Universal Journal of Agricultural Research 12(2): 429-444, 2024 http://www.hrpub.org
DOI: 10.13189/ujar.2024.120221

On Huber's Robust Technique and Quantile Regression Models for the Total Production of Field Crops in Oman

Sanjith Bharatharajan Nair, Zuhair A. Al-Hemyari*

Department of Mathematical and Physical Sciences, University of Nizwa, Oman
*Corresponding Author: alhemyari@unizwa.edu.om

Received February 21, 2024; Revised April 16, 2024; Accepted April 21, 2024

Cite This Paper in the Following Citation Styles

(a): *[1] Sanjith Bharatharajan Nair, Zuhair A. Al-Hemyari , "On Huber's Robust Technique and Quantile Regression Models for the Total Production of Field Crops in Oman," Universal Journal of Agricultural Research, Vol. 12, No. 2, pp. 429 - 444, 2024. DOI: 10.13189/ujar.2024.120221.*

(b): *Sanjith Bharatharajan Nair, Zuhair A. Al-Hemyari (2024). On Huber's Robust Technique and Quantile Regression Models for the Total Production of Field Crops in Oman. Universal Journal of Agricultural Research, 12(2), 429 - 444. DOI: 10.13189/ujar.2024.120221.*

Copyright©2024 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In this paper, the problem of modelling the data of the agriculture total production of the Field Crops, agricultural land area, temperature and humidity is studied, the data is collected (for the period 1999-2020) and two different models are developed, i.e. the univariate and multivariate models based on the Huber loss robust technique and quantile regression. A transformed data set is used to mitigate the impact of skewness and stabilize variance. This study evaluates the significance of goodness of fit of the above random variables. Moreover, several assumptions and testing of hypotheses are conducted to identify the behaviour of the data. Due to the size of the collected data, the bootstrap approach is utilized to verify the predictability, goodness of fit, and uniqueness of the estimations. Furthermore, the prediction accuracy of the univariate/multivariate models of Huber loss was resilient when compared to each quantile model, and it was discovered that both models fit the data well. A 95% prediction interval of production is generated and is shown to be valid for the models. Based on real data example, it turned out that the predictive accuracy of robust regression and representative quantile regression models was not significant for univariate/multivariate models. When comparing goodness of fit indices in univariate and multivariate models, the RR and sixth QR models were found to be the best fits for the data. The recommendation, limitations and future research are discussed.

Keywords Huber Regression, Quantile Regression, Robust Regression, Diebold-Mariano Test, Multivariate Model, Field Crops

1. Introduction

The accuracy and efficiency are two important principles in techniques of estimation, evaluation, modelling and forecasting for any data set of applied sciences (agriculture, health, production, finance, etc.). In order to get accurate and efficient estimation, evaluation, modelling and forecasting, the statistical assumptions should be checked and satisfied for any statistical technique.

It is well known that the Crop models are many, and each model has its assumptions and limitations - some of which are the statistical models. In general, the Crop models with classifications are Descriptive Model, Deterministic Model, Stochastic Model, Dynamic Models, Static Model, Simulation Models, and Explanatory Model [1-2]. In addition, the Crop models are based on statistical methods, where the models describe the relationship between the random variables of Crop production, yield, and meteorological variables.

The above models use statistical approaches to measure relationships within a system and are expressed as regression equations which directly describe observable data. As a result, they require less data. These models are effective for discovering major relationships in historical data sets. Statistical models are employed for estimating agricultural yield in specific environments, providing policymakers with valuable information on management and production options.

Crop production studies have generally relied on traditional experiential agronomic research when crop production functions are generated from statistical analysis without regard for the fundamental biological or physical concepts involved. The use of correlation and regression analysis has offered some qualitative understanding of the factors and interactions involved in cropping systems, contributing to the advancement of agricultural research. Crop modelling allows researchers to test and evaluate adaptation techniques for lowering climate change's influence on crop production [3-4].

The modelling and analyzing of the Simple Linear Regression Model (SLRM) and Multiple Linear Regression Model (MLRM) are the most important techniques for practical problems of applied sciences, like the Crop modelling and analysing. In addition, the regression techniques have wide interests for many researchers in applied fields due to their simplicity and forecasting purposes. Moreover, the statistical assumptions of regression techniques are many; and in most practical problems these assumptions cannot be satisfied completely [5-8].

Also, the assumptions of the regression techniques should be tested in order to assure the suitability of the model to the given data, and then they should confirm the accuracy of the modelling, estimation, and forecasting; in addition, they should confirm the reliability and validity of the conclusions drawn from the estimation, and they should forecast using the regression techniques.

Unfortunately, the literature shows that there is a gap in published papers in the field of modelling and analysing the data of the crop using SLRM and MLRM, which addresses the assumptions and tests mentioned above (ibid).

It is well-known that numerous fundamental presumptions were required in regression analysis, including the independence of the dependent variable's observations, the homogeneity of variance of the residuals, independence of observations, the normality of the random variables, and the absence of perfect multicollinearity of the independent variables (ibid).

Moreover, the violation of any of the above assumptions will affect the accuracy of the results. For example, the outliers have multiple effects, and they may affect the overall assumptions of the statistical model. If it is not distributed randomly, it may lead to the data set not being normally distributed; and it may also increase the variance of errors and thus the power of statistical tests to be reduced. In addition, it may lead to biased or non-efficient estimates.

Additionally, in many practical problems, the regression analysis findings can be unnecessarily impacted by outliers or significant data points. Thus, it is critical to look for outliers that might have an impact on the regression model and to take appropriate action if found. In such cases, the regression techniques should not be implemented, otherwise, the processes of estimation, evaluation, modelling, forecasting, etc. are useless and the results are not accurate and not useful (ibid).

From the above observations, it is clear that the justification for this research very briefly, comes, from the fact that models for studying the agricultural production in general, and crops production specifically, have multiple requirements, conditions, and numerous statistical properties, and their data are greatly affected by many characteristics of the data, such that the homogeneity of outliers, missing data and autocorrelation. In addition, developing the methods “Robust Regression” and “Quantile Regression” is excellent options as they are not affected by any of the problems mentioned above (ibid).

The well-known techniques of quantile and robust regression are the most critical and significant techniques for detecting the outliers and getting robust estimators, and to be considered as two complementary topics which are very famous in treating the problems of outliers in different contexts. In addition, these procedures give a deeper understanding of conditional distributions and robustness against outliers. Even though they have inherent drawbacks, continuous research is tackling these problems and looking for fresh uses. For academics and practitioners looking for accurate estimates in the midst of complicated data structures, these strategies are essential. The relationship that exists between robust and quantile regression is anticipated to make a substantial contribution to our knowledge of linkages in a variety of domains as statistical approaches continue to advance [9-12].

It is well-known that the technique of Robust Regression (RR) is designed to solve the problem of outliers if they exist [13]. In addition, this method aims to minimize the impact of outliers, influential observations that can unduly affect parameter estimates. In addition, outliers can distort the classical Ordinary Least Squares (OLS) estimates, leading to inaccurate inferences. In the presence of outliers, RR, in particular the Huber regression is intended to deliver more accurate parameter estimations [14].

In addition, the statistical bounds and convergence characteristics for Huber estimation and inference have been further examined; and RR is less sensitive to outliers than OLS. Moreover, the influence of extreme observations is mitigated, and allowing for more accurate parameter estimation. It may be added here that different robust loss functions, such as Huber, Tukey's biweight, and Andrew's wave, offer flexibility in addressing the specific characteristics of the data.

Moreover, the problem of not satisfying the normality assumption is particularly famous in practical problems; but the RR methods can be applied to data suffering from the above assumption (non-Gaussian distribution). One of the best advantages of the method of RR over other techniques of estimation is that it can work properly with such data due to its efficiency in handling diverse datasets (ibid).

It may be necessary to mention that the applications of RR are extensive - some of these are: assessing the associations between financial variables in financial modelling when severe events (e.g. market collapses) can dramatically affect the data. Because of stochastic events, environmental data frequently includes outliers. Also, in predicting the effect of environmental influences on health outcomes, it can be implemented in biomedical research to describe correlations between biological data, where outliers may signify abnormal or unhealthy situations [15]. In addition, it is less impacted by irregular drawbacks, and it aids in the identification of factors impacting product quality in manufacturing and quality control [16].

The qualitative random variables are commonly used in practical problems, and modelling such random variables using the technique of MLRM is not proven statistically, and thus has many limitations. In addition, any model with only qualitative (dummy) independent variables and many sets of the data will be useless for generating estimates of coefficients of MLRM because the dummy variables reduce the amount of available data. The problem of reducing the amount of data will display the estimators to a bad condition, which is called "the estimator's breakdown point". If the MLRM model contains quantitative and few qualitative random variables, then the "OLS Method fits the MLRM model in **a non-robust way**" [6,7]. However, it is well known that the Least Square method is very sensitive to outliers. Then, the well-known robust estimation method denoted by RDL1 [17-18] should be developed.

The second topic to be considered in this paper is also one well-known statistical technique, which is called "Quantile Regression" (QR). QR is a statistical technique that extends traditional regression by estimating not just the mean but various quantiles of the response variable. While OLS focuses on the conditional mean, more information and understanding of the conditional distribution of the response variable are provided by QR [19]. In addition, QR is inherently robust to outliers just like RR as it focuses on estimating quantiles, which are less sensitive to extreme values. Unlike OLS, QR does not assume constant variance of errors, thus making it suitable for data with heteroscedasticity [20]. Moreover, QR technique is flexible in handling asymmetric distributions and permits for a more concrete understanding of the relationship between variables.

Just like RR, the applications of QR are extensive. One such application involves examining the effects of numerous variables on academic outcomes, such as the test scores at various percentiles of the distribution in order to spot variables that can affect pupils differently depending on their academic attainment levels. In addition, the QR is used in income distribution studies to determine how various variables affect income at different percentiles [21]. In addition, it is applied extensively in financial modelling to evaluate the risk and return profiles at various quantiles of the distribution. Besides the above applications, QR is applied in health economics, and is used to examine how various factors taking into account various quantiles of the health distribution affect the results of health care [22].

In addition, QR can be developed to examine on how to eliminate disparities or discriminatory patterns which fluctuate across various quantiles of the wage distribution in studies of job market discrimination. Regarding the problems of Heteroscedasticity and outliers in environmental information, QR frequently exists in practical data/problems; and QR is very useful in this regard. Additionally, the problems of modelling the effects of environmental variables on health and varying levels of susceptibility are taken into account by developing QR. Moreover, QR aids in the knowledge of how various variables affect earnings at various percentiles of the pay distribution.

QR may be used in credit scoring models to evaluate an individual's employability at various risk levels; and to examine demographic trends, such as income distribution across age groups or the effects of numerous variables on population health at various quantiles, QR can be used. In addition, QR is very useful in the insurance sector, and the risk evaluation can be conducted using QR. In addition, in the modelling of the distribution of insurance claims and the estimation of the impacts of various variables on the tails of the distribution, QR may be prone to severe occurrences [23].

In modelling the conditional distribution of a time series variable, QR can be used for this problem, i.e. in time series analysis, and how the predictor can effect change throughout different time-periods. Thus, developing QR analysis in practical problems is extensive, and the above uses illustrated the adaptability and wide applicability of QR in tackling research topics where the effects of variables may change across several segments or levels of a distribution. Its applicability is anticipated to increase across a range of areas such as computer techniques.

It is well-known that the OLS method is one of the best estimation methods, and usually developed in traditional regression techniques; but it is highly sensitive to outliers and may produce estimates that are skewed in the presence of extreme values. These problems exist in real data sets of several problems [24], and can be neglected in alternative techniques like Huber Robust and Quantile Regression.

The methods of RR and QR offer flexibility and robustness in estimating statistical parameters. A non-Gaussian distribution data is thoroughly analysed to study the underlying properties of quantile and robust regression using Huber loss function methods for univariate/multivariate models. In addition, the Huber loss can be made less susceptible to outliers than the conventional least squares loss by selecting the right value of δ. The above two methods for parameter estimation can offer a differentiable function that can be tuned using different numerical techniques such as gradient descent.

In this paper, the production of Field Crops (the production measured in tons) and related random variables (the area was measured in feddans, temperature in degree Celsius and humidity in percentage) for the period

1999-2020 are collected, and two different models, univariate and multivariate regression models using the RR and QR were analysed for the study purpose. The univariate model was comprised of total agricultural production and agricultural land area, where total agricultural production was the response variable. Climate variables, temperature, and humidity, were added to the univariate model as predictors to form the multivariate model. This study explores the significance of goodness of fit in two advanced regression models, RR and QR, for moderately skewed data values. Huber loss robust and QR techniques were applied and modelled to identify the link between agricultural land area and agricultural productivity for univariate and multivariate regression models. Model estimation, goodness of fit, and the bootstrapping method were used to check the validity of the models' estimates, and for forecasting the outcome accuracy, which was assessed for the study analysis.

The need for this research is coming from several real national and international reasons. The first reason is the international direction of Crops Models which is based on the selection of the best model for modelling the data which should not be affected by the outliers, missing data and autocorrelation. Secondly, due to the importance of Crops production studies in the life, and because of the urgency of these studies, the precise crop production management including data analysis and modelling are necessary and beneficial to any country, agriculture sector, and many other industrial activities, such as Food Industries.

This paper is divided into nine sections. The literature review, which is contained in Section 2, is a comprehensive examination and analysis of recent academic literature and scholarly sources that are relevant to the research question. The methodology, which outlined the methodological procedures followed to address the research questions and achieve the objectives of the study, is covered in Sections 3, 4, 5 and 6. These sections give a comprehensive and understandable description of the methods used to assess the reliability and validity of the results. The findings and conclusion are presented in Sections 7 and 8. Section 9 contains the study's limitations and future research objectives.

2. Literature Review

The literature of this paper includes three main topics, which are as follows: some papers, which deal with some applied problems by applying some statistical models including the regression models to be reviewed. Then, some papers on the RR and QR techniques will also be reviewed. In addition, the available literature on the above three topics is very extensive and so it will be difficult to review all of them in this paper. Therefore, for the purpose of keeping within the limitational boundaries of this paper, we will summarize only some of them.

Irz [25] explained why agricultural growth should be expected to relieve poverty. Several realistic and compelling justifications exist, including the creation of jobs on the land, links from farming to the rest of the rural sector, and a decrease in the real cost of food for the entire economy, but the degree of influence is always qualified by specific conditions. Bravo-Ortega and Lederman [26] made three contributions to the literature field on agricultural productivity using ordinary regression models.

Pavlov and Karmyshova [27] used a systematic approach, enriched by the multifunctionality paradigm of agricultural production to develop an algorithm that consists of a series of steps aimed at creating a regression model that reflects the significance of selected factors and their impact on the development of the regional agricultural system

Zwolak [28] conducted research based on the functional dependence of net final output on land, labor, and fixed assets in agriculture, and it was concluded that the growth rate of the value of output was greater than proportionate. Diana and Corraya [29] created an advanced and efficient system for forecasting future agricultural crop prices and yields while accounting for inflation and other relevant considerations.

Zeru [30] examined the factors that influence the performance of agricultural goods in the Amahara area national state in order to identify the most important input parameters for producing high-quality agricultural outputs.

In Northern China, Sheng [31] explored the relationship between maize yield and farm size. Young [32] reviewed the key methods currently utilized to generate official statistics, such as surveys, remote sensing, and the integration of these with meteorological, administrative, or other data.

Deepa [33] conducted a thorough investigation into the prediction of cotton prices using five distinct machine learning regression algorithms: Poisson, Boosted Decision Tree, Decision Forest, Linear, and Bayesian Linear. In Kenya, Madagascar, and Mozambique, Fitawek and Hendriks [34] calculated the effects of Large-scale Agricultural Investments (LSAIs) on household food security in one community each. To account for potential selection bias due to unobserved factors, they adopted an endogenous switching regression model. Using regression modelling, Hayat [35] investigated the effect of soil and other factors on wheat yield.

Irmeilyana [36] used binary logistic regression analysis to discuss the elements affecting the land productivity of coffee farms in Kota Pagar Alam. Spicka [37] summarized and compared knowledge from 51 studies to provide a complete discussion on several methods for measuring economic viability and sustainability in order to properly provide income support for farms in places with environmental limits. Ustaoglu [38] applied regression analysis to figure out agricultural land area and grain production for unknown data points from Ottoman Archives, as well as a cropland suitability map, accessibility, and geophysical attributes such as ancillary data, to estimate non-irrigated crop production and its corresponding cultivation area in the 1840s Bursa Region.

Wójcik-Leń [39] reviewed the terminology and classification of land unfit for agriculture in Poland and several European and Asian countries. Iuliia [40] researched the creation of regression models that describe the operations of farming businesses throughout Russia's regions. Xie [41] performed regression analysis to estimate how to increase the overall value of Sichuan Province's agricultural output with varied factor inputs. It may be mentioned here that we are not sure about the accuracy of modelling of some of above papers using SLRM and MLRM, since the conditions of SLRM and MLRM were not verified. Yu [42] researched robust estimation. Nugroho [15] sought to compare the robust estimation approach and the OLS method on data with various levels of significance. Alshqaq [43] researched robust estimation of spread and smooth regression functions, as well as their derivatives. Using real data from a geological mapping project, Boogaart [44] treated both classical least-squares regression and robust MM regression, and they are contrasted inside several regression models.

Wasim [45] suggested some brand-new, reliable ridge Minimax-estimators (M-estimators). They used a Monte Carlo simulation analysis to compare the new estimators' effectiveness. In order to accommodate temporal correlation within each individual, Yoon and Galvao [46] created cluster resilient inference techniques for panel QR models with individual fixed effects. Tian [47] used machine-learning techniques to identify the connection between the utilization of cultivable land and poverty from a global perspective. Using QR, Yahya and Lee [48] examined the asymmetric impact of agriculturalization on climate neutrality goals.

For censored QR, Wang [49] presented a new estimate technique based on a multiple robust propensity score. Akgun [50] employed the QR approach to analyze the factors that influence the happiness index scores throughout eight periods in European countries. In a predictive QR model, Cai [51] put forth an innovative approach to provide a robust inferential theory across all varieties of persistent regressors.

Hao [52] researched smoothed QR using nonparametric inference. Kai [53] used a New MM Algorithm to investigate Nonparametric and Semiparametric QR. For QRs, Lee and Shin [21] suggested a brand-new full subset averaging (CSA)-based conditional quantile prediction approach. Liu [54] created a QR predictability test that may be used independently of the existence of an intercept or the persistence of a predictor.

Nguyen-Tang [55] investigated the Composite QR Neural Network for Forecasting of Water Quality Extremes. Suhail [56] looked at the new quantile-based ridge M-estimator for multi-collinear and outlier linear regression models. Zhang [57] studied the penalized QR for a geographic panel model with fixed effects, which uses a tuning parameter to reduce each fixed effect to a standard value in order to control excessive variability.

3. The Methodology

Because of the research problem of this paper, as explained in Section 1, which is related to a sensitive applied issue, the data for such problems often, suffers from some deficiencies in the characteristics required by the Statistical Methods. In order to model the data of this paper efficiently, the methodology of this paper will take into account several important issues and will consist of a number of stages, as follows:

3.1. The Research Problem and its Variables

The problem of this research is modelling the data of the variables: agricultural total production of the Field Crops, agricultural land area, temperature and humidity by developing two models, i.e. the univariate and multivariate models based on the Huber loss robust and QR, agriculture production output, agricultural land area, temperature, and humidity make up the data variables. It may be mentioned that the area was measured in feddans, the production in tons, the temperature in degrees Celsius, and humidity in percentage.

This research was proposed because of the many problems which affected the crops models, and due to the difficulties, that accompanied them. The literature related to this topic indicates that the majority of field crop studies are carried out according to models that are negatively affected by many problems and statistical properties, including, but not limited to, the problems of outliers, missing data, and autocorrelation. The proposed models in this research are little affected by the above problems [10,58].

3.2. The Data Resources

"The National Centre for Statistics and Information Oman (NCSI)" and "Directorate General of Meteorology" provided the climate variables' study data from year 1999 to 2020 (https://www.caa.gov.om/caa/directorates/directorate-general-of-meteorology). In addition, the agricultural data is collected (for the period of 1999-2020) from "The National Centre for Statistics and Information Oman (NCSI)". For the study, daily data for the period were analyzed.

3.3. Data Studying and Identifying Data Shortage

The natural log of agricultural production and cultivated area was used in the analysis to reduce the influence of skewness and to stabilize the variance of the agriculture data.

3.4. The Proposed Methods: Prior to the Analysis, the Following Assumption Tests for Regression Models Were Performed

- Quantile-Quantile(Q-Q) plot tool used to determine whether dataset follows a normal distribution or not
- Normality of residuals using Shapiro-Wilk test
- Linearity of residuals performed using Rainbow test
- Homoscedasticity of residuals checked using Breusch-Pagan-Godfrey test
- Multicollinearity of independent variables examined using Tolerance and variance inflation factor
- Effect of outliers of the model inspected using Cook's distance.

3.5. The Software Selection

For the study analysis, IBM SPSS 22 as well as R software packages 'ggplot2', 'quantreg', 'lmtest', 'robustbase', 'MASS', 'CAR' 'boot', and 'forecast' were utilized, as they provide accurate results for study analysis.

3.6. Interpretation of Numerical Results

The significance of goodness of fit in two advanced regression models, RR and QR, for moderately skewed data values is investigated in this paper. In univariate/multivariate models, RR and the sixth QR models were identified as the most suited models for the data when comparing the goodness of fit measures. Based on a real-data example, the prediction accuracy of RR and representative QR models was found to be insignificant for univariate/multivariate models.

4. The Proposed Methods

Two of the main methods to be developed in this paper for the given research problem are RR and QR, and the methods of RR and QR are to be discussed in this section.

4.1. Robust Regression using Huber Loss Function

A statistical method known as RR seeks to give accurate estimates of the regression coefficients even when the data don't conform to the assumptions of conventional least squares (OLS) regression [4,7]. In addition, RR methods reduce the influence of outliers or leverage points, making the estimation procedure more robust to their influence, as objected to concentrating only on minimizing the sum of squared residuals. As we discuss in Section 1, RR can lead to more accurate and reliable statistical inference by taking into account a larger range of data features [15].

The Huber loss function is often employed as the M-estimation goal function in RR and is one of the resilient loss functions utilized to mitigate the impact of outliers on the model setting process. The M-estimation is a general method for parameter estimation that minimizes an objective function.

It may be mentioned that the squared loss and the absolute loss are combined to form the Huber loss function. The following is a general expression of MLRM for RR using M-estimation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \ , \qquad (1)$$

where y_i is the i-th data point's observed value, x_{ij} is the i-th data point's j-th predictor value, β_j is j-th predictor's coefficient and ε_i explains the deviation between the observed value, y_i , $\varepsilon_i{}'s$ are the residuals and the forecasted value is $f(x_i;\beta) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$.

In order to minimize the influence of outliers or leverage points in the data, we employ a robust loss function, i.e. the Huber loss function denoted by H(e) that penalizes large residuals less than the squared loss function in place of the squared loss function (used in OLS). In addition, the Huber loss function is well-known as a reliable loss function utilized in M-estimation, and given by,

$$H(e) = \begin{cases} \frac{1}{2} e^2 & \text{for } |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right) & \text{for } |e| > \delta, \end{cases} \qquad (2)$$

where the residual, or the variation between the observed and estimated values, is represented by the letter e, and the tuning parameter δ determines the point in the loss function where squared loss changes to absolute loss. It specifies the point at which the loss switches from being quadratic (squared) to being linear (absolute).

The goal of finding the model parameters that lessen the sum of the Huber loss function across all data points is the goal of M-estimation. This goal can be satisfied by solving the optimization problem,

$$\min_{\beta} \sum_i H_\delta\big(y_i - f(x_i;\beta)\big), \qquad (3)$$

where β stands for the model parameters that must be calculated, the observed value for the i^{th} data point is represented by y_i , $f(x_i;\beta)$ is the i^{th} data point's projected value determined by the model and the parameter.

The well-known numerical optimization methods like gradient descent, Newton-Raphson, or Iteratively Reweighted Least Squares (IRLS) can be used to minimize the objective function in Eq. (3), in order to obtain the ideal coefficients. In addition, depending on the unique characteristics of the data and the desired level of robustness, the robust loss function and its tuning parameters will be selected. For the case of the data, if affected by noise or outliers, RR will employ M-estimation, which offers a more accurate estimation of the model parameters. When the OLS presumptions, such as normality and constant variance of errors, are not satisfied, it is especially helpful [58-60].

4.2. Quantile Regression (QR) Approach

It is well-known that the QR is a statistical method which focuses on assessing the association between variables at various quantiles of the distribution of the dependent variable. Also, QR offers insights into the conditional distribution as opposed to OLS regression, which estimates the conditional mean [61], and these results are very important as they provide a more thorough knowledge of the relationship between the variables.

Over other regression models, QR has a number of benefits like the ability to capture heterogeneity, to be resistant to outliers, handle skewed distributions, and do subgroup analysis. The QR model can be defined as follows [62].

Let $(x_i, y_i), i = 1,2,\dots,n$, be a sample from some population, where xi is a $K \times 1$ vector of regressors. It is assumed that $P_r(y_i \le \tau | x_i) = F_{u\theta}(\tau - x_i'\beta_\theta | x_i), i = 1,2,\dots,n$. This relation can be expressed as,

$$y_i = x_i'\beta_\theta + u_{\theta i}\ ,\ \mathrm{Quant}_\theta(y_i|x_i) = x_i'\beta_\theta, \qquad (4)$$

where $\mathrm{Quant}_\theta(y_i|x_i)$ denotes the θ-th conditional quantile of the observed value of the response variable y_i for the i^{th} observation given the values of predictor variables x_i for the i^{th} observation, β_θ stands for the vector of coefficients to be estimated. If $F_{u\theta}(.)$ was known, then various techniques could be used to estimate β_θ. However, here the distribution of the error term $u_{\theta i}$ is left unspecified, and as is implied by Eq. (2), it is only assumed that $u_{\theta i}$ satisfies the quantile restriction $\mathrm{Quant}_\theta(u_{\theta i}|x_i) = 0$. Then, in general, the θ^{th} sample quantile $(0 < \theta < 1)$ of y, say $\hat{\mu}_\theta$, which minimizes the weighted sum,

$$\mathrm{Min}_b \begin{Bmatrix} \sum_{i:y_i \ge b} \theta |\, y_i - b| + \\ \sum_{i:y_i \ge b} (1-\theta) |\, y_i - b| \end{Bmatrix}. \qquad (5)$$

Estimating various conditional quantiles of the response variable using Eq. (5) above, i.e. QR offers a useful addition to conventional regression techniques [63].

5. Developing Goodness of Fit

In Section 4, the estimation methods, denoted by RR and QR, are developed; and in this section, several information criteria for goodness of fit and Diebold-Mariano forecasting test are to be discussed in order to assure the accuracy of RR and QR.

5.1. General Goodness of Fit Measures

It is necessary to append the estimation techniques of the regression model proposed in the previous section by the goodness of fit techniques - some of which are: the log-likelihood, the Mean Absolute Error (MAE), the Watanabe-Akaike Information Criterion (WAIC), and the Deviance Information Criterion (DIC) used to assess the goodness of fit of the regression model.

The log-likelihood function measures how well a model matches the available data. A higher log-likelihood number suggests a more accurate model fit. It may be mentioned here that Lower MAE indicates better forecast accuracy. Finally, the WAIC is utilized to assess the model fit by taking into consideration both goodness of fit and model complexity. Lower WAIC values indicate a better fit for the model. The DIC is an extra model complexity and fit measure. Lower DIC values indicate an improved fit for the model. The formulas for these measures are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \hat{x_i}| \qquad (6)$$

$$\mathrm{DIC} = \text{Mean deviance} + 2 \cdot (\text{Mean deviance - Deviance at posterior mean}) \qquad (7)$$

$$\mathrm{WAIC} = -2\left(\sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1} p(y_i \mid \theta_s)\right) - \sum_{i=1}^{n} \mathrm{Var}\big(\log p(y_i \mid \theta)\big)\right) \qquad (8)$$

5.2. Diebold-Mariano Forecasting Accuracy Test

The statistical test, Diebold-Mariano test (D-M test) is applied to compare the predictive accuracy between two regression models [64]. This test analyzes the mean squared forecast errors of two comparing predicting regression models to see if one performs considerably superior to the other [65].

6. Developing Bootstrapping and Real Data Study

In order to perform a statistic's sampling distribution, a resampling approach known as bootstrapping is performed. It is applied for model viability assessments and model aggregating methodologies, as well as for calculating the error of model selection procedures that require tuning parameters. This approach is very helpful when utilizing standard analytical approaches, which might be challenging. It provides a flexible and model-independent way of calculating confidence intervals.

Initially, the original dataset is employed fitting the Huber loss robust and QR models for getting RR and QR parameter estimates. Following that, B = 5000 bootstrap samples were generated from the original dataset. Each bootstrap sample was fitted with the RR and QR models for the specified quantiles, and the parameter estimates were captured [55]. The bootstrap sample distribution of parameter estimates is then utilized to generate confidence intervals for RR and QR models. The generated confidence intervals for the models provide knowledge of the uncertainty surrounding quantile-specific parameter values, allowing for a more extensive understanding of the conditional distribution.

7. Results

This Section presented and discussed the results of the approaches developed in Sections 3, 4, 5 and 6.

7.1. Preliminary Results

A general comprehension of the data's fundamental statistical features is required prior to data analysis. The mean and standard deviation of agricultural production, agricultural land area, temperature and humidity, were found to be 14.25(0.393), 12.14(0.183), 28.2093(0.0731) and 48.5872(0.2905) respectively. In addition, the median, range, skewness and kurtosis were (14.06, 1.12, 0.83, -0.97) for production, (12.07, 0.55, 0.91, -0.583) for cultivated area, (28.2242, 1.24, -0.812, 0.349) for temperature and humidity (48.8423, 4.48, -0.568, -0.695) respectively.

The above results showed that the distribution of all the variables was platykurtic and moderately asymmetric. Moreover, the results of Quantile-Quantile plot (Q-Q plot) are displayed in Figure 1. For all the above variables, which indicated that the variables' data points slightly deviated from the diagonal line and resulted in a moderate skewness of the variables' data values.

In addition, the linearity of the residuals is validated by the rainbow test, which produced a p-value of 0.0431, and showed a marginal significant linearity. The studentized Breusch-Pagan test's p-value of 0.08875, which was higher than the typical significance criteria of 0.05, indicated that heteroscedasticity does not exist in the regression model. The variance inflation factor and tolerance for the predictor variables area, temperature and humidity were (1.250255, 1.361314, 1.571801) and (0.7998366, 0.7345844, 0.6362128) respectively. It revealed that multicollinearity is not a major issue. Finally, Cook's distance of the model is evaluated and the highest value 0.3 is obtained, which indicates that outliers have a moderate impact on the predicted regression coefficients.

The above results indicated that the techniques of RR and QR can be developed for modelling the data of given variables.

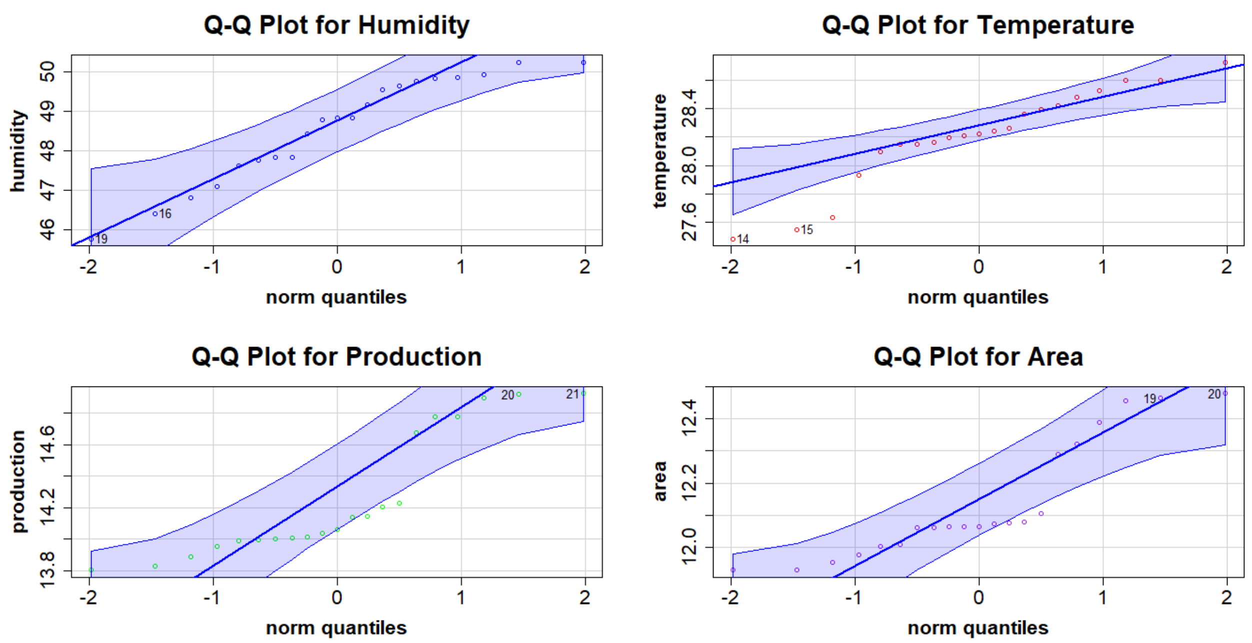

Figure 1. Normal Q -Q plots for the study variables

Table 1. Regression estimates of all univariate models

	Regression Estimates for area		Regression Estimates for temperature		Regression Estimates for humidity	
Regression Model	**Intercept**	**Coefficient**	**Intercept**	**Coefficient**	**Intercept**	**Coefficient**
Robust	-11.222	2.098	20.825	-0.236	20.957	-0.139
Quantile						
tau = 0.1	-12.922**	2.231**	23.320	-0.333	17.855	-0.081
tau = 0.2	-12.442**	2.193**	20.245	-0.222	17.437	-0.072
tau = 0.25	-12.438**	2.192**	20.059	-0.215	18.898	-0.101
tau = 0.3	-12.115**	2.167**	19.520	-0.196	18.482	-0.091
tau = 0.4	-11.226**	2.096**	18.876	-0.172	18.352	-0.089
tau = 0.5	-10.463**	2.034**	20.061	-0.211	17.487	-0.070
tau = 0.6	-10.829**	2.067**	19.029	-0.174	23.127	-0.182
tau = 0.7	-10.271**	2.023**	24.783	-0.376	24.167	-0.202
tau = 0.75	-10.271**	2.023**	20.110	-0.189	22.615	-0.169
tau = 0.8	-10.752**	2.069**	20.522	-0.203	22.615	-0.169
tau = 0.9	-9.7605**	1.992**	15.72	-0.029	14.63	0.006

* and ** denote statistical significance at the 5and 1% level, respectively

Table 2. Goodness of fit of univariate models

	Area				Temperature				Humidity			
Regression Model	**Log-lik**	**MAE**	**WAIC**	**DIC**	**Log-lik**	**MAE**	**WAIC**	**DIC**	**Log-lik**	**MAE**	**WAIC**	**DIC**
Robust	18.62		0.20	0.23	-9.73		2.96	3.4	-8.34		2.60	2.98
Quantile												
tau = 0.1	11.03	0.102	0.41	1.29	-15.95	0.35	51.46	9.78	-14.39	0.34	45.59	8.40
tau = 0.2	12.57	0.098	0.35	1.11	-13.67	0.30	37.79	5.22	-13.03	0.30	37.43	5.68
tau = 0.25	12.59	0.096	0.35	1.11	-13.54	0.29	37.00	4.96	-11.59	0.28	28.81	2.81
tau = 0.3	14.32	0.10	0.30	0.94	-13.17	0.29	34.79	4.22	-11.21	0.27	26.49	2.04
tau = 0.4	16.68	0.097	0.24	0.75	-12.75	0.28	32.24	3.37	-11.10	0.27	25.81	1.81
tau = 0.5	18.02	0.097	0.21	0.66	-11.06	0.28	22.12	2.24	-10.19	0.28	20.38	1.45
tau = 0.6	18.74	0.090	0.19	0.62	-10.60	0.28	19.36	0.92	-8.72	0.27	11.59	0.56
tau = 0.7	18.09	0.103	0.21	0.66	-9.90	0.30	25.16	2.34	-9.51	0.29	16.35	0.64
tau = 0.75	18.09	0.120	0.21	0.66	-20.74	0.55	79.01	18.96	-10.57	0.34	22.64	0.75
tau = 0.8	10.86	0.151	0.42	1.31	-20.74	0.56	80.18	19.35	-10.57	0.34	22.64	0.75
tau = 0.9	5.55	0.196	0.69	2.17	-24.04	0.66	100.04	25.97	-24.22	0.66	104.6	28.08

7.2. Univariate/Multivariate Models

The result of the RR and QR estimates of the univariate models between Agricultural production and the predictor variables (Agricultural land area, Temperature and Humidity) are calculated (estimators, Quantiles, R-Squared, Residual, Standard Error, and related measures) for tau= 0.1 to 0.9 and summarized in Table 1. The estimated representative univariate sixth QR models are given by,

$$Q_{0.6}(Y \mid X) = -10.829 + 2.067 \times \text{Agricultural land area} \quad (9)$$

$$Q_{0.6}(Y \mid X) = 19.029 + (-0.174) \times \text{Temperature, and} \quad (10)$$

$$Q_{0.6}(Y \mid X) = 23.127 + (-0.182) \times \text{ Humidity} \quad (11)$$

The results of Table 1 showed that the estimators of the regression parameters are significant, and R-Squared values were highly significant for Agricultural land area and moderate for temperature and humidity. This means that the model is appropriate for fitting the data.

The results of modelling the multivariate models are given in Table 3, which provides an overview of the RR and QR approaches' outcomes. The estimated representative sixth QR model of the above variables is given by,

$$Q_{0.6}(Y \mid X) = 1.86 + 1.85 \text{ x Agricultural area} - 0.284 \times \text{Temperature} - 0.04 \text{ x Humidity} \quad (12)$$

Table 3. Regression estimates of multivariate models

Regression Estimates						
Regression Model	**Intercept (SE)**	**Area (SE)**	**Temperature (SE)**	**Humidity (SE)**	**Multivariate Adjusted R-Squared**	**Residual Standard Error**
Robust	-0.2631 (3.21)	1.898 (0.115)**	-0.235 (0.066)*	-0.039 (0.018)	0.959**	0.0748
Quantile						
tau = 0.1	-7.967	2.173**	-0.113*	-0.021	0.888**	0.13834
tau = 0.2	-4.604	2.076**	-0.169*	-0.034	0.917**	0.12479
tau = 0.25	-3.243	2.031**	-0.187*	-0.040	0.926**	0.12124
tau = 0.3	-3.243	2.031**	-0.187*	-0.040	0.926**	0.11365
tau = 0.4	0.6823	1.926**	-0.256*	-0.054	0.941**	0.11267
tau = 0.5	2.523	1.930**	-0.319*	-0.056	0.945**	0.10917
tau = 0.6	1.860	1.845**	-0.284*	-0.041	0.951**	0.09871
tau = 0.7	1.063	1.798**	-0.234*	-0.041	0.946**	0.11467
tau = 0.75	0.101	1.773**	-0.210*	-0.035	0.937**	0.13124
tau = 0.8	2.434	1.750**	-0.217*	-0.066	0.902**	0.15059
tau = 0.9	-4.002	2.151**	-0.212*	-0.036	0.837**	0.19876

* and ** denote statistical significance at the 5% and 1% level, respectively

Table 4. Goodness of fit of multivariate models

Regression Model	**Log-likelihood**	**MAE**	**WAIC**	**DIC**
Robust	23.558	0.063	0.124	0.143
Quantile				
tau = 0.1	13.89	0.084	0.312	0.982
tau = 0.2	16.996	0.074	0.232	0.731
tau = 0.25	18.164	0.075	0.208	0.654
tau = 0.3	18.164	0.075	0.208	0.654
tau = 0.4	20.682	0.071	0.163	0.515
tau = 0.5	21.374	0.079	0.153	0.482
tau = 0.6	22.605	0.094	0.136	0.428
tau = 0.7	21.489	0.100	0.151	0.476
tau = 0.75	19.873	0.118	0.136	0.556
tau = 0.8	15.333	0.135	0.272	0.856
tau = 0.9	9.9631	0.140	0.464	1.428

It can be observed, that based on the results of Table 3, the estimators of the regression parameters are significant, and R-Squared values were significant except for humidity.

7.3. Goodness of Fit for Univariate/Multivariate Models

When comparing several models for the same problem, the best fitting estimation model is chosen based on the goodness of fit measures. The measures Log-likelihood, MAE WAIC, and DIC computed for RR and QR univariate/multivariate models are given in Tables 2 and 4.

Results showed that the sixth quantile is the representative model for all QR models.

The Log-likelihood, WAIC, and DIC values for the RR model were observed to be 18.62032, 0.1988451, and 0.228608, respectively (Table 2). The goodness of fit values (Log-likelihood, MAE, WAIC, DIC) for the QR were displayed in Table 2.

According to the results of RR values for Log-likelihood, MAE, WAIC, and DIC were 23.5582, 0.0628, 0.1423 and 0.1428 respectively for multivariate model (Table 4). The quantile multivariate model's goodness of fit measures Log-likelihood, MAE, WAIC, and DIC values are shown in Table 4.

7.4. Bootstrapping for Univariate/Multivariate Models

The bootstrapping results of univariate models are represented in Table 5; and it can be observed that the estimates of the regression models found within the limits of 95% bootstrapping confidence intervals satisfied the validity of the regression estimates.

In addition, Table 6 displayed the bootstrapping results of multivariate models, and it was identified that all models' estimates were within the 95% bootstrapping confidence interval, which satisfied the validity of all models' estimates.

7.5. Forecasting Accuracy for Univariate/Multivariate Models

Prediction intervals of agricultural production for the univariate model when area ranges are between 12.5 and 12.7 by employing the Huber loss regression and the candidate quantile models are displayed in Table 8. Intervals between predictions for the candidate models were determined to be suitable for data forecasting. The D-M test is conducted for all combinations of Huber loss regression with QR univariate models to compare the predictive accuracy of the model pairs in order to compare the predictive accuracy between any two regression models. Except for the quantiles with tau= 0.1, 0.2 and 0.25, all pairs resulted in a non-significant predictive accuracy between the models including the representative QR model (Table 7). It indicated that the candidate QR model and the Huber loss RR model are equivalent in terms of predictive accuracy.

Prediction intervals of agricultural production for the multivariate model when area ranges are between 12.5 and 12.7, temperature ranges are between 27 and 29 and humidity ranges are between 46 and 50 using the Huber loss regression and the candidate quantile multivariate models are displayed in Table 9. Prediction intervals of all multivariate models are found to be appropriate for forecasting the data. The D-M test is used to examine the predictive accuracy of each model pair for every combination of the QR and Huber loss regression multivariate models. With the exception of the quantiles with tau=0.1, the prediction accuracy between the models was non-significant for every pair (Table 7). This implied that there is no predictive accuracy difference between the Huber loss robust and the candidate QR model.

Table 5. Bootstrap Confidence Interval of univariate model regression estimates

95% bootstrap Confidence Interval of Regression estimates		
Model	**Intercept**	**Coefficient**
Robust	(-13.4025, -2.5577)	(2.05, 98.06)
Quantile		
tau = 0.1	(-14.2173, -4.7078)	(1.9671, 3.0314)
tau = 0.2	(-22.0824, - 9.4147)	(1.9508, 2.8913)
tau = 0.25	(-20.8543, -2.7429)	(1.9902, 2.9724)
tau = 0.3	(-14.1013, -3.8107)	(2.0236, 2.8767)
tau = 0.4	(-17.8543,- 4.7078)	(2.0567, 2.9876)
tau = 0.5	(-13.8091, -7.2682)	(1.9967, 2.7678)
tau = 0.6	(-14.5217, -6.3178)	(1.9875, 2.4585)
tau = 0.7	(-14.6201, - 5.4076)	(1.8863, 2.6672)
tau = 0.75	(-14.2275, -4.5326)	(1.9175, 2.7658)
tau = 0.8	(-19.9210, -5.2334)	(1.8375, 2.8394)
tau = 0.9	(-15.2732, -3.1869)	(1.8175, 2.8145)

Table 6. Bootstrap Confidence Interval of multivariate model regression estimates

95% bootstrap Confidence Interval of Regression estimates				
Model	Intercept	area	temperature	humidity
Robust	(-25.046, 9.204)	(0.509, 2.510)	(-0.426, 0.324)	(-0.090, 0.081)
Quantile				
tau = 0.1	(-12.766, 14.38)	(0.881, 2.322)	(-0.416, -0.002)	(-0.068, -0.017)
tau = 0.2	(-15.423, 17.48)	(0.874, 2.465)	(-0.432, 0.015)	(-0.068, -0.017)
tau = 0.25	(-14.934, 8.476)	(1.73, 2.464)	(-0.435, 0.113)	(-0.068, 0.067)
tau = 0.3	(-14.934, 8.475)	(1.797, 2.463)	(-0.413, -0.063)	(-0.067, 0.067)
tau = 0.4	(-16.213, 8.406)	(1.637, 2.579)	(-0.406, -0.028)	(-0.078, 0.004)
tau = 0.5	(-18.972, 8.528)	(1.632, 2.562)	(-0.413, 0.059)	(-0.088, 0.010)
tau = 0.6	(-16.212, 16.99)	(0.687, 2.699)	(-0.367, 0.069)	(-0.073, 0,029)
tau = 0.7	(-20.737, 8.799)	(1.630, 2.415)	(-0.384, -0.007)	(-0.098, -0.0008)
tau = 0.75	(-20.737, 8.799)	(1.633, 2.451)	(-0.376, -0.0071)	(-0.103, 0.0075)
tau = 0.8	(-17.419, 6.824)	(1.612, 2.488)	(-0.329, -0.007)	(-0.103, 0.0008)
tau = 0.9	(-25.436, 6.17)	(1.65, 2.598)	(-0.413, 0.314)	(-0.142, 0.00009)

Table 7. Diebold-Mariano predictive accuracy for univariate/ multivariate pairs

Regression model pairs	**Univariate**		**Multivariate**	
Huber loss robust Vs Quantile	**D-M statistic**	**P-value**	**D-M statistic**	**P-value**
tau = 0.1	-3.738	0.0013**	-2.45	0.024*
tau = 0.2	-2.883	0.0092**	1.01	0.326
tau = 0.25	-2.861	0.0096**	1.21	0.241
tau = 0.3	1.682	0.108	1.30	0.209
tau = 0.4	2.150	0.054	1.42	0.170
tau = 0.5	1.867	0.077	2.44	0.169
tau = 0.6	1.574	0.131	1.44	0.166
tau = 0.7	1.125	0.274	1.19	0.247
tau = 0.75	1.125	0.274	0.84	0.412
tau = 0.8	-0.567	0.579	-0.82	0.419
tau = 0.9	-1.701	0.104	-1.11	0.282

* and ** denote statistical significance at the 5% and 1% level, respectively

Table 8. Prediction intervals of agriculture production when area ranges between 12.5 and 12.7 by employing the candidate univariate models

	95% Prediction Interval of regression models	
Model	Lower bound	Upper bound
Robust	14.85969	15.83197
Quantile (tau = 0.6)	14.9612	15.62181

Table 9. Prediction intervals of agriculture production when area ranges between 12.5 and 12.7 temperature ranges between 27 and 29 and humidity ranges between 46 and 50 by employing the candidate multivariate models

	95% Prediction Interval of regression models	
Model	Lower bound	Upper bound
Robust	14.503	16.498
Quantile(tau=0.6)	14.524	16.317

8. Conclusions

In this paper, two models, univariate and multivariate, using the Huber loss robust and QR were analysed for the study's purpose. The univariate model comprised of agricultural total production and agricultural land area, where agricultural total production was the response variable. Climate variables, temperature, and humidity were added to the univariate model as predictors to form the multivariate model.

Model estimation, goodness of fit, the bootstrapping method to check the validity of the models' estimates, and forecasting-outcome accuracy were assessed for the study analysis. The results of this paper show that all the univariate models showed a very high significance (p-value < 0.001) connection variables with very high R-squared values which concluded that the association between the variables was significantly high. In addition, all multivariate models fit the data well. In addition, the humidity did not appear significant for all the models. The results of multivariate models showed that the RR model had the lowest residual standard error. In addition, the adjusted R-squared value was significant and very similar in all multivariate models. In addition, the RR showed the highest association between the variables. In order to identify the most suited model, goodness of fit measures were examined for each of the univariate models and multivariate models.

Based on the goodness of fit values, the univariate/multivariate models, Robust and the sixth QR (tau = 0.6) were observed as the candidate models for the data. For multivariate models, all models showed similar goodness of fit values. The bootstrapping method was applied to all regression models to validate the estimates. Within the 95% bootstrapping confidence intervals, observed regression model estimates were found, which satisfied the validity of the regression estimates. A 95% prediction interval of production was observed to be valid for the models. In order to examine the prediction accuracy of the model pairs, the D-M test was run for every combination of QR univariate/multivariate models and Huber loss regression models. In the univariate model, with the exception of the quantiles with tau=0.1, 0.2, and 0.25, the prediction accuracy between the models was non-significant for every pair, while for the multivariate model, an exception was for tau=0.1. Finally, based on the real data example, it was found that the predictive accuracy of RR, and representative QR models did not result significantly for univariate/multivariate models.

The practical importance of this paper is great. From the international perspective, it gave a clear picture of some of the shortcomings of Crops models and clarified the possibility of choosing other models (RR and QR) that are more accurate and not affected by the problems of Crops models. Regarding the national perspective, this is the first study in Oman, and it sheds light on the Crops data problems.

It is necessary to clarify the importance of the developed models (RR and QR) from a practical perspective. Any researcher in the field of agricultural production who is familiar with ready-made statistical programs - which is an essential part of the general knowledge of the researchers - can apply the proposed models in this paper to his data with complete ease and convenience.

Finally, the findings of the models of QR and RR are very useful for modelling the data of the variables of the study, i.e., the agricultural total production and agricultural land area, climate variables, temperature, and humidity. In addition, the proposed models can assist the development of farming systems that are resilient to extreme weather, can minimize waste, and maximize input utilization, all of which contribute to sustainability. A better comprehension and prediction of crop yields using QR and RR under a variety of situations can aid in the preparation and implementation of food security measures when considering the effects of climate change and rising population demands.

9. Limitations and Future Research Directions

- In order to assure the quality of data collection, the data on agricultural production will be collected from the Ministry of Agriculture, and the data on humidity and temperature will be obtained using the NASA Power data.
- The size of the data collection is for the period 1999 to 2020; but for the future, we will extend the above period.
- Develop and compare different robust methods with small and heavy Breakdown Points (robust and efficient weighted least squares estimator (REWLSE); LMS Estimates; MM-estimates, GM-estimate; etc).
- Develop the two-stage estimation technique for some selected r robust methods and compare them.

REFERENCES

[1] P. Oteng-Darko, S. Yeboah, S. N. T. Addy, S. Amponsah, and E. O. Danquah, "*Crop modeling: A tool for agricultural research–A review*," 2013.

[2] M. Sekar, "Crop Simulation Model - *Types and Its Application in Agriculture*," 2023, pp. 29–37.

[3] D. A. Kasampalis, T. K. Alexandridis, C. Deva, A. Challinor, D. Moshou, and G. Zalidis, "*Contribution of remote sensing on crop models: a review*," J. Imaging, vol. 4, no. 4, p. 52, 2018.

[4] J. M. G. Dait, "Impact of Climate Change and Economic Activity on Philippine Agriculture: A Cointegration and Causality Analysis," Univers J Agric Res, vol. 10, no. 4, pp.

405–416, 2022.

[5] W. Mendenhall and T. Sincich, "*A Second Course in Statistics: Regression Analysis*," J. Am. Stat. Assoc., vol. 92, Jun. 1997, doi: 10.2307/2965740.

[6] P. N. Kephe, K. K. Ayisi, and B. M. Petja, "*Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa*," Agric. Food Secur., vol. 10, pp. 1–24, 2021.

[7] C. Vannier, T. A. Cochrane, P. Zawar Reza, and L. Bellamy, "*An analysis of agricultural systems modelling approaches and examples to support future policy development under disruptive changes in New Zealand*," Appl. Sci., vol. 12, no. 5, p. 2746, 2022.

[8] M. Reynolds et al., "*Role of modelling in international crop research: overview and some case studies*," Agronomy, vol. 8, no. 12, p. 291, 2018.

[9] S. K. Jha et al., "*Sugarcane yield prediction using vegetation indices in Northern Karnataka, India*," Univ J Agric Res, vol. 10, no. 6, pp. 699–721, 2022.

[10] A. F. Ramsey, "*Probability Distributions of Crop Yields: A Bayesian Spatial Quantile Regression Approach*," Am. J. Agric. Econ., vol. 102, no. 1, 2020, doi: 10.1093/ajae/aaz029.

[11] D. Timlin, K. Paff, and E. Han, "*The role of crop simulation modeling in assessing potential climate change impacts*," Agrosystems, Geosci. Environ, vol. 7, no. 1, p. e20453, 2024.

[12] J. W. Jones et al., "*Brief history of agricultural systems modeling*," Agric. Syst., vol. 155, pp. 240–254, 2017.

[13] R V Hogg and P. J. Huber, "Robust Statistical Procedures.," J. Am. Stat. Assoc., vol. 73, no. 364, 1978, doi: 10.2307/2286307.

[14] L. Wang, C. Zheng, W. Zhou, and W.-X. Zhou, "*A New Principle for Tuning-Free Huber Regression*," Stat. Sin., 2021, doi: 10.5705/ss.202019.0045.

[15] W. H. Nugroho, N. W. S. Wardhani, A. A. R. Fernandes, and Solimun, "*Robust regression analysis study for data with outliers at some significance levels*," Math. Stat., vol. 8, no. 4, 2020, doi: 10.13189/ms.2020.080401.

[16] R. Chen and I. C. Paschalidis, "*A robust learning approach for regression models based on distributionally robust optimization*," J. Mach. Learn. Res., vol. 19, 2018.

[17] H. Draper, N. R., & Smith, *Applied regression analysis*. USA: John Wiley & Sons, Inc., 1998.

[18] M. Hubert and P. J. Rousseeuw, "*Robust regression with both continuous and binary regressors*," J. Stat. Plan. Inference, vol. 57, no. 1, 1997, doi: 10.1016/s0378-3758(96)00041-9.

[19] H. Saulo, R. Vila, G. V. Borges, M. Bourguignon, V. Leiva, and C. Marchant, "*Modeling Income Data via New Parametric Quantile Regressions: Formulation, Computational Statistics, and Application*," Mathematics, vol. 11, no. 2, 2023, doi: 10.3390/math11020448.

[20] J. Alejo and G. Montes-rojas, "*Quantile Regression under Limited Dependent Variable*," pp. 1–38, 2021.

[21] J. H. Lee and Y. Shin, "COMPLETE SUBSET AVERAGING FOR QUANTILE REGRESSIONS," Econom. Theory, vol. 39, no. 1, 2023, doi: 10.1017/S0266466621000402.

[22] P. Iuliia et al., "*Measuring heterogeneity in hospital productivity: a quantile regression approach*," J. Product. Anal., vol. 59, no. 1, pp. 474–490, Feb. 2023, doi: 10.1007/s11123-022-00650-3.

[23] N. Nwakuya, M. T, and O. B. O, "*Quantile Regression-based Multiple Imputation of Skewed Data with Different Percentages of Missingness*," Sch. J. Physics, Math. Stat., vol. 9, no. 4, 2022, doi: 10.36347/sjpms.2022.v09i04.002.

[24] L. Merlo, L. Petrella, N. Salvati, and N. Tzavidis, "*Marginal M-quantile regression for multivariate dependent data*," Comput. Stat. Data Anal., vol. 173, 2022, doi: 10.1016/j.csda.2022.107500.

[25] X. Irz, L. Lin, C. Thirtle, and S. Wiggins, "*Agricultural productivity growth and poverty alleviation*," Dev. Policy Rev., vol. 19, no. 4, 2001, doi: 10.1111/1467-7679.00144.

[26] C. Bravo-Ortega and D. Lederman, "*Agricultural productivity and its determinants: revisiting international experiences*," Estud. Econ., vol. 31, no. 2, 2004.

[27] A. Y. Pavlov and J. V. Karmyshova, "*Identification of the factors of sustainable development of regional agricultural systems using regression model*," Int. J. Econ. Financ. Issues, vol. 5, no. 2, 2015.

[28] J. Zwolak, "*Regression relationship between production and factors of production in Polish agriculture*," Bulg. J. Agric. Sci., vol. 22, no. 6, 2016.

[29] A. Diana and S. Corraya, "*Regression based Price and Yield Prediction of Agricultural Crop*," Int. J. Comput. Appl., vol. 152, no. 5, 2016, doi: 10.5120/ijca2016911804.

[30] M. A. Zeru, "*Assessment of Factors That Affect the Performance of Agricultural Production, in the Case of Amhara Region, Ethiopia*," Agric. Sci., vol. 09, no. 08, 2018, doi: 10.4236/as.2018.98073.

[31] Y. Sheng, J. Ding, and J. Huang, "*The Relationship between Farm Size and Productivity in Agriculture: Evidence from Maize Production in Northern China*," American Journal of Agricultural Economics, vol. 101, no. 3. 2019. doi: 10.1093/ajae/aay104.

[32] L. J. Young, "*Agricultural crop forecasting for large geographical areas*," Annual Review of Statistics and Its Application, vol. 6. 2019. doi: 10.1146/annurev-statistics-030718-105002.

[33] S. Deepa, A. Alli, Sheetac, and S. Gokila, "*Machine learning regression model for material synthesis prices prediction in agriculture," in Materials Today: Proceedings*, 2021. doi: 10.1016/j.matpr.2021.04.327.

[34] W. Fitawek and S. L. Hendriks, "*Evaluating the impact of large-scale agricultural investments on household food security using an endogenous switching regression model*," Land, vol. 10, no. 3, 2021, doi: 10.3390/land10030323.

[35] A. Hayat, M. Amin, S. Afzal, A. H. Muse, O. M. Egeh, and H. S. Hayat, "*Application of Regression Analysis to Identify the Soil and Other Factors Affecting the Wheat Yield*," Adv.

Mater. Sci. Eng., vol. 2022, 2022, doi: 10.1155/2022/7793187.

[36] Irmeilyana, Ngudiantoro, M. N. Samsuri, and B. Suprihatin, "*Logistic regression model on land productivity of Pagar Alam coffee farming*," in Journal of Physics: Conference Series, 2021. doi: 10.1088/1742-6596/1943/1/012135.

[37] J. Spicka, T. Hlavsa, K. Soukupova, and M. Stolbova, "Approaches to estimation the farm-level economic viability and sustainability in agriculture: A literature review," *Agricultural Economics* (Czech Republic), vol. 65, no. 6. 2019. doi: 10.17221/269/2018-AGRICECON.

[38] E. Ustaoglu, M. E. Kabadayi, and P. J. Gerrits, "*The estimation of non-irrigated crop area and production using the regression analysis approach: A case study of Bursa Region (Turkey) in the mid-nineteenth century*," PLoS One, vol. 16, no. 4 April, 2021, doi:10.1371/journal.pone.02510 91.

[39] J. Wójcik, P. Leń, A. Doroż, and B. Kępowicz, "Hierarchisation of needs with regard to land consolidation works in the villages of southern Poland on the example of the Haczów municipality," Geomatics, *Landmanagement Landsc*., vol. 3, pp. 31–43, Jan. 2022, doi: 10.15576/GLL/2022.3.31.

[40] P. Iuliia, A. Natalya, T. Tatyana, and B. Natalia, "Regression Models of Agricultural Production: Evaluation of Data on Russian Regions," in Lecture Notes in Networks and Systems, 2023. doi: 10.1007/978-3-031-08954-1_71.

[41] L. Xie, H. He, J. Han, and G. Wen, "*Regression Analysis on the Influencing Factors of Total Value of Agricultural Output in Sichuan Province*," 2023. doi: 10.4108/eai.18-11-2022.2326763.

[42] H. Yu, Y. Shen, L. Yang, and Y. Nie, "*Robust M-estimation using the equivalent weights constructed by removing the influence of an outlier on the residuals*," Surv. Rev., vol. 51, no. 364, 2019, doi: 10.1080/00396265.2017.1360436.

[43] S. Alshqaq, A. Abuzaid, and A. Ahmadini, "*Robust estimators for circular regression models*," J. King Saud Univ. - Sci., vol. 33, no. 7, 2021, doi: 10.1016/j.jksus.2021.101576.

[44] K. G. van den Boogaart, P. Filzmoser, K. Hron, M. Templ, and R. Tolosana-Delgado, "*Classical and Robust Regression Analysis with Compositional Data*," Math. Geosci., vol. 53, no. 5, 2021, doi: 10.1007/s11004-020-09895-w.

[45] D. Wasim, S. A. Khan, and M. Suhail, "*Modified robust ridge M-estimators for linear regression models: an application to tobacco data*," J. Stat. Comput. Simul, vol. 93, no. 15, 2023, doi: 10.1080/00949655.2023.2202913.

[46] J. Yoon and A. F. Galvao, "*Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects*," Quant. Econom., vol. 11, no. 2, 2020, doi: 10.3982/qe802.

[47] F. Tian, B. Wu, H. Zeng, G. R. Watmough, M. Zhang, and Y. Li, "*Detecting the linkage between arable land use and poverty using machine learning methods at global perspective*," Geogr. Sustain., vol. 3, no. 1, pp. 7–20, Mar. 2022, doi: 10.1016/j.geosus.2022.01.001.

[48] F. Yahya and C.-C. Lee, "The Asymmetric Effect of Agriculturalization Toward Climate Neutrality Targets," SSRN Electron. J., 2022, doi: 10.2139/ssrn.4145490.

[49] X. Wang, G. Qin, X. Song, and Y. Tang, "*Censored quantile regression based on multiply robust propensity scores*," Stat. Methods Med. Res., vol. 31, no. 3, 2022, doi: 10.1177/09622802211060520.

[50] A. İ. Akgun, S. P. Türkoğlu, and S. Erikli, "*Investigating the determinants of happiness index in EU-27 countries: a quantile regression approach*," Int. J. Sociol. Soc. Policy, vol. 43, no. 1–2, 2023, doi: 10.1108/IJSSP-01-2022-0005.

[51] Z. Cai, H. Chen, and X. Liao, "*A new robust inference for predictive quantile regression*," J. Econom., vol. 234, no. 1, 2023, doi: 10.1016/j.jeconom.2021.10.012.

[52] M. Hao, Y. Lin, G. Shen, and W. Su, "*Nonparametric inference on smoothed quantile regression process*," Comput. Stat. Data Anal., vol. 179, 2023, doi: 10.1016/j.csda.2022.107645.

[53] B. Kai, M. Huang, W. Yao, and Y. Dong, "*Nonparametric and Semiparametric Quantile Regression via a New MM Algorithm*," J. Comput. Graph. Stat., 2023, doi: 10.1080/10618600.2023.2184374.

[54] C. Liu, G. Wang, X. Guan, and C. Huang, "*Robust M-estimation-based maximum correntropy Kalman filter*," ISA Trans., 2022, doi: 10.1016/j.isatra.2022.10.025.

[55] T. Nguyen-Tang, M. Yin, S. Gupta, S. Venkatesh, and R. Arora, "*On Instance-Dependent Bounds for Offline Reinforcement Learning with Linear Function Approximation*," in Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023, 2023. doi: 10.1609/aaai.v37i8.26116.

[56] M. Suhail, S. Chand, and M. Aslam, "*New quantile based ridge M-estimator for linear regression models with multicollinearity and outliers*," Commun. Stat. Simul. Comput., vol. 52, no. 4, 2023, doi: 10.1080/03610918.2021.1884715.

[57] Y. Zhang, J. Jiang, and Y. Feng, "*Penalized quantile regression for spatial panel data with fixed effects*," Commun. Stat. - Theory Methods, vol. 52, no. 4, 2023, doi: 10.1080/03610926.2021.1934028.

[58] S. S. Paul, B. Heung, and D. H. Lynch, "*Modeling of total and active organic carbon dynamics in agricultural soil using digital soil mapping: a case study from Central Nova Scotia*," Can. J. Soil Sci., vol. 103, no. 1, 2023, doi: 10.1139/cjss-2022-0012.

[59] W. W. Wang, P. Yu, L. Lin, and T. Tong, "*Robust estimation of derivatives using locally weighted least absolute deviation regression*," J. Mach. Learn. Res., vol. 20, pp. 1–49, 2019.

[60] D. et all Montgomery, "Introduction to Linear Regression Analysis," WILEY Ser. Probab. Stat., vol. 5, 2012.

[61] K. Yu, Z. Lu, and J. Stander, "*Quantile regression: Applications and current research areas*," J. R. Stat. Soc. Ser. D Stat., vol. 52, no. 3, pp. 331–350, 2003, doi: 10.1111/1467-9884.00363.

[62] R. Koenker, "Quantile regression for longitudinal data," J. Multivar. Anal., vol. 91, no. 1, 2004, doi: 10.1016/j.jmva.2004.05.006.

[63] I. A. Canay, "A simple approach to quantile regression for panel data," Econom. J., vol. 14, no. 3, 2011, doi: 10.1111/j.1368-423X.2011.00349.x.

[64] F. X. Diebold and R. S. Mariano, "*Comparing predictive accuracy*," J. Bus. Econ. Stat., vol. 13, no. 3, 1995, doi: 10.1080/07350015.1995.10524599.

[65] J. Zhou, H. Li, and W. Zhong, "*A modified Diebold–Mariano test for equal forecast accuracy with clustered dependence*," Econ. Lett, vol. 207, 2021, doi: 10.1016/j.econlet.2021.110029.