Journal of Scientific Agriculture 2025, 9: 1-15
doi: 10.25081/jsa.2025.v9.9411
https://updatepublishing.com/journal/index.php/jsa

Research Article

ISSN: 2184-0261

Economic and financial viability of wheat production in Cameroon

Adama Farida[1]*, Ngonkeu Mangaptche Eddy Leonard[2], Jean Marie Gankou[1]

[1]Faculty of Economic Sciences and Management, University of Yaoundé II. P.O. Box 1365 Yaoundé, and P.O. Box 18 Soa, Centre Region, Republic of Cameroon, [2]Department of Plant Biology, Faculty of Science, University of Yaoundé I. P.O. Box 812 Yaoundé, Centre Region, Republic of Cameroon

Received: October 22, 2024
Revised: January 16, 2025
Accepted: January 17, 2025
Published: January 28, 2025

***Corresponding author:**
Adama Farida
E-mail: djanaboudjanatou@yahoo.fr

ABSTRACT

This study evaluated the economic and financial profitability of wheat production in Cameroon using data from 300 individuals in Adamawa, North-West, and West regions. Key factors influencing profitability were identified through correlation heatmaps, pair-plot diagrams, and modeling algorithms (Generalized Least Squares, Random Forest, and Least Absolute Shrinkage and Selection Operator). Positive factors included production volume, packaging, and transportation costs, while negative factors included production workforce, experience, and fertilizer costs. The net margin for wheat production was positive at 76,691,000 FCFA, but financial profitability was low, with an import-to-export ratio of 0.16. The study highlights the need to enhance wheat production to reduce importation.

KEYWORDS: Economic Profitability, Financial Profitability, Wheat Production, Cameroon

INTRODUCTION

Agriculture is the main source of income in rural Cameroon, employing about 70% of the workforce and contributing 22.9% to national wealth in 2013. With significant agricultural potential, Cameroon could enhance its production to feed its growing population and meet regional demand. Arable land is estimated at 7.2 million hectares, but only 1.8 million hectares are cultivated, and only 33,000 out of 240,000 hectares of irrigable land are used (MINADER/AFD/IRAM, 2015). Agriculture accounts for over a quarter of GDP and half of non-oil export revenues. Approximately 70% of the population depends on agriculture and livestock, and 90% of rural households engage in agricultural activities, with a third earning from cash crops. Cameroon is the largest agricultural producer and exporter in the CEMAC region, though characterized by smallholder farming with 63% working on less than 2 hectares. Modernizing agriculture could reduce unemployment, food insecurity, and poverty while boosting farm incomes, reducing food prices, cutting import bills, and improving the balance of trade, thereby enhancing the well-being of small rural farmers.

Cereal growing is a key sector of Cameroon's agricultural economy, vital for human consumption, animal feed, and industrial uses (Guinness, Cameroon Breweries). With the global population expected to reach 9.6 billion by 2050, Cameroon faces the challenge of feeding its 27.2 million inhabitants amid a worsening cereal shortage. From 2005 to 2007, 21% of Cameroonians (3.9 million) were undernourished. In 2011, the cereal deficit was 640,000 tons, with 25% of cereals imported (FAO, 2011). Despite significant agricultural potential, national wheat production was only 800-950 tons annually from 2010-2014, barely reaching 900 tons in 2016 (FAO, 2016). This highlights Cameroon's reliance on imports and the need for agricultural development to meet future food needs.

Cameroon imports over a billion USD worth of agricultural products annually, with 725,000 tons of wheat imported in the past decade. This dependency undermines government efforts due to foreign currency flight. The bankruptcy of the wheat development corporation (SOciété de DÉveloppement de BLÉ, SODÉBLÉ) increased wheat imports, making it crucial for the government to revive the wheat sector to meet rising demand. Various wheat varieties have been tested across the country to reduce imports. Investigating the causes for wheat production abandonment is vital, as sustainable production impacts socio-economic and environmental well-being. This study analyzes the economic and financial profitability of wheat production in Cameroon through surveys of producers in Adamawa, North-West, and West regions.

MATERIALS AND METHODS

Study Areas

The Adamawa region

Geographical location

The Adamawa Region (6°-8° N, 11°-15° E) covers 63,701 km². It consists of highlands stretching from Nigeria to the Central African Republic, positioned between Cameroon's southern and northern parts. It borders the Centre and East Regions to the south, the West and North-West Regions to the southwest, and the North Region to the north. Administratively, it includes 5 Divisions and 21 Sub-divisions (Decree N° 2008/376). The region's diverse landscapes are divided into two agroecological zones: the Sudano-Sahelian (8°-13° N) and the Guinean high savannah (4°-8° N) (Figure 1).

Relief

The Adamawa plateau is a high block of bedrock with small volcanoes, ranging from 900 to 1,500 m in altitude, peaking at 1,800 m, with U-shaped valleys.

Climate

The region has a tropical climate with bimodal rainfall in central and eastern low savannahs, and monomodal rainfall in the north. The climate types include bimodal equatorial Guinean, monomodal tropical Sudan, and monomodal equatorial Cameroonian. The high altitude ensures relatively cool weather with temperatures from 22 °C to 34 °C, dropping to 10 °C at times.

Population

The population grew from 359,334 in 1976 to 495,185 in 1987 (RGPH-2), and 884,289 in 2005 (RGPH-3) with 49.6% men and 50.4% women. The urban population was 343,490, and the rural 540,799, with a 38.8% urbanization rate. The average growth rate increased from 2.9% (1976-1987) to 3.2% (1987-2005). The projected population in 2018 is 1,274,325, with 43% under 15. The average age was 20.8 years. The sex ratio was 98.5 men per 100 women in 2005, estimated at 95.7 women per 100 men in 2018. Population density was 20 inhabitants/km² in 2018, up from 13.9 in 2005.

Agricultural activities

Over 30% of the population engages in agriculture, mainly cultivating maize, cassava, potatoes, yams, and groundnuts, primarily for local consumption. Tropical fruits and vegetables, and beekeeping are also significant. We covered two departments (Vina and Djerem) and five arrondissements: Ngan-ha, Ngaoundéré 1er, Mbé, Nyambaka, and Tibati.

The North-West region

Geographical location

The North-West Region, part of former West Cameroon, covers 17,300 km². It borders Nigeria to the north, the South-West

Figure 1: Administrative map of the Adamawa region of Cameroon

Region to the west, the Adamawa Region to the east, and the West Region to the south. Its capital, Bamenda, is one of the country's two English-speaking regions. In 2012, it had 7 Divisions and 34 Sub-divisions (Figure 2). The north-western part of the region is mountainous, with altitudes up to 3,000 meters.

Relief

The region features panoramic contrasts with plains, mountains, deep valleys, streams, waterfalls, and crater lakes.

Climate

The climate is mild and cool year-round, except for the rainy season (July to October), with temperatures around 22 °C.

Population

In November 2005, the population was 1,728,953. By 2014, it was 1.93 million, growing at a strong demographic trend. The population is very young, with more than half under 20 years old and 44% under 15. Seniors (60+) made up 5.5% in 2014. In 2010, the gender distribution was 49.47% men and 50.53% women.

Agricultural activity

Agriculture is a major economic activity, despite being affected by the security crisis. It employs 70% of the population, with the region being the main Arabica coffee producer (over 70% of national production) and also producing cocoa, tea, oil palm, rubber, and food crops. It is ideal for market garden crops like lettuce, leeks, radishes, and peppers. The Ndop plain is a significant rice-growing area. Surveys covered two departments, Boyo and Mezan, including the Njinikom and Tubah Sub-divisions.

The West region

Geographical location

The West Region (5°S-16°N, 10°E-11°W) covers 13,892 km^2 (3% of national territory). It borders the North-West and Adamawa Regions to the north, Coastal Region to the south, Centre Region to the east, and South-West Region to the west. It is a key transit and trade zone, with eight Divisions, about forty Subdivisions, and forty towns (Figure 3).

Relief

The *Hauts-Plateaux de l'Ouest* features high-altitude terrain (average 1,100 m), and diverse landscapes with mountains, plateaus, and basins. Dominated by meadows and volcanic soils, the region also has alluvial soils in valleys. The Noun River and Lake Bamendjin, formed by a hydroelectric dam, are notable features.

Climate

The high altitude results in cool temperatures (around 25 °C) and heavy rainfall, making the air quality good. The Dschang Climate Center at 1,380 m is the only temperate location south of the Sahara in Africa.

Population

In 2015, the West Region had over 180,000 people, almost 9% of Cameroon's population. The urban population was about 49% (sex ratio 88.1). The average age was 23, with 46.8% men and 53.2% women. Ethnic groups include Bamilékés, Bamouns, Tikars, Mbô, Mbororos, and Peuls. Population density was 137 inhabitants/km^2, the second highest in the country.

Agricultural activity

Agriculture is the main activity, with cash crops like cocoa, tea, banana plantain, robusta and arabica coffee. The region

Figure 2: Administrative map of the North-West region of Cameroon

Figure 3: Administrative map of the West region of Cameroon

also grows food crops and markets produce such as tomatoes, potatoes, manioc, macabo, corn, carrots, and green beans. Livestock and poultry farming are significant, with many products destined for export. Surveys covered eight divisions and nineteen sub-divisions with support from the Institute of Agricultural Research for Development (IRAD) and the Ministry of Agriculture and Rural Development (MINADER).

General Information on Wheat

Wheat (*Triticum sp.*) is a key cereal from the grass family *Poaceae*, feeding over 70% of the global population. It's cultivated for its starch-rich seeds (70%), protein (10-15%), and pentosans (8-10%), widely used in human nutrition (58% for bread flour, cookies, pasta, etc.,), animal feed (34%), and industrial purposes (8%). A wheat-rich diet can help prevent cardiovascular disease, manage type 2 diabetes, and avoid colon cancer and other age-related diseases, but can cause symptoms of celiac disease, allergies, and kidney stones.

Reviving Cameroon's wheat industry is a government priority. The Institute of Agricultural Research for Development (IRAD) collaborates with the Ministry of Agriculture's PROSAPVA project to improve yields and expand wheat cultivation by distributing quality seeds and fertilizers to producer groups and cooperatives.

In Cameroon, national wheat production is 4-5 tons/ha. Agricultural experts believe production is feasible in five agroecological zones. Modern wheat cultivation has been established in the country. Recently, IRAD developed and tested several wheat varieties on over 305 experimental plots at high and low altitudes, listed in Cameroon's 2018 official catalog (Figure 4). However, dissemination through MINADER to farmers in adapted sites has not yet shown good results.

Data Collection

We collected both quantitative and qualitative data from primary and secondary sources. Primary data was collected via surveys of wheat growers in three regions, using questionnaires. This included data on wheat production inputs and outputs: fixed cost items (farm tools and equipment, their price and lifespan), variable cost items (inputs, labor, their availability and use, their cost), and farm income items (products and their sale prices). Secondary data was mainly drawn from FAOSTAT and World Bank data, books, dissertations, theses, articles, and experimental reports. The survey sample was designed to provide reliable estimates for several indicators relevant to our study. We surveyed 300 growers, 100 per region.

Determination of Variables for Analyzing Economic and Financial Profitability

Several authors have attempted to define profitability. Measuring profitability assesses the optimal allocation of production factors. Sion (2020) states profitability is more complex than profitability; it relates results to invested capital. Financially, investment aims to maximize added value or return on investment. A study of PARSE's community businesses in Adamawa shows profitability as the relationship between income and used resources. Profitability includes financial and socio-economic factors, related to sourcing, production, and marketing (GIZ/PARSE-AD, 2017). Profitability is a company's ability to generate profits over time (Makelele, 2014). It assesses capital performance and management efficiency. Pirou (2005) defines profitability as capital's ability to generate income. Comparing profit with employed capital calculates percentage gain. Batsch (2003) reserves profitability for income-to-capital relationships, while income-to-sales is the margin or rate of gain. Turnover or return on capital refers to sales or investment return.

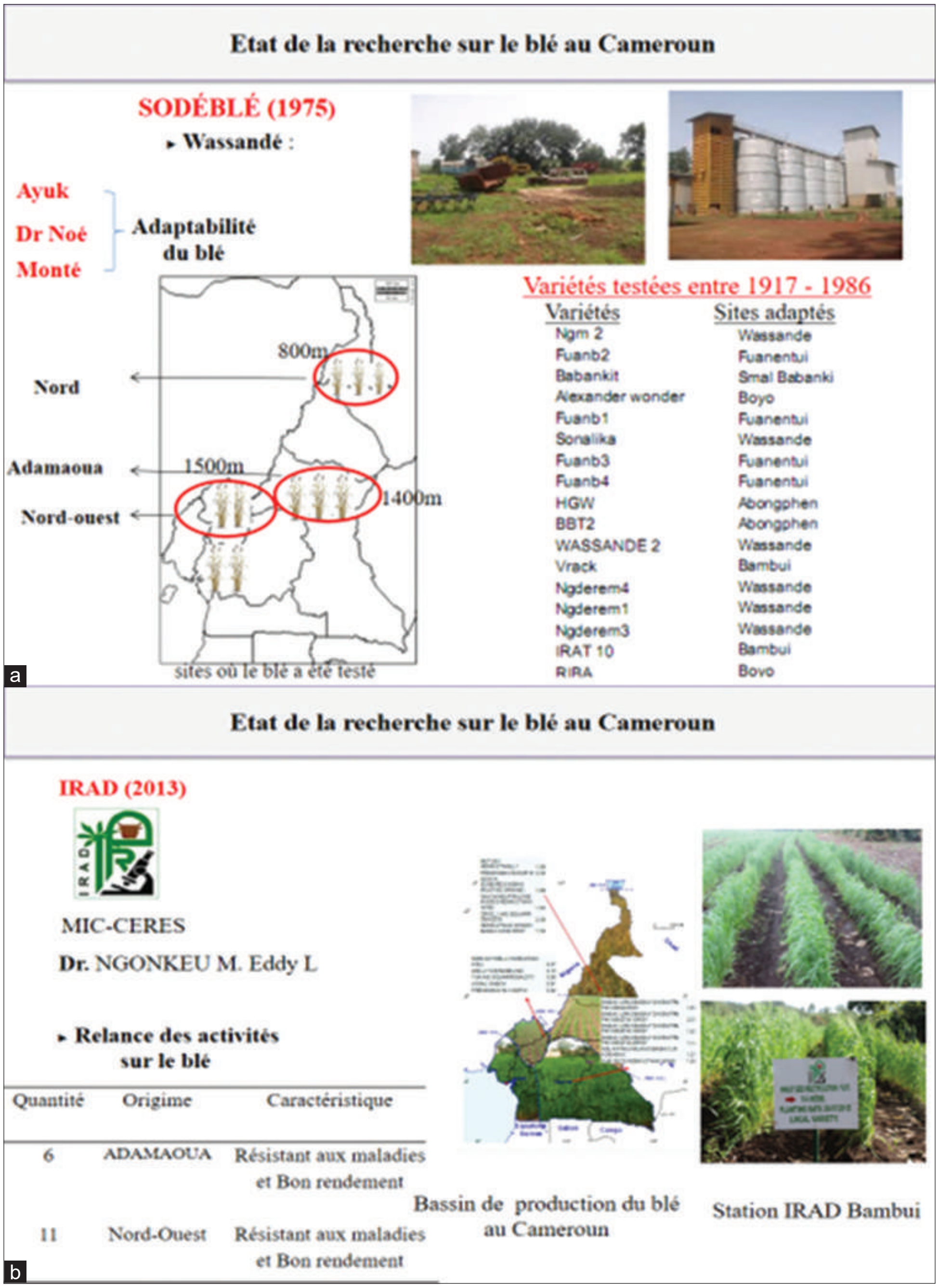

Figure 4: Wheat seed a) testing and b) production sites in Cameroon

Economic profitability indicator

Pirou (2005) also defines economic profitability or return on investment (ROI) as the comparison of income generated by a company with the capital committed to production, regardless of equity or debt. It measures asset performance and the relationship between revenue and expenditure. The six economic rates of return depend on income and expenses. The net production margin, or net profit, is calculated by subtracting total costs (fixed and variable) from total value (Paraïso *et al.*, 2010). If the net margin is positive, production is economically viable; if negative, it's not. High total costs and low production often lead to negative margins, mainly due to high variable costs or significant fixed costs in high-investment situations (Paraïso *et al.*, 2010).

Financial profitability indicator

The Profit-Cost ratio is the sole financial profitability indicator studied (Paraïso *et al.*, 2010). It measures monetary benefits from investments in units like FCFA. Let B be the total benefits after investment T, and Rf is the financial profitability indicator. According to Darbelet and Laugine (1990): Rf=B/CT. In agricultural economics, B is the total product value, and CT is the total costs, including family labor. If PBV is gross product in monetary terms, CP is total cost excluding family labor, and MOV is total family labor value, then: B/CT=PBV/(MOV+CP). If B/CT>1, each franc invested generates more than 1 FCFA in profit, making the activity financially profitable. If B/CT<1, it generates less than 1 FCFA in profit, making the activity financially unprofitable, meaning the producer earns less than invested.

Variables

The literature review found that profitability, economic profit, and financial rates of return are key indicators of a company's performance. These rates assess efficiency by comparing the result obtained with the capital invested (or gross product in monetary terms) and total cost, aiming to maximize added value or return on investment. The noise product or total sales result is used to determine the net margin, which helps identify signs of the calculated rates. The model explores the relationship between the dependent variable, Y, and 14 independent variables, X (Table 1), excluding the cost of renting land due to its minor representation. This relationship predicts the target variable values as a linear combination of parameters, focusing on variables that affect profitability in order to improve producers' net margin.

Data Analysis

Mindful of the analytical objective to identify the factors/variables which positively or negatively influence the annual sales the most, three modeling algorithms were employed – Generalized Least Squares (GLM), Random Forest (RF) and Least Absolute Shrinkage and Selection Operator (LASSO). Before this, both a Correlation heat map and a complementary pair plot were generated.

The Generalized Least Squares (GLS) model is an extension of the Ordinary Least Squares (OLS) regression method, designed to address specific violations of the assumptions underlying OLS. While OLS assumes that the error terms are homoskedastic (having constant variance) and uncorrelated, GLS is employed when these assumptions are violated, particularly in the presence of heteroskedasticity (non-constant variance) or autocorrelation (correlated errors). The GLS estimator is formulated to provide the Best Linear Unbiased Estimator (BLUE) under these conditions, making it a robust alternative when OLS fails to yield reliable estimates (Browne, 1973; Chronopoulos *et al.*, 2023; Jones *et al.*, 2023; Li *et al.*, 2024).

In a typical linear regression framework, the model can be expressed as $Y=X\beta + \varepsilon$, where Y represents the dependent variable, X is the matrix of independent variables, β is the vector of coefficients to be estimated, and β is the vector of errors. In GLS, it is assumed that the variance of the errors can be represented by a known covariance matrix Σ, which captures both the variances and covariances of the error terms. The key idea behind GLS is to transform the original model into one that satisfies the assumptions of OLS by applying a linear transformation that accounts for this covariance structure. This transformation involves pre-multiplying both sides of the regression equation by Σ exponent -0.5, leading to a new model where the transformed errors have constant variance and are uncorrelated (Hu *et al.*, 2009; Roch & Rohe, 2017; Chang & Goplerud, 2023; Lai & Bernstein, 2024).

GLS minimizes a weighted sum of squared residuals, where each residual is weighted inversely by its variance. When Σ is diagonal, indicating uncorrelated errors with different variances, GLS reduces to Weighted Least Squares (WLS), which also minimizes a weighted sum of squared residuals. Importantly, in practice, researchers often do not know Σ and must estimate it from data. This leads to what is known as Feasible Generalized Least Squares (FGLS), where initial estimates from OLS are used to approximate the covariance structure (Maggin *et al.*, 2011; Buhl & Klüppelberg, 2018; Gafarov, 2023; Li & Sonthalia, 2024).

GLS has broad applications across various fields such as econometrics, finance, and environmental studies. It is particularly useful in time series analysis where autocorrelation among residuals is common. By providing efficient estimates even in the presence of correlated or heteroskedastic errors, GLS enhances the reliability of statistical inference in regression models. Overall, while OLS remains a fundamental tool in regression analysis, GLS serves as an essential method for addressing more complex data structures where traditional assumptions do not hold (Lenoir, 2013; Nobari & Gibberd, 2024).

Random Forest Regression is a powerful and widely-used machine learning algorithm that excels in predicting continuous outcomes. It operates by constructing a multitude of decision trees during the training phase and outputting the average prediction from these trees for any new data point. This ensemble approach leverages the concept of "wisdom of the crowds," where the combined predictions of multiple models generally yield better accuracy than any single model alone. The process begins with the selection of random samples from the training dataset. Each sample is used to build a separate decision tree, which is a model that makes predictions based on a series of questions about the input features. For instance, in predicting real estate prices, a decision tree might ask questions regarding the number of bedrooms, location, or square footage to arrive at an estimated price. However, rather than using all available features for every tree, Random Forest introduces randomness by selecting a subset of features for each tree. This feature randomness helps to ensure that the trees are uncorrelated, reducing the risk of overfitting - a common issue in machine learning where models perform well on training data but poorly on unseen data (Louppe, 2015; Gaïffas *et al.*, 2023; Hiabu *et al.*, 2023; Broutin *et al.*, 2024; Curth *et al.*, 2024).

Once all decision trees are constructed, each tree provides its prediction for a given input. In the case of regression tasks, these predictions are averaged to produce the final output. This averaging process smooths out individual errors from the trees and leads to more stable and reliable predictions. The algorithm's robustness is further enhanced by using a technique called bagging

(Bootstrap Aggregating), where each tree is trained on a random subset of data points drawn with replacement from the training set. This means that some data points may be used multiple times while others may not be included at all, adding another layer of diversity among the trees (Probst *et al.*, 2019; Izza & Marques-Silva, 2021; Cattaneo *et al.*, 2024; Kim *et al.*, 2024; Waltz, 2024).

One of the key advantages of Random Forest Regression is its ability to handle large datasets with high dimensionality without significant loss in performance. It can also manage missing values effectively, as individual trees can still make predictions based on available features. Additionally, Random Forest provides insights into feature importance, allowing practitioners to understand which variables contribute most significantly to predictions (Biau & Scornet, 2015; Utkin & Konstantinov, 2022; Boström, 2023; Rhodes *et al.*, 2023; Surve & Pradhan, 2024; Zhang *et al.*, 2024).

Hyperparameters play a crucial role in tuning the performance of Random Forest models. Important hyperparameters include the number of trees in the forest (often referred to as n_estimators), which generally correlates with model accuracy - more trees typically lead to better performance but also increased computational cost. Other hyperparameters control aspects such as the maximum number of features considered for splitting nodes and the minimum samples required to split an internal node (Probst *et al.*, 2019).

LASSO Regression, which stands for Least Absolute Shrinkage and Selection Operator, is a powerful statistical technique used primarily in linear regression analysis. Its main purpose is to enhance the prediction accuracy and interpretability of statistical models by incorporating a regularization method that penalizes the absolute size of the coefficients of the regression model. This approach is particularly beneficial in scenarios where there are many predictors, potentially leading to overfitting, especially when multicollinearity exists among them (Emmert-Streib & Dehmer, 2019; Chintalapudi *et al.*, 2022; Alshqaq & Abuzaid, 2023; Li *et al.*, 2023; Shah *et al.*, 2023; Bir-Jmel *et al.*, 2024; Chen & Wang, 2024).

The core principle behind Lasso Regression is its ability to perform variable selection and shrinkage simultaneously. By applying a penalty to the regression coefficients, Lasso encourages simpler models that contain fewer predictors. This is achieved by shrinking some coefficients towards zero, effectively eliminating less significant variables from the model. As a result, Lasso Regression not only helps in reducing the complexity of the model but also improves its generalization capabilities when applied to new data (Lee *et al.*, 2014; Li *et al.*, 2020; McEligot *et al.*, 2020; Liu *et al.*, 2023a; Mokrišová & Horváthová, 2023; Wang *et al.*, 2023; Christakis *et al.*, 2024; Zhang & Drissi-Habti, 2024).

One of the key features of LASSO Regression is its use of L1 regularization. This technique adds a penalty term to the loss function that is proportional to the absolute values of the coefficients. The strength of this penalty is controlled by a hyperparameter known as lambda (λ). A higher value of λ increases the penalty applied to the coefficients, leading to more coefficients being pushed to zero. Conversely, a lower value of λ reduces the penalty, allowing more variables to remain in the model. This trade-off between bias and variance is crucial; as λ increases, bias tends to increase while variance decreases, resulting in a simpler model with potentially fewer parameters (Reid *et al.*, 2014; Lloyd-Jones *et al.*, 2016; Freijeiro-González *et al.*, 2020; Lee *et al.*, 2021; Yang *et al.*, 2023; Zhao & Huo, 2023; Cheng & Wong, 2024; Mei & Shi, 2024).

LASSO Regression is particularly advantageous in high-dimensional datasets where traditional linear regression may struggle due to overfitting. By effectively reducing the number of predictors through automatic feature selection, Lasso enhances model interpretability and can lead to better predictive performance. This makes it a popular choice in various fields such as machine learning, bioinformatics, and economics, where understanding which variables are most influential can be as important as making accurate predictions (Fontanarosa & Dai, 2011; Meng *et al.*, 2021; Lupton-Smith *et al.*, 2022; Ng & Newton, 2022; Andriopoulos & Kornaros, 2023; Khanna *et al.*, 2023; Wong *et al.*, 2023; Hong *et al.*, 2024).

Practically implementing LASSO Regression involves several steps: preprocessing data to standardize features, fitting the model using training data while iteratively adjusting coefficients based on minimizing errors with respect to both prediction accuracy and regularization penalties, and finally evaluating model performance using metrics like Mean Squared Error (MSE) or R-squared. Cross-validation techniques are often employed to select an optimal value for λ that balances model complexity with predictive accuracy (Cui & Gong, 2018; Rajaratnam *et al.*, 2019; van Egmond *et al.*, 2021; Li *et al.*, 2022; Pourhamidi & Moslemi, 2023; van der Wurp & Groll, 2023; Hsu *et al.*, 2024; Wyss *et al.*, 2024).

RESULTS AND DISCUSSION

Socio-demographic Characteristics of Respondents

Age

The majority of growers in the study are very young, with over half aged between 23 and 42; the minimum age was 23, the maximum was 63, and the average is 39. Cameroon's population is predominantly young, with nearly 50% under 18 and about 64% under 25. This aligns with findings by Silue *et al*. (2019) that most growers are between 25 and 55 years old, and those by Tiama *et al*. (2018). This age range indicates active and mature yam growers and shows young people's interest in yam growing in two rural communes, unlike in Passoré province, where over 82% of growers are over 50 years old (Tiama *et al*. 2018).

Gender

The majority of wheat growers surveyed were men (61.67%), compared with women (38.33%).

Marital status

The majority of producers were married, most of them monogamous, i.e. (46%) versus (26%) polygamous, followed by 13.33% cohabiting, (7.67%) divorced and the remainder (4.33%) widowed and (2.66%) single.

Level of education

The literature highlights education as a key determinant of the standard of living, affecting female reproductive behavior through mechanisms such as delayed age at first birth, higher infertility, and fewer offspring (McDonald, 2000; Esping-Andersen, 2009; Bhrolcháin & Beaujouan, 2012; Peri-Rotem, 2020). Becker's (1991) theory suggests that as women's potential income increases, the demand for children decreases due to higher opportunity costs. In the survey, 28.86% attended high school, 21.81% primary school, 21.48% university, and 20.13% secondary school. Only 7.718% were illiterate, with illiteracy being a notable issue in Adamawa, which has a school enrolment rate of 2%.

Years of experience

The average experience of wheat producers was 10 years. The minimum experience was 2 years and the maximum was 15 years.

Group membership

The majority of the respondents belonged to a farmers' association or cooperatives, which made it easier to obtain financing, agricultural inputs and training, inter alia.

Characteristics of the Different Farms Surveyed

Cultivated area

Despite insecurity reducing agricultural production, Northwest producers sowed 296 ha, yielding 1,087.7 tons, with a yield range of 2.5-5 t/ha and experience spanning 2-13 years. In the West, 245 ha were sown, producing 869.9 tons, with a yield range of 1.2-5.2 t/ha and experience between 2-5 years. Adamawa, growing wheat for the first time, sowed 201 ha, yielding 532.4 tons, with a yield range of 1.2-3.5 t/ha.

Method of land acquisition

Inheritance or the family field is the dominant mode, with 149 producers using or inheriting land. Other modes include buying (52), borrowing or lending with no compensation (39), renting with financial compensation (29), receiving plots as gifts/shares (18), and sharecropping with payment in kind (12).

Family workforce

Production units have an average of 7 people, with family dependents making up 70% and farm labor 30%. The average number of farm workers is 3. Temporary hired labor is used for various activities, especially ploughing, and workers are paid based on services rendered. Prices vary by locality.

Source of income for wheat production

The majority of respondents (64%) use their own funds or personal savings to grow wheat. To finance the rest of the activity, 18% take loans from tontine groups, often interest-free or with low interest. Another 10% seek agricultural loans from banking institutions at low interest rates, while 8% use help from family or friends.

Intermediate consumption

Various equipment, including hoes, rakes, machetes, sprayers, shovels, picks, and others, were used to grow the wheat. Fertilizers included organic types such as cow dung, chicken droppings, and compost, as well as chemical types like NPK 20:10:10 and urea. These were used by 95% of growers and applied at least three times during the growing period. All 300 growers had received training and good-quality seeds for free. MINADER, IRAD, and cooperatives provided the seeds. Phytosanitary products like herbicides, insecticides, and pesticides were used and applied on average twice during the campaign period.

Correlation Analysis

The correlation analysis (Figure 5) revealed the following – There was no correlation between the number of dependents (x7) and the total cost of transport (x1). There was a weak negative correlation between the total cost of fertilizers (x14) and age (x6) (-0.06), the number of dependents (x7) (-0.11) and the number of farm workers (x8) (-0.03). There was also a weak negative correlation between the amount paid for labor in FCFA (x13) and the number of dependents (x7) (-0.04) and the number of farm workers (x8) (-0.01). Furthermore, there was a weak negative correlation between the number of dependents (x7) and the annual sales in FCFA (y) (-0.01), the total cost of equipment (x2) (-0.02), the total cost of seeds (x3) (-0.03), the total cost of Phyto (x4) (-0.01), packaging cost (x5) (-0.01) and the production in tons (x10) (-0.01). There was a weak positive correlation between the total cost of fertilizers (x14) and the annual sales in FCFA (y) (0.21), the total cost of transport (x1) (0.16), the total cost of equipment (x2) (0.27), the total cost of seeds (x3) (0.29), the total cost of Phyto (x4) (0.41) and packaging cost (x5) (0.25).

There was a strong positive correlation between the annual sales in FCFA (y) and the yield in tons per hectare (x12) (0.91), the amount paid for labor in FCFA (x13) (0.92), the production surfacer area in hectares (x9) (0.91), the total cost of transport (x1) (0.99), the total cost of equipment (x2) (0.92), the total cost of seeds (x3) (0.91) and the total cost of Phyto (x4) (0.79). There was also a strong positive correlation between the production in tons (x10) and the total cost of transport (x1) (0.99), the total cost of equipment (x2) (0.92), the total cost of seeds (x3) (0.91),

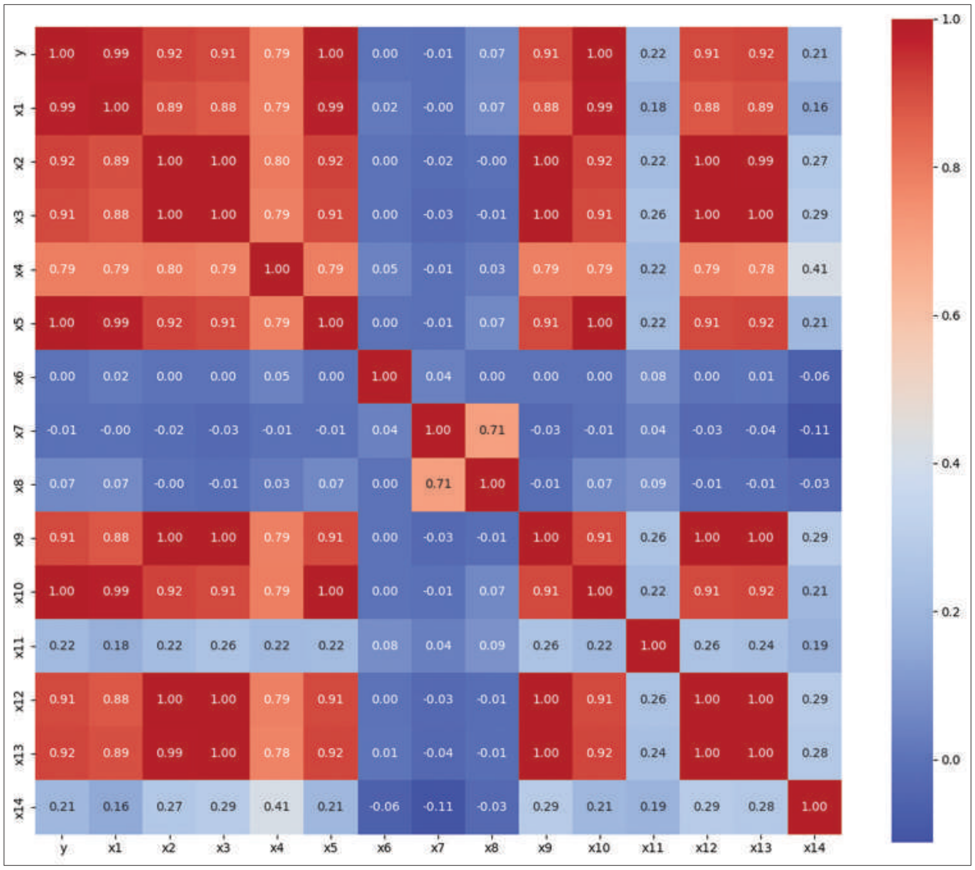

Figure 5: Correlation analysis

the yield in tons per hectare (x12) (0.91) and the amount paid for labor in FCFA (x13) (0.92). Furthermore, there was a strong positive correlation between the yield in tons per hectare (x12) and the total cost of transport (x1) (0.88), the total cost of Phyto (x4) (0.79) and packaging cost (x5) (0.91). Additionally, there was a strong positive correlation between the amount paid for labor in FCFA (x13) and the total cost of equipment (x2) (0.99), the total cost of Phyto (x4) (0.78) and packaging cost (x5) (0.92).

Finally, there was a perfect correlation between the production surfacer area in hectares (x9) and the total cost of equipment (x2), the total cost of seeds (x3), the yield in tons per hectare (x12) and the amount paid for labor in FCFA (x13). There was also a perfect correlation between the yield in tons per hectare (x12) and both the total cost of equipment (x2) and the total cost of seeds (x3). Additionally, there was a perfect correlation between the following two sets of variables - the amount paid for labor in FCFA (x13) and both the total cost of seeds (x3) and the yield in tons per hectare (x12); the annual sales in FCFA (y) and both the packaging cost (x5) and the production in tons (x10).

Generalized Least Squares Modeling Results

The Generalized Least Squares Equation is presented below:

Annual Sales (Turnover) in FCFA = 13,140 – 1.1887 Total Cost of Transport + 0.9785 Total Cost of Equipment – 1.0878 Total Cost of Seeds + 0.5558 Total Cost of Phyto + 2.2132 Packaging Cost + 237.6065 Age – 426.4714 Number of Dependents + 212.1882 Number of Farm Workers – 0.0000136 Production Surface Area in Hectares + 496,800 Production in Tons – 170.8518 Years of Experience – 0.5235 Yield in Tons Per Hectare + 0.1814 Amount Paid for Labor in FCFA – 0.0594 Total Cost of Fertilizers.

Figure 6 visualizes the model coefficients as summarized in Table 2. The top three positively influential features were the Production in Tons (x10) (p-value=0.000, t-value=8.520, lower-upper limits= [3.82e+05, 6.12e+05]), Age (x6) (p-value=0.339, t-value=0.958, lower-upper limits= [-250.667, 725.888]) and the Number of Farm Workers (x8) (p-value=0.923, t-value=0.096, lower-upper limits= [-4128.687, 4553.064]). Conversely, the top three negatively influential features were the Number of dependents (x7) (p-value=0.711, t-value=-0.371, lower-upper limits= [-2688.111, 1835.168]), the Years of Experience (x11) (p-value=0.893, t-value=-0.135, lower-upper limits= [-2662.235, 2320.532]) and the Total Cost of Transport (x1) (p-value=0.000, t-value=-3.695, lower-upper limits= [-1.822, -0.556]).

Figure 7 visualizes the relationship between the residuals and the fitted values. The fitted values (predicted values from the model) lie on the X-axis and the residuals (differences between observed and predicted values) on the Y-axis. The plot helps in diagnosing the fit of the GLS model and validating its assumptions. One key aspect to consider is homoscedasticity,

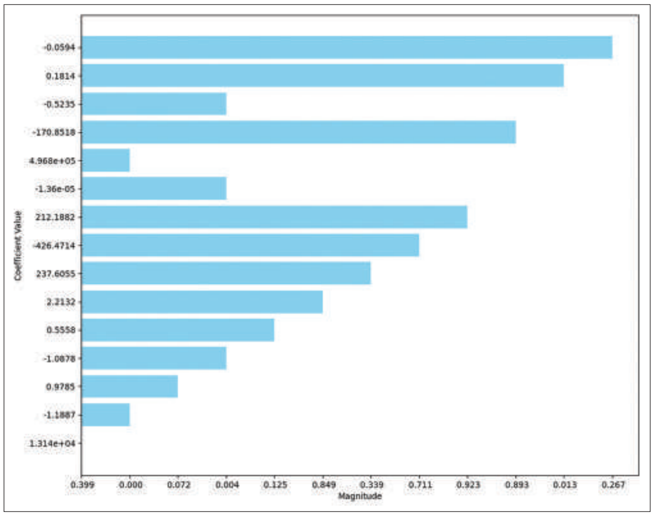

Figure 6: Generalized Least Squares Coefficient Plot

which means that the variance of the residuals is constant across all levels of the fitted values. If the residuals are randomly scattered around the horizontal line at zero without any clear pattern, it suggests homoscedasticity, indicating that the model's errors are evenly distributed. In Figure 7, the residuals appear to be fairly evenly distributed around zero, suggesting that the assumption of homoscedasticity is reasonably met, although the presence of a few extreme values might indicate some heteroscedasticity. Another important aspect is linearity, which means that the relationship between the predictors and the response variable is linear. The residuals should not show any systematic pattern. If there is a pattern, it suggests that the model is not capturing some aspect of the data. In Figure 4, there does not appear to be a clear pattern, which suggests that the linearity assumption is likely satisfied. Outliers are data points that are significantly different from the rest of the data. Points that are far from the zero line are potential outliers. These points can have a large influence on the model and may indicate areas where the model does not fit well. In Figure 7, there are a few points that are significantly far from the zero line, which may warrant further investigation. Independence means that the residuals are not correlated with each other. Figure 7 does not directly show independence, but if there were a pattern, it might suggest a lack of independence. The plot does not show any obvious patterns that would suggest a lack of independence. Based on the foregoing, Figure 7 suggests that the GLS model fits the data reasonably well, but there are a few potential outliers that may need further investigation.

Random Forest Modeling Results

Figure 8 visualizes the random Forest Feature Importance Scores summarized in Table 3. The top three positively influential features were the Production in Tons (x10) (Importance Score= 0.452244578698737, Permutation Importance= 0.0752147990577783), Packaging Cost (x5) (Importance Score= 0.365316105261279, Permutation Importance= 8.99534840538655E-06) and Total Cost of Transport (x1) (Importance Score= 0.18182466378679, Permutation Importance= 7.95533374117937E-06). Conversely, the top three negatively influential features were Yield in Tons Per

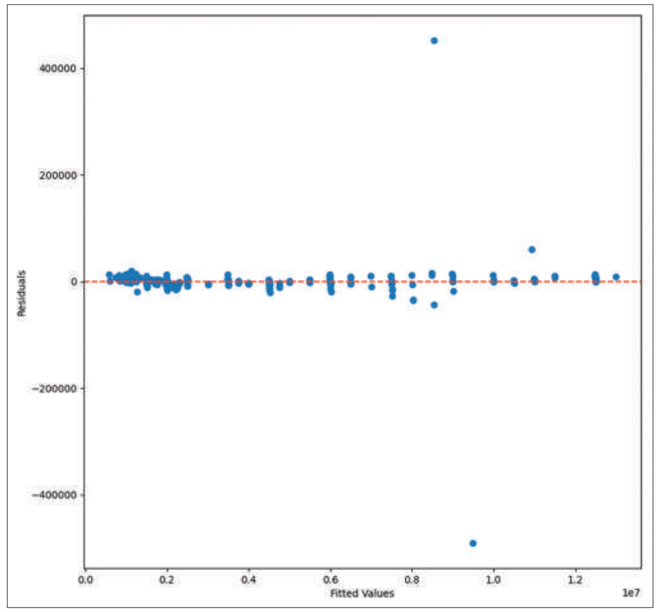

Figure 7: Generalized Least Squares Residuals versus Fitted Values visualization

Table 1: Variables used in this study and their respective codes

Variable	Code
Annual sales (turnover) in FCFA	y
Total cost of transport	x1
Total cost of each equipment used	x2
Total cost of seeds	x3
Total cost of phyto	x4
Packaging	x5
Age	x6
Number of dependents	x7
Number of farm workers	x8
Production surface area in hectares	x9
Production in tons	x10
Years of experience	x11
Yield in tons per hectare	x12
Amount paid for labor in FCFA	x13
Total cost of fertilizers	x14

Table 2: Generalized least squares model summary

	Coef	Std err	t	P>|t|	[0.025	0.975]
const	1.314e+04	1.56e+04	0.844	0.399	-1.75e+04	4.38e+04
x1	-1.1887	0.322	-3.695	0.000	-1.822	-0.556
x2	0.9785	0.542	1.807	0.072	-0.088	2.045
x3	-1.0878	0.375	-2.899	0.004	-1.826	-0.349
x4	0.5558	0.361	1.539	0.125	-0.155	1.267
x5	2.2132	11.642	0.190	0.849	-20.702	25.129
x6	237.6055	248.078	0.958	0.339	-250.677	725.888
x7	-426.4714	1149.053	-0.371	0.711	-2688.111	1835.168
x8	212.1882	2205.433	0.096	0.923	-4128.687	4553.064
x9	-1.36e-05	4.69e-06	-2.899	0.004	-2.28e-05	-4.36e-06
x10	4.968e+05	5.83e+04	8.520	0.000	3.82e+05	6.12e+05
x11	-170.8518	1265.777	-0.135	0.893	-2662.235	2320.532
x12	-0.5235	0.181	-2.899	0.004	-0.879	-0.168
x13	0.1814	0.072	2.511	0.013	0.039	0.324
x14	-0.0594	0.053	-1.111	0.267	-0.165	0.046

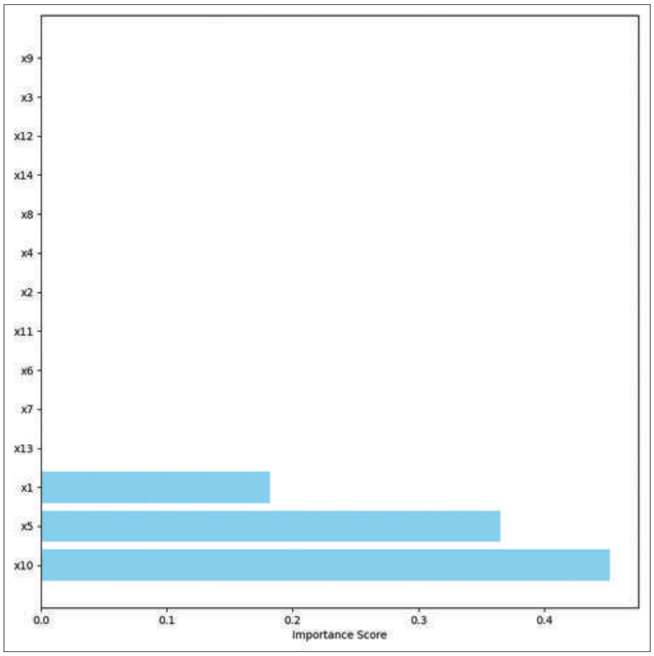

Figure 8: Horizontal Bar plot of Random Forest Feature Importance Scores

Hectare (x12) (Importance Score= 0.0000137617736842951, Permutation Importance= -8.14024295892798E-06), Total Cost of Seeds (x3) (Importance Score= 0.0000111250675734569, Permutation Importance= 0.0002515091615465) and Production Surface Area in Hectares (x9) (Importance Score=8.46470176599585E-06, Permutation Importance=-0.0000330621371149105).

Figure 9 presents the Partial Dependence Plots for all of the predictors. Partial Dependence Plots (PDPs) are a powerful visualization tool used in machine learning, particularly with models like Random Forests. They help illustrate the relationship between one or two features and the predicted outcome of a model while controlling for the effects of all other features. By averaging the predictions over the distribution of the other features, PDPs provide insights into how changes in a specific feature influence the model's predictions. When creating a PDP, the process involves selecting a feature of interest and varying its values while keeping all other features constant. For each value of the selected feature, the model predicts outcomes based on the training data. These predictions are then averaged to create a smooth curve that represents the expected prediction for each value of the feature. This allows analysts to see not only the strength of the relationship but also its shape - whether it is linear, monotonic, or exhibits more complex behaviors. The utility of PDPs extends beyond mere visualization; they facilitate a deeper understanding of model behavior. For instance, they can reveal whether an increase in a particular feature leads to an increase or decrease in predicted outcomes, thereby indicating the direction and nature of the relationship. This is particularly valuable in complex models like Random Forests, which can behave like "black boxes." By using PDPs, practitioners can gain insights into which features are most influential and how they interact with each other. However, it's important to note that while PDPs provide significant insights, they are primarily effective for low-dimensional analyses - typically involving one or two features at a time. In cases with many features, interpreting interactions can become challenging. Despite this limitation, PDPs remain an essential tool for interpreting machine learning models and enhancing their transparency. The PDPs for variables x10, x5 and x1 all show a characteristically steady increase in model partial dependence with an increase in variable magnitude. These PDPs thus serve as explainability tools to justify their importance.

Table 3: Random forest feature importance and permutation importance scores

Feature	Importance	Permutation Importance
x10	0.452244579	0.075214799
x5	0.365316105	8.99535E-06
x1	0.181824664	7.95533E-06
x13	0.000298511	1.29643E-05
x7	8.5347E-05	0.287477617
x6	7.59776E-05	1.18507E-05
x11	3.347E-05	-0.000257773
x2	3.01683E-05	-3.85188E-06
x4	2.42163E-05	-4.34115E-06
x8	1.8346E-05	0.393264871
x14	1.5265E-05	-8.99175E-06
x12	1.37618E-05	-8.14024E-06
x3	1.11251E-05	0.000251509
x9	8.4647E-06	-3.30621E-05

Table 4: LASSO feature important scores and feature P-values

Feature	Importance	P-value
x1	7423.599283	-3.35311E+11
x2	0	-1.90684E+12
x3	0	-2.14622E+12
x4	0	-4.54152E+12
x5	3295656.619	-1587686402
x6	402.4447588	-1.17117E+13
x7	0	-1.1649E+13
x8	0	-1.16711E+13
x9	0	-2.14622E+12
x10	77254.2683	-1060870369
x11	0	-1.09122E+13
x12	0	-2.14622E+12
x13	4390.959043	-1.90833E+12
x14	-1237.996281	-1.11949E+13

Table 5: Summary of correlation heatmap results and these three models

Models	Top 3 positively influential features	Top 3 negatively influential features
Correlation Heatmap	x10, x5, x1	x7, x6, x8
Generalized Least Squares	x10, x11, x1	x7, x11, x1
Random Forest	x10, x5, x1	x12, x14, x3
LASSO	x10, x5, x13	x14, x11, x7

LASSO Modeling Results

Table 4, visualized in Figure 10, summarizes the feature importance scores. The top three positively influential features were the Packaging Cost (x5) (Importance Score=3295656.619),

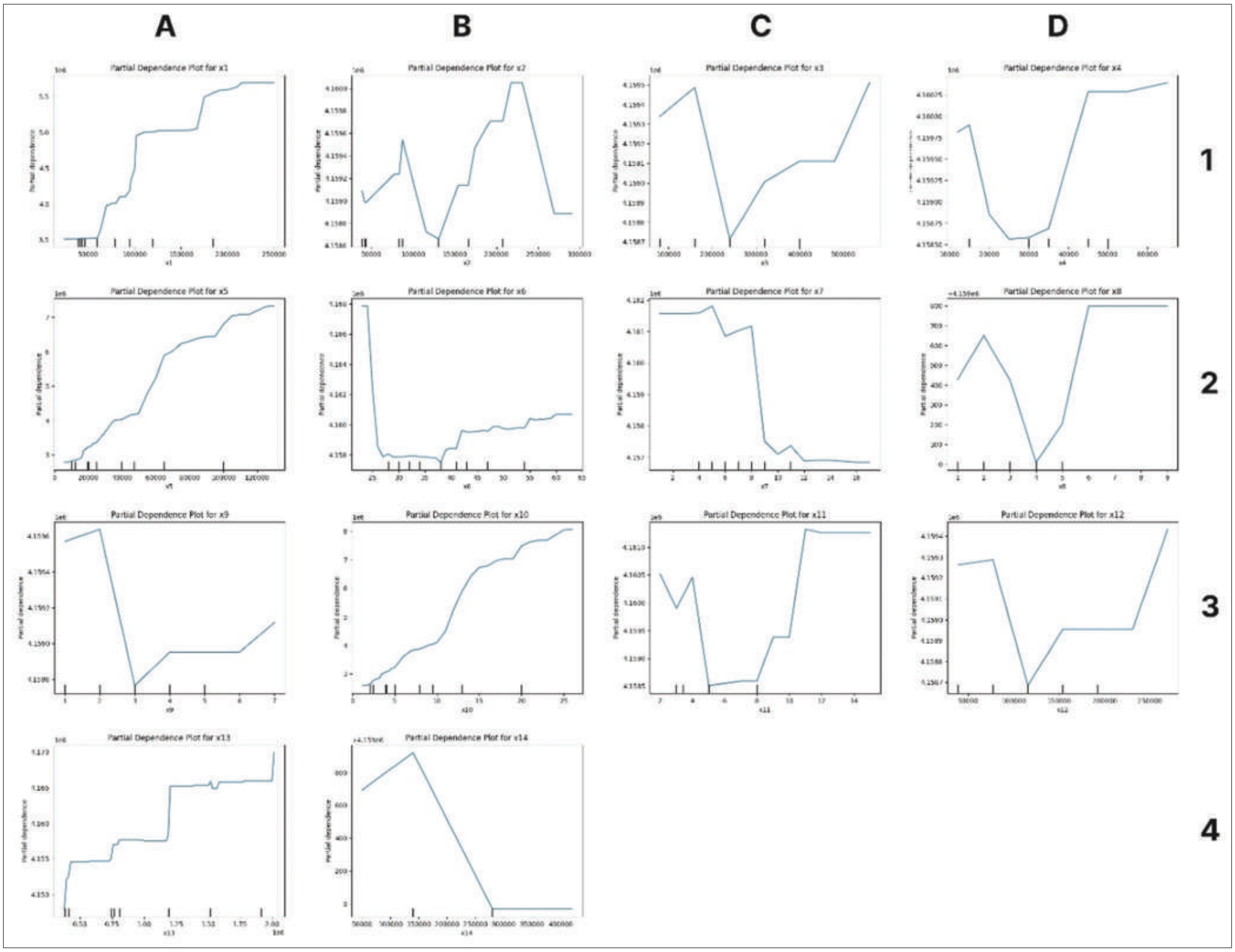

Figure 9: Random Forest Partial Dependence Plots. A1=x1; B1=x2; C1=x3: D1=x4; A2=x5; B2=x6; C2=x7; D2=x8; A3=x9; B3=x10; C3=x11; D3=x12; A4=x13; B4=x14

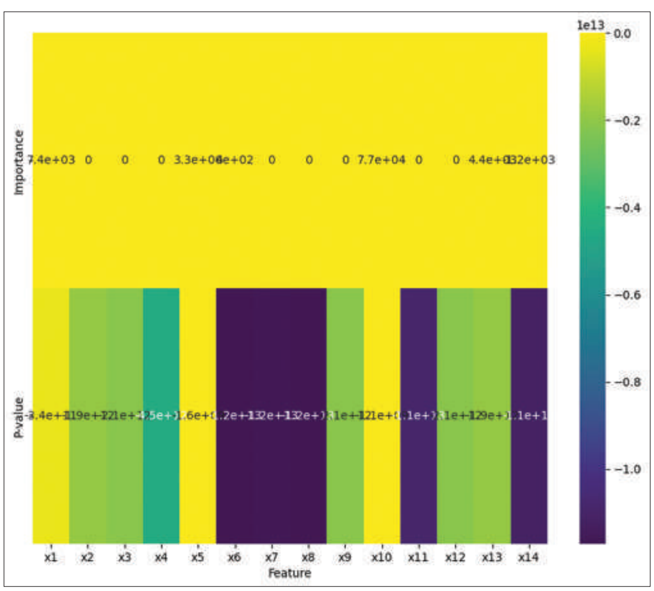

Figure 10: Visualization of the LASSO Feature Importances and P-values

the Production in Tons (x10) (Importance Score=77254.2683) and the Amount Paid for Labor in FCFA (x13) (Importance Score=4390.959043). The only negatively influential feature was the Total Cost of Fertilizers (x14) (Importance Score=-1237.996281).

Key Summary

According to the 3 models, gross income is positively impacted by production in tons, packaging costs and transport costs. On the other hand, the number of people in charge, the number of years of experience and the cost of fertilizers has a negative impact on the result (Table 5). The net margin is positive for all wheat growers. The Profit/Cost ratio is below 1, meaning that wheat growers with a Profit/Cost ratio of 0.160 earn less than they invest.

CONCLUSION

A survey in three regions (Adamawa, North-West, West) among 300 wheat growers identified variables affecting profitability using three modeling algorithms: generalized least squares, random forest, and least absolute shrinkage and selection operator. Positive influences on net margin include production in tons, packaging costs, and transport costs. Negative influences include the number of people in charge, years of experience, and fertilizer costs. The net margin for wheat production in Cameroon is positive at 76,691,000 FCFA, but the crop is economically profitable yet financially unprofitable, with a ratio of 0.160. Growers receive seeds, fertilizers, and technical assistance from government-endowed organizations, but state-facilitated sales remain a significant obstacle.

REFERENCES

Alshqaq, S. S., & Abuzaid, A. H. (2023). An Efficient Method for Variable Selection Based on Diagnostic-Lasso Regression. *Symmetry, 15*(12), 2155. https://doi.org/10.3390/sym15122155

Andriopoulos, V., & Kornaros, M. (2023). LASSO Regression with Multiple Imputations for the Selection of Key Variables Affecting the Fatty Acid Profile of *Nannochloropsis oculata*. *Marine Drugs, 21*(9), 483. https://doi.org/10.3390/md21090483

Becker, G. S. (1991). *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bhrolcháin, M. N., & Beaujouan, É. (2012). Fertility Postponement is Largely due to rising Educational Enrolment. *Population Studies, 66*(3), 311-327. https://doi.org/10.1080/00324728.2012.697569

Biau, G., & Scornet, E. (2015). A Random Forest Guided Tour. *arXiv*, arXiv:1511.05741. https://doi.org/10.48550/arXiv.1511.05741

Bir-Jmel, A., Douiri, S. M., Bernoussi, S. E., Maafiri, A., Himeur, Y., Atalla, S., Mansoor, W., & Al-Ahmad, H. (2024). GFLASSO-LR: Logistic Regression with Generalized Fused LASSO for Gene Selection in High-Dimensional Cancer Classification. *Computers, 13*(4), 93. https://doi.org/10.3390/computers13040093

Boström, H. (2023). Example-Based Explanations of Random Forest Predictions. *arXiv*, arXiv:2311.14581. https://doi.org/10.48550/arXiv.2311.14581

Broutin, N., Devroye, L., Lugosi, G., & Oliveira, R. I. (2024). Subtractive random forests. *arXiv*, arXiv:2210.10544. https://doi.org/10.48550/arXiv.2210.10544

Browne, M. W. (1973). Generalized Least Squares Estimators in the Analysis of Covariance Structures. *ETS Research Bulletin Series, 1973*(1), i-36. https://doi.org/10.1002/j.2333-8504.1973.tb00197.x

Buhl, S., & Klüppelberg, C. (2018). Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes. *arXiv*, arXiv:1704.05656. https://doi.org/10.48550/arXiv.1704.05656

Cattaneo, M. D., Klusowski, J. M., & Underwood, W. G. (2024). Inference with Mondrian Random Forests. *arXiv*, arXiv:2310.09702. https://doi.org/10.48550/arXiv.2310.09702

Chang, Q., & Goplerud, M. (2023). Generalized Kernel Regularized Least Squares. *arXiv*, arXiv:2209.14355. https://doi.org/10.48550/arXiv.2209.14355

Chen, L., & Wang, D. (2024). Cost Estimation and Prediction for Residential Projects Based on Grey Relational Analysis–Lasso Regression–Backpropagation Neural Network. *Information, 15*(8), 502. https://doi.org/10.3390/info15080502

Cheng, S.-W., & Wong, M. T. (2024). A Fast Method for Lasso and Logistic Lasso. *arXiv*, arXiv:2402.02463. https://doi.org/10.48550/arXiv.2402.02463

Chintalapudi, N., Angeloni, U., Battineni, G., di Canio, M., Marotta, C., Rezza, G., Sagaro, G. G., Silenzi, A., & Amenta, F. (2022). LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining. *Bioengineering, 9*(3), 124. https://doi.org/10.3390/bioengineering9030124

Christakis, I., Sarri, E., Tsakiridis, O., & Stavrakas, I. (2024). Investigation of LASSO Regression Method as a Correction Measurements' Factor for Low-Cost Air Quality Sensors. *Signals, 5*(1), 60-86. https://doi.org/10.3390/signals5010004

Chronopoulos, I., Chrysikou, K., & Kapetanios, G. (2023). High Dimensional Generalised Penalised Least Squares. *arXiv*, arXiv:2207.07055. https://doi.org/10.48550/arXiv.2207.07055

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage, 178*, 622-637. https://doi.org/10.1016/j.neuroimage.2018.06.001

Curth, A., Jeffares, A., & van der Schaar, M. (2024). Why do Random Forests Work? Understanding Tree Ensembles as Self-Regularizing Adaptive Smoothers. *arXiv*, arXiv:2402.01502. https://doi.org/10.48550/arXiv.2402.01502

Darbelet, M., & Laugine, J. M. (1990). Economie d'entreprise. *Foucher*, 374-375.

Emmert-Streib, F., & Dehmer, M. (2019). High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection. *Machine Learning and Knowledge Extraction, 1*(1), 359-383. https://doi.org/10.3390/make1010021

Esping-Andersen, G. (2009). *The Incomplete Revolution: Adapting to Women's New Roles*. Cambridge, UK: Polity Press.

FAO. (2011). L'État de l'insécurité alimentaire dans le monde 2011. Quelles sont les conséquences de l'instabilité des cours internationaux pour l'économie et la sécurité alimentaire des pays ? Rome.

FAO. (2016). Produire plus avec moins en pratique le maïs, le riz et le blé. ORGANISATION DES NATIONS UNIES POUR L'ALIMENTATION ET L'AGRICULTURE. Rome.

Fontanarosa, J. B., & Dai, Y. (2011). Using LASSO regression to detect predictive aggregate effects in genetic studies. *BMC Proceedings, 5*, S69. https://doi.org/10.1186/1753-6561-5-S9-S69

Freijeiro-González, L., Febrero-Bande, M., & González-Manteiga, W. (2020). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. *arXiv*, arXiv:2012.11470. https://doi.org/10.48550/arXiv.2012.11470

Gafarov, B. (2023). Generalized Automatic Least Squares: Efficiency Gains from Misspecified Heteroscedasticity Models. *arXiv*, arXiv:2304.07331. https://doi.org/10.48550/arXiv.2304.07331

Gaïffas, S., Merad, I., & Yu, Y. (2023). WildWood: A new Random Forest algorithm. *arXiv*, arXiv:2109.08010. https://doi.org/10.48550/arXiv.2109.08010

GIZ/PARSE-AD. (2017). Etude sur les métiers rentables: Projet d'Appui à la Résilience Socio-économique des jeunes vulnérables (Extrême-Nord, Nord et Adamaoua). Ministère Fédéral de la Coopération économique et du Développement (BMZ), 2-100

Hiabu, M., Mammen, E., & Meyer, J. T. (2023). Random Planted Forest: A directly interpretable tree ensemble. *arXiv*, arXiv:2012.14563. https://doi.org/10.48550/arXiv.2012.14563

Hong, C., Xiong, Y., Xia, J., Huang, W., Xia, A., Xu, S., Chen, Y., Xu, Z., Chen, H., & Zhang, Z. (2024). LASSO-Based Identification of Risk Factors and Development of a Prediction Model for Sepsis Patients. *Therapeutics and Clinical Risk Management, 20*, 47-58. https://doi.org/10.2147/TCRM.S434397

Hsu, C.-L., Wu, P.-C., Wu, F.-Z., & Yu, H.-C. (2024). LASSO-derived model for the prediction of lean-non-alcoholic fatty liver disease in examinees attending a routine health check-up. *Annals of Medicine, 56*(1), 2317348. https://doi.org/10.1080/07853890.2024.2317348

Hu, X. J., Lagakos, S. W., & Lockhart, R. A. (2009). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statistica Sinica, 19*, 561-580.

Izza, Y., & Marques-Silva, J. (2021). On Explaining Random Forests with SAT. *arXiv*, arXiv:2105.10278. https://doi.org/10.48550/arXiv.2105.10278

Jones, A. M., Bosler, P. A., Kuberry, P. A., & Wright, G. B. (2023). Generalized moving least squares vs. Radial basis function finite difference methods for approximating surface derivatives. *arXiv*, arXiv:2309.04035. https://doi.org/10.48550/arXiv.2309.04035

Khanna, A., Lu, F., Raff, E., & Testa, B. (2023). Sparse Private LASSO Logistic Regression. *arXiv*, arXiv:2304.12429. https://doi.org/10.48550/arXiv.2304.12429

Kim, Y., Kim, S. Y., & Kim, H. (2024). Heterogeneous Random Forest. *arXiv*, arXiv:2410.19022. https://doi.org/10.48550/arXiv.2410.19022

Lai, B., & Bernstein, D. S. (2024). Generalized Forgetting Recursive Least Squares: Stability and Robustness Guarantees. *arXiv*, arXiv:2308.04259. https://doi.org/10.48550/arXiv.2308.04259

Batsch, L. (2003). Rentabilité Économique, Linéarité de l'investissement et Retour Sur Dépenses Cumulées: Quelques Problèmes de Mesure. Cereg, Université Paris-Dauphine Mai 2003.

Lee, J. H., Shi, Z., & Gao, Z. (2021). On LASSO for Predictive Regression. *arXiv*, arXiv:1810.03140. https://doi.org/10.48550/arXiv.1810.03140

Lee, T.-F., Chao, P.-J., Ting, H.-M., Chang, L., Huang, Y.-J., Wu, J.-M., Wang, H.-Y., Horng, M.-F., Chang, C.-M., Lan, J.-H., Huang, Y.-Y., Fang, F.-M., & Leung, S. W. (2014). Using Multivariate Regression Model with Least Absolute Shrinkage and Selection Operator (LASSO) to Predict the Incidence of Xerostomia after Intensity-Modulated Radiotherapy for Head and Neck Cancer. *Plos One, 9*(2), e89700. https://doi.org/10.1371/journal.pone.0089700

Lenoir, B. (2013). A general approach of least squares estimation and optimal filtering. *arXiv*, arXiv:1305.6324. https://doi.org/10.48550/arXiv.1305.6324

Li, J., & Sonthalia, R. (2024). Generalization for Least Squares Regression With Simple Spiked Covariances. *arXiv*, arXiv:2410.13991. https://doi.org/10.48550/arXiv.2410.13991

Li, J., Pollinger, F., & Paeth, H. (2020). Comparing the Lasso Predictor-Selection and Regression Method with Classical Approaches of Precipitation Bias Adjustment in Decadal Climate Predictions. *Monthly Weather Review, 148*(10), 4339-4351. https://doi.org/10.1175/MWR-D-19-0302.1

Li, R., Yan, Q., & Jiang, S. W. (2024). Generalized Moving Least-Squares for Solving Vector-valued PDEs on Unknown Manifolds. *arXiv*, arXiv:2406.12210. https://doi.org/10.48550/arXiv.2406.12210

Li, Y., Lu, F., & Yin, Y. (2022). Applying logistic LASSO regression for the diagnosis of atypical Crohn's disease. *Scientific Reports, 12*, 11340. https://doi.org/10.1038/s41598-022-15609-5

Li, Y., Yang, R., Wang, X., Zhu, J., & Song, N. (2023). Carbon Price Combination Forecasting Model Based on Lasso Regression and Optimal Integration. *Sustainability, 15*(12), 9354. https://doi.org/10.3390/su15129354

Liu, L., Gao, J., Beasley, G., & Jung, S.-H. (2023a). LASSO and Elastic Net Tend to Over-Select Features. *Mathematics, 11*(17), 3738. https://doi.org/10.3390/math11173738

Liu, M., Li, Q., Zhang, J., & Chen, Y. (2023b). Development and Validation of a Predictive Model Based on LASSO Regression: Predicting the Risk of Early Recurrence of Atrial Fibrillation after Radiofrequency Catheter Ablation. *Diagnostics, 13*(22), 3403. https://doi.org/10.3390/diagnostics13223403

Lloyd-Jones, L. R., Nguyen, H. D., & McLachlan, G. J. (2016). A globally convergent algorithm for lasso-penalized mixture of linear regression models. *arXiv*, arXiv:1603.08326. https://doi.org/10.48550/arXiv.1603.08326

Louppe, G. (2015). Understanding Random Forests: From Theory to Practice. *arXiv*, arXiv:1407.7502. https://doi.org/10.48550/arXiv.1407.7502

Lupton-Smith, C., Stuart, E. A., McGinty, E. E., Dalcin, A. T., Jerome, G. J., Wang, N.-Y., & Daumit, G. L. (2022). Determining Predictors of Weight Loss in a Behavioral Intervention: A Case Study in the Use of Lasso Regression. *Frontiers in Psychiatry, 12*, 707707. https://doi.org/10.3389/fpsyt.2021.707707

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*(3), 301-321. https://doi.org/10.1016/j.jsp.2011.03.004

Makelele, L. (2014). Analyse de la Performance Financière d'une Entreprise Hôtelière: cas de l'Hôtel Mangwe/Uvira. TFC, UCB.

Mcdonald, P. (2000). Gender Equity, Social Institutions and the Future of Fertility. *Journal of the Australian Population Association, 17*, 1-16. https://doi.org/10.1007/BF03029445

McEligot, A. J., Poynor, V., Sharma, R., & Panangadan, A. (2020). Logistic LASSO Regression for Dietary Intakes and Breast Cancer. *Nutrients, 12*(9), 2652. https://doi.org/10.3390/nu12092652

Mei, Z., & Shi, Z. (2024). On LASSO for High Dimensional Predictive Regression. *arXiv*, arXiv:2212.07052. https://doi.org/10.48550/arXiv.2212.07052

Meng, L., Zheng, T., Wang, Y., Li, Z., Xiao, Q., He, J., & Tan, J. (2021). Development of a prediction model based on LASSO regression to evaluate the risk of non-sentinel lymph node metastasis in Chinese breast cancer patients with 1–2 positive sentinel lymph nodes. *Scientific Reports, 11*, 19972. https://doi.org/10.1038/s41598-021-99522-3

MINADER/AFD/IRAM. (2015). Etude de Préfaisabilité d'un Programme de Gestion des Risques Agricoles, Rapport Final, Décembre 2015.

Mokrišová, M., & Horváthová, J. (2023). Domain Knowledge Features versus LASSO Features in Predicting Risk of Corporate Bankruptcy—DEA Approach. *Risks, 11*(11), Article 199. https://doi.org/10.3390/risks11110199

Silue, N. Z., Dao, D., Valerie Kouame, H. K., & Kone, M. (2019). Analyse de la Rentabilité Economique des Systemes de Production a Base d'Igname: cas des Sites de Leo et Midebdo au Burkina Faso. *Institut de Gestion Agropastorale/Université Péléforo Gon Coulibaly Côte d'Ivoire Agronomie Africaine, 31*(1), 1-14.

Ng, T. L., & Newton, M. A. (2022). Random weighting in LASSO regression. *arXiv*, arXiv:2002.02629. https://doi.org/10.48550/arXiv.2002.02629

Nobari, K. S., & Gibberd, A. (2024). l_1-Regularized Generalized Least Squares. *arXiv*, arXiv:2405.10719. https://doi.org/10.48550/arXiv.2405.10719

Paraïso, A., Sossou, A., Yègbémey, R., & Biaou, G. (2010). Analyse de la Rentabilité de la Production du Fonio (Digitaria Exilis) dans la Commune de Boukombe au Bénin. IVème Journée Scientifique de l'Université de Lomé, Togo, 25-28 octobre 2010.

Peri-Rotem. (2020). Écarts de fécondité en fonction du niveau d'instruction: le rôle de la religion en Grande-Bretagne Et En France. *Population, 75*, 9-38.

Pirou, J. P. (2005). Mesure de la Rentabilité des Entreprise. Bulletin de la Banque de France, N° 134 de Février 2005.

Pourhamidi, R., & Moslemi, A. (2023). Using the Precision Lasso for gene selection in diffuse large B cell lymphoma cancer. *Journal of the Egyptian National Cancer Institute, 35*(1), 19. https://doi.org/10.1186/s43046-023-00172-5

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and Tuning Strategies for Random Forest. *arXiv*, arXiv:1804.03515. https://doi.org/10.48550/arXiv.1804.03515

Rajaratnam, B., Roberts, S., Sparks, D., & Yu, H. (2019). Influence Diagnostics for High-Dimensional Lasso Regression. *Journal of Computational and Graphical Statistics, 28*(4), 877-890. https://doi.org/10.1080/10618600.2019.1598869

Reid, S., Tibshirani, R., & Friedman, J. (2014). A Study of Error Variance Estimation in Lasso Regression. *arXiv*, arXiv:1311.5274. https://doi.org/10.48550/arXiv.1311.5274

Rhodes, J. S., Cutler, A., & Moon, K. R. (2023). Geometry- and Accuracy-Preserving Random Forest Proximities. *arXiv*, arXiv:2201.12682. https://doi.org/10.48550/arXiv.2201.12682

Roch, S., & Rohe, K. (2017). Generalized least squares can overcome the critical threshold in respondent-driven sampling. *arXiv*, arXiv:1708.04999. https://doi.org/10.48550/arXiv.1708.04999

Shah, I., Naz, H., Ali, S., Almohaimeed, A., & Lone, S. A. (2023). A New Quantile-Based Approach for LASSO Estimation. *Mathematics, 11*(6), 1452. https://doi.org/10.3390/math11061452

Sion, M. (2020). Ratio de rentabilité économique: quelle est son importance?. Par Michel Sion le 31 mars 2020, Analyse financière.

Surve, T., & Pradhan, R. (2024). Example-based Explanations for Random Forests using Machine Unlearning. *arXiv*, arXiv:2402.05007. https://doi.org/10.48550/arXiv.2402.05007

Tiama, D., Kaboré, B., Nanéma, K. R., Dabiré, M., & Sawadogo, N. (2018). Systèmes de culture et caractérisation paysanne des ignames du Passoré au Burkina Faso. *International Journal of Innovation and Scientific Research, 38*(2), 203-211.

Utkin, L. V., & Konstantinov, A. V. (2022). Attention and Self-Attention in Random Forests. *arXiv*, arXiv:2207.04293. https://doi.org/10.48550/arXiv.2207.04293

van der Wurp, H., & Groll, A. (2023). Introducing LASSO-type penalisation to generalised joint regression modelling for count data. *AStA Advances in Statistical Analysis, 107*, 127-151. https://doi.org/10.1007/s10182-021-00425-5

van Egmond, M. B., Spini, G., van der Galien, O., IJpma, A., Veugen, T., Kraaij, W., Sangers, A., Rooijakkers, T., Langenkamp, P., Kamphorst, B., van de L'Isle, N., & Kooij-Janic, M. (2021). Privacy-preserving dataset combination and Lasso regression for healthcare predictions. *BMC Medical Informatics and Decision Making, 21*, 266. https://doi.org/10.1186/s12911-021-01582-y

Waltz, N. (2024). Grafting: Making Random Forests Consistent. *arXiv*, arXiv:2403.06015. https://doi.org/10.48550/arXiv.2403.06015

Wang, J., Xu, Y., Liu, L., Wu, W., Shen, C., Huang, H., Zhen, Z., Meng, J., Li, C., Qu, Z., he, Q., & Tian, Y. (2023). Comparison of LASSO and random forest models for predicting the risk of premature coronary artery disease. *BMC Medical Informatics and Decision Making, 23*, 297. https://doi.org/10.1186/s12911-023-02407-w

Wong, A., Kramer, S. C., Piccininni, M., Rohmann, J. L., Kurth, T., Escolano, S., Grittner, U., & Domenech de Cellès, M. (2023). Using LASSO

Regression to Estimate the Population-Level Impact of Pneumococcal Conjugate Vaccines. *American Journal of Epidemiology, 192*(7), 1166-1180. https://doi.org/10.1093/aje/kwad061

Wyss, R., van der Laan, M., Gruber, S., Shi, X., Lee, H., Dutcher, S. K., Nelson, J. C., Toh, S., Russo, M., Wang, S. V., Desai, R. J., & Lin, K. J. (2024). Targeted learning with an undersmoothed LASSO propensity score model for large-scale covariate adjustment in health-care database studies. *American Journal of Epidemiology, 193*(11), 1632-1640. https://doi.org/10.1093/aje/kwae023

Yang, S., Yan, Y., Zhu, X., & Sun, Q. (2023). Online Linearized LASSO. *arXiv*, arXiv:2211.06039. https://doi.org/10.48550/arXiv.2211.06039

Zhang, Q., & Drissi-Habti, M. (2024). Electric Cable Insulator Damage Monitoring by Lasso Regression. *Machines, 12*(1), 50. https://doi.org/10.3390/machines12010050

Zhang, Y., Ji, W., & Bradic, J. (2024). Adaptive Split Balancing for Optimal Random Forest. *arXiv*, arXiv:2402.11228. https://doi.org/10.48550/arXiv.2402.11228

Zhao, Y., & Huo, X. (2023). A Survey of Numerical Algorithms that can Solve the Lasso Problems. *arXiv*, arXiv:2303.03576. https://doi.org/10.48550/arXiv.2303.03576