Quantile DEA: A Direct Linear Programming Based Approach to Obtaining Quantile Efficiency or Quantile Group Benchmarking Performance Estimates

Joseph Atwood
309E Linfield Hall
Montana State University
Bozeman, MT 59717 USA
Tel: 1-406-994-5614
Fax: 1-406-994-4838
jatwood@montana.edu

Saleem Shaik
Dept. of Agribusiness and Applied Economics
North Dakota State University
Fargo, ND USA

Selected Poster prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Quantile DEA: A Direct Linear Programming Based Approach to Obtaining Quantile Efficiency or Quantile Group Benchmarking Performance Estimates

Joseph Atwood, 208 Linfield Hall. Dept. of Agricultural Economics and Economics MSU, Bozeman, MT 59717-2920 Phone: (406) 994 5614 E-mail: jatwood@montana.edu
Saleem Shaik, 504 Richard H Barry Hall. Dept. of Agribusiness and Applied Economics NDSU, Fargo, ND 58108-6050 Phone: (701) 231 7459 E-mail: saleem.shaik@ndsu.edu

INTRODUCTION

Data envelope analysis (DEA) was introduced by Charnes, Cooper and Rhode (1978) in operational research and popularized in a more informative and easily applied way by Fare et al. (1994).

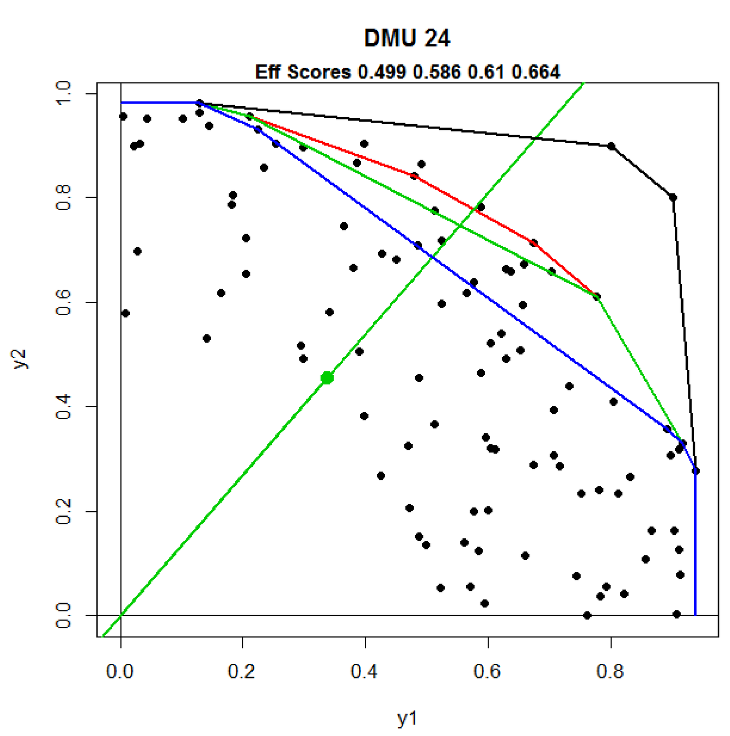
Since its introduction, the number of applications of DEA to firms, industry, state and countries have exploded. Along with the applications, researchers are beginning to identify and address issues associated with the empirical application of DEA including observational outliers and robust estimators.

In this research, a new robust procedure with good statistical properties is developed using a linear lower partial moment stochastic inequality. The procedure weights deviations from constraints in such a way that guarantees that no more than q x 100 percent of points lie external to hull.

The new procedure is "Quantile DEA (qDEA)" and simultaneously identify the constraints to relax in a second stage conventional DEA model. Similar but less efficient procedures date as far back as Timmer's sequentially deleting constraining observations (JPE-1971).

MOTIVATION FOR QDEA

The qDEA procedure utilizes partial moment (Fishburn-1977) stochastic inequalities presented by Berck-Hihn (1982) and Atwood (1985). Atwood's linear PM inequality can be utilized to implement chance constraints within a conventional LP model.



DEA

$$\max_{p,w} \; y^j p$$
$$st. \quad x^j w \le 1$$
$$Y p - X w \le 0$$
$$p, w \ge 0$$

qDEA Stage I

$$\max_{p,w,T,d,\theta} \; y^j p$$
$$st. \quad x^j w \le 1$$
$$Y p - X w + \underline{1} T - I d^+ \le 0$$
$$\left(\tfrac{1}{N}\right)' d^+ - \theta = 0$$
$$T - \left(\tfrac{1}{q}\right)\theta \ge 0$$
$$p, w, T, d, \theta \ge 0 \text{ and continuous}$$

OBJECTIVES

- **Quantile-DEA (qDEA) model to obtain more robust outlier resistant efficiency estimates**
- **Enhance the usefulness of DEA in quantile-based bench marking performance metric comparisons.**
- **Examine the statistical properties of the qDEA efficiency estimates.**

METHODS

Lower Partial Moment Inequality: Atwood (1985) generalized Berck-Hihn (1982)

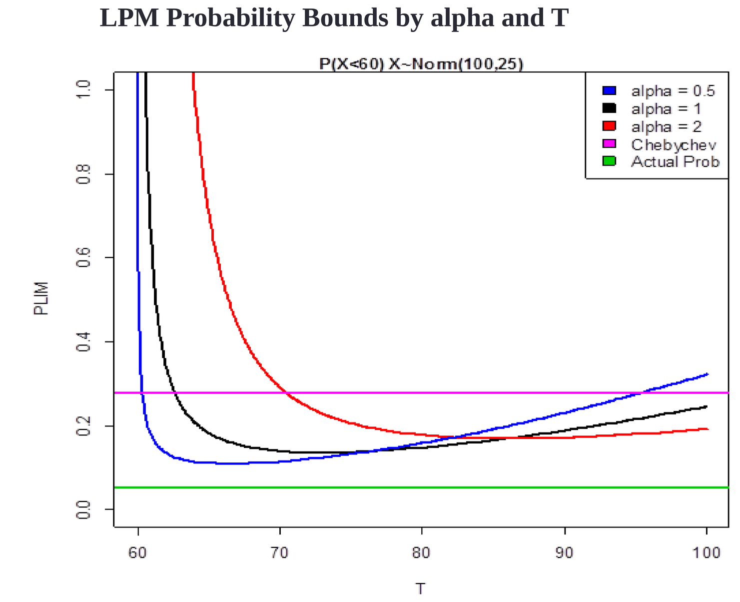
$$\rho(\alpha,T)=\int_{-\infty}^{T}(T-x)^{\alpha}dF(x)=\int_{-\infty}^{g}(T-x)^{\alpha}dF(x)+\int_{g}^{T}(T-x)^{\alpha}dF(x)\ge\int_{-\infty}^{g}(T-x)^{\alpha}dF(x)\ge\int_{-\infty}^{g}(T-g)^{\alpha}dF(x)$$

$$=(T-g)^{\alpha}F(g)\Rightarrow\rho(\alpha,T)\ge(T-g)^{\alpha}F(g)\Rightarrow F(g)\le\frac{\rho(\alpha,T)}{(T-g)^{\alpha}}\;\forall\,T>g,\;\alpha>0.$$

$$\text{Defining } \theta(\alpha,T)=\left[\rho(\alpha,T)\right]^{1/\alpha}\ge 0\Rightarrow F(g)\le\left[\frac{\theta(\alpha,T)}{(T-g)}\right]^{\alpha}\;\forall\,\alpha>0,\,T>g$$

The resulting probability limit will usually be less conservative that Chebychev's one-sided probability limit for an appropriate choice of alpha and T. When alpha=1, we can allow an LP model to simultaneously select T, compute the LPM, and use the above inequality in the first stage of the LP Model to identify which points are allowed to be external to the DEA hull.

LPM Probability Bounds by alpha and T

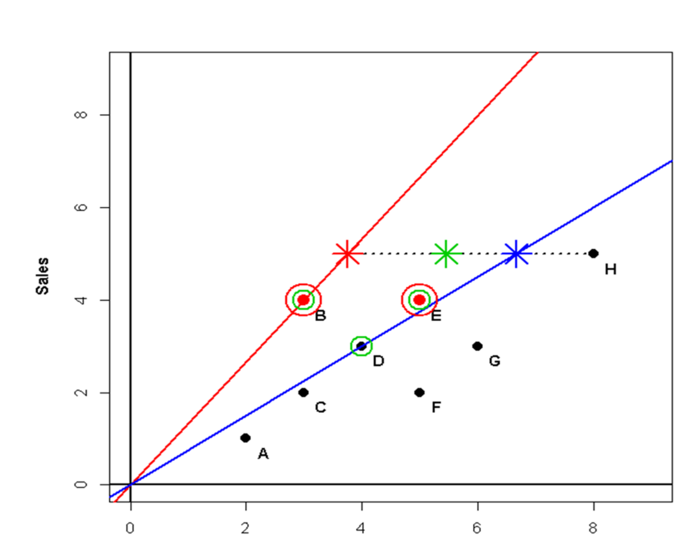


qDEA Stage 1 – Using Linear LPM Stochastic Inequality

$$\max_{p,w,T,d,\theta} \; y^j p$$
$$st. \quad x^j w = 1$$
$$Y p - X w + \underline{1} T - I d \le 0$$
$$\left(\tfrac{1}{N}\right)' d - \theta = 0$$
$$T - \left(\tfrac{1}{q}\right)\theta \ge 0$$
$$p, w, T, d, \theta \ge 0 \text{ and continuous}$$

LP duality theory shows that the qDEA Stage 1 efficiency scores are projected from both the desired support points and the external points. Stage 2 in qDEA relaxes the DEA LP constraints for the stage 1 external points. The qDEA stage 2 problem generates the desired projections and quantile efficiency distances and efficiencies. The following table presents estimated efficiency scores for a slightly modified problem from Cooper, Seiford, and Tone (2007). The following table presents the estimated efficiency scores with 0, 1, and 2 external points.

Modified CST Example for DMU H and Two External Points



Estimated Efficiency Scores for Modified CST Example with 0, 1, and 2 External Points

DMU	EFF0	EFF-Q1	EFF-Q2
A	0.3750	0.6250	0.6667
B	1.0000	1.6667	1.7778
C	0.5000	0.8333	0.8889
D	0.5625	0.9375	1.0000
E	0.6000	1.0000	1.0667
F	0.3000	0.5000	0.5333
G	0.3750	0.6250	0.6667
H	0.4688	0.7813	0.8333

STATISTICAL PROPERTIES

Difficult to obtain closed form expressions for statistical properties of DEA and qDEA estimators.

Subsample bootstrapping has proven useful in estimating confidence intervals for DEA and qDEA estimates.

nCm SUBSAMPLE BOOTSTRAPPING

Under a reasonable set of assumptions, subsample or nCm bootstrapping can be used to estimate the significance, bias corrections, and confidence intervals of parameter estimates. (Politis et. al. (1999, 2001) ; Simar and Wilson (2011); Geyer (2015). With the nCm bootstrap B subsamples (X_m, Y_m) of size m << n are sampled from the original input-output sample (X_n, Y_n) and used to construct a simulated set of parameter realizations: $\tilde{\theta}_{m,b}=\hat{\theta}_n-\left(\frac{m}{n}\right)^{\beta}\left(\hat{\theta}_{m,b}-\hat{\theta}_n\right)$ for $b=1,\cdots,B$ .The mean or median of the B resulting values can be used as a bias corrected parameter estimate. The B values $\tilde{\theta}_{m,b}$ are also bias corrected and their quantile values can be used to construct confidence intervals. A complication arises in that there is no known way to select the appropriate level for m and common practice is to estimate a set of $\tilde{\theta}_{m,b}$ values for multiple m values and use suggestions by Politis et. al.(2001) and Simar and Wilson to select an appropriate level for m. Computation time was reduced substantially by using only five m values in the set mlist where

mlist = floor(exp(seq(log(sqrt(n),log(5sqrt(n),length.out = 5).

For example, when n = 1000, the set of m values = (31,47,70,105, 158).

A large number of Monte Carlo simulations were run for various sample sizes and data generating processes. Empirical coverage levels were then computed as the proportion of times that the estimated confidence interval contained the population parameter. 95% qDEA coverage levels similar to those obtained by Simar and Wilson for the DEA model were obtained.

			COVERAGE LEVELS					
DGP			INPUT ORIENTATION COVERAGE		OUTPUT ORIENTATION COVERAGE		INOUT ORIENTATION COVERAGE	
nDMU	# inputs	BETA	NO REPLACE	REPLACE	NO REPLACE	REPLACE	NO REPLACE	REPLACE
100	1	0.500	0.845	0.876	0.844	0.869	0.833	0.845
100	2	0.500	0.888	0.911	0.880	0.897	0.853	0.869
100	3	0.400	0.937	0.941	0.915	0.929	0.902	0.904
100	5	0.286	0.935	0.932	0.917	0.916	0.914	0.924
100	6	0.250	0.964	0.954	0.944	0.935	0.914	0.913
		BETA	NO REPLACE	REPLACE	NO REPLACE	REPLACE	NO REPLACE	REPLACE
250	1	0.500	0.822	0.834	0.805	0.824	0.778	0.821
250	2	0.500	0.903	0.909	0.856	0.877	0.824	0.816
250	3	0.400	0.935	0.936	0.920	0.932	0.890	0.900
250	5	0.286	0.916	0.911	0.918	0.915	0.880	0.906
250	6	0.250	0.923	0.901	0.916	0.914	0.904	0.917
		BETA	NO REPLACE	REPLACE	NO REPLACE	REPLACE	NO REPLACE	REPLACE
500	1	0.500	0.845	0.864	0.845	0.864	0.811	0.835
500	2	0.500	0.898	0.905	0.887	0.897	0.853	0.858
500	3	0.400	0.940	0.939	0.941	0.947	0.927	0.932
500	5	0.286	0.939	0.938	0.943	0.936	0.939	0.937
500	6	0.250	0.938	0.935	0.925	0.923	0.935	0.938
		BETA	NO REPLACE	REPLACE	NO REPLACE	REPLACE	NO REPLACE	REPLACE
1000	1	0.500	0.8200	0.8400	0.829	0.844	0.772	0.786
1000	2	0.500	0.8740	0.8980	0.858	0.878	0.832	0.845
1000	3	0.400	0.9550	0.9600	0.958	0.964	0.948	0.954
1000	5	0.286	0.9375	0.9390	0.938	0.939	0.942	0.943
1000	6	0.250	0.9480	0.9470	0.941	0.939	0.952	0.950

CONCLUSIONS

- **Possible to estimate quantile benchmark DEA and DDEA efficiency metrics while allowing up to q x 100 percent of points to lie external to the efficiency hull.**
- **Procedures are useful in the presence of outliers and observational noise.**
- **Good success with nCm subsample bootstrapping of confidence intervals.**

References

Cooper, W. W., Seiford, L. M., Tone, K., 2007. "Data Envelopment Analysis". Springer. New York.
Geyer. C. J. "The Subsampling Bootstrap." http://www.stat.umn.edu/geyer/5601/notes/sub.pdf
Politis, D.N., Romano, J.P., Wolf, M., 1999. "Subsampling". Springer. New York.
Politis, D.N., Romano, J.P., Wolf, M., 2001. "On the asymptotic theory of subsampling." Statistica Sinica 11, 1105-1124.
Simar, L., Wilson, P.W., 2011. "Inference by the m Out of n Bootstrap in Nonparametric Frontier Models." Journal of Productivity Analysis 36,33-53.