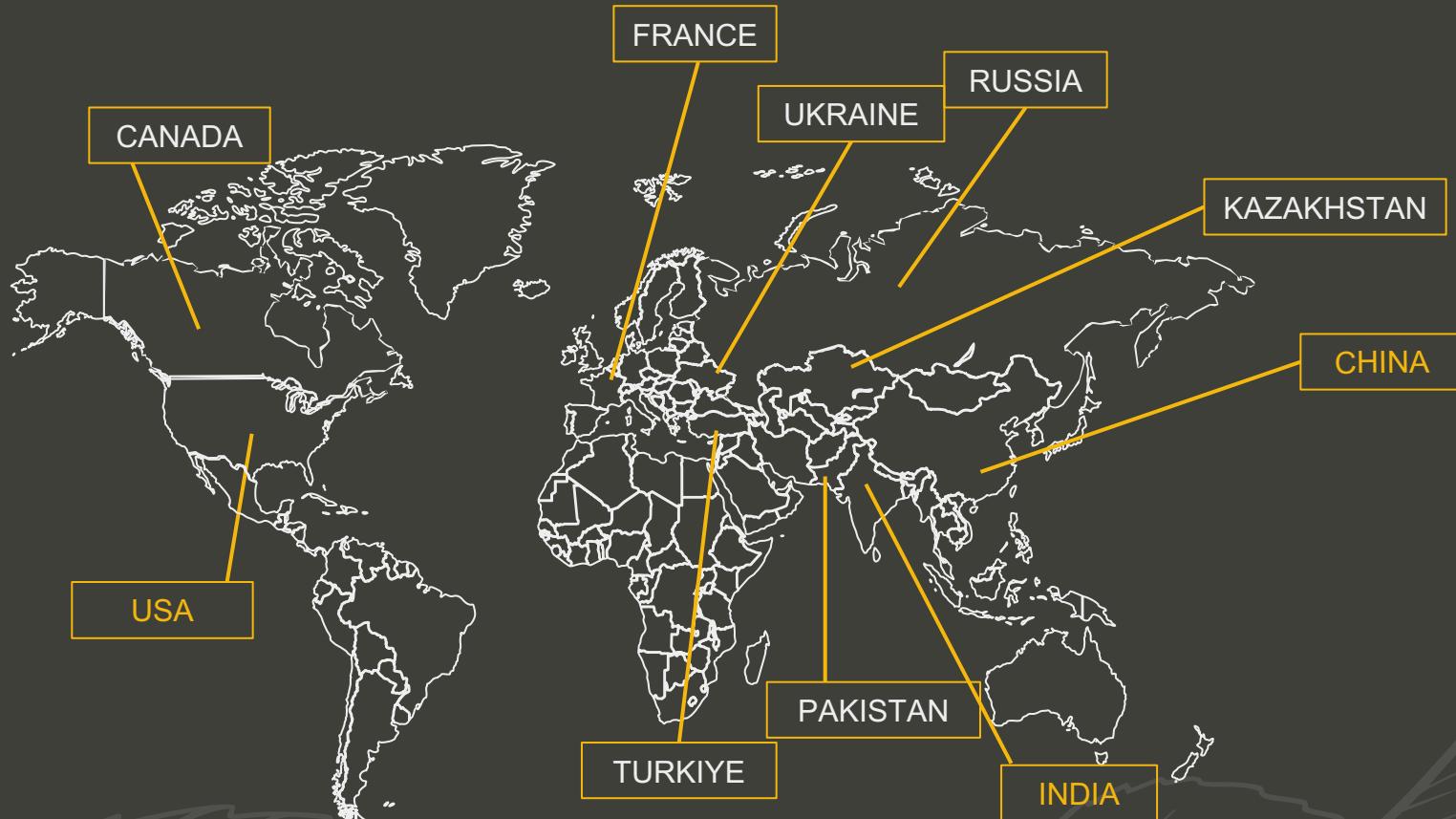
Top 10 Wheat Exporters by Quantity in 2022



2022

# TOP PRODUCER COUNTRIES

---





A wide-angle photograph of a wheat field at sunset. The sky is filled with dramatic, colorful clouds ranging from deep blue to bright orange and yellow. The sun is low on the horizon, casting a warm glow over the entire scene. In the foreground, the golden-brown wheat stalks are visible, swaying slightly. A large, stylized yellow arrow points from the left side of the image towards a yellow rectangular box on the right.

**CLIMATE DATA** 1990-2023

**INTERNATIONAL TRADE DATA** 1990-2022

**SUPPLY AND DISAPPEARANCES** 2003-2023

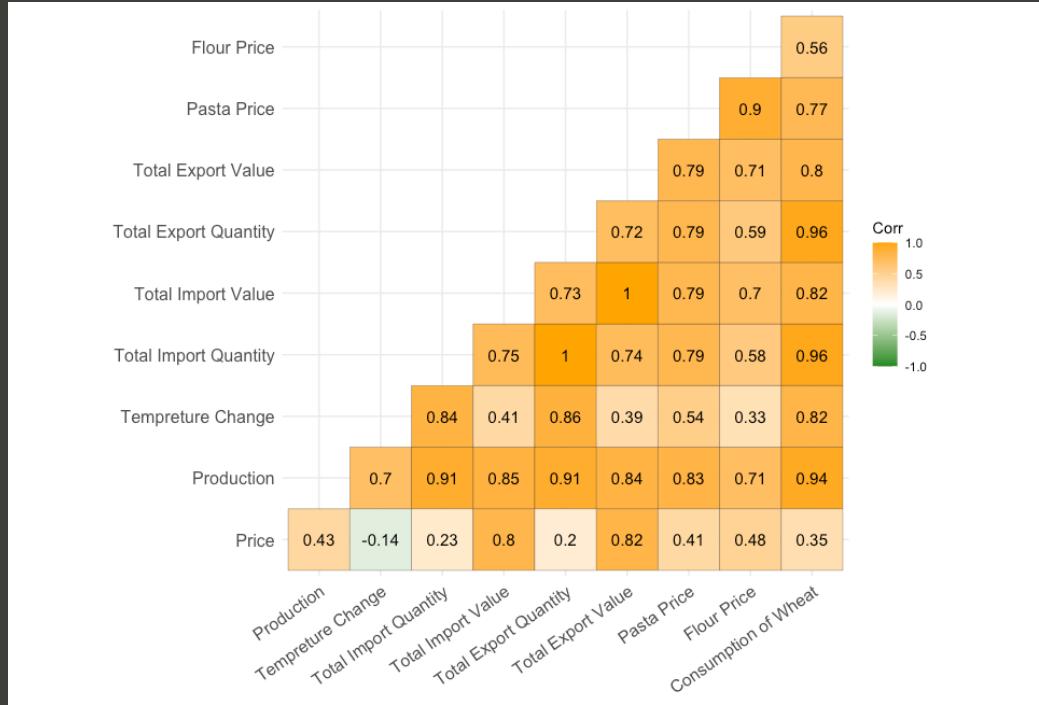
**COMMODITY PRICE DATA** 1990-2022

**CROP PRODUCTION DATA** 1990-2023

**COMBINED DATASET**

**2003-2022**

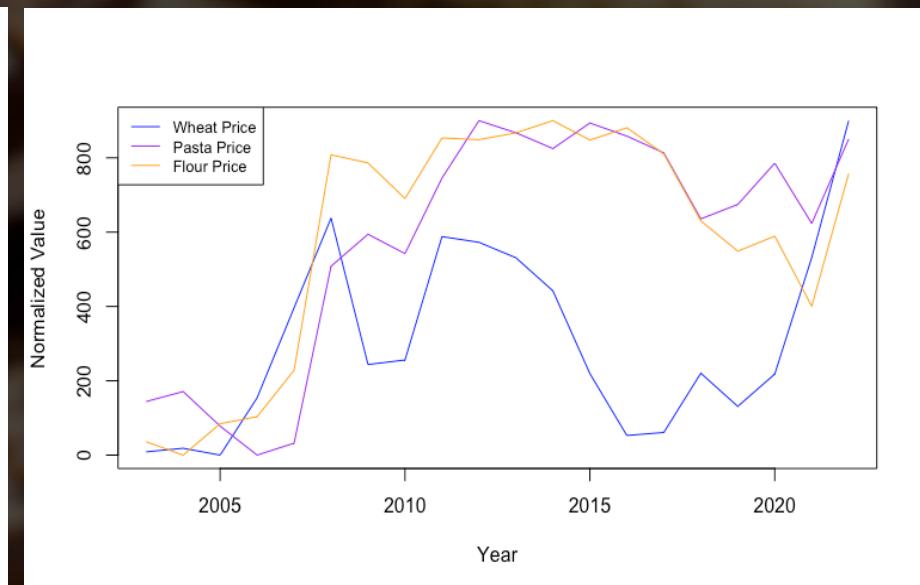
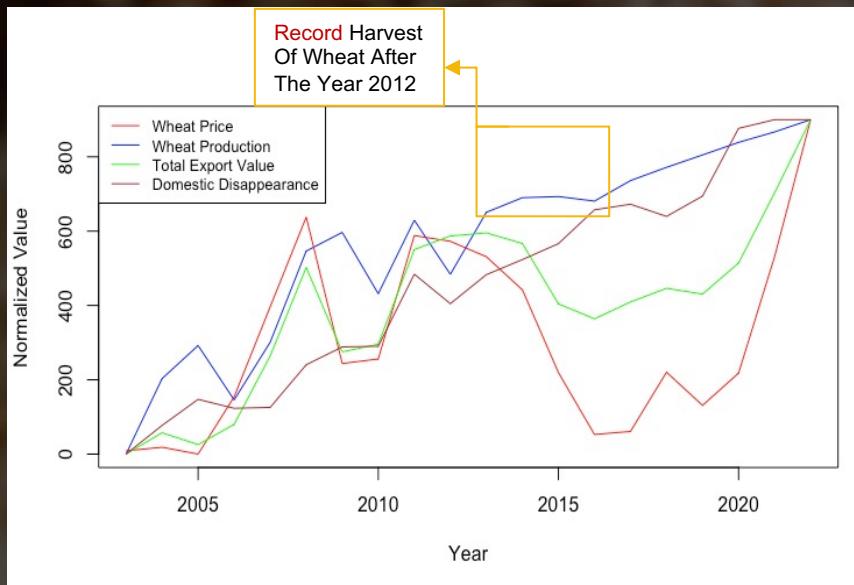
# CORRELATION MATRIX



## FEATURE SELECTION

- Price
- Total Export Value
- Pasta Price
- Flour Price
- Consumption of Wheat
- Production

# RELATION BETWEEN SELECTED VARIABLES



# 3

## DATA MODELING



# MODELS

---



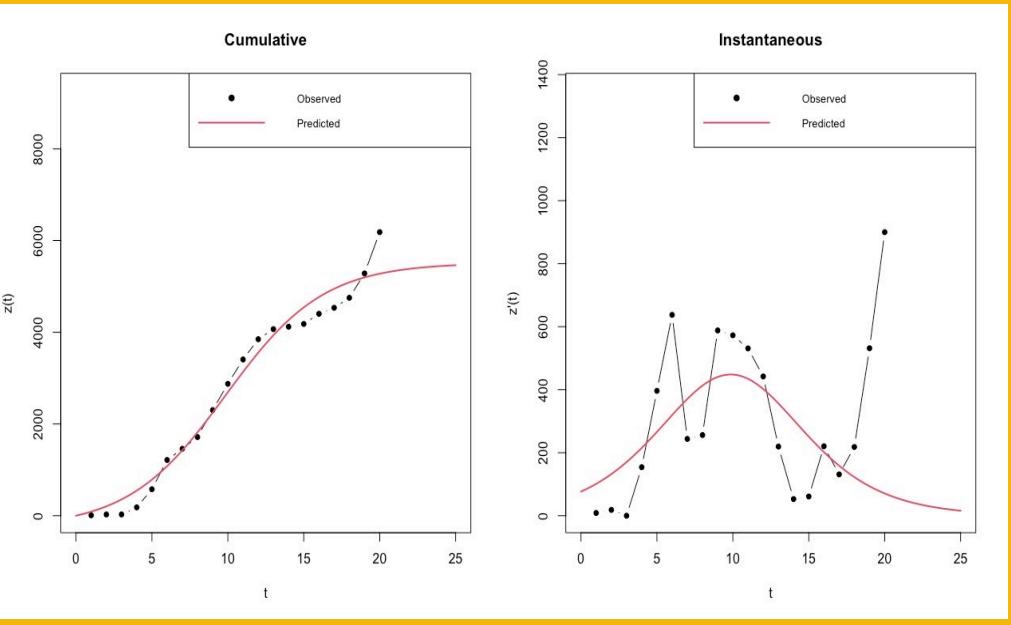
## PREDICTION

- LINEAR REGRESSION MODEL
- BASS MODEL
- GENERALIZED BASS MODEL
- LOCAL REGRESSION AND SPLINES
- LASSO REGRESSION

## FORECASTING

- ARIMA
- ARMAX
- EXPONENTIAL SMOOTHING
- GENERALIZED ADDITIVE MODEL

# BASS MODEL



```
> summary(bass_model_selected)
```

```
Call: ( Standard Bass Model )
```

```
BM(series = selected_dataset$Price, display = TRUE)
```

Residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-414.64	-329.93	-91.11	-39.07	134.00	905.79

Coefficients:

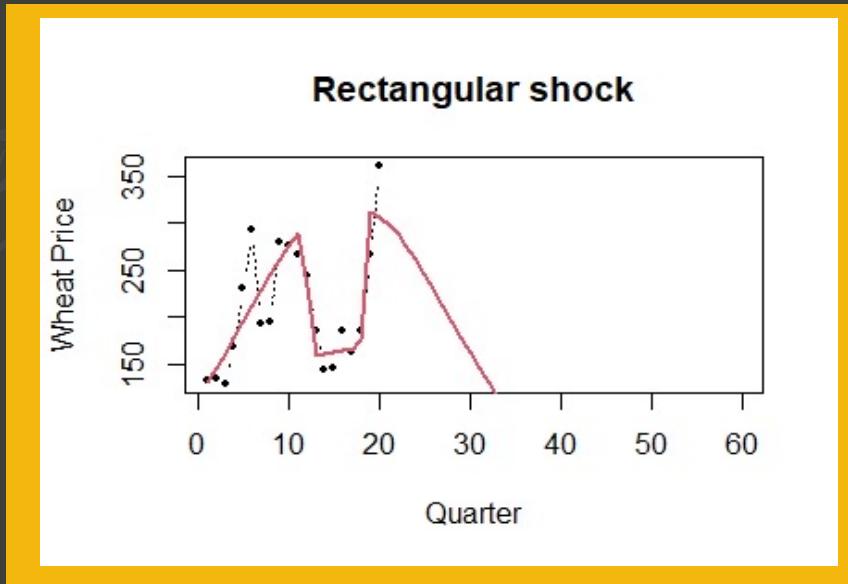
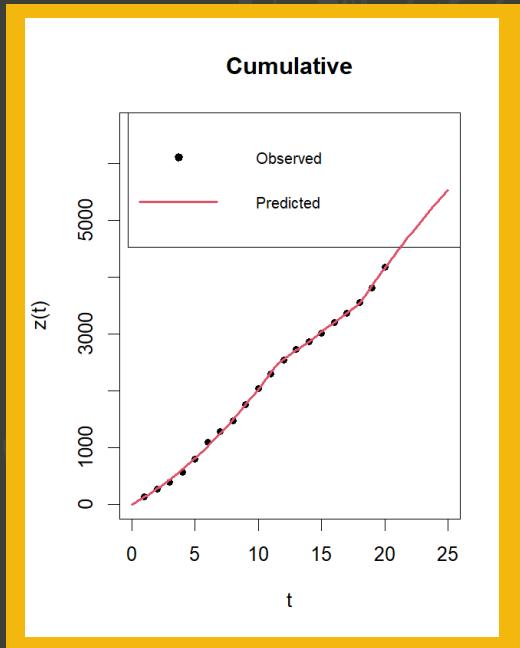
	Estimate	Std.Error	Lower	Upper	p-value						
m	5.514264e+03	3.279420e+02	4.87151e+03	6.157019e+03	4.99e-12 ***						
p	1.387943e-02	4.166958e-03	5.71234e-03	2.204651e-02	3.96e-03 **						
q	2.967984e-01	5.669769e-02	1.85673e-01	4.079238e-01	6.73e-05 ***						
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'..'	0.1	' '	1

Residual standard error 340.1046 on 17 degrees of freedom

Multiple R-squared: 0.989884 Residual sum of squares: 1966409

- Cumulative data is almost an S-shaped curve
- Instantaneous data is a Bell-shaped curve
- The parameters are significant

# GENERALIZED BASS MODEL: RECTANGULAR SHOCK

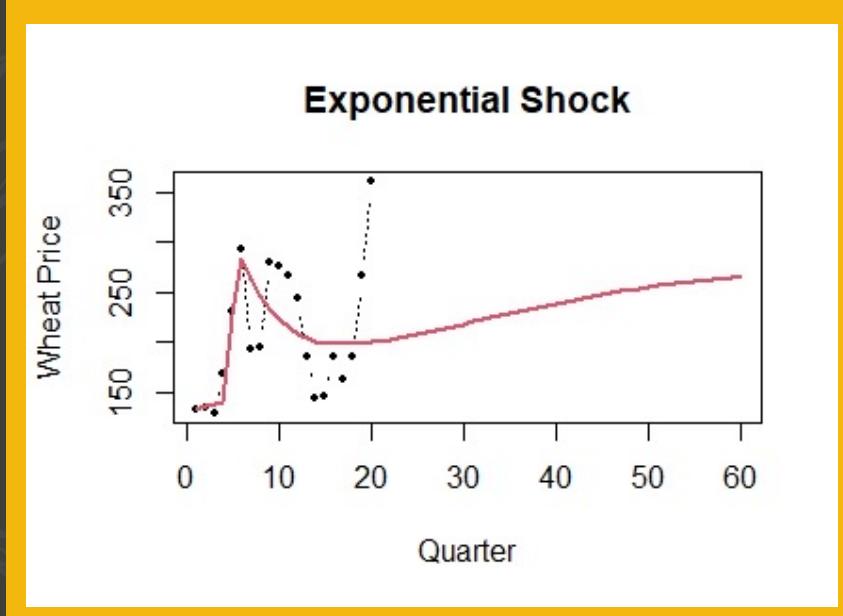
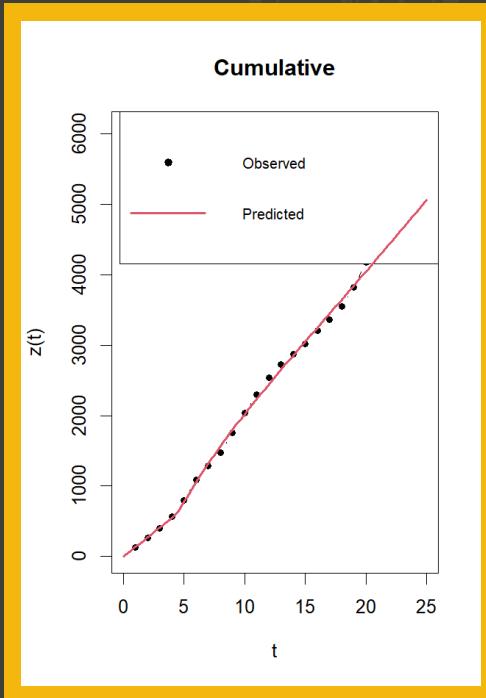


(Rectangular shock (2016-2021) due to mass production, followed by the Russian invasion of Ukraine.)

## ERROR MEASURES

- MSE: 19673.17
- RMSE: 140.2611

# GENERALIZED BASS MODEL: EXPONENTIAL SHOCK



• The **2008 global food crisis** was a pivotal event that led to a substantial increase in food prices worldwide.

## ERROR MEASURES

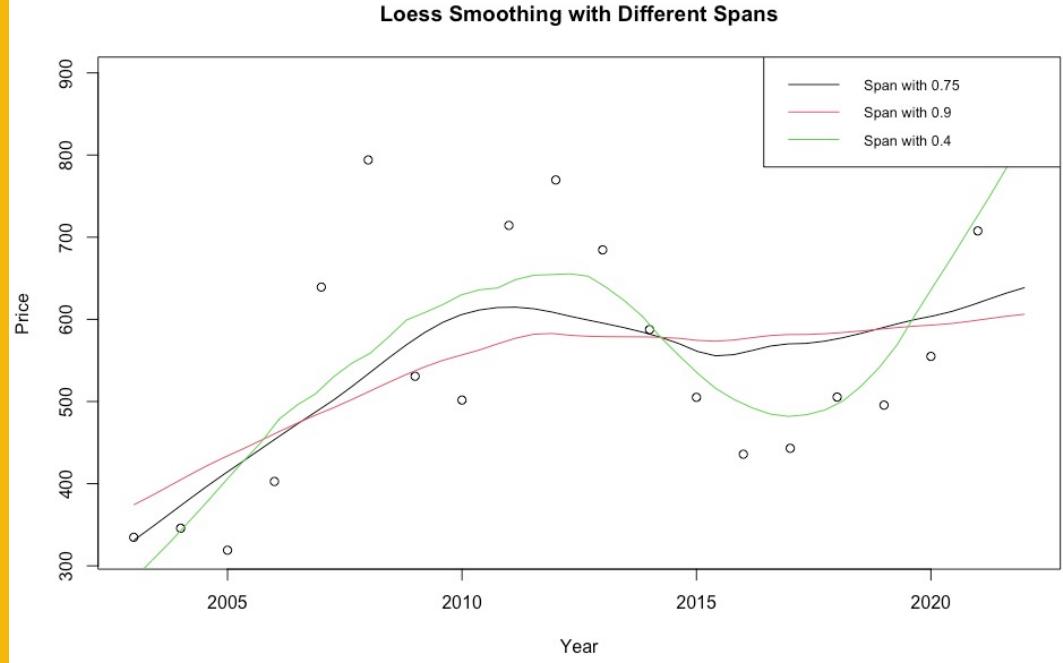
- **MSE:** 4106.724
- **RMSE:** 64.08373

# LOCAL REGRESSION (LOESS)

Higher the span the higher level of smoothing.

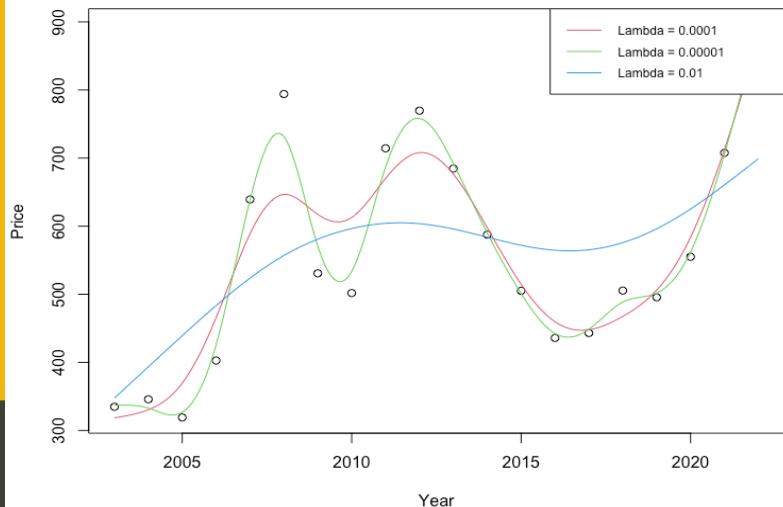
Error Metrics (span=0.4):

- MSE: 1651.43
- RMSE: 40.64
- MAE: 35.41



# REGRESSION SPLINE

Smoothing Splines with Different Lambda Values



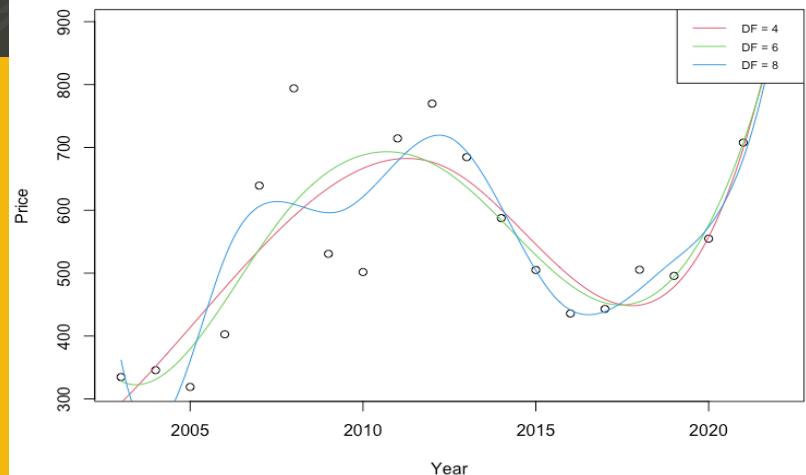
Error Metrics (lambda=0.01):

MSE: 41599.09

RMSE: 203.96

MAE: 123.57

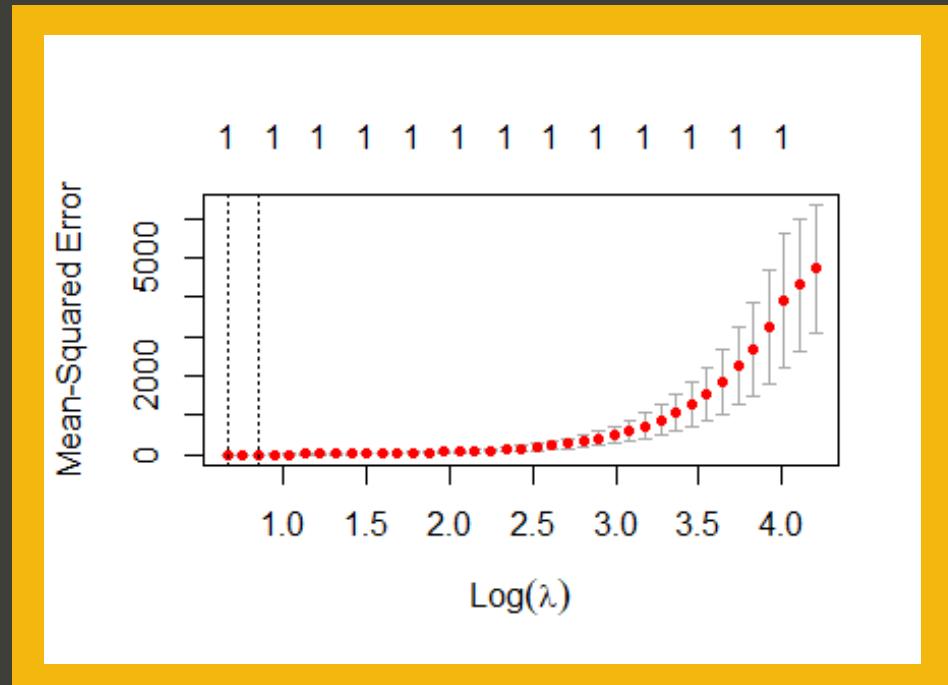
Regression Splines with Different Degrees of Freedom



# LASSO REGRESSION

Prediction with lambda=1.945 (best):

- MSE: 368.238
- RMSE: 19.189



# 4

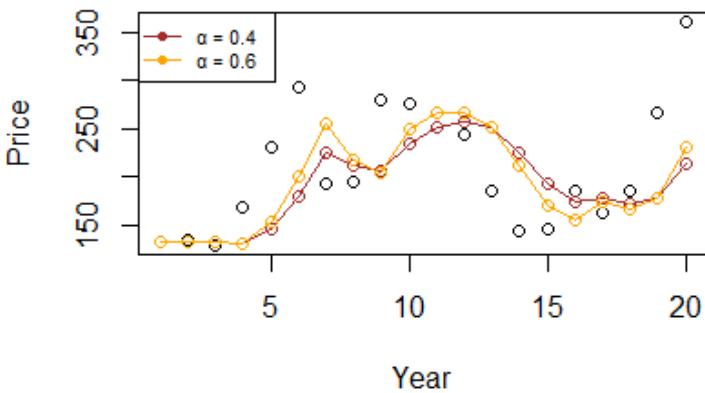
## FORECASTING



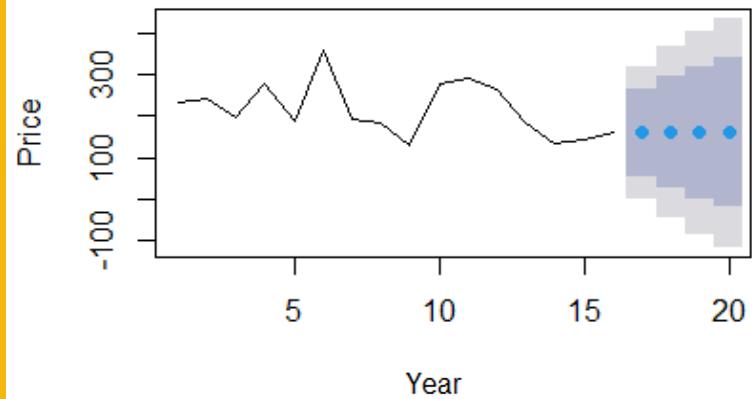
# SIMPLE EXPONENTIAL SMOOTHING



Non-Linear Exponential smoothing



Exponential Smoothing Forecasting



Error Metrics:

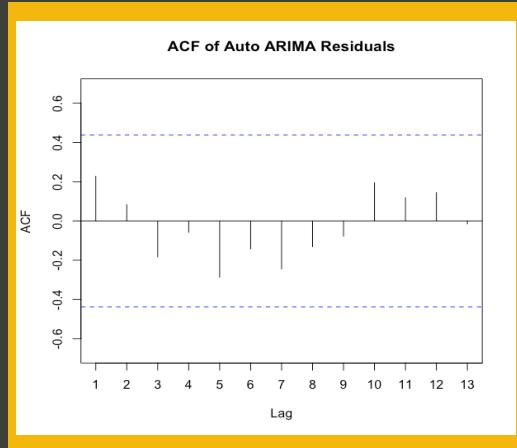
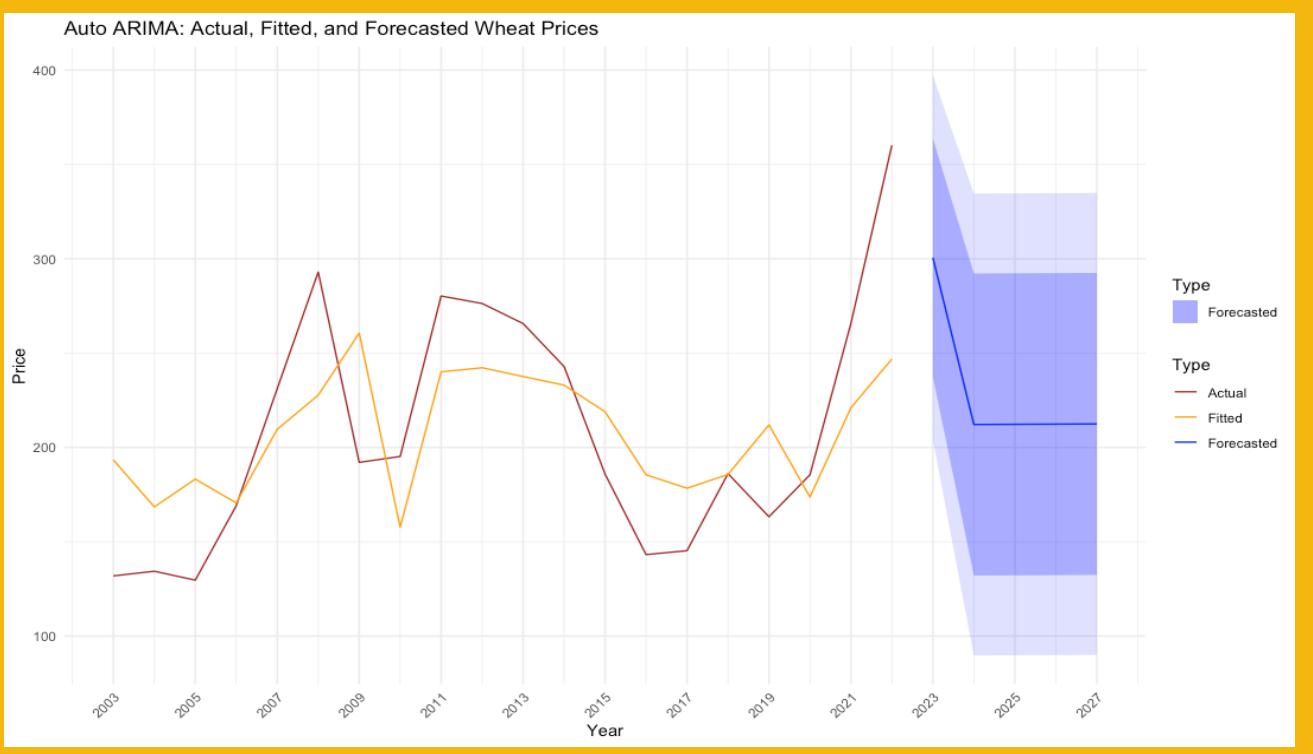
MSE: 3090.756

RMSE: 55.594

AIC: 187.416

$\alpha = 0.6$  shows a better fit → lower error

# FORECASTING WHEAT PRICES WITH ARIMA MODELS



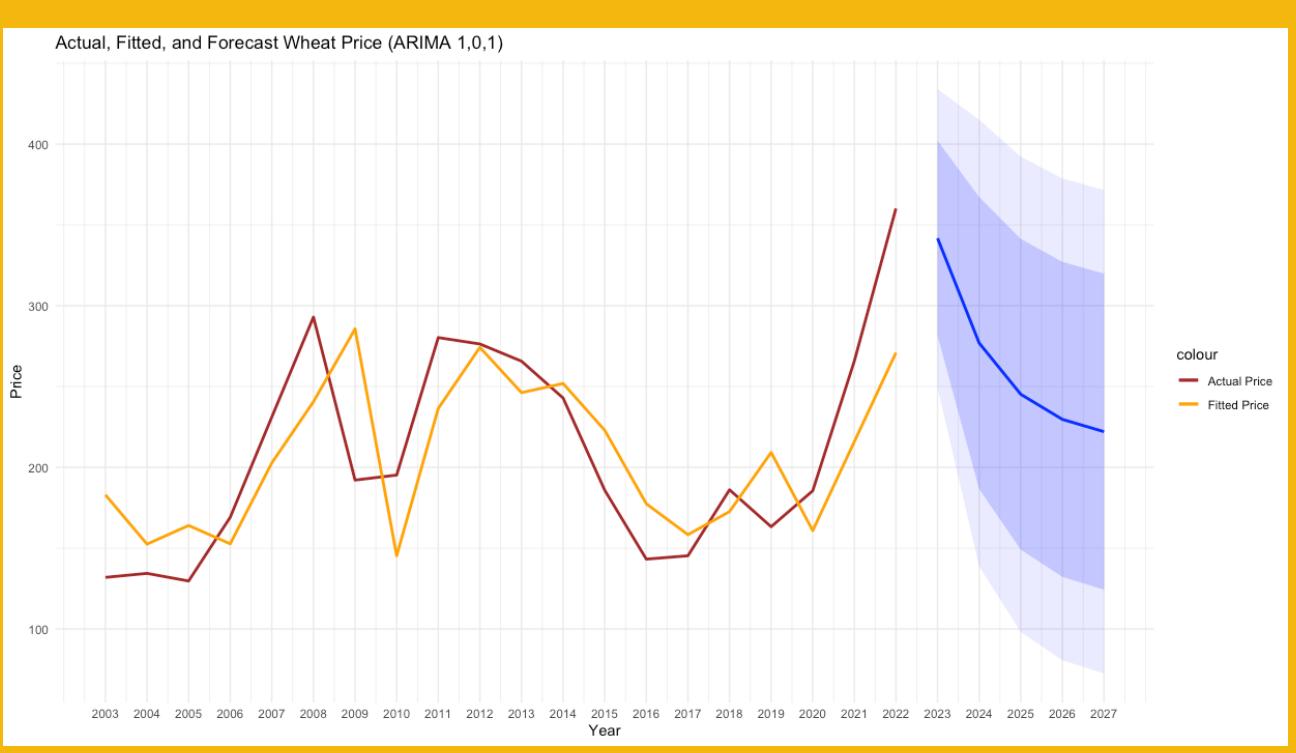
## Auto ARIMA (0,0,1) Model Metrics:

- RMSE: 46.838
- MAPE: 19.903
- AIC: 217.568

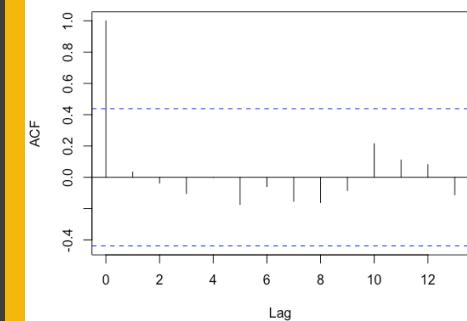
# FORECASTING WHEAT PRICES WITH ARIMA MODELS



Actual, Fitted, and Forecast Wheat Price (ARIMA 1,0,1)



Series residuals\_wheat\_price\_101



## ARIMA(1,0,1) Model Metrics:

- RMSE: 43.381
- MAPE: 18.268%
- AIC: 216.862

# FORECASTING WHEAT PRICES WITH ARMAX MODEL

Actual vs Fitted Prices for ARMAX Model with Confidence Intervals



## ARMAX Model Metrics:

- RMSE: 40.577
- AIC: 216.098
- MAPE: 16.467

# GENERALIZED ADDITIVE MODEL (GAM)

```
Family: gaussian  
Link function: identity  
  
Formula:  
Price ~ s(Year, bs = "cs")
```

```
Parametric coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) 208.852 5.188 40.26 6.22e-13 ***  
---  
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Approximate significance of smooth terms:  
edf Ref.df F p-value  
s(Year) 8.406 9 14.87 4.15e-05 ***  
---  
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
R-sq.(adj) = 0.871 Deviance explained = 92.8%  
GCV = 1016.3 Scale est. = 538.32 n = 20
```

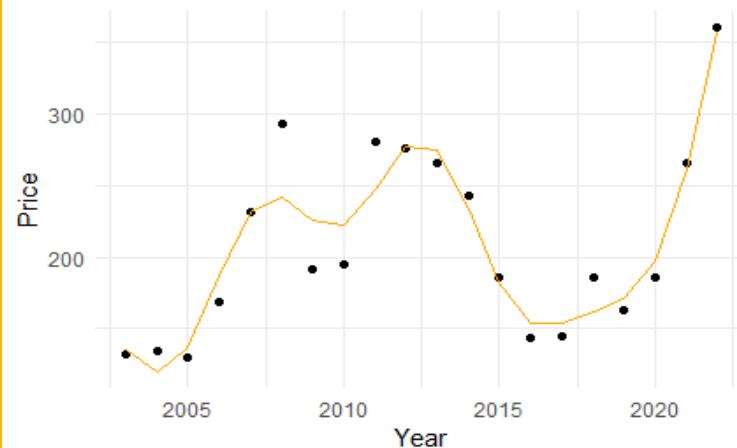
Error Metrics:

RMSE: 19.24

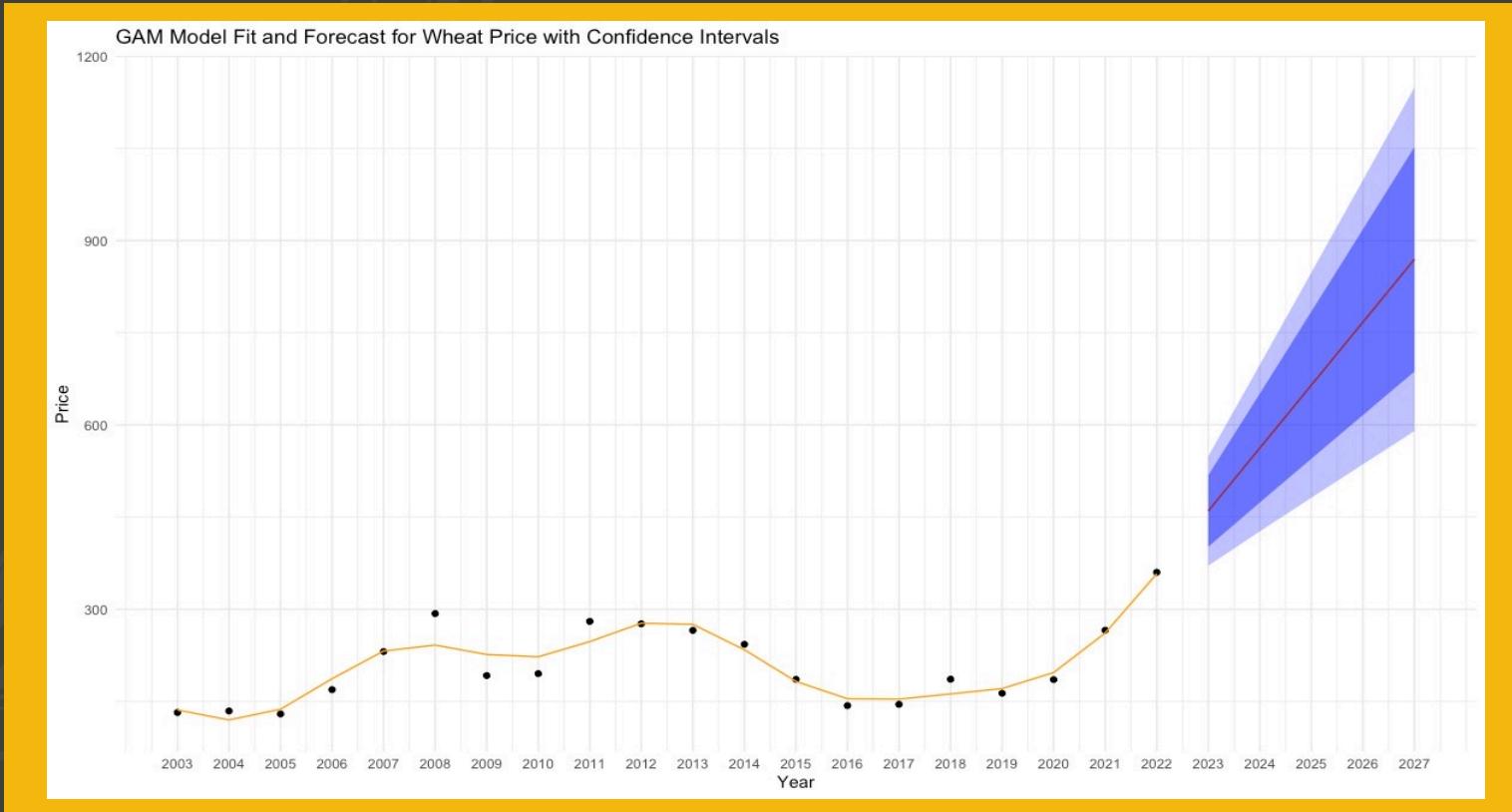
AIC: 195.25

MAPE : 7.045 %

GAM Model Fit for Wheat Price



# GENERALIZED ADDITIVE MODEL (GAM)



# MODEL COMPARISON: FORECASTS



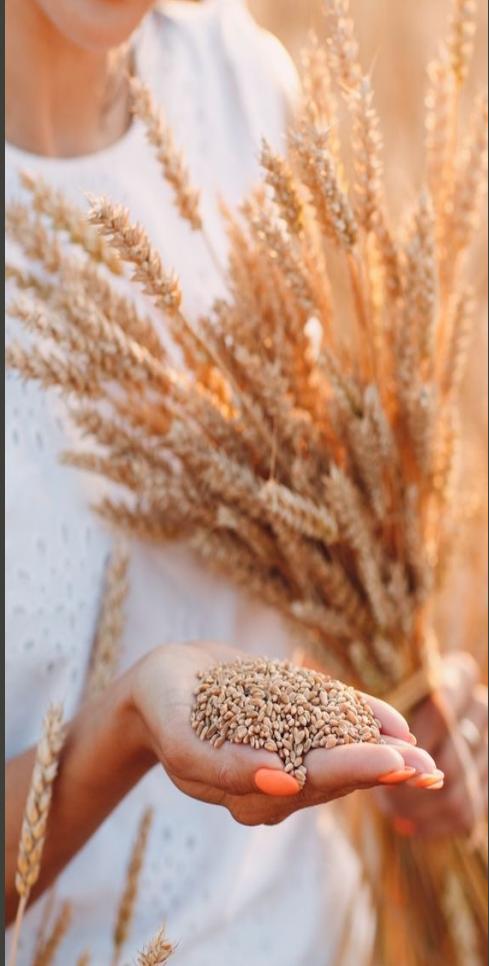
	MSE	RMSE	MAE	MAPE
ARIMA	1881.914	43.381	36.312	18.268
AUTO ARIMA	2193.844	46.838	39.278	19.903
ARMAX	1646.556	40.577	33.166	16.467
GAM	370.151	19.24	14.160	7.045
Exponential smoothing	3090.756	55.594	NA	NA



# CONCLUSION



- Forecasting with GAM has the best performance with an error of 19\$.
  - The best model for prediction was Lasso Regression.
  - Production and total import quantities are the most important factors influencing wheat prices.
- 
- **\$19 per tonne**
  - **\$0.019 per kilogram error with GAM**



# LIMITATIONS & FUTURE WORK



- Future work can develop models like regression trees with more data.
- Include more diverse data such as satellite images for crop assessment.
- Examine the level of poverty for potential food crises.

When pasta prices rise, it's not just the dough you're paying for! It's a **noodle-knotty tale** – wheat prices swaying like spaghetti, production in a pasta mess, and exports taking surprising turns, and occasionally, the '**pasta-bilities**' of geopolitical affairs adding a pinch of unpredictability.





THANK YOU!