Glen Lewis, Sohrab Rajabi, Rosalie Sherry, Tamlyn Tamura

OMSBA-5112

Data Translation - Draft

11/20/2020

DATE: November 20, 2020

TO: The ACME Corporation

FROM: OMSBA5112 division DTC-4

SUBJECT: re: Agricultural Profits in Ghana

Claus,

Thank you for tasking us with this project. After weeks of hard work, we have concluded that the factors that have the largest effects on agricultural profit per an acre is workforce composition, location, and education. The impact of these three factors lead us to believe that the average expected profit per acre of land is \_\_\_\_\_\_\_\_\_\_\_.

Included below are ***some*** graphs and charts that will illustrate the magnitude of impact these factors have on agricultural profit in Ghana. Beginning with the \_\_\_\_\_ you can see….

Moving onto our next \_\_\_\_ we examine the \_\_\_....

If you wish for information beyond what has been stated here please be sure to read through the Statistical Analysis we have attached that details our methodology and provides greater insight into our findings.

Best,

The OMSBA DTC-4 Team

Glen Lewis Sohrab Rajabi

Rosalie Sherry Tamlyn Tamura

Abstract

In determining what drives agricultural profit, three massive surveys were conducted in Ghana. Both at the individual and household level, participants were able to give us a first-hand look at their day-to-day. While some of the data was focused on household, much of it was focused on land and farming. Analyzing the data from a birds-eye perspective generated several key insights related to profitability in Ghana and whether it is a strategic business move to invest here. This paper is currently incomplete due to the need to finish merging the two datasets together.

Background

The ACME Corporation has entrusted the DTC-4 Group with finding the factors that determine agricultural profit in Ghana. Since Ghana is so geographically and culturally different than the United States, DTC-4 members were not too sure what to expect before diving into the data; they hypothesized that education and literacy were two key factors that would drive profitability as well as the number of hands working on the land. In terms of land, they hypothesized that ecological/locality zones and revenues from sale of cash crops would explain around 70% of the profitability model.

For the DTC-4 group to draw an accurate conclusion, they have spent the last couple of weeks parsing through files in order to get a full picture of what drives profit. DTC-4 has determined that there are two main groups that all factors can be split into that determine profit, land and household. Each factor was split into one of these two groups and analyzed in comparison to the profit equation, known as AGGI1 (taken from the dataset AGG2). The equation is:

AGRI1 **=** CRPINC1 + CRPINC2 + ROOTINC + INCOTHAG + TRCRPINC +HOMEPRO - EXPLAND - EXCROP - EXLIV - EXPFDPR1 - EXPFDPR2

The equation and its parts will be explained in full later on in this paper. Looking initially at the household portion it was determine that both the number of male and the number of female household members affects profit as well as the number of paid male workers. For land, it was determined that \_, \_, and \_ were the three biggest determining factors. As we progress in combining these groups together, we expect to see education take a more prominent role as well as the type of crops on the land. We did notice that overall district 10, and 11 (in reference to district 1) did seem to be the most profitable in general and will further dive into how these two groups affect these districts in general.

Implementation

In order to split the data into their respective groupings, it first needed to be decided which variables from the data sets were relevant. The DTC-4 Group gathered to determine that 7 datasets were relevant to the household portion and 13 were relevant to the land portion. The datasets for household being SEC1: The Household Roster, SEC2A: Education: General Education, SEC2C: Education: Literacy/Apprenticeship, EXP3: LANDEXP = Expenditure on renting farmland, EXP4: CROPEXP = Expenditure on crop inputs, EXP6: which is made up of two subgroups: FDPREXP1: Labour[[1]](https://word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en%2DUS&rs=en%2DUS&wopisrc=https%3A%2F%2Fredhawks-my.sharepoint.com%2Fpersonal%2Ftamurat_seattleu_edu%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F2c192678bcec422aa36999a757256e1e&wdorigin=OFFICECOM%2dWEB%2eSTART%2eREC&wdenableroaming=1&mscc=1&wdodb=1&hid=235B8F9F-60EB-B000-A5D6-429E151D79E9&wdhostclicktime=1605839971744&jsapi=1&jsapiver=v1&newsession=1&corrid=07039743-8060-4cc2-90e6-801b23f162c4&usid=07039743-8060-4cc2-90e6-801b23f162c4&sftc=1&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush&rct=Medium&ctp=LeastProtected" \l "_ftn1) costs on food processing, FDREXP2: Other costs on food processing, and finally EXP7: HP = Consumption of home production. If one looks at *Table 1* below, one will see some definitions of the various variables used, how they appear, their data type, and their definition.

Table 1: Household Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Alternate Name | Type | Definition |
| District | N/A | Factor | District where household resides |
| s2aq2 | N/A | Factor | Highest level of schooling completed |
| av\_hh\_age | agey | Numeric | Average age in years for household.  Determined by grouping individuals by  household id then - agey =  round(mean(agey), 1) |
| male | sex | Numeric | Total number of males in household  Determined by grouping individuals by  household id then counting the number  applicable factor levels |
| female | sex | Numeric | Total number of females in household |
| male\_help | s8cq17a | Numeric | Taken directly from sec8c data set |
| female\_help | s8cq17b | Numeric | Taken directly from sec8c data set |

The datasets used for the land variables are as follows:

INC10: Revenue from sale of cash crops, INC11: ROOTSV = Revenue from sale of roots/fruits/vegetables, INC12: OTHAGINC = Revenue from other agricultural income, and INC13: INCTRCRP = Revenue from transformed crops. On the expense side, DTC-4 chose to use the same EXP3 through EXP7, as listed earlier in the household section. Other important datasets include SEC8A1: Land, SEC8A2: Livestock and Fishing, SEC8A3: Agriculture equipment, and SEC8B: Plot details.

Table 2: Land Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Alternate Name | Type | Definition |
| income10\_final | N/A | Numeric | Shows how much revenue was produced from each cash crop, labeled by the unique cluster/id “key” |
| income11\_final | N/A | Numeric | Shows how much revenue ws produced from each type of root/fruit/vegetable |
| income12\_final | N/A | Numeric | Similar to INC12, but with united “nh” and “clust” to produce “key” |
| income13\_final | N/A | Numeric | Shows how much revenue is generated from transformed or processed foods |
| agri\_land | sec8a1 |  |  |
| agri\_livestock\_Fishing | sec8a2 |  |  |
| agri\_equipment | sec8a3 |  |  |
| agri\_plot | sec8b |  |  |

In order to combine the two categories of data, DTC-4 first created a universal key that could combine data from different datasets. They did so by uniting the variables “nh” and “clust”. Doing this gives a more specific explanation of where certain parts of the data are coming from.

INC10\_FINAL was created by combining two variables within the dataset, CROPSV1 and CROPSV2, and then mutating the data so that the crop code (variable name “cropcd”) displayed a table with each unique name of the crop as its own column, instead of the numeric code associated with the crop in one column. After that, the crops with no values across the board were eliminated, and INC10\_FINAL was created.

INC11\_FINAL was created in a similar manner, by mutating the data so that the “root\_name” code was transposed to the actual name of the root/fruit/vegetable having its own column.

INC13\_FINAL was also created in a similar manner, by mutating the data so that the “trans\_crop\_name” was transposed to displaying the actual name of the transformed crop assigned to its own column, instead of a series of numbers displayed under one variable “trans\_crop\_name”.

[[1]](https://word-edit.officeapps.live.com/we/wordeditorframe.aspx?ui=en%2DUS&rs=en%2DUS&wopisrc=https%3A%2F%2Fredhawks-my.sharepoint.com%2Fpersonal%2Ftamurat_seattleu_edu%2F_vti_bin%2Fwopi.ashx%2Ffiles%2F2c192678bcec422aa36999a757256e1e&wdorigin=OFFICECOM%2dWEB%2eSTART%2eREC&wdenableroaming=1&mscc=1&wdodb=1&hid=235B8F9F-60EB-B000-A5D6-429E151D79E9&wdhostclicktime=1605839971744&jsapi=1&jsapiver=v1&newsession=1&corrid=07039743-8060-4cc2-90e6-801b23f162c4&usid=07039743-8060-4cc2-90e6-801b23f162c4&sftc=1&instantedit=1&wopicomplete=1&wdredirectionreason=Unified_SingleFlush&rct=Medium&ctp=LeastProtected#_ftnref1) The datasets and documentation was written in the Queen’s English, when quoting said documentation this paper will use the language as it appears in the documentation, when writing it will continue to be written in American English style.

In assessing the impact, the DTC-4 assessed 7 hypotheses to identify the influences on annual agricultural profits:

1. Education positively affects expected annual profits.

2. Age positively affects expected annual profits.

3. More male members in a household positively affects expected annual profits.

4. More female members do not have a positively affects expected annual profits.

5. More male hire helped positively affects expected annual profits.

6. More female hired help positively affects expected annual profits.

7. The location of the household does have an affect on expected annual profits.

Review of different combination of household variables lead to the following model (bkwd\_1\_a) being adopted as representing their influence on profit.

*lm(profit ~ factor(district) + factor(loc2) +*

*av\_hh\_age + male + female + male\_help +*

*I(male\_help^2) + female\_help, data = nh\_profile\_base)*

The base model (bkwd\_base) that is used for comparison is:

*lm(profit ~ factor(district) + factor(ez) + factor(loc2)*

*av\_hh\_age + male + female + male\_help +*

*female\_help, data = nh\_profile\_base)*

The model was selected because:

F-Statistics, p –value and R-squared:

The selected model – bkwd\_1\_a – produces a F-statistic of 86.43, with a p-value: < 2.2e-16 while the base model produces an F-statistic of 49.96 and a p-value: < 2.2e-16. While the p-values are equal and statistically significant at a .001 level equal, bkwd\_1\_a’s higher F-statistic indicates the model result are even less likely to be a random result than the bkwd\_base model.

The bwlwd\_1\_a model had an adjusted R-square of 0.1061, with the base having a .1086. However, due to the poor residual v. fit of bkwd\_1, DTC-4 felt that the .0025 difference in the R-square was not significant enough to justify the poorer model.

Residual Distribution:

The error distribution for both models – while showing an ever so slight right tail – are basically identical.

NEED GRAPH

A review of the homo/heterskedasticity shows is where the difference in the models is seen. The bkwd\_base visual shows a significant downward bend in the mean, indicating that the errors are not constantly varied with a mean of 0.

NEED Graph

The bkwd\_base\_1\_a model not only reduces the profit range of residual outliers but also flattens the mean indicating a better variance across the errors. While this model is not perfect, it provides the best fit in determining the influences of household variables on profit.

NEED GRAPH

Results

In reviewing our hypothesis there were a couple of surprises. The first being the influence education has on agricultural profit. Only 2 factor levels of completed education had statistically significant influence on profit at the 10% level: middle school completion and the koranic stage completion. Of interest, while having completed middle school has a positive influence (coefficient: ¢193867 GHS), the koranic stage training led to a significant negative influence on profit (coefficient: ¢ -401125). That could result from the time dedicate to such training takes away from agricultural work. An interesting trend, though not significant, it appears as household members gained more education, the influence of education on profit turns increasingly negative.

The second surprise was the minimal impact the average household age had on profit. The DTC-4 hypothesized that an older household could expect higher annual profits due to the buildup of knowledge and experience in growing crops over time. However, the average household age (variable: av\_hh\_age) while statistically insignificant (.0973) at the .10% level in influencing profit, the influence was only about ¢3279 GHS per additional year of age.

Household Workforce Profile:

To begin all models showed that regardless of the variable makeup, the composition of the household workforce was statistically significant at the .01% level. Of interest, when the model removed the female help, the female household member’s coefficient did not increase in parallel to the changes in the male household member and male help.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Variable | Coefficient | Percent Change |
| male | 197340 | male | 202402 | 2.5% |
| female | 96083 | female | 96713 | .65% |
| male\_help | 22233 | male\_help | 24537 | 9.4% |
| female\_help | 41004 |  |  |  |

This indicates that ...

The male help variable needs to be consider carefully thought. The majority of the household do not have hired help (median: 0.000, mean: 2.081, max: 262.000). This can be interpreted that hired help is not used by most of the household farms and thus not a major contributor to profit. The majority appears to work on a few farms, potential large, corporate type farms where there help does have a significant impact on profits.

Conclusion

Based on our findings, the ACME Corporation should (not) invest in Ghana based on Agricultural land per an acre. The DTC-4 team will provide more insight on this once they merge the land and household data.

Notice of Possible Errors

It should be noted that this conclusion was drawn using data that was inherently biased as well as did not have consistent methods of collection. Should the ACME Corporation want a more indepth review or wishes to further pursue Ghana for various business interests the DTC-4 team would recommend finding a different surveying company as the previous one left many areas for error to be found in the DTC-4's analysis. Within the provided documentation it noted that many of the numbers were calculated to be on an annual scale but using data colleting from varying date ranges to get to make that estimation. For more information the errors that may have occurred in this analysis please look through the Aggregate.pdf pages 13-16.