Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK

DISSERTATION

MEAT CONSUMPTION PREDICTION: A DATA SCIENCE PERSPECTIVE

DIOGO DE SÃO BRÁS SIMÃO

JULY-2023

Lisbon School of Economics & Management
Universidade de Lisboa

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER´S FINAL WORK
DISSERTATION

MEAT CONSUMPTION PREDICTION: A DATA SCIENCE PERSPECTIVE

DIOGO DE SÃO BRÁS SIMÃO

SUPERVISION:
PROF. CARLOS MANUEL JORGE DA COSTA

JULY-2023

ACKNOWLEDGMENTS

The culmination of an enriching journey over the last few years, always in new areas that have always posed a constant challenge for me.

A special thanks to all the math teachers who shaped and made my path happier, namely Professor Pedro Alves and my great mentor Professor Ângela Luís.

To all the people who stayed and impacted my life in some way: family and friends without exception.

I also thank ISEG and my supervisor not only for the advice, patience, motivation and but for all the sharing over these months. This whole phase was highly enriching and I will take it with me for the rest of my life.

GLOSSARY

AIC - Akaike Information Criterion

ARIMA – Autoregressive Integrated Moving Average

BIC - Bayesian Value of Information Criterion

CRISP-DM - CRoss Industry Standard Process for Data Mining

GDP – Gross Domestic Product.

JEL – Journal of Economic Literature.

Kg - Kilogram

LSTM - Long Short-Term Memory

MAE - Mean Absolute Error

MAPE - Mean Absolute Percentage Error

MLP - Multi-layer perceptron

RMSE - Root Mean Square Error

RNN - Recurrent Neural Network

SARIMA - Seasonal Autoregressive Integrated Moving Average

SARIMAX - Seasonal Autoregressive Integrated Moving Average with Exogenous factors

ABSTRACT

The growth registered in the meat industry has presented some challenges. The price component, while revealing market trends, can be affected by external factors that cause fluctuations in other types of meat. The dissertation focused on an analysis over the period 1990-2022 of the evolution assumed by meat prices and for its different types: poultry, pig, bovine and ovine. A series of causes justified the oscillating behavior of prices since the financial crisis of 2007-2008, the rise in the price of grains (a central element in animal feed) or government policies that restricted trade between countries. Projections were made for 5 years using different models, in order to determine which, one had a better performance for predicting the meat price index. The comparison of the Prophet, SARIMAX and LSTM RNN models showed better overall accuracy in the Prophet. It stood out for its ease of implementation and adjustment of the different components, as well as the adaptation to strong seasonal patterns. When considering projections for the future, possible shocks in supply and demand, extreme weather conditions, exchange rate variability or health concerns must be taken into account.

The complementary analysis on predicting pig meat consumption involved a causal analysis on the statistical inference of several variables and possible impacts on the increase/decrease in pig consumption. Variables such as mean years of schooling, GDP per capita and female labour participation showed a direct and significant relationship in the increase in consumption, as opposed to the religious component, which had an inverse impact. Additionally, approaches from different Machine Learning models (AdaBoost, Random Forest, Multi-Layer Perceptron) complementing the robust linear model (Huber Estimator) were combined to verify which was better in the predictability of pig consumption, with AdaBoost standing out from the others.

KEYWORDS: Meat Price Prediction; Prophet; volatility; Machine Learning models; Meat Consumption

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Markets are economic systems where two major forces interact: demand and supply. In the presence of a balance between them, efficient markets are generated, without the affectation of considerable external factors. The price should reflect the forces surrounding the market. With the growing uncertainty surrounding political, economic, or environmental matters, inefficiencies may emerge in the market (Ben Abdallah, Fekete Farkas and Lakner, 2020). In this sense, the analysis of the price component is particularly important.

The present work focuses on the analysis of the meat price index in general and for its different types, including poultry, pig, bovine and ovine, and consequent analysis in detail of meat consumption. The analysis around prices makes it possible to identify market trends, possible inefficiencies and how a better match between supply and demand can be made. In addition, the predictability of price trends helps in this task. On the other hand, the detail associated with the consumption of meat, in particular pig, motivated above all by a greater incidence of religious/cultural issues (Milford et al., 2019) offers the possibility of identifying the main factors that impact the side of search.

The work will have a structure where it will first focus on analysing the evolution of prices in the period from 1990 to 2022 and identifying the main causes affecting their volatility; followed by a decomposition of the time series to verify its components: trend and seasonality; 5-year projections for the variable under study and assess the main points of uncertainty for the future of the meat sector; comparison between statistical models to assess which of them has a better predictive character for the variable of interest: meat price index, as well as the comparison between different types of meat; a causal analysis with the incorporation of a regression to study which factors have a direct influence on higher pig consumption, followed by a pipeline with machine learning models (AdaBoost, Random Forest and Multi-layer Perceptron), as a way to identify the model that best suits pork consumption.

My contribution to this work will focus on providing good predictive models for the meat price index (poultry, pig, bovine and ovine), presenting the one that shows greater performance and accuracy; 5-year projections, being able to predict trends in the market; present the potential of Prophet for this specific case and reflect the main factors impacting pig consumption with the incorporation of a new variable in the literature: Mean years of schooling.

2. LITERATURE REVIEW

The present study will be based on four major topics: Meat volatility prices; Forecasting models; Regression Models and Machine Learning approaches in predictive analyses. These topics are presented by different researchers who provide theoretical evidence regarding the main investigation problem, meat consumption prediction based on meat price fluctuations.

2.1. MEAT VOLATILITY PRICES

Gilbert and Morgan (2010) studied the main causes associated with food volatility, namely meat, identifying the primary role of grains as animal feed. They can generate increasing pressures on the production side with their rising prices. The consequences and main uncertainties for the future relate to factors linked to shocks in demand or supply; lower demand/supply elasticities and exchange rate variability (the devaluation of the dollar could have a huge impact on rising prices on the producer side).

As analyzed by von Braun et al (2012), the global financial crisis that occurred in 2007-2008 and which extended to the food market generated several political upheavals. The adoption of impulse trade measures limiting exports and imports and increasing subsidies predicted effects that could have an extremely negative impact in the long run. The relief generated in the short term had some effects on trade between countries, leading to major market players: China and India, having to depend on their domestic sources.

Over time, growing health concerns have been raised and have been felt in the price of meat. The example of covid-19 raised concerns in the sector, emphasizing food safety and the repercussions it may have on health. Price declines have been felt since 2022, mirroring malfunctions and limitations in supply chains (Guo and Tanaka, 2020).

There may be discrepancies between the price paid by consumers and the price at which meat is quoted on the market, and in addition to changes that occur on the supply or demand side, the availability of raw materials, production and distribution costs, taxes and profit margins of intermediary companies are examples that support this difference (Hobbs, 2021; Akbar et al., 2021).

2.2. FORECASTING MODELS

In forecasting time series, both Prophet, Arima and LSTM (Long Short-Term Memory) RNN (Recurrent Neural Network) are widely used, statistical models. The choice of model to be used will depend on the characteristics of the time series in question and the forecast objectives.

The Prophet model is a non-parametric approach, which was developed by Facebook in 2017 and corresponds to an extension of the classic decomposition model, with the incorporation of additional components such as seasonality, trend and holidays. On the other hand, the ARIMA model (Autoregressive Integrated Moving Average) is a parametric model widely used to model time series (Zhao et al., 2022). It has some assumptions such as the linearity of time series and the following of normal distributions. The model has a standard ARIMA(p,d,q) notation, where p corresponds to the autocorrelation component of the time series (uses previous values to predict future values); d represents the differentiation component of the time series, where in cases of non-stationarity it is necessary to differentiate the series to ensure adequate modelling; q corresponds to the moving average component of the time series (uses a moving average of past values to project future values) (Chakraborty, Corici and Magedanz, 2020).

Studies developed by Duarte and Faerman (2019) show that the use of the Prophet model in the presence of seasonality presents significant improvements compared to the

ARIMA model. The practical case provided by the authors shows improvements in the order of 69% of the occurrences, providing greater accuracy of the model for the variables under study.

The representative function of the Prophet model is the following (Taylor and Letham, 2018):

$$y(t) = g(t) + s(t) + h(t) + e_t \tag{1}$$

where:

- y(t) corresponds to the value observed in the time series in period t;
- g(t) corresponds to the general trend of the time series, which can be assumed to be linear or non-linear;
- s(t) comprises the seasonality component of the model, reflecting repetitive variations in time periods (e.g., weekly, monthly or yearly). Depending on how seasonality affects the time series, it can be modelled as additive or multiplicative;
- h(t) is the function that represents the influence of holidays and special events;
- e(t) represents the random error of the model which controls for uncertainties in the predictions and any unforeseen changes in the model. Assumes a normal distribution given the assumption of normality required in statistical models to model data uncertainty.

In the presence of data that show seasonal patterns, ARIMA can assume an extension in SARIMA or SARIMAX, contributing to more accurate forecasts and better modelling performance. The difference between both lies in the incorporation of exogenous variables to the time series, which in the case of the meat price index, factors such as inflation, price of substitute foods or the availability of animals for slaughter can impact the variable under study. The SARIMAX model (Seasonal Autoregressive Integrated Moving Average with exogenous factors) has the following notation (p, d, q) * (P, D, Q)s, which, in addition to the previous notation of the ARIMA model, incorporates the seasonal component (Fathi et al., 2019; Manigandan et al., 2021).

The auto_arima() function of the pmdarima package in python allows to automate of the process of selection and best fit of parameters for an ARIMA model based on information such as the Akaike Information Criterion (AIC) and the Bayesian Value of Information Criterion (BIC) (Alavi et al., 2020; Awan and Aslam, 2020).

LSTM (Long Short-Term Memory) is a type of recurrent neural network that was developed to solve the problem of gradient fading. Its architecture contains memory cells that allow the storage of relevant information for long periods of time. In this structure there are three ports (input port, forgetting port and output port), which in the first phase, in the input port there is the management of which relevant information to transmit to the memory cell, later determining the information to be transmitted. preserve and those that will be discarded at the door of oblivion. The output port is where the relevant stored information is used to calculate the current output. Derived from memory cells being endowed with these complex structures, they enable the model to capture long-term dependencies in time series (Li et al., 2021).

The Prophet model stands out from other time series models for some reasons (Li et al., 2021):

o Simplicity of use and processing speed, which in the presence of a large volume of data has special preponderance;

o Incorporation of seasonal factors and holiday effects, in some cases complex seasonality, with the easy adjustment of parameters depending on the different assumptions and objectives of the analysis;

o Less sensitivity to outliers and missing data, enabling more accurate forecasts even in time series with extreme or unusual values.

2.3. REGRESSION MODELS

In the case of traditional regressions, they follow some assumptions such as linearity, independence, normality and homoscedasticity, which when some of them are not met, can generate inaccurate or even invalid results in the regression. In these cases, more appropriate analysis can be performed using robust regressions. However, although they are less sensitive to violations of the assumptions of traditional regressions, they require that some of these assumptions are met, namely, linearity; independence; the absence of multicollinearity between the independent variables and the absence of outliers (Yaffee, 2002; Filzmoser and Nordhausen, 2020).

It should be noted that although robust regression presents itself as a solution to the violation of traditional assumptions, it generally has a lower statistical power at the level of the relationship between the independent variables and the dependent variable (Yaffee, 2002).

According to Macêdo (2014), the Huber M-Estimator statistical technique is adequate when the heteroscedasticity of the data and the presence of outliers is not extreme.

2.4. MACHINE LEARNING APPROACHES IN PREDICTIVE ANALYSES

Regarding the predictive analysis of a certain variable, several Machine Learning models can be selected, each of them employing a different approach, such as:

o AdaBoost: boosting algorithm commonly used in classification tasks, but which can also be adapted for regression problems. The goal is to implement a consistent model that appropriately generalizes the data, thus minimizing the general training error. The algorithm is executed through several iterations, where several models are conjugated according to their performance during training, thus adjusting their sample weight (Cao and Li, 2022);

o Random Forest: Corresponds to a machine learning algorithm widely used both in classification tasks and in regression problems. One of its advantages is the ability to reduce data overfitting, which happens when the model is perfectly adjusted to the training data and has difficulty predicting future data. The model combines several decision trees that are trained on random subsets of data and features, enabling greater model robustness and more accurately projecting future data. Another advantage is the high dimension of the data, allowing sets with many characteristics. It adapts to missing data and outliers and identifies which variables have the most impact on the regression model (Biau and Scornet, 2016).

o Multi-layer Perceptron (MLP): This is an artificial neural network composed of multiple layers of neurons. The MLP has interconnected layers, which interact with each other and learn the relationship between data set inputs and outputs, and projections for new future scenarios. Depending on the network configuration, it is possible to learn linear or non-linear relationships between input and output variables (Masci et al., 2011; Guarino et al., 2022).

3. METHODOLOGY

In this section, the models and adjustments made to the data will be presented, where the CRISP-DM methodology was used.

The methodology enables large data mining projects to be more manageable, more reliable and with greater operational efficiency. In the initial phase, the project's central objectives are defined, data collection, and checking the first trends and behaviors to arrive at a set of final data for analysis and modelling. Subsequently, this step provides a broad spectrum of the models' quality and specifies which is most appropriate to respond to the initial problem. The lifecycle of this methodology is shown in the graph A-10 (Wirth and Hipp, 2000).

3.1. MEAT PRICE INDEX (DATA AND APPROACH)

In order to identify some of the main factors associated with the meat industry, including its consumption and production levels, global data on the general meat price index were used, subdivided into poultry, pig, bovine and ovine. The data used are monthly and cover the period from 1990 to the end of 2022. The Meat Price Index is calculated from the average prices of four types of meat and weighted by the averages of the world export trade shares from 2014 to 2016. The index includes two poultry products, three bovine products, three pig products and two ovine products. In situations where there are several quotations for the same type of meat, they are weighted by fixed shares assumed in the market. Meat prices are commonly measured in dollars per ton, which is a unit of measurement used for internationally traded commodities (FAO, 2022).

Monthly meat price index data are from FAOSTAT (Markets and Trade: Commodity Data, FAO, 2022).

The data format is adapted to the needs with numerical values for the various meat price indices and for the time period from 1990 to 2022 in date time form. When using the Prophet model, there has to be an adaptation of the data frame by renaming the columns 'data' and the various 'meat price index' in ds and y and reindexing sequentially.

Thus, using the Prophet statistical model, a time series forecast was performed, including an exploratory analysis of the evolution of the variable during the period considered, its trend and seasonality components, projections for the next 5 years and a performance comparison between the Prophet models, SARIMAX and LSTM RNN.

3.2. PIG MEAT CONSUMPTION (DATA AND APPROACH)

Meat consumption is one of the driving forces in the sector. As such, and complementing the study of the price of meat, it is seen in some detail which factors can lead to greater consumption of meat and which is the best predictive model for the variable under study. The analysis will focus on the year 2020, an atypical year marked by the covid-19, which boosted meat prices (FAO, 2022). The study was also centered on one of the types of meat that have suffered the most external influences over time: pig meat.

A study of the impact of some variables on pig consumption will be incorporated. Studies carried out by several authors have shown some relevance in variables associated with economic factors, such as GDP per capita (Milford et al., 2019; Whitton et al., 2021), social factors such as female labor participation (Manrique and Jensen 1997) and urbanization rates in a given country (York and Gossard, 2004) and religious/cultural factors (Milford et al., 2019).

The transition that was felt in the labor market with greater participation of women has a potential influence on the composition of the family diet, either through faster-prepared foods, greater convenience or greater awareness of the universe of meat consumption and their surrogates (Connor, 1994; Manrique & Jensen, 1997; Schmidhuber & Shetty, 2005).

Additionally, as a result of covid-19, the education system was affected, as reported by Moscoviz and Evans (2022). The average years of schooling will be analyzed, included in the regression and verified if they have any impact on the variable under study.

The data that make up the dataset are aggregated by country and come from different entities:

- From Pew Research Center for variables associated with the percentage of a certain religion (Buddhists, Jews, Christians, Muslims, Hindus);
- From the International Labor Organization for the variable female labor force (%);
- From the United Nations for urbanization indices (%);

o From FAOSTAT for values concerning pig consumption per capita (kg/capita), GDP per capita (dollars at constant 2015 prices) and population (in millions);

o From Global Data Lab Area Database (includes data from statistical offices and Eurostat) for the mean years of schooling variable.

It should be noted that mean years of schooling focus on an adult population over 25 years old, reflecting the years of schooling completed. This variable forced a projection of the values for 2019 for the year 2020, assuming two scenarios: maintaining the levels of 2019, and in most cases, a decrease of 1% compared to the values recorded in 2019.

The female labor force variable corresponds to the proportion of the female population aged 15 or over who are considered economically active.

Urbanization indices refer to the population living in areas called urban. The data via projections from the United Nations in 2014 show some inaccuracies in some countries, in which cases urbanization rates equal to 100% were adopted.

In the case of religions, the adjustments were implemented only in two situations:

- o <1 (considered as zero);
- o >99 (considered as 100%).

After the adjustments made to the data, using the 'merge' function in python (based on the column: 'Country') it was possible to standardize the dataset and prepare the predicted analyses. Just to highlight the fact that, as the data come from different entities, small adjustments were made in the name of the countries for a similar format in all.

Once the solidity of the final data set was assured, in the modelling stage, the statistical inference of the variables was initially evaluated as a way to understand their significance at the 5% level in pig meat consumption (using robust regression), followed by the complementary analysis, incorporated with a Machine Learning pipeline in to identify the best predictive model for the variable of interest.

In the GitHub repository (https://github.com/Diogodsbs/Thesis2023), I have all the materials used and python codes incorporated in my analysis in a more detailed way and for various scenarios.

4. RESULTS

This is the section dedicated to the analysis and discussion in detail of the different topics.

4.1. MEAT PRICE INDEX (POULTRY, PIG, BOVINE AND OVINE) EVOLUTION BETWEEN 1990-2022

The meat price index measures changes in the prices of different types of meat: bovine, ovine, poultry and pig.

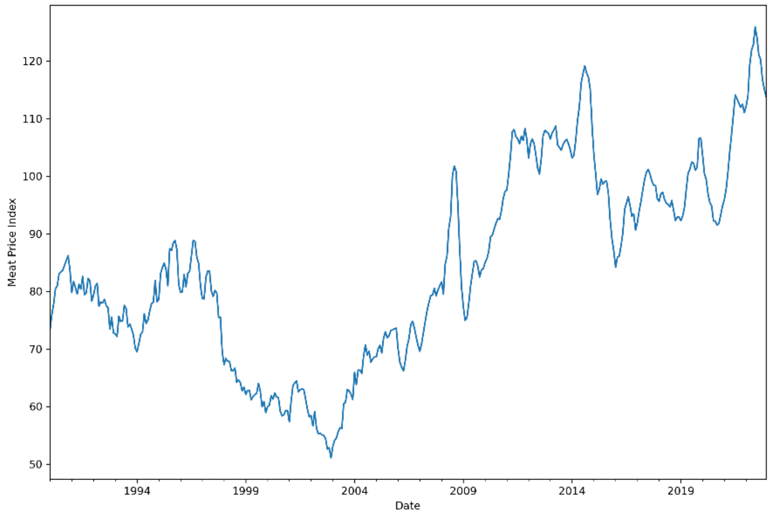
During the period from 1990 to 2022, it has shown some considerable fluctuations, with a clear trend division, that is, from 1990 to 2003 it showed a general downward trend in prices, from values between 80/90 to the 50 recorded in 2003. This decline was mainly due to imbalances in the market, between supply and demand. There was rapid economic growth in the Asian continent, particularly in China, which was responsible for 57% of this increase in the meat supply side for developing countries. In clear contrast are the developed countries that have shown some stagnation in the market (FAO, 2005). The supply side has mainly felt lower productivity margins, leading to lower stocks, some ecological and environmental restrictions, as well as more expensive fuels and fertilizers. The demand side saw a slowdown in consumption in some countries due to health concerns leading to changes in food preferences (Woertz, 2013; Harrigan, 2014).

Since 2004, there has been a reversal of the trend, moving towards a clear upward movement culminating in the current value of 115. During this period of increase, between the period from 2004 to 2022, there were several notable events, including:

- The years 2007-2008 are reflected by the financial and global crisis affecting the food market, leading to a general rise in meat prices. The price of inputs is considered a central element, serving as a food ration, and such an increase has led to higher production costs on the part of producers, providing climates of high volatility and instability in the market (Tejeda and Goodwin, 2009; Guo and Tanaka, 2020);

- The years 2014-2016 exposed the slowdown in demand from emerging countries and major oil exporters, leading to stagnation in world trade. The weakening of supply, with considerable declines in meat exports, mainly from Brazil. Some trade restriction policies namely reduced imports to the Russian Federation, as well as a loss of trade in North America, led to this decrease in the meat price index (OECD/FAO, 2016);
- since 2020 prices have risen to historic highs in June 2022, as a result of restricted supply and a slowdown in the economy with global import demands in smaller countries. Factors such as the rise in prices of grains for animal feed, the emergence of infectious diseases in animals and some stoppage in the functioning of the meat industry value chains due to Covid-19 boosted such price increases. The recorded meteorological phenomena also affected the production side (Guo and Tanaka, 2020; FAO, 2022).

Graph 1: Evolution of Meat Price Index between 1990-2022 [Prophet Output]



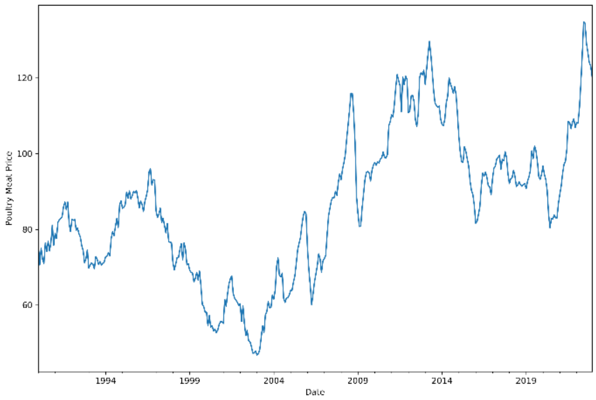
Source: Own elaboration based on FAO (2022)

The particular analysis of each type of meat reflects the general fundamentals associated with fluctuations in the meat price index, reflecting certain characteristics and trends over time that each type of meat assumes.

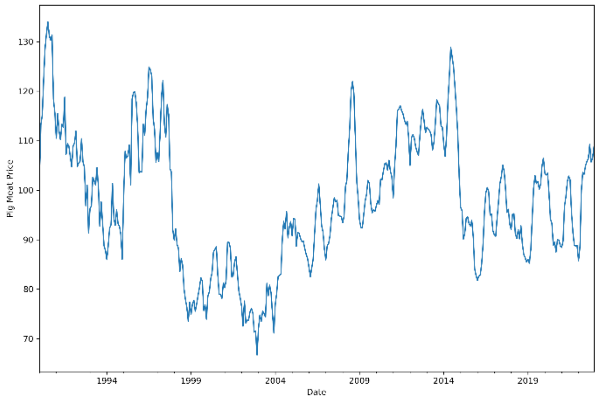
Poultry meat assumes a behaviour similar to the general meat price index, with emphasis on the aforementioned periods. Of note is the abrupt rise that took place in the years 2006-2008 from values around 60 to 120, followed by a sharp decline in the years 2008-2009 from 120 to 80. The period from 2015 to 2020 reflected some slowdown in the trend ascending (contrary to the meat price index), to subsequently resume an upward trend since the beginning of the pandemic, reaching maximum values similar to those practised in 2012 (between 120/125).

Pig prices show huge fluctuations over time, following a clear downward trend. The beginning of the 90s was the one that showed maximum values of around 130, followed by downward cycles where the market tried to counter such declines. After the decline between the period from 1990 to 1994, the market acted upwards reaching values of 120. Subsequent behaviour followed the trends of the meat price index until 2016. It should be noted that the minimum value reached for pig meat stands at 70 as opposed to the 50 recorded by the meat price index. Since 2016, the volatile character that the price index has continued to show has revealed a null trend in values around 100. All types of meat have shown price increases since covid-19, except for pork. In short, it is a type of meat that has been losing importance in the market where there are changes in consumer food preferences, environmental and health reasons, and equally enormous pressure from government policies to restrict trade between some countries.

Graph 2: Evolution of Poultry (1) and Pig (2) Meat Price Index between 1990-2022 [Prophet Output]



(1)

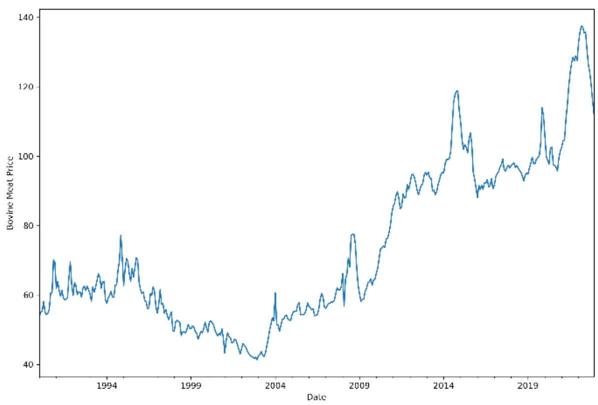


(2)

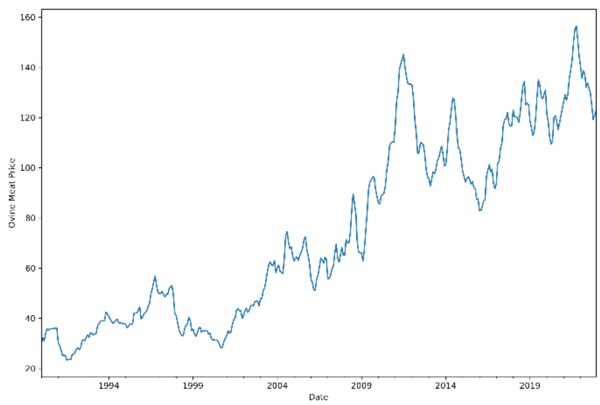
Source: Own elaboration based on FAO (2022)

The bovine and ovine meat price indexes behave similarly throughout the period. Throughout the 1990s, they did not show any fluctuations in the name, being somewhat out of line with other types of meat. In the case of bovine meat, prices varied between 80 and 40 and in ovine meat prices were between 60 and 30. Prices were considerably lower in this period compared to the rest. The boom in prices occurred at the beginning of the year 2000 projecting continuous upward trends without major slowdowns. It should be noted that in 2017 the ovine meat price index distanced itself from other types of meat up to its historical maximum recorded in 2021 (values around 160). However, since 2022 it has registered some slowdown, with a drop to values around 120, at the levels of bovine and poultry meats.

Graph 3: Evolution of Bovine (3) and Ovine (4) Meat Price Index between 1990-2022 [Prophet Output]



(3)



(4)

Source: Own elaboration based on FAO (2022)

4.2. STATISTICAL COMPARISON BETWEEN PROPHET, SARIMAX AND LSTM RNN FOR PREDICTING MEAT PRICE INDEX

Table 1: Statistical comparison between Prophet, SARIMAX and LSTM RNN for predicting meat price index.

Performance Metric	**Prophet**	**SARIMAX** (2,1,2) (1,0,[1],12)	**LSTM RNN**
RMSE (Root Mean Square Error)	6.39	19.38	7.21
MAE (Mean Absolute Error)	5.22	17.83	5.83

To compare simulated and observed values, RMSE and MAE were used to test performance between models. This study uses Prophet, SARIMAX and LSTM RNN to compare their performance and accuracy on the meat price index variable dataset. In this sense, the dataset was divided into a training set from 1990 to 2020 and a test set from 2021 to 2022, as a way to generalize the new data, avoiding overfitting.

Table 1 reflects which model performs better, illustrating that Prophet, having lower values for both RMSE and MAE, is the most appropriate choice in predicting the meat price index variable.

4.3. PROPHET MODELS AMONG DIFFERENT TYPES OF MEAT

Once Prophet was identified as the best predictive model for the variable under study, 5 models were incorporated that follow time series for the following variables: meat price index, poultry meat price index, pig meat price index, bovine meat price index and ovine meat price index. Subsequently, the training-test split was performed to evaluate the performance of each of the models, encompassing a training set from 1990 to 2020 and a test set from 2021 to 2022.

The performance measures presented in Table 2 concern the minimization of precision errors in the case of MAE and RMSE, in the evaluation of precision in percentage terms, in the case of MAPE and R^2, which verifies the proportion of data that is explained by the models on the various variables of interest.

Models with lower values of MAE, RMSE and MAPE and higher values of R^2 are considered better models in terms of performance. In the example, the meat price index is the one that presents the best result, given that it has a lower mean and mean squared error (MAE and RMSE, respectively) and a lower mean percentage error (MAPE). The MAPE has a value of 0.06, which indicates that, on average, the model's predictions presented an error of about 6% in relation to the values observed in the meat index price.

At the level of data adjustment in the different models, the bovine meat price index presents the best value (0.91), indicating that 91% of the variation in the bovine meat price index is explained in this proportion by the model. It should be noted that the pig meat price index has the lowest value (0.55), showing that this type of meat is easily more volatile and influenced by exogenous factors.

Table 2: Statistical comparison between Prophet models (Meat Price Index, Poultry Meat Price Index, Pig Meat Price Index, Bovine Meat Price Index and Ovine Meat Price Index)

Performance Metric	**Prophet** (y=Meat Price Index)	**Prophet** (y=Poultry Meat Price Index)	**Prophet** (y=Pig Meat Price Index)	**Prophet** (y=Bovine Meat Price Index)	**Prophet** (y=Ovine Meat Price Index)
RMSE (Root Mean Square Error)	6.39	9.51	9.19	6.96	11.54
MAE (Mean Absolute Error)	5.22	7.68	7.34	5.29	8.35
MAPE (Mean Absolute Percentage Error)	0.06	0.09	0.08	0.07	0.13
$\mathbf{R^2}$ (R-squared)	0.86	0.76	0.55	0.91	0.90

4.4. PROPHET FORECASTING

In this section, 5-year projections are made up to the year 2027 for the different variables: meat price index (poultry, pig, bovine and ovine). A 95% confidence interval was used, in which the real value of the different variables is expected to be within the range estimated by the model.

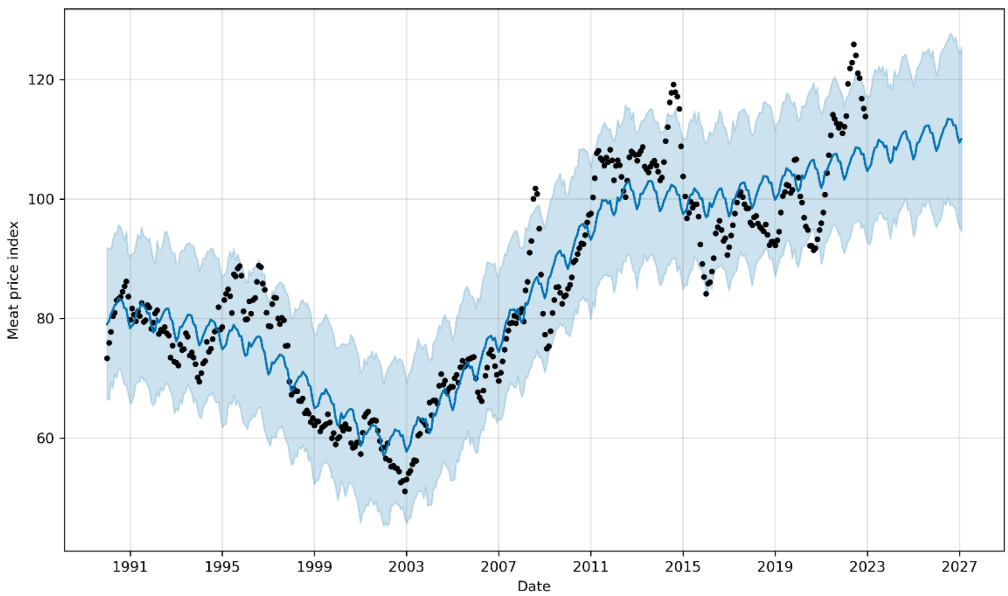
The previously analyzed performance metrics provide us with solid bases for future projections, taking into account the quality of the models, however, it is necessary to take into account external factors that may occur in the market, as is the case, for example, of extreme weather events, which would impact meat prices (Whitton et al., 2021).

4.4.1. MEAT PRICE INDEX FORECASTING

The 5-year future projections indicate a continued trend towards increasing values. As graph 4 demonstrates, it is expected to reach values around 110, subjugated in the 95% confidence interval (minimum expected values between 100 and maximum expected to reach 120). External factors can push the meat price index to values outside the confidence interval, but with a very low probability that will be attenuated by the trend of the graph, offering short/medium-term reliability.

In seasonal terms, as shown in Graph A-1, it is reflected that the first two months of a given year (mainly February) are those that register the greatest declines in the index, however, cancelled out by the increases observed in the months of March.

Graph 4: Projections of Meat Price Index between 2023-2027 [Prophet Output]



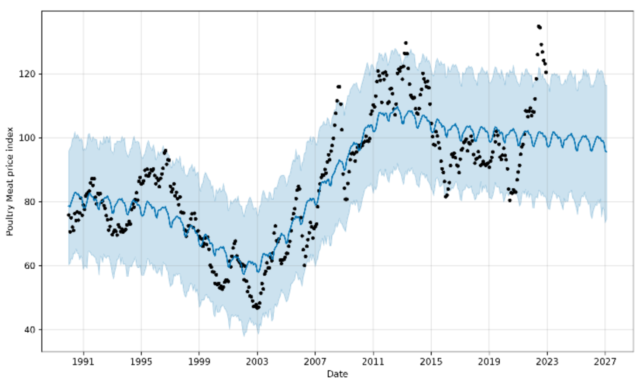
Source: Own elaboration based on FAO (2022)

4.4.2. POULTRY AND PIG MEAT PRICE INDEX FORECASTING

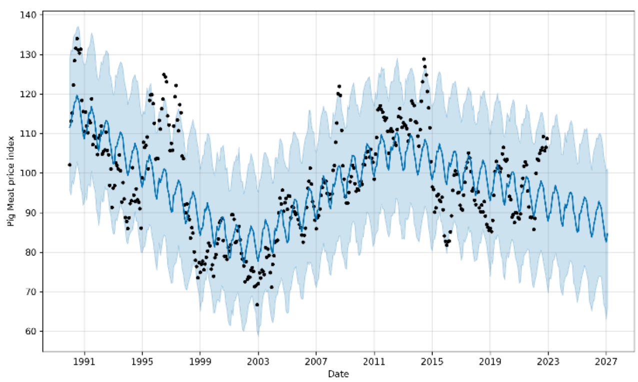
Graph 5 reflects the projections associated with the poultry meat price index, which points to some neutrality in the market, stabilizing prices around 100 in 2027, and in the case of the pig meat price index, the general downward trend in pig meat is mirrored that has been felt in recent years, where this downward cycle is expected to be accompanied by values around 85/90.

With regard to the seasonal component of these variables, the poultry meat price index demonstrates volatile behavior throughout all months, with the first two months of the year being the ones with the greatest decrease, followed by the months with the greatest impact in terms of increase of the prices that are the case of the month of March and mainly April. With regard to the pig meat price index, it shows a monthly behavior that is minimally uniform, despite the fact that the month of February is considered an atypical month, with uneven behavior, accentuating the biggest decreases with some scale. In March, there will be some recovery to rising values of the pig price index (graph A-2).

Graph 5: Projections of Poultry (5) and Pig (6) Meat Price Index between 2023-2027 [Prophet Output]



(5)



(6)

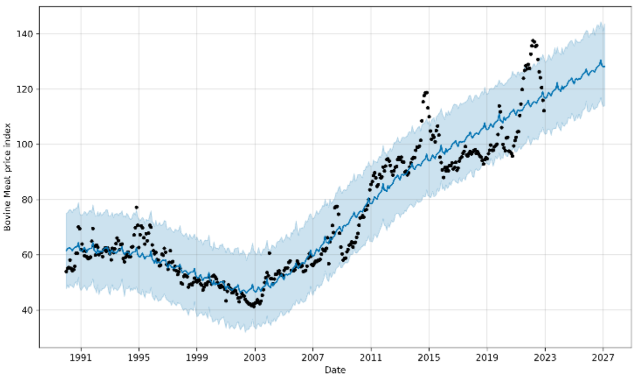
Source: Own elaboration based on FAO (2022)

4.4.3. BOVINE AND OVINE MEAT PRICE INDEX FORECASTING

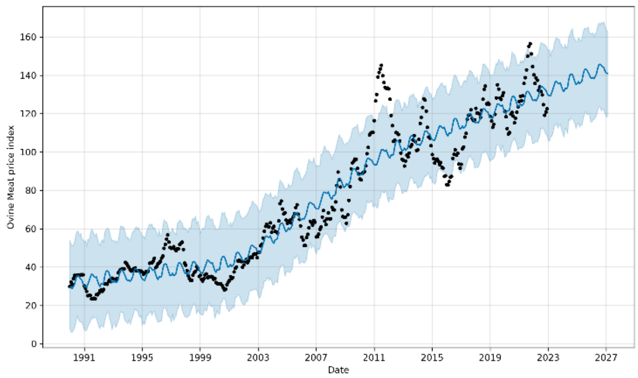
The bovine and ovine price index has a similar pattern even at the 5-year projection level, following the continued growth trends of recent years, moving towards values around 125 for the bovine price index and 140 concerning ovine meat (Graph 6).

In seasonal terms, in addition to the trends already mentioned, the bovine price index increased in March, in contrast to January, when the index recorded the greatest declines. On the other hand, the ovine meat price index records a sharper rise in October and falls in January and December (Graph A-3).

Graph 6: Projections of Bovine (7) and Ovine (8) Meat Price Index between 2023-2027 [Prophet Output]



(7)



(8)

Source: Own elaboration based on FAO (2022)

4.5. UNCERTAINTIES ON THE FUTURE

The projections made must consider the uncertainties that the future entails from:

o Trade policies, especially in relation to export and import (imposition of tariffs and trade barriers);

o Environmental safety and the consequent search for more sustainable options;

o Health concerns;

o Extreme weather conditions;

o Shocks in demand and supply (production costs - grains; speculation; population growth; income; productivity growth rate);

o Exchange rate variability (in case of a devalued dollar)

4.6. ROBUST LINEAR MODEL USING HUBER ESTIMATOR'S

In this specific case, there was some heteroscedasticity of the data and the non-existence of extreme outliers (graphs A-5 to A-9 and table A-2).

The R^2 is a useful measure of how well the model fits the data. The regression presents an R^2 of around 0.56, representing that 56% of the dependent variable (pig consumption) can be explained by the independent variables. It is considered a moderately strong relationship, with factors external to the model that may explain the variable under study.

As a way of evaluating the general quality of the model, the analysis of the statistical significance of the regression coefficients is carried out in parallel. Table 3 shows that all variables with the exception of % urbanization are statistically significant at a significance level of 5% (pvalues<0.05), with the variables GDP per capita, mean years of schooling and female labor participation having a positive association with consumption of pig. On the other hand, the composition by religion in a given country assumes an inverse relationship. In absolute terms, the coefficient of average years of schooling has the most impact since an increase in one year in average schooling would impact meat consumption by 2 kg per capita in the year.

Table 3: Robust linear model using Huber Estimator's

```
                    Robust linear Model Regression Results
==============================================================================
Dep. Variable:     Pigmeat Consumption   No. Observations:                  156
Model:                             RLM   Df Residuals:                      147
Method:                           IRLS   Df Model:                            8
Norm:                           HuberT
Scale Est.:                        mad
Cov Type:                           H1
Date:                 Thu, 09 Mar 2023
Time:                         14:13:42
No. Iterations:                     18
==========================================================================================================
                                           coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------
Mean years of schooling                  2.0234      0.288      7.015      0.000       1.458       2.589
% of Buddhists                          -0.1322      0.057     -2.309      0.021      -0.244      -0.020
% of Christians                         -0.1131      0.035     -3.260      0.001      -0.181      -0.045
% of Hindus                             -0.2378      0.066     -3.629      0.000      -0.366      -0.109
% of Jews                               -0.4131      0.118     -3.495      0.000      -0.645      -0.181
% of Muslims                            -0.1824      0.029     -6.311      0.000      -0.239      -0.126
Female Labour force participation rate (%)  0.0870   0.042      2.087      0.037       0.005       0.169
GDP_per_capita                           0.0001   4.81e-05      2.357      0.018    1.91e-05       0.000
% urbanization                           0.0045      0.037      0.120      0.904      -0.069       0.078
==========================================================================================================

If the model instance has been used for another fit with different fit parameters, then the fit options might not be the correc
t ones anymore .
R² for the Huber estimator's model: 0.5649288728213724
```

4.7. STATISTICAL COMPARISON BETWEEN ADABOOST, RANDOM FOREST, MULTI-LAYER PERCEPTRON AND HUBER ESTIMATORS FOR PREDICTING PIG MEAT CONSUMPTION

After the inference analysis of the variables carried out using the Huber estimator method, a pipeline was created with several Machine Learning models, such as AdaBoost, Random Forest and Multi-Layer Perceptron (MLP). The pipeline encompasses a series of steps, from data pre-processing, a division into training and test sets and consequent modelling in the training phase and evaluation of the test set. This approach allows, in the regression task, to compare which of the models has a better predictive performance for the variable under study: consumption of pig meat.

In this sense, the complementarity of the two analyses is important, as the variables with the greatest influence on the dependent variable are identified and the model with the greatest predictive capacity is also identified.

Table 4 shows which of the models presents better predictive performance. AdaBoost, by virtue of demonstrating a lower value to minimize the mean squared error (9.04) and a higher R^2 (0.66) is considered the best model. It should be noted that the R^2 in the case of the Huber estimator decreased compared to the previous example from 0.56 to 0.48, showing that the full model generalizes the data better than the trained model (Note: the training set for all models was 20%).

Table 4: Statistical comparison between AdaBoost, Random Forest, Multi-Layer Perceptron and Huber Estimators for predicting pig meat consumption.

Performance Metric	**AdaBoost**	**Random Forest**	**Multi-Layer Perceptron**	**Huber Estimator's**
RMSE (Root Mean Square Error)	9.04	9.55	9.80	11.16
MAE (Mean Absolute Error)	7.29	6.77	6.57	7.69
$\mathbf{R^2}$ (R-squared)	0.66	0.62	0.60	0.48

5. CONCLUSIONS

This paper examined the evolution that has occurred in meat price indices and different types over the past three decades. During this period, similar patterns were demonstrated between some of these types of meat, such as beef and lamb. It should be noted that the beginning of the 90s was mainly dominated by poultry meat and pig meat, and there was a change in consumer preferences reflected in prices. As Gilbert and Morgan (2010) studied there are several factors that make the price of meat highly volatile from shocks in demand and supply, extreme weather conditions, imposition of tariffs and trade barriers or health concerns.

With the incorporation of predictive models and evaluating the individual performance of each one of them, it was possible to identify the one that reflects a better adjustment of the data and a better minimization of precision errors. Prophet performed better compared to SARIMAX and LSTM RNN and provided us with 5-year projections for the meat price index and its different types. Projections point to continued growth in the general price index, mainly driven by beef and lamb, in contrast to the poultry meat price index, which shows some slowdown, and the pig meat price index, which shows a downward trend.

Complementing the analysis of prices, and the specific analysis of the demand side, in this specific case the consumption of pig meat, it was possible to identify factors that impact and enhance higher levels of consumption. In line with what was demonstrated in the loss of importance of the pig sector, variables incorporated in the regression model have some relevance. A causal analysis of statistical inference of the variables was adopted with a good fit of data in the model (R^2 =0.56) but indicating that there are external factors impacting pig consumption. For the future, in addition to the events that may impact price volatility and consequently pressure on supply and demand, it is necessary to pay attention to how the variables that impact higher levels of consumption evolve and whether this importance remains. Higher levels of consumption are justified by economic factors such as GDP per capita or by social factors, including female labour participation or middle years of schooling. Projections for the composition of religions in countries also have preponderance as they have proved to have an impact in the opposite direction.

Finally, the creation of a pipeline with Machine Learning models provides evidence that the model has a better predictive character to explain pork consumption. The model that stood out was the AdaBoost having a better R^2 and RMSE. Analyses with a greater focus on issues related to the existing waste in the industry, productivity or migration to meat substitute products will be a good complement to these analyses presented focusing on the producer's side.

6. REFERENCES

Akbar, M., Hasnudi., & Supriana, T. (2021). Analysis Factors Affecting the Demand and Supply of Free-range Chicken Meat in Pakpak Bharat Regency. Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, 4(1), p.986–998. https://doi.org/10.33258/birci.v4i1.1718.

Araújo, P. D., Araújo, W. M. C., Patarata, L., & Fraqueza, M. J. (2022). Understanding the main factors that influence consumer quality perception and attitude towards meat and processed meat products. Meat Science, 193, p.5-8. https://doi.org/10.1016/j.meatsci.2022.108952

Awan, T.M. and Aslam, F. (2020). Prediction of daily COVID-19 cases in European countries using automatic ARIMA model. Journal of Public Health Research, 9(3). https://doi.org/10.4081/jphr.2020.1765.

Ben Abdallah, M., Fekete Farkas, M. & Lakner, Z. (2020). Analysis of meat price volatility and volatility spillovers in Finland. Agricultural Economics (Zemědělská ekonomika), 66(No. 2), p.84–91. https://doi.org/10.17221/158/2019-agricecon.

Biau, G. and Scornet, E. (2016). A random forest guided tour. Test, 25(2), pp.197–227. https://doi.org/10.1007/s11749-016-0481-7.

Cao, Y. and Li, X. (2022). Multi-Model Attention Fusion Multilayer Perceptron Prediction Method for Subway OD Passenger Flow under COVID-19. Sustainability, 14(21), p.14420. https://doi.org/10.3390/su142114420.

Chakraborty, P., Corici, M. & Magedanz, T. (2020). A comparative study for Time Series Forecasting within software 5G networks. 2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS). https://doi.org/10.1109/icspcs50536.2020.9310033.

Connor, J.M. (1994). North America as a precursor of changes in Western European food-purchasing patterns. European Review of Agricultural Economics, 21(2), p.155–173. https://doi.org/10.1093/erae/21.2.155.

Duarte, D. and Faerman, J. (2019). Comparison of Time Series Prediction of Healthcare Emergency Department Indicators with ARIMA and Prophet. 9th International Conference on Computer Science, Engineering and Applications (ICCSEA 2019). https://doi.org/10.5121/csit.2019.91810.

FAO, 2005. Global perspective. Food and Agriculture Organisation, Rome. https://www.fao.org/3/a0255e/a0255e01.pdf

FAO, 2022. Food Outlook – Biannual Report on Global Food Markets. Rome. https://doi.org/10.4060/cb9427en

FAO, 2022. Meat Market Review: Emerging trends and outlook. 2022. Rome.

Fathi, M.M., Awadallah, A.G., Abdelbaki, A.M. & Haggag, M. (2019). A new Budyko framework extension using time series SARIMAX model. Journal of Hydrology, 570, p.827–838. https://doi.org/10.1016/j.jhydrol.2019.01.037.

Filzmoser, P. and Nordhausen, K. (2020). Robust linear regression for high-dimensional data: An overview. WIREs Computational Statistics. https://doi.org/10.1002/wics.1524.

Gilbert, C.L. and Morgan, C.W. (2010). Food price volatility. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), p.3023–3034. https://doi.org/10.1098/rstb.2010.0139.

Guarino, A., Grilli, L., Santoro, D., Messina, F. & Zaccagnino, R. (2022). To learn or not to learn? Evaluating autonomous, adaptive, automated traders in cryptocurrencies financial bubbles. 34(23), p.20715–20756. https://doi.org/10.1007/s00521-022-07543-4.

Guo, J. and Tanaka, T. (2020). The Effectiveness of Self-Sufficiency Policy: International Price Transmissions in Beef Markets. Sustainability, 12(15), p.6073. https://doi.org/10.3390/su12156073.

Alavi, H. S., Zhong, S. & Lalanne, D. (2020). Predictive Models of Indoor Carbon Dioxide Concentration to Prevent Daily Decay of Productivity and Well-Being in Shared Offices. CEUR-WS Open-Access (2020). in press.

Harrigan, J. (2014). The Political Economy of Arab Food Sovereignty, Basingstoke; New York, Palgrave Macmillan.

Hobbs, J.E. (2021). The Covid-19 pandemic and meat supply chains. Meat Science, 181, p.108459. https://doi.org/10.1016/j.meatsci.2021.108459

Li, Y., Yang, Y., Zhu, K. & Zhang, J. (2021). Clothing Sale Forecasting by a Composite GRU–Prophet Model with an Attention Mechanism. IEEE Transactions on Industrial Informatics, [online] 17(12), p.8335–8344. https://doi.org/10.1109/TII.2021.3057922.

Macêdo, T. de L. (2014). Estatística robusta aplicada ao mercado de ações e ao índice Bovespa. bdm.unb.br. [online] https://bdm.unb.br/handle/10483/14029.

Manigandan, P., Alam, M.S., Alharthi, M., Khan, U., Alagirisamy, K., Pachiyappan, D. & Rehman, A. (2021). Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models. Energies, 14(19), p.6021. https://doi.org/10.3390/en14196021.

Manrique, J. and Jensen, H.H. (1998). Working Women and Expenditures on Food Away-From-Home and At-Home in Spain. Journal of Agricultural Economics, 49(3), p.321–333. https://doi.org/10.1111/j.1477-9552.1998.tb01275.x.

Manrique, J. and Jensen, H. (1997). Working women and expenditures on food away from home and at home in Spain. American Journal of Agricultural Economics, 79(5), p.1719-1719.

Masci, J., Meier, U., Cireşan, D. & Schmidhuber, J. (2011). Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. Lecture Notes in Computer Science, pp.52–59. https://doi.org/10.1007/978-3-642-21735-7_7.Milford, A.B., Le Mouël, C., Bodirsky, B.L. and Rolinski, S. (2019). Drivers of meat consumption. Appetite, [online] 141, p.104313. https://doi.org/10.1016/j.appet.2019.06.005.

Moscoviz, L. and Evans D. (2022). “Learning Loss and Student Dropouts during the COVID-19 Pandemic: A Review of the Evidence Two Years after Schools Shut Down.” CGD Working Paper 609. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/learning-loss-and-student-dropouts-during-covid-19-pandemic-review-evidence-two-years

OECD/FAO (2016), “Meat”, in OECD-FAO Agricultural Outlook 2016-2025, OECD Publishing, Paris. https://doi.org/10.1787/agr_outlook-2016-10-en.

Wirth, R. and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Vol. 1, p. 29-40.

Schmidhuber, J. and Shetty, P. (2005). The nutrition transition to 2030. Why developing countries are likely to bear the major burden. Food Economics - Acta Agriculturae Scandinavica, Section C, 2(3-4), p.150–166. https://doi.org/10.1080/16507540500534812.

Taylor, S. J., and Letham, B. (2018). ”Forecasting at scale.” The American Statistician 72.1. p.37-45.

Tejeda, H., and Goodwin, B. (2009). Price volatility, nonlinearity, and asymmetric adjustments in corn, soybean, and cattle markets, implications of ethanol-driven (market) shocks. In Paper presented at the 2009 NCCC-134 Conference on applied commodity price analysis, forecasting, and market risk management, St. Louis, MO, April p.20-21. https://ageconsearch.umn.edu/record/53039

von Braun, J. and Tadesse G. (2012). : Global food price volatility and spikes: An overview of costs, causes, and solutions, ZEF Discussion Papers on Development

Policy, No. 161, University of Bonn, Center for Development Research (ZEF), Bonn, https://doi.org/10.22004/ag.econ.120021

Whitton, C., Bogueva, D., Marinova, D. & Phillips, C.J.C. (2021). Are We Approaching Peak Meat Consumption? Analysis of Meat Consumption from 2000 to 2019 in 35 Countries and Its Relationship to Gross Domestic Product. Animals, 11(12), p.3466. https://doi.org/10.3390/ani11123466.

Woertz, E. (2013). Oil for Food. The Global Food Crisis and the Middle East, Oxford ; New York, Oxford University Press.

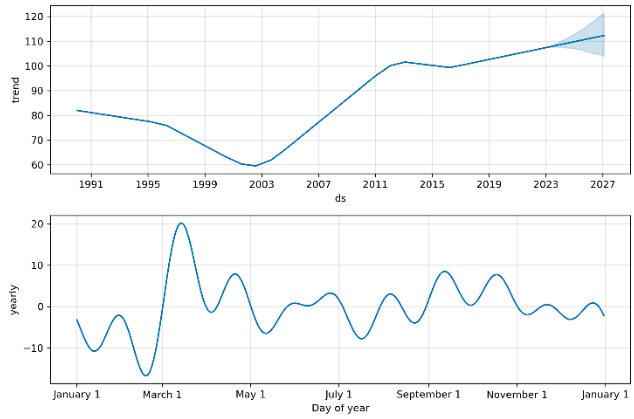
Yaffee, R. A. (2002). Robust Regression Analysis: Some Popular Statistical Package Options, New York.

York, R. and Gossard, M.H. (2004). Cross-national meat and fish consumption: exploring the effects of modernization and ecological context. Ecological Economics, 48(3), p.293–302. https://doi.org/10.1016/j.ecolecon.2003.10.009.

Zhao, D., Zhang, R., Zhang, H. & He, S. (2022). Prediction of global omicron pandemic using ARIMA, MLR, and Prophet models. Scientific Reports, [online] 12(1), p.18138. https://doi.org/10.1038/s41598-022-23154-4.

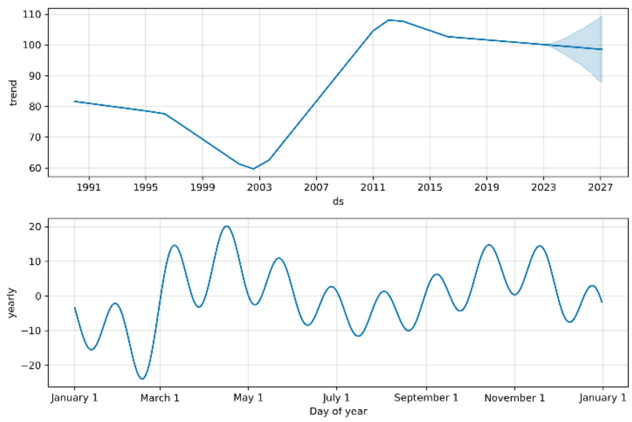
7. APPENDICES

Graph A- 1: Trend and seasonality components of the Meat Price Index

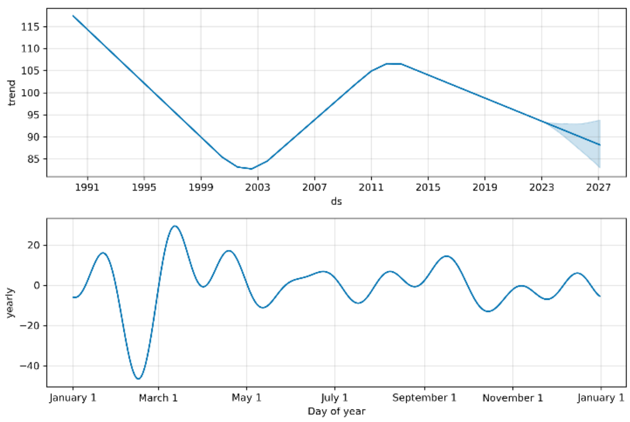


Source: Own elaboration based on FAO (2022)

Graph A- 2: Trend and seasonality components of the Poultry Meat Price Index (9) and Pig Meat Price Index (10).



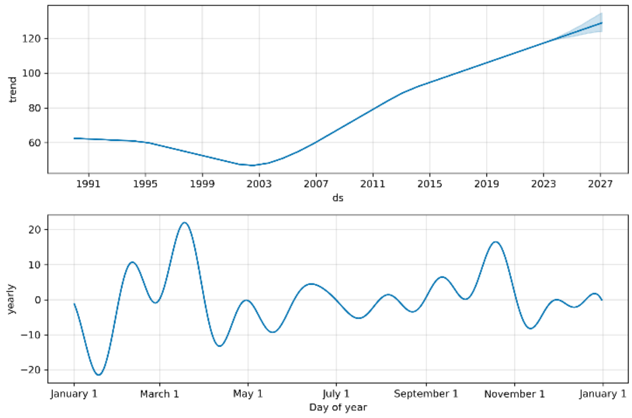
(9)



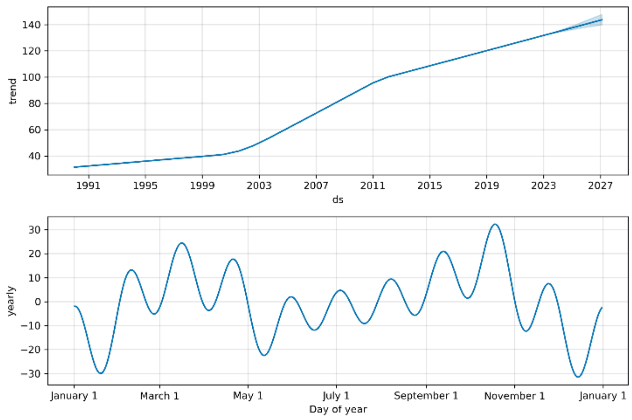
(10)

Source: Own elaboration based on FAO (2022)

Graph A- 3: Trend and seasonality components of the Bovine Meat Price Index (11) and Ovine Meat Price Index (12).



(11)



(12)

Source: Own elaboration based on FAO (2022)

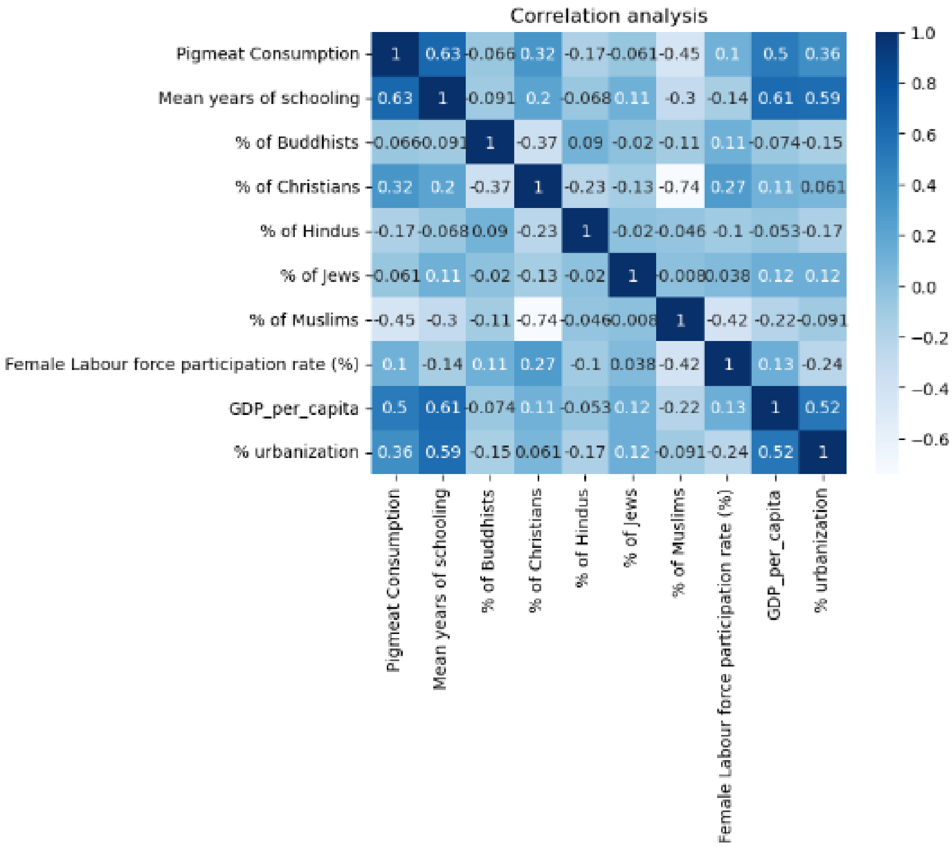
Table A- 1: Prophet Time Series analysis – Meat Price index (poultry, pig, bovine and ovine)

	Trend	Seasonality	Projections (5 years)
Meat Price Index	The 90s were marked by a downward movement, followed by an upward trajectory from the lowest point in the year 2003 until 2022 (from 50 to 110 USD per ton).	March registers the biggest rises.	Rise, following the trend (from 100 to 110 USD per ton), at the 95% significance level.
Poultry Meat Price Index	The 90s were marked by a downward movement until 2003, followed by an upward trend, which	April has the greatest impact, in contrast to February, which has the opposite	Some neutrality, setting prices around 100 USD per ton, at the 95% significance level.

	has been neutralizing since 2015, despite the huge impact felt since 2020 (from 80 to 130 USD per ton).	effect when compared to the others.	
Pig Meat Price Index	Downward trend since the 90s, only mitigated from 2003 to 2011, but then returns to the downward pattern, expecting for the following years an approach to the values practiced in 2003 (85 USD per ton).	Uniformity at the monthly level, with the exception of February which is the month that registers sharp drops in prices, followed by the most significant month in terms of increases: March.	Price decline, following the trend (values around 85/90 USD per ton), at the 95% significance level.
Bovine Meat Price Index	Solid uptrend movement since 2003 (from 40 to 115 USD per ton).	January register the biggest decreases in prices, as opposed to the month of March.	Continuous growth projections, following the trend, at a 95% significance level.
Ovine Meat Price Index	Solid uptrend movement since 2003 (from 40 to 140 USD per ton).	Huge monthly volatility: October as the most impactful month in terms of increases in contrast to December and January in the opposite direction.	Continuous growth projections, following the trend, at a 95% significance level.

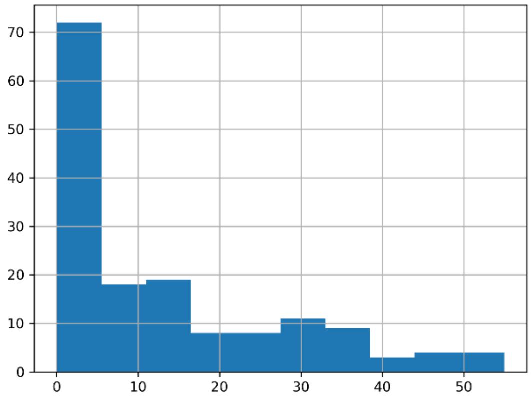
Source: Own elaboration based on FAO (2022)

Graph A- 4: Correlation matrix between variables for the pig meat consumption regression [Prophet Output]

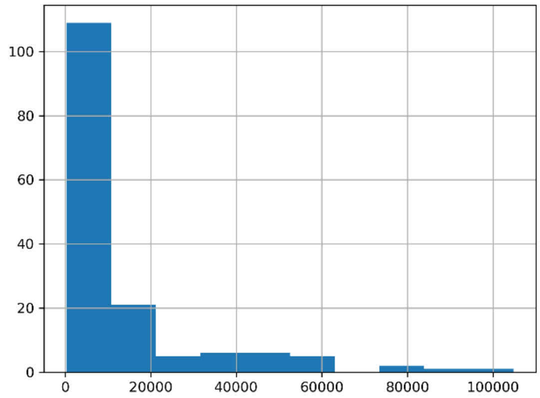


Source: Own elaboration based on FAO (2022)

Graph A- 5: Histograms of the variables pig meat consumption (13) and GDP per capita (14) [Prophet Output]



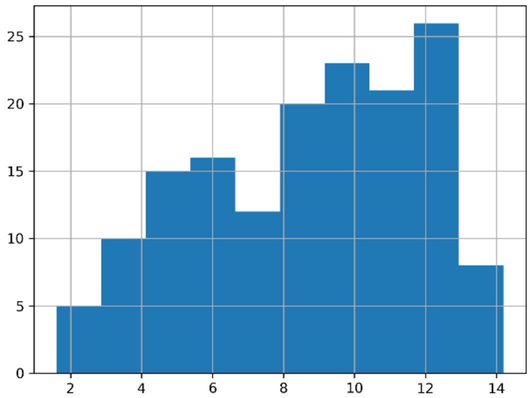
(13)



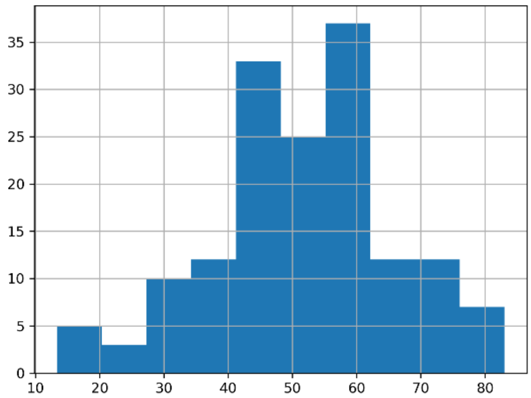
(14)

Source: Own elaboration based on FAO (2022)

Graph A- 6: Histograms of the variables mean years of schooling (15) and female labour force participation rate (16) [Prophet Output]



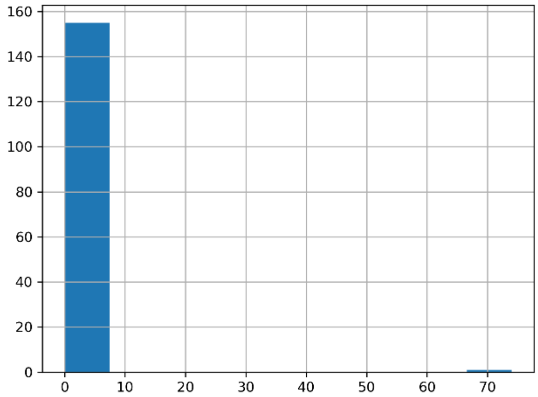
(15)



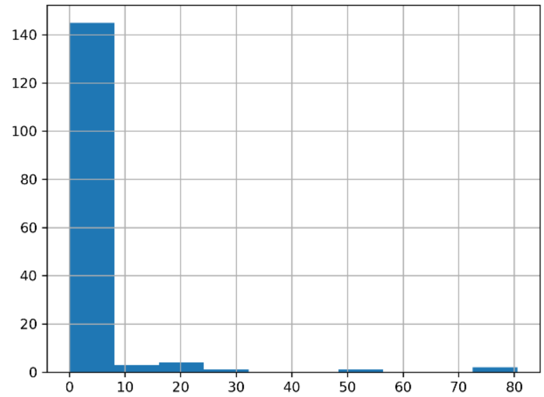
(16)

Source: Own elaboration based on FAO (2022)

Graph A- 7: Histograms of the variables % of Jews (17) and % of Buddhists (18) [Prophet Output]



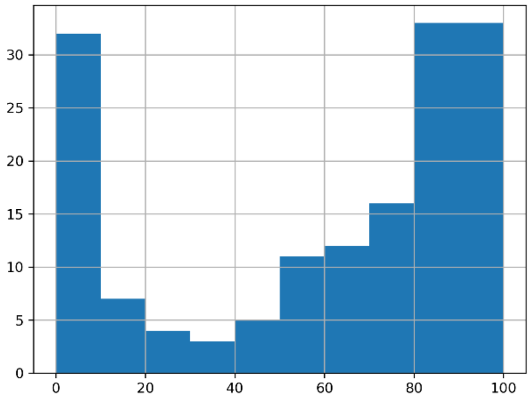
(17)



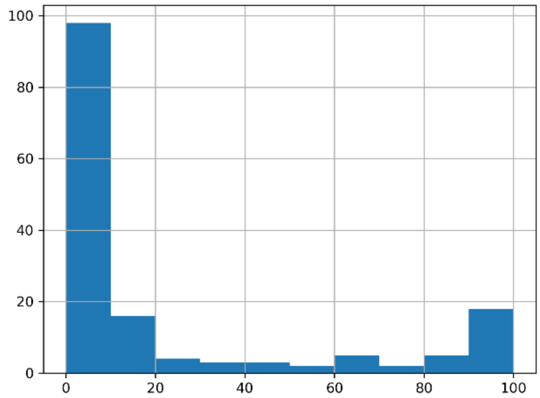
(18)

Source: Own elaboration based on FAO (2022)

Graph A- 8: Histograms of the variables % of Christians (19) and % of Muslims (20) [Prophet Output]



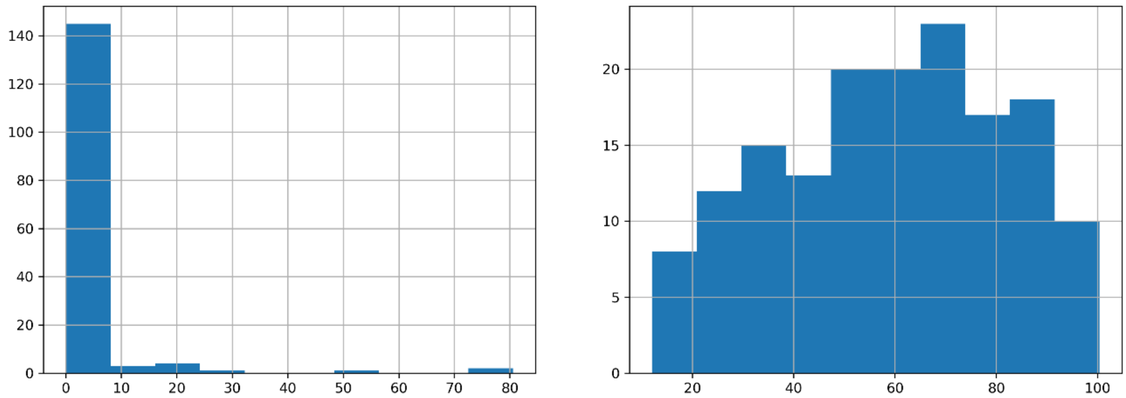
(19)



(20)

Source: Own elaboration based on FAO (2022)

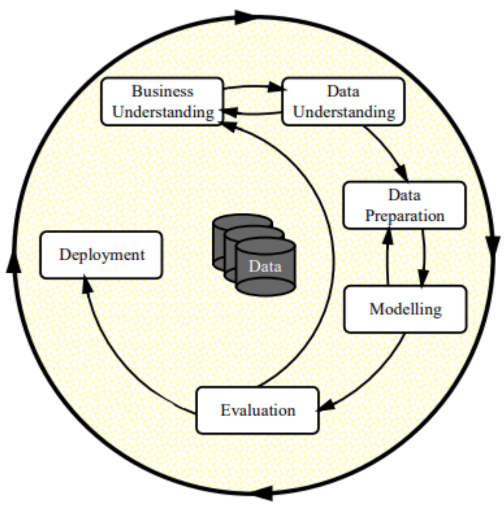
Graph A- 9: Histograms of the variables % of Hindus (21) and % of Urbanization (22) [Prophet Output]



(21) (22)

Source: Own elaboration based on FAO (2022)

Graph A- 10: Phases of the CRISP-DM methodology



Source: Adapted from Wirth and Hipp (2000)

Table A- 2: Python code implementation - White test to assess the existence of Heteroscedasticity in pig meat consumption regression

```
#Importing required libraries
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
import statsmodels.stats.diagnostic as sm_diag
# Define as independent (X) and dependent (Y) variables
X = df_merge3[['Mean years of schooling', '% of Buddhists', '% of Christians', '% of Hindus', '% of Jews', '% of Muslims', 'Female Labour force participation rate (%)', 'GDP_per_capita', '% urbanization']]
Y = df_merge3['Pigmeat Consumption']
# Fit a regression model using the statsmodels library
model = smf.ols('Y ~ X', data=df_merge3).fit()
# Perform White's test
white_test = sm_diag.het_white(model.resid, model.model.exog)
# Print test results
print('White Test Statistics:', white_test[0])
print('White test p-value:', white_test[1])
print('White test critical value:', white_test[3])
```

#Output from jupyter python

```
White Test Statistics: 54.77497861195796
White test p-value: 0.2330860924936047
White test critical value: 0.21171109588937995
```