CENTRE FOR HEALTH
PROGRAM EVALUATION

WORKING PAPER 133

On Generating Confidence Intervals for Nonparametric Methods

Dr John Wildman
Research Fellow, Health Economics Unit, Monash University

Dr Bruce Hollingsworth
Senior Research Fellow, Health Economics Unit, Monash University

May, 2002
ISSN 1325 0663
ISBN 1 876662 53 0

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding of these disciplines within the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434 **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au
Or by downloading from our website
Web Address http://chpe.buseco.monash.edu.au

ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

Table of Contents

List of Figures

Abstract

Nonparametric techniques provide no analytical solutions for confidence intervals. The bootstrap and jackknife methods are applied to Data Envelopment Analysis to generate different confidence intervals, which affect the probability of committing a type I error.

Keywords

Nonparametric methods, bootstrap, jackknife, efficiency measurement.
JEL classification: C14; C12

On Generating Confidence Intervals for Nonparametric Methods

Introduction

The use of nonparametric analysis has increased rapidly, one important method is Data Envelopment Analysis (DEA). In nonparametric applications analytical methods of deriving standard errors for estimates may not exist, preventing researchers from investigating the accuracy of an estimate. Even if asymptotic results exist, in finite samples the performance of the asymptotic results may be poor, eg. in partially linear models (Jones, 2000). Computer intensive resampling methods have been developed to investigate the accuracy of estimators that are too complex for traditional statistical analysis. This paper provides a note on these methods using DEA analysis on OECD data.

This paper uses different bootstrap and jackknife methods to compare the performance of these approaches in determining the accuracy of estimates. Different methods produce different confidence intervals, altering the chance of making a type I error (rejecting the null hypothesis when it is true). We use resampling methods to produce confidence intervals around the efficiency scores for each country. We also use these methods to generate confidence intervals around the mean efficiency score. The standard error for the mean efficiency score can be easily calculated using standard statistical techniques but it illustrates the case for statistics, such as the median, for which no statistical method can be applied. The use of the mean allows us to demonstrate the bias that can be introduced by using different bootstrap methods.

The Estimation Procedure

A full description of DEA can be found in (Hollingsworth et al., 1999 and Cooper et al., 2000), only a brief description is given here. DEA applies a linear programme to solve one of the following problems: to maximise weighted output with weighted input equal to unity or minimise weighted input with weighted output equal to unity.
Thus, for a country with *n* outputs and *m* inputs the ratio is:

$$Efficiency = \frac{\sum_{r=1}^{n} u_r * y_r}{\sum_{i=1}^{m} v_i * x_i} \qquad (1)$$

where: y_r = quantity of output *r*; u_r = weight attached to output *r*; x_i = quantity of input *i*; v_i = weight attached to input *i;* and a value equal to unity implies complete efficiency. The weights are specific to each unit so that: $0 \prec Efficiency \leq 1$.

We use data from OECD countries in 1997 to measure the efficiency of health production within countries given health care expenditure and education levels. Conceptually countries wish to maximise health output given their resource inputs, so we use the output maximisation model. The health output is measured in DALEs and the health inputs are health care expenditure and literacy rates.

The constant returns to scale model estimated produces efficiency scores for the 30 countries, Turkey is the most efficient, scoring 100, and the US is the least efficient, scoring 42.53. The distributions of these results are unknown, preventing the analytical derivation of confidence intervals. The mean level of efficiency from the sample is 56.96. We use this statistic in order to compare bootstrap and jackknife methods with statistical methods, demonstrating that for other summary statistics, for which no analytical solution exits, care must be taken.

The Confidence Interval

We generate confidence intervals around two statistics; the estimate of the efficiency score for each country, *è*, and the mean level of efficiency of the whole sample, $\bar{q}$.

Confidence intervals around efficiency scores

Two methods are used to generate confidence intervals around the efficiency score for each country, *è*, the jackknife and the bootstrap.

The jackknife, which predates the bootstrap but is similar in method. To generate confidence intervals around the efficiency estimates we reestimate the DEA models omitting one observation from each model. Let:

$$q_{(ij)} = s(x_{(ij)}) \,, \tag{2}$$

be the *i*th jackknife replication for country *j* of the estimated efficiency score, *è*. For our sample size, this involves running thirty DEA models omitting one different country in each model. The estimates are then used to generate standard errors (and confidence intervals) for the efficiency estimate from the full model. For country *n* we have 29 estimates of efficiency, by taking the standard error of these 29 estimates we derive an estimate of the standard error. Adding and subtracting twice the standard error from the original estimate gives a 95% confidence interval around the estimate (Efron and Tibshirani, 1993).

The bootstrap (Efron, 1979) provides an alternative method of producing confidence intervals. The bootstrap is implemented by sampling with replacement from the data to produce a new data set each time, from this new data new estimates are calculated. The procedure is replicated a number of times, the larger the number of replications the more accurate the statistic, although between 50 and 200 replications are generally sufficient (Efron and Tibshirani, 1993).

To generate the confidence interval for the efficiency score for each country we use 100 replications.

$$q_{kj}(b) = s(X_{kj}(b)) \tag{3}$$

Where $è_{kj}(b)$ is the efficiency score for the *j*th country from the *k*th replication. Taking the standard error from these efficiency scores for each country allows us to create confidence intervals around the original efficiency estimate.

Confidence intervals around the mean efficiency score

To generate confidence intervals around the mean efficiency score of the whole sample, $\bar{q}$ we use four methods. With data intensive nonparametric methods it is tempting to bootstrap the estimated scores as in (i) but this disregards the estimation procedure. We then bootstrap (ii) and jackknife (iii) equation 1, reestimating the model with each replication. We calculate the standard error from the estimated efficiency scores from equation (1) (iv), such techniques are not applicable to other summary statistics, eg the median. In the limit the standard error estimated from (ii) will equal the result from (iv).

(i) Bootstrap the efficiency scores estimated from model (1) and take the standard error. The average of $\bar{q}$ from the resampled efficiency scores is:

$$\bar{q}_k^* = \frac{1}{R}\sum \bar{q}_k \quad (4)$$

where $\bar{q}_k = \frac{1}{N}\sum q_{jk}$ and R is the number of replications. The standard error is given by:

$$se_{\bar{q}_k^*} = \{\frac{1}{R-1}\sum[\bar{q}_k - \bar{q}_k^*]^2\}^{1/2} \quad (5)$$

(ii) Bootstrap the whole of equation (1) to produce new average efficiency scores from each replication. These new efficiency scores are denoted as *è*$_{kj}$*(b).* The average of these bootstrapped scores is denoted as:

$$\bar{q}_k(b) = \frac{1}{N} q_{kj}(b)\ . \quad (6)$$

The standard error is calculated as:

$$se_{\bar{q}_k(b)} = \{\frac{1}{R-1}\sum[\bar{q}_k(b) - \bar{q}_k^*(b)]^2\}^{1/2} \quad (7)$$

where $\bar{q}_k^*(b) = \frac{1}{R}\sum \bar{q}_k(b)$.

(iii) Jackknife the whole of equation (1) to produce new average efficiency scores from each replication and calculate standard errors.

$$\bar{q}_i = \frac{1}{N}\sum q_{ij}\ . \quad (8)$$

The standard error is calculated as:

$$se_{\bar{q}_i} = \{\frac{1}{N}\sum [\bar{q}_i - \bar{q}_i^*]^2\}^{1/2} \quad (9)$$

where $\bar{q}_i^* = \frac{1}{R}\sum \bar{q}_i$.

(iv) Calculate the standard error from the efficiency scores estimated

$$se_{\bar{q}} = \{\frac{1}{N}\sum [q_j - \bar{q}]^2\}^{1/2} \quad (10)$$

Results

Figure 1 plots the efficiency scores for each country, the confidence intervals produced by the jackknife and bootstrap. The confidence intervals are larger from the bootstrap than from the jackknife. In DEA models countries are scored relatively to the most efficient country, here Turkey is always the most efficiency country, unless it is omitted. With only 1 observation removed each time there is little variation in the efficiency scores estimated in each of the jackknife replications, leading to narrow confidence intervals and increasing the chance of a type I error. The bootstrap is seen as more efficient and robust. Although an examination of Figure 1 demonstrates that the jackknife and the bootstrap detect similar patterns in the data providing validity to the confidence intervals estimated.

Figure 1 Confidence Intervals around country efficiency scores

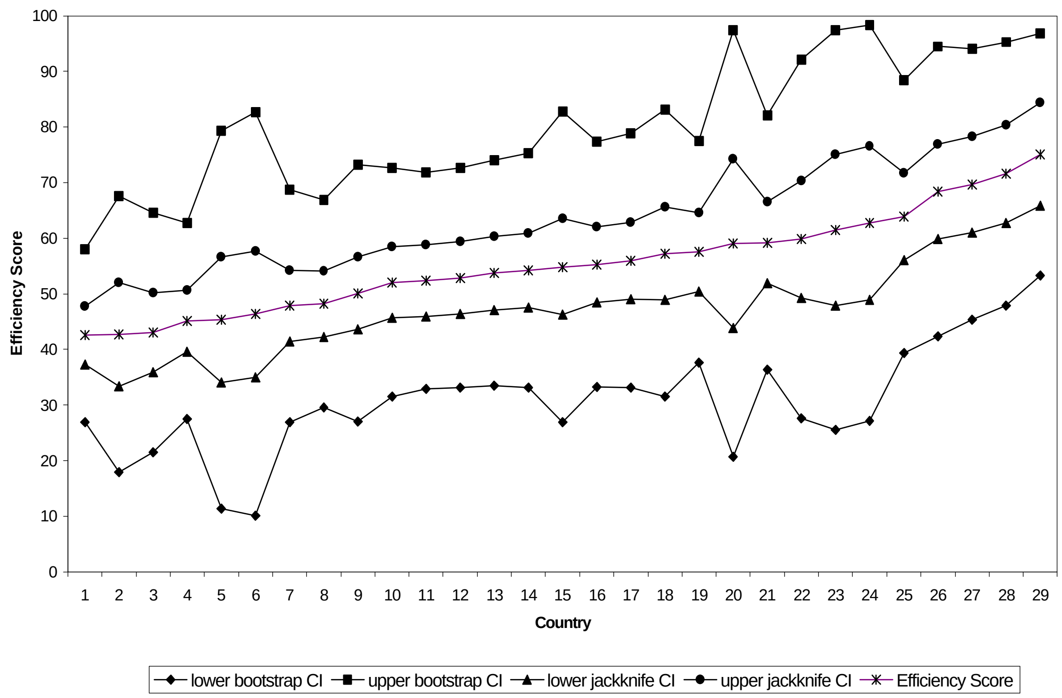
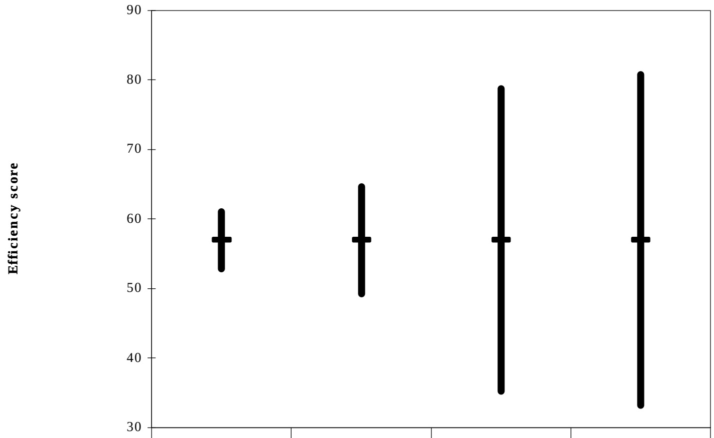


Figure 2 illustrates the confidence intervals produced around the mean scores. The smallest confidence interval is produced by resampling from the estimated efficiency scores (i), the jackknife (iii) produces the next smallest confidence intervals, bootstrapping the model fully (ii) produces the next widest confidence intervals. The application of the usual standard error formula (iv) produces the widest confidence interval. The results from (ii) and (iv) are very similar suggesting, in this case, that 100 replications are sufficient. The bootstrap method is considered more robust and efficient than the jackknife procedure and is preferable. The different methods do demonstrate that with very large data sets, where running the a new DEA model is time consuming, that attempting to shortcut the bootstrap process by resampling from the estimated scores (as in (i)) produces confidence intervals that are very narrow. It is preferable to produce a confidence interval by using the jackknife or the full bootstrap. For this sample size the jackknife produces narrower confidence intervals than the bootstrap but is less computer intensive.

Figure 2 Confidence intervals around mean efficiency score



	i	ii	iii	iv
lower confidence interval	52.88	49.26	35.25	33.2
upper confidence interval	61.03	64.65	78.74	80.71
— mean	56.995	56.995	56.995	56.995

Method

Conclusion

We have demonstrated that different resampling methods produce different confidence intervals and that researchers must be careful in their choice of method. The method of producing confidence intervals for summary statistics is especially important. Researchers may be tempted to bootstrap from the estimated efficiency scores to generate standard errors. We demonstrate, using the mean, that this approach reduces the size of the confidence intervals, increasing the chance of a type I error. When attempting to assess the accuracy of summary statistics from nonparametric estimation the jackknife or bootstrap should be performed fully.

References

Cooper, W., Seiford, L.M., Tone, K. (2000) *Data Envelopment Analysis*. Kluwer: Dordrecht.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annuls of Statistics* 7: 1-26.

Efron, B. and Tibshirani, R.J. (1993)*. An Introduction to the Bootstrap*. Chapman & Hall, New York.

Hollingsworth, B., Dawson, P., Maniadakis, N. Efficiency (1999). Measurement of Health Care: A review of non-parametric methods and applications. *Health Care Management Science* 2: 161-172.

Jones, A. (2000) Health Econometrics. In Culyer, A. and Newhouse, J. eds, *The Handbook of Health Economics*. Elsevier, Oxford.