Jurnal Sains dan Matematik Vol.3 No.1 (2011) 32-38
ISSN 1985-7918

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

Penanggaran Parameter Fungi Pengeluaran Cobb-Douglas Melalui Kuasa Dua Terkecil Separa Teguh

Maryouma Enaami, Sazelli Abdul Ghani and Zulkifley Mohamed
Department of Mathematics, Faculty of Science and Mathematic
Universiti Pendidikan Sultan Idris
[1]Corresponding author e-mail:

Abstract

The Cobb-Douglas production function (Cobb and Douglas, 1928) is still today the most ubiquitous form in theoretical and empirical analyses of growth and productivity. The estimation of the parameters of aggregate production functions is central to much of today's work on growth, technological change, productivity, and labour. Empirical estimates of aggregate production functions are a tool of analysis essential in macroeconomics, and important theoretical constructs, such as potential output, technical change, or the demand for labour, are based on them. It is usually fitted by first linearizing the models through logarithmic transformation and then applying method of least squares (Prajneshu, 2008), but the ordinary least squares (OLS) is not the best estimation method (Kahane, 2001). In statistics and econometrics, more and more attention is paid to techniques that can deal with data containing atypical observations, which can arise from outliers, miscoding, or heterogeneity and not captured or presumed in a model. This is of very high importance especially in (non) linear regression models and time series as the least squares (LS) and maximum likelihood estimators (MLE) are heavily influenced by data contamination (Pavel, 2007). In addition, multicollinearity often exists between the economic factors and could greatly affect parameter estimation. The seriousness of multicollinearity will affect the results mostly negatively. Partial least squares (PLS) are especially good in dealing with small sample data, plenty of variables and multicollinearity. It can greatly improve reliability and precision of model (Zhang & Shang, 2009). While Robust Partial Least Squares (RPLS) is used to solve the problems of multicollinearity and outliers. This can be done through Minimum Covariance Determinant (MCD) and the reweighted MCD (RMCD) estimator. This method is called RSIMPLS (Branden & Hubert, 2003). The purpose of this article is to suggest the best method in overcoming the outliers and multicollinearity problems of Cobb-Douglas production function. This is done by using the robust partial least squares (RPLS) method. This developed methodology will be illustrated in the contacts of its theoretical background.

Keywords Cobb-Douglas production function, Minimum Covariance Determinant (MCD), Partial Least Squares (PLS), Robust Partial Least Squares (RPLS).

Abstrak

Fungsi pengeluaran Cobb-Douglas (Cobb-Douglas 1928) masih lagi menjadi fungsi yang digunakan dalam mana-mana analisis mengenai pertumbuhan dan daya pengeluaran baik secara teoritikal mahupun secara empirikal. Pada masa kini, penganggaran bagi parameter-parameter bagi kesemua agregat fungsi-fungsi pengeluaran adalah perkara yang utama dalam bidang kajian mengenai pertumbuhan, perubahan teknologi, daya pengeluaran dan tenaga buruh. Anggaran-anggaran empirikal bagi agregat fungsi-fungsi pengeluaran merupakan alat penganalisisan yang utama dalam bidang makro ekonomi. Konstruk –konstruk teoritikal yang penting seperti hasil keluaran yang mungkin, perubahan teknikal atu daya permintaan tenaga buruh adalah berdasarkan kepada anggaran-anggaran empirikal ini. Anggaran-anggaran ini disesuaikan dengan menjadikan model-model yang digunakan sebagai linear melalui transformasi logaritma dan kemudian mengaplikasikan kaedah kuasa dua terkecil (Prajneshu, 2008). Perlu diingatkan yang kaedah kuasa dua terkecil yang biasa bukanlah kaedah penganggaran yang terbaik (Kahane, 2001). Dalam bidang statistik dan ekonometrik, banyak perhatian telah ditumpukan kepada teknik-teknik yang mampu mengendalikan data yang mengandungi cerapan-cerapan yang atipikal. Ini mungkin timbul kerana wujudnya data yang terpinggir (outliers), kesilapan kod atau keheterogenan dan yang tidak dapat dibataskan atau diandaikan oleh sesuatu model itu. Perkara ini adalah tersangat mustahak terutamanya dalam model-model bukan linear dan siri masa oleh kerana kaedah kuasa dua terkecil dan penganggar-penganggar kemungkinan maksimum dipengaruhi hebat oleh pencemaran data (Pavel, 2007). Ini ditambah pula dengan faktpr multicollinearity yang selalunya wujud di antara faktor-faktor ekonomi dan sangat mempengaruhi penganggaran parameter. Masalah collinearity yang serius akan mempengaruhi hasil keputusan yang diperolehi secara negatif. Kaedah kuasa dua terkecil separa adalah satu kaedah yang bagus dalam mengendalikan data yang kecil, mempunyai banyak pembolehubah dan berhadapan dengan masalah multicollinearity. Penggunaan kaedah ini dapat membantu dengan baik sekali faktor kebolehpercayaan dan kejituan model (Zhang & Shang, 2009) manakala penggunaan kaedah kuasa dua terkecil separa teguh dapat memyelesakan masalah data yang terpinggir (outliers) dan multicollinearity. Kaedah kuasa dua terkecil separa teguh diperolehi melalui Penentu Kovarians Minimum (MCD) dan penganggar MCD yang diberi pemberatan berlainan (reweighted) Kaedah ini kemudiannya dipanggil RSIMPLS (Braden & Hubert, 2003). Tujuan utama artikel ini adalah untuk mencadangkan kaedah terbaik bagi mengatasi masalah-masalah data terpinggir (outliers) dan multicollinearity dalam fungsi pengeluaran Cobb-Douglas. Ini dapat dilakukan dengan menggunakan kaedah kuasa dua terkecil separa teguh. Metodologi yang telah dibangunkan adalah ditujukkan di dalam artikel ini berdasarkan konteks latar belakang teoritikalnya.

Katakunci Fungsi pengeluaran Cobb-Douglas, Penentu Kovarians Minimum (MCD), Kuasa dua terkecil separa (PLS), Kuasa dua terkecil separa teguh (RPLS)

Introduction

Economical problems underlie many events or problems that seem to be 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 $Q = Ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \ldots x_n^{\alpha_n}$ 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, it is usually fitted by first linearizing the models through logarithmic transformation and then applying method of least squares (Prajneshu, 2008), but the 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 a small sample size. In addition, multicollinearity often exists between the economic factors and could greatly affect parameter estimation (Aguirregabiria, 2009).

Jurnal Sains dan Matematik Vol.3 No.1 (2011) 32-38
ISSN 1985-7918

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 significance actually 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, the estimation of parameters in Cobb-Douglas production function, through Robust Partial Least Squares (PLS) is introduced.

Partial Least Squares Regression (PLS)

Partial least squares is designed to cope with problems in data specifically, small data sets, missing values and multicollinearity. In contrast, ordinary least squares (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 partial least squares is to predict dependent variable Y from independent variable X and to describe the common structure underlying the two variables (Abdi, 2003). Partial least squares 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.

The term partial least squares specifically means the computation of the optimal least squares fit to part of a correlation or covariance matrix (McIntosh et al, 2004; Wold, 1982). The part of the correlation or covariance matrix that the least squares are fit to is the "cross-block" correlation between the exogenous or X variable and the dependent measures or Y variable. Partial least squares 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 (McIntosh et al, 1996). Partial least squares is sometimes called soft modeling because while OLS regression makes hard assumptions such as non multicollinearity in the independent variable, soft modeling refers to softening of these assumptions.

Robust Partial Least Squares Regression (RPLSR)

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 estimator is the Minimum Covariance Determinant (MCD) by Rousseeuw (1984). The objective of the MCD estimator is to find h observations (out of 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) proposes the forward search algorithm which also permits the detection of multiple outliers. More recently, Rousseeuw and Driessen (1999) present 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.

Significance of the Study

This study will benefit 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 mean reason why we attempted to develop best methods to obtain estimation in the context of Libya and other countries' agricultural sector by using robust methods.

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 in Libya. In this paper we attempt to present a model for evaluating Libya agricultural production that incorporates agriculturally induced resource externalities.

The relationship between agricultural inputs and outputs is documented in the various studies. We have been using 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 (two of the main agricultural crops in Libya) are local admissive: water, electrics, urea fertilizer, seeds and other local entailed; importer admissive: chemical fertilizer, exterminators and other importer entailed; and additive values: human labour and auto labour.

In the old empirical Y study (Cobb-Douglas production function) we used the following model:

$$Y_i = \alpha_0 \; x_{11}^{\alpha_1} \; x_{12}^{\alpha_2} \; x_{13}^{\alpha_3} \; x_{14}^{\alpha_4} \; x_{15}^{\alpha_5} \; x_{21}^{\alpha_6} \; x_{23}^{\alpha_7} \; x_{24}^{\alpha_8} \; x_{31}^{\alpha_9} \; x_{32}^{\alpha_{10}} \; e.$$

Y_i=α0i=A, B …(i)

Where Y 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 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}$ are production elasticities. And 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. And x_{31} = human labour and x_{32} = auto labour.

$$\log Y = \log\alpha_0 + \alpha_1 \log x_{11} + \alpha_2 \log x_{12} + \alpha_3 \log x_{13} + \alpha_4 \log x_{14} + \alpha_5 \log x_{15}$$
$$+\alpha_6 \log x_{21} + \alpha_7 \log x_{22} + \alpha_8 \log x_{23} + \alpha_9 \log x_{31} + \alpha_{10} \log x_{32} + e. \quad ...(2)$$

From the equation (1) relationship between output and input is nonlinear; however after using log transformation (conventional method) the linear model is as follows:
The above parameters can then be estimated by applying the OLS method. But this method is not the best method to estimate the parameter.

Theroetical Framewor

Ideally, we would like to able to detect such multicollinearity and outliers effects easily and more importantly provide an accurate estimate of these the effect more than the conventional 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 weights the indicators such that a resulting latent variables estimate can be obtained.

We conform to an underlying in this scenario, the latent variables would be the dependent variable (Yi) would be 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 improvement method known as Robust Partial Least Square Cobb-Douglas Production Function (RPLS-CDF) was illustrated by applying it to the information data from the important input for product output (wheat and barley) Libya Agriculture sector. In our theoretical study we have been used the additive following model

$$\eta = \delta_0 + \delta_1\xi_1 + \delta_2 \xi_2 + \delta_3 \xi_3 + e . \quad ... \quad (3)$$

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 suggested model is tested using data from Libyan agriculture sector (The main source of data for this study is based from the Agricultural Research Center, Tripoli (Libya)). The log transformation method was then compared with the new developed method. The improvement of RPLS-CDF is done by replacing the conventional variance-covariance matrix of the data with the corresponding robust variance-covariance matrix. The method of determining robust variance-covariance matrix is based on minimum covariance determinant (MCD) method, which is known to be robust in the presence of outliers. 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 with 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.

Conclusion

In this paper, we have discussed the problems that are often either exaggerated, or are uncritically glossed over, are that of multicollinearity and outliers. And this study seeks a way to possible application of the robust partial least squares (RPLS) methods in developing Cobb-Douglas production function, to solve the problems of multicollinerity and outliers in Cobb-Douglas production function. The improvement method suggested is RPLSR-CDF.

Reference

Abdi, H. (2003). Partial least squares regression (PLS regression), In M. Lewis-Beck, A. Bryman and T. Futing (eds): *Encyclopedia for research methods for the social sciences*. Thousand Oaks, CA: Sage.

Aguirregabiria, V. (2009). Econometric issues and methods in the estimation of production functions. *MPRA.* Department of Economics. University of Toronto.

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 robust methods for the detection of multiple outliers. *Journal of the American Statistical Association,* 89: 1329 - 1339.

Baltagi, B.H. (2008). *Econometrics*. Fourth Edition. New York: Springer.

Bhanumurthy, K.V. (2002). Arguing a case for the Cobb-Douglas production function. *Review of Commerce Studie,* 1(20 – 21).

Branden K. V., & Hubert, M. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics,* 17: 537 – 549.

Camminatiello, I. (2006). *Robust methods for Partial Least Squares Regression: methodological contributions and applications in environmental field.* FedOA. Universita Degli Studi Di Napoli Federico II.

Cobb, C.W., & Douglas, P.H. (1928). A Theory of Production. *American Economic Review* 18(1): 139 – 165.

Field, A. (2000). *Discovering statistics using SPSS for windows*. London: Sage

Kahane, L. H. (2001). Regression basics. Thousand Oaks, CA: Sage.

Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* 2(1): 49-55.

ISSN 1985-7918

McIntosh, A.R., Bookstein, F., Haxby, J. & Grady, C. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage,* 3: 143–157.

McIntosh, A.R. & Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: Applications and advances. *Neuroimage,* 23: 250-263.

Pavel, C. (2007). Efficient robust estimation of regression models. CentER Discussion Paper. Department of Econometrics & Operational Rsearch. Faculty of Economics and Business Administration, Tilburg University, Tilburg: Netherlands.

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*. Amsterdam: Academic Press.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In P. R. Krishnaiah, (eds.). *Multivariate analysis*. New York, NY: Academic Press.

Wold, H.(1982). Soft modeling-The basic design and some extensions. In K. Jöreskog & H. Wold, (eds.). *Systems under indirect observation II*. Amsterdam: North-Holland Press.

Zhang, Y. & 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.