ACME CORPORATION

Pascal Ntaganda, Kate Imbrone, Jill Thompson.
March 18, 2022
STATISTICAL ANALYSIS OF AGRICULTURE INPUTS IN GHANA.
git@github.com:omsba-5112-2022-winter/dtc-dtc-2.git

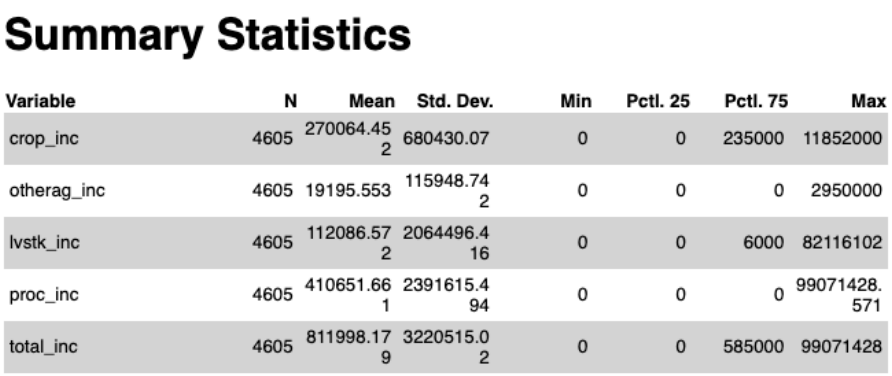
1. Background and Motivation

The agricultural industry is one of the most difficult sectors, affected by numerous natural and human induced factors. The expansion into the industry therefore requires tremendous caution. To target profit, agricultural productivity should be complemented by other inputs into the channel of distribution of the products. These include identification of markets, skilled labor sources, supporting infrastructure, water sources, machinery, and numerous other aspects. Most importantly, the analysis of the geography is important as a fundamental building block of the project. This statistical analysis will create an assessment of the profitability of ACME Corporation moving into agricultural inputs in Ghana.

2. Descriptive Statistics

Figure 1 through 4 below describe the general descriptive statistics of the agricultural data at the household level. They include all the variables used to create the regression models used to determine the overall profitability of the investment. In addition to this, some variables used are because of their utilization to create indices used to calculate incomes and expenditures of the involved activities.

Figure 1: Household agricultural income descriptive statistics in Ghanaian Cedi



Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
crop_inc	4605	270064.452	680430.07	0	0	235000	11852000
otherag_inc	4605	19195.553	115948.742	0	0	0	2950000
lvstk_inc	4605	112086.572	2064496.416	0	0	6000	82116102
proc_inc	4605	410651.661	2391615.494	0	0	0	99071428.571
total_inc	4605	811998.179	3220515.02	0	0	585000	99071428

currency.

Data collection dates: 1998 - 1999

Figure 2: Household agricultural expenditures descriptive statistics in Ghanaian Cedi

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
crop_exp	4337	167532.035	806361.447	0	15000	143000	34600000
home_cons	4337	601438.837	990495.498	0	73000	786453.33	31713633.33
land_rent_exp	4337	7438.15	35371.019	0	0	0	840000
livestock_exp	4337	32167.679	503018.869	0	0	0	21260000
proc_exp	4337	225652.264	1368847.971	0	0	91250	37876571.43
total_exp	4337	1034228.965	2130788.826	24.33	221243.66	1193267.44	43177000

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
total_inc	837	2043171.691	4883632.272	12000	460000	2346428	82638602
total_exp	837	879292.554	1659672.167	0	210000	1117853.01	40106679.05
profit	837	1163879.137	4348892.63	166.67	113467.24	1186309.77	81641102
relig	837	0.903	0.296	0	1	1	1
ecowas	837	0.998	0.049	0	1	1	1
urban	837	0.154	0.361	0	0	0	1
coastal	837	0.182	0.386	0	0	0	1
forest	837	0.527	0.5	0	0	1	1
savannah	837	0.292	0.455	0	0	1	1
educyrs	837	2.534	3.148	0	0	4	17
everschool	837	1.406	0.506	0	1	2	2
highest_cert	837	1.706	1.702	0	1	2	12
read_gha	837	0.388	0.488	0	0	1	1
read_eng	837	0.378	0.485	0	0	1	1
fund_source	837	0	0	0	0	0	0
water_dist_m	837	612.646	8652.679	0	30	400	250000

currency.

Figure 3: Household descriptive statistics in Ghanaian Cedi currency.

Figure 4: Household agricultural profit descriptive statistics in Ghanaian Cedi currency.

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
total_inc	4612	810765.745	3218224.826	0	0	584000	99071428
total_exp	4612	972560.932	2080737.567	0	170290.775	1127141.675	43177000
profit	4612	-161795.187	2847419.13	-37177000	-595085.178	0	97324642.29

Figure 5: Prominent cash crops in Ghana

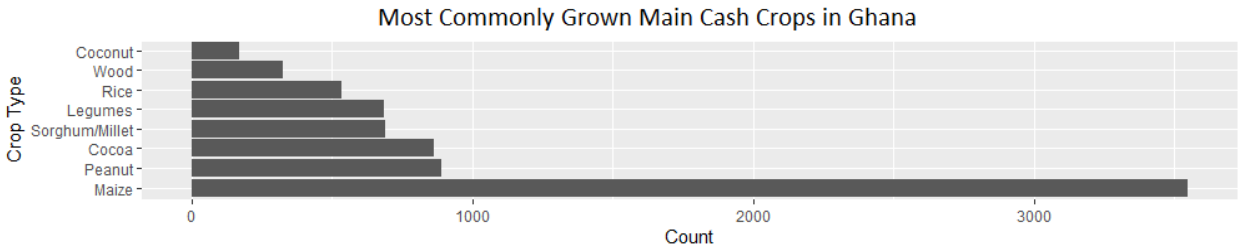


Figure 5 visualizes the most common cash crops in Ghana, with maize being the predominant crop in the country by a huge margin.

3. Methodology and Model

To estimate the agricultural profit in Ghana, we developed two multiple regression models that considered the gender, marital status, nationality, literacy language, source of funding, education years, relationship to head of household, distance from water source and climate of the regions. Most companies use profit as the metric to decide on where to invest. Our team therefore compliments to this metric by ascertaining a sustainable approach to analyzing agricultural profit in Ghana. Using multiple regression, we intend to control these shortcomings quantitatively for comprehensive exploration of profit in this industry and region.

Initially, we created datasets with corresponding income and expense variables from different activities involved, at both household and individual levels, from the original survey data. This was followed by calculating profit as the difference between the relevant income values and expenses involved (Profit = Income - Expenses) at the household level, which is the focus of the study. All income, expense, and profit numbers are measured in Ghanian cedis. We could include all relevant variables in the regression models and determine the relevant profits

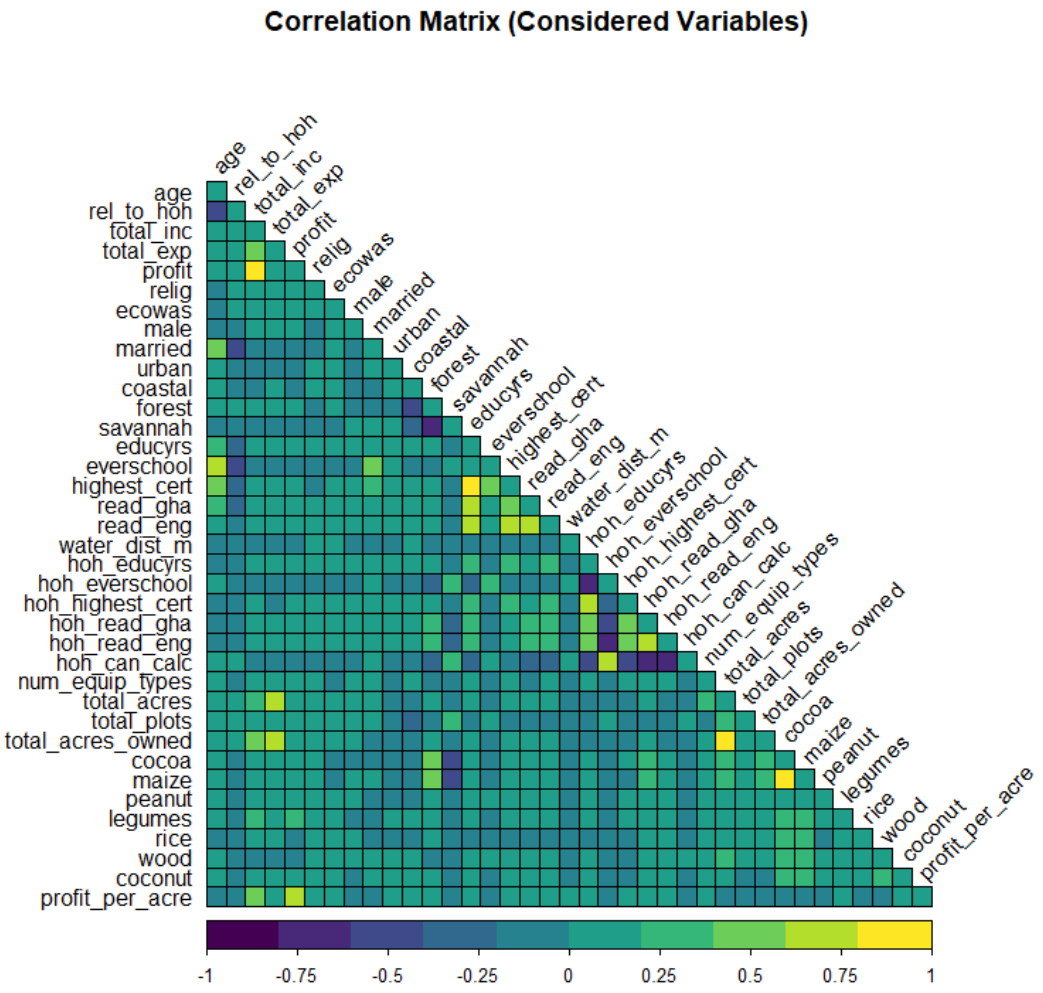
involved. Reviewing the standardized residual plots, or data error terms (assumptions for regression analysis), we found the results were not normally distributed and with traces of heteroscedasticity, meaning they would produce unreliable tests from the selected variables.

For efficient analysis of the variables that affect profit, our team created dummy variables for geographic zones of forestry, savannah, and coastal as either 0 or 1(fitting or not fitting). This was followed by using the correlation matrix to decide on what variables are a best fit for the models. Incessantly, our team adjusted the different variables though creating their respective logarithms to ensure that they are normally distributed. In addition to this, a multiple regression model was created to further assess the crop with the highest profit potential among the top cash crops in Ghana. Finally, the respective regression models were assessed to ensure dependable results by using the residual plots. Residual plots check for the assumptions of the Ordinary Least Squares Regression Model used in this statistical analysis.

4. Results

The results from the correlation analysis can be found in figure 6 below, presenting the correlation matrix of the important variables used at both household and individual levels of the analysis

Figure 6: Correlation matrix of household variables



The multiple regression results can be found below in figure 7. Model 1 described the relationship between the log of profit per acre as a dependent variable and independent variables (explanatory variables) namely: individual ever being in school, log of total acres owned, cocoa cash crop, marital status, coastal geography, forest region, maize crop, and peanut crop. This relationship demonstrates that there is an 92.3% increase in profit for those who have attended any amount of school versus those who have not attended any school, provided other exploratory variables are held constant. In model 2, there is a 91.5% increase in agricultural income for those who have attended some school versus those who have not attended school when all other factors remain constant.

In model 1, there is approximately a 21 percent decrease in profit per acre for approximately every 1 percent increase in total acres owned, when all other factors are held constant. On the

other hand, there is approximately a 42% percent increase in the total individual income, for every 1 percent increase in the total acres owned, provided all other explanatory variables remain constant. Now focusing on the effect of prominent cash crops to the profit per acre and the total income; in model 1, cocoa crops do not have a statistically significant effect. As for the effect in model 2, there is an 23% increase in income for those who grow cocoa as opposed to those who do not, when all other exploratory variables are held constant.

Looking at the maize cash crop, there 74% increase in profit for individuals who grow maize versus individuals who do not grow maize, provided other factors are held constant in model 1. As for model 2, there is a 16% increase in agricultural income for those who grow maize as opposed to those who don't, when all other predictors are held constant. As for the peanut crop, no statistical significance was found in Model 1. Evaluating the effect of the peanut crop on income as observed in model 2, growing peanut crops results in a 63% increase in agricultural income when all other predicting variables remain constant.

Bringing into light the effect of geographical location to the total income and profit per acre, for areas in the coastal region there is a 103% higher profit per acre shown in model 1. In model 2, there is a 63% increase in agricultural income over non-coastal areas, when all other predictors are held constant. Observing the effect of the forest regions, in model 1, there is a 101.3% increase in the profit per acre when all other factors are held constant. In model 2, agricultural income is 6% higher in forest areas as compared to non-forest areas, when all other predictors are held constant.

Figure 7: Multiple regression results

Variable	Interpretation	Model 1: Coefficient Estimate	Model 2: Coefficient Estimate
coastal	1 = Coastal 0 = Not coastal	0.71 *** (0.09)	-0.45 *** (0.04)
forest	1 = Forest 0 = Not Forest	0.69 *** (0.08)	0.06 '.' (0.04)

log_total_land_acres	Log of total land size in acres	-0.21 *** (0.03)	0.28 *** (0.01)
everschool	1 = Has received some form of formal education 0 = No formal education	-0.08 ‘.’ (0.04)	-0.09 *** (0.02)
log_profit_per_acre	Log of profit (Ghanian cedi) per acre of land owned	X	0.42 *** (0.01)
maize	1 = Grows maize 0 = Does not grow maize	-0.29 * (0.13)	0.14 * (0.06)
peanut	1 = grows peanut 0 = does not grow peanut	-0.43 ‘ ‘ (0.28)	0.49 *** (0.13)
cocoa	1 = grows cocoa 0 = does not grow cocoa	-.07 ‘ ‘ (0.13)	0.21 ‘ ’ (0.06)
R-squared		0.05	0.51
Adjusted R-squared		0.05	0.51

Sample size: 4800 people split between 837 households

Standard errors are represented in parenthesis in the table

Significance level keys: * significant at 0.1%, ** significant at 1%, * significant at 5%, . significant at 10%, ‘ ‘ not significant**

5. Limitations and Conclusions

What is profit? What impact does profit have on lucrative potential?

While the definition of profit in a rudimentary sense is simply the difference between income and expenditures, real world implications are far more complex. Agribusiness in Ghana in 1998-1999 was experiencing incredible change due to recent legislation and progressive introduction of government subsidy. As such, a household’s profit may vary from one year to the next, while their actual gross income stays relatively consistent. Many households in the GLSS4 data have

members who receive additional income from outside employment and as such, a financial deficit in agricultural income/expense may not result in an overall lack of funds in the household.

Are there better ways to determine which demographic or area is a better target for sales ventures by ACME Corporation?

In the real world and within provided data, there are far too many factors playing into what defines profit with any sort of certainty. A household's buying power or lucrative potential cannot be ascertained through the lens of profit as we are able to define it with the inherent constraints of these data. It is for this reason our group chose to begin investigating overall agricultural income over rudimentarily calculated estimations of profit. Once we began regressing variables against income, we were able to determine correlated, statistically significant regional and educational factors.

Conclusion

Both regression models used produced relatively similar results, but Model 2 produced results with a higher level of statistical significance. Overall, our analysis shows that households where at least one member has attended school have both higher profit per acre and higher agricultural income. Households in forest regions have higher profit per acre. Households in the forest and coastal region both have a higher agricultural income. Growing peanuts, maize, or cocoa correlate with higher income, but not with higher agricultural profit.

When selling agricultural inputs, ACME Corporation should focus on coastal households, households where at least one member has formal education, and households that grow either peanuts or cocoa as a main cash crop.