ECON 5100 18FQ Statistical App & Quant Methods
Group 05
December 9, 2018

Introduction :

The Ghana Living Standards Survey (GLSS), with its focus on the household as a key social and economic unit, provides a valuable insights into living conditions in Ghana. The purpose of this report is to find factors that impacts agricultural profit since ACME Corporation is considering to move into agricultural inputs in Ghana. ACME Corp. is looking for factors in their sales efforts to maximize their profit.

Initial Hypothesis :

Region, District, EZ : Region segregates the household data into different regions and hence this information is useful in finding the relation with maximum profit.

Loc2 : This segregates households into urban and rural. In this report, we are considering only rural regions.

agri1c : agri1c is the amount of agricultural profit obtained from aggregate data (agg2.dta). The primary objective of this report is to analyse agricultural profit as the dependent variable.

s2aq2 : Education attainment is one of the explanatory variables affecting the agricultural profit. This variable is "highest education level of a person". The variable has been used in two ways : first, highest education level of household and second, education level of household head.

Sex : Represents gender of head of household. The variable is used as a dummy variable is_male to determine the effect of sex of head of household has on the agricultural profit.

s4aq3 : Signifies if a person has ever worked on a farm or not. Assumption is that if a person has prior experience in the farm, it might impact the agricultural profit. So for the purpose of analysis, observations where the experience is a yes has been considered.

s4aq6 : Represents the agricultural activity done by a household member and its impact on agricultural profit. For the analysis, following data has been considered: first, if cash crops are used in agriculture and second, if food crops are used in agriculture.

s4bq8 : This tells us about the employment status of household member. An employment status is categorized as self-employed with/without employees or unpaid family worker. A self-employed member having worked in agricultural practices will have positive impact on agricultural profit.

S8bq4a : This gives us the size of land owned by every household. We have assumed here that the land size is directly proportional to the agricultural profit.

Imprt : This gives relation of imputed income with agricultural profit.

S2q1a : This is the primary occupation of the community. Communities where major activity is farming has been considered.

S2q4, s2q23 : In order to determine how does the presence of motorable road and public transport within a community affects the agricultural profit, these two variables have been taken into consideration.

S2q8, s2q10 : In order to determine how does the presence of electricity and pipe borne water within a community affects the agricultural profit, these two variables have been taken into consideration.

S5bq5, s5bq10 : In order to determine how does the presence of an extension centre and cooperative within a community affects the agricultural profit, because it might help them to procure and sell their goods, these variables have been taken into consideration.

S5bq13, s5bq14 : These variables have been considered in order to determine how does the presence of modern equipments (tractors and rice husking machine) within a community affects the agricultural profit.

S5bq15, s5bq16 : These variables have been considered in order to determine how does the pesticides impact the agricultural growth and so does the profit.

S5bq17, s5bq18 : These variables have been considered in order to determine how is the rainfall within a community and how does it impact the agricultural profit.

Input Variables and their Significance :

From Data	Variable Name	Significance	Renamed As	Type of Variable
Section 0	Region	Region ID - All household		
Section 0	District	District ID - All household		
Section 0	nh	Household ID - All household		Quantitative
Section 0	clust	Enumeration area number		
Section 0	ez	Ecological zone number		
Section 0	loc2	Locality classification		
Section 0	eanum	Enumeration area number		

Aggregate	agri1c	Income from employment	agriculture_profit	Quantitative
Section 1	sex	Gender of HH member	hh_is_male	Categorical
Section 1	rel	Relationship to HH head		Categorical
Section 2	pid	Person ID		Quantitative
Section 2	s2aq1	Ever attended school	attended_school	Categorical
Section 2	s2aq2	Highest level completed	hh_highest_educ- Head of HH highest_educ_level - in HH	Categorical
Section 2	s2aq3	Highest education qualification	highest_educ_qualification	Categorical
Section 4	s4aq3	Worked on farm	worked_on_farm	Categorical
Section 4	s4aq6	Which activity	agricultural_activities	Categorical
Section 4	s4bq8	Employment status	employment_status	Categorical
Section 8	s8bq4a	Farm land size	farm_land_size	Quantitative
Section 8	s8bq4b	Unit of measure	unite_of_measure	Categorical
Aggregate	imprt	Actual and imputed rental income	imprt	Quantitative
Community	s2q1a	Major economic activity	primary_occupation	Categorical
Community	s2q4	Motorable road pass	motorable_road_y	Categorical
Community	s2q8	Electricity or Generator	electricity_y	Categorical
Community	s2q9	Electricity to Most or few HH	electricity_most_few	Categorical
Community	s2q10	HH have pipe-borne water	water_y	Categorical
Community	s2q11	pipe-borne water to Most/few HH	water_most_few	Categorical
Community	s2q23	Public transport pass by	public_transport_y	Categorical
Community	s5bq5	Agriculture extension center in community	extension_center_y	Categorical
Community	s5bq10	Co-operative in community	cooperative_y	Categorical
Community	s5bq13	No of tractors in community	no_of_tractors	Quantitative
Community	s5bq14	Rice-husking machine in community	rice_husking_machine_y	Categorical

Community	s5bq15	Community use chemical fertilizers	chemical_fertilizer_y	Categorical
Community	s5bq16	Community use insecticides or herbicides	insecticides_herbicides_y	Categorical
Community	s5bq17	Irrigated fields in community	irrigated_fields_y	Categorical
Community	s5bq18	Rain in last 12 months compared to 12 months before	rainfall_more_less	Categorical
Community	s5bq24	Community measure Land plots as	plot_measure_equip	Categorical

Data Exploration and Prediction

Model

As agricultural profit has non linear curve, we are taking Logistic Regression on agricultural profit while doing the analysis and in order to get linear relationship between dependent and independent variables. Also the agricultural profit has been offseted with +2500000 cedis to remove the negative values. The log transformation reduces the influence of outliers on both dependent and independent variable.

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.99635 -0.19745 -0.07265  0.13199  1.76127

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  1.498e+01  3.022e-02 495.606  < 2e-16 ***
highest_educ_level           1.897e-02  3.919e-03   4.841 1.37e-06 ***
I(hh_highest_educ^2)        -1.772e-03  3.766e-04  -4.704 2.69e-06 ***
factor(hh_is_male)1          7.151e-02  1.554e-02   4.603 4.38e-06 ***
factor(empl_cash_crops_y)1   1.224e-01  2.057e-02   5.954 3.00e-09 ***
imputed_income               1.204e-07  4.893e-08   2.461 0.013926 *
motorable_road_y            -1.202e-01  1.995e-02  -6.025 1.95e-09 ***
electricity_y               -1.204e-02  2.140e-02  -0.563 0.573750
factor(water_y)1            -7.887e-02  2.987e-02  -2.640 0.008334 **
public_transport_y          -5.858e-03  1.603e-02  -0.365 0.714866
extension_centre_y           3.473e-02  2.250e-02   1.544 0.122729
cooperative_y               -6.470e-02  1.627e-02  -3.976 7.22e-05 ***
no_of_tractors              -2.274e-03  8.588e-03  -0.265 0.791155
insecticides_herbicides_y    3.611e-02  1.596e-02   2.263 0.023747 *
rice_husking_machine_y       4.469e-02  2.501e-02   1.787 0.074113 .
chemical_fertilizer_y        1.908e-02  1.469e-02   1.299 0.194059
irrigated_fields_y          -6.937e-02  2.987e-02  -2.322 0.020307 *
region                      -1.454e-02  3.789e-03  -3.837 0.000128 ***
factor(ez)2                  1.426e-01  2.263e-02   6.302 3.48e-10 ***
factor(ez)3                  1.045e-01  2.689e-02   3.887 0.000104 ***
farm_area_acres              8.171e-03  5.400e-04  15.132  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3217 on 2467 degrees of freedom
Multiple R-squared:  0.2577,    Adjusted R-squared:  0.2517
F-statistic: 42.82 on 20 and 2467 DF,  p-value: < 2.2e-16
```

We are using polynomial regression for household head education since it has non linear relationship with agricultural profit. The p-value of household head education level is very low , hence it has a significant effect on the agricultural profit. As the value of household head education level increases by an year, the agricultural profit decreases by 0.08 square cedis.

The p - value for electricity_y, public_transport_y, extension_centre_y, no_of_tractors, chemical_fertilizer_y are greater than than the significance level of 0.1. This implies that these variables are not statistically significant in the determination of agricultural profit.

The p - value for Highest_educ_level, (hh_highest_educ)square, factor(hh_is_male)1, factor(empl_cash_crops_y)1, motorable_road_y, cooperative_y, region, factor(ez)2, factor(ez)3, farm_area_acres are less than the significance level of 0.001. Each of these variables are significant with respect to agricultural profit while holding other dependent variables in the model constant .

The p - value for factor(water_y)1 are less than than the significance level of 0.01. Each of these variables are significant with respect to agricultural profit while holding other dependent variables in the model constant .

The p - value for imputed_income, insecticides_herbicides_y, irrigated_fields_y is less than than the significance level of 0.05. This variable is significant with respect to agricultural profit while holding other dependent variables in the model constant .

The p - value for rice_husking_machine_y is less than than the significance level of 0.1. This variable is significant with respect to agricultural profit while holding other dependent variables in the model constant .

Here value of multiple $R-squared$ is 25.77%. This statistic tells us that there is 26% correlation between the independent variable and response variable.
Adjusted $R-squared$ is 25.17% . It is smaller than R - squared because it fits all the independent variables to adjust the model.

Keeping other variables constant, agricultural profit increases by 0.02 % per increase in highest education level in an household.

Keeping other variables constant, agricultural profit increases by 0.1224 % per increase in cash crops grown.

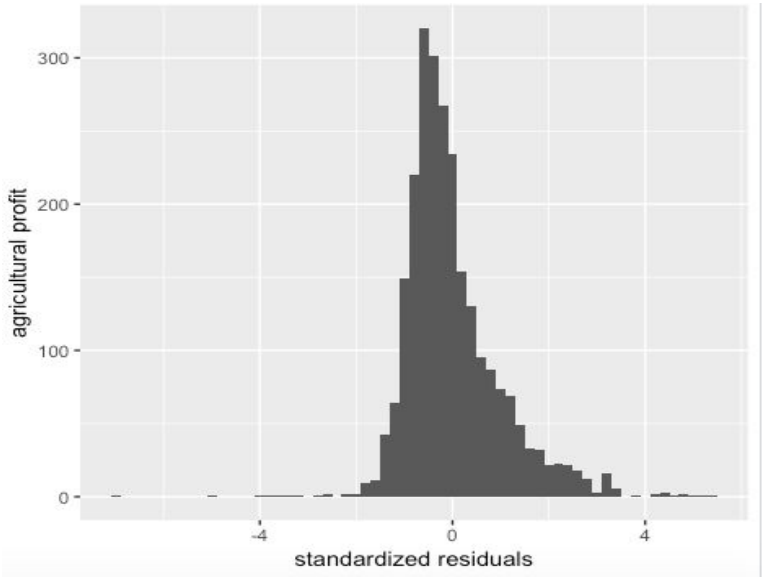
Dummy Variables :

- factor(hh_is_male)1 : when (household head is male) = 1, then agricultural profit increases by 0.96 cedis. As compared with (household head is male) = 0, then its effect on agricultural profit will be 0.
- factor(empl_cash_crops)1 : when (empl_cash_crops) = 1, then agricultural profit increases by 0.45 cedis. As compared with (empl_cash_crops) = 0, then its effect on agricultural profit will be 0.
- factor(water_y)1 : when (water_y) = 1, then agricultural profit decreases by 1.067 cedis. As compared with (water_y) = 0, then its effect on agricultural profit will be 0.
- factor(ez)2 : when (ez) = 2, then agricultural profit increases by 0.522 cedis. As compared with (ez) != 2, then its effect on agricultural profit will be 0.
- factor(ez)3 : when (ez) = 3, then agricultural profit increases by 0.384 cedis. As compared with (ez) != 3, then its effect on agricultural profit will be 0.

Results

Standardized Residuals

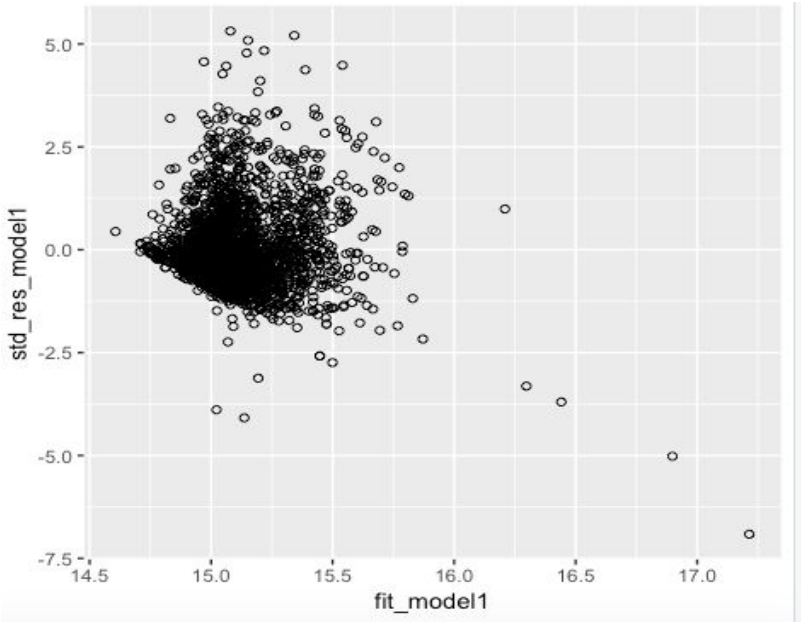
The standardized residual is a measure of the strength of the difference between observed and expected values. The histogram plot with x axis as the standardized residual appears slightly skewed to the right. This means that we do not have normal distribution which implies that the condition that the error terms are normally distributed is met.



Residuals vs Fitted Values

The residuals appear on the y axis and the fitted values appear on the x axis. Also, the residual = 0 line corresponds to the estimated regression line.Following are the interpretations from the scatter plot of Residuals and Fitted values:

- The residuals are heteroscedastic.
- Most of the residuals are scattered around the '0' line. This means that the relationship is somewhat linear.
- Many residual stands out from the basic random pattern of residuals, this means that there are outliers.



Conclusion:

As per our analysis, we arrived at a conclusion that the more acres of farmland area a household has, the more is the profit. Also education level of the household has a positive effect on agricultural profit. Also the effect is positive if the household is in community where they have access to modern machinery like “rice-husking machine” which reduces human labor and provides support to farmers through extension centers. This could be one domain where ACME can look into.

Moreover, If ACME corporation invests in cash crops and related activities, if will be profitable for them because cash crops are more profitable than the food crops. Insecticide, herbicides and chemical fertilizers can also be considered as another domain where the investment will yield profit because the analysis shows that these factors are affecting agricultural profit in positive way.

Cooperative societies and areas with motorable roads seem to have a marginal negative effect on agriculture profit. Hence, they could be avoided by the corporation.