Available online at http://scik.org
J. Math. Comput. Sci. 5 (2015), No. 6, 848-856
ISSN: 1927-5307

EVALUATION OF DETERMINANTS OF EMPLOYMENT EFFICIENCY USING STOCHASTIC FRONTIER ANALYSIS AND BETA REGRESSION

EMRE DÜNDER, SERPIL GÜMÜŞTEKIN* AND MEHMET ALI CENGIZ
Ondokuz Mayıs University Department of Statistics, 55139 Atakum SAMSUN TURKEY

Abstract. One of the main issues in today's world is the fact that the world economy is not creating enough jobs, particularly for young people. Unemployment is one of the biggest issues in this modern life. This study investigates the most important determinants of employment efficiency for countries. There are two steps in this study. First we measure countries' employment efficiency by Stochastic Frontier Analysis (SFA) using the data from the Database and cover the period between 2009 and 2012. SFA, which is a parametric method, is used frequently to estimate the boundary functions and to measure the production effectiveness. At second stage, we expose the multiple relationships between the employment efficiency and the other variables using Beta regression approach which is performed in which the dependent variable is restricted to the standard unit interval such as rates and proportions. The most important factors which affect the employment efficiency were determined as Industry Value Added and Net Taxes.

Keywords : unemployment; stochastic frontier analysis; beta regression approach.
2000 AMS Subject Classification: 92B05, 37C23, 37G15, 65L99, 70K50.

1. Introduction

According to International Labour Organization Report [5], more than 197 million people globally or 6% of the world's workforce were without a job in 2012. The fact that large numbers of people are unemployed has a negative effect on subsequent long-run economic growth. Unemployment can harm growth not only because it is a waste of resources, but also because it generates redistributive pressures and subsequent distortions, drives people to poverty, constrains liquidity limiting labour mobility, and erodes self-esteem promoting social dislocation, unrest and conflict [3].

In the last two decade, SFA, which was first introduced by [1] and [7], has been used intensively for determining the efficiency in many areas. SFA establishes a functional

*Corresponding author
Received June 8, 2015

relationship between output variables such as cost, profit and production and input variables such as explanatory variables and environmental factors. For panel data stochastic frontier model considers the estimated measurement error productivity and estimates a more reliable measure of productivity. The maximum likelihood method, which is the most popular estimation method, is used to derive parameter estimates of the SFA.

In many research areas, it is very common to encounter that the dependent variable takes values in the standard unit interval $(0,1)$ such as rates, proportions, percentages and fractions. There are some drawbacks when the linear regression model is applied to this kind of dependent variable. The major drawback is the fitted values for dependent variable that exceeds its lower and upper bounds. To overcome these drawbacks one way is to transform the dependent variable to values on the real line. However, this approach has also its own obstacles. The major obstacle is that the model parameters cannot be interpreted in terms of the original dependent variable. [4] introduced a beta regression model which is based on the assumption that the dependent variable is beta distributed. They used the modelling and inferential methods in the form of Generalized Linear Models (GLM) introduced by[6]. Since then, focus on beta regression models have been increased in many fields such as medicine, forest science, education, economics and political science. [11] modelled the mean and the precision parameters in beta regression using linear regression. [10] used linear or nonlinear regression to estimate the mean and the precision parameters in Beta regression. [8] and [2] proposed beta regression models allowing zero and/or one occurrences by incorporating degenerate distributions to model the extreme values. [9] used beta regression models allowing truncation in a subset of the unit interval.

First aim of this study is to estimate employment efficiency for 50 countries using Stochastic Frontier Analysis (SFA). The second aim is to select the factors affecting the employment efficiency using Beta regression analysis.

2. Stochastic Frontier Analysis

SFA approach proposed by [1] can be specified as follows:

$$y_i = X_i\beta + \varepsilon_i \quad , \qquad \varepsilon_i = v_i - u_i \quad , \quad u_i \geq 0 \tag{1}$$

Where y_i is log output, X_i is a vector of input measures, β is the vector of coefficients, v_i is independent and identically distributed error term and $u_i \geq 0$ is technical inefficiency. Error term $\varepsilon_i = v_i - u_i$ has a symmetric distribution.

3. Beta Regression Model

Let y is continuous variable that takes values in unit interval (0, 1). The variable is assumed to be beta distributed with the following parameterization:

$$f(y;\mu;\varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)}\, y^{\mu\varphi-1}(1-y)^{(1-\mu)\varphi-1} \qquad 0 < y < 1 \tag{2}$$

where $0 < \mu < 1$ and $\varphi > 0$. Regression model is obtained with the parameterization of the mean and precision as $E(y) = \mu$ and $Var(y) = \mu(1-\mu)/(1+\varphi)$

The mean response is related to linear predictors through a monotonic and twice differentiable link function such that,

$$g(\mu_i) = \sum_{k=1}^{p} x_{ik}\beta_k \tag{3}$$

where $\beta = (\beta_1,\beta_2,\ldots\ldots,\beta_p)$ represent the beta coefficients on regression parameters. $X = (x_1, x_2,\ldots\ldots\ldots,x_p)$ denotes the predictor variables in mean model. g(.) shows the link function. It is possible to choose link functions between several functional forms such as logit and probit.

Beta coefficients are estimated with the log-likelihood function of the model. Log-likelihood function of the model is

$$\iota_t(\mu_t, \varphi)=\log\Gamma(\varphi) - \log\Gamma(\mu_t\varphi) - \log\Gamma((1-\mu_t)\,\varphi) + (\mu_t\varphi-1)\log y_t+\{(1-\mu_t)\varphi-1\}\,\log(1-y_t) \tag{4}$$

We can estimate the beta coefficients based for mean model on scoring functions derived from the log-likelihood function with numerical optimization methods. Most common used methods are Fisher scoring or Newton Rapson algorithms for the maximum likelihood [12].

Dataset

Our sample covers 50 countries selected from four income groupings. Those income groups, which are obtained by using Gross National Income (GNI) per capita, in U.S. dollars, converted from local currency using the World Bank Atlas method, are low, lower-middle, upper middle and high. All variables used in this study are taken from The World Bank Database and cover the period between 2009 and 2013. The measurement of productive efficiency is based on the relationship between output produced and inputs required for production. In this paper we consider the number of working population as output. Population, Manufacturing and Gross Domestic Product (GDP) per capita were considered as inputs. At the second step, the efficiency scores obtained from SFA are used as dependent variable and Industry Value Added, Net Taxes, Inflation, Gross domestic savings and General Government final Consumption are taken as independent variables for Beta regression.

4. SFA Model

The proposed model is specified as follow:

$$Y_{it} = \beta_0 + \beta_1 X_{1_{it}} + \beta_2 X_{2_{it}} + \beta_3 X_{3_{it}} + \beta_4 t + v_{it} - u_{it} \tag{5}$$

$$u_i \overset{\text{i.i.d.}}{\sim} N^+(0, \lambda)$$

where for i th country in t th year, Y_{it} is the logarithm of the number of working population, $X_{1_{it}}$ is the logarithm of Population, $X_{2_{it}}$ is the logarithm of the manufacturing, $X_{3_{it}}$ is the logarithm of GDP and t is a time trend. v_{it} is a symmetric disturbance representing the effect of noise and u_{it} is a term of the inefficient employment.

Table 1. Maximum likelihood estimates of the parameters in stochastic frontier model

	Estimate	Standard Error	z-value	p
Constant	3,661	1,002	3,654	0
Log(Population)	1,162	0,073	15,846	0
Log(Manufacturing)	-0,51	0,108	-4,714	0
Log(GDP)	0,252	0,126	1,991	0,046
Sigma2	0,295	0,055	5,339	0
Gamma	0,972	0,004	240,048	0
Time	3,661	1,002	3,654	0

Table 1 shows the estimation results of the parameters for stochastic frontier model. Parameter estimate is statistically significant for all inputs. Gamma parameter is also significant and the value is one. It means that all the deviation from the frontier occurs because of the technical inefficient performance of countries. Technical efficiency scores for all countries are shown in Table 2. Minimum efficiency score is 0.009761 and maximum is 0,834397. Average efficiency of the countries is 0.1815.

Table 2. Technical efficiency scores for all countries

Level	Country	Efficiency
Lower-middle-income economies	Armenia	0,834397
Upper-middle-income economies	Ecuador	0,629504
High-income economies	Croatia	0,494221
High-income economies	Germany	0,473110
High-income economies	Korea, Rep.	0,466711
High-income economies	Denmark	0,437691
High-income economies	Singapore	0,423306
High-income economies	Austria	0,391380
High-income economies	Sweden	0,349860
High-income economies	Netherlands	0,336582
Upper-middle-income economies	Dominican Republic	0,311694
High-income economies	Finland	0,309729
High-income economies	United Kingdom	0,300666
High-income economies	Italy	0,299286
High-income economies	Norway	0,278649
High-income economies	France	0,272671
Upper-middle-income economies	Belarus	0,266284
High-income economies	Spain	0,266030
High-income economies	Portugal	0,244206
Upper-middle-income economies	Kazakhstan	0,228625
High-income economies	Estonia	0,194494
Upper-middle-income economies	Turkey	0,191958
Lower-middle-income economies	Ukraine	0,190923
High-income economies	Czech Republic	0,181567
High-income economies	Greece	0,180218
High-income economies	Luxembourg	0,172458
Lower-middle-income economies	El Salvador	0,168823
Upper-middle-income economies	Costa Rica	0,167140
Upper-middle-income economies	Mauritius	0,163675
Lower-middle-income economies	Guatemala	0,159883
Lower-middle-income economies	Honduras	0,130301
Upper-middle-income economies	South Africa	0,129923
Lower-middle-income economies	Georgia	0,124733
High-income economies	Australia	0,109461
Upper-middle-income economies	Macedonia, FYR	0,099965
Upper-middle-income economies	Jordan	0,097980
Upper-middle-income economies	Namibia	0,097425
Upper-middle-income economies	Albania	0,097085
Low-income economies	Nepal	0,085915
Lower-middle-income economies	Mongolia	0,083589
Low-income economies	Tajikistan	0,082402
Lower-middle-income economies	Zambia	0,082140
Lower-middle-income economies	Moldova	0,078456
Low-income economies	Uganda	0,070915
Upper-middle-income economies	Brazil	0,065411
Low-income economies	Afghanistan	0,055055
High-income economies	Belgium	0,020167
Upper-middle-income economies	Colombia	0,019937
Lower-middle-income economies	Egypt, Arab Rep.	0,013105
Low-income economies	Bangladesh	0,009761

The average efficiencies for income groups are also presented in Table 3. High-income economies have maximum average efficiency score and Low-income economies have the minimum average efficiency score.

Table 3. The average efficiencies for income groups

Level	Average efficiency
High-income economies	0,295
Upper-middle-income economies	0,187
Lower-middle-income economies	0,183
Low-income economies	0,061

In order to select the factors affecting employment efficiency, Beta regression analysis is employed in this study. Beta regression approach is based on the assumption that the mean is related to a set of explanatory variables through a linear predictor with unknown coefficients and a link function. For Beta regression, technical efficiency scores given in Table 2 are taken as dependent variable and Industry Value Added, Net Taxes, Inflation, Gross domestic savings and General Government final Consumption are taken as independent variables.
In Beta regression, the selection of an appropriate link function can greatly improve the model Fit. Therefore the different link functions are employed to select the best link in terms of the AIC and the BIC criteria. The AIC and BIC scores for different link functions in Beta regression are demonstrated in Table4.

Table 4. The AIC and BIC scores for different link functions

Link function	**AIC**	**BIC**
Logit	-218,212	-178,632
Probit	-218,361	-178,781
C log-log	**-218,931**	**-179,351**
Cauchy	-218,517	-178,937
Log	-217,659	-178,080
Log-log	-217,987	-178,408

We select the C log-log link as an appropriate link function since it gives bigger the AIC and BIC scores than the other link functions. For comparison, Tobit and Truncated regression model are also applied to our dataset. The AIC and BIC scores for different regression models are shown in Table 5.

Table 5. The AIC and BIC scores for different regression models

	AIC	BIC
Tobit	-165.184	-131.291
Truncated regression	-192.259	-158.583
Beta regression	**-218,931**	**-179,351**

We prefer Beta regression with the C log-log since it gives bigger The AIC and BIC scores than the other two regression models. Table 6 gives the coefficients for the mean model with C log-log link.

Table 6. Coefficients for the mean model with the C log-log link

	Estimate	Standard Error	z-value	p
Constant	-1,250	0.1613	-7,749	0
Industry Value Added	0.003771	0.001437	2,625	0.009
Net Taxes	-0.001784	0.0004	-3,983	0.005
Inflation	-0.6325	1,074	-0.589	0.556
Gross domestic savings	0.5974	0.5930	1,007	0.314
General Government final Consumption	0.004083	0.002193	1,862	0.063

In Beta regression, since the parameter accounting for the precision of the data is not assumed to be constant across observations but it is allowed to vary, the variable dispersion beta regression model is constructed. The coefficients for the precision model are demonstrated in Table 7.

Table 7. Coefficients for the precision model

	Estimate	Standard Error	z-value	p
Constant	1,491	0.1732	8,611	0
Industry Value Added	-0.004962	0.001782	-2,785	0.0053
Net Taxes	0.00238	0.000557	4,15	0.00332
Inflation	2,010	1,599	1,257	0.2089
Gross domestic savings	0.7395	0.6737	1,098	0.2723
General Government final Consumption	-0.005341	0.002671	-2,00	0.04552

According to Table 6 while Inflation, Gross domestic savings and General Government final Consumption were found to be insignificant parameters, Industry Value Added and Net Taxes were found as statistically significant parameters.

Conclusions

In this study we used a stochastic frontier model to determine the employment performance of 50 countries. We proposed a stochastic frontier model with errors in variables. Results from Stochastic Frontier analysis shows that Population, Manufacturing, Gross Domestic Product (GDP) and Time were significant. According to SFA, technical efficiency scores of the countries were estimated and according to these scores, rankings of efficiency were obtained. Based on technical efficiency while the most effective country was Armenia, the lowest effective is Bangladesh. The average efficiencies for income groups were obtained. The average efficiencies are very low even for high-income economies.

Beta regression approach was employed to select the factors affecting the employment efficiency for 50 countries. Industry Value Added and Net Taxes were found as statistically significant parameters while Inflation, Gross domestic savings and General Government final Consumption were found to be insignificant.

Although the employment efficiency for high-income economies is relatively greater than the other class economies, this study showed that the employment efficiency is very low even for high-income economies. Decision-makers must take precautions to improve the efficiency of employment. This study suggests that especially the policies on the industry Value Added and Net Taxes should be reconsidered.

Conflict of Interests

The authors declare that there is no conflict of interests.

REFERENCES

[1] Aigner, D., Lovell, C. K., and Schmidt, P. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1977), 21-37.

[2] Calabrese, R. *Regression model for proportions with probability masses at zero and one*. Vita e pensiero. (2012).

[3] Castells Quintana, D., & Royuela Mora, V. Unemployment and long-run economic growth: The role of income inequality and urbanization. *Investigaciones Regionales*, *24*(2012), 153-173.

[4] Cribari-Neto, F., & Souza, T. C. Testing inference in variable dispersion beta regressions. *Journal of Statistical Computation and Simulation*, 82(2012), 1827-1843.

[5] Global employment trends Recovering from a second jobs dip /International Labour Office. Geneva: ILO. (2013).

[6] McCullagh, P., Nelder, J. A., & McCullagh, P. *Generalized linear models*, 2. London: Chapman and Hall. (1989).

[7] Meeusen, W., & Van den Broeck, J. Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, (1977) 435-444.

[8] Ospina, R., & Ferrari, S. L. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(2012), 1609-1623.

[9] Pereira, G. H., Botter, D. A., & Sandoval, M. C. A regression model for special proportions. *Statistical Modelling*, 13(2013), 125-151.

[10] Simas, A. B., Barreto-Souza, W., & Rocha, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, 54(2010), 348-366.

[11] Smithson, M., & Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(2006), 54.

[12] Zhao, W., Zhang, R., Lv, Y., & Liu, J.. Variable selection for varying dispersion beta regression model. *Journal of Applied Statistics*, 41(2014), 95-100.