# Maximum entropy estimators to assess technical efficiency with state-contingent production frontiers

Macedo, Pedro
*Center for Research and Development in Mathematics and Applications*
*Department of Mathematics, University of Aveiro*
*Campus Universitário de Santiago*
*Aveiro, 3810-193, Portugal*
*E-mail: pmacedo@ua.pt*

Silva, Elvira
*Faculty of Economics, University of Porto*
*Rua Dr. Roberto Frias*
*Porto, 4200-464, Portugal*
*E-mail: esilva@fep.up.pt*

Scotto, Manuel
*Center for Research and Development in Mathematics and Applications*
*Department of Mathematics, University of Aveiro*
*Campus Universitário de Santiago*
*Aveiro, 3810-193, Portugal*
*E-mail: mscotto@ua.pt*

In the last decade, the work of Chambers and Quiggin (2000) inspired a remarkable research in the production economics literature. Although the theory of state-contingent production is nowadays well-established, the empirical implementation of this approach is still in an infancy stage. Among others, there are two important difficulties: (a) the real number of states of nature may be very large (creating ill-posed models); and (b) with the increasing number of states, it is very likely to find few observations for some states of nature as well as collinearity problems. Some authors claim the urgent need to develop robust estimation techniques to overcome these difficulties.

Golan et al. (1996) developed the Generalized Cross Entropy (GCE) and the Generalized Maximum Entropy (GME) estimators, which are widely used in linear and nonlinear regression models, in particular in models with small size samples, non-normal errors and affected by collinearity, and in models where the number of parameters to be estimated exceeds the number of observations available. Later, Golan and Perloff (2002) defined the GME-$\alpha$ estimators by replacing the Shannon entropy measure with Tsallis and Rényi entropies in the objective function of the GME estimator. So far there have been only few attempts to estimate state-contingent production functions using the GME estimator. However, looking at the advantages of these estimators it becomes clear that they could be the solution for some problems in the estimation of technical efficiency with state-contingent production frontiers, with the following additional important advantage: the traditional parametric assumptions on the error distributions, specifically the error inefficiency component, are not needed.

In this work, we illustrate the performance of these estimators when compared with the Maximum Likelihood (ML) estimator through several simulation studies (including models affected by collinearity and ill-posed models). Small mean squared error loss and small differences between the true and the estimated mean of technical efficiency reveal that the GCE, GME and GME-$\alpha$ estimators perform better than the ML estimator in most of the cases analyzed.

**State-contingent production frontier and maximum entropy estimators**

The state-contingent production frontier model used in this work is defined by

$$\ln q = \sum_{s=1}^{S}\sum_{p=1}^{P} d_s b_{ps}^{-1} \ln x_{ps} - \sum_{s=1}^{S}\sum_{p=1}^{P} d_s b_{ps}^{-1} \ln a_s - \sum_{s=1}^{S}\sum_{p=1}^{P} b_{ps}^{-1} \sum_{k=1}^{K} \alpha_k z_k + v - u, \tag{1}$$

with $0 < b_{ps}^{-1} \leq 1$, $a_s > 0$ and $0 < \sum_p b_{ps}^{-1} \leq P$, for all $s$ $(s = 1, 2, \ldots, S)$. The observed output is denoted by $q$, $S$ is the number of states of nature, $P$ is the number of state-allocable inputs, $K$ is the number of exogenous variables, $d_s$ is a dummy variable that assumes the value 1 when the state $s$ is chosen and 0 otherwise, $b_{ps}$ are the parameters that account for the possibility of output substitution between states, $x_{ps}$ are the state-allocable inputs, $a_s$ are ex post realizations of an unobservable random variable under Nature's control, $z_k$ are the exogenous variables, $\alpha_k$ are parameters to be estimated, $v$ is a random variable representing statistical noise and $u \geq 0$ is a one-sided random variable representing technical inefficiency; see Nauges et al. (2009) and O'Donnell et al. (2010) for further details concerning this state-contingent production frontier model.

Considering the state-contingent production frontier model in (1) defined by the matricial form

$$\ln \boldsymbol{q} = f(\boldsymbol{X}; \boldsymbol{\beta}) + \boldsymbol{v} - \boldsymbol{u}, \tag{2}$$

the reparameterization of the $(R \times 1)$ vector $\boldsymbol{\beta}$ and the $(N \times 1)$ vector $\boldsymbol{v}$ in the GME-$\alpha$ and GME estimators follows the same procedure as in the traditional regression model. Each parameter is treated as a discrete random variable with a compact support and $1 < M < \infty$ possible outcomes and each error is defined as a finite and discrete random variable with $1 < J < \infty$ possible outcomes. The reparameterization is given by

$$\boldsymbol{\beta} = \boldsymbol{Z}\boldsymbol{p} = \begin{bmatrix} \boldsymbol{z}_1' & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{z}_2' & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{z}_R' \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \vdots \\ \boldsymbol{p}_R \end{bmatrix}, \tag{3}$$

with $\boldsymbol{Z}$ a $(R \times RM)$ matrix of support points and $\boldsymbol{p}$ a $(RM \times 1)$ vector of unknown probabilities, and

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{w} = \begin{bmatrix} \boldsymbol{a}_1' & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{a}_2' & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{a}_N' \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_N \end{bmatrix}, \tag{4}$$

with $\boldsymbol{A}$ a $(N \times NJ)$ matrix of support points and $\boldsymbol{w}$ a $(NJ \times 1)$ vector of unknown probabilities.

This traditional approach can be extended to the vector $\boldsymbol{u}$. The reparameterization is similar to the one conducted for $\boldsymbol{v}$ taking only into account that $\boldsymbol{u}$ is a one-sided random variable which implies that the lower bound for the supports (with $1 < L < \infty$ points) is zero for all error values (the full efficiency case). The reparameterization of $\boldsymbol{u}$ can be defined by $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{\rho}$, with $\boldsymbol{B}$ a $(N \times NL)$ matrix of support points and $\boldsymbol{\rho}$ a $(NL \times 1)$ vector of unknown probabilities.

The maximum entropy principle is applied to estimate the unknown $\boldsymbol{p}$, $\boldsymbol{w}$ and $\boldsymbol{\rho}$ vectors that maximize

$$H(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{\rho}) = H_{\alpha_1}^{e}(\boldsymbol{\beta}) + H_{\alpha_2}^{e}(\boldsymbol{v}) + H_{\alpha_2}^{e}(\boldsymbol{u}), \tag{5}$$

subject to the model constraint and the additivity constraints

$$\begin{aligned} \ln \boldsymbol{q} &= \boldsymbol{X}\boldsymbol{Z}\boldsymbol{p} + \boldsymbol{A}\boldsymbol{w} - \boldsymbol{B}\boldsymbol{\rho}, \\ \boldsymbol{1}_R &= (\boldsymbol{I}_R \otimes \boldsymbol{1}_M')\boldsymbol{p}, \\ \boldsymbol{1}_N &= (\boldsymbol{I}_N \otimes \boldsymbol{1}_J')\boldsymbol{w}, \\ \boldsymbol{1}_N &= (\boldsymbol{I}_N \otimes \boldsymbol{1}_L')\boldsymbol{\rho}, \end{aligned} \tag{6}$$

where $\otimes$ represents the Kronecker product, $H^e_{\alpha_1}(\cdot)$ and $H^e_{\alpha_2}(\cdot)$ are entropy measures (Shannon, Rényi or Tsallis), and $\alpha_1$, $\alpha_2$ are the order of the entropy measure applied when $e$ represents the Rényi or Tsallis entropies.

The support matrices $\boldsymbol{Z}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined by the researcher based on prior information. Since the vector $\boldsymbol{v}$ is a two-sided random variable representing statistical noise, the supports in the matrix $\boldsymbol{A}$ can be defined symmetrically and centered around zero, using the three-sigma rule with the empirical standard deviation of the noisy observations. It is important to note that the traditional distributional assumptions concerning the error inefficiency component (Half Normal, Truncated Normal, Exponential and Gamma distributions, among others) have been used in several empirical studies, since it is expected a particular behavior in the distribution of technical inefficiency estimates (Kumbhakar and Lovell (2000)). These distributional assumptions are not necessary with the maximum entropy estimators, but the same beliefs can be expressed in the model through the error supports. In order to specify the matrix $\boldsymbol{B}$ we suggest a similar approach as the one developed and tested by Campbell et al. (2008) with Stochastic Frontier Analysis (SFA), which includes five points and assumes a negative skewness for the estimates of technical efficiency. Campbell et al. (2008) recognize that the choice of the supports is somewhat arbitrary and suggest the use of the mean efficiency of the Data Envelopment Analysis (DEA) and SFA estimates to define the supports for the error inefficiency component. To reduce the subjectivity in this support choice, we suggest a simpler approach in which the supports of matrix $\boldsymbol{B}$ can be defined as

$$\text{(7)} \qquad \boldsymbol{b}'_n = [0, 0.01, 0.02, 0.03, -\ln(\text{DEA}_n)]',$$

where $\text{DEA}_n$ represents the technical efficiency estimate provided by DEA for the production unit $n$ ($n = 1, 2, \ldots, N$). Since in the DEA method all deviations from the production frontier are due to inefficiency, this approach provides lower levels of efficiency than the SFA and the state-contingent approach, and can be used to define an upper bound for the supports. Specific supports can be defined for each production unit (the desirable approach) or, for simplicity, equal supports can be defined for all observations considering $\text{DEA}_n$ as the lower technical efficiency estimate obtained by DEA in the $N$ observations.

In the context of state-contingent production and using only the Shannon entropy measure, the GCE estimator can be defined as the minimization of

$$\text{(8)} \qquad H(\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{\rho}, \boldsymbol{q_3}) = \boldsymbol{p}' \ln \boldsymbol{p} + \boldsymbol{w}' \ln \boldsymbol{w} + \boldsymbol{\rho}' \ln(\boldsymbol{\rho}/\boldsymbol{q_3})$$

subject to conditions (6), where the vector $\boldsymbol{q_3}$ represents prior information about the unknown $\boldsymbol{\rho}$. With the GCE estimator, the supports in matrix $\boldsymbol{B}$ can follow a similar structure as in (7), although with equally spaced points in the range $(0, -\ln(\text{DEA}_n))$. We suggest that the set of subjective probability distribution might also take the form $\boldsymbol{q_3} = [0.40, 0.30, 0.15, 0.10, 0.05]'$ for each observation, where the cross entropy objective shrinks the posterior distribution in order to have more mass near zero.

**Simulation study**

In this work we are particularly interested with models in which the ML estimator should be avoided. Thus, particular attention is given to models with few observations in some states of nature and models with a large number of states of nature (which are usually models affected by collinearity), illustrating potential real problems when using the state-contingent production approach. Note that few observations per state restrict the use of traditional estimators and, even with a reasonable number of observations per state, the collinearity problem arises with an increasing number of states of nature.

The 2-norm condition number (i.e., the ratio of the largest singular value of $\boldsymbol{X}$ with the smallest singular value) is considered to evaluate the collinearity in the design matrix. Since the 2-norm condition number increases with the number of (state and non state-allocable) inputs and considering that the number of inputs lies between four and six in empirical studies, we define $P = 2$ and $K = 3$, representing a very likely empirical scenario. Depending also on the number of observations per state, we verify that models with a 2-norm condition number greater than $5 \times 10^3$ are very likely for $P \geq 3$ and $K \geq 4$. We consider $S = 2, 5, 10$ to define the number of states of nature in this study.

The matrix $\boldsymbol{X}$ and the parameter vector $\boldsymbol{\beta}$ defined in (2) are generated according to the state-contingent production frontier model in (1). The state-allocable inputs and the exogenous variables are generated randomly using the absolute values of Normal distributions with means between one and ten, and standard deviations equal to one. The parameters are generated using Uniform distributions, namely $b_{ps}^{-1}$ is generated using $U(10^{-3}, 1)$; for the constant parameter $U(-4, 4)$ is used; and the parameters of the exogenous variables are generated using $U(-2, 2)$. The random variable representing statistical noise, $\boldsymbol{v}$, is generated using a standard Normal distribution. The one-sided random variable representing technical inefficiency, $\boldsymbol{u}$, is generated considering two possible distributions: (i) the Half Normal distribution with zero mean and standard deviation $\sigma_u = 0.1$; (ii) the Exponential distribution with parameter $\lambda = 10$. In case (i), we generate the Normal-Half Normal model, which is a common model used in SFA. Note that the selected values for the parameters of the previous distributions are arbitrary. Nevertheless, we choose values that we consider reasonable in empirical studies.

The supports in matrix $\boldsymbol{Z}$ are defined in the interval $[0, 1]$ for the parameters of state-allocable inputs and $[-10, 10]$ for the other parameters. For all supports in $\boldsymbol{Z}$, we consider five points ($M = 5$). The support interval for the other parameters could be narrowed although it is defined in this way to illustrate real cases where prior information is scarce. The supports in the matrix $\boldsymbol{A}$ are defined symmetric and centered on zero, using the three-sigma rule with the empirical standard deviation of the noisy observations. The number of support points are three in each support ($J = 3$). The supports in the matrix $\boldsymbol{B}$ are defined using (7). In the Normal - Half Normal models, we consider that the lowest estimate of technical efficiency obtained using DEA in the samples is approximately 67%. Note that, using the Half Normal distribution defined in (i) to generate the vector $\boldsymbol{u}$, the probability of finding a value greater than 0.3 is less than 1%. Thus, by considering the lowest technical efficiency estimate obtained by DEA as 67%, we are defining the supports in matrix $\boldsymbol{B}$ by $\boldsymbol{b}_n' = [0, 0.01, 0.02, 0.03, 0.4]'$. In the Normal - Exponential models, the supports in the matrix $\boldsymbol{B}$ are defined by $\boldsymbol{b}_n' = [0, 0.01, 0.02, 0.03, 0.8]'$, considering that the lowest estimate of technical efficiency obtained by DEA is 45%. In all the simulations performed with the Normal - Exponential models, the maximum value generated in $\boldsymbol{u}$ is 0.62. Note that we are using conservative supports in both models (i.e., supports with an upper bound larger than necessary), illustrating that DEA provides lower levels of efficiency than the SFA and the state-contingent approach.

The GCE estimator is performed using the supports $[0, 0.4]$ and $[0, 0.8]$ respectively in the Normal - Half Normal models and the Normal - Exponential models, with five equally spaced points in both supports. We define two cases: the GCE1 with $\boldsymbol{q_3} = [0.40, 0.30, 0.15, 0.10, 0.05]'$ and the GCE2 with $\boldsymbol{q_3} = [0.50, 0.40, 0.05, 0.03, 0.02]'$, where the cross entropy objective shrinks the posterior distribution in order to have more mass near zero than in the first case.

The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are replicated in order to create 50 different versions of each model. To evaluate the performance of the estimators, we use two measures: the mean squared error loss (MSEL) and the diference between the true and the estimated mean of technical efficiency (DMTE).

## Some results

Under severe empirical conditions, a first conclusion is that the GCE, GME and GME-$\alpha$ estimators perform, in general, better than the ML estimator in terms of MSEL and DMTE. Even in models with only two states of nature, these estimators exhibit, in general, a better performance than the ML estimator, except in models with a small 2-norm condition number and with a reasonable number of observations in each state of nature (i.e., a model with ideal empirical conditions).

As the number of observations per state decreases and/or the number of states of nature increases, the 2-norm condition number increases substantially. In particular, the 2-norm condition number can increase exponentially when the number of observations per state approaches to one, illustrating the collinearity problem in the estimation of state-contingent production frontiers. As expected, the MSEL of the ML estimator increases when the 2-norm condition number increases. However, for the GCE, GME and GME-$\alpha$ estimators, the increase in MSEL is smaller and, in general, the MSEL increases only slightly.

Generally the maximum entropy estimators have a lower DMTE than the ML estimator in most of the models tested. Among the maximum entropy estimators, the GME estimator provides, in general, the highest DMTE and the GME-$\alpha$ estimators generates the lowest DMTE. The GME and the GME-$\alpha$ estimators provide higher values of DMTE in the Normal - Exponential models than in the Normal - Half Normal models. This result is due to the different supports in the matrix $\boldsymbol{B}$ used in these models, with different prior means: 91.2% for the Normal-Half Normal models and 84.2% for the Normal-Exponential models. Taking into account the parameters defined in the Half Normal and the Exponential distributions in this simulation study, the positive skewness value of the Exponential distribution is much higher than the positive skewness value of the Half Normal distribution. Thus with a lower prior mean and a higher degree of positive skewness, the GME and the GME-$\alpha$ estimators generate higher values of DMTE in the Normal - Exponential models.

In the Normal - Half Normal models the GCE1 estimator provides, in general, a smaller DMTE than the GCE2 estimator. However, in the Normal - Exponential models, the GCE2 estimator presents a smaller DMTE than the GCE1 estimator. This discrepancy is due to the different vectors $\boldsymbol{q_3}$ used in both estimators (the cross entropy objective in GCE2 shrinks the posterior distribution in order to have more mass near zero than in GCE1). Since the positive skewness value of the Exponential distribution is much higher than the positive skewness value of the Half Normal distribution in this simulation study, the GCE2 (GCE1) estimator has a smaller DMTE than the GCE1 (GCE2) estimator in the Normal - Exponential models (Normal - Half Normal models). However, it is important to note that even with no fully correct prior beliefs expressed in vector $\boldsymbol{q_3}$ (i.e., for the GCE1 estimator in the Normal - Exponential models and for the GCE2 estimator in the Normal - Half Normal models), the differences in DMTE between the GCE1 and GCE2 estimators are small in most of the cases, revealing the important feature of the GCE estimator that wrong prior beliefs do not have a significant impact on the estimates (Golan et al. (1996)).

**Concluding remarks**

The results achieved in this work indicate that the maximum entropy estimators are powerful alternatives to the ML estimator in the estimation of state-contingent production frontiers under severe empirical conditions. Some advantages of these estimators seem to be clear: (1) the possibility of considering easily prior information on the parameters and errors components; (2) the traditional assumptions on the errors' distributions are not necessary; and (3) these estimators can be used in models with a large number of states of nature and even with only one observation per state (which represent usually models affected by severe collinearity). Note that few observations per state restrict the use of traditional estimators and, even with a reasonable number of observations per state, the collinearity problem arises with an increasing number of states of nature. In these cases, researchers can reduce the number of states (increasing the number of observations per state, but losing important information) or use the maximum entropy estimators that are robust in the presence of collinearity and can be used even if there is only one observation per state.

## References

Campbell, R., Rogers, K. and Rezek, J. (2008). Efficient frontier estimation: a maximum entropy approach. *Journal of Productivity Analysis* 30:213–221.

Chambers, R. G. and Quiggin, J. (2000). *Uncertainty, Production, Choice, and Agency: The State-Contingent Approach.* Cambridge University Press, Cambridge.

Golan, A. and Perloff, J. M. (2002). Comparison of maximum entropy and higher-order entropy estimators. *Journal of Econometrics* 107:195–211.

Golan, A., Judge, G. and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* John Wiley & Sons, Chichester.

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis.* Cambridge University Press, Cambridge.

Nauges, C., O'Donnell, C. J. and Quiggin, J. (2009). Uncertainty and technical efficiency in Finnish agriculture. Australian Agricultural and Resource Economics Society, 53rd Annual Conference, Cairns, Australia.

O'Donnell, C. J., Chambers, R. G. and Quiggin, J. (2010). Efficiency analysis in the presence of uncertainty. *Journal of Productivity Analysis* 33:1–17.

## Acknowledgments

This work was supported by *FEDER* funds through *COMPETE*–Operational Programme Factors of Competitiveness ("Programa Operacional Factores de Competitividade") and by Portuguese funds through the *Center for Research and Development in Mathematics and Applications* (University of Aveiro) and the Portuguese Foundation for Science and Technology ("FCT–Fundação para a Ciência e a Tecnologia"), within project PEst-C/MAT/UI4106/2011 with COMPETE number FCOMP-01-0124-FEDER-022690.
Pedro Macedo is also supported by the grant SFRH/BD/40821/2007 from FCT.