The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0002-1466.htm

Received 8 May 2024
Revised 30 June 2024
Accepted 30 June 2024

Implications of crop yield distributions for multiperil crop insurance rating in Ghana: a lasso model application

Kwame Asiam Addey
North Dakota State University, Fargo, North Dakota, USA, and
John Baptist D. Jatoe
Department of Agricultural Economics and Agribusiness, University of Ghana, Legon, Ghana

Abstract

Purpose – The objective of this paper is to examine crop yield predictions and their implications on MPCI in Ghana. Farmers in developing countries struggle with their ability to deal with agricultural risks. Providing aid for farmers and their households remains instrumental in combatting poverty in Africa. Several studies have shown that correctly understanding and implementing risk management strategies will help in the poverty alleviation agenda.
Design/methodology/approach – This study examines the importance of crop yield distributions in Ghana and its implication on multiperil crop insurance (MPCI) rating using the Lasso regression model. A Bonferroni test was employed to test the independence of crop yields across the regions while the Kruskal-Wallis H test was conducted to examine statistical differences in mean yields of crops across the ten regions. The Bayesian information criteria and k-fold cross-validation methods are used to select an appropriate Lasso regression model for the prediction of crop yields. The study focuses on the variability of the threshold yields across regions based on the chosen model.
Findings – It is revealed that threshold yields differ significantly across the regions in the country. This implies that the payment of claims will not be evenly distributed across the regions, and hence regional disparities need to be considered when pricing MPCI products. In other words, policymakers may choose to assign respective weights across regions based on their threshold yields.
Research limitations/implications – The primary limitation is the unavailability of regional climate data which could have helped in a better explanation of the variation across the regions.
Originality/value – This is the first study to examine the implications of regional crop yield variations on multiperil crop insurance rating in Ghana.

Keywords Crop insurance, lasso regression, MPCI, Risks, Threshold yield, Planting for food and jobs, Machine learning
Paper type Research paper

1. Introduction

Managing risk continues to be a challenge in every endeavor of life. The risks are ever-increasing in agriculture, given the changing climate and environmental conditions. This makes it difficult for farmers and insurers in their businesses (Breustedt *et al.*, 2008). Furthermore, the spatial covariate nature of agriculture risks sometimes makes the magnitude of losses significant. The success of agricultural risk management programs depends on understanding the nature of risks being managed – distribution of crop yields (Yang *et al.*, 1992; Ramsey, 2020; Addey *et al.*, 2023). Previous studies (Duarte *et al.*, 2017; Jensen *et al.*, 2018; Park *et al.*, 2019) have shown that asymmetric information problems

JEL Classification — C52, C53, O10, O55
The authors are grateful to two anonymous referees for very useful comments on an ealier version of the paper. Any remaining errors are entirely the authors' alone.

0002-1466
DOI 10.1108/AFR-05-2024-0078

(adverse selection and moral hazard) are significant challenges to the success of insurance programs. This makes the information on the distribution of yields across space and over time crucial for crop insurance programs.

The objective of this paper is to examine crop yield predictions and their implications on MPCI in Ghana. In developing countries where the government's ability to subsidize premiums is marginal, farmers are more likely to survive these risks if insurance companies are viable. In a review of the literature on index insurance for developing countries, Miranda and Farrin (2012) iterated that the inability of farmers in developing countries to manage agricultural related risk properly makes the objective of poverty reduction difficult. The Ghana Agricultural Insurance Pool (GAIP), in collaboration with the German Society for International Cooperation (GIZ) and some Ghanaian insurance companies, has championed the course of agricultural insurance in Ghana over the past decade.

The four categories of agricultural insurance under their program are drought index insurance, MPCI, poultry insurance, and area-yield insurance. The MPCI protects against yield losses by allowing producers to insure a certain percentage of historical crop production. It is advantageous because a single policy protects farms against all perils. Given the eclectic nature of agricultural risks and that most farmers in Ghana practice mixed cropping, farmers are likely to benefit significantly from such a scheme. Recent studies have shown that Ghanaian farmers are interested in acquiring crop insurance (Adjabui *et al.*, 2019; Addey *et al.*, 2021). This prognosis presents an opportunity for the insurance industry to invest in the agricultural sector. However, Barnett (2014) stated that MPCI has had a checkered history in countries where it was introduced as most of the programs were eventually discontinued due to poor actuarial performance. In addition, yield distributions vary over time and space with many moments (Ramsey, 2020). This bleak history of MPCI casts a doubt on its expected success in Ghana.

On grounds of actuarial performance, one would begin with the assumption of heterogenous yields across space (regions). However, the government of Ghana has been implementing a fertilizer subsidy program since 2008. Although the universal subsidy was changed in 2013 to officially target smallholder farmers with subsidized input that is sufficient to cover not more than 2 hectares of farmland, it allocates a fixed quantity of fertilizer to beneficiaries irrespective of location and specific soil-plant requirements (Pauw, 2021). In addition, Houssou *et al.* (2017) noted that despite these reforms of 2013, many recipients of subsidized fertilizer were larger-scale, wealthier farmers. Given such evidence of government's reliance on uniform farm support programs we contend that policy makers, analysts and insurance companies will need strong evidence on the implications of spatially varying yields to justify non-uniform crop insurance policies, especially where such programs require government's direct support.

To overcome these challenges, this study uses a Lasso regression model to model the probability distribution of crop yields in Ghana. Quantifying the degree of systemic risk is essential for computing an actuarially fair premium which motivates farmers to purchase insurance and keep insurers viable. However, these systemic risks vary by region and geographical location due to variations in climate and agronomic characteristics. One way to capture these variations is by including these regions as covariates. This can lead to having several covariates, which causes the ordinary least squares (OLS) model to be overfitted with misleading parameter coefficients and significance levels. The Lasso model is appealing under conditions where the model could be overfitted because it is a straightforward extension of the linear regression and can shrink the regression coefficients and automatically performs variable selection by setting some coefficients to zero (Ahrens *et al.*, 2020). This is done through imposing a regularization penalty aimed at limiting model complexity, improving predictive performance, and enhancing parsimony.

This article further investigates the implications of these yield distributions on MPCI rating. We evaluate the variability of threshold (trigger) yields across the regions of Ghana and examine if they are substantially different from the overall national threshold yields. Threshold yields are essential for MPCI because they determine the yield levels at which the payment of claims could be triggered. A variation of threshold yields across regions implies that the risk of claims will also vary, potentially reducing the uniformity of the program across regions. Our simulations show, for instance, that using a 70% coverage level the threshold yields for the Volta region could be 68.8% higher than the national threshold. In comparison, that of the Western region is 38.4% lower than the national threshold. Such a problem, if neglected, may cause insurance providers to focus on regions where the risks of claims are lower. On the other hand, any government supported MPCI program would put insurance providers in the regions with higher risk at a disadvantage with likely more frequent and or larger claims.

We use crop yield data for six field crops across all districts in the ten regions from 1993 to 2017. Considering the spatial variation of weather, soil, and other conditions among the ten regions, the fundamental question this study seeks to address is whether predicted yields are similar or dissimilar across the regions. The analysis of spatiotemporal data presents two econometric issues; spatial heterogeneity and spatial autocorrelation (Farzammehr *et al.*, 2020). Spatial heterogeneity of crop yields is mainly due to differences in climate among regions and time points. On the other hand, spatial autocorrelation arises due to spatial dependence due to similarities of crop growing conditions between neighboring regions. The ten administrative regions of Ghana are nested within six agroecological zones. This presents a natural question of whether the yield risks will differ substantially if modeled independently by region or jointly for the whole country.

The contribution of this study is twofold. Primarily, this is the first study to examine the implications of regional crop yield variations on multiperil crop insurance rating in Ghana. Hence, it presents an opportunity for farmers, government, and major stakeholders to understand the distribution of crop yields and examine its implications on MPCI rating. There is an existing hypothesis that trust by farmers could help in the uptake of insurance products (Stein, 2018). While we do not formally test this hypothesis, we believe that farmers' understanding of these concepts through studies like this will enable them to re-evaluate their risk-mitigating strategies and decide which best suits their farming business. Secondly, this study will provide the government with appropriate information to enhance the sustainability of the 'planting for food and jobs program.' In the government of Ghana's quest to reduce poverty, it introduced the 'planting for food and jobs program' in 2017. To make this program sustainable, it is ideal to have a comprehensive risk-mitigating strategy. However, there has not been a profound risk-mitigating strategy to our knowledge. An understanding of crop yield distributions will enable the government to make relevant decisions regarding the success and sustainability of agricultural insurance.

The rest of the paper is organized as follows. Section 2 presents a brief review of the literature related to the modeling of crop yields. In Section 3, we discuss the empirical framework of the study. This section presents the framework of the Lasso regression model, the data employed for the study, and its sources. The results are presented in Section 4, while the conclusions are presented in Section 5.

2. Literature review

Farmers in developing countries struggle with their ability to deal with agricultural risks. Providing aid for farmers and their households remains instrumental in combatting poverty in Africa. Several studies have shown that correctly understanding and implementing risk management strategies will help in the poverty alleviation agenda. According to Karlan *et al.* (2014), there is a vicious cycle between agricultural risks and poverty in Ghana. They explained that farmers in developing countries fail to invest in potentially profitable inputs due to their wariness of the risk involved in adopting new technologies. For instance, if they invest and their crops fail, they may not have enough capital for the next season. Meanwhile, not investing in innovations will likely result in low crop yields. Agricultural insurance remains a risk-mitigating strategy that encourages farmers to invest more in production and promotes their confidence in adopting new and improved farming technologies (Gehrke, 2019; Addey *et al.*, 2021).

Miranda and Farrin (2012) reviewed the theoretical and empirical research on index insurance for developing countries. In summarizing the lessons learned, one striking point was that efforts to design contract indemnity schedules to correlate indemnities with policyholder losses are frustrated by the scarcity of farm-level yield and loss data. They indicated that most of the existing models to predict crop yields have been unable to explain yield variability at the farm level, where production is subject to idiosyncratic shocks that are diversified away in aggregate yields because of data limitations. This makes the prediction models for crop yields a vital issue for crop insurance. Prior to this study, Goodwin and Mahul (2004) identified the key challenges that arise in the design and rating of crop yield insurance plans with an emphasis on production risk modeling. They showed that the availability of data affects insurance and rate-making procedures. This challenge causes most studies to be based on aggregate data and assumed probability distributions of crop yields.

There have been several studies on the probability distribution of crop yields over the years (Atwood *et al.*, 2003; Shaik, 2010; Harri *et al.*, 2011; Annan *et al.*, 2014; Duarte *et al.*, 2017; Ramsey, 2020). Existing studies have adopted diverse methods to address the pivotal issues associated with yield predictions – heteroskedasticity and spatial dependence. Yang *et al.* (1992) examined the effect of three alternative models on wheat yield. These models include the time trend, a generalized autoregressive conditional heteroskedasticity (GARCH) process, and an econometric model that includes the potential sources of heteroskedasticity. Based on non-nested test results, their findings revealed that modeling the sources of heteroskedasticity is the preferred procedure. While the authors suggest that identifying the sources of heteroskedasticity may not be too difficult, we argue that this assumption will likely fail under conditions when multiple crops are being analyzed. Increasing the number of crops and spatial components of the yield data could primarily lead to aggregation bias issues. Other studies have addressed issues of heteroskedasticity with different methods (Harri *et al.*, 2011; Goodwin and Hungerford, 2015; Park *et al.*, 2019; Liu and Ker, 2020). Among the numerous methods, the approach proposed by Harri *et al.* (2011) has been adopted by the U.S. Department of Agriculture's (USDA) Risk Management Agency (RMA). It employs a process of normalization of crop yield distributions based on two-stage applications to correct yields for heteroskedasticity. This method models the changes in temporal and spatial variations using a two-knot linear spline with estimators that are robust to outliers.

Park *et al.* (2019) considered using Bayesian kriging to smooth the parameters of generalized Pareto distribution (GPD). In addition, the GPD was employed to fit the tail of crop yield distributions. This model provides estimates of the spatial structure across U.S. regions in the similitude of the maximum distance of the spatial effects. According to the authors, the Bayesian kriging method is advantageous over other spatial smoothing methods. Furthermore, the GPD allows flexible forms of the tail of the distribution and hence provides a higher likelihood of accurate tail probability. The study concluded, based on an out-of-sample performance game between a private insurance company and the federal agency, that the proposed model was a considerable improvement, especially when rating deeper tail probability.

Ramsey (2020) examined the effect of weather across U.S. corn and soybean yield distribution quantiles using a semiparametric Bayesian spatial quantile regression for the conditional distribution of yields. The contribution of this study was to overcome the difficulty in modeling yield distributions emanating from their time-varying nature and indeterminate parametric forms. Measuring the effects of weather across quantiles, the author found that the adverse effects of extreme heat are the largest in the lower tail of the conditional yield distribution. It was further found that the effects of weather varied spatially. In addition, the study estimated crop insurance premium rates based on the Bayesian spatial quantile regression and compared them to other existing methods for modeling time-varying yield distributions.

Liu and Ker (2020) employed a nonparametric Bayesian model averaging (BMA) to incorporate extraneous information into estimated premium rates of corn, soybean, cotton, and winter wheat. This method is also advantageous because it does not make any assumptions about the parametric form or similarities of yield distributions. Furthermore, it decreases error and enables statistically significant and economically essential rents to be captured. The authors concluded that the nonparametric BMA estimator is more efficient at estimating premium rates than the current RMA methodology or individual kernel method. In addition, the efficiency gains with BMA are in small samples, with relatively little information in the individual sample. Their results were robust to restrictions on the distance of spatial smoothing.

From the crop insurance and yield distribution literature reviewed, we find a common theme of the recurring accuracy of crop yield prediction and its impact on the success of crop insurance programs. Evidence from the well-established U.S. federal crop insurance programs suggests that policy reforms and methodological improvements can alleviate some of the challenges associated with crop insurance. From the policy perspective, subsidies have played a vital role in developing the U.S. crop insurance sector (Tack *et al.*, 2018; Yu *et al.*, 2017). However, understanding the distributions of yields is crucial for such decisions. Issues of heteroskedasticity and spatial dependence of crop yields have been addressed significantly. Despite this, the literature on crop insurance and yield distributions in developing countries like Ghana remains limited. Meanwhile, the prospects for the agricultural sector in Ghana is significantly being hampered by risks (Addey, 2018; Addey *et al.*, 2021).

The present paper differs from other studies by analyzing crop yields across regions in Ghana over 25 years using the Lasso regression model. A consensus exists that nonparametric methods may yield better results when predicting crop yields. Typically, the use of crop yields with short periods may provide inconclusive results regarding the choice of optimal distribution and hence may be problematic for nonparametric estimators with slow convergence rates (Ramsey, 2020). Using the least absolute shrinkage and selection operator (lasso) regression model, a parametric method, is beneficial because the bias-variance trade-off leads to better predictions.

3. Methodology and data

To model the distribution of different crop yields across spatially varying regions, we require a model that will account for the spatial and temporal variation of the crop yields. Given that we have six crops across ten regions over 25 years, interactions among these will lead to many predictors. Typically, nonparametric density methods are preferred to parametric prediction methods. However, Goodwin (2008) cautions against using nonparametric methods when data history is limited. Thus, we forgo the nonparametric approaches and instead employ a machine-learning technique to predict crop yields. Machine learning has grown appealing across various scientific disciplines, including economics and applied econometrics (Kleinberg *et al.*, 2018). The regularized linear regression is a machine learning technique that is often regarded as an extension of the standard ordinary least squares (OLS) method. Their similarity is that both minimize the sum of squared deviations between the observed and predicted dependent variables. The extension involves the imposition of a regularized penalty to limit model complexity.

Researchers and data analysts, particularly in insurance, are primarily interested in prediction accuracy. However, the standard OLS estimates minimize the residual squared error and are characterized by low bias but large variance. In this article, we use the Lasso regression model, which employs the L_1-norm of penalizing coefficients for modeling the yield distribution. Several penalization techniques have been proposed to improve the OLS (Zou and Hastie, 2005). One such method is the ridge regression method. Despite its efficacy in prediction through a bias-variance trade-off achieved from minimizing the residual sum of squares subject to a bound on the L_2-norm [1], the ridge regression method fails to produce a parsimonious model.

3.1 The Lasso regression model

Tibshirani (1996) introduced the Lasso regression model as an innovative variable selection method for regression and a well-known sparse regression method that regularizes the parameter $\boldsymbol{\beta}$ under the sparse assumption. It minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. There are diverse mathematical representations of this model. This study follows the formulation by Wang *et al.* (2018). Consider a simple random sample of size N. We wish to model the response of y_i as a function of p covariates, $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$, which is a vector for the i^{th} case $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$. Given this problem, the standard Lasso regression model is given as:

$$\arg \min_{\beta_0, \beta \in \mathcal{R}^p} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^T \boldsymbol{\beta}\right)^2$$

$$s.t. \sum\nolimits_{j=1}^{P} |\beta_j| \le t \tag{1}$$

where $t \ge 0$ is the upper bound for the sum of the coefficients. It is a pre-specified free parameter that determines the amount of regularization. Under conditions where t is large, all the coefficients are shrunk to almost zero, while smaller values of t allow Lasso to shrink some of the estimated coefficients to zero. Suppose that $\mathbf{X}$ represents the $N \times p$ covariate matrix and $\boldsymbol{y}$ is the response vector, equation 1 can be written in a compact form as

$$\arg \min_{\beta_0, \beta \in \mathcal{R}^p} \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{y} - \beta_0 \boldsymbol{I} - \boldsymbol{X}\boldsymbol{\beta}^T \right\|_2^2$$

$$s.t. \|\boldsymbol{\beta}\|_1 \le t \tag{2}$$

Prior knowledge of $\widehat{\beta}_0 = \bar{y} - \bar{x}^T \boldsymbol{\beta}$ warrants:

$$y_i - \widehat{\beta}_0 - x_i^T \boldsymbol{\beta} = y_i - \left(\bar{y} - \bar{x}^T \boldsymbol{\beta}\right) - x_i^T \boldsymbol{\beta} \tag{3}$$

Simplification of equation (3) yields

$$\left(y_i - \bar{y}\right) - \left(x_i - \bar{x}\right)^T \boldsymbol{\beta} \tag{4}$$

Based on equation (4), equation (2) can be rewritten as

$$\arg \min_{\boldsymbol{\beta} \in \mathcal{R}^p} \left\{ \frac{1}{N} \left\| y - \boldsymbol{X}\boldsymbol{\beta}^T \right\|_2^2 \right\}$$

$$s.t.\ \|\boldsymbol{\beta}\|_1 \leq t \tag{5}$$

Following Zou (2006) and Wang *et al.* (2018), we write the Lasso estimator in a Lagrangian form

$$L(\beta, \lambda) = \min_{\boldsymbol{\beta} \in \mathcal{R}^p} \left\{ \frac{1}{N} \left\| y - \boldsymbol{X}\boldsymbol{\beta}^T \right\|_2^2 + \lambda \|\boldsymbol{\beta}\| \right\} \tag{6}$$

where $\lambda \geq 0$ is the tuning parameter and controls the empirical error and the sparsity of the model parameter. The relationship between λ and the upper bound t is data dependent.

3.2 Empirical model specification

We first consider the Harri *et al.* (2011) model for normalizing crop yield observations over time. Following Liu and Ker (2020), the model assumes y_t for crop yield in time t as

$$y_t = \theta_1 + \theta_2 t + \alpha_1 d_1 (t - k_1) + \alpha_2 d_2 (t - k_2) + \varepsilon_t \tag{7}$$

where $d_1 = 1$ if $t \geq k_1$ and $d_2 = 1$ if $t \geq k_2 \in (1 + \bar{k}, \ldots, T - \bar{k})$, and $k_2 - k_1 \geq k$. The $k, \bar{k} \geq 10$ are *a priori* imposed bounds that prevent the knots from locating too close to each other ($\underline{k}$) or to the endpoints $\bar{k}$. The k_i are the knot locations and selected from least squares criterion (grid search). The degree of heteroskedasticity in the residuals is estimated using the method by Harri *et al.* (2011). The RMA then combines the prediction from the spline model with the heteroskedasticity-adjusted residuals from their temporal model to recover the assumed *i.i.d* yields from the required conditional yield density.

The vital concern for the empirical framework is to test whether crop yields, and trend parameters are spatially invariant across regions in Ghana. Having obtained the heteroskedasticity-adjusted predicted crop yields, we formally test this with a Lasso estimating equation which allows for region-specific and crop-specific parameters. To perform the analysis, a set of assumptions are required. These are the independence of crop yields by regions and statistical differences in mean yields of crops across regions. Testing for the significance of differences in the mean yields of crops across regions leads to many simultaneous t-tests. Thus, we performed a one-way Bonferroni test to examine the independence of crop yields across the regions. The Bonferroni test is a multiple comparison test used for several dependent or independent statistical tests on a single dataset. The fundamental problem with multiple comparisons is the possibility of errors or false positives. Using this test helps prevent data from incorrectly appearing statistically significant. Alternatives to this test include Scheffe's test and the Tukey-Kramer method test.

3.3 Data and descriptive statistics

This study focuses on crop yield distributions in Ghana. The dataset employed is annual data, at the district-level, from 1993 to 2017 on six crops for the ten regions of Ghana: Ashanti, Brong Ahafo, Central, Eastern, Greater Accra, Northern, Upper East, Upper West, Volta, and Western. The crops include cassava, cocoyam, maize, plantain, rice, and yam. To minimize the effects of outliers, the data was winzorized between 1 and 99% following Yao *et al.* (2022). The crop yield data was collected by the Ministry of Food and Agriculture (MOFA) and is freely available on Ghana's open data initiative website [2].

The descriptive statistics of the yields for the six crops are presented in Table 1. From the table, cassava has an average yield of 14.11 tons/ha with a standard deviation of 4.54 tons/ha. Cocoyam and maize have average yields of 6.53 tons/ha and 1.59 tons/ha, with standard deviations of 1.53 tons/ha and 0.38 tons/ha. Plantain, rice, and yam have average yields of 8.35 tons/ha, 1.98 tons/ha, and 11.21 tons/ha with 2.15 tons/ha, 0.92 tons/ha, and 4.15 tons/ha as their respective standard deviations. Other statistics, including the 25th percentile (p25), 50th percentile (p50), 75th percentile, minimum, and maximum, are presented in the table.

Table 2 further presents the descriptive statistics of the crop yields by regions. This table shows that the different regions have variations in yields for the diverse crops. For instance, it can be seen average maize yields were 1.83 tons/ha and 1.95 tons/ha in the Brong Ahafo and

Crop	N	Mean	Std Dev	p25	p50	p75	Min	Max
Cassava	7083	14.11	4.54	10.39	13.85	16.28	5.00	41.00
Cocoyam	5232	6.23	1.53	5.12	6.10	7.00	2.27	11.50
Maize	6720	1.59	0.38	1.32	1.60	1.80	0.38	3.45
Plantain	5520	8.35	2.15	6.46	8.52	9.73	0.95	19.43
Rice	5659	1.98	0.92	1.24	1.66	2.60	0.29	6.38
Yam	5985	11.21	4.15	8.00	11.00	14.52	2.75	27.20

Source(s): Authors' own creation/work

Table 1. Descriptive statistics of crop yields

Crop	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
	Ashanti region					Brong Ahafo region				
Cassava	1508	12.92	4.55	6.10	41.00	1438	16.31	2.92	9.99	29.60
Cocoyam	1503	7.32	1.54	4.20	11.50	1129	6.06	0.74	3.05	8.50
Maize	1474	1.49	0.23	0.90	2.73	1423	1.83	0.21	0.90	2.70
Plantain	1497	9.46	1.70	4.81	15.27	1253	8.50	2.66	0.95	19.43
Rice	1515	1.86	0.65	0.29	5.44	780	1.36	0.36	0.50	2.50
Yam	1502	11.38	1.98	4.89	18.36	1415	13.54	2.82	4.90	21.68
	Central Region					Eastern Region				
Cassava	584	14.41	3.81	8.64	32.00	792	16.04	5.07	9.90	34.50
Cocoyam	379	4.74	0.68	2.27	7.25	771	7.22	0.90	4.50	10.00
Maize	587	1.67	0.53	0.88	3.40	754	1.95	0.38	1.13	3.45
Plantain	577	7.04	2.06	1.51	13.84	742	9.09	0.89	6.84	11.01
Rice	375	1.60	0.65	0.90	5.00	528	2.37	0.71	1.39	4.40
Yam	551	5.50	1.03	2.95	11.41	738	16.25	1.92	7.41	22.50
	Greater Accra Region					Northern Region				
Cassava	307	19.76	6.34	12.67	38.20	665	11.50	4.15	5.00	25.60
Cocoyam	1	5.67		5.67	5.67	23	6.39	1.22	4.35	8.87
Maize	7	1.59	0.16	1.30	1.81	13	1.65	0.15	1.40	1.85
Plantain	6	8.50	1.86	5.83	11.42	13	7.88	0.95	6.36	9.26
Rice	5	1.55	0.20	1.27	1.81	19	1.85	0.45	1.20	2.64
Yam	6	10.53	2.86	6.37	14.51	13	11.20	1.78	8.71	14.34
	Upper East Region					Upper West Region				
Cassava	18	13.97	1.40	11.81	16.42	18	14.03	2.01	10.73	17.44
Cocoyam	9	6.33	0.70	5.62	7.73	14	6.50	0.88	5.04	7.74
Maize	462	1.18	0.38	0.38	2.29	333	1.37	0.39	0.50	3.42
Plantain	7	8.45	0.68	7.69	9.28	5	8.68	1.62	6.75	10.81
Rice	467	2.45	0.68	0.71	5.48	340	1.42	0.50	0.53	2.72
Yam	15	11.34	2.51	5.60	14.51	309	13.29	4.25	6.50	27.20
	Volta Region					Western Region				
Cassava	916	14.46	2.84	7.00	21.93	837	10.04	1.53	7.44	21.50
Cocoyam	595	4.81	1.48	2.59	10.68	808	5.20	0.67	3.35	7.59
Maize	869	1.56	0.29	0.73	2.45	798	1.33	0.13	0.96	2.25
Plantain	596	6.27	0.87	4.79	10.30	824	7.83	1.82	3.27	13.10
Rice	852	3.33	0.87	0.88	6.38	778	1.21	0.09	0.94	1.62
Yam	616	10.68	2.96	3.00	19.88	820	5.78	1.38	2.75	12.35

Source(s): Authors' own creation/work

Table 2. Crop yields across regions in Ghana (MT/Ha)

Eastern regions, respectively. On the other hand, low yields of 1.18 tons/ha and 1.33 tons/ha were observed for the Upper East and Western regions.

A kernel distribution of the crop yields for the six crops across Ghana is presented in Figure 1. The kernel density estimate is obtained from the summation of the weighted values of a kernel function K, and is generally given as $\widehat{f}_K = \frac{1}{qh}\sum_{i=1}^{n} w_i K\left(\frac{y - Y_i}{h}\right)$, where $q = \sum_i w_i$ if weights are frequency or analytic weights and $q = 1$ if the weights are importance weights. This estimator approximates the density of the yields, $f(y)$, from our yield observations y. We employ the Epanechnikov kernel density, which is the most efficient in minimizing the mean integrated squared error. The specific mathematical formulation for this kernel function is given as $K[z] = \begin{cases} \dfrac{\frac{3}{4}\left(1 - \frac{1}{5}z^2\right)}{\sqrt{5}} & \text{if } |z| < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$.

4. Empirical results

4.1 Statistical tests for regional yield distributions

To understand the implications of crop yield predictions on MPCI in Ghana, we first perform the statistical tests required to ensure the robustness of our results. The results of the one-way Bonferroni test are presented in Table 3. It consists of ANOVA tests for the crop yields across their cultivation regions and a Bartlett test for equal variances. The between-region

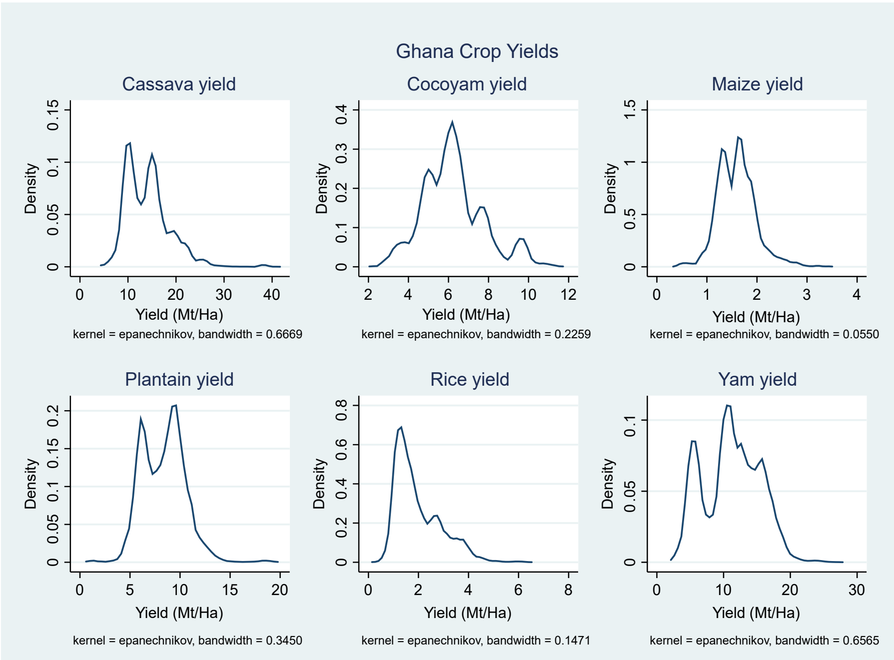


Figure 1.
Kernel distributions of crop yields in Ghana

Source(s): Authors' own creation/work

Source	SS	ANOVA test DF	MS	F-Prob	Bartlett's test for equal variance Chi-Sq
Cassava					
Between	40469	9	4497	301.99***	1600***
Within	105314	7073	15		
Total	145783	7082	21		
Cocoyam					
Between	5487	9	610	475.26***	1300***
Within	6698	5222	1		
Total	12185	5231	2		
Maize					
Between	343	9	38	414.85***	1900***
Within	617	6710	0		
Total	961	6719	0		
Plantain					
Between	6078	9	675	192.04***	1400***
Within	19375	5510	4		
Total	25453	5519	5		
Rice					
Between	2676	9	297	793.08***	2900***
Within	2118	5649	0		
Total	4794	5658	1		
Yam					
Between	70125	9	7792	1405.73***	1400***
Within	33118	5975	6		
Total	103244	5984	17		

Source(s): Authors' own creation/work

Table 3. Bonferroni test for mean crop yields across regions

sum of squares for the cassava yield is 404,469 with 9 degrees of freedom. The corresponding F statistic is 301.99 and has a significance level of 1%. The test statistic for Bartlett's test of equal variances is 1600 at a significance of 1%. Hence, we reject the assumption that cassava yields are homogeneous across the ten regions. Cocoyam yields obtained a between-region sum of squares of 5487 with 9 degrees of freedom at 1% significance. The test statistic for its equality of variance is significant at 1%. Similarly, the table revealed a 1% significance for the Barttlet's test of equal variances for maize, plantain, rice and yam, implying heterogeneity of yields across the ten regions.

A Kruskal-Wallis H test was conducted to determine if the mean yields of the six crops were statistically different across the ten regions. For each of the crops, the test showed that there were statistical differences in the mean crop yields among the regions. Cassava had a $X_9^2 = 2562, p = 0.0001$. Cocoyam and plantain had $X_9^2 = 2747, p = 0.0001$ and $X_9^2 = 1519$, $p = 0.0001$ respectively. Maize, rice and yam had chi-square values of 2827, 3054 and 3954 respectively at 1% significance. Table 4 presents the results (regional rank sums, chi-square, and probability distributions) of the six crops by columns.

4.2 Regression outcomes and yield predictions

Following this, the crop yields are adjusted for heteroskedasticity based on Harri *et al.* (2011). The descriptive statistics of the heteroskedasticity-adjusted crop yields across the country are presented in Table 5. From the table, the average cassava, cocoyam, and maize heteroskedasticity-adjusted yields across the country are 18.26 tons/ha, 8.21 tons/ha, and 2.09 tons/ha, respectively, with standard deviations of 5.18 tons/ha, 2.24 tons/ha, and 0.47 tons/ha. The average heteroskedasticity-adjusted yields for plantain, rice, and yam across the country are 10.81 tons/ha, 2.58 tons/ha, and 14.54 tons/ha, respectively. These crops have standard

Table 4.
Kruskal-Wallis H test of difference in mean yields of crops across regions

Region	Cassava	Cocoyam	Maize	Plantain	Rice	Yam
Ashanti	4350000	5430000	4060000	5430000	4240000	4520000
Brong Ahafo	6990000	2920000	7060000	3570000	1320000	5670000
Central	2240000	335587	2010000	992260	811417.5	383002
Eastern	3430000	3010000	3820000	2540000	2030000	3790000
Greater Accra	1750000	1799.5	23839	17021	12004	15986
Northern	1460000	64002.5	50694	30728.5	57335	38210.5
Upper East	67712.5	25522	693698.5	19590.5	1850000	45606.5
Upper West	68616	43141.5	669431	15567	624820	1120000
Volta	3640000	719831	2780000	652620.5	4120000	1660000
Western	1090000	1140000	1410000	1970000	946370.5	669904.5
Chi-Sq (DF)	2562 (9)	2747 (9)	2827 (9)	1519 (9)	3054 (9)	3954 (9)
Probability	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Source(s): Authors' own creation/work

Table 5.
Descriptive statistics of heteroskedasticity-adjusted yields (HAY) in MT/Ha

Crop	N	Mean	Std Dev	p25	p50	p75	Min	Max
Cassava	7083	18.27	5.18	14.38	18.06	21.48	5.00	59.97
Cocoyam	5232	8.21	2.24	6.54	7.94	9.46	2.40	16.79
Maize	6720	2.09	0.47	1.75	2.08	2.39	0.55	4.46
Plantain	5520	10.81	2.30	9.05	11.00	12.52	1.04	24.74
Rice	5659	2.58	1.17	1.69	2.23	3.10	0.41	8.90
Yam	5985	14.54	5.07	10.83	14.77	17.93	3.44	38.62

Source(s): Authors' own creation/work

deviations of 2.30 tons/ha, 1.17 tons/ha, and 5.07 tons/ha, respectively. Other descriptive statistics, such as the percentiles, minimum and maximum, are presented in the table.

Having obtained these two [3] groups of yield, we then perform a Lasso regression model on them. We compare the prediction performance of four models based on the Bayesian information criteria (Schwarz, 1978; Ahrens *et al.*, 2020) and the K-fold cross-validation (Geisser, 1975; Ahrens *et al.*, 2020). The Bayesian information criteria, BIC $(\lambda,\alpha) = nlog\,(\widehat{\sigma}^2(\lambda,\alpha)) + df(\lambda,\alpha)\log(n)$, where $\widehat{\sigma}^2(\lambda,\alpha) = n^{-1}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2$ and $\widehat{\varepsilon}_i$ are the residuals. The effective degrees of freedom $(df(\lambda,\alpha))$ measures the model complexity. The rule of thumb is that the model with the least BIC is the best. The K-fold cross-validation assesses model performance by splitting the data into approximately equal K groups.

Consider that we have a data set with observations K_k in the kth fold and let n_k be the size of data partition k for $k = 1,\ldots,K$. The kth fold is treated as the validation data, while the $K-1$ folds are treated as the training data. An estimate $\widehat{\beta}_k(\lambda,\alpha)$, is generated based on all data except the observations in the validation data. To compare the performance of the models, a K-fold cross-validation estimate of a mean square prediction error (MSPE) is given by

$$\widehat{L}^{CV}(\lambda,\alpha) = \frac{1}{K}\sum_{k=1}^{K} MSPE_k(\lambda,\alpha) \tag{8}$$

$$\text{where } MSPE_k(\lambda,\alpha) = \frac{1}{n_k}\sum_{i\in K_k}\left(y_i - x_i'\widehat{\beta}_k(\lambda,\alpha)\right)^2 \tag{9}$$

From Table 6, Models 1 and 2 were conducted using the unadjusted crop yields as the dependent variable, while Models 3 and 4 were conducted with the heteroskedasticity-adjusted yields. In Model 2, the crop type and year interaction are considered a prediction, while it is not considered in Model 1. Similarly, this interaction is considered in Model 4, while it is not considered in Model 3. The BIC for Models 1, 2, 3, and 4 are 42.19, 20.05, 65.88, and 41.37, respectively. The rule of thumb indicates that the least BIC gives the best model fit. Therefore, Model 2 is the best model, given its low BIC value. Furthermore, we performed a 10-fold cross-validation to ascertain the validity of the BIC results. The MSPEs are 3.94, 3.10, 5.70 and 5.24 for Models 1, 2, 3, and 4, respectively.

Since Model 2 presents the least value for both BIC and MSPE, we selected this model for further analysis and discussion. We, therefore, predicted the crop yields based on Model 2 (the predicted yields for all four models are presented in Table 7). The average yield for cassava, cocoyam, maize, plantain, rice, and yam are 14.11 tons/ha, 6.23 tons/ha, 1.59 tons/ha, 8.35 tons/ha, 1.98 tons/ha, and 11.21 tons/ha, respectively. Their standard deviations are 3.29 tons/ha, 1.06 tons/ha, 0.26 tons/ha, 1.67 tons/ha, 0.76 tons/ha, and 3.69 tons/ha, respectively. The kernel densities from the predictions of the four models are presented in Figures 2–5.

The descriptive statistics for the predicted crop yield by regions based on Model 2 are presented in Table 8. Despite varying average yields across regions, the risks (measured by standard deviation) are normalized across regions for the crops. Table 9 presents a Kruskal-Wallis test which shows that the predicted yield for the respective crops differ across regions.

4.3 Implications of predicted crop yields on MPCI rating

The motivation for this study is to evaluate the implications of crop yield distributions for MPCI rating in Ghana using the Lasso regression model. Threshold (trigger) yields are essential for crop insurance rating. We evaluate threshold yields across the regions for the six crops as a proxy for the MPCI rating. To perform this task, we employ the formulation of threshold yield employed by Clarke *et al.* (2012).

$$Threshold\ Yield = Indemnity\ level \times Probable\ Yield \tag{10}$$

The indemnity levels are set at 70%, 80%, and 90%. Table 10 presents the results across the regions, crops, and coverage levels. The average threshold yield across the country is 9.88 tons/ha, 4.36 tons/ha, 1.11 tons/ha, 5.84 tons/ha, 1.38 tons/ha, and 7.85 tons/ha for cassava, cocoyam, maize, plantain, rice, and yam respectively at a 70% coverage level. At the 80% coverage level, these yields are 11.29 tons/ha, 4.98 tons/ha, 1.27 tons/ha, 6.68 tons/ha, 1.58 tons/ha, and 8.97 tons/ha, respectively. Using a 90% coverage level, the average threshold yields across the six crops over the country are 12.70 tons/ha, 5.60 tons/ha, 1.43 tons/ha, 7.51 tons/ha, 1.78 tons/ha, and 10.09 tons/ha respectively.

Components	Model 1	Model 2	Model 3	Model 4
Dependent Variable	Unadjusted Yield	Unadjusted Yield	Heteroskedasticity Adjusted Yield	Heteroskedasticity Adjusted Yield
Crop x Region	Yes	Yes	Yes	Yes
Crop x Year	No	Yes	No	Yes
BIC	42.19	20.05	65.88	41.37
MSPE ($\widehat{\lambda}_{lse}$)	3.94 (1094.91)	3.10 (626.55)	5.70 (1419.26)	5.24 (812.16)
MSPE ($\widehat{\lambda}_{opt}$)	3.88 (35.03)	3.05 (20.05)	5.61 (72.30)	5.15 (31.30)

Source(s): Authors' own creation/work

Table 6. Model comparison for four models

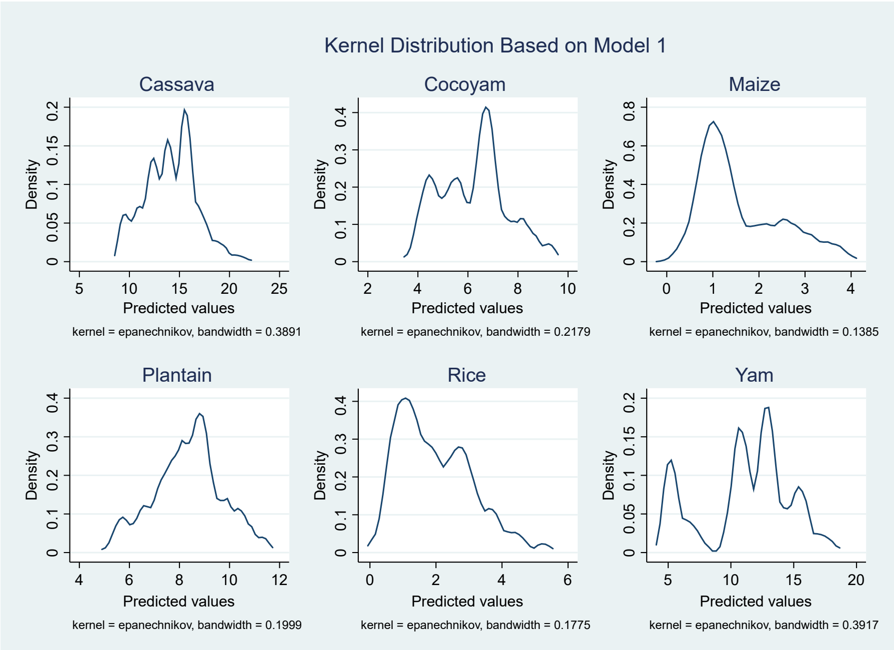
Crop	N	Mean	SD	p25	p50	p75	Min	Max
Model 1								
Cassava	7083	14.11	2.55	12.29	14.04	15.73	8.92	21.80
Cocoyam	5233	6.23	1.34	5.14	6.51	6.97	3.67	9.39
Maize	6720	1.59	0.90	0.92	1.26	2.22	0.09	3.98
Plantain	5520	8.35	1.37	7.43	8.39	9.11	5.10	11.53
Rice	5659	1.98	1.11	1.13	1.80	2.75	0.11	5.37
Yam	5985	11.21	3.53	9.92	11.90	13.26	4.45	18.28
Model 2								
Cassava	7083	14.11	3.29	11.67	13.92	15.99	7.73	25.08
Cocoyam	5233	6.23	1.06	5.15	6.23	7.20	4.32	7.88
Maize	6720	1.59	0.26	1.41	1.56	1.79	0.96	2.28
Plantain	5520	8.35	1.67	7.27	8.27	9.73	4.30	11.68
Rice	5659	1.98	0.76	1.41	1.78	2.38	0.68	3.93
Yam	5985	11.21	3.69	9.22	12.01	13.64	3.05	19.18
Model 3								
Cassava	7083	18.27	3.08	16.03	18.82	20.54	13.27	26.34
Cocoyam	5233	8.20	1.30	7.10	7.94	9.57	6.12	9.67
Maize	6720	2.09	0.29	1.88	2.04	2.35	1.33	2.61
Plantain	5520	10.81	1.26	10.42	10.91	12.01	8.10	12.11
Rice	5659	2.58	0.93	1.79	2.32	3.10	1.43	4.52
Yam	5985	14.54	4.41	13.80	14.56	17.48	7.05	21.50
Model 4								
Cassava	7083	18.27	3.21	15.62	18.68	20.45	11.84	28.02
Cocoyam	5233	8.20	1.71	7.04	8.22	9.69	4.08	11.19
Maize	6720	2.09	0.31	1.87	2.05	2.33	1.28	2.74
Plantain	5520	10.81	1.53	9.84	10.93	11.82	6.80	13.74
Rice	5659	2.58	0.95	1.85	2.36	3.16	1.02	4.81
Yam	5985	14.54	4.43	13.43	14.75	17.53	6.09	22.31

Source(s): Authors' own creation/work

Table 7.
Predicted crop yield based on 4 models in MT/Ha

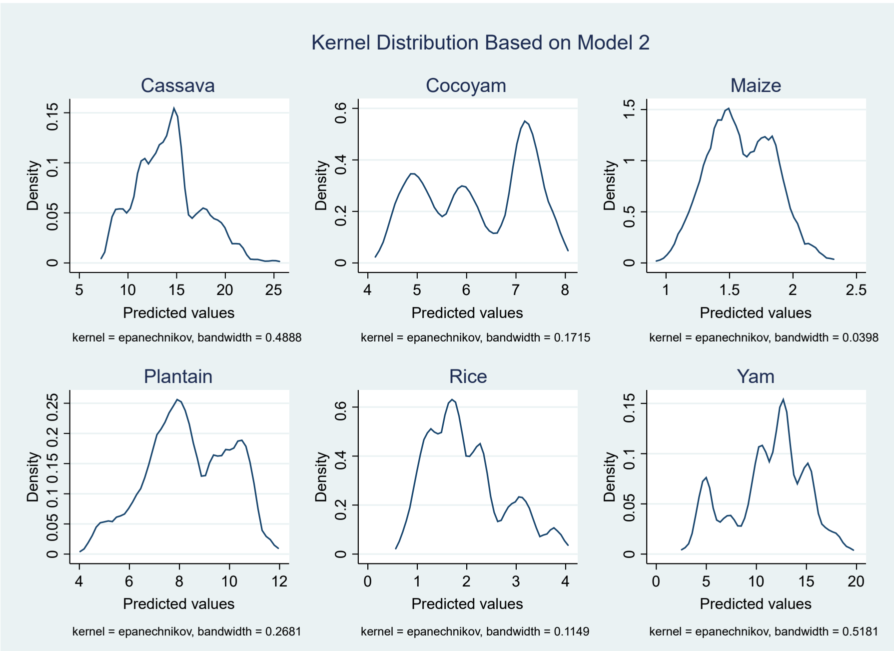
Across the ten regions studied, the range of the threshold yields for the 70% coverage level are 7.03 tons/ha to 11.42 tons/ha, 3.32 tons/ha to 5.13 tons/ha, 0.83 tons/ha to 1.37 tons/ha, 4.39 tons/ha to 6.62 tons/ha, 0.85 tons/ha to 1.71 tons/ha and 3.85 tons/ha to 11.37 tons/ha for cassava, cocoyam, maize, plantain, rice, and yam respectively. At the 80% coverage level, the threshold yield ranges from 8.03 tons/ha to 13.05 tons/ha (cassava), 3.79 tons/ha to 5.86 tons/ha (cocoyam), 0.95 tons/ha to 1.56 tons/ha (maize), 5.02 tons/ha to 7.57 tons/ha (plantain), 0.97 tons/ha to 2.66 tons/ha (rice) and 4.40 tons/ha to 13.00 tons/ha (yam). The range of threshold yields for cassava, cocoyam, maize, plantain, rice, and yam at the 90% coverage level is 9.03 tons/ha to 14.68 tons, 4.27 tons/ha to 6.59 tons/ha, 1.20 tons/ha to 1.76 tons/ha, 5.64 tons/ha to 8.52 tons/ha, 1.09 tons/ha to 3.00 tons/ha and 4.95 tons/ha to 14.62 tons/ha respectively.

In summary, the findings of this paper are useful for policymakers for development of government crop insurance programs. First, the variations in threshold yields could be due to several factors. Some important factors could be production inefficiency among producers in the various regions, lack of capital and climate issues. This study serves as a starting point for policymakers to seek ways to link the challenges producers face to government crop insurance programs based on the geographical peculiarity of the specific production problem. For instance, Tsiboe *et al.* (2022) found heterogeneities in farm management practices and production technology across the regions for cereal production in Ghana. In another study, Asravor *et al.* (2019) concluded that heterogeneous production technologies are employed across regions in Ghana due to differences in their production environments.



Source(s): Authors' own creation/work

Figure 2.
Kernel distributions of crop yield in Ghana based on Model 1



Source(s): Authors' own creation/work

Figure 3.
Kernel distributions of crop yields in Ghana based on Model 2

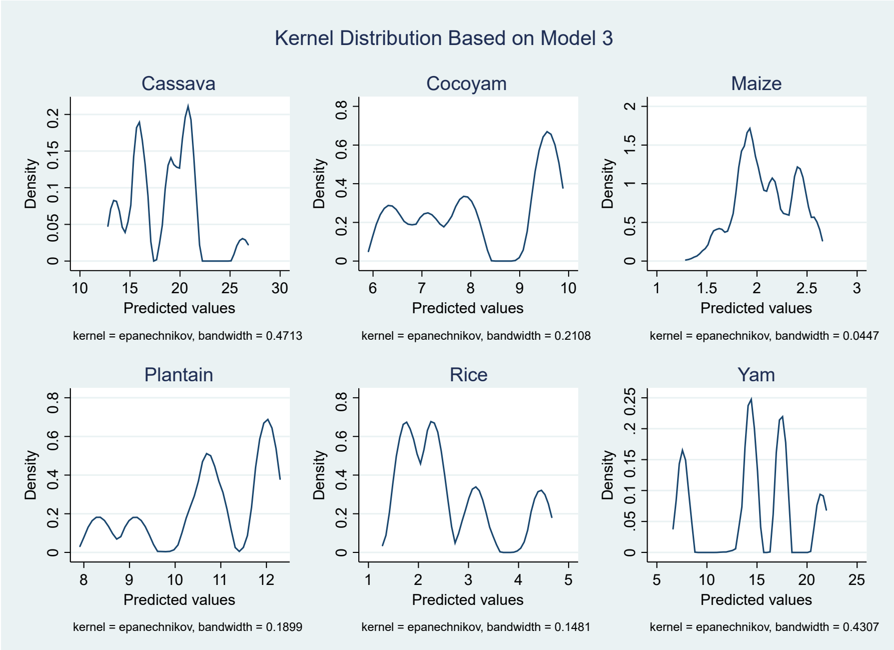


Figure 4.
Kernel distributions of crop yields in Ghana based on Model 3

Source(s): Authors' own creation/work

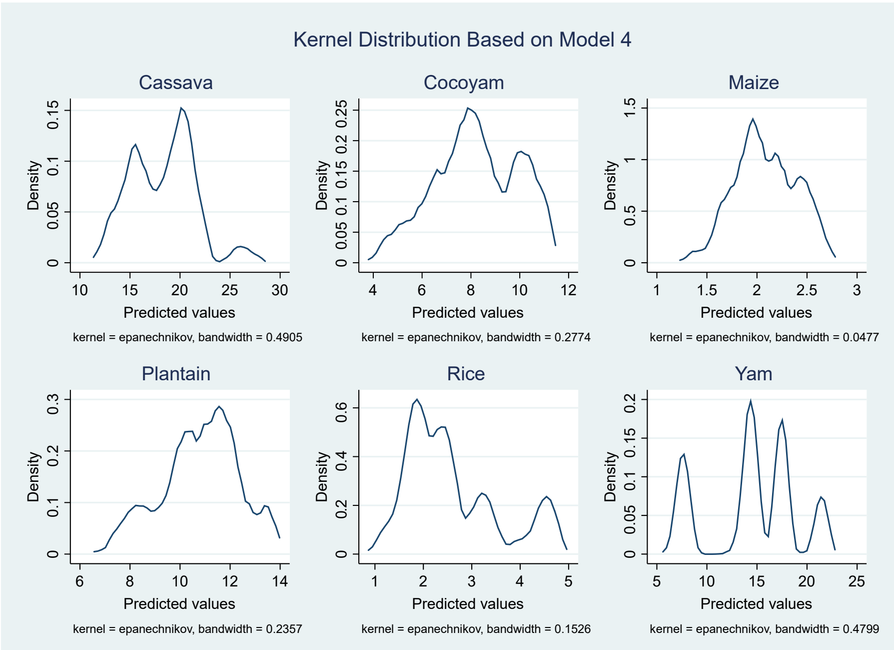


Figure 5.
Kernel distributions of crop yields in Ghana based on Model 4

Source(s): Authors' own creation/work

Crop	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
	Ashanti Region					Brong Ahafo Region				
Cassava	1508	12.92	2.25	10.58	18.26	1438	16.31	2.29	13.92	21.61
Cocoyam	1503	7.32	0.26	6.93	7.88	1129	6.06	0.26	5.67	6.62
Maize	1474	1.49	0.12	1.26	1.82	1423	1.83	0.12	1.60	2.16
Plantain	1497	9.46	1.29	7.55	11.68	1253	8.50	1.30	6.57	10.70
Rice	1515	1.86	0.33	1.32	2.46	780	1.36	0.34	0.80	1.95
Yam	1502	11.38	1.37	8.82	14.36	1415	13.54	1.37	10.98	16.51
	Central Region					Eastern Region				
Cassava	584	14.41	2.22	12.15	19.83	792	16.04	2.32	13.61	21.30
Cocoyam	379	4.74	0.26	4.32	5.27	771	7.22	0.26	6.83	7.78
Maize	587	1.67	0.13	1.45	2.01	754	1.95	0.13	1.72	2.28
Plantain	577	7.04	1.30	5.16	9.30	742	9.09	1.32	7.17	11.31
Rice	375	1.60	0.32	1.08	2.22	528	2.37	0.33	1.82	2.97
Yam	551	5.50	1.36	3.05	8.59	738	16.25	1.42	13.64	19.18
	Greater Accra Region					Northern Region				
Cassava	307	19.76	2.28	17.40	25.08	665	11.50	2.29	9.11	16.80
Cocoyam	1	5.67		5.67	5.67	23	6.39	0.14	6.14	6.80
Maize	7	1.59	0.19	1.36	1.82	13	1.65	0.14	1.45	1.86
Plantain	6	8.50	0.58	7.57	9.38	13	7.88	0.54	6.90	8.67
Rice	5	1.55	0.09	1.40	1.62	19	1.85	0.32	1.11	2.16
Yam	6	10.53	1.27	8.36	11.99	13	11.20	1.41	9.21	12.84
	Upper East Region					Upper West Region				
Cassava	18	13.97	2.76	10.12	16.65	18	14.03	2.37	9.84	16.65
Cocoyam	10	6.31	0.14	6.09	6.57	14	6.50	0.15	6.32	6.80
Maize	462	1.18	0.12	0.96	1.52	333	1.37	0.12	1.16	1.72
Plantain	7	8.45	1.15	6.03	9.42	5	8.68	0.21	8.51	9.03
Rice	467	2.45	0.32	1.92	3.06	340	1.42	0.32	0.90	2.05
Yam	15	11.34	0.94	9.79	12.64	309	13.29	1.43	10.80	16.34
	Volta Region					Western Region				
Cassava	916	14.46	2.24	12.14	19.82	837	10.04	2.24	7.73	15.41
Cocoyam	595	4.81	0.26	4.43	5.38	808	5.20	0.26	4.80	5.75
Maize	869	1.56	0.12	1.34	1.89	798	1.33	0.12	1.11	1.67
Plantain	596	6.27	1.31	4.30	8.43	824	7.83	1.30	5.94	10.08
Rice	852	3.33	0.33	2.79	3.93	778	1.21	0.32	0.68	1.83
Yam	616	10.68	1.42	8.00	13.54	820	5.78	1.38	3.25	8.79

Source(s): Authors' own creation/work

Table 8. Predicted regional crop yields based on Model 2

It is essential to accurately capture the sources of these variations for the specific regions because MPCI is often subsidized in countries where they have been introduced. Hence, the identification of the sources of variations of threshold yields can help policymakers to make informed decisions on subsidy allocations across regions and crops for the rates examined in this study. In addition, policymakers will be able to identify areas where specific improvement tools to mitigate crop yield risks could be allocated. Some improvement tools may be extension training, provision of irrigation materials, and machinery.

5. Conclusions

Farmers in developing countries struggle to deal with the risks associated with the agricultural sector. Providing aid for farmers and their households remains instrumental in combatting poverty in Africa. Several studies have shown that correctly understanding and implementing risk management strategies will help in the poverty alleviation agenda. Even as agricultural insurance continues to grow in Africa, there is growing research to

Table 9.
Kruskal-Wallis test for predicted yields based on Model 2

Region	Cassava	Cocoyam	Maize	Plantain	Rice	Yam
Ashanti	4350000	5430000	4060000	5430000	4240000	4520000
Brong Ahafo	6990000	2920000	7060000	3570000	1320000	5670000
Central	2240000	335587	2010000	992260	811417.5	383002
Eastern	3430000	3010000	3820000	2540000	2030000	3790000
Greater Accra	1750000	1799.5	23839	17021	12004	15986
Northern	1460000	64002.5	50694	30728.5	57335	38210.5
Upper East	67712.5	25522	693698.5	19590.5	1850000	45606.5
Upper West	68616	43141.5	669431	15567	624820	1120000
Volta	3640000	719831	2780000	652620.5	4120000	1660000
Western	1090000	1140000	1410000	1970000	946370.5	669904.5
Chi-Sq	2562 (9)	2747 (9)	2827 (9)	1519 (9)	3054 (9)	3954 (9)
Probability	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Source(s): Authors' own creation/work

Table 10.
Threshold yields for coverage levels across regions* and Ghana

Crop	N	Ghana	Ashanti	Brong Ahafo	Central	Eastern	Upper East	Upper West	Volta	Western
Coverage level of 70%										
Cassava	7083	9.88	9.04	11.42	10.09	11.23	9.78	9.82	10.12	7.03
Cocoyam	5233	4.36	5.13	4.24	3.32	5.06	4.41	4.55	3.37	3.64
Maize	6720	1.11	1.04	1.28	1.17	1.37	0.83	0.96	1.09	0.93
Plantain	5520	5.84	6.62	5.95	4.93	6.37	5.92	6.08	4.39	5.48
Rice	5659	1.38	1.30	0.95	1.12	1.66	1.71	0.99	2.33	0.85
Yam	5985	7.85	7.97	9.48	3.85	11.37	7.94	9.30	7.48	4.05
Coverage level of 80%										
Cassava	7083	11.29	10.34	13.05	11.53	12.84	11.17	11.22	11.57	8.03
Cocoyam	5233	4.98	5.86	4.85	3.79	5.78	5.05	5.20	3.85	4.16
Maize	6720	1.27	1.19	1.46	1.34	1.56	0.95	1.10	1.25	1.07
Plantain	5520	6.68	7.57	6.80	5.63	7.28	6.76	6.95	5.02	6.27
Rice	5659	1.58	1.49	1.09	1.28	1.89	1.96	1.14	2.66	0.97
Yam	5985	8.97	9.11	10.83	4.40	13.00	9.07	10.63	8.54	4.62
Coverage level of 90%										
Cassava	7083	12.70	11.63	14.68	12.97	14.44	12.57	12.62	13.01	9.03
Cocoyam	5233	5.60	6.59	5.45	4.27	6.50	5.68	5.85	4.33	4.68
Maize	6720	1.43	1.34	1.64	1.51	1.76	1.06	1.24	1.40	1.20
Plantain	5520	7.51	8.52	7.65	6.34	8.18	7.61	7.82	5.64	7.05
Rice	5659	1.78	1.67	1.22	1.44	2.13	2.20	1.28	3.00	1.09
Yam	5985	10.09	10.24	12.18	4.95	14.62	10.20	11.96	9.61	5.20

Note(s): *Greater Accra and Northern regions are not included due to small sample sizes
Source(s): Authors' own creation/work

understand how its advantages can be fully gained across the continent. This study examines crop yield distributions in Ghana and their implications on MPCI rating using the Lasso regression model. The appropriate Lasso model is selected based on the BIC criteria and k-fold cross-validation method, following which crop yields are predicted with a focus on the variability of the threshold yields across regions.

Based on the Kruskal-Wallis statistical test of differences among the regional threshold crop yields, the study concludes that regional crop yields vary across the ten regions for the six crops studied. Understanding this difference is essential for policymakers and insurance providers to enhance the sustainability of agricultural insurance programs. Considering this variability will prevent the overpricing of crop insurance products in regions where trigger yields are substantially high, as this is likely to disincentivize farmers. On the other hand, underpricing of insurance products in areas where threshold yields are substantially low could lead to the payment of huge claims by insurance providers. The knowledge of this variation of crop yields across regions presents an opportunity for the government to know the areas to provide support to make it sustainable.

Considering that this is a new area of research in Ghana's agricultural insurance sector, there are few issues that we could not address. For instance, several models have been employed in predicting crop yields in other countries. The predictions from the alternative models in this study shows that, selection of a wrong prediction model could lead to serious under or over estimation of yields and standard deviations (risks). The sector will benefit if future studies could also evaluate crop yield predictions based on alternative models. In addition, the impact of climate on agriculture is in continuous evolution. It is likely to play a major role in crop insurance rating and hence will be a relevant area of research.

Notes

1. Suppose $\boldsymbol{a}$ is a vector of dimension m with elements a_j for $j = 1, \ldots, m$, an L_1-norm is defined as $\|\boldsymbol{a}\|_1 = \sum_{j=1}^{m} |a_j|$ while the L_2-norm is $\|\boldsymbol{a}\|_2 = \sqrt{\sum_{j=1}^{m} |a_j|^2}$.
2. PRODUCTION ESTIMATES 1993–2017 - AGRICULTURAL PRODUCTION ESTIMATES, 1993–2017 | Ghana Open Data Initiative
3. The two groups are the unadjusted yields (presented in Table 1) and the heteroskedasticity-adjusted yields (presented in Table 5)

References

Addey, K.A. (2018), "Determinants of rice farmers' risk-attitudes in rural Ghana: an operational domain scale approach", *Journal of Agricultural Studies*, Vol. 6 No. 1, pp. 116-130, doi: 10.5296/jas.v6i1.12314.

Addey, K.A., Jatoe, J.B.D. and Kwadzo, G.T.-M. (2021), "Adoption of crop insurance in Ghana: an application of the complementary log-log truncated Poisson hurdle model", *Agricultural Finance Review*, Vol. 81 No. 1, pp. 76-93, doi: 10.1108/afr-06-2019-0062.

Addey, K.A., Shaik, S., Nganje, W. and SenGupta, I. (2023), "Implications for the dirichlet processes mixture linear model on U.S. Crop yield predictions", doi: 10.2139/ssrn.4382591, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4382591

Adjabui, J.A., Tozer, P.R. and Gray, D.I. (2019), "Willingness to participate and pay for index-based crop insurance in Ghana", *Agricultural Finance Review*, Vol. 79 No. 4, pp. 491-507, doi: 10.1108/afr-01-2019-0001.

Ahrens, A., Hansen, C. and Schaffer, M.E. (2020), "lassopack: model selection and prediction with regularized regression in Stata", *The Stata Journal*, Vol. 20 No. 1, pp. 176-235, doi: 10.1177/1536867x20909697.

Annan, F., Tack, J., Harri, A. and Coble, K. (2014), "Spatial pattern of yield distributions: implication for crop insurance", *American Journal of Agricultural Economics*, Vol. 96 No. 1, pp. 253-268, doi: 10.1093/ajae/aat085.

Asravor, J., Wiredu, A.N., Siddig, K. and Onumah, E.E. (2019), "Evaluating the environmental-technology gaps of rice farms in distinct agro-ecological zones of Ghana", *Sustainability*, Vol. 11 No. 7, p. 2072, doi: 10.3390/su11072072.

Atwood, J., Shaik, S. and Watts, M. (2003), "Are crop yields normally distributed? A reexamination", *American Journal of Agricultural Economics*, Vol. 88 No. 2, pp. 365-381, doi: 10.1111/j.1467-8276.2006.00864.x.

Barnett, B. (2014), “Multiple-peril crop insurance: successes and challenges”, *Agricultural Finance Review*, Vol. 74 No. 2, pp. 200-216, doi: 10.1108/afr-11-2013-0040.

Breustedt, G., Bokusheva, R. and Heidelbach, O. (2008), “Evaluating the potential of index insurance schemes to reduce crop yield risk in an arid region”, *Journal of Agricultural Economics*, Vol. 59 No. 2, pp. 312-328, doi: 10.1111/j.1477-9552.2007.00152.x.

Clarke, D.J., Mahul, O. and Verma, N. (2012), “Index based crop insurance product design and ratemaking: the case of modified NAIS in India”, Policy Research Working Paper WPS5986, World Bank.

Duarte, G.V., Braga, A., Miquelluti, D.L. and Ozaki, V.A. (2017), “Modelling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance”, *Journal of Applied Statistics*, Vol. 45 No. 11, pp. 1920-1937, doi: 10.1080/02664763.2017.1406902.

Farzammehr, M.A., Zadkarami, M.R., McLachlan, G.J. and Lee, S.X. (2020), “Skew-normal Bayesian spatial heterogeneity panel data models”, *Journal of Applied Statistics*, Vol. 47 No. 5, pp. 804-826, doi: 10.1080/02664763.2019.1657812.

Gehrke, E. (2019), “An employment guarantee as risk insurance? Assessing the effects of the NREGS on agricultural production decisions”, *The World Bank Economic Review*, Vol. 33 No. 2, pp. 413-435, doi: 10.1093/wber/lhw067.

Geisser, S. (1975), “The predictive sample reuse method with applications”, *Journal of the American Statistical Association*, Vol. 70 No. 350, pp. 320-328, doi: 10.1080/01621459.1975.10479865.

Goodwin, B.K. (2008), “Climate variability implications for agricultural crop production and risk management: discussion”, *American Journal of Agricultural Economics*, Vol. 90 No. 5, pp. 1263-1264, doi: 10.1111/j.1467-8276.2008.01214.x.

Goodwin, B.K. and Hungerford, A. (2015), “Copula-based models of systemic risk in U.S. agriculture: implications for crop insurance and reinsurance contracts”, *American Journal of Agricultural Economics*, Vol. 97 No. 3, pp. 879-896, doi: 10.1093/ajae/aau086.

Goodwin, B.K. and Mahul, O. (2004), “Risk modeling concepts relating to the design and rating of agricultural insurance contracts”, Word Bank Policy Research Working Paper 3392.

Harri, A., Coble, K.H., Ker, A.P. and Goodwin, B.J. (2011), “Relaxing heteroscedasticity assumptions in area-yield crop insurance rating”, *American Journal of Agricultural Economics*, Vol. 93 No. 3, pp. 707-717, doi: 10.1093/ajae/aar009.

Houssou, N., Andam, K. and Asante-Addo, C. (2017), *Can Better Targeting Improve the Effectiveness of Ghana’s Fertilizer Subsidy Program? Lessons from Ghana and Other Countries in Africa South of the Sahara*, IFPRI Discussion Paper No. 01605, International Food Policy Research Institute, Washington, DC.

Jensen, N.D., Mude, A.G. and Barrett, C.B. (2018), “How basis risk and spatiotemporal adverse selection influence demand for index insurance: evidence from northern Kenya”, *Food Policy*, Vol. 74, pp. 172-198, doi: 10.1016/j.foodpol.2018.01.002.

Karlan, D., Osei, R., Osei-Akoto, I. and Udry, C. (2014), “Agricultural decisions after relaxing credit and risk constraints”, *The Quarterly Journal of Economics*, Vol. 129 No. 2, pp. 597-652, doi: 10.1093/qje/qju002.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018), “Human decisions and machine predictions”, *The Quarterly Journal of Economics*, Vol. 133 No. 1, pp. 237-293, doi: 10.1093/qje/qjx032.

Liu, Y. and Ker, A.P. (2020), “Rating crop insurance contracts with nonparametric Bayesian model averaging”, *Journal of Agricultural and Resource Economics*, Vol. 45 No. 2, pp. 244-264.

Miranda, M.J. and Farrin, K. (2012), “Index insurance for developing countries”, *Applied Economic Perspectives and Policy*, Vol. 34 No. 3, pp. 391-427, doi: 10.1093/aepp/pps031.

Park, E., Brorsen, B.W. and Harri, A. (2019), “Using Bayesian Kriging for spatial smoothing in crop insurance rating”, *American Journal of Agricultural Economics*, Vol. 101 No. 1, pp. 330-351, doi: 10.1093/ajae/aay045.

Pauw, K. (2021), *A Review of the Ghana Planting for Food and Jobs Program: 2017-2020. Implementation, Impact, and Further Analysis*, Ghana Strategy Support Program (GSSP) Working Paper 53, International Food Policy Research Institute, Washington, DC.

Ramsey, A.F. (2020), "Probability distribution of crop yields: a Bayesian spatial quantile regression approach", *American Journal of Agricultural Economics*, Vol. 102 No. 1, pp. 220-239, doi: 10.1093/ajae/aaz029.

Schwarz, G. (1978), "Estimating the dimension of a model", *The Annals of Statistics*, Vol. 6 No. 2, pp. 461-464, available at: http://www.jstor.org/stable/2958889

Shaik, S. (2010), "Importance of panel analysis in examining normality of crop yield distributions", *Journal of Quantitative Economics*, Vol. 8 No. 1, pp. 55-68.

Stein, D. (2018), "Dynamics of demand for rainfall index insurance: evidence from a commercial product in India", *The World Bank Economic Review*, Vol. 32 No. 3, pp. 692-708.

Tack, J.B., Coble, K. and Barnett, B. (2018), "Warming temperatures will likely induce higher premium rates and government outlays for the U.S. crop insurance program", *Agricultural Economics*, Vol. 49 No. 5, pp. 635-647, doi: 10.1111/agec.12448.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society, Series B*, Vol. 58 No. 1, pp. 267-288, doi: 10.1111/j.2517-6161.1996.tb02080.x.

Tsiboe, F., Asravor, J., Owusu, V. and Mensah-Bonsu, A. (2022), "Production technology, efficiency, and productivity of cereal farms: prospects for enhancing farm performance in Ghana", *Agricultural and Resource Economics Review*, Vol. 51 No. 3, pp. 579-609, doi: 10.1017/age.2022.16.

Wang, S., Ji, B., Zhao, J., Liu, W. and Xu, T. (2018), "Predicting ship fuel consumption based on lasso regression", *Transportation Research Part D*, Vol. 65, pp. 817-824, doi: 10.1016/j.trd.2017.09.014.

Yang, S.-R., Koo, W.W. and Wilson, W. (1992), "Heteroskedasticity in crop yield models", *Journal of Agricultural and Resource Economics*, Vol. 17 No. 1, pp. 103-109.

Yao, S., Qin, Y., Cheng, F., Wu, J.G. and Goodell, J.W. (2022), "Missing momentum in China: considering individual investor preference", *Finance Research Letters*, Vol. 49, 103110, doi: 10.1016/j.frl.2022.103110.

Yu, J., Smith, A. and Sumner, D.A. (2017), "Effects of crop insurance premium subsidies on crop acreage", *American Journal of Agricultural Economics*, Vol. 100 No. 1, pp. 91-114, doi: 10.1093/ajae/aax058.

Zou, H. (2006), "The adaptive lasso and its oracle properties", *Journal of the American Statistical Association*, Vol. 101 No. 476, pp. 1418-1429, doi: 10.1198/016214506000000735.

Zou, H. and Hastie, T. (2005), "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society, Series B*, Vol. 67 No. 2, pp. 301-320, doi: 10.1111/j.1467-9868.2005.00503.x.

Further reading

Ker, A.P. and Tolhurst, T.N. (2019), "On the treatment of heteroscedasticity in crop yield data", *American Journal of Agricultural Economics*, Vol. 101 No. 4, pp. 1247-1261, doi: 10.1093/ajae/aaz004.

Shen, Z., Odening, M. and Okhrin, O. (2018), "Adaptive local parametric estimation of crop yields: implications for crop insurance rate making", *European Review of Agricultural Economics*, Vol. 45 No. 2, pp. 173-203, doi: 10.1093/erae/jbx028.

Tack, J.B. and Ubilava, D. (2015), "Climate and agricultural risk: measuring the effect of ENSO on U.S. crop insurance", *Agricultural Economics*, Vol. 46 No. 2, pp. 245-257, doi: 10.1111/agec.12154.

Corresponding author
John Baptist D. Jatoe can be contacted at: jjatoe@ug.edu.gh