Kallio, Markku; Halme, Merja

Individual Efficient Frontiers in Performance Analysis

Published in:
Advances in Efficiency and Productivity Analysis

Published: 01/01/2021

Document Version
Peer reviewed version

Please cite the original version:
Kallio, M., & Halme, M. (2021). Individual Efficient Frontiers in Performance Analysis. In C. Parmeter, & R. Sickles (Eds.), *Advances in Efficiency and Productivity Analysis* (pp. 195-209). (Springer Proceedings in Business and Economics). SPRINGER.

Individual Efficient Frontiers in Performance Analysis e-Appendix

Markku Kallio[1] *and Merja Halme*[2]

[1,2] Department of Information and Service Management
Aalto University School of Business
POB 21210, FI-00076 Aalto, Finland
[1]E-mail: markku.kallio@aalto.fi, tel. +358403538087
[2]Corresponding author, E-mail: merja.halme@aalto.fi, tel. +358-40 353 8081

June 14, 2019

A Evolutionary optimization for price computation

For solving (2) we use an implementation of the evolutionary optimization algorithm PCX-$G3$ (Deb et al., 2002). Our implementation deals with admissible price vectors only. Therefore the fitness function of $\mu \in P$ is the objective function value in (2).

Before stating the algorithmic steps, we introduce some further notation. Recall that the set P of admissible prices μ is defined by linear equations and linear inequalities. Let P_0 denote the null space of these linear equations. Then $P \subset \mu^0 + P_0$ for some price vector μ^0. Let columns of matrix Δ be (non-normalized) orthogonal coordinates of P_0. For $\bar{w} \geq 0$ define in P_0 a rectangular box $B(\bar{w}) = \{\mu; \mu = \Delta w, 0 \leq w \leq \bar{w}\}$. Then there exist μ^0 and $\bar{w}$ such that $P \subset \mu^0 + B(\bar{w})$; i.e., P is in the box $B(\bar{w})$ translated by μ^0. Consider the smallest such box. Then after rescaling of the columns of Δ, for the smallest box $\bar{w} = e$ with $e^t = (1, \ldots, 1)^t$. Considering such scaled matrix Δ and the box $\mu^0 + B(e)$ containing P the steps of our PCX-$G3$ implementation are as follows:

1. **Initialization**. Initiate with a random population $M = \{\mu^s\}$ from uniform distribution over P. Here we generate uniformly distributed points μ in the box $\mu^0 + B(e)$ which are accepted in M if $\mu \in P$. Population size is 200.[1]

2. **Begin iteration**. Using tournaments[2], select three candidates from M. These are denoted by μ^r, $r = 1, 2, 3$, in ranking order with $r = 1$ referring to the best one.

3. **Create offsprings**. Denote $d^1 = \mu^1 - (\mu^1 + \mu^2 + \mu^3)/3$ and $d^2 = \mu^2 - \mu^3$ and scale both so that the norm is one. Generate two offsprings μ^a and μ^b as follows. Get a direction $d = \alpha d^1 + (1 - \alpha) d^2 + \xi$ where weight α is drawn from uniform distribution $U(0, 1)$ and the mutation increment is $\xi = \omega \times \Delta w$ where $\omega = 0.07$ and the components of vector w are drawn from normal distribution $N(0, 1)$. An offspring (μ^a or μ^b) is then given by $\mu^1 + \nu d$, where the step size ν is randomly chosen using a uniform distribution between 0 and the maximum feasible step.

4. **Update population**. Select randomly μ^c and μ^d from M, and replace them in M by two of the best candidates among $\{\mu^a, \mu^b, \mu^c, \mu^d\}$.

[1] An alternative approach is to start with each candidate μ in the initial population M to find local maxima and pick the best one found. Then with population size 1000 we obtain equally good results as with PCX-$G3$; however, the number of function evaluations needed by solver $MINOS$ is three-fold.

[2] In a tournament two points are randomly drawn from M and the best is selected.

5. **End iteration**. If the fitness difference of the best and worst element in M is less than $\tau = 10^{-2}$ or no improvement is achieved in 2000 iterations or an iterations limit 10000 is met then proceed to step 6, otherwise return to step 2.

6. **Closing**. Starting with the best candidate $\hat{\mu}$ in M, find a local maximum using $AMPL/MINOS$. Replace $\hat{\mu}$ by the local optimum, if it provides an improvement.

Given the small dimension of problem (2), for parameter choices in PCX-$G3$, the population size of 200, iterations limit 10000 and stopping if no improvement 2000 iterations can be regarded as safe. However, for mutation parameter ω and fitness tolerance τ we carry out sensitivity tests. For these tests we use from Section 4the bank branches case only and consider two alternatives of κ (profit or return) and admissible prices P (constrained or un-constrained). For density estimates $\hat{\psi}$ we study the log-normal case and kernel density estimate employed in Section 4

Because the heuristics in PCX-$G3$ does not necessarily guarantee a global optimum, the best solution found may depend on the initialization of the seed for random number generators. Among tests based on ten alternative initializations, the standard deviation of the objective function value relative to the average value (over two alternatives for κ) is 1.8 % for log-normal cases and 0.01 % for kernel density estimate; as expected, smooth objective functions in the latter case yield small variation.

To test the sensitivity of tolerance τ we decrease its value from the initial level 10^{-2} to 0. Thereby the objective function values may increase. For both density estimation cases we run a single test with two alternatives for κ. The average increase in the objective function value was 1.3 % for log-normal and 0.004 % for the kernel case.

For the mutation parameter ω we test the values 0.01, 0.04, 0.07 and 0.10. Again for the two density estimation cases we run one test with the two alternative measures κ, and determine the relative difference in the objective function in comparison with the best found using the four values of ω. In the log-normal case, on the average such percentage differences were 1.2, 1.2, 2.0 and 1.9 %, for $\omega = 0.01, \ldots, 0.10$, respectively. In the kernel density case, such average figures were 0.01, 0.02, 0.01 and 0.01 %. Given the variation due to alternative choices of seeds of random number generators, we conclude that choice of the mutation parameter ω in the range 0.01-0.10 does not have a major impact in the performance.

B Simulation examples

To illustrate PA, we experiment with simulations. For data generation, the netput vector z^j for DMU^j, $j = 1, 2, \ldots, n$, with input vector x^j and output vector y^j, are drawn randomly and independently from a multivariate log-normal distribution with expected value a and covariance matrix C for the logarithms of inputs and outputs.

For value and return performance, we assume a convex $CET\text{-}GD$ transformation function $g^j(\xi, \eta)$ for each DMU^j so that the convex PPS^j is defined by input vectors ξ and output vectors η satisfying $g^j(\xi, \eta) \leq 0$; see Kumbhakar et al. (2015). Hence g^j for DMU^j is given by

$$g^j(\xi, \eta) = [\sum_k \gamma_{jk} \eta_k^{c_j}]^{1/c_j} - \alpha_j \Pi_i \xi_i^{\beta_{ji}}$$

for some non-negative parameters α_j, β_{ji}, $\sum_i \beta_{ji} \leq 1$, γ_{jk}, $\sum_k \gamma_{jk} = 1$ and $c_j \geq 1$. Given a price vector μ, optimality conditions for a profit maximizing $(\xi, \eta) = (x^j, y^j)$ imply $\mu_y y^j / \mu_x x^j \geq 1$. For all DMUs, we assume that a safe upper bound for the profit margin is 75 % which implies an upper bound 4 for the return. Additionally, we require non-negativity $\mu \geq 0$ and scaling $\mu_x \bar{x} = 1$, where $\bar{x}$ is the average of input vectors x^j in the sample. Hence, the set of admissible prices for PA is

$$P = \{\ \mu \mid \mu \geq 0,\ \mu_x \bar{x} = 1,\ 1 \leq \mu_y y^j / \mu_x x^j \leq 4\ \forall\ j\ \}$$

and the estimated equilibrium price vector $\hat{\mu}$ is in P. Note that for any $\mu \in P$, there are parameters for the transformation functions as well as for supply and demand functions such that $\hat{\mu} = \mu$ is an equilibrium price. Therefore, we assume that the unknown equilibrium price vector μ^* is in P.

We let $n = 20$ and consider three cases: (a) one input and one output, (b) one input and two outputs and (c) two inputs and two outputs. In each case we generate 20 independent samples of netput vectors $\{z^j\}$ setting all expected values in a equal to 2 and using the following covariance matrices C

$$C_a = \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{bmatrix} \quad C_b = \begin{bmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & -0.1 \\ 0.2 & -0.1 & 0.5 \end{bmatrix} \quad C_c = \begin{bmatrix} 0.5 & -0.1 & 0.1 & 0.1 \\ -0.1 & 0.5 & -0.1 & 0.2 \\ 0.1 & -0.1 & 0.5 & -0.1 \\ 0.1 & 0.2 & -0.1 & 0.5 \end{bmatrix}. \qquad (1)$$

We accept a sample $\{z^j\}$ only if the set P of admissible prices is nonempty. For each eligible sample we draw randomly 20 equilibrium price vectors μ^* from P to be used for comparison with the estimated price vector $\hat{\mu}$ which is obtained by likelihood maximization

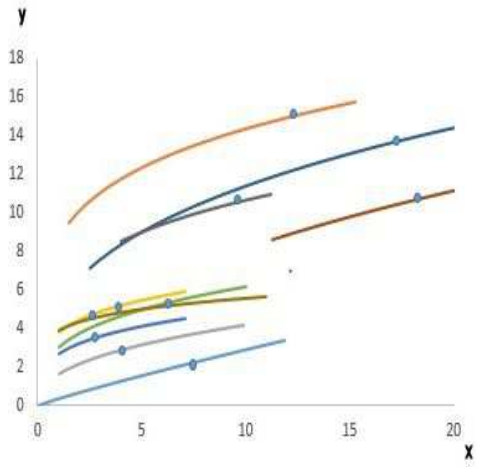

Figure 1: Inputs (x^j) and outputs (y^j) of ten DMUs j in an equilibrium with price vector $\hat{\mu} = (1, 3.79)$. Part of the efficient frontier is shown in the neigborhood of equilibrium point (x^j, y^j) for each DMU.

(2). We use approximation (8) for the pdf $\hat{\psi}(\kappa; \mu)$ and assume that the random netput vector $\tilde{z}$ has a multivariate log-normal distribution whose parameters are estimated using the sample $\{z^j\}$. For the comparisons, we report the average (in the 20×20 cases of) Spearman rank correlation for rankings based on $\hat{\mu}$ and μ^*. Both return and profit based comparisons are presented.

The case (i) of one input and one output serves merely for graphical illustration; Figure 1 illustrates a sample of netput vectors with $n = 10$ obtained with $a = (2, 2)$ and $C = C_i$ given in (1). Given price estimate $\hat{\mu}_x = 1$ and $\hat{\mu}_y = 3.79$ in VPA, transformation functions $g^j(\xi, \eta) = \eta - \alpha_j \xi^{\beta_j}$ are revealed with $\beta_j = \hat{\mu}_x x^j / \hat{\mu}_y y^j$ and $\alpha_j = y^j (x^j)^{-\beta_j}$. Parts of the efficient frontiers satisfying $g^j(x, y) = 0$ are shown in Figure 1 together with the equilibrium points (x^j, y^j). For $n = 20$ DMUs, for 20 randomly chosen samples $\{z^j\}$ of netput vectors and for each sample 20 randomly chosen equilibrium prices μ^*, the average of Spearman rank correlation of the ranking based on VPA and on equilibrium prices is 99.5 % (standard deviation 0.4 % points). Because there is only one input and one output, such average is naturally 100 % for RPA. In case (b) of one input and two outputs, average rank correlation was 95.3 % (3.3 %) for profit based and 98.1 % (1.6 %) for return based analysis. In case (c) of two inputs and two outputs, average rank correlation is 87.5 % (7.4 %) for profit based and 92.0 % (5.3 %) for return based analysis. Thus, rank correlation is high. Of course, due to scarce data perfect ranking replication cannot expected.

C Bank branches' case: data and results

Table 1: Input and output data for 25 bank branches B_j of Helsinki OP Bank. The figure of inputs $I_1, \ldots, I_5$ are total work times in full-time years per annum in five sales force categories. Both outputs, financing services J_1 and investment services J_2, are given in average number of aggregated transactions per annum.

	inputs					outputs	
DMU	I_1	I_2	I_3	I_4	I_5	J_1	J_2
B_1	5.29	2.08	0.57	7.16	1.31	1090	497
B_2	9.48	6.68	1.09	11.61	1.75	2633	1111
B_3	14.11	8.02	1.33	13.76	1.90	3320	1477
B_4	5.29	1.95	0.60	5.59	1.51	1147	353
B_5	3.09	3.29	0.46	4.17	1.41	1180	540
B_6	16.15	10.09	1.64	14.61	3.41	3821	1769
B_7	7.73	4.04	1.06	8.04	1.58	1574	716
B_8	4.38	2.61	0.63	9.00	1.71	1171	1004
B_9	4.31	2.58	0.67	6.48	0.78	1174	449
B_{10}	5.06	2.56	0.78	6.92	1.42	1203	568
B_{11}	3.70	3.06	0.66	5.71	0.90	928	384
B_{12}	14.81	9.37	1.92	12.48	2.28	4393	2210
B_{13}	7.68	3.65	1.13	11.36	1.51	2642	931
B_{14}	12.71	5.61	1.55	12.65	1.68	3362	1505
B_{15}	7.61	2.65	0.71	1.55	1.62	2263	541
B_{16}	16.36	5.53	1.75	18.16	2.95	3619	1541
B_{17}	15.85	10.94	1.78	12.81	3.09	4163	1594
B_{18}	11.72	5.24	1.15	9.31	1.69	3075	805
B_{19}	18.82	11.68	1.92	23.57	3.67	5757	2601
B_{20}	5.35	3.06	0.58	6.78	1.69	1763	496
B_{21}	19.20	8.47	1.71	19.54	3.47	3825	1961
B_{22}	11.00	4.79	0.87	8.09	1.59	2354	792
B_{23}	21.21	14.69	3.04	23.13	3.65	5289	3160
B_{24}	4.20	2.52	0.60	5.76	1.44	1108	332
B_{25}	3.98	2.39	0.57	5.46	1.36	743	354

Table 2: Bank branches' return performance RP (%) and ratio measure of return efficiency RE by RPA (with log-Normal and kernel density), and ratio measure of return efficiency REA by traditional DEA. In RPA, the ratio measure of efficiency RE of B^j is $\rho_j^* / \max_i \rho_i^*$ where ρ_j is the return of bank B_j.

DMU	log-Normal		Kernel		REA
	RP	RE	RP	RE	
B_1	13.5	0.617	25.9	0.696	0.808
B_2	70.9	0.803	56.2	0.798	0.901
B_3	46.9	0.726	42.9	0.752	0.984
B_4	9.0	0.598	21.3	0.679	0.758
B_5	90.4	0.912	84.0	0.945	1.000
B_6	36.6	0.694	39.8	0.742	0.881
B_7	12.0	0.609	12.3	0.638	0.679
B_8	75.3	0.827	87.1	0.963	1.000
B_9	68.2	0.791	45.1	0.759	0.833
B_{10}	32.4	0.687	37.4	0.734	0.766
B_{11}	31.3	0.683	12.3	0.638	0.738
B_{12}	91.1	0.915	80.8	0.927	1.000
B_{13}	97.4	1.000	92.7	1.000	1.000
B_{14}	78.8	0.839	71.3	0.872	1.000
B_{15}	77.7	0.835	87.0	0.962	1.000
B_{16}	34.0	0.689	46.6	0.764	0.982
B_{17}	52.9	0.744	44.4	0.757	0.858
B_{18}	59.6	0.763	54.2	0.790	1.000
B_{19}	89.4	0.903	83.0	0.939	1.000
B_{20}	81.4	0.851	79.3	0.918	1.000
B_{21}	16.7	0.629	28.8	0.707	0.874
B_{22}	20.5	0.644	26.6	0.699	0.925
B_{23}	65.2	0.780	52.5	0.784	0.959
B_{24}	27.9	0.673	29.2	0.708	0.756
B_{25}	1.2	0.513	3.4	0.561	0.582

Table 3: Bank branches' value performance VP (%) and difference measures of value efficiency VE by VPA (with Normal, log-Normal and kernel density) and difference measure of value efficiency VEA by traditional DEA. In VPA, the difference measure of efficiency VE of B^j is $\max_i \pi_i^* - \pi_j^*$ where π_j is the profit of bank B_j.

DMU	log-Normal		Kernel		VEA
	VP	VE	VP	VE	
B_1	49.6	0.237	46.6	0.398	0.073
B_2	37.1	0.274	20.2	0.469	0.091
B_3	45.6	0.248	38.5	0.419	0.017
B_4	40.8	0.263	53.1	0.381	0.086
B_5	74.9	0.150	81.5	0.272	0.000
B_6	30.4	0.298	40.4	0.414	0.167
B_7	13.2	0.379	27.7	0.447	0.243
B_8	85.6	0.082	83.3	0.259	0.000
B_9	34.5	0.281	38.2	0.419	0.069
B_{10}	36.0	0.279	50.4	0.388	0.121
B_{11}	18.9	0.349	20.5	0.468	0.101
B_{12}	91.2	0.018	95.7	0.000	0.000
B_{13}	70.2	0.174	56.7	0.371	0.000
B_{14}	77.2	0.140	87.0	0.223	0.000
B_{15}	90.7	0.020	95.5	0.005	0.000
B_{16}	31.2	0.292	17.8	0.477	0.018
B_{17}	14.8	0.370	30.2	0.440	0.219
B_{18}	43.7	0.255	53.5	0.380	0.000
B_{19}	92.2	0.000	63.1	0.353	0.000
B_{20}	70.2	0.174	59.5	0.363	0.000
B_{21}	34.5	0.280	17.3	0.479	0.195
B_{22}	52.3	0.231	65.0	0.347	0.058
B_{23}	10.0	0.406	46.6	0.398	0.083
B_{24}	30.4	0.299	30.5	0.439	0.107
B_{25}	23.8	0.325	29.6	0.442	0.176

Table 4: Bank branches' ranking based on VPA and RPA with log-Normal and Kernel density, and ranking based on traditional VEA and REA.

DMU	Rank by value			Rank by return		
	log-N	Kern	VEA	log-N	Kern	REA
B_1	10	13	14	22	21	19
B_2	14	23	17	9	9	14
B_3	11	16	10	14	15	10
B_4	13	11	16	24	22	21
B_5	6	5	1	3	4	1
B_6	19	15	21	15	16	15
B_7	24	21	25	23	24	24
B_8	4	4	1	8	2	1
B_9	17	17	13	10	13	18
B_{10}	15	12	20	17	17	20
B_{11}	22	22	18	18	23	23
B_{12}	2	1	1	2	6	1
B_{13}	7	9	1	1	1	1
B_{14}	5	3	1	6	8	1
B_{15}	3	2	1	7	3	1
B_{16}	18	24	11	16	12	11
B_{17}	23	19	24	13	14	17
B_{18}	12	10	1	12	10	1
B_{19}	1	7	1	4	5	1
B_{20}	8	8	1	5	7	1
B_{21}	16	25	23	21	19	16
B_{22}	9	6	12	20	20	13
B_{23}	25	14	15	11	11	12
B_{24}	20	18	19	19	18	22
B_{25}	21	20	22	25	25	25