

# Double Bootstrap Confidence Intervals in the Two-Stage DEA Approach<sup>♦</sup>

Dimitris K. Chronopoulos<sup>a\*</sup>, Claudia Girardone<sup>b</sup> and John C. Nankervis<sup>b</sup>

<sup>a</sup> School of Management, University of St Andrews

<sup>b</sup> Essex Business School, University of Essex

**Forthcoming in the Journal of Time Series Analysis**

## Abstract

Contextual factors usually assume an important role in determining firms' productive efficiencies. Nevertheless, identifying them in a regression framework might be complicated. The problem arises from the efficiencies being correlated with each other when estimated by Data Envelopment Analysis (DEA), rendering standard inference methods invalid. Simar and Wilson (2007) suggest the use of bootstrap algorithms that allow for valid statistical inference in this context. This paper extends their work by proposing a double bootstrap algorithm for obtaining confidence intervals with improved coverage probabilities. It also offers an empirical application using bank-level data. Moreover, acknowledging the computational burden associated with iterated bootstrap procedures, we provide an algorithm based on deterministic stopping rules which is less computationally demanding. Monte Carlo evidence shows considerable improvement in the coverage probabilities after iterating the bootstrap procedure. The results also suggest that percentile confidence intervals perform better than their basic counterpart. The empirical application of the double bootstrap to a sample of European banks suggests that diversified financial conglomerates are more efficient than specialised institutions and the findings are robust to alternative double bootstrap confidence interval methods.

*JEL Classification:* C14, C15, C24, G21

*Keywords:* Data Envelopment Analysis; Double Bootstrap; Confidence Intervals; Stopping Rules; Two-Stage Approach.

---

<sup>♦</sup> This paper is dedicated to the memory of John C. Nankervis, an inspiring mentor and colleague. He was a true scholar and a gentleman.

<sup>\*</sup> *Corresponding Author:* Dimitris K. Chronopoulos, School of Management, University of St Andrews, St Andrews, KY16 9RJ. E-mail: [dc45@st-andrews.ac.uk](mailto:dc45@st-andrews.ac.uk).

## 1. INTRODUCTION

Production analysis is of key interest to both managers and policy-makers alike. It is therefore only natural that considerable research effort has been invested in the development of frontier methods for evaluating the efficiency of a productive unit relative to its peers. Firms are typically evaluated in terms of their ability to maximise output production (minimise input usage) with given inputs (output levels) relative to other firms' performance in some comparison set. Data Envelopment Analysis (DEA) offers a non-parametric alternative to the parametric frontier production function analysis and has become one of the most commonly used frontier methodologies (Berger and Humphrey, 1997).

DEA relies upon linear programming methods for obtaining an estimate of the true production frontier, also known as the "best practice" frontier, relative to which the performance of a productive unit is measured (Charnes et al., 1978). A firm's failure to produce the estimated optimal output levels using the minimum inputs required results in a less than optimal efficiency level. Moreover, it is often the case that not all of the sampled firms are exposed to the same environmental or contextual factors. Thus efficiency measures also incorporate additional variation, which should not be attributed to management performance, as these exogenous factors are often beyond its control, at least in the short term. As a result, the efficiency estimates are typically subjected to a second stage regression analysis to determine the impact upon efficiency of these environmental factors. Knowledge about these relationships can help managers improve their firms' performance, while policy makers can develop a more comprehensive understanding of the costs that new policies can impose on the markets.

This approach, though, is not without problems. In particular, a serious concern, raised by Xue and Harker (1999), stems from the dependency of the efficiency measures generated by DEA.<sup>1</sup> They point out that the DEA efficiency estimate of a firm is a function not only of its own input-output set but of the input-output sets of all sampled firms. This creates a problem of correlation

---

<sup>1</sup> Another problem relates to the bounded nature of the efficiency measure and the appropriate model to be used for estimating the second stage regression (e.g. Tobit or Truncated models). However, this issue is not of concern in this paper (see Simar and Wilson, 2007).

among the efficiency estimates.<sup>2</sup> Simar and Wilson (2007) (henceforth SW) point out that maximum likelihood estimation of the second stage regression will still yield consistent coefficient estimates, since the correlation among the DEA efficiency estimates disappears asymptotically. Nevertheless, this occurs at a rate that might invalidate conventional inference procedures. This is especially true in those cases where the production technology, assumed in the first stage for estimating the efficiency levels, is defined in an input-output space with more than three dimensions.

SW (2007) propose two algorithms for constructing confidence intervals for the parameters of the second stage regression. Both are based on a bootstrap resampling scheme applied to the second stage of an otherwise typical two-stage DEA approach. The main distinction between the two algorithms is that the second one incorporates an additional procedure for bias-correcting the efficiency estimates before being fitted to a truncated model. This however, could take its toll on the statistical efficiency of the second stage regression parameters. That is, unless the bias being corrected is large, the bias reduction procedure implemented in the second algorithm is likely to add noise to the estimation. Monte Carlo experiments have shown that there are cases where the second algorithm does not perform as well as the first one in terms of coverage error (Simar and Wilson, 2007).

In this paper, first we investigate the coverage accuracy of double bootstrap confidence intervals for the parameters of a truncated model within the context of the two-stage DEA approach. Simulating data from a process that justifies regression of efficiency estimates on some exogenous factors, we examine improvements in the coverage errors of confidence intervals generated with SW (2007) first algorithm after calibrating the procedure using a second level bootstrap (see Hall, 1986; Beran, 1987; Hall and Martin, 1988; Shi, 1993). The Monte Carlo evidence indicates that iterating the bootstrap principle results in confidence intervals with much improved coverage probabilities, especially as the dimensionality of the first stage problem increases.

The paper also examines the convergence properties of the coverage rates of both the basic bootstrap intervals and its simpler percentile counterpart, following the terminology of Davison and Hinkley (1997). The Monte Carlo evidence suggests that the dimensionality of the input-output space,

---

<sup>2</sup> In an attempt to remedy this dependency problem Xue and Harker (1999) suggest a non-parametric bootstrap approach. This is based on re-sampling from the inputs and outputs of the originally sampled firms and then fitting an OLS model to the efficiency estimates after reapplying DEA to each of the bootstrap samples.

within which the production process is defined, has an adverse effect on the convergence rate of both the basic and the percentile bootstrap intervals. Moreover, the simulation results indicate a non-monotonic convergence to their nominal significance levels for both bootstrap interval methods investigated in this study when the more realistic 3-input 3-output production process is considered.

In recent years sophisticated efficiency and productivity models are often applied to the banking sector. In particular, DEA studies often rely on bootstrapping techniques to obtain confidence intervals for the estimated efficiency scores as well as in the second stage analysis of the determinants of estimated efficiencies (see for a survey e.g., Fethi and Pasiouras, 2010). In this paper we provide an empirical application of the double bootstrap methodology to a sample of European banks and investigate whether greater diversification can lead to improved efficiency. Our evidence suggests that larger and more diversified banks are also more efficient, a finding which is robust to different confidence intervals.

This paper is organised as follows. In section 2 we describe the concept of the bootstrap procedure, how it is currently applied to the two-stage DEA approach and an algorithm to calibrate the bootstrap confidence intervals. Section 3 describes the Monte Carlo experiments, presents an algorithm for reducing the computational burden induced from iterating the bootstrap and discusses the empirical results. An application of the methodology to a sample of European banks is described in section 4. Finally, the last section concludes.

## **2. DEA AND THE BOOTSTRAP**

### **2.1 Second Stage Regression and the Bootstrap Conjecture**

Let  $T(X) = \hat{\theta}$  be an estimator of  $\theta$ , the parameter of interest, and  $X$  be a random sample drawn from the population with c.d.f.  $F_\theta$ . Now, let  $X^*$  denote the bootstrap samples and  $T(X^*) = \hat{\theta}^*$ . If  $F_\theta$  is unknown, one can use the empirical distribution function to construct the bootstrap samples. Otherwise, if  $F_\theta$  is well defined apart from the parameter  $\theta$ , instead of drawing with replacement

from the original sample, one can draw from the distribution  $F_{\hat{\theta}}$ , where  $\hat{\theta}$  is the estimated value of  $\theta$  obtained from the original sample. The latter version is known as the parametric bootstrap.

Suppose we wish to construct a two-sided confidence interval for  $\theta$ . The  $100(1-\alpha)\%$  basic bootstrap confidence interval with nominal coverage of  $1-2\alpha$  is

$$I_{BET} = \left[ \hat{\theta} - (\hat{\theta} - \theta)_{1-\alpha}, \hat{\theta} - (\hat{\theta} - \theta)_{\alpha} \right] \quad [1]$$

where  $(\hat{\theta} - \theta)_{\alpha}$  is the  $\alpha$ -quantile of  $\hat{\theta} - \theta$ .

Alternatively, the two-sided percentile confidence interval of nominal coverage  $1-2\alpha$  is

$$I_{PCTL} = \left[ \hat{\theta}_{\alpha}, \hat{\theta}_{1-\alpha} \right] \quad [2]$$

where  $\hat{\theta}_{\alpha}$  is the  $\alpha$ -quantile of  $\hat{\theta}$ .

Consequently, knowledge of the sampling distributions of  $\hat{\theta} - \theta$  and  $\hat{\theta}$  is required for constructing [1] and [2], respectively. However, such information is not available in practice.

Bootstrap procedures, in this situation, can be a valuable tool. The bootstrap principle implies that information on the relationship between the true parameter  $\theta$  and its estimator  $\hat{\theta}$  can be obtained by treating  $T(X)$  as the true parameter value in question and looking at the relationship between  $T(X)$  and  $T(X^*)$ , its bootstrap estimates. Based on this conjecture, in the case of the basic confidence interval method, for instance, the distribution of  $T(X^*) - T(X)$  could be used to approximate the sampling distribution of  $\hat{\theta} - \theta$ , provided that the bootstrap is consistent for  $\hat{\theta}$ .

In practice, the theoretical confidence intervals defined in [1] and [2] can be approximated by drawing  $B$  independent bootstrap samples  $X_b^*$  and estimating the  $\hat{\theta}_b$ ,  $b=1, \dots, B$ . Thus the Monte Carlo approximations to  $I_{BET}$  and  $I_{PCTL}$  are defined to be

$$\hat{I}_{BET} = \left[ \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{((1-\alpha)(B+1))}, \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{(\alpha(B+1))} \right] \quad [3]$$

and

$$\hat{I}_{PCTL} = \left[ \hat{\theta}_{(\alpha(B+1))}^*, \hat{\theta}_{((1-\alpha)(B+1))}^* \right] \quad [4]$$

where subscripts in parentheses denote ordered values and  $a(B+1)$  is an integer.

SW (2007) laid the foundation for the use of a consistent bootstrap in the two-stage DEA approach context. Accordingly, confidence intervals for the parameters of the second stage regression appropriate for inference can be constructed with the following algorithm:

- 1) Solve the output oriented DEA problem to obtain the efficiency estimate  $\hat{\delta}_i$  for each firm  $i=1, \dots, n$ .
- 2) Regress  $\hat{\delta}_i$  on the environmental variables  $z_i$  using the truncated regression model to obtain estimates  $\hat{\beta}$  and  $\hat{\sigma}_\varepsilon$  of the parameters  $\beta$  and  $\sigma_\varepsilon$  respectively.
- 3) Repeat the following three steps  $B$  times to obtain a set of bootstrap estimates

$$B^* = \left\{ (\hat{\beta}^*, \hat{\sigma}_{\varepsilon}^*)_b \right\}_{b=1}^B$$

- a) For each  $i=1, \dots, n$ , draw  $\varepsilon_i^*$  from the  $N(0, \hat{\sigma}_\varepsilon^2)$  distribution with left truncation at  $(1 - z_i \hat{\beta})$ .
  - b) For each  $i=1, \dots, n$ , compute  $\delta_i^* = z_i \hat{\beta} + \varepsilon_i^*$ .
  - c) Estimate  $\hat{\beta}_b^*$  and  $\hat{\sigma}_{\varepsilon, b}^*$ , by regressing  $\delta_i^*$  on  $z_i$ , again using the truncated regression model.
- 4) Construct the confidence intervals [3] or [4] as described above.

The actual coverage probabilities of these bootstrap confidence interval methods however, may not coincide with their nominal coverage probabilities. The difference between the actual coverage rate of a confidence interval and the claimed value  $1-2a$  is defined as the coverage error. The primary source of this error, provided that the number of bootstrap replications is sufficiently large to render Monte Carlo error negligible, is the use of an approximation of  $\hat{\theta} - \theta$  in  $I_{BET}$  or similarly the substitution of  $\hat{\theta}_\alpha$  with  $\hat{\theta}_{(\alpha(B+1))}^*$ , its bootstrap approximation, in  $I_{PCTL}$ .

The double bootstrap is designed to estimate this coverage error and adjust, based on this estimate, the procedure in such a way that the coverage probability improves.

## 2.2 Iterating the Bootstrap Principle

There are various ways of calibrating the bootstrap confidence intervals through iteration of the bootstrap principle (see Efron and Tibshirani, 1993). Here we use the double bootstrap with no pivot as described in Shi (1993).

First, we consider the basic double bootstrap confidence interval for the parameter  $\theta$

$$I'_{BET} = \left[ \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{q(1-\alpha)}, \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{q(\alpha)} \right] \quad [5]$$

In practice, this theoretical interval can be approximated by means of a Monte Carlo simulation.

Therefore, its empirical estimate is defined as

$$\tilde{I}_{BET} = \left[ \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{\hat{q}(1-\alpha)}, \hat{\theta} - (\hat{\theta}^* - \hat{\theta})_{\hat{q}(\alpha)} \right]. \quad [6]$$

$\tilde{I}_{BET}$  can be constructed by drawing a set of  $M$  second level bootstrap samples for each of the  $B$  ones obtained from the first level bootstrap and calculating the estimates  $\hat{\theta}_b^*$  and  $\hat{\theta}_{bm}^{**}$  based on the first and second level bootstrap samples respectively,  $b=1, \dots, B$  and  $m=1, \dots, M$ . Subsequently, estimates of  $\hat{q}(\alpha)$  and  $\hat{q}(1-\alpha)$  can be obtained by  $u_{(a)(B+1)}^*$  and  $u_{((1-a)(B+1))}^*$  respectively with  $u_b^*$  given by

$$u_b^* = \frac{\sum_{m=1}^M I(\hat{\theta}_{bm}^{**} \leq \hat{\theta}_b^* - \hat{\theta})}{M} \quad [7]$$

(see Davison and Hinkley, 1997; p225). Here  $I(\cdot)$  denotes the indicator function, equal to one if the argument is true and zero otherwise.

Similarly, the percentile double bootstrap confidence interval is

$$I'_{PCTL} = \left[ \hat{\theta}_{q(a)}^*, \hat{\theta}_{q(1-\alpha)}^* \right] \quad [8]$$

with its empirical estimate given by

$$\tilde{I}_{PCTL} = \left[ \hat{\theta}_{\hat{q}(a)}^*, \hat{\theta}_{\hat{q}(1-\alpha)}^* \right]. \quad [9]$$

Again, values for  $\hat{q}(\alpha)$  and  $\hat{q}(1-\alpha)$  can be obtained using the algorithm described above. However, in the percentile double bootstrap confidence interval method  $u_b^*$  is defined differently than for the basic double bootstrap (Davison and Hinkley, 1997; p250). In this case we have

$$u_b^* = \frac{\sum_{m=1}^M I(\hat{\theta}_{bm}^{**} \leq \hat{\theta})}{M}. \quad [10]$$

The idea of calibrating the confidence intervals for the parameters of the DEA second stage regression can be implemented, by introducing an additional bootstrap level after step (3) as follows:

- 1) Solve the output oriented DEA problem to obtain  $\hat{\delta}_i$  for each firm  $i = 1, \dots, n$ .
- 2) Regress  $\hat{\delta}_i$  on the environmental variables  $z_i$  using the truncated regression model to obtain estimates  $\hat{\beta}$  and  $\hat{\sigma}_\varepsilon$  of the parameters  $\beta$  and  $\sigma_\varepsilon$  respectively.
- 3) Repeat the following three steps  $B$  times to obtain a set of bootstrap estimates

$$A^* = \left\{ (\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)_b \right\}_{b=1}^B$$

- a) For each  $i = 1, \dots, n$ , draw  $\varepsilon_i^*$  from the  $N(0, \hat{\sigma}_\varepsilon^2)$  distribution with left truncation at  $(1 - z_i \hat{\beta})$ .
  - b) For each  $i = 1, \dots, n$ , compute  $\hat{\delta}_i^* = z_i \hat{\beta} + \varepsilon_i^*$ .
  - c) Estimate  $\hat{\beta}_b^*$  and  $\hat{\sigma}_{\varepsilon,b}^*$ , by regressing  $\hat{\delta}^*$  on  $z$  using again the truncated regression model.
- 4) For each set of bootstrapped estimates  $A_b^*$ ,  $b = 1, \dots, B$  do the following:
    - a. For each  $i = 1, \dots, n$ , draw  $\varepsilon_i^{**}$  from the  $N(0, \hat{\sigma}_\varepsilon^{*2})$  distribution with left truncation at  $(1 - z_i \hat{\beta}^*)$ .
    - b. For each  $i = 1, \dots, n$ , compute  $\hat{\delta}_i^{**} = z_i \hat{\beta}^* + \varepsilon_i^{**}$ .
    - c. Estimate  $\hat{\beta}_{bm}^{**}$  and  $\hat{\sigma}_{\varepsilon bm}^{**}$ , by regressing  $\hat{\delta}^{**}$  on  $z$  using again the truncated regression model.
  - 5) Repeat steps (4a) – (4c)  $M$  times to obtain a set of double bootstrap estimates

$$A^{**} = \left\{ \left\{ (\hat{\beta}^{**}, \hat{\sigma}_\varepsilon^{**})_m \right\}_{m=1}^M \right\}_{b=1}^B.$$



- 6) Obtain the  $u_b^*$ ,  $b=1, \dots, B$  using either [7] or [10] together with the set of double bootstrap estimates  $A^{**}$ .

### 3. SIMULATIONS

#### 3.1 Design of the Experiment

Let  $\delta_i$  denote the true efficiency level of firm  $i$  and  $x_{ip}$  and  $y_{iq}$  be respectively the vectors of inputs used and outputs produced by the same firm, with  $i=1, \dots, n$ ,  $p=1, \dots, P$  and  $q=1, \dots, Q$ . Finally, let  $z_i$  be a vector of environmental variables faced by the  $i^{th}$  firm.

The data generating process used for the simulations follow assumptions A1 to A8 of SW (2007). These concern the distribution of the observations, the production technology and some regularity conditions necessary for the consistency of the DEA estimators. The first three assumptions referring to the distribution of inputs, outputs and environmental variables faced by the firms are briefly:

A1. That all the  $n$  observed input output and environmental factor bundles  $(x, y, z)$  with a joint distribution function  $f(x, y, z)$  are identically and independently distributed on the production possibility set defined by  $T = \{(x, y) : y \text{ can be produced by } x\}$

A2. The efficiency levels are conditioned on the environmental factors faced by the firms as described by the model:

$$\delta_i = \psi(z_i, \beta) + \varepsilon_i \geq 1. \quad [11]$$

A3. The error term  $\varepsilon_i$  of the model is independently and identically distributed, normal with zero mean and left truncated at  $1 - \psi(z_i, \beta)$ .

By A1 above the environmental variable is generated randomly as  $z_{i2} \sim N(\mu_z, \sigma_z)$ , where  $z_{i1} = 1$ . More specific than A2, the mechanism through which firm efficiency is affected by environmental factors is explained by the linear model:

$$\delta_i = z_i \beta + \varepsilon_i \quad [12]$$

where, as in A3, the innovations  $\varepsilon_i$  are drawn from a  $N(\mu_\varepsilon, \sigma_\varepsilon)$  left truncated at  $1 - z_i\beta$ . The inputs are assumed uniformly distributed over the interval  $[6,16]$ , i.e.  $x_{ip} \sim \text{uniform}(6,16)$ . In the single output case, the quantity produced by the  $i^{\text{th}}$  firm is determined by the production function showing the maximum output quantity producible from  $x_{ip}$  after adjusting it to reflect the firm's level of efficiency,  $y_i = \delta_i^{-1} \sum_{p=1}^P x_{ip}^{3/4}$ . In the multiple output case, the output level for each "product-line" is obtained by disaggregating the total production of the firm,  $\zeta_i = \delta_i^{-1} \sum_{p=1}^P x_{ip}^{3/4}$  into its constituent components. If  $Q=2$  we draw  $\alpha_1 \sim \text{uniform}[0,1]$ , whereas if  $Q \geq 3$ ,  $\alpha_1$  is generated as before and we additionally draw  $\alpha_\ell \sim \text{uniform}(0, 1 - \sum_{K=1}^{\ell-1} \alpha_K)$ , for each  $\ell = 2, \dots, Q-1$ . Finally, the disaggregation is performed in the following manner: for  $q=1, \dots, Q-1$  we set  $y_{iq} = \alpha_q \zeta_i$ , whereas the output level of the final "product-line" is given by  $y_{iQ} = (1 - \sum_{k=1}^{Q-1} \alpha_k) \zeta_i$ .

Each experiment consists of 1000 trials. Following SW (2007), we consider three distinct cases of technology, where the number of inputs and outputs are  $p=q=\{1, 2, 3\}$ . Moreover, for comparison purposes we also set the values of the parameters as  $\mu_\varepsilon = 0$ ,  $\sigma_\varepsilon = 1$ ,  $\mu_z = 2$ ,  $\sigma_z = 2$ ,  $\beta_1 = \beta_2 = 0.5$ . For each trial in a Monte Carlo experiment we keep track of whether or not the constructed confidence intervals cover the true values of  $\beta_1$ ,  $\beta_2$  and  $\sigma_\varepsilon$ .

### 3.2 A Fast Double Bootstrap Algorithm

Monte Carlo experiments involving the bootstrap can become rather expensive since the number of estimations of the statistic of interest is  $R * B$ , where  $R$  is the number of the Monte Carlo replications and  $B$  is the number (usually large) of the bootstrap replications. Evaluating the accuracy of double bootstrap confidence intervals by means of Monte Carlo simulation is even more computationally demanding as each Monte Carlo trial now involves a multiple  $M$  of the first layer bootstrap replications.

A plausible approach to ease the simulation's computational burden would be to avoid calculations that would have no impact on the final results. Such an approach is adopted here by incorporating a set of deterministic stopping rules, in the spirit of Nankervis (2005), in the Monte Carlo experiments. These stopping rules are deterministic in the sense that the obtained results are the same as if all the double bootstrap computations had been performed. The stopping rules used in the Monte Carlo experiments conducted for this paper are presented below.

**Double bootstrap two sided equal-tailed basic (percentile) confidence intervals using stopping rules.**

- 1) Do all  $B$  single bootstrap calculations to obtain the set of bootstrap estimates

$A^* = \left\{ (\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\sigma}_\varepsilon^*)_{b_1} \right\}_{b=1}^B$  needed for obtaining double bootstrap samples  $X_{bm}^{**}$ ,  $m = 1, \dots, M$  and double bootstrap statistics.

- 2) Sort each of the  $\hat{\beta}_{1b}^*$ ,  $\hat{\beta}_{2b}^*$  and  $\hat{\sigma}_b^*$  in an increasing order (decreasing in the percentile case) while preserving at the same time the information on their correspondence between each other.

- 3) Let  $K$  be  $\hat{\beta}_{1b}^*$ .

- 4) Do an initial  $2\alpha(R+1)$  sets of double bootstrap replications using the smallest  $\alpha(R+1)$  and largest  $\alpha(R+1)$  values of  $K$ . Use [7] (or [10]) to obtain values of  $u_{wb}^*$ , where

$$w = \hat{\beta}_{1b}^*, \hat{\beta}_{2b}^*, \hat{\sigma}_{\varepsilon b}^*.$$

- 5) Order the values of  $u_{wb}^*$  in an increasing fashion. Define

$$LOW_w = \left\{ U_{w,(1)}^*, \dots, LOW_{w,\max} = U_{w,(\alpha(R+1))}^* \right\}, \quad UP_w = \left\{ UP_{w,\min} = U_{w,(\alpha(R+1)+1)}^*, \dots, U_{w,(2\alpha(R+1))}^* \right\}.$$

For each  $w$ ,  $LOW_{w,\max}$  and  $UP_{w,\min}$  are correspondingly the upper and lower bounds for  $\hat{q}(\alpha)$  and  $\hat{q}(1-\alpha)$ .

- 6) If  $LOW_{K,\max} = 0$  and  $UP_{K,\min} = 1$  go to step 12.

- 7) If just  $LOW_{K,\max} = 0$  (or just  $UP_{K,\min} = 1$ ) then do the remaining bootstrap calculations in an order corresponding to decreasing (increasing) values of  $K$ .
- 8) If neither  $LOW_{K,\max} = 0$  nor  $UP_{K,\min} = 1$  then carry out the remaining sets of bootstrap replications in an order corresponding to alternating high and low values of  $K$ .
- 9) In each set of double bootstrap replications stop further calculations if for all  $w = \hat{\beta}_{1b}^*, \hat{\beta}_{2b}^*, \hat{\sigma}_{eb}^*, u_{wb}^*$  is at least as large as  $LOW_{w,\max}$  and cannot exceed  $UP_{w,\min}$ . Put differently, we would stop after  $M_b^*$  double bootstrap replications if the following expressions
- $$\sum_{m=1}^{M_b^*} I(\hat{\theta}_{b,m}^{**} \leq 2\hat{\theta}_b^* - \hat{\theta}) \geq M \times LOW_{\hat{\theta}^*_{\max}} \quad \text{and} \quad \sum_{m=1}^{M_b^*} I(\hat{\theta}_{b,m}^{**} \leq 2\hat{\theta}_b^* - \hat{\theta}_i) + M - M_b^* \leq M \times UP_{\hat{\theta}^*_{\min}}, \quad \text{hold true}$$
- for every  $\theta = \beta_1, \beta_2, \sigma$ .
- 10) Add every newly estimated  $u_w^*$ , to those already obtained at step 4 and perform step 5 once again.
- 11) Check whether both  $LOW_{K,\max} = 0$  and  $UP_{K,\min} = 1$ . If not go back to step 7.
- 12) If no remaining sets of bootstrap replication are left to be done and the above condition holds: stop.
- 13) Otherwise continue to the next step.
- 14) Substitute  $\hat{\beta}_{2b}^*$  for  $K$  and repeat steps 7-12. Start estimations using the remaining  $2\alpha(R+1)$  largest and smallest values of  $K$  obtained at step 2, if any are still left, before using any other remaining values of  $K$ . This time apply the stopping rule described at step 9 only to  $w = \hat{\beta}_{2b}^*, \hat{\sigma}_{eb}^*$ .
- 15) If the condition in step 11 holds and there are still sets of bootstrap replications left to be done continue to the next step, otherwise stop.
- 16) Substitute  $\hat{\sigma}_{eb}^*$  for  $K$  and repeat steps 7-12. Start the estimations using the remaining  $2\alpha(R+1)$  largest and smallest values of  $K$  of step 2, if any are still left, before using any other remaining values of  $K$ . This time apply the stopping rule described at step 9 only to  $w = \hat{\sigma}_{eb}^*$ .

### 3.3 Results

Results are presented for the empirical coverage probabilities of the basic and percentile bootstrap confidence intervals and their double bootstrap versions as well as for the asymptotic intervals. Four confidence levels are considered; .80, .90, .95 and .99. Our discussion centres on the slope parameter, which is the parameter of interest in the typical two-stage type of analysis. Thus, for the sake of brevity, we present results only for the slope parameter. The bootstrap results are based on 1999 replications, and in the case of the iterated bootstraps each second-level bootstrap is based on a maximum 250 replications.

Table 1 reports Monte Carlo results on the accuracy of two types of bootstrap confidence intervals we have already discussed, the basic bootstrap and its percentile counterpart, as well as those constructed based on the inverse of the negative Hessian of the log-likelihood function. Average widths of all three methods are shown in Table 2.

< **Insert Table 1 about here** >

< **Insert Table 2 about here** >

A clear pattern emerges when looking at Table 1. Given the sample size, the estimated coverage probabilities of the confidence intervals of both algorithms deteriorate as the dimensionality of the problem in the first stage increases. This is consistent with the “curse of dimensionality” that plagues the DEA efficiency estimator (Kneip et al., 1998). Looking at the coverage probabilities across the different confidence interval methods, the performance of the basic bootstrap interval is broadly comparable to that of the asymptotic one. Table 2, however, indicates that the basic bootstrap produces less precise intervals than the conventional method of inference. Nevertheless, the latter is consistent only in the simple case of  $p = q = 1$ . As noted earlier, conventional inference methods are rendered inconsistent in those cases where  $p + q > 3$  (see Simar and Wilson, 2007). Comparing the simpler percentile method discussed by Efron and Tibshirani (1993) with the basic bootstrap interval reveals that the former performs better in terms of coverage probability. Moreover, the percentile bootstrap unlike the basic one exhibits significantly better coverage rates than the asymptotic method.

That is true especially for sample sizes that are relatively small given the dimensions of the problem in the first stage.

Nevertheless, the accuracy of either the basic or the percentile bootstrap interval in both multi-input multi-output cases considered in this study is not ensured even with 10,000 observations; a sample size that is not easily within reach of applied researchers. What is more, there is evidence that the coverage rate of the basic confidence interval, in the case of  $p = q = 2$ , and for all but the .99 level of confidence are statistically different from the nominal rate even with 15,000 observations. Finally, the last panel of Table 1 presents the results for a more complex production process with  $p = q = 3$ . In this case, and contrary to expectations, the coverage errors of both bootstrap intervals appear to increase as the sample size increases with no sign of convergence even when the maximum number of observations considered in this study is used. These results are suggestive of the sample sizes that may be required in production processes more complicated than  $p = q = 3$ .

Turning to the double bootstrap methods, Table 3 shows, for three different sample sizes,  $n = \{100, 200, 400\}$ , the coverage probabilities of the estimated confidence intervals over 1000 Monte Carlo trials. Average widths of these methods are shown in Table 4.<sup>3</sup>

< **Insert Table 3 about here** >

< **Insert Table 4 about here** >

We observe that both iterated bootstrap confidence interval methods significantly improve coverage accuracy over their single bootstrap counterparts. Careful inspection of the table reveals that this improvement is even more pronounced in the case of multi-input multi-output technologies, especially for the basic double bootstrap over its single bootstrap counterpart. In the special case of  $p = q = 1$ , the percentile double bootstrap method generates confidence intervals that achieve nominal coverage probabilities, even with a sample size of only 100 observations. The basic double bootstrap confidence intervals have a similar performance. The only exception is at .99 significance, for which

---

<sup>3</sup> Following Nankervis (2005) we also estimate the computational efficiencies achieved by the stopping rules. Monte Carlo results for the more realistic scenario of 3-input 3-output production process with a sample size of 400 firms indicate that the mean computational efficiencies due to the stopping rules are 1.8 and 2.2, respectively, for the estimation of basic and percentile double bootstrap intervals. These efficiencies mean that with a maximum number of double bootstrap replications  $M=250$  only the equivalent of 139 (113) double bootstrap computations are required for estimating the basic (percentile) interval.

there is evidence that the true coverage rate is statistically different from the nominal one in two occasions; at  $n=100$  and  $n=200$ . Nevertheless, the accuracy of both double bootstrap methods is still affected by the dimensionality of the production technology in a manner similar to that observed in Table 1.

#### 4. EMPIRICAL APPLICATION

We apply the SW (2007) method and the double bootstrap refinement suggested above to investigate the output efficiency for a sample of European financial conglomerates and specialised banks, after controlling for environmental factors that may influence their performance.<sup>4</sup>

##### 4.1 Measure of Diversification and Data

Different views exist as to what constitutes a financial conglomerate (e.g. Verweire, 1999). EU legislation defines it as a group of entities comprised of at least an insurance undertaking and a banking or investment services company.<sup>5</sup> Following Laeven and Levine (2007), we use the income diversity index to classify banks into specialised and diversified institutions. The index measures the diversity of a bank's income sources and is defined as:

$$DIVIN = 1 - \left| \frac{(Net\ Interest\ Income - Other\ Operating\ Income)}{Total\ Operating\ Income} \right| \quad [13]$$

Net interest income is defined as interest income minus interest expenses and the other operating income is the sum of net fee income, net commission income and net trading income. The range value of the index is between 0 and 1, with higher values indicating greater income diversification.

The analysis is based on a sample of 1557 commercial and savings banking institutions operating in seventeen European countries drawn from the Bankscope database.<sup>6</sup> The data are extracted from consolidated balance sheets and income statements for the year 2004.<sup>7</sup> This allows for

---

<sup>4</sup> The programming code is available from the authors upon request.

<sup>5</sup> Directive 2002/87/CE.

<sup>6</sup> The countries are the 15 EU member states before the fifth enlargement of 2004 – namely, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom- as well as Norway and Switzerland.

<sup>7</sup> All amounts are expressed in Euro.

capturing the diversification activity of banking groups that is undertaken through their subsidiaries. Table 5 presents the summary statistics for the sampled banks and their country composition. The table shows the variation in size and income sources both across the different types of institutions as well as countries. The average commercial bank appears significantly larger in size measured by total assets and more diversified than the corresponding savings bank. However, the size difference is moderate when looking at the median figures, whereas commercial banks still seem considerably more diversified than savings banks. A closer look at the cross-country differences also reveals that the average Swiss bank has less diverse income sources than the average bank in the sample. Despite Switzerland being considered as a traditional universal banking country, there are only a relatively small number of banks fully engaged in universal banking activities (Rime and Stroh, 2003). Hence, this lends further support to our choice of a diversification measure based on financial data instead of classifying banks for example on the basis of whether they are headquartered in a country that prohibits universal banking or not.

**<Insert Table 5 about here>**

#### **4.2 Inputs and Outputs Definition**

One of the most important steps in evaluating relative efficiency in banking is the selection of appropriate inputs and outputs. Following Berger and Humphrey (1997), we adopt the “intermediation” approach, which views banks as institutions that employ labour, physical capital and deposits to produce loans and other assets. Accordingly, we consider personnel expenses, total fixed assets and deposits as inputs and total customer loans and other earning assets as outputs. Since this study deals with banking institutions that have a wider scope of activities than the traditional bank, capturing the non-traditional aspect of banks’ production technology is essential. Hence, a measure of off-balance sheet business is included as one of the outputs to capture banks’ non-traditional activities (e.g. Clark and Siems, 2002). This corresponds to the aggregate nominal value of a bank’s contingent liabilities including, for example, acceptances, documentary credits and guarantees.

However, choosing a parsimonious model to describe the production process is equally important with selecting the right inputs and outputs, as over-specification of the production model



may result in rendering some of the banks efficient by default (Leibenstein and Maital, 1992). Therefore, we investigate whether information embedded in the off-balance sheet activities is also incorporated into the other two output variables by means of the Efficiency Contribution Measure (ECM) test (Pastor, Ruiz and Sirvent 2002).<sup>8</sup>

The data reject the null hypothesis that the off-balance sheet variable is superfluous at the 1% significance level.<sup>9</sup> Thus, the variable should be included in the model as the marginal information on the banks' operating efficiency contained within is significant.

### 4.3 Second Stage Regression and Results

In the second stage, we investigate how a bank's degree of income diversification affects its level of efficiency by regressing the first stage estimates on a financial conglomeration index as well as other environmental factors. Specifically, we estimate the following regression equation.

$$TE_j = \beta_1 DIVIN_j + \beta_2 DIVIN_j^2 + \beta_3 DIVIN_j * \ln AST_j + \beta_4 EQAST_j + \beta_5 \ln AST_j + \beta_6 ROAE_j + \beta_7 MS_j + \beta_8 LLP_j + \beta_9 COMM_j + \sum_{i=1}^{17} d_i Country_{ij} + u_j \quad [14]$$

In this regression equation the dependent variable  $TE_j$  is the output oriented technical efficiency estimated in the first stage. As covariates, we include DIVIN to capture a bank's degree of income diversification,  $\ln AST$ , which is the logarithm of total assets, to control for bank size and a multiplicative interaction term.<sup>10, 11</sup> Following Pallich et al. (2000), Elsas et al. (2010) and Berger et al. (2010), among others, we include  $DIVIN^2$ , a quadratic term for income diversification, in our specification to capture any non-linearities between the level of bank efficiency and the banks' degree of diversification.<sup>12</sup> We also include the EQAST and the ROAE variables. The former, defined as the

---

<sup>8</sup> The ECM test is based on estimating the DEA model twice. First with all the variables included in the model (full model) and subsequently with omitting the candidate variable from the model (reduced model). We set the tolerance level of efficiency change at  $\bar{p}=1.1$  and postulate that the variable is regarded as influential if the proportion of banks with a change in their efficiency levels exceeding the tolerance level due to omitting the off balance sheet variable is greater than 5%.

<sup>9</sup> The ECM statistic is  $T=125$  and follows the binomial distribution with  $p=0.05$  and  $N=1556$ .

<sup>10</sup> Both DIVIN and  $\ln AST$  variables have been demeaned to facilitate the interpretation of the results.

<sup>11</sup> To investigate the robustness of our results with regard to the measurement of diversification we also employ the Asset Diversification measure as defined in Laeven and Levine (2007). Our findings indicate that the relationship between banks' efficiency levels and their degree of diversification remains fairly robust to the employed measures of diversification.

<sup>12</sup> We thank a referee for suggesting the inclusion of a quadratic term for the diversification index.

ratio of equity over total assets, measures the level of capitalisation, while the latter is defined as the return on average equity and measures bank profitability. In addition, we include each bank's market share of deposits (MS) as a measure of market conditions, as well as the LLP variable, defined as the loan loss provision over total customer loans, to capture the quality of each bank's loan portfolio. A dummy variable (COMM) takes the value of 1 if a bank is classified as commercial and 0 otherwise. Finally, we incorporate country dummies to control for systematic differences in operating performance across countries due to features and conditions specific to each country's banking sector and economy.

The results of the estimation are given in Table 6. Column 2 reports the parameter estimates. Columns 3-12 report the endpoints of 95% confidence intervals, respectively generated with asymptotic normal approximation, basic and percentile bootstrap methods and basic and percentile double bootstrap methods with the basic bootstrap confidence interval being the suggested method by SW (2007). As mentioned above the dependent variable,  $TE_j$  is bounded between unity and infinity, with unity representing perfect efficiency. Therefore, estimated parameters with negative signs indicate sources of efficiency and vice versa.

**<Insert Table 6 about here>**

Diversifying into financial services like insurance and securities underwriting and the economic benefits associated with it are often cited as the main driving forces behind mergers between firms specialising into different financial services. The results presented in Table 6 corroborate this view.

The negative sign of the coefficient on the interaction term between size and diversification suggests that for a given level of diversification larger banks be able to achieve higher efficiency levels. A plausible explanation is that larger institutions are operating with some slack capacity which can be set in good use by diversifying into other financial activities as well. This finding is statistically significant at the 5% level for all four bootstrap confidence interval estimates, but according to the simulation results, our preferred method for conducting hypothesis tests is the double bootstrap confidence intervals. Incidentally, the coefficient is also found statistically significant by the

normal approximation method, although this method does not enable consistent inference in this context, given the dimensionality of the problem in the first stage.

Looking at the lower-order interaction-term coefficients, these are statistically significant according to both double bootstrap interval methods. The coefficients for DIVIN and DIVIN<sup>2</sup> suggest a nonlinear (concave) relationship between diversification and operating (in)efficiency for the averaged sized bank in our sample. On the other hand the negative sign of the size coefficient indicates that larger banks are also more efficient, implying that they enjoy scale economy advantages over their smaller counterparts, conditional on being as diversified as the average diversified bank in the sample.

Moreover, we observe that the technical efficiency decreases with the banks' level of capitalisation. In particular the sign of the EQAST coefficient is positive. This finding is in line with other studies in Europe, and suggests that efficient banks finance their assets by relying less on their capital. This is also in accordance with the notion of pro-cyclical leverage (Adrian and Shin, 2008). That is, raising capital for funding asset growth is relatively expensive, especially in periods when money markets are characterised by excess liquidity, as in 2004, resulting in banks turning to alternative more cost-effective funding sources.

This view is corroborated by the negative sign of the LLP coefficient. It implies that efficient banks also have a reduced loan portfolio quality, a finding in line with the skimping hypothesis (Berger and DeYoung, 1997). In other words, it is likely that resources previously devoted to screening and monitoring of loans are transferred to other activities with higher growth prospects. That is, managers faced with the trade-off between short-term performance and future loan quality choose to focus on the former by putting emphasis on growth opportunities in general and asset growth in particular. Moreover, the negative sign on MS variable is consistent with the efficient structure theory (Berger, 1995), lending further support to the above argument. However, the two double bootstrap methods yield different outcomes, hence, the results on the significance of this coefficient are inconclusive. Moreover the coefficient on the ROAE variable is not significant at the 5% level. As such there appears to be no strong relationship between variation in profitability and

efficiency levels. Finally, the results suggest that the commercial banks are more efficient than their savings counterparts.

## 5. SUMMARY AND CONCLUSION

The two-stage approach provides a framework for determining those factors that significantly affect the efficiency of a firm or organisation. This methodology can be a useful tool in the hands of managers and policy makers who consider production analysis to be of key interest. In most realistic cases, though, where production technology is defined in a multidimensional input-output space, classical inference methods are no longer consistent. This is due to the inherent dependency of the efficiency estimates. To overcome this problem, SW (2007) put forward two alternative algorithms for generating confidence intervals, based on bootstrap resampling. The main distinction between the two is that the second algorithm incorporates an additional procedure for bias-correcting the efficiency estimates. Use of bias estimates, however, is likely to introduce additional noise in the procedure. Hence, in this paper we have considered a refinement of SW (2007)'s first algorithm by iterating the bootstrap principle and investigated how the basic interval fares relative to its simpler percentile counterpart.

Our Monte Carlo experiments indicate, in the case of  $p = q = 3$  that convergence of the confidence intervals towards their nominal significance levels are non-monotonic. That is, the coverage error will most likely increase with greater sample size, before it starts converging towards zero. Consequently, this finding has important implications for empirical applications, where the dimensionality of the production problem is similar or greater than this one. In addition, the Monte Carlo evidence confirms that the convergence rate of confidence intervals is adversely affected by the number of dimensions assumed for the production process in the first stage. The simulation results also show that the percentile bootstrap interval method performs better than its basic counterpart and is even better than the asymptotic interval, whenever the latter is valid. Despite its apparent advantage over the basic bootstrap method, the results suggest that in more realistic cases neither the percentile nor the basic bootstrap intervals can approximate well the nominal coverage probabilities. In these

cases, iterating the principle of bootstrapping provides a correction to the single bootstrap intervals, offering a considerable improvement in the coverage errors.

Finally we applied the SW (2007)'s first algorithm and its refinement as suggested in this paper to a sample of European banks. The results indicate that even after controlling for various bank and country specific factors, the degree of diversification has a positive effect on bank efficiency. Moreover, this finding is robust to alternative double bootstrap confidence interval methods.

## **ACKNOWLEDGEMENTS**

We would like to thank Fotios I. Papadimitriou, Stella Papageorgiou, two anonymous referees, the journal Editor-in-Chief Robert Taylor and the guest editors Jerry Coakley and Neil Kellard as well as participants at the University of St Andrews ESRC seminar series, the Computational and Financial Econometrics network, the Multinational Finance Society, and the International DEA Society meetings for useful comments and suggestions.

## **References**

- Adrian, T., and Shin, H.S. (2008). Liquidity and leverage. *Journal of Financial Intermediation* 19, 418–437.
- Beran, R. (1987). Prepevoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.
- Berger, A.N. (1995). The profit-structure relationship in banking: Tests of market power and efficient-structure hypothesis. *Journal of Money Credit and Banking* 27, 404–431.
- Berger, A.N., and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance* 21, 849–870.
- Berger, A.N., Hasan, I., Korhonen, I., and Zhou, M. (2010). Does diversification increase or decrease bank risk and performance? Evidence on diversification and the risk-return tradeoff in banking. *Bank of Finland Discussion Papers Series No. 9/2010*.
- Berger, A.N., and Humphrey, D. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research* 98, 175–212.

- Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* 2, 429–444.
- Clark, J.A., and Siems, T.F. (2002). X-Efficiency in banking: Looking beyond the balance sheet. *Journal of Money Credit and Banking* 34, 987–1013.
- Davison, A.C., and Hinkley, D.V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- Efron, B., and Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.
- Elsas, R., Hackethal, A., and Holzhauser, M. (2010). The anatomy of bank diversification. *Journal of Banking and Finance* 34, 1274-1287.
- Fethi, M.D., Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey'. *European Journal of Operational Research* 204, 189–198.
- Hall, P. (1986). On the bootstrap and confidence intervals. *The Annals of Statistics* 14, 1431–1452.
- Hall, P., and Martin, R. (1988). On bootstrap resampling and iteration. *Biometrika* 75, 661-671.
- Kneip, A., Park, B., and Simar, L. (1998). A note on the convergence of nonparametric DEA efficiency measures. *Econometric Theory* 14, 783–793.
- Laeven, L., and Levine, R. (2007). Is there a diversification discount in financial conglomerates?. *Journal of Financial Economics* 85, 331–67.
- Leibenstein, H., and Maital, S. (1992). Empirical estimation and partitioning of X-Inefficiency: A Data-Envelopment Approach. *American Economic Review* 82, 428–433.
- Nankervis, J.C. (2005). Computational algorithms for double bootstrap confidence intervals. *Computational Statistics and Data Analysis* 49, 461–475.
- Palich, L.E., Cardinal, L.B., and Miller, C.C. (2000). Curvilinearity in the diversification – performance linkage: An examination of over three decades of research. *Strategic Management Journal* 21, 155-174.
- Pastor, J.T., Ruiz, J.L., and Sirvent, I. (2002). A statistical test for nested radial DEA models. *Operations Research* 50, 728–735.

- Rime, B., and Stiroh, K. (2003). The performance of universal banks: Evidence from Switzerland. *Journal of Banking and Finance* 27, 2121–2150.
- Shi, S. (1993). Accurate and efficient double bootstrap confidence limit method. *Computational Statistics and Data Analysis* 13, 21–32.
- Simar, L., and Wilson, P.W. (2007). Estimation and Inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* 136, 31–64.
- Verweire, K. (1999). Performance consequences of financial conglomerates and empirical analysis in Belgium and The Netherlands. PhD Thesis, The Netherlands: Erasmus University.
- Xue, M., and Harker, P.T. (1999). Overcoming the inherent dependency of DEA efficiency scores: a bootstrap approach. Wharton Financial Institutions Center, University of Pennsylvania, Working Paper No. 99-17.

Table 1. Estimated coverages of confidence intervals generated by conventional and single bootstrap methods.

n	Basic Bootstrap Alg. Nominal significance				Percentile Bootstrap Alg. Nominal significance				Asymptotic Normal Apr. Nominal significance			
	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
p=q=1												
100	0.76	0.83	0.89	0.95	0.77	0.88	0.94	0.98	0.77	0.85	0.90	0.96
200	0.78	0.86	0.91	0.95	0.78	0.88	0.93	0.98	0.78	0.87	0.92	0.96
400	0.78	0.88	0.93	0.98	0.77	0.90	0.94	0.99	0.78	0.89	0.93	0.98
600	0.79	0.89	0.94	0.99	0.80	0.90	0.95	1.00	0.79	0.90	0.95	0.99
1200	0.79	0.88	0.94	0.98	0.80	0.89	0.94	0.99	0.79	0.89	0.94	0.98
3400	0.82	0.91	0.96	0.99	0.82	0.92	0.96	0.99	0.82	0.91	0.96	0.99
5000	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
10000	0.80	0.90	0.96	0.99	0.80	0.90	0.96	0.99	0.80	0.90	0.96	0.99
15000	0.81	0.91	0.95	0.99	0.81	0.91	0.95	0.99	0.81	0.91	0.95	0.99
p=q=2												
100	0.70	0.76	0.81	0.87	0.72	0.82	0.88	0.96	0.71	0.77	0.83	0.90
200	0.71	0.80	0.84	0.91	0.73	0.83	0.89	0.97	0.72	0.80	0.86	0.92
400	0.70	0.80	0.87	0.94	0.71	0.83	0.90	0.97	0.70	0.81	0.87	0.95
600	0.74	0.84	0.89	0.95	0.75	0.86	0.92	0.98	0.75	0.86	0.90	0.96
1200	0.71	0.83	0.88	0.96	0.72	0.85	0.90	0.97	0.71	0.84	0.90	0.97
3400	0.72	0.84	0.91	0.97	0.72	0.85	0.93	0.98	0.72	0.84	0.92	0.97
5000	0.73	0.82	0.90	0.96	0.73	0.84	0.91	0.97	0.73	0.83	0.90	0.96
10000	0.73	0.85	0.91	0.97	0.73	0.85	0.92	0.98	0.73	0.85	0.91	0.98
15000	0.77	0.87	0.93	0.98	0.77	0.88	0.93	0.98	0.77	0.88	0.93	0.98
p=q=3												
100	0.63	0.69	0.73	0.81	0.66	0.79	0.85	0.94	0.64	0.71	0.76	0.84
200	0.68	0.76	0.80	0.87	0.70	0.81	0.87	0.96	0.69	0.76	0.81	0.90
400	0.62	0.72	0.78	0.87	0.64	0.77	0.85	0.95	0.62	0.73	0.80	0.89
600	0.67	0.76	0.82	0.91	0.69	0.79	0.86	0.96	0.68	0.77	0.84	0.93
1200	0.57	0.69	0.79	0.89	0.59	0.73	0.83	0.95	0.58	0.71	0.81	0.92
3400	0.54	0.67	0.77	0.89	0.54	0.69	0.79	0.93	0.54	0.67	0.78	0.91
5000	0.52	0.65	0.75	0.88	0.53	0.66	0.77	0.90	0.52	0.66	0.76	0.89
10000	0.47	0.58	0.71	0.87	0.47	0.59	0.74	0.90	0.47	0.59	0.72	0.88
15000	0.47	0.60	0.70	0.86	0.46	0.60	0.71	0.89	0.46	0.60	0.70	0.87

Note: Results are based on 1,000 Monte Carlo trials.



Table 2. Widths of estimated confidence intervals generated by conventional and single bootstrap methods

n	Basic Bootstrap Alg. Mean width - Nominal sign.				Percentile Bootstrap Alg. Mean width - Nominal sign.				Asymptotic Normal Apr. Mean width - Nominal sign.			
	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
p=q=1												
100	0.266	0.347	0.422	0.587	0.266	0.347	0.422	0.587	0.270	0.347	0.414	0.544
200	0.184	0.238	0.286	0.384	0.184	0.238	0.286	0.384	0.185	0.237	0.283	0.372
400	0.128	0.165	0.198	0.263	0.128	0.165	0.198	0.263	0.128	0.165	0.196	0.258
600	0.104	0.134	0.160	0.213	0.104	0.134	0.160	0.213	0.104	0.134	0.159	0.210
1200	0.074	0.095	0.113	0.150	0.074	0.095	0.113	0.150	0.074	0.095	0.113	0.148
3400	0.043	0.056	0.066	0.088	0.043	0.056	0.066	0.088	0.043	0.056	0.066	0.087
5000	0.036	0.046	0.055	0.072	0.036	0.046	0.055	0.072	0.036	0.046	0.055	0.072
10000	0.025	0.032	0.038	0.051	0.025	0.032	0.038	0.051	0.025	0.032	0.038	0.050
15000	0.020	0.026	0.031	0.041	0.020	0.026	0.031	0.041	0.020	0.026	0.031	0.041
p=q=2												
100	0.326	0.437	0.549	0.864	0.326	0.437	0.549	0.864	0.340	0.436	0.520	0.683
200	0.211	0.274	0.331	0.452	0.211	0.274	0.331	0.452	0.212	0.272	0.325	0.427
400	0.137	0.177	0.212	0.284	0.137	0.177	0.212	0.284	0.136	0.175	0.209	0.275
600	0.112	0.144	0.172	0.228	0.112	0.144	0.172	0.228	0.111	0.142	0.170	0.223
1200	0.076	0.098	0.116	0.154	0.076	0.098	0.116	0.154	0.075	0.097	0.115	0.152
3400	0.044	0.057	0.068	0.089	0.044	0.057	0.068	0.089	0.044	0.056	0.067	0.088
5000	0.036	0.047	0.056	0.073	0.036	0.047	0.056	0.073	0.036	0.046	0.055	0.073
10000	0.025	0.032	0.038	0.051	0.025	0.032	0.038	0.051	0.025	0.032	0.038	0.051
15000	0.020	0.026	0.031	0.041	0.020	0.026	0.031	0.041	0.020	0.026	0.031	0.041
p=q=3												
100	0.467	0.669	0.925	1.985	0.467	0.669	0.925	1.985	0.580	0.745	0.888	1.167
200	0.268	0.353	0.433	0.624	0.268	0.353	0.433	0.624	0.271	0.348	0.415	0.546
400	0.164	0.212	0.254	0.343	0.164	0.212	0.254	0.343	0.163	0.209	0.249	0.327
600	0.129	0.166	0.198	0.265	0.129	0.166	0.198	0.265	0.127	0.164	0.195	0.256
1200	0.084	0.108	0.129	0.171	0.084	0.108	0.129	0.171	0.083	0.107	0.127	0.167
3400	0.047	0.060	0.072	0.095	0.047	0.060	0.072	0.095	0.046	0.060	0.071	0.093
5000	0.038	0.049	0.058	0.077	0.038	0.049	0.058	0.077	0.038	0.048	0.058	0.076
10000	0.026	0.033	0.040	0.053	0.026	0.033	0.040	0.053	0.026	0.033	0.040	0.052
15000	0.021	0.027	0.032	0.042	0.021	0.027	0.032	0.042	0.021	0.027	0.032	0.042

Table 3. Estimated coverages of confidence intervals generated by double bootstrap methods.

n	Basic Double Bootstrap Alg. Nominal significance				Percentile Double Bootstrap Alg. Nominal significance			
	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
	$\rho=q=1$							
100	0.77	0.88	0.94	0.98	0.78	0.90	0.95	0.99
200	0.79	0.89	0.95	0.98	0.79	0.90	0.95	0.99
400	0.79	0.91	0.95	0.99	0.80	0.91	0.96	0.99
	$\rho=q=2$							
100	0.74	0.84	0.91	0.94	0.75	0.86	0.93	0.99
200	0.75	0.84	0.91	0.95	0.76	0.86	0.92	0.99
400	0.74	0.86	0.92	0.98	0.74	0.86	0.93	0.99
	$\rho=q=3$							
100	0.72	0.80	0.87	0.88	0.71	0.83	0.90	0.98
200	0.72	0.83	0.91	0.93	0.73	0.85	0.92	0.99
400	0.68	0.80	0.88	0.95	0.69	0.80	0.90	0.99

Note: Results are based on 1,000 Monte Carlo trials

Table 4. Widths of estimated confidence intervals generated by double bootstrap methods.

n	Basic Double Bootstrap Alg. Nominal significance				Percentile Double Bootstrap Alg. Nominal significance			
	0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
	$\rho=q=1$							
100	0.281	0.363	0.429	0.479	0.292	0.392	0.498	0.819
200	0.190	0.244	0.303	0.356	0.192	0.251	0.307	0.475
400	0.131	0.168	0.203	0.264	0.131	0.170	0.206	0.312
	$\rho=q=2$							
100	0.305	0.443	0.504	0.608	0.405	0.603	1.124	2.009
200	0.218	0.283	0.345	0.389	0.224	0.295	0.367	0.582
400	0.141	0.181	0.220	0.277	0.141	0.184	0.223	0.341
	$\rho=q=3$							
100	0.532	0.686	0.847	1.641	0.891	1.772	4.490	6.425
200	0.280	0.358	0.410	0.452	0.300	0.409	0.540	1.040
400	0.168	0.216	0.268	0.313	0.170	0.223	0.273	0.423

Note: Results are based on 1,000 Monte Carlo trials

Table 5. Descriptive statistics

	No of Observations	Total Assets (Size) in €.th.		Income Diversity index	
		Mean	median	mean	median
AUSTRIA	92	4,446,463	409,997	0.44	0.43
BELGIUM	16	2,370,268	1,465,646	0.53	0.40
DENMARK	62	6,155,734	304,102	0.46	0.46
FINLAND	6	24,227,542	9,832,681	0.53	0.51
FRANCE	113	4,615,236	1,538,488	0.59	0.64
GERMANY	588	6,484,665	1,142,362	0.42	0.39
GREECE	10	11,273,294	2,625,859	0.48	0.46
IRELAND	8	27,835,580	3,237,596	0.31	0.31
ITALY	68	1,969,861	725,086	0.54	0.56
LUXEMBOURG	52	6,797,045	1,418,101	0.60	0.64
NETHERLANDS	16	53,875,885	1,926,433	0.44	0.44
NORWAY	40	3,721,192	365,842	0.40	0.38
PORTUGAL	7	21,780,662	6,609,877	0.55	0.60
SPAIN	38	29,077,412	5,758,037	0.49	0.47
SWEDEN	78	2,103,156	101,215	0.43	0.43
SWITZERLAND	303	4,814,006	188,708	0.26	0.20
UNITED KINGDOM	60	41,829,887	1,295,279	0.57	0.58
Commercial Banks	649	16,249,536	807,461	0.51	0.50
Savings Banks	908	2,187,380	644,445	0.36	0.37

Table 6. Truncated regression analysis of European bank efficiency estimates (95% Confidence Intervals,  $p=q=3$ ).

	$\hat{\beta}$	Asymptotic CI		Basic Bootstrap CI		Percentile Bootstrap CI		Basic Double Bootstrap CI		Percentile Double Bootstrap CI	
		low	high	low	high	low	high	low	high	low	High
Constant	13.171	13.164	13.177	11.704	14.342	11.999	14.637	11.499	14.404	11.938	14.842
DIVIN	17.591	17.463	17.720	12.453	24.459	10.724	22.729	12.394	27.756	7.427	22.788
DIVIN <sup>2</sup>	-8.316	-9.250	-7.382	-10.596	-5.849	-10.783	-6.036	-10.709	-5.614	-11.018	-5.923
lnAST	-0.155	-0.155	-0.154	-0.336	0.137	-0.447	0.0264	-0.338	0.305	-0.616	0.028
DIVIN*lnAST	-1.047	-1.053	-1.041	-1.554	-0.650	-1.444	-0.540	-1.673	-0.629	-1.465	-0.421
EQAST	0.030	0.028	0.031	0.017	0.044	0.015	0.043	0.018	0.045	0.015	0.042
ROAE	0.0006	0.00005	0.0011	-0.0049	0.0065	-0.0053	0.0059	-0.0048	0.0053	-0.0041	0.0060
MS	-13.127	-13.315	-12.939	-26.056	-12.494	-13.761	-0.198	-28.392	-12.776	-13.479	2.137
LLP	-1.175	-1.176	-1.174	-1.479	-1.146	-1.204	-0.871	-1.561	-1.148	-1.202	-0.789
COMM	-0.334	-0.343	-0.324	-0.600	-0.076	-0.592	-0.067	-0.613	-0.080	-0.587	-0.054
$\sigma_u^2$	1.714	1.713	1.714	1.650	1.805	1.622	1.777	1.649	1.814	1.613	1.778

Notes: (i) The dependent variable is the DEA estimate of the unobserved inefficiency score of firm  $j$ . (ii) The technology describing the production process is defined in a 3-input 3-output space. (iii) The regression also includes country dummy variables which are not reported. (iv) The independent variables are defined as follows: DIVIN is a measure of diversification calculated as the  $1 - |(\text{Net Interest Income} - \text{Other Operating Income}) / \text{Total Operating Income}|$ . lnAST is the logarithm of total assets. EQAST is the ratio of equity over total assets. ROAE is the ratio of return on average equity. MS is the market share of deposits. LLP is a measure of loan portfolio quality measured as the Loan Loss Provisions over Total Customer Loans. COMM is a dummy variable that equals 1 if a bank is commercial and zero otherwise. (v) The basic bootstrap confidence interval is the Simar and Wilson (2007)'s suggested confidence interval.