COMPARING OLS AND QUANTILE REGRESSION ESTIMATION TECHNIQUES FOR PRODUCTION ANALYSIS: AN APPLICATION TO GHANAIAN MAIZE FARMS

Acquah De-Graft Henry, Associate Professor
Department of Agricultural Economics and Extension, University of Cape Coast,
Cape Coast, Ghana
E-mail: henrydegraftacquah@yahoo.com

ABSTRACT
Relying on the mean approach in the estimation of the Cobb-Douglas production function may cloud the empirical linkage between farm specific inputs and outputs in production analysis. In order to address this issue, this paper introduces the quantile regression estimation method to modelling the Cobb-Douglas production function as an alternative to the least squares approach in maize production analysis. The quantile regression results are compared to the widely used ordinary least squares regression. Results of the OLS estimation find an insignificant linkage between labour and maize output as well as fertilizer and maize output. However, the quantile regression estimation reveals a significant positive link between labour and maize output as well as fertilizer and maize output. OLS underestimate and overestimate the true relationship between inputs and output depending on the position of the quantile. These results suggest that the quantile regression methods might provide a more suitable framework for exploring the input output relationship in production analysis than the least squares method when the coefficient of inputs varies significantly across the distribution of farm output.

KEY WORDS
Production, function, quantile regression estimation, ordinary least squares estimation.

The use of the Cobb-Douglas (1928) econometric methodology has dominated the production analysis literature as a tool for estimating the relationship between inputs and outputs. This methodology relies on Ordinary Least Squares (OLS) to estimate the Cobb Douglas Model.

Previous studies (Bagi and Huang (1983), Battese and Coelli (1988), Bravo-Ureta and Rieger (1991), Onumah and Acquah (2010), Hoyo, Espino and Toribio (2004) estimating the relationship between inputs and output relies on OLS or Maximum Likelihood method to estimate the marginal effect of the specific input on the conditional mean function of the output. Some studies (Acquah (2016), Essilfie, Asiamah and Nimoh (2011), Kuwornu, Amoah and Wayo (2013)) estimating the relationship between farm specific inputs and maize output relied on OLS or Maximum Likelihood method to estimate the marginal effect of the farm specific input on the conditional mean function of the maize output. Such estimates ignore the potential heterogeneous patterns of the influence of the covariates in the conditional distribution and provides limited information on the relationship between the inputs and outputs.

Quantile regression approach has been developed as a robust alternative to the least square approach that does not ignore the potential heterogeneous patterns of the influence of the covariates in the conditional distribution. The quantile regression unlike the least squares method has the advantage of providing a detailed picture of the relationship between the covariates and the dependent variable. Applications of the quantile regression in linear modelling has been met with success and demonstrated in some studies (Koenker and Bassett (1978, 1982) and Buchinsky (1998).

Though the mean approach has been extensively used in empirical production analysis research, it does not provide a detailed picture of the relationship between inputs and outputs or alternatively clouds the empirical linkage between inputs and outputs. However, very little
research has been undertaken to demonstrate that empirical linkage between farm specific inputs and outputs can be clouded by the use of a mean approach.

A fundamental research question which remains unanswered is does the mean approach (OLS) provide limited information on the input-output relationship in maize production analysis? Alternatively, is the empirical linkage between input and output clouded by the use of the mean approach in maize production analysis? The foregoing discussion points to the fact that there is the need to employ methods that provide a detailed picture of the relationship between farm specific inputs and outputs.

Some studies (Acquah (2016), Essilifie, Asiamah and Nimoh (2011), Kuwornu, Amoah and Wayo (2013)) employed the Cobb Douglas production function to analyse the effect of farm specific factors on maize output in Ghana using least squares method or maximum likelihood estimation. These studies found insignificant relationship between fertilizer and farm output. Will an alternative approach such as the quantile regression lead to the same conclusion or find an insignificant relationship between fertilizer and farm output when applied to the data used in Acquah (2016)? Notably, the least squares approach provides limited information on the effects of input on output. Damiani and Ricci (2011) note that quantile regression method offers notable advantages over the least-squares method when the coefficient of inputs varies significantly across the distribution of firm productivity. They claim that estimation of the Cobb-Douglas production function using the mean approach may mask substantial heterogeneity in the effects of input variables on firm output.

The purpose of this paper is to support the claim that the mean approach (OLS) may cloud the empirical linkage between farm specific inputs and outputs, and in so doing, demonstrate the usefulness of the quantile regression approach in providing a detailed picture of the relationship between farm specific inputs and output in maize production analysis.

MATERIALS AND METHODS OF RESEARCH

The present study is interested in comparing the OLS and Quantile regression models in the estimation of input output relationship. The study therefore uses the input output data used by Acquah (2016). The data consist of farm output and farm specific inputs (labour, land, equipment, agrochemical, fertilizer, and seeds) of 306 randomly selected maize farmers from the Ejura Sekyedumase District in Ashanti region of Ghana.

The Cobb-Douglas functional form of production functions is widely used to represent the relationship of an output to inputs and dates back to Von Thünen (1850), Wicksell (1923) and Wicksell (1934) but is credited to Cobb and Douglas (1928) and given by:

\[ y = f(\mathbf{x})e^u = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}e^u \]  

with \( \beta_1 + \beta_2 = 1 \)

Then the log model can be specified as:

\[ \ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + u \]  

The ordinary least squares (OLS) technique or the quantile regression model can be employed in the estimation of the Cobb-Douglas regression model. 

Estimation with Ordinary Least Squares (OLS). Assuming a given data consist of explanatory and outcome variables and there exist a linear functional relationship between the outcome variable and the explanatory variable, then the general linear model can be of the matrix form:
Where: $y = (y_1, ..., y_n)^T$, $u = (u_1, ..., u_n)^T$, $\beta = (\beta_1, ..., \beta_n)^T$ and

$$X = \begin{pmatrix} 1 & x_{11} & x_n \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{nn} \end{pmatrix}$$

Ordinary Least Squares (OLS) minimizes the square distances between the observed and the predicted dependent variable $y$:

$$S(\beta) = \sum_{i=1}^{N} (y_i - x_i^T \beta)^2 = (y - x \beta)^T (y - x \beta) \rightarrow \min_{\beta}$$

(4)

Differentiating with respect to $\beta$, setting it to zero, we find $\hat{\beta}$ satisfies:

$$X^T X \hat{\beta} = X^T y$$

(5)

The OLS method of estimating $\beta$ can be estimated by $\hat{\beta}$ which is given by the explicit formula:

$$\beta = (X^T X)^{-1} X^T y$$

(6)

The dependent variable can be predicted by: $\hat{y} = x_i^T \hat{\beta}$ and the error term by $\hat{u}_i = y_i - x_i^T \hat{\beta}$. $\hat{u}_i$ is the residual. The OLS assumes that the errors are normally distributed with a mean of zero, constant variance and are uncorrelated. Additionally, observations are independently and identically distributed (iid) and there is a linear functional relationship between dependent and explanatory variables.

Quantile Regression Estimation. Koenker and Bassett (1978, 1982) first introduced the quantile regression approach as a robust alternative to the ordinary least squares regression. Estimating conditional quantiles at various points of the distribution of the dependent variable will allow us to trace out different marginal responses of the dependent variable to changes in the explanatory variables at these points.

Buchinsky (1998) asserts that two additional characteristics of the quantile regression are noteworthy. First, the classical properties of efficiency and minimum variance of the least squares estimator are obtained under the restrictive assumptions of independently and identically normally distributed (iid) errors. When the distribution of the errors is non normal, the quantile regression estimator may be more efficient than the least squares estimator.

Secondly, since the objective function for the quantile regression estimator is a weighted sum of absolute deviations, the parameter estimates are robust to outliers. The quantile regression model can be written as:

$$y_{it} = X_{it} \beta_q + u_{it}$$

(7)

with $\text{Quant}_q(y_{it} / x_{it}) = X_{it} \beta_q$

Where: $y$ is the dependent variable, $x$ is a vector of regressors, $\beta$ is the vector of parameters to be estimated and $u$ is a vector of residuals. $\text{Quant}_q(y_{it} / x_{it})$ denotes the q-th conditional quantile of $y$ given $x$. The q-th regression quantile, $0 < q < 1$ solves the following problem.
RESULT AND DISCUSSION

Quantile Regression and Ordinary Least Squares for Estimation of Cobb-Douglas Model

Ordinary Least Squares (OLS) estimation results of the Cobb-Douglas regression model presented in Table 1, reveals significant and positive relationship between land and equipment as explanatory variables and maize yield as the dependent variable. There is also a significant but negative relationship between the use of agrochemicals and maize yield as dependent variable. There is also a positive and insignificant relationship between labour and fertilizer as explanatory variables and maize yield as dependent variable.

Alternatively, quantile regression estimation results of the Cobb-Douglas model presented in Table 1, shows a significant positive relationship between labour and maize output at the 10th, 20th and 30th quantile. Again, the quantile regression show a significant positive relationship between fertilizer and maize output at the 20th and 50th quantile.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Quantile (0.10)</th>
<th>Quantile (0.20)</th>
<th>Quantile (0.30)</th>
<th>Quantile (0.40)</th>
<th>Quantile (0.50)</th>
<th>Quantile (0.60)</th>
<th>Quantile (0.70)</th>
<th>Quantile (0.80)</th>
<th>Quantile (0.90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.259***</td>
<td>(0.321)</td>
<td>4.471***</td>
<td>(0.455)</td>
<td>4.358***</td>
<td>(0.321)</td>
<td>4.741***</td>
<td>(0.496)</td>
<td>5.431***</td>
<td>(0.462)</td>
</tr>
<tr>
<td>log(Labour)</td>
<td>0.093</td>
<td>(0.046)</td>
<td>0.102</td>
<td>(0.062)</td>
<td>0.128**</td>
<td>(0.044)</td>
<td>0.134**</td>
<td>(0.069)</td>
<td>0.147</td>
<td>(0.064)</td>
</tr>
<tr>
<td>log(Land)</td>
<td>1.256***</td>
<td>(0.062)</td>
<td>1.235**</td>
<td>(0.084)</td>
<td>1.210**</td>
<td>(0.057)</td>
<td>1.237**</td>
<td>(0.073)</td>
<td>1.275**</td>
<td>(0.086)</td>
</tr>
<tr>
<td>log(Equipment)</td>
<td>0.069**</td>
<td>(0.024)</td>
<td>0.123**</td>
<td>(0.033)</td>
<td>0.086**</td>
<td>(0.025)</td>
<td>0.078**</td>
<td>(0.034)</td>
<td>0.045</td>
<td>(0.032)</td>
</tr>
<tr>
<td>log(Agrochemicals)</td>
<td>-0.140</td>
<td>(0.065)</td>
<td>-0.175</td>
<td>(0.090)</td>
<td>-0.099</td>
<td>(0.062)</td>
<td>-0.128</td>
<td>(0.087)</td>
<td>-0.099</td>
<td>(0.083)</td>
</tr>
<tr>
<td>log(Fertilizer)</td>
<td>0.051</td>
<td>(0.054)</td>
<td>0.067</td>
<td>(0.082)</td>
<td>0.096</td>
<td>(0.053)</td>
<td>0.068</td>
<td>(0.078)</td>
<td>0.017</td>
<td>(0.066)</td>
</tr>
<tr>
<td>log(Seed)</td>
<td>-0.151</td>
<td>(0.078)</td>
<td>-0.065</td>
<td>(0.107)</td>
<td>-0.144</td>
<td>(0.071)</td>
<td>-0.246</td>
<td>(0.104)</td>
<td>-0.245</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Returns to scale</td>
<td>1.140</td>
<td>1.300</td>
<td>1.300</td>
<td>1.143</td>
<td>1.941</td>
<td>1.163</td>
<td>1.093</td>
<td>1.067</td>
<td>1.146</td>
<td>1.024</td>
</tr>
</tbody>
</table>

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

Land remains positive and significant at all quantiles whilst equipment remains positive and significant at all quantiles except the 40th quantile. There is also a significant negative quantile relationship between agrochemicals and maize output at the 10th, 50th, 60th, and 70th and 80th quantile and a significant negative relationship between seeds and maize output at the 20th, 30th, 40th, 50th, 60th, 70th and 90th quantile.

Noticeably, the positive but insignificant relationship between labour and maize outputs revealed by the OLS technique is not supported by the quantile regression. The quantile estimation reveals a totally different picture. As displayed in Table 1, there is a significant positive relationship between labour and maize output at lower quantiles. The results suggests that the mean approach of the OLS has clouded the role of labour on maize output at the different points. Notably, these results suggest that labour is an important determinant for maize farms with low output.

Again, results from the traditional OLS reveals positive and insignificant relationship between fertilizer and maize output. Whereas the quantile regression show a significant positive relationship between fertilizer and maize output at the 20th and 50th quantile. In effect, fertilizer is an important determinant in maize farms with low or moderate output. These results are consistent with Abawiera and Dadson (2016), who noted that, maize farmers can increase their output by employing more resources (fertilizer and labour).
Similarly, Essilfie, Asiamah, and Nimoh (2011) in an empirical analysis found labour to have a positive and significant effect on maize yield. Kuwornu, Amoah and Wayo (2013) also found that hired labour has a positive and significant effect on maize yield.

Results from the traditional OLS and quantile regression reveals significant positive relationship between maize output and explanatory variables (land, equipment) and a significant negative relationship between maize output and explanatory variables (agrochemical and seed). Like the OLS, the quantile regression results suggest that an increase in each of these variables (land, equipment) will lead to an increase in the output of maize. And an increase in each of these variables (agrochemical and seed) will lead to a decrease in maize output. Thus when the seeds used by farmers are higher than the recommended seed rate, yield will decline. This may lead to overcrowding which makes seedlings compete for nutrients, space and air. This result is consistent with the studies by Battese and Hassan (1999).

In addition, the quantile regression reveals a clearer representation depicting defined quantiles wherein each variable maintain a significant effect on maize output. For example, land remains positive and significant at all quantiles whilst equipment remains positive and significant at all quantiles except the 40th quantile. These results suggest that land is an important determinant of maize farm output regardless of the levels or distribution of maize outputs. Also equipment remains an important factor in maize production at all levels of output except the 40th quantile. There is a significant negative relationship between agrochemicals and maize output at the 10th, 50th, 60th, and 70th and 80th quantile. These results also show that agrochemicals significantly reduce the output of maize farmers with very low output and also for those with medium to high levels of output. Again, the significant negative relationship between maize output and seeds show that increases in seeds reduce maize output for farmers with low, moderate and high levels output. In effect, excessive use of agrochemicals and seeds could lead to a decline in yield. This result is consistent with the studies by Battese and Hassan (1999) who found yield to decline due to competition for nutrients by overcrowded seeds.

Generally, these results show that, the input-output relationship can be clouded by the mean approach. Furthermore, depending on the levels of production of the farmers' output, specific inputs could be important determinants or not. Hence the need to consider the effects of the covariates/inputs at the various levels or distribution of maize output remains imperative.

A comparison of the estimation result from the Cobb-Douglas model using the least squares method and the quantile regression approach indicates that the estimates obtained in the alternative methods are different. For example results from farm-level production analysis in Table 1 show that, the magnitude of each coefficient is different in each quantile as well as in the OLS estimation. Furthermore, not all statistically significant input variables in the quantile regression were statistically significant in the OLS regression. The difference between OLS and Quantile regression results in Table 1, suggests that OLS underestimate and overestimate the true relationship between inputs and output depending on the position of the quantile. These results are consistent with Vu, Holmes, Lim and Tran (2014) who demonstrated that the mean approach (OLS) can mask the differences in the effects of inputs on output in an export profitability relationship. Similarly, Damiani and Ricci (2011) note that the estimation of the Cobb-Douglas production function using the mean approach may mask substantial heterogeneity in the effects of input variables on firm output.

Using the OLS and quantile regression approaches, the productivity level of the maize farmers were studied by investigating their output elasticities and returns to scale. On the basis of the OLS and quantile regression, it is noted that if the farmers increase input (labour, equipment, agrochemicals, fertilizer and seed) by one percent, output changes by less than one percent, whilst if farmers increase input (land) by one percent, output increases by more than one percent. Prominently, land which is positive and significantly related to output had the highest elasticity. This suggest that, increasing land used in maize production will lead to increases in maize output. The importance of land in production is also noted by Rahman, Wiboonpongse, Sriboonchitta and Chaovanapoonphol (2009).
Notably, both the OLS and the quantile regression estimation techniques suggest that the maize farmers were exhibiting increasing returns to scale. Thus output grows more than proportionately with any increase in input. This evidence is consistent with Wu, Devadoss and Lu (2003). This means that farmers could increase output by using more of the inputs that are significantly positively related to the output at various distribution of output.

CONCLUSION

The main point of this paper is that empirical linkage between inputs and outputs in production analysis should not be based solely on the mean approach. This is because the empirical linkage between inputs and outputs can be clouded by the use of the mean approach. In order to address the issue this paper proposes a quantile regression estimation method to modelling the Cobb-Douglas production function as an alternative to the ordinary least squares estimation approach. A comparison of the result from the Cobb-Douglas model using the least squares method and the quantile regression approach indicates that the estimates obtained in the alternative methods differ. Notably, OLS underestimate and overestimate the true relationship between inputs and output depending on the position of the quantile. 

On the basis of the quantile regression estimation, farm inputs such as labour and fertilizer had a significant positive effect on maize output whilst the OLS found an insignificant positive relationship between labour and fertilizer as explanatory variables and maize output as the dependent variable. On the basis of the OLS and quantile regression, land and equipment had significant positive effect on maize output, whilst agrochemicals and seed had a significant negative effect on output. Furthermore, the OLS and quantile regression suggest that the farmers were operating at a increasing returns to scale though the magnitude varied across quantiles. In summary, this paper has demonstrated that quantile regression offers an alternative and a more detailed picture of the relationship between inputs and outputs in a maize production analysis.

REFERENCES


