Exploratory Data Analysis and Forecasting of Coffee Prices

Vittorio Balestrieri

1 Introduction

Coffee, one of the most traded commodities globally, is subject to various factors influencing its price fluctuations. From weather conditions to geopolitical events, numerous elements impact the supply and demand dynamics of coffee in the market. Analyzing the historical price data and identifying the underlying patterns are crucial for stakeholders in the coffee industry, including producers, traders, and consumers.

In this document, we delve into the time series data related to coffee prices. Our objective is twofold: first, to perform an in-depth exploratory data analysis (EDA) to uncover underlying patterns, trends, and relationships within the data, and second, to utilize vector autoregression (VAR) modeling to forecast future coffee prices. By combining statistical techniques with domain knowledge, we aim to provide actionable insights and reliable forecasts to stakeholders in the coffee industry.

2 Data

The data used in this analysis comprises historical time series data related to coffee prices and various factors that may influence them. The dataset consists of multiple variables, including:

- **Coffee Prices**: The primary variable of interest, representing the price of coffee over time. This variable serves as the target for our forecasting models. Resource : yfinance python library.
- **Weather Data**: Information on weather conditions in coffee-producing regions, such as temperature, precipitation, humidity, and sunlight exposure. Weather patterns can significantly impact coffee production and, consequently, prices. Resource : meteostat python library.

The dataset covers a significant time span, allowing for the analysis of long-term trends and seasonal patterns in coffee prices and related variables. Moreover, the data may exhibit non-stationarity, seasonality, and correlation structures that need to be addressed during the modeling process.

Prior to analysis, the dataset undergoes preprocessing steps, including cleaning, normalization, and transformation, to ensure data quality and compatibility with the modeling techniques employed. Missing values, outliers, and inconsistencies are addressed through appropriate methods to maintain the integrity and reliability of the data.

3 Exploratory Data Analysis (EDA)

3.1 Understanding Patterns and Trends

Before diving into the modeling phase, it's essential to gain a comprehensive understanding of the data through exploratory data analysis (EDA). We will explore various aspects of the data, including descriptive statistics, visualizations, and correlation analysis, to identify patterns, trends, and anomalies.

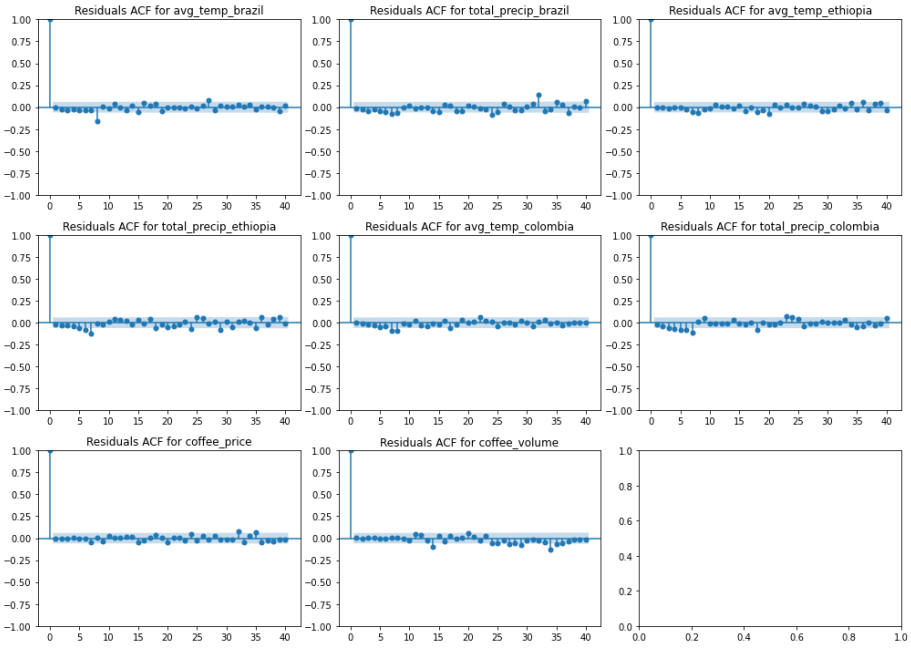

Figure 1: Autocorrelation Function (ACF)

3.2 Identifying Influential Factors

Coffee prices are influenced by a multitude of factors, both internal and external. In this phase of analysis, we will investigate the relationship between coffee prices and other relevant variables such as weather conditions, production volume, economic indicators, and geopolitical events. Understanding the

causal relationships between these factors and coffee prices is vital for accurate forecasting.

3.3 Granger Causality Test

We perform a Granger causality test to identify causal relationships between different variables. Figure 2 presents the heatmap of the Granger causality matrix, highlighting significant causal links between coffee prices and other relevant factors.

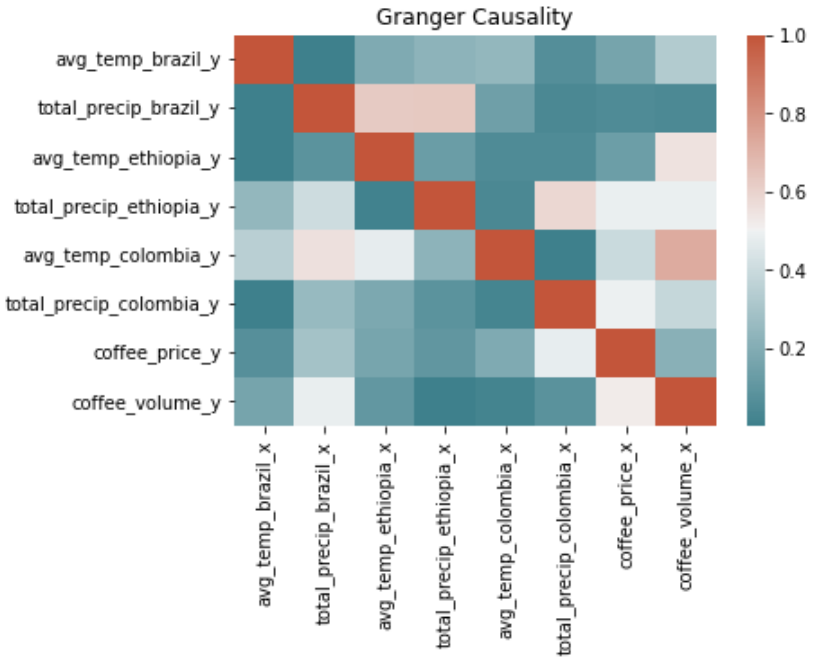

Figure 2: Granger Causality Matrix

4 Vector Autoregression (VAR) Modeling

4.1 Modeling Approach

Vector autoregression (VAR) modeling is a powerful technique used for analyzing the dynamic relationship between multiple time series variables. By capturing the interdependencies among these variables, VAR models enable us to generate forecasts and assess the impact of exogenous factors on the target variable—in this case, coffee prices.

4.2 Forecasting Future Prices

Using historical price data and other relevant variables, we will develop a VAR model to forecast future coffee prices. By leveraging the insights gained from the exploratory data analysis and incorporating relevant factors into the model, we aim to generate accurate and reliable forecasts for stakeholders in the coffee industry.

5 Vector Autoregression (VAR) Forecasting

5.1 Predicted vs. Actual Values

After fitting a VAR model to the data, we forecast future coffee prices and compare them with the actual prices for a specific time period. Figure 3 shows the comparison between predicted and actual coffee prices, providing insights into the forecast accuracy.

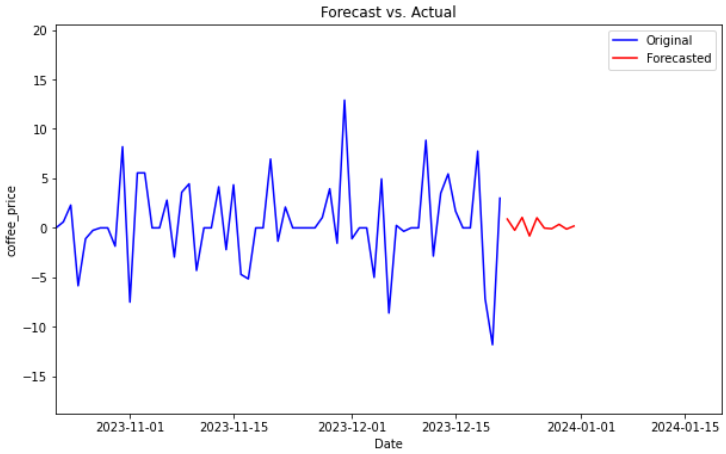

Figure 3: Predicted vs. Actual Coffee Prices