Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP04/2018

Profit Efficiency, DEA, FDH and Big Data

Valentin Zelenyuk

Date: June 2018

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Profit Efficiency, DEA, FDH and Big Data

by Valentin Zelenyuk*

June 25, 2018

Abstract

The goal of this article is to outline a very simple way of estimating profit efficiency in the DEA and FDH frameworks, but avoiding the computational burden of linear programming. With this result it is possible to compute profit efficiency even when dimension of inputs and outputs are larger than the dimension of number of decision making units (firms, individuals, etc.), as is often the case in the 'big data'.

*School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Australia. The author acknowledges the financial support from ARC Discovery Grant(DP130101022). Email: v.zelenyuk@uq.edu.au

1 Theoretical Underpinnings

Profit of a business activity is considered to be the main goal of any for-profit organization. For this very reason, profit maximization criterion is also the corner stone of virtually any model in neoclassical economic theory. The profit-oriented framework in economics typically starts with a premise that the key benchmark is the (maximal) profit function, defined as

$$\pi(w, p|\Psi) = \sup_{x,y} \{py - wx \ : \ (x, y) \in \Psi\}, \tag{1.1}$$

where $x = (x_1, ..., x_N) \in \mathbb{R}^N_+$ and $y = (y_1, ..., y_M) \in \mathbb{R}^M_+$ are inputs and outputs, respectively, and $w = (w_1, ..., w_N) \in \mathbb{R}^N_+$ and $p = (p_1, ..., p_M) \in \mathbb{R}^M_+$ are their corresponding market prices, and Ψ is the relevant technology set characterizing business activities, defined in very general terms as

$$\Psi \equiv \{(x, y) \ : \ x \text{ can produce } y\}, \tag{1.2}$$

which is typically assumed to satisfy some standard regularity conditions of production theory (e.g., see Shephard (1953); Färe and Primont (1995)).

Many profit efficiency measures were introduced in the literature, e.g., Cooper et al. (2011); Aparicio et al. (2013, 2017), to mention a few (see Färe et al. (2018) for a review and references).

Recently, Färe et al. (2018) developed a general approach that unified many other approaches. Specifically, they proposed a general measure of *profit efficiency*, defined for an observed quantity vector (x^j, y^j) and price vector (w, p), as

$$\begin{aligned} \mathscr{E}(x^j, y^j; w, p|\Psi) = \sup_{\theta,\lambda,x,y} & \{f(\theta, \lambda) : \\ & \sum_{m=1}^{M} p_m \theta_m y^j_m - \sum_{i=1}^{N} w_i \lambda_i x^j_i \leq py - wx, \\ & (x, y) \in \Psi, \ \theta = (\theta_1, \ldots, \theta_M), \ \lambda = (\lambda_1, \ldots, \lambda_N)\} \end{aligned} \tag{1.3}$$

where $f(\theta, \lambda)$ is an objective function chosen by a researcher, to be optimized jointly over $\theta = (\theta_1, \ldots, \theta_M)$ and $\lambda = (\lambda_1, \ldots, \lambda_N)$ and (x, y).[1]

This general profit efficiency measure can be viewed as a dual analogue of a technical efficiency measure that satisfies Pareto-Koopmans efficiency criterion, but in addition to taking into account efficiency with respect to technology it also takes into account the quintessence from the input and output markets, i.e., the prices that reflect valuation of both buyers and

[1]Some constraints may be needed on Ψ (in addition to standard regularity conditions) to regularize the non-decreasing returns to scale cases, where profit maximization may yield profit to be ∞ or 0.

sellers on those markets. In this sense, this measure can be viewed as one that satisfies a superior criterion relative to the primal Pareto-Koopmans efficiency that only considers technical efficiency.

By selecting different forms of $f(\theta, \lambda)$ many useful efficiency measures can be obtained from (1.3), as detailed in Färe et al. (2018). A particularly interesting case of (1.3) they show is derived by setting $\lambda_1, \ldots, \lambda_N = 1$ and $\theta_1, \ldots, \theta_M = \theta$ so that $f(\lambda_1, \ldots, \lambda_N; \theta_1, \ldots, \theta_M) = \theta$, to obtain

$$\begin{aligned} \mathscr{E}_o(x^j, y^j; w, p|\Psi) = &\sup_{\theta,x,y} \{\theta : \\ &p\theta y^j - wx^j \leq py - wx, \, (x, y) \in \Psi\} \end{aligned} \tag{1.4}$$

which in turn can be stated in terms of the sup-sup (or "maxi-max"), as

$$\mathscr{E}_o(x, y; w, p|\Psi) = \sup_{\theta} \left\{ \sup_{x,y} \left\{ \frac{py - wx + wx^j}{py^j} \ : \ (x, y) \in \Psi \right\} \geq \theta \right\}, \tag{1.5}$$

While looking very complicated and 'too theoretical', Färe et al. (2018) also derive several very intuitive versions of this measure, that are composed of simple and intuitive notions often used in business analysis. Specifically, they showed that

$$\mathscr{E}_o(x^j, y^j; w, p|\Psi) = \frac{c^j}{r^j} + \frac{\pi(p, w|\Psi)}{r^j} \tag{1.6}$$

where $c^j = wx^j$, $r^j = py^j$ are the observed total costs and total revenue of the firm at the allocation (x^j, y^j) that face prices (w^j, p^j).

That is, intuitively, (1.6) says that the profit efficiency measure defined in (1.4) or (1.5) can be decomposed into two key performance indicators used in business analysis: (i) the realized cost-revenue ratio and the best possible profit margin for the firm with allocation (x^j, y^j) that faces prices (w, p). Note that the first component is the reciprocal of the "return to the dollar" measure of performance advocated by Georgescu-Roegen (1951).[2]

In practice, researcher does not observe Ψ and so cannot obtain the true value of $\pi(w, p|\Psi)$ and thus of $\mathscr{E}_o(x^j, y^j; w, p|\Psi)$, but the analytical developments in the next sub-section allow an easy estimation.

[2] Färe et al. (2018) also show that this profit efficiency measure can be further decomposed into three sources: (i) revenue efficiency, (ii) Farrell technical efficiency (output oriented here) and (iii) a new allocative efficiency measure measuring the gap between profit maximization and revenue maximization.

2 DEA and FDH Estimation

Let $x = (x_1^k, ..., x_N^k) \in \mathbb{R}_+^N$ and $y^k = (y_1^k, ..., y_M^k) \in \mathbb{R}_+^M$ be observations on inputs and outputs for a decision making unit $k \in \{1, ..., n\}$, then it is well-known that the DEA formulation for *the maximal profit function* is given by

$$\begin{aligned} \widehat{\pi}(w, p|DEA - VRS) \;\; &\equiv \max_{\substack{x_1,...,x_N,\; y_1,...,y_M \\ z^1,...,z^n}} \sum_{m=1}^{M} p_m y_m - \sum_{l=1}^{N} w_l x_l, \\ \text{s.t.} \\ &\sum_{k=1}^{n} z^k y_m^k \;\; \geq y_m^j,\; m = 1, ..., M, \\ &\sum_{k=1}^{n} z^k x_l^k \;\; \leq x_l,\; l = 1, ..., N, \\ &\sum_{k=1}^{n} z^k = 1, \\ &z^k \geq 0,\; k = 1, ..., n. \end{aligned} \tag{2.1}$$

Because the optimization is done over all x and all y, with strictly positive prices for inputs and outputs, there must exists an optimum where all the inequalities turn to equalities (i.e., no slacks), reaching Pareto-Koopmans efficiency. Thus, one can multiply each m^{th} output equality constraint by the corresponding output price p_m, $(m = 1, \ldots, M)$ and sum these inequality constraints over m; similarly, one can multiply each i^{th} input equality constraint (2.1) by the corresponding input price w_i, $(i = 1, \ldots, N)$, while keeping the other constraints the same, to get the following optimization problem

$$\begin{aligned} \widehat{\pi}(w, p|DEA - VRS) \;\; &\equiv \max_{\substack{x_1,...,x_N,\; y_1,...,y_M \\ z^1,...,z^n}} \sum_{m=1}^{M} p_m y_m - \sum_{l=1}^{N} w_l x_l, \\ \text{s.t.} \\ &\sum_{k=1}^{n} z^k \sum_{m=1}^{M} p_m y_m^k \;\; = \sum_{m=1}^{M} p_m y_m^j,\; m = 1, ..., M, \\ &\sum_{k=1}^{n} z^k \sum_{l=1}^{N} w_l x_l^k \;\; = \sum_{l=1}^{N} w_l x_l,\; l = 1, ..., N, \\ &\sum_{k=1}^{n} z^k = 1, \\ &z^k \geq 0,\; k = 1, ..., n. \end{aligned}$$

which is equivalent to the following problem

$$\begin{aligned}
\widehat{\pi}(w,p|DEA-VRS) &= \max_{z^1,\ldots,z^n} \sum_{k=1}^{n} z^k r^k - \sum_{k=1}^{n} z^k c^k = \max_{z^1,\ldots,z^n} \sum_{k=1}^{n} z^k \pi^k, \\
\text{s.t.} & \\
& \sum_{k=1}^{n} z^k = 1 \\
& z^k \geq 0, \ k = 1, ..., n.
\end{aligned}$$

which in turn implies that

$$\widehat{\pi}(w,p|DEA-VRS) = \max\{\pi^1, ..., \pi^n\}. \tag{2.2}$$

That is, one can estimate $\widehat{\pi}(w^j, p^j)$ without information on (w, p) and even without information on (x^j, y^j), but simply having the aggregated information about costs and revenue, (c^j, r^j), for each j to compute the observed profit for each j, rank the computed profits across all j and select the highest value, which will be $\widehat{\pi}(w^j, p^j)$ for the given sample. To the best of our knowledge, this appears to be a new result, although very simple, yet very useful for practice. This result is especially useful for cases when dimension of x and y are very large, e.g., including 'big data' cases, because the DEA estimator is known to be not immune from the so-called 'curse of dimensionality' problem.

Also note that the same result holds for the case of DEA with decreasing returns to scale, i.e., when $\sum_{k=1}^{n} z^k = 1$ is replaced with $\sum_{k=1}^{n} z^k \leq 1$. Similar result can also be derived for the DEA with increasing returns to scale (i.e., $\sum_{k=1}^{n} z^k = 1$ is replaced with $\sum_{k=1}^{n} z^k \geq 1$) and with constant returns to scale (i.e., when $\sum_{k=1}^{n} z^k = 1$ is removed from the formulation) when additional constraints are imposed (e.g., maximal bounds on inputs) that will prevent the objective function going to infinity. Moreover, this results also holds for the Free Disposal Hull (FDH) estimator, proposed by Deprins et al. (1984), because it can be represented as the DEA-VRS problem where $z^k \geq 0$ is replaced with $z^k \in \{0, 1\}$, and so by similar logic as above we arrive to the following formulation

$$
\begin{aligned}
\widehat{\pi}(w,p|FDH) &= \max_{z^1,\dots,z^n} \sum_{k=1}^{n} z^k r^k - \sum_{k=1}^{n} z^k c^k = \max_{z^1,\dots,z^n} \sum_{k=1}^{n} z^k \pi^k, \\
\text{s.t.} & \\
& \sum_{k=1}^{n} z^k = 1 \\
& z^k \in \{0,1\}, \ k = 1,...,n.
\end{aligned}
$$

which also implies that

$$\widehat{\pi}(w,p|FDH) = \max\{\pi^1,...,\pi^n\}. \tag{2.3}$$

3 Concluding Remarks

While looking fairly simple (after seen the proof), this result opens the door for many applications that would have been infeasible due to the absence of information on all the outputs or when the dimensionality of the output space is too large in comparison with the available sample size, including the 'big data' cases.

With this result, a researcher can substitute all the unobserved inputs and outputs (however 'big' their dimensions are) by the observed total cost and total revenue, compute profits, and then obtain the estimate of *overall* output efficiency. The resulting estimates, in fact, might even have more valuable information about efficiency from an economic point of view, since they incorporate such economically important information as output prices, the corresponding allocation of inputs and the underlying behavior of the decision making units.

For example, for the context of the general framework of measuring efficiency developed by Färe et al. (2018) briefly described above, if one wants to use the DEA framework for estimating Ψ, then the estimated output oriented Farrell-type profit efficiency measure can be easily obtained as

$$\hat{\mathscr{E}}_o(x^j, y^j; w, p|\widehat{\Psi}) = \frac{c^j}{r^j} + \frac{\widehat{\pi}(w,p)}{r^j}, \tag{3.1}$$

where $\widehat{\pi}(w,p)$ is the estimated profit function from (2.2), obtained without actual computation of DEA or FDH model, and possible even if some of x, y, p, w are not observed or their dimension is too large for DEA and FDH to handle, as long as total revenue and total costs are observed for the firms of interest. In a similar fashion, this result may be also adapted for estimating other profit efficiency measures (e.g., see Cooper et al. (2011); Aparicio et al. (2013, 2017)).

References

Aparicio, J., Ortiz, L., and Pastor, J. T. (2017). Measuring and decomposing profit inefficiency through the slacks-based measure. *European Journal of Operational Research*, 260(2):650 – 654.

Aparicio, J., Pastor, J. T., and Ray, S. C. (2013). An overall measure of technical inefficiency at the firm and at the industry level: The lost profit on outlay. *European Journal of Operational Research*, 226(1):154 – 162.

Cooper, W. W., Pastor, J. T., Aparicio, J., and Borras, F. (2011). Decomposing profit inefficiency in DEA through the weighted additive model. *European Journal of Operational Research*, 212(2):411–416.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labour efficiency in post offices. In Marchand, M., Pestieau, P., and Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurement*, pages 243–267. Amsterdam, NL: Springer.

Färe, R., He, X., Li, S. K., and Zelenyuk, V. (2018). A unifying framework for farrell profit efficiency measurement.

Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications*. New York, NY: Kluwer Academic Publishers.

Georgescu-Roegen, N. (1951). The aggregate linear production function and its applications to von Neumann's economic model. In Koopmans, T., editor, *Activity Analysis of Production and Allocation*. New York, NY: Wiley.

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton, NJ: Princeton University Press.