Emir. J. Food Agric. 2013. 25 (1): 81-88
doi: 10.9755/ejfa.v25i1.9060
http://www.ejfa.info/

AGRICULTURAL ECONOMICS

Model development for wheat production: Outliers and multicollinearity problem in Cobb-Douglas production function

Maryouma E. Enaami[1*], Zulkifley Mohamed[2] and Sazelli A. Ghani[2]

[1]*Department of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya*
[2]*Department of Mathematics, Faculty of Science and Mathematics, Sultan Idris University of Education, Perak, Malaysia*

Abstract

Despite the important role that production function has played in growth literature, few attempts have been made to change the methodology to estimate it. The Cobb-Douglas functions are among the best known production functions utilized in applied production analysis. This paper describes development of a new model based on Cobb-Douglas production function with the used of robust method and partial least squares path modeling for parameter estimation. The new model attempted to solve two main problems in modeling namely the issue of multicollinearity and outliers. Each issue was handled separately but using the same method of least square for parameter estimations. This paper goes on to provide an overview of the measurements and structural criteria needed for model development and, also to introduce a robust partial least squares-path modeling for the Cobb-Douglas production function (RPLS-PM-CD). The researcher hypothesizes that utilization of the minimum covariance determinant (MCD) provides an estimate by the measurement model and expresses the structural relationships between the latent variables through the partial least squares-path modeling (PLS-PM). The inputs and outputs of the RPLS-PM-CD were based on agricultural wheat production data pertaining to Al-Kufra Agricultural Production Project. This paper is more theoretical and should be seen as a new way to estimate Cobb-Douglas production function.

Key words: Cobb-Douglas production function, Wheat production, The minimum covariance determinant (MCD), Partial least squares-path modeling (PLS-PM), Multicollinearity, Outliers, Ordinary least squares (OLS)

Introduction

Production functions are a basic component of all economics domains. As such, estimation of production functions have a long history in applied economics, starting in the early 1800's. Unfortunately, this history cannot be deemed an unqualified success, as many of the econometric problems that hampered early estimation are still an issue today (Ackerberg et al., 2006). A large number of research papers dealing with Cobb-Douglas production functions published in the area of agricultural economics is a testimony to the important role played by these models, by using ordinary least squares regression methodology (Prajneshu, 2008). But the OLS is not the best method. In economic theory, the multicollinearity arising in least squares estimation of the Cobb-Douglas model is not new. It is a problem that emerged with the model itself, when evaluating the 1928 work of Cobb and Douglas (Cobb and Douglas, 1928). On the other hand, in applied economics and econometrics, it has always been highlighted that even if a small amount of data behaves differently from the vast majority of the observations, classical estimations may be affected, leading to results that are not representative of the population. In other words, the presence of outliers might bias the results. The researchers proposed the development of Cobb-Douglas production function parameter through the partial least squares- Path modeling (RPLS-PM) method to be applied on Libyan Agriculture sector data. It attempted to solve two main problems in the model namely the issues of multicollinearity and outliers. Each of the issues was handled separately using the same method of least square for parameter estimations.

Received 27 September 2011; Revised 17 March 2012; Accepted 27 March 2012; Published Online 24 November 2012

*Corresponding Author

Maryouma E. Enaami
Department of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya

Email: reem521@vahoo.com

Multicollinearity Problem in Cobb-Douglas Production Function

Murthy (2002) suggestion for access to the best estimate of the Cobb Douglas is to take dissimilar industries such that capital/labour varies, but would be fitted across a heterogeneous group of industries. Olarinde and Manyong (2008) stated that in the economic theory multicollinearity commonly occurs because of the nature of aggregation of economic data (Wethrill, 1986; Gujarati, 2006). They used of ridge regression which overcomes the problem of multicollinearity by adding a small quantity to the diagonal of X^1. Zhang and Shang (2009) suggested that the relationships between influential factors and rural infrastructure are often not so complex. Zhang and Shang adopted the PLS method in real data (rural infrastructure) to avoid the preceding limitations of OLS by using Cobb-Douglas production function after log-transformation. In order to verify the precision of the PLS method, they compared this method with the classic methods and found that the PLS methods are obviously more precise than the others.

The minimum covariance determinant (MCD)

The minimum covariance determinant (MCD) method of Rousseeuw (1985) is a highly robust estimator of multivariate location and scatter. Its objective is to find h observations (out of n) whose covariance matrix has the lowest determinant.

Partial least squares- path modeling (PLS-PM)

Partial Least Square- Path Modeling (PLS-PM) is a statistical approach for modeling complex multivariable relationships among observed and latent variables (Vinzi at el., 2009). The structural (or path) coefficients are estimated through ordinary least squares (OLS) multiple/ simple regressions among the estimated latent variable scores. Partial least squares regression (PLSR) can nicely replace OLS regression for estimating path coefficients whenever one or more of the following problems occur: missing latent variable scores, strongly correlated latent variables, a limited number of units as compared to the number of predictors in the most complex structural equation. A PLS-PM is described by two models: (1) a measurement model relating the manifest variables (MVs) to their own latent variables (LVs) and (2) a structural model relating some endogenous LVs to other LVs. The structural model is called the inner model and the measurement model is also called the outer model (Tenenhaus et al., 2005).

Sources of Data and Construction of Variables

The most important projects of agricultural productivity in Libya are in three regions. The first region is the southwest, area of Fezzan, which has four most important agricultural projects, Mknusp Project, Barjuj Project, Aldboat Project, and Deeseh Project. The second region is the area of the central region, Abouchebp which has the Abouchebp project. The third region is the South East region which has the largest agricultural projects in Libya namely, the Alsrero project and the Alkufra project. The latter is one of the most important agricultural project and the largest in Libya. In this paper, data were collected from the Al- Kufra project, involving agricultural production of the wheat crop from 1960 to 2010. This data covered almost all important economic activity inputs. The following variables have been used to estimate the model:

Output Items: the unit for wheat production is tone per hectare.

Input Items: -Land data requirements include: *Harvested area* thousand hectare of land farmed in wheat crop and *water use*: Million cubic meters of water / ha. Agricultural inputs: Average per hectare: *seeds* Ton/ ha, *Chemical fertilizers*: DAP, Urea 46%, TSP, Mixed, Micro elements (Tons / ha), and *Pesticides*: Herbicide, Insecticide, Insecticide (Liter/ ha). Hours of operation hours / day (*Labour* use): Labour input is measured in person - year equivalent of workers directly engaged in production in farming, and *wages*, salaries & admin cost: Average cost per season: Wages & salaries, Social security, Camping costs, Administrative costs. Operating and maintenance: Average cost per season: *Electricity, Fuel, Spare parts, and Oil & lubrications.*

Description of the Model

The production function is specified as a Cobb-Douglas function in the form of:

$$W=\alpha_0\, y_1^{\alpha_1}\, y_2^{\alpha_2}\, y_3^{\alpha_3}\, y_4^{\alpha_4}\, y_{15}^{\alpha_5}\, y_6^{\alpha_6}\, y_7^{\alpha_7}\, y_8^{\alpha_8}\, y_9^{\alpha_9}\, y_{10}^{\alpha_{10}}\, y_{11}^{\alpha_{11}} \quad (1)$$

where W is the wheat crop output; the coefficient α_0 is the total factor efficiency parameter for composite primary factor inputs in sector i; the parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9$ *and* α_{10} are production elasticities; y_1 = water; y_2 = land; y_3 = seeds; y_4 = chemical fertilizers; y_5 = pesticides; y_6 = operation hours; y_7 = wages; y_8 = spare parts; y_9 = fuel; y_{10} = oil and lubricants; and y_{11} = electricity. Equation (1) demonstrates that the relationships between the

output and inputs are non-linear. However, by taking the logs of both sides of the equation, the following log-linear relationship is obtained:

$$\log W = \log \alpha_0 + \alpha_1 \log y_1 + \alpha_2 \log y_2 + \alpha_3 \log y_3 + \alpha_4 \log y_4 + \alpha_5 \log y_5 + \alpha_6 \log y_6 + \alpha_7 \log y_7 + \alpha_8 \log y_8 + \alpha_9 \log y_9 + \alpha_{10} \log y_{10} + \alpha_{11} \log y_{11} + e \quad (2)$$

This equation is described as log-linear because both the dependent variable and the regressors have been log-transformed. The coefficients in log-linear equations are elasticities. The Cobb-Douglas production function is estimated using the OLS method. The strength of this paper is embodied chiefly by the development of a Cobb-Douglas production function model based on a robust method and PLS-PM for parameter estimation. The presumption of this investigation is the presence of significant, strong, relationships between the inputs and the wheat production outputs. Besides that, the researcher created four separate data sets, and named groups of data according to variables using the names of the underlying data that represent a particular group.

The researcher denoted; the endogenous latent variable of the *land* and *water* (*LW*) is denoted by η_1. *LW* is formed by two indicator variables y_1, and y_2. The measurement error for *LW* is represented by δ_1 and δ_2. Also, endogenous latent variable for agriculture inputs, *AR* is labeled as η_2. *AR* is formed by three indicator variables $y_3, y_4,$ and y_5 (seed, chemical fertilizer and, pesticides). The measurement error for this indicator variables are represented by $\delta_3, \delta_4,$ and δ_5 respectively. According to Khan and Hossain (2007), wheat output is affected by seed, chemical fertilizer and, pesticides. While; the *Hour operation and Wages (HW*) latent variables consist of exogenous latent variable *HW*, which is symbolized as ξ_1, and two indicator variables x_1 and x_2. and, the measurement error is represented by ε_1 , and ε_2. The *Electricity, Fuel, Spare parts* and, *Oil & lubricants (OP)* model consists of exogenous latent variables and the four indicators, represented by x_3, x_4, x_5 and, x_6. On the other hand, the measurement error is represented by $\varepsilon_3, \varepsilon_4, \varepsilon_5$ *and*, ε_6. The *Output* in the final set deals with relationship between the indicator variables (y_6) and an endogenous latent variables (η_3), and is the model's output. The output model is the form of formative measurement. One way to develop a model of wheat production is based on the concept of inputs and finding correlations between them. Formulation of this model is based on the results of research conducted by Kherallah (2000), Keyzer et al. (2000), Boisvert and Chang (2006), and Khana and Hossaib (2007) confirming a significant correlation between the *Output* and *LW, AR, HW*, and *OP*.

The structure matrix of the measurement model *LW, AR, HW, OP*, and *Output* are respectively as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} [\eta_1] + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

(3)The measurement model of *LW*

$$\begin{bmatrix} y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} [\eta_2] + \begin{bmatrix} \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

(4) The measurement model of *AR*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} [\xi_1] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

(5)The measurement model of *HW*

$$\begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \end{bmatrix} [\xi_2] + \begin{bmatrix} \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

(6)The measurement model of *OP*

$$[y_6] = [\lambda_6][\eta_3] + [\delta_6]$$

(7)The measurement model of *Output*

The *Output*, *AR, LW* structural model describes the relationships between the *Output* endogenous latent variables and the *LW, AR, HW*, and *OP* exogenous latent variables. The *Output, AR, and LW* structural model is hence described as:

$\eta_1 = \Lambda_{21}\eta_2 + \Omega_{11}\xi_1 + \Omega_{21}\xi_2 + \zeta_1$ $\eta_2 = \Omega_{12}\xi_1 + \Omega_{22}\xi_2 + \zeta_2$

(8)The measurement model of *LW* (9)The measurement model of *AR*

$\eta_3 = \Lambda_{13}\eta_1 + \Lambda_{23}\eta_2 + \Omega_{13}\xi_1 + \Omega_{23}\xi_2 + \zeta_3$

(10) The measurement model of *Output*

The overall structural and measurement models of *LW, AR, HW, OP*, and the *Output* will be described in Figures 1-9 all as the following:

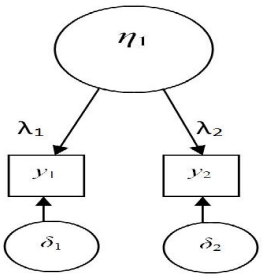

Figure 1. The *LW* Measurement Model

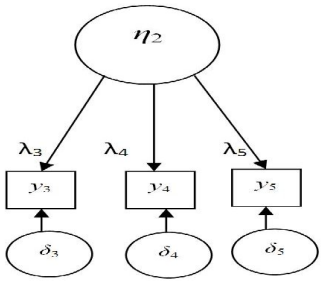

Figure 2. The *AR* Measurement Model

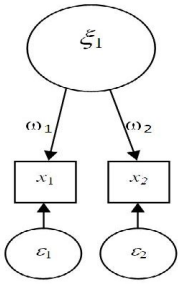

Figure 3.The *HW* Measurement Model

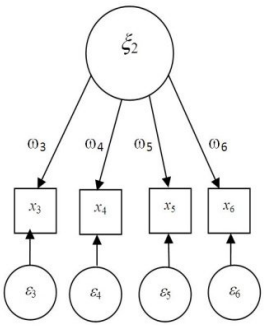

Figure 4. The *OP* Measurement Model

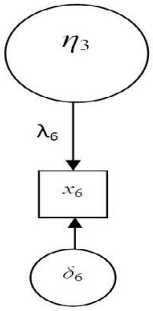

Figure 5. The Measurement Model of *Output*

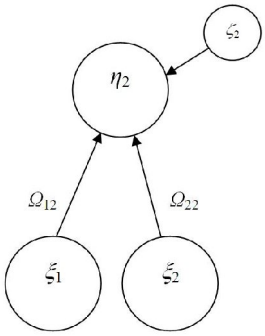

Figure 6. The Structural Model of *AR*

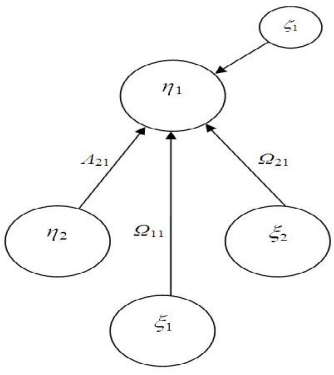

Figure 7. The Structural Model of *LW*

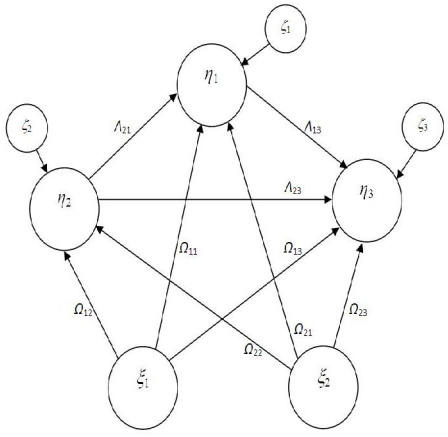

Figure 8. The Structural Model of *Output*

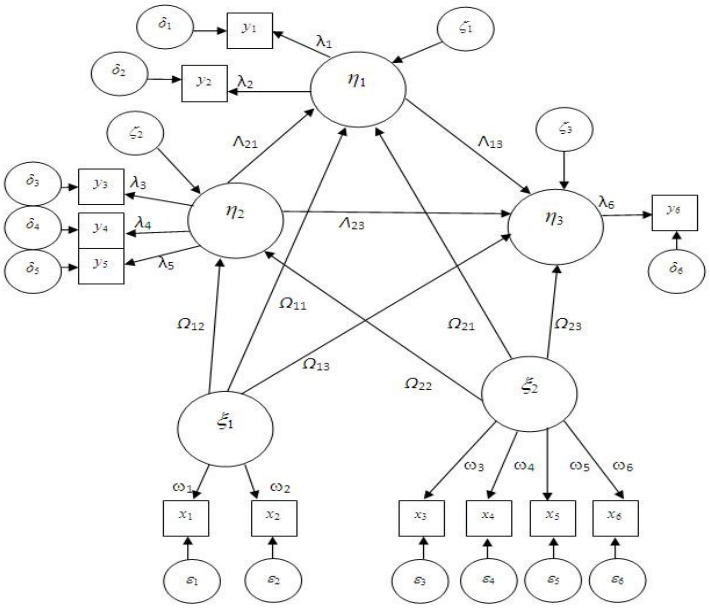

Figure 10. Structural Model and Measurement Model *LW*, *AR*, *HW, OP*, and *Output*.

Results

The Cobb Douglas production function results can be written as follows:

$$\begin{aligned} LogW &= 0.89 - 0.213\ Log\ y_1 + 0.959\ Log\ y_2 - 0.124\ Log\ y_3 \\ &\quad - 0.464\ Log\ y_4 - 0.016\ Log\ y_5 - 0.269\ Logy_6 + 0.259\ Logy_7 \\ &\quad + 1.51\ Logy_8 + 0.001\ Log\ y_9 - .296\ Log\ y_{10} + 0.101\ Log\ y_{11}. \end{aligned} \quad (3)$$

The negative coefficients for the pesticide, water, seeds, hours of operation, and oil and lubrications variables indicate that these variables have an inverse impact on the dependent variable which is in conflict with the logic of production functions.

Outliers and Multicollinearity

In this paper, the distance-distance plot (D-D plot) can be employed for detection of outliers and multicollinearity as was suggested by Field (2000) who considered the correlation coefficients between the independent variables and also the VIF in the diagnosis of multicollinearity. The distance plot showed in Figures 10 below the MCD versus the MD plots which indicate that more than 15 points clearly stand out, whereas the classical analysis shows no points stand out. While, the variance inflation factor (VIF) values are all greater than 10 except for the wages variable. Also, the correlation coefficients results between the independent variables are very high which considered a strong evidence of the existence of multicollinearity. In this paper the researcher seeks to determine a viable technique that can be employed to overcome the outliers and multicollinearity problems in the Cobb-Douglas production function. Therefore, the RPLS-PM is proposed to overcome the process of outliers and multicollinearity in this paper.

Cobb-Douglas Production Function Through Partial Least Squares- Path Modeling (PLS-PM-CD) Model

Evaluation of a research model using PLS-PM analysis consists of two distinct steps. The first step includes assessment of the measurement model and deals with the evaluation of the characteristics of the latent variables and measurement items that represent them. The second step involves the assessment of the structural model and involves evaluation of the relationships between the LVs. And the overall structure model according to the PLS-PM structure provides three different fit indices: the communality index, the redundancy index, and the Goodness of Fit (GOF) index. Results of the assessment demonstrated that the research model successfully passed the test of composite reliability where the Cronbach's Alpha

coefficient was greater than 0.70. Furthermore, the results of the assessment of reliability of the individual measures illustrated that the individual loadings of all items were greater than 0.87, which indicates that the proposed model satisfied the reliability of the individual items as well. On the other hand, the results of the assessment pointed out that the research model successfully passed the test of Average Variance Extracted (AVE) assessment which implies high reliability of the measures. The results presented in Table 1.

Assessment of the structural model: The essential criterion for this assessment is the coefficient of determination (R^2) of the endogenous latent variables and estimates for path coefficients. The results described the relationship between the LVs by correlation matrix. The correlation for the model was greater than 0.70, which implies a fairly strong positive relationship as shown in Table 2. The researcher found that there was a relationship between the LVs which indicates that the developed model performs well. The estimated values for path relationships in the structural model should be evaluated in terms of sign, magnitude, and significance. Assessment of the structural model includes testing for significance of the hypothesized relationships between the research model constructs. Once the path coefficients between the two constructs in the model have been calculated, their significance and the significance level of the path can be evaluated. The results of the bootstrap re-sampling procedure with different numbers of re-samples were very stable with respect to the number of re-samples. The results in the case of 200 re-samples pointed out that the model worked well.

Finally, Assessment of the overall structural model. A global criterion of goodness of fit has been proposed by Tenenhaus et al. (2005), which is the goodness of fit (GoF) index. The GoF index is obtained by finding the geometric mean of the average communality index and the average R^2 value. The value of the GoF index was 0.6882 presented in Table3. This index is bounded between 0 and 1. Since 0.688 is within that range; it is concluded that the structural model works well.

Summary

The body of this work yielded the Cobb-Douglas production function which is used extensively in theoretical and applied research. The design of a Cobb-Douglas production function model consists of few steps. First, the general model structure should be determined; input and output parameters as well as their mutual relationships. Then, parameter values should be determined by first linearizing the models through logarithmic transformation and then applying the method of least squares to the linearized parameters. But the log transformation process does not get rid of all the outliers and multicollinearity problems. At this point the researcher attempted to identify the problems facing this approach of research and to find solutions to identified problems. This research is focused on tackling the problems of outliers and multicollinearity by using MCD and PLS-PM. RPLS-PM-CD model shows that the model consists of five measurement and structure models, and describes the structural relationships between the *LW, AR, HW* and, *OP*, besides their structural relationships with the *Output* based on output and inputs of agricultural wheat production in Al- Kufra Agricultural Production Project, Libya. The path diagram in Figure 9 describes an ideal RPLS-PM-CD model. The terms simple structure and unidimensional measurement are used in the general factor analysis framework to signify a model that meets two conditions: (i) Each latent variable is defined by a subset of indicator variables which are strong indicators of that latent variable, and (ii) each indicator variable is strongly related to the other latent variables. Finally, the findings of the model correspond to very good results and provide important new insights. This supports the point that if it solves problems, not only in agriculture sector but may be the most effective path to macroeconomic development.

Table 1. Reliability assessment of the various interested components.

Construct	Composite Reliability	AVE	Squared Root of AVE	Cronbach's Alpha
LW	0.851186	0.741321	0.861000	0.772577
AR	0.941525	0.843237	0.918279	0.755356
HW	0.932941	0.874349	0.935066	0.858261
OP	0.981022	0.928194	0.963428	0.686936

Table 2. Coefficients of the correlations between the latent variables.

	output	*LW*	*AR*	*HW*
LW	0.888			
AR	0.802	0.925		
HW	0.871	0.845	0.865	
OP	0.751	0.890	0.935	0.914

Table 3. The structural model.

Block	Mult.RSq	AvResVar	AvCommun	AvRedund	Goodness of Fit
Output	0.8148	0.0000	1.0000	0.8148	
LW	0.8632	0.2587	0.7413	0.6399	
AR	0.8797	0.1568	0.8432	0.7417	
HW	0.0000	0.1257	0.8743	0.0000	
OP	0.0000	0.0718	0.9282	0.0000	
Average	0.5115	0.1272	0.8728	0.3600	0.6882
	Inner model		Outer model		

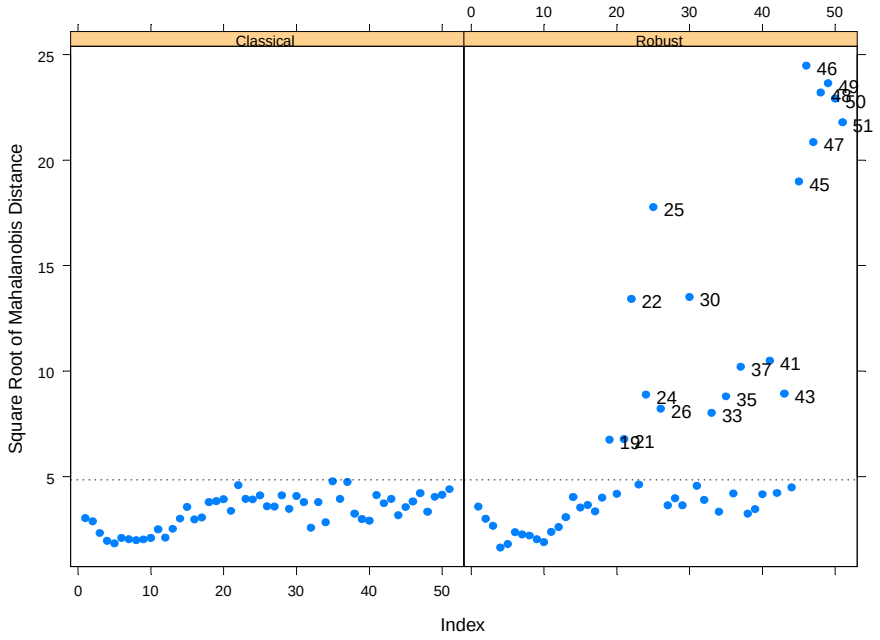

Figure 10. Robust Distances.

Reference

Ackerberg, D., K. Caves and G. Frazer. 2006. Structural Identification of Production Functions, mimeo, R & R Econometrica.

Anderson, Sweeney, D. J., T. A. Williams, J. Freeman and E. Shoesmith. 2009. Statistics for Business and Economics. Cengage Learning EMEA.

Boisvert, N. R. and H. H. Chang. 2006. Multifunctional agricultural policy, reduced domestic support and liberalized trade: an empirical assessment for Taiwanese rice (IWMI, 2006).

Cobb, C. W. and P. H. Douglas. 1928. A Theory of Production. American Econ. Rev. 18:139-165.

Field, A. 2000. Discovering Statistics Using SPSS for Windows. London: Sage Bartlett, J. Gani, Ed. London: Academic Press.

Gujarati, D. N. 2006. Econometria basica. (4th ed.). Rio de Janeiro: Elsevier.

Keyzer, M., M. Merbis and G. Overbosch, Food and Agriculture Organization of the United Nations, Centre for World Food Studies. (2000). WTO, agriculture, and developing countries: the case of Ethiopia. Food and Agriculture Org.

Khan, M. and S. M. A. Hossain. 2007. Study on energy input, output and energy use efficiency of major jute based cropping pattern. Bangladesh J. Sci. Indust. Res. 2(42):195-202.

Kherallah, M., H. Lofgren, P. Gruhn and M. M. Reeder. 2000. Wheat policy reform in Egypt: adjustment of local markets and options for future reforms. Intl Food Policy Res Inst. Business and Economics.

Murthy, K. V. B. 2002. Arguing a Case for Cobb-Douglas Production Function. SSRN eLibrary. (1):20-21.

Olarinde, L. and V. M. Manyong. 2008. Risk aversion and sustainable maize production in Nigeria: Some challenges and prospects for agricultural and economic development. African Association of Agricultural Economists (AAAE)>2007 Second International Conference, Accra, Ghana, Advancing Technical Change in African Agriculture, pp.67-72.

Prajneshu, 2008. Fitting of Cobb-Douglas Production Functions: Revisited. Agric. Econ. Res. Rev. 21:289-292

Rousseeuw, P. J. 1985. Least Median of Squares Regression. J. American Stat. Assoc. 79(388):871-880.

Tenenhaus, M. 2008. Component-based Structural Equation Modelling. Total Qual. Manage. Business Excellence 19(7):871.

Tenenhaus, M., V. E. Vinzi, Y. Chatelin and C. Lauro. 2005. PLS path modeling. Comput. Stat. Data Anal. 48(1):159-205.

Vinzi, V. E., W. W. Chin, J. Henseler and H. Wang. 2009. Handbook of Partial Least Squares: Concepts, Methods and Applications (1st ed.). Springer. Berlin.

Wethrill, P. 1986. Evaluation of ordinary ridge regression. Bull. Math. Stat. 18:1-35.

Shang, W. and Y. Zhang. 2009. The Relationship between Rural Infrastructure and Economic Growth Based on Partial Least-Squares Regression. In Networking and Digital Society, International Conference on (Vol. 2, pp. 127-130). Los Alamitos, CA, USA: IEEE Computer Society.