

Determinants of Wheat Yield Variation of Smallholders in South Eastern Ethiopia: Application of Ordered Logistic Regression

Tolesa Alemu*

Ethiopian Institute of Agricultural Research, Kulumsa Research Center, P.O. Box: 489, Asella, Ethiopia

Abstract	Article Information
<p>There is high wheat yield variation among smallholders in the study area. The yield gap is mainly due to socio-economic, farm inputs utilization and institutional service characteristics of the farmers. This study identifies the determinant factors that affect the probability of wheat yield tends to be relatively low, medium or high given a set of yield influencing factors. Cross-sectional data collected from randomly selected farm households and ordered logistic regression were used to attain the objective of the study. The study found out that seed rate, rates of herbicides and fungicides applications, use of hand weeding and agro-ecological location of household were significant factors that determine the probability of wheat yield level to be low, medium or high. The findings imply that agro-ecology based seed rate, weed and disease management efforts are needed to improve wheat yield provided that the recommended rates of seed, herbicides and fungicides are properly applied.</p>	<p>Article History: Received : 01-07-2015 Revised : 09-09-2015 Accepted : 14-09-2015</p> <hr/> <p>Keywords: Wheat Determinants for yield variation Ordered logit Ethiopia</p> <hr/> <p>*Corresponding Author: Tolesa Alemu</p> <p>E-mail: tolesaalemu@yahoo.com</p>

Copyright©2015 STAR Journal, Wollega University. All Rights Reserved.

INTRODUCTION

Low crop yield is one of the challenges of Ethiopian to ensure food security and self sufficiency. Wheat is one of the major crops with low yield and high demand. The country imports wheat every year or every other year from abroad because of the low level of wheat production and the high demand from consumers. Increasing farm-level production and yield i.e. enhancing the current national mean yield of 24.5 quintals per hectare (CSA, 2014), is of prime importance to overcome the low level of wheat production, and meet the increasing consumer demand. Moreover, increasing wheat production and yield mitigates the effect of higher food prices, and it is one of the long term solutions to overcome wheat food problems and meet the demand of the growing population of the country.

Wheat is one of the prominent food and cash crops in the country in general and in the south eastern part in particular. It is the first important cereal crop in the south eastern part (especially in Arsi and West Arsi zones) with annual production of about 9.63 million quintals produced on area of 0.31 million hectares (CSA, 2014). Based on CSA data of 2014, wheat occupied about 39% of the total cereal area in Arsi and West Arsi zones with an average yield of about 30 quintals per hectare in 2014 cropping season. This is still low yield compared to the world average of over 40 quintals per hectare. The low yield has made the country unable to meet the high demand and the country remained net importer of wheat despite its good potential for wheat production.

Given that wheat is such an essential crop in south eastern Ethiopia, it is of critical importance to identify

ways how to enhance wheat yield at smallholders' farm levels. Smallholder wheat production is mainly characterized by traditional and subsistence production with low input use and low yield. Therefore, increasing wheat yield requires adoption of improved agricultural practices and technologies as well as efficient utilization of farm inputs (Dorosh and Rashid, 2013).

Agricultural extension activities have focused on improving the adoption, promotion and scaling-up of improved agricultural technologies and farming practices in the study area. Though great efforts have been made to improve wheat yield for many years, the national mean yield is about 24.5 quintal per hectare with higher yield variations among farmers. There are considerable numbers of farmers who get yield much below or above the mean national yield. Hence, it is crucial to identify the socioeconomic, institutional and biophysical factors that determine the relatively low, medium and high wheat yield levels among farmers. Empirical knowledge is limited on the comparative analysis of the factors that are responsible for the low, medium and high wheat yield levels in the country in general and in the study area in particular. Therefore, identification of determinant factors for variation of yield level among farmers helps for targeted agricultural extension activities and for the successful adoption, promotion and scaling-up of improved agricultural technologies and farming practices. To contribute to this knowledge gap, therefore, this study was under taken with the objective of identifying determinant socio-economic and institutional factors for the relatively low, medium and high wheat yield levels among smallholder farmers.

MATERIALS AND METHODS

Description of the Study Area

The study area, Arsi zone, is found in Oromia Regional State and south eastern part of Ethiopia. The zone is mainly divided into five agro-climatic zones mainly due to variation in altitude. It is dominantly characterized by moderately cool followed by cool and warm annual weather. Cool/cold type of weather is found in the highland areas of Kaka, Chilalo, Hankolo and *Gugu* mountains with altitudes of 4245, 4005, 3850 and 3625 meters above sea level, respectively. Moderately warm weather is found in the lowland areas of Dodota, Ziway-dugda, Jeju, Amigna, Seru and Merti districts. Some highland districts include Lemu-bilbilo, Dodola, Chole and Hankolo-wabe, whereas Hetosa, Tiyo and Munesa districts mainly fall in midland.

The rainy season in the zones starts in June with the highest concentration in July and August while *Belg* season extends from February to April. There is a variation in amount from one place to another. The central highlands receive a mean annual rainfall between 1000-1400mm where as the lowland receives a mean annual rainfall between 700-1000mm. Moreover, the rainy days of the zone varies from 120 to 200 days in the highland parts and slightly decreases as one goes down to the lowland areas. These characteristics make the zones good potential for production of various agricultural crops.

The major cereal crops grown in the zone are wheat, barley, maize, teff and sorghum. However, the contribution of wheat to annual total cereal output is the highest (44%), and it also accounts for about 38% of the total cereal area cultivated in the zone in 2014 (CSA, 2014). This shows the higher relative importance of wheat in food crop cultivation in the zone. However, nationally wheat accounted for about 16% of the total cereal area cultivated in 2014 cropping season (CSA, 2014).

Sampling Methods

A combination of purposive and three stages probability sampling procedures were used for sample selection. In purposive sampling, Arsi zone which is high potential wheat producers in Ethiopia was selected. In the first stage of probability sampling procedure, a list of major wheat producing districts within the zone was prepared based on the level of agricultural extension activities on wheat production. From the lists of districts, three districts were randomly selected. The selected districts were, namely, Dodota, Hetosa, and Lemu-Bilbilo.

In the second stage of the probability sampling, a list of major wheat growing Farmers' Associations (*kebeles*) within the selected districts were obtained from district agricultural offices, and then two *kebeles* were randomly selected from each district from the lists prepared. In the third and final stage, a list of farm households was obtained from the selected *kebeles'* agricultural extension offices. The list of farm households was assigned consecutive serial numbers. Sample farm households were selected by simple random sampling technique.

To determine a representative sample size for the study, population size of wheat farm households was taken from respective district agricultural offices. Hence, using 95% level of confidence and chi-square value for one degree of freedom, and proportion of population assumed to be 0.5 with degree of accuracy of 0.05, the

sample size was determined based on the formula given by Krejcie and Morgan (1970):

$$n = \frac{\chi^2 NP(1-P)}{d^2 (N-1) + \chi^2 P(1-P)} \dots\dots\dots(1)$$

Where:

n = required sample size

χ^2 = tabulated value of chi-square for 1 degree of freedom at 5% significance level (3.841)

N = the population size which is the size of wheat farm households

P = proportion of population assumed to be 0.5 since this would provide maximum sample size

d = the degree of accuracy expressed as proportion (0.05) i.e. standard error

The total number of wheat producer households in the three districts was 49,546 (Arsi Arsi zone and respective districts agricultural development offices, unprocessed data obtained through personal communication). Using equation (1) and the given values, the sample size (n) was calculated to be 381. Allocation of sample size to each district was determined proportional to the size of wheat farm household population of each district.

The sample size for each district was determined as:

$$n_i = n \left(\frac{N_i}{N} \right), \text{ and } n = \sum n_i \dots\dots(2)$$

Where n_i is the sample size from each selected district (i = districts, n is total sample size of the study which is the sum of the sample size of the three districts, and N_i is total wheat farm households in respective district, and N is the total population or wheat farm households of the three districts combined).

Methods of Data Collection

The data was collected from both primary and secondary sources. Cross-sectional data were collected from the survey of randomly selected sample farm households. Structured and pre-tested questionnaire was used to collect primary data. Enumerators were employed and trained and the interview was conducted in 2014. Both quantitative and qualitative information were collected. The data collection included households' demographic and socioeconomic characteristics (household sizes, age and sex structures, education, etc), land holding (agricultural, grazing, wheat land, and others), farm inputs utilization (seeds, fertilizers, herbicides and fungicides, labor utilization, credit, extension services), farm outputs, input and output prices, livestock holding, income sources, row planting and hand weeding, etc. Secondary information on rainfall amounts (annual mean and cropping season), temperature, etc. were also collected. Published and unpublished documents were used as sources of secondary data.

Method of Data Analysis

The data were analyzed using descriptive and inferential statistics and econometric methods. Descriptive and inferential statistics such as mean, percent, standard deviation, F-test, Chi-square test, likelihood ratio test, and analysis of variance (ANOVA) were used to analyze the data collected from sample households. Ordered logistic regression was employed to identify the factors influencing wheat yield at different levels. For analyses, STATA software version 12 was used.

Wheat yield varies among farm households in the study area. There are farm households who get yield exceeding 50 quintals per hectare whereas other get much more below the average yield of sample households (Table 2). The yield difference among households could be the differences in socioeconomic and institutional as well as biophysical circumstances of farm households. To identify determinant factors for yield variation, ordered logistic regression could be used. In ordered logistic regression, we have a qualitative dependent variable, that is, the response variable is ordinal. The model considers the ordered aspect of the outcome. An ordered logit model provides a means to exploit the ordering information.

There is more than one "ordered logit" model. The model fit by ologit in STATA, which referred as the ordered logit model, is also known as the proportional odds model. Another popular choice, not fit by ologit, is known as the stereotype model. The proportional-odds ordered logit model is so called because, if we consider the odds $(k) = P(Y \leq k) / P(Y > k)$, then odds (k_1) and odds (k_2) have the same ratio for all independent variable combinations. The model is based on the principle that the only effect of combining adjoining categories in ordered categorical regression problems should be a loss of efficiency in estimating the regression parameters (McCullagh 1980). The stereotype model rejects the principle on which the ordered logit model is based. Anderson (1984) argues that there are two distinct types of ordered categorical variables: "grouped continuous", such as income, where the "type a" model applies; and "assessed", such as extent of pain relief, where the stereotype model applies. The present study used ordered logit model with grouped continuous categorical variable since wheat yield of households was a "grouped continuous" ordered dependent variable.

To identify the determinant factors for each yield category, ordered logistic regression model of ordinal variable ($Y =$ yield category of low, medium and high) was regressed on independent variables (socioeconomic and institutional factors). Ordered logit models are used to estimate relationships between an ordinal dependent variable and a set of independent variables. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes. This study assumed Y values as 1, 2 and 3 for the yield categories of low, medium and high, respectively. For the ordered categorization of yield, the mean and standard deviation of sample households' yield was used. Accordingly, yield levels falling within range of mean plus or minus one standard deviation were grouped as medium, and those falling in the range of mean minus more than one standard deviation grouped as low, and those with mean plus more than one standard deviation grouped as high yield level.

In ordered logit, an underlying score is estimated as a linear function of the independent variables and a set of cutpoints. The probability of observing outcome i corresponds to the probability that the estimated linear function, plus random error, is within the range of the cutpoints estimated for the outcome:

$$\Pr(\text{outcome } j = i) = \Pr(K_{i-1} < B_1X_{1j} + B_2X_{2j} + \dots + B_kX_{kj} + U_j \leq K_i). U_j \quad (3)$$

This is assumed to be logistically distributed in ordered logit. We estimate the coefficients B_1, B_2, \dots, B_k together with the cutpoints K_1, K_2, \dots, K_{k-1} , where K is the number of possible outcomes. K_0 is taken as $-\infty$, and K_k is taken as $+\infty$.

Based on Greene and Hensher (2009) and Greene (2012), let there is an observed ordinal variable, Y , and Y , in turn, is a function of another variable, Y^* , that is not measured (which is a continuous, unmeasured latent variable Y^* , whose values determine what the observed ordinal variable Y equals). The continuous latent variable Y^* has various threshold points (k). The value on the observed variable Y depends on whether or not you have crossed a particular threshold. Therefore, in the population, the continuous latent variable Y^* is equal to:

$$Y_i^* = \sum_{k=1}^K \beta_k X_{ki} + \varepsilon_i = Z_i + \varepsilon_i \quad (4)$$

Where ε is a random disturbance term with standard logistic distribution.

The Ordered Logit Model estimates part of the above:

$$Z_i = \sum_{k=1}^K \beta_k X_{ki} = E(Y_i^*) \quad (5)$$

The K β s and the $M-1$ k s are parameters that need to be estimated.

Since there is no intercept term, we then use the estimated $M-1$ cutoff terms to estimate the probability that Y will take on a particular value. The estimation equation is:

$$P(Y_i > J) = \frac{\exp(X_i\beta - k_j)}{1 + [\exp(X_i\beta - k_j)]}, \quad J = 1, 2, \dots, M - 1, \quad (6)$$

This implies

$$P(Y_i = 1) = 1 - \frac{\exp(X_i\beta - k_1)}{1 + [\exp(X_i\beta - k_1)]} \quad (7)$$

$$P(Y_i = J) = \frac{\exp(X_i\beta - k_{j-1})}{1 + [\exp(X_i\beta - k_{j-1})]} - \frac{\exp(X_i\beta - k_j)}{1 + [\exp(X_i\beta - k_j)]}, \quad J=2, \dots, M-1, \quad (8)$$

$$P(Y_i = M) = \frac{\exp(X_i\beta - k_{M-1})}{1 + [\exp(X_i\beta - k_{M-1})]} \quad (9)$$

In this case, since Y has three categories, we have $M - 3$ that can be simplified to:

$$P(Y = 1) = \frac{1}{1 + \exp(Z_i - k_1)} \quad (10)$$

$$P(Y = 2) = \frac{1}{1 + \exp(Z_i - k_2)} - \frac{1}{1 + \exp(Z_i - k_1)} \quad (11)$$

$$P(Y = 3) = 1 - \frac{1}{1 + \exp(Z_i - k_2)} \quad (12)$$

Hence, using the estimated value of Z and the assumed logistic distribution of the disturbance term, the ordered logit model can be used to estimate the probability that the unobserved variable Y^* falls within the various threshold limits.

There are various farmers' related socioeconomic and institutional service factors that could affect wheat yield levels of smallholders in Ethiopia. They mainly include age and educational level of household head, access to and use of improved seed, size of agricultural land holding, household and livestock holding sizes, levels of

farm inputs (seed, chemical fertilizers, manures, herbicides, fungicides) utilization in wheat production, hand weeding, practices of row planting and crop rotation, off-farm income and agro-ecological location of farm households. These explanatory (independent) variables

were hypothesized to influence the probability that the each outcome of variable Y falls within the various threshold limits (low, medium and high yield levels). The descriptions and measurements of the variables have been given in Table 1.

Table 1: Descriptions and measurement of variables used in the ordered logit model

Variables	Measurement	Descriptions
Dependent (Y)	Ordinal (low, medium, and high)	Y is low if yield < 16 q/ha, medium if it falls within $yield \geq 16 \text{ and } \leq 32 \text{ q/ha}$, and high if yield > 32 q/ha
Independent (Xs):		
Age	Years	Age of household head
Education	Grades	Educational level of household head in grades
Seed	Yes/No	Access and use of improved seed, 1 if yes and 0 otherwise
Land	Hectares	Total wheat land cultivated
Household size	Adult equivalent	Household members involved in farming
Livestock	TLU	Livestock holding size in tropical livestock unit (TLU)
Seed rate	Quintal	Quantity of wheat seed used per hectare
Off-farm income	ETB	Household annual off-farm income in thousands ETB
Fertilizers	Kilogram	Quantity of chemical fertilizers used per hectare
Herbicides	Liters	Quantity of herbicides used per hectare
Fungicides	Liters	Quantity of fungicides used per hectare
Manure	Quintal	Quantity of manure used per hectare
Hand weeding	Yes/No	Use of hand weeding of wheat farm
Row planting	Yes/No	Practice of planting wheat in row, 1 if yes, and 0 otherwise
Crop rotation	Yes/No	Household practice of crop rotation, 1 if yes 0 otherwise
Agro-ecology	low/mid/highland	Household agro-ecological setting, 1 if lowland, 2 if midland and 3 if highland

RESULTS AND DISCUSSION

Descriptive Results

The average wheat yield of sample households was about 25 q/ha with minimum and maximum yield levels of 6 and 57 q/ha, respectively (Table 2). This shows that there is high yield difference among sample households with variance and standard deviation of about 70 and 8 q/ha, respectively. The yield distribution is close to normal with median yield and coefficient of skewness of 24 and 0.29, respectively. It can be concluded that there is a high yield variation among sample households, and great efforts are required to reduce the high yield gap between the low and high yield levels of smallholder producers.

Table 2: Descriptive statistics of wheat yield of households

S.N	Statistics	Yield in quintals per hectare
1	Mean	24.95
2	Standard deviation	8.38
3	Minimum	6.00
4	Maximum	57.28
5	Median	24.00
6	Variance	70.16
7	Skewness	0.29
8	Kurtosis	3.23
9	Observations	381

Table 3 depicts the number and percent of households falling within yield category of low, medium and high. Majority of the farmers (about 62%) produced wheat yield of medium level i.e. wheat yield falling within the range of 16 to 32 q/ha. High wheat yield (> 32 q/ha) producers comprise about 24% of the sample households whereas the low yield category sample households constitute 14

percent. This implies that about 75% of sample households have a chance of pushing (increasing) their wheat yield in to the high yield class if appropriate farm technologies and input levels are used given their current socio-economic and institutional contexts. Minimizing constraints related to farm technologies, socio-economic and institutional services could enhance the yield levels of low and medium category producers.

Table 3: Percent and number of households within different yield category

Yield category	Frequency	Percent	Cumulative frequency
Low	54	14.17	14.17
Medium	235	61.68	75.85
High	92	24.15	100
Total	381	100	

The mean yield difference among the yield categories was judged through analysis of variance (Table 4). The results of the analysis shows significant F statistic (560.59) with p -value of 0.0001, implying statistically significant mean yield difference among low, medium and high yield categories of smallholder producers at 1% significance level. Bartlett's test for equal variance also resulted in significant chi-square statistic (24.28) with p -value of less than 0.01, depicting statistically significant yield variance among the three yield categories. Therefore, it is very essential to identify the factors that are responsible for households' yield variation that make yield fall in the low, medium and high yield categories. Enhancing wheat yield needs targeting the factors that are responsible for making wheat yield relatively low or high.

Table 4: Analysis of mean yield variance of the yield categories

Source	Analysis of variance				
	SS	df	MS	F	Prob > F
Between groups	19,939.51	2	9969.75	560.59*	0.000
Within groups	6,722.53	378	17.78		
Total	26,662.03	380	70.16		

Bartlett's test for equal variance: $\chi^2(2) = 24.28^*$; Prob> $\chi^2=0.000$

*Significant at 1% significance level

The next econometric estimation result provides the factors that determine the probability of the yield being relatively low, medium or high given a set of socioeconomic and institutional characteristics of farm households.

Econometric Estimation Results

The result of ordered logistic regression (using ologit command in Stata) after the fifth iteration is shown in Table 5. The model chi-square is 140.83 with 16 degrees of freedom (d.f). Its value is used to test whether all the coefficients in the model are different from zero. It is highly significant chi-square statistic with p -value less than 0.01. This tells that the explanatory variables included in the model have significant effect on wheat yield level of smallholder producers. Similarly, the test for the hypothesis that each coefficient is different from zero has been indicated as z with two tail p -values listed as $p>|z|$ (Table 5). The McFadden R^2 or Pseudo R^2 is 0.20. The test for model assumption i.e. test of proportional odds assumption using Brant's test gave insignificant overall chi-square value suggesting that the ordered logistic regression (ologit) assumptions are met. The positive coefficients for variables show that on all other things being equal basis, households with more units of positive variables tend to have higher level of wheat yield.

Similarly, the negative coefficients imply that households with more units of the variables tend to have low level of wheat yield.

However, only six variables are statistically significant at 1, 5 or 10 percent significance level. The significant variables with positive coefficients tell us that all other things being equal, households that practice hand weeding, use adequate seed and fungicides rates tend to have higher yield in wheat production. Likewise, households agro-ecological location, amount of herbicide applied per hectare and access to improved seed have significant effect on the level of wheat yield. When household's agro-ecological location increases or shifts to highland, there is going to be a probability of having low yield level. This is in line with the fact that wheat yield is relatively higher in mid altitude than in highland agro-ecology in the study area. Use of improved seed alone could not increase yield unless other appropriate input packages and crop management practices are used along with the improved seed. Most farmers consider the use improved seed alone increase yield and disregard the use of other inputs and management practices. This has negative effect on the level of wheat yield. Similarly, use of inappropriate type and rate of herbicides could have negative effect on the level of yield.

Table 5: Determinant factors for low, medium and high wheat yield levels

Ordered logistic regression		Number of obs =	381
		LR $\chi^2(16)$ =	140.83***
		Prob > χ^2 =	0.0000
Log likelihood =	-279.377	Pseudo R2 =	0.2013
Variables	Coef.	Std. Err.	z p>z
Age of household head	0.007	0.012	0.610 0.539
Educational level	-0.008	0.035	-0.240 0.808
Wheat area	0.059	0.128	0.460 0.643
Crop rotation	0.153	0.275	0.560 0.576
Row Planting	-0.483	0.601	-0.800 0.422
Hand weeding	1.308	0.617	2.120** 0.034
Access to improved seed	-0.442	0.259	-1.710* 0.087
Seed rate	1.135	0.293	3.880*** 0.000
Manure application	0.000	0.010	-0.010 0.995
Herbicide rate	-1.448	0.382	-3.790*** 0.000
Fungicide rate	5.927	1.216	4.870*** 0.000
Household size	-0.036	0.072	-0.490 0.621
Livestock holding	-0.010	0.029	-0.330 0.741
Off-farm income	-0.008	0.017	-0.470 0.635
Fertilizers rate	-0.002	0.004	-0.410 0.685
Agro-ecological location	-1.220	0.247	-4.940*** 0.000
/cut1	-2.709	1.128	
/cut2	1.293	1.111	

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6 presents the predicted probabilities for each yield category when all independent variables are set to their mean values. The mean probability of observing low yield category was 0.1427 in the sample at the mean values of independent variables. The probability of wheat yield being in medium yield category was higher with mean value of 0.6175 ranging from minimum probability of 0.0399 to 0.7618 in the sample.

Table 6: Predicted probabilities for yield categories after ordered logistic regression

Yield	Mean	Std.dev	Minimum	Maximum
Low	0.1427	0.1759	0.0008	0.8911
Medium	0.6175	0.1566	0.0399	0.7618
High	0.2398	0.1978	0.0022	0.9593

CONCLUSION

This study attempted to identify the determinant factors that affect the probability of wheat yield to be relatively low, medium or high among farm households in south eastern Ethiopia given a set of various yield influencing farmers related socio-economic, farm inputs, institutional service and biophysical factors. Cross-sectional data collected from randomly selected farm households and ordered logistic regression were used to identify the determinant factors that were responsible for yield variation among farmers. The study identified that there were high yield differences among farmers mainly due to inappropriate farm inputs utilization and management practices. Wheat seed rate, herbicides and fungicides application rates, practice of hand weeding, access to and use of improved seed and agro-ecological location of the farmers were the determinant factors that affected the probability of wheat yield being relatively low, medium or high among the farmers. The findings imply that there is a need to properly manage seed rate, weeds and wheat diseases to improve wheat yield in the study area. Agricultural extension efforts are required to provide agro-ecology based extension services that focus on the

management and proper application of seed rate, herbicides and fungicides rates, and use of hand weeding.

Conflict of Interest

Conflict of interest none declared.

Acknowledgement

I am grateful to Ethiopian Institute of Agricultural Research (EIAR), Kulumsa Agricultural Research Center and East African Agricultural Productivity Project (EAAPP) for their financial and material support for farm households' data collection. I sincerely thank data enumerators for their kind cooperation for field data collection. I also thank to the field assistant for the painstaking involvement in data entry.

REFERENCES

Anderson, J.A. (1984). Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society, Series B* 46(1): 1-30.

CSA (Central Statistical Agency of Ethiopia). (2014). Report on area and production of major crops, private peasant holdings, *meher* season, Addis Ababa.

Dorosh, P., and Rashid. S. (2013). Food and Agriculture in Ethiopia: Progress and Policy Challenges. University of Pennsylvania Press. Philadelphia, USA.

Greene, W.H. and Hensher, D.A. (2009). Modeling ordered choices. Stern school of business, New York University, New York.

Greene, W.H. (2012). Econometric analysis 7th edition. USA, Pearson education, Inc.

Krejcie, R.V. and Morgan, D.W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement* 30: 607-610.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B* 42: 109-142.