The Estimation of Cobb-Douglas Production Function Parameter through A Robust Partial Least Squares

Penganggaran Parameter Fungsi Pengeluaran Cobb-Douglas melalui Kaedah Kuasa Dua Terkecil Separa Teguh

Maryouma Enaami, Sazelli Abdul Ghani & Zulkifley Mohamed
Department of Mathematics, Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris
E-mail: reem521@yahoo.com

Abstract

The paper discusses the development of the agricultural production model for the agricultural sector based on the model suggested by the Libyan Government underlining the country policy in promoting this sector. The development of this agricultural production model is through the method of Cobb Douglas production function. As a step of precaution, the robust partial least squares is incorporated in the Cobb Douglas production model to avoid the effect of multicollinearity and outliers in the data. The new model is known as Robust Partial Least Squares of Cobb Douglas Production Function (RPLS-CDF). This study shows that the RPLS-CDF overcomes the problem in Cobb Douglas production model when multicollinearity and the outliers exist in the data as compared to the conventional method namely the log transformation. It is found that the pernicious effect of the outliers have been reduced considerably by the proposed robust estimators.

Keywords Cobb Douglas Production Function, Robust Partial Least Squares, Robust Estimators, Multicollinearity and Outliers

Abstrak

Kertas ini membincangkan pembangunan model pengeluaran pertanian untuk sektor pertanian berdasarkan model yang dicadangkan oleh Kerajaan Libya sebagaimana yang ditekankan oleh dasar negara Libya dalam menggalakkan sektor ini. Pembangunan model pengeluaran pertanian ini adalah melalui kaedah fungsi pengeluaran Cobb Douglas. Sebagai langkah berjaga-jaga, kaedah kuasa dua terkecil separa teguh digabungkan dalam model pengeluaran Cobb Douglas untuk mengelakkan kesan multi kolinearan dan data pencilan. Model baharu dikenali sebagai Kuasa Dua Terkecil Separa Teguh bagi Fungsi Pengeluaran Cobb Douglas (RPLS-CDF). Kajian ini menunjukkan bahawa RPLS-CDF mengatasi masalah dalam model pengeluaran Cobb Douglas apabila wujud multi kolinearan dan data pencilan jika dibandingkan dengan kaedah konvensional iaitu transformasi log. Didapati bahawa kesan yang merosakkan daripada data pencilan telah dikurangkan oleh penganggar teguh yang dicadangkan.

Kata kunci Fungsi Pengeluaran Cobb Douglas, Kuasa Dua Terkecil Separa Teguh, Penganggar Teguh, Multi Kolinearan dan Data Pencilan

Jurnal Sains dan Matematik Vol.3 No.2 (2011) 59-65
ISSN 1985-7918

Introduction

Economical problems underlie many events or problems that seem hard to explain and solve (Webster, 2003). The efforts for economical development have increasingly become important. In applied work, Generalized Cobb-Douglas production function is very much capable of handling multiple inputs (Bhanumurthy, 2002). It can be represented as

$$Q = Ax_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3}\ldots x_n^{\alpha_n} \qquad \text{(i)}$$

where, $x_1, x_2, \ldots, x_n$ are n inputs, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are their elasticities of output with respect to inputs. It is important to note that researchers in many studies use the set of Cobb-Douglas production functions, where it is usually fitted by first linearizing the models through logarithmic transformation and then applying method of ordinary least squares (Prajneshu, 2008). However the Ordinary Least Square (OLS) is not the best estimation method. The OLS method has many problems that might occur. For example, least squares estimates are very sensitive to outliers, particularly in small samples. In addition, in the context of the economic factor, multicollinearity often exists between the economic factors and could greatly affect parameter estimation (Aguirregabiria, 2009). The seriousness of multicollinearity will affect the results mostly negatively, such as increase in the OLS variance, reduce reliability of model and lack of stationarity. Since the rank of parameter estimation is close to zero, the diagonal data of covariance matrixes will become too big, which means the variance inflating factor (VIF) will be infinite. This will eliminate some important explaining variables and reduce reliability of model (Zhang & Shang, 2009). All of these to a great extent decrease the OLS precision and cannot reflect the actual significance while the error variances are not constant or "heteroscedasticity" (Baltagi, 2008). The partial least square (PLS) method is adopted here in analyzing agriculture data to avoid the preceding limitations of OLS. To overcome the above-mentioned problems, robust partial least squares will be improvised by means of robust statistics**.** In this paper, improving the estimation of parameters in Cobb-Douglas production function, through Robust Partial Least Squares (RPLS) is introduced.

Partial Least Squares (PLS) Regression

PLS was designed to cope with problems in data specifically, small datasets, missing values and multicollinearity. In contrast, OLS regression yields unstable results when data has small sample size, missing values and multicollinearity between predictor. In OLS regression increases standard error of their estimated coefficients (Field, 2000). High multicollinearity increases risk of theoretically sound predictor to be rejected from regression model as non-significant variable. This method was first developed by Wold (1966). The goal of PLS is to predict *Y* from *X* and to describe the common structure underlying the two variables (Abdi, 2003). PLS is a regression method that allows for the identification of underlying factors, which are a linear combination of the explanatory variables or *X* (also known as latent variables) which best model the response or *Y* variables (Chin et al., 1996).

The term partial least squares specifically means the computation of the optimal least squares fit to part of a correlation or covariance matrix. The part of the correlation or covariance matrix that the least squares are fitted to is the "cross-block" correlation between the exogenous or *X* variables and the dependent measures or *Y* variables. PLS measures covariation between two or more block of variables and creates a new set of variables that is optimized for maximum covariance (not maximal correlation) using the fewest dimensions . PLS is sometimes called soft modeling because while OLS regression makes hard assumptions such as no multicollinearity in the independent variable, soft modeling refers to softening of these assumptions.

Robust Partial Least Squares (RPLS) Regression

The PLS regression solves the problem of multicollinearity, but the results are affected by outliers. A first class of robust alternatives for PLS regression involves the application of a robust regression method to the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. A second class includes methods which use a robust cross-covariance matrix and a robust regression method. Branden and Hubert (2003) proposed to replace the empirical variance-covariance matrix in the SIMPLS algorithm by the Minimum Covariance Determinant (MCD) and the reweighted MCD (RMCD) estimator. They called this method RSIMPLS (see Camminatiello, 2006).

Minimum Covariance Determinant (MCD)

A well known covariance determinant is the Minimum Covariance Determinant (MCD) by Rousseeuw (1984).The objective of the MCD estimator is to find h observations (out of $\underline{n}$) whose covariance matrix has the lowest determinant. The MCD mean estimator is then the sample mean of those h points and the MCD covariance estimator is their sample covariance matrix. To compute the MCD, one needs an algorithm for finding the best subset of h points, which usually involves the repeated computation of the sample mean and covariance as well as Mahalanobis distances (introduced by Mahalanobis in (1936)).

Atkinson (1993, 1994) proposed the forward search algorithm which also permits the detection of multiple outliers. More recently, Rousseeuw and Driessen (1999) presented a new algorithm called FAST-MCD supposed to be even faster than the forward search algorithm and able to deal with very large data sets. A key idea of the FAST-MCD algorithm is the fact that starting from any approximation to the MCD, it is possible to find an approximation with a lower determinant.

Statistical Methodology

The importance of agriculture sector in the process of economic development is indispensable. With the recognition of this fact, Libya planners have emphasized on the development of agricultural and allied sectors right from the beginning of the economic planning process of Libya. In this paper we attempt to present a model for evaluating Libya agricultural production that incorporates agriculturally induced resource externalities.

We are going to use Cobb Douglas production function to determine the contribution of a particular input in the total production. Studies made earlier in Libya about the agriculture production function (Harvest Report 2008, from National Center for the final results of improved seeds to harvest, and Agricultural Research Center, Tripoli (Libya) (2009)) explain important inputs in wheat and barley crops (some of the main agricultural crops in Libya). These inputs can be categorized as (a) Local admissive: water, electrics, urea fertilizer, seeds and other local entails and Importer admissive: chemical fertilizer, exterminators and other importer entails, and (b) Additive value: human labour and auto labour. In the old empirical Y study (Cobb-Douglas production function) we used the following model:

$$Y_i = \alpha_0 \, x_{11}^{\alpha_1} \, x_{12}^{\alpha_2} \, x_{13}^{\alpha_3} \, x_{14}^{\alpha_4} \, x_{15}^{\alpha_5} \, x_{21}^{\alpha_6} \, x_{22}^{\alpha_7} \, x_{23}^{\alpha_8} \, x_{31}^{\alpha_9} \, x_{32}^{\alpha_{10}} \, e. \qquad i = A, B \qquad \text{(ii)}$$

where Y_i is outputs of crops (A = wheat and B = Barley) and coefficient α_0 is the total factor efficiency parameter for composite primary factor inputs in sector i. 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. And x_{11} = water, x_{12} = electrics, x_{13} = urea fertilizer, x_{14} = seeds and x_{15} = other local entails. x_{21} = chemical fertilizer, x_{22} = exterminators and x_{23} = other importer entails. And x_{31} = human labour and x_{32} = auto labour. From equation (ii), the relationship between output and input is nonlinear; however after using log transformation (traditional method) for this model, we have the following linear model:

$$\log Y = \log \alpha_0 + \alpha_1 \log \alpha_{11} + \alpha_{12} \log x_{12} + \alpha_3 \log x_{13} + \alpha_{14} \log x_{14} + \alpha_5 \log x_{15} + \alpha_6 \log x_{21}$$
$$+ \alpha_7 \log x_{22} + \alpha_8 \log \alpha_{23} + \alpha_9 \log \alpha_{31} + \alpha_{10} \log \alpha_{32} + e. \qquad \text{(iii)}$$

and then apply the OLS methods. However this method is not the best method to estimate the parameter.

Suggested Model

Ideally, we would like to be able to detect problems as such multicollinearity and outliers effects easily and more importantly to provide an accurate estimate of these effects more than the traditional method. The PLS procedure is then used to estimate the latent variables as an exact linear combination of its indicators with the goal of maximizing the explained variance for the indicators and latent variables. Following a series of ordinary least squares analyses, PLS optimally weighs the indicators such that a resulting latent variables estimate can be obtained. We conform to an underlying assumption in this scenario where the latent variables (un observation) would be the dependent variables (*Yi*) and also would be the output of product (wheat and barley) and the predictors would be local admissive (ξ_1), Importer admissive (ξ_2) and Additive value (ξ_3). And independent variables (observation) would be x_{11} = water, x_{12} = electrics, x_{13} = urea fertilizer, x_{14} = seeds and x_{15} = other local entailed, x_{21} = chemical fertilizer, x_{22} = exterminators and x_{23} = other importer entailed, x_{31} = human labour and x_{32} = auto labour and Output (η): y_{11} = wheat, y_{12} = barley. Our suggested model methodology is illustrated by applying it to the information data from the important input for product output (wheat and barley) in Libya Agriculture sector. In our theoretical study we used the following additive model:

$$\eta = \delta_0 + \delta_1 \xi_1 + \delta_2 \xi_2 + \delta_3 \xi_3 + e. \qquad \text{(iv)}$$

where Coefficient δ_0 is the total factor efficiency parameter for composite primary factor inputs in sector i. Parameters δ_1, δ_2 and δ_3 are production elasticities.

The formation of the production model of Libya agricultural sector was based on the recommendations by Kherallah et al. (2000), Mahagayu et al. (2007) and Carver (2009) who opined that there is a correlation between production output (PO) and each of production input (land and water (LAW)); seeds, chemical fertilizers, and pesticide (SCP); hours of operation and wages (HAW); and fuel, spare parts, oil and lubrications and electricity (FSE). The model can be divided into two parts. The external model or measurement model which relates indicator variables to latent variables. The internal model or structural model that relates latent variables to one another. In this research the external model consists of five measurement models namely measurement models of LAW, SCP, HAW, FSE, and PO. The latent variables for measurement models of LAW, SCP, HAW, FSE, and PO were marked as η_1, η_2, η_3, η_4 and η_5 respectively. The indicator variables of these latent variables were land, water, seeds, chemical fertilizers, pesticide, hours of operation, wages, fuel, spare parts, oil and lubrications and electricity, and wheat crop, represented by $x_{11}, x_{12},\ldots, x_{51}$ respectively. While the measurement error for each indicator variable was marked as δ_{12}, δ_{12}, ..., δ_{51} and Λ_{15}, Λ_{12}, Λ_{31}, Λ_{32}, Λ_{35}, Λ_{42}, Λ_{52} were the regression coefficients relating the exogenous latent variables and endogenous latent variables. The overall structural model of Libya Agricultural Production Model of wheat is illustrated in Figure 1.

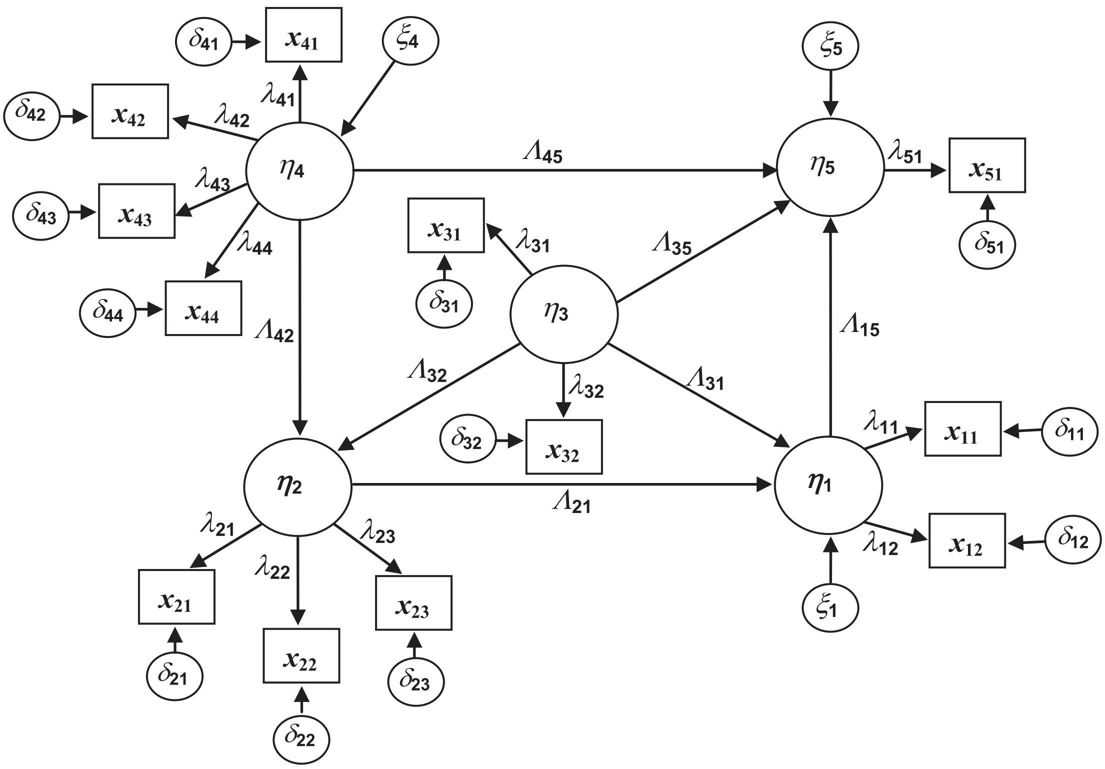

Figure 1 Structural Equation Model of Libya Agricultural Production of Wheat

Jurnal Sains dan Matematik Vol.3 No.2 (2011) 59-65
ISSN 1985-7918

The main structural equation for the Libya Agricultural Production model is as follows:

$$\eta_5 = \Lambda_{15}\eta_1 + \Lambda_{35}\eta_3 + \Lambda_{45}\eta_4 + \xi_5 \quad \text{(v)}$$

Conclusion

In this paper, we have discussed the problems that are often either exaggerated, or are uncritically glossed over, that of multicollinearity and outliers. This study seeks a way to possible application of the robust partial least squares regression (RPLSR) methods in developing Cobb-Douglas production function, to solve these problems. This study benefited the estimation problems of Cobb-Douglas production functions. Empirical studies of production functions, by and large, employ a function of the Cobb-Douglas type for reasons of computational economy. Typically, in this paper; we concentrate on the problems of outliers and multicollinearity, and on solutions that have been proposed to deal with these issues. In addition, 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. That is the main reason why we attempt to develop best methods to obtain estimation in the context of Libya and other countries' agricultural sector by using robust methods.

References

Abdi, H. (2003). Partial Least Squares Regression (PLS regression), In M. Lewis-Beck, A. Bryman, T. Futing (Eds). *Encyclopedia for Research Methods for the Social Sciences*. Sage.Thousand Oaks:CA.

Aguirregabiria. V (2009), Econometric Issues and Methods in the Estimation of Production Functions. *Munich Personal RePEc Archive,* Online at http://mpra.ub.uni-muenchen.de/15973/MPRA Paper No. 15973, posted 30. June 2009 / 01:49.

Atkinson, A. C. (1993). Stalactite Plots and Robust Estimation for the Detection of Multivariate Outliers. In S. Morgenthaler, E. Ronchetti, and W. A. Stahel (Eds.). *New Directions in Statistical Data Analysis and Robustness.* Basel: Birkhäuser.

Atkinson, A. C. (1994). Fast Very Robust Methods for the Detection of Multiple Outliers. *Journal of the American Statistical Association.* 89, 1329-1339.

Baltagi, B. H. (2008). *Econometrics*. *4th edition*. *Springer.* New York: NY..

Bhanumurthy, K.V. (2002). Arguing a Case for the Cobb-Douglas Production Function. *Review of Commerce Studies*. 1, 20-21.

Branden K. V., & Hubert, M. (2003). Robust Methods for Partial Least Squares Regression, *Journal of Chemometrics.* 537 – 549.

Camminatiello. I. (2006). A Robust Approach for Partial Least Squares regression.

Carver, B.F. (2009). *Wheat: Science and Trade.* John Wiley. California: CA.

Chin. W. W., Marcolin. B. L. and Newsted. P. R. (1996). A Partial Least Squares Latent Variable Modeling Approach For Measuring Interaction Effects: Results From A Mote Carlo Simulation Study And Voice Mail Emotion/ Adoption Study. *Proceedings of The Seventh International Conference On Information Systems*. 21-41

Field, A. (2000). *Discovering Statistics Using SPSS for Windows*. Academic Press. London..

Kherallah, M., Lofgren, H., Gruhn, P., and Reeder, M.M. (2000). Wheat Policy Reform in Egypt: Adjusment of Local Markets and Options for Future Reforms. Research Reports No. 115. *International Food Policy Research Institute.* 170.

Mahagayu, M.C., Kamwaga, J., Cndiema, A., Kamundia, J., and Gamba, P. (2007). Wheat Productivity, Constraints Associated in The Eastern Parts of Kenya Timau Division. *Afr. Crop Sc. Conference Proc.* 8, 1211-1214.

Mahalanobis, P. C. (1936). On the Generalised Distance in Statistics . *Proceedings of the National Institute of Sciences of India.* 2 (1): 49–55. Retrieved 2008-11-05.

Prajneshu, (2008). Fitting of Cobb-Douglas Production Functions: *Revisited. Agricultural Economics Research Review.* 21, 289-292.

Rousseeuw, P.J. (1984). Least Median of squares Regression, *Journal of the American statistical association*. 76, 871-880.

Rousseeuw, P. J., & Van Driessen, V. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics.* 41, 212-223.

Webster, T.J. (2003). *Managerial Economics, Theory and Practice*. Elsevier. California: CA.

Wold H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*. Academic Press. New York: NY.

Zhang. Y and Shang, W. (2009). The Relationship Between Rural Infrastructure and Economic Growth Based on Partial Least-Squares Regression. 2009 *International Conference on Networking and Digital Society.* 127-130.