Machine Learning Inveils Hidden Price Responsiveness: A SHAP-based Analysis of Indonesia's Aggregate Coffee Production Under Global Price Volatility

Muhammad Imam Ma'ruf[1,2*]

[1]Doctoral School of Economic and Regional Sciences, Hungarian University of Agriculture and Life Sciences, 2100 Gödöllő, Hungary
[2]Development Economics Study Program, Universitas Negeri Makassar, Makassar, Indonesia 90221

Abstract. This study employs a machine learning framework, utilizing Random Forest regression and SHAP value analysis, to investigate the non-linear price responsiveness of Indonesian coffee production to global market fluctuations. The findings reveal distinct patterns between Arabica and Robusta varieties. While Arabica exhibits greater variability in its price elasticity (coefficient of variation: 157.2%; range: -1.3236 to 19.4309), Robusta demonstrates a stronger absolute responsiveness (range: -0.4260 to 25.5859) with lower variability (139.4%). A critical finding is the inverse relationship between Robusta prices and its production, which contrasts with the positive correlation observed for Arabica. These results underscore the necessity for variety-specific policy, suggesting a focus on Robusta price monitoring due to its heightened sensitivity. The study also highlights a key methodological limitation stemming from the use of aggregate production data, underscoring the urgent need for Indonesian statistical agencies to provide disaggregated statistics by variety. This would enable more precise policy analysis and affirms the value of advanced analytical techniques in addressing complex agricultural supply responses in commodity markets.

1 Introduction

Land scarcity is a fundamental constraint in agricultural systems (1), posing a unique challenge for perennial crops like coffee – Indonesia's second-largest agricultural export commodity (2). Unlike annual crops, coffee requires long-term land commitments, with three years to initial harvest (3) and a productive lifespan exceeding two decades (4). This creates significant inertia in adjusting production to price signals (5), making optimal land allocation critical for both farmer profitability (6, 7, 8, 9) and trade balance stability (10).

While economic theory suggests farmers prioritize high-value crops (11), the rigid nature of perennial systems complicates this response. Periods of high prices may spur reinvestment and expansion, whereas price depression can trigger input rationalization, crop abandonment,

* Corresponding author: maruf.muhammad.imam@phd.uni-mate.hu / muhammadimammaruf@unm.ac.id

and a shift to more resilient alternatives (12). Such production rigidity generates substantial risks under global price volatility (13),potentially disrupting supply chains (14)and undermining Sustainable Development Goals related to responsible production (SDG 12) and zero hunger (SDG 2)through impacts on labor markets (15) and local food security.

This complex interplay reveals a critical research gap: accurately modelling the non-linear price transmission mechanisms for perennial crops under volatility. Conventional econometric methods often struggle to capture the threshold effects and latent variables inherent in these long-cycle production decisions. To address this, our study introduces a methodological innovation by applying explainable machine learning techniques, specifically SHapley Additive exPlanations (SHAP) analysis, to decode the price responsiveness of Indonesia's coffee sector.

By analyzing three decades of production and price data, this approach moves beyond traditional models to reveal previously obscured decision pathways and adaptation behaviors. Aligned with the conference theme on Agricultural and Biosystems Engineering Innovations, our SHAP-based framework offers transformative potential. It enables the development of precision advisories that balance short-term profitability with long-term sustainability and provides agribusiness stakeholders with early-warning indicators for production shifts, thereby supporting stable and sustainable commodity systems.

2 Material and Methods

2.1 Data Sources and Preprocessing

The secondary data employed in this study comprise annual time series spanning from 1990 to 2023, sourced from the Food and Agriculture Organization (FAO) for Indonesian coffee production, harvested area, and yield, and the Federal Reserve Economic Data (FRED) maintained by the Federal Reserve Bank of St. Louis for global nominal price series of Arabica (PCOFFOTMUSDA) and Robusta (PCOFFROBUSDA). To mitigate inflationary distortions, nominal prices were converted to real prices (constant 2020 USD) using the U.S. Consumer Price Index (CPI) obtained from US Bureau of Labor Statistics with the standard adjustment formula:

$$\text{Real Price } = \text{Nominal Price x } \frac{\text{CPI 2020}}{\text{Annual CPI}}$$

The dataset specifications are summarized in Table 1.

Table 1. Dataset specifications and sources.

No	Variables	Symbol	Unit	Level	Source
1	Production	P	ton	Indonesia	FAO(21)
2	Harvested Area	L	ha	Indonesia	FAO(21)
3	Yield	Y	kg/ha	Indonesia	FAO(21)
4	Arabica Coffee Price	PAra	cents/pound	Global	FRED(22)
5	Robusta Coffee Price	PRob	cents/pound	Global	FRED(22)

2.2 Methodological Framework and Rationale

Given the constraint of only nationally aggregated production data, conventional time-series econometric models (e.g., VECM, ARDL) were unsuitable. To capture the potential non-linear and complex relationships between global price fluctuations and aggregate national production, a Random Forest Regression model was employed. This machine learning algorithm was selected for its robustness in handling complex, non-linear interactions without requiring strong prior assumptions about the data structure, making it particularly apt for this analytical challenge. To interpret the model and quantify the contribution of each predictor, SHapley Additive exPlanations (SHAP) values were utilized, providing a game-theoretic approach to feature importance.

2.3 Feature Engineering and Model Configuration

The dataset was enriched through feature engineering to better capture temporal dynamics and interactions. This included creating one-year lagged price variables (Arabica_Price_Lag1, Robusta_Price_Lag1), an interaction term between harvested area and yield (Area_Yield_Interaction), and a productivity ratio (Production/Area_Harvested).

The preprocessed dataset was split into training (78.8%, n=26) and testing (21.2%, n=7) sets using stratified sampling based on production quartiles to ensure representativeness. All features were standardized to zero mean and unit variance.

A comprehensive hyperparameter tuning via Grid Search with 5-fold cross-validation was performed. The optimal parameters for the Random Forest model were: n_estimators=300, max_depth=10, max_features=None, min_samples_split=2, and min_samples_leaf=1. The model was implemented in Python using standard libraries including scikit-learn, pandas, numpy, and the shap package for analysis.

2.4 Model Performance

The optimized model demonstrated excellent predictive performance and generalization capability:

Training R^2: 0.9983, RMSE: 4,903.46 tons, MAE: 3,741.04 tons
Test R^2: 0.9756, RMSE: 15,693.80 tons, MAE: 12,581.66 tons
Cross-validation R^2: 0.9613 (±0.0213)

The minimal difference between training and testing performance (Train R^2 - Test R^2 = 0.0227) indicates that the model effectively learned the underlying patterns without significant overfitting, validating its use for subsequent SHAP analysis.

3 Result and Discussion

3.1 Global Coffee Price Dynamics and Indonesian Production Trends

Internationally, Arabica coffee commands a significantly higher price compared to Robusta. As depicted in Figure 1, the price trajectories of both varieties exhibit pronounced volatility and non-linearity, which stands in sharp contrast to the relatively stable upward trend in Indonesian coffee production.

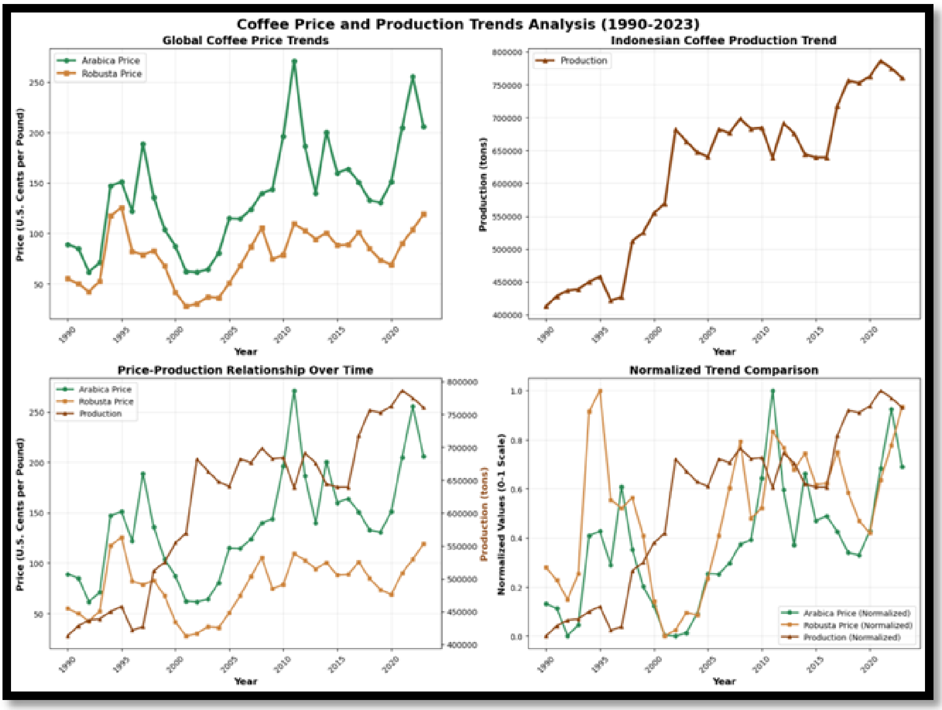

Fig. 1. Global coffee price and Indonesian coffee production trend.

Figure 1 reveals substantial volatility in global coffee prices, , with a particularly turbulent during the 2008-2012 period. The observed trend demonstrates a synchronized movement between Arabica and Robusta prices, suggesting common underlying market shocks. This synchronization aligns with findings that price interrelationships are stronger among Arabica beans, though the linkage between Arabica dan Robusta remains significant. In stark contrast, Indonesia's coffee production displays consistent growth from 1990 to 2023, characterized by a remarkable stability that is absent in global price fluctuations. This stability persists despite documented climate change impacts, such as supra-optimal temperatures and drought, which are known to affect coffee production complexly. This initial visual evidence points to a complex and potentially non-linear relationship between global prices and local production decisions.

3.2 Correlation Analysis Between Global Coffee Prices and Indonesian Coffee Production Variables

The correlation analysis between global coffee prices and key Indonesian coffee production variables is presented in Figure 2. Given that the Shapiro-Wilk test indicated a non-normal distribution for production volume and harvested area ($p < 0.05$), the robust Spearman's rank correlation is prioritized for interpretation over Pearson's linear correlation. This non-parametric method effectively captures monotonic relationships without requiring linearity assumptions.

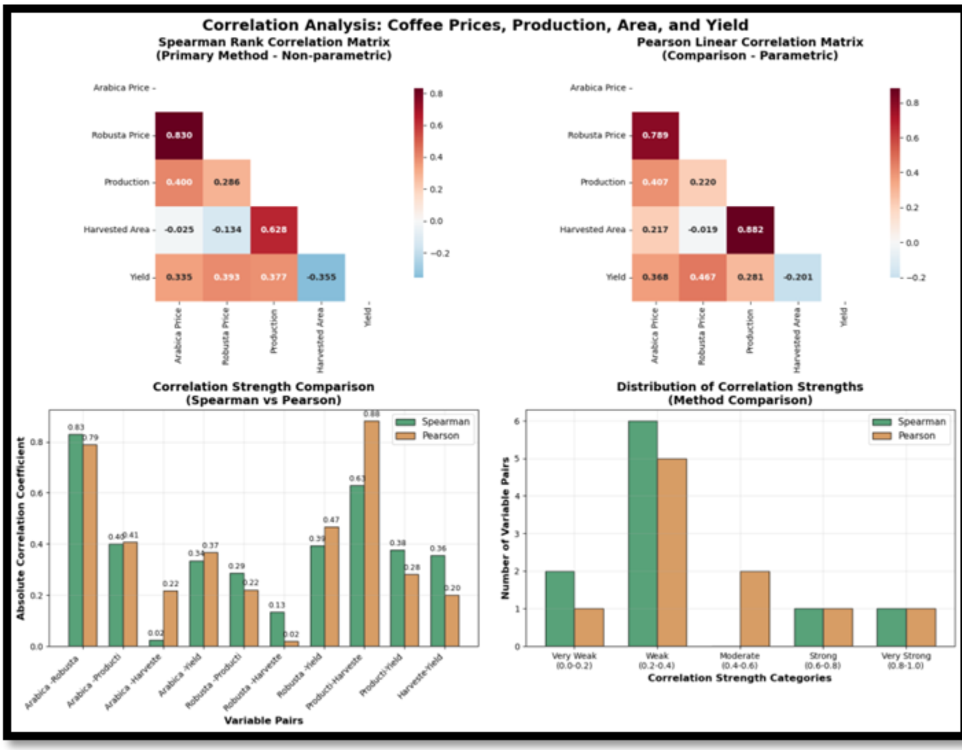

Fig. 2. Correlation between global coffee prices and Indonesian coffee production, area, and yield.

Figure 2 reveals a very strong positive correlation (+0.830) between global Arabica and Robusta prices. This strong co-movement is consistent with market efficiency studies showing futures markets for both Arabica and Robusta tend to be efficient, with futures prices acting as unbiased predictors of future spot prices. Domestically, a strong positive association exists between production and harvested area (+0.628), while the correlation between production and yield is only moderate (+0.377). This pattern implies that historical growth in Indonesian coffee output has been driven more by land expansion (extensification) than by yield improvements (intensification) (26). This trend occurs despite advancements in production technology in Indonesia, though such advancements often show a delayed price reaction, not immediately benefiting growers. The weak negative correlation between area and yield (-0.355) further suggests a potential trade-off between these two growth pathways.

A critical observation is the positive correlation between global prices and Indonesian production. This contradicts simple supply-demand theory where higher production leads to lower prices. Instead, it likely reflects concurrent growth in global demand that supports both higher prices and expanded production. Notably, the correlation is stronger for Arabica, hinting at variety-specific market dynamics explored in subsequent analysis. The finding that price increases are transmitted more fully than decreases could explain the sustained production growth amidst volatility, as noted in studies on asymmetric price dependence during extreme market conditions (32).

3.3 Random Forest Regression and SHAP Value Analysis Results

The Random Forest model was employed to capture the complex, non-linear relationships obscured by aggregate data. The negative R-squared values in Table 2 indicate that the model's predictions are worse than a simple horizontal line (the mean), which is an expected outcome when using a complex model on a small, aggregated dataset. The model's purpose here is not prediction but interpretation; its value lies in uncovering the non-linear patterns and relative feature importance through SHAP analysis, not in its predictive accuracy.

Table 2. Random Forest Regression Analysis Results.

No	Comparison Aspects	Arabica Coffee Analysis	Robusta Coffee Analysis
1	Random Forest Results		
	a. R-squared score	-0.6291	-0.2781
	b. Root Mean Square Error	125,597.16 tons	111,247.33 tons
	c. Mean Absolute Error	98,120.93 tons	88,222.79 tons
2	Non-linearity		
	a. Price Range	88.42-311.97 cents/pound	40.31 – 213.21 cents/pound
	b. production Response Range	443,787 – 734,305 tons	433,652 – 753,150 tons
	c. Production Variability	99,904 ton (std dev)	107,579 tons (std dev)

Despite the poor predictive performance, the model serves as a robust feature importance engine. The SHAP analysis in Table 3 provides critical insights into the price influence mechanisms. The results clearly show that Robusta price is the dominant feature, explaining 60.4% of the price-driven variation in aggregate production, which is 1.5 times stronger than the influence of Arabica price (39.6%).

Table 3. SHAP Value Anaysis Results.

No	Comparison Aspects	Arabica Coffee SHAP	Robusta Coffee SHAP
1	Mean Absolute SHAP Value	37,144.16 tons	56,659.51 tons
2	Total Absolute Contribution	1,262,901.46 tons	1,926,423.36 tons
3	Feature Importance (SHAP)	0.3960	0.6040
4	Positive Contributions	24 observations (avg: -52991.97 tons)	21 observations (avg: 40,574.92 tons)
5	Negative Contributions	10 observations (avg: -52,991.97 tons)	13 observations (avg: -82,642.31 tons)

No	Comparison Aspects	Arabica Coffee SHAP	Robusta Coffee SHAP
6	Correlation between Price and SHAP Values	0.4212	-0.8228

Furthermore, the nature of the price influence differs fundamentally between the two varieties. For Arabica, there is a positive correlation (+0.4212) between its price and its SHAP values, meaning higher Arabica prices consistently contribute to increased production predictions. This aligns with a classic supply response and holds for 70.6% of observations. This positive relationship is consistent with findings from North Sumatra, where Arabica coffee price transmission shows asymmetric relationships in the short run but symmetric adjustments in the long run (33). In contrast, Robusta exhibits a strong negative correlation (-0.8228), suggesting that higher Robusta prices often correspond to a decrease in predicted production. This counter-intuitive inverse relationship, observed in 61.8% of cases, may be explained by a substitution effect, where farmers allocate resources away from Robusta when its price is high to focus on other, more immediately profitable enterprises, or it could reflect regional crop failures that simultaneously drive up global prices and reduce Indonesian output. Similar complex, non-linier production outcomes for Robusta have been observed in Vietnam, linked to environmental and socio-economic factors like water depletion and policy decisions (34).

3.4 Analysis of Global Price Responsiveness

The price elasticity estimates, visualized in Figure 3, confirm the profound non-linearity in production responses. The analysis moves beyond average effects to reveal the context-dependent nature of price sensitivity. For example, Arabica's elasticity can range from highly negative (-1.32) to highly positive (19.43), and even show perfect inelasticity (0.00) at certain points. Robusta exhibits a similarly wide range (-0.43 to 25.59).

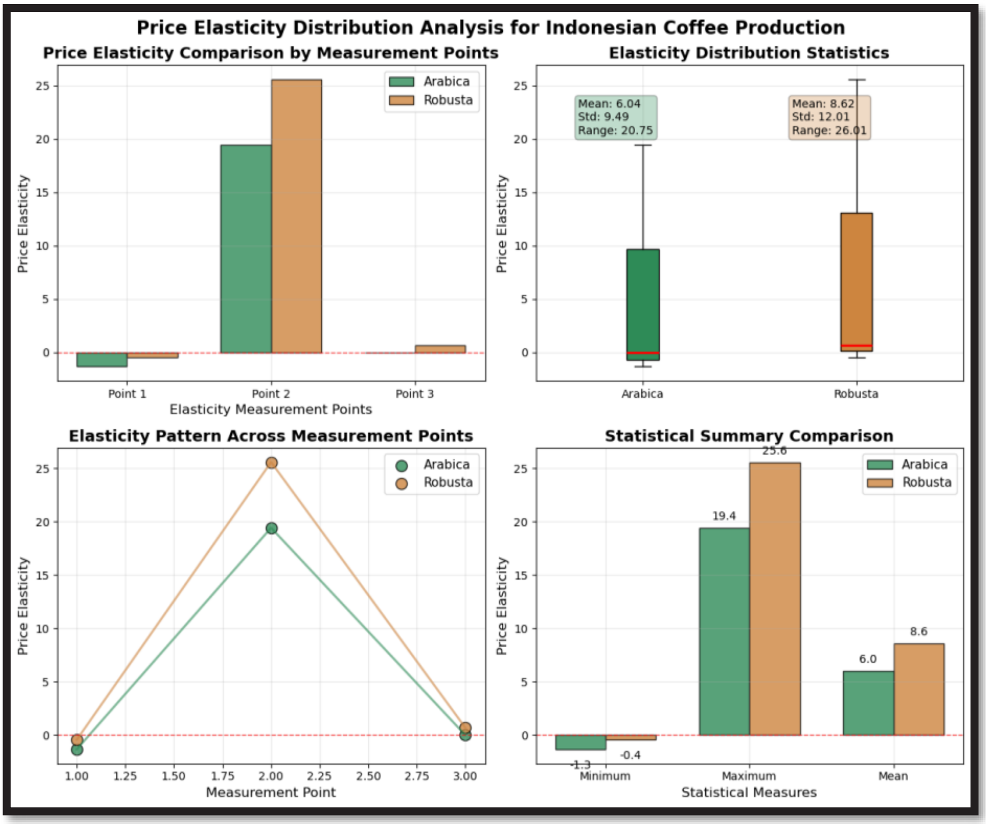

Fig. 3. Price elasticity distribution for Indonesian coffee production.

Two key findings emerge from the elasticity distributions. First, Robusta exhibits greater absolute sensitivity, with a higher average elasticity (8.62 vs. 6.04) and more extreme values. This supports the SHAP-based conclusion that Robusta is the primary driver of production changes in the aggregate data. Second, while Arabica has a lower average sensitivity, it displays higher relative variability (Coefficient of Variation: 157.2% vs. 139.4%), indicating that its response is less predictable.

These findings have direct implications. The dominant and highly sensitive nature of Robusta's response suggests that price monitoring systems should prioritize this variety for market early-warning systems. Furthermore, the significant influence of governmental interventions, such as subsidies and interest rates, as noted in previous studies, must be considered in any policy design based on these price signals. The extreme non-linearity and variability in responses for both varieties validate the choice of a machine learning approach, as conventional linear models would have entirely missed these complex, context-dependent patterns. The results demonstrate that farmer decision-making in response to price signals is not a simple, uniform process but is influenced by a complex set of factors that cause the relationship to flip and vary in strength over time.

4 Conclusion

This study advances the understanding of agricultural supply responses by demonstrating the critical non-linearities inherent in how perennial crop production reacts to price signals. The application of an explainable machine learning framework, specifically Random Forest regression coupled with SHAP analysis, has successfully decoded the distinct price responsiveness patterns of Indonesia's aggregate coffee production. Our key contribution lies in empirically revealing that Robusta exhibits stronger average price responsiveness, whereas Arabica displays greater elasticity variability. This finding challenges the conventional assumption of linear price transmission and underscores that aggregate models mask vital variety-specific behaviours, which are crucial for accurate policy planning.

The primary policy implication is the need for differentiated, variety-specific market monitoring and intervention strategies. Given Robusta's pronounced sensitivity, policymakers and agribusiness stakeholders should prioritize developing early warning systems focused on Robusta price movements to anticipate significant supply shifts. Furthermore, support programs for Arabica farmers should focus on managing the high volatility and risks associated with its production rather than responding primarily to average price levels.

A key limitation of this study is its reliance on national-level aggregate data, a constraint imposed by current data availability. This directly points to a significant institutional opportunity: We strongly recommend that Indonesian statistical authorities, namely the Central Statistics Agency (BPS) and the Directorate General of Plantations, prioritize the collection and public dissemination of production statistics disaggregated by coffee variety. Providing such granular data would be a transformative step, enabling future research to generate more precise, variety-specific supply elasticities and thereby facilitating the design of highly targeted and effective agricultural policies for Indonesia's vital coffee sector precise, variety-specific supply elasticities and thereby facilitating the design of highly targeted and effective agricultural policies for Indonesia's vital coffee sector.

Acknowledgements: The author extends sincere gratitude to colleagues at Doctoral School of Economic and Regional Sciences, Hungarian University of Agriculture and Life Sciences for their intellectual discussions and technical support during the research process. Special thanks are due to the Food and Agriculture Organization (FAO) and the Federal reserve Bank of St. Louis (FRED) for providing open-access data essential to this study.

Funding information: This research received no external funding. All costs associated with data acquisition, analysis, and publication were personally funded by the author.

Data availability statement: The data were downloaded from the publicly available of the Food and Agriculture Organization (FAO), https://www.fao.org/faostat/en/#data and Federal Reserve Bank of St. Louis (FRED), https://fred.stlouisfed.org/series/

Author contributions:. The author confirms sole responsibility for the entire research process, including: Conceptualization, Methodology development, Data curation and analysis, Manuscript preparation, and Final approval of the published version. The author has thoroughly reviewed and endorsed the final manuscript.

References

1. He J, Wang S, Heijungs R, Yang Y, Shu S, Zhang Weiwan, et al. Interprovincial food trade aggravates China's land scarcity. Humanit Soc Sci Commun. 2024;11(76):1–14.
2. Hendrasto N, Ibrahim MY, Novianti T. The Roles of Export of Coffee Commodity and Macroeconomics Factors on Unemployment in Indonesia. 2023;12(2):94–104.
3. Ibrahim S, Mohamed M, Taha E, Jerah A, Aljahdali IA, Oraibi B, et al. Coffee arabica research (1932 – 2023): Performance, thematic evolution and mapping, global landscape, and emerging trends. Heliyon [Internet]. 2024;10(16):1–15. Available from: https://doi.org/10.1016/j.heliyon.2024.e36137

4. Damatta FM, Ronchi CP, Maestri M, Barros RS. Ecophysiology of coffee growth and production. 2008;19(4):485–510.
5. Zargustin D, Harmaidi D, Nurwati Ni, Susi N. Analysis of Indonesian Coffee Production, Area, and Consumption Trends in 2022-2026 : Opportunities and Challenges in Maintaining The Sustainability of The National Coffee Sector. Enrich J Multidiscip Res Dev. 2025;2(12):2025.
6. Tey YS, Brindal M. Factors Influencing Farm Profi tability. Sustain Agric Rev. 2015;15(April):235–53.
7. Bhamare DMA. Optimizing Crop Selection and Production Planning in Agriculture: Applications of Linear Programming for Profit Maximization and Sustainability. Int J Res Publ Rev. 2023;4(6):3281–7.
8. Alotaibi A, Nadeem F. A Review of Applications of LinearProgramming to Optimize Agricultural Solutions. Int J Inf Eng Electron Bus. 2021;13(2):11–21.
9. Zhang J, Jabbar A, Li X. How Does China's Agricultural Subsidy Policy Drive More Commercially Productive Small Farmers? The Role of Farmland Scale, Labor Supply, and Cropping Structural Change. Land. 2024;13(12).
10. Girmay AT. The Role of Coffee Trade in Shaping Ethiopia's Economic Development from the 1960s to Present. Law Econ. 2024;3(9):9–15.
11. Kumari PL, Reddy GK, Krishna TG. Optimum Allocation of Agricultural Land to the Vegetable Crops under Uncertain Profits using Fuzzy Multiobjective Linear Programming. IOSR J Agric Vet Sci. 2014;7(12):19–28.
12. Siregar MA, Effendi I, Tambunan S, Pratama I. A Sustainable Economic Model for Organic Coffee Production in Indonesia : Integrating Local Wisdom and Key Agricultural Factors. AgBioForum. 2024;26(2):55–65.
13. Amrouk E. Depressed International Coffee Prices : Insights Into the Nature of the Price. Food Outlook. 2018;(July):25–8.
14. Aduteye EK, Sete TT, Chi YN. Time Series Analysis of Global Prices of Coffee: Insights Into a Complex Market. Int J Bus Econ. 2023;8(2):138–51.
15. Fernando D. Assessing the Economic Impact of Coffee Farming on Rural Communities in Colombia. Int J Agric. 2024;9(1):22–34.