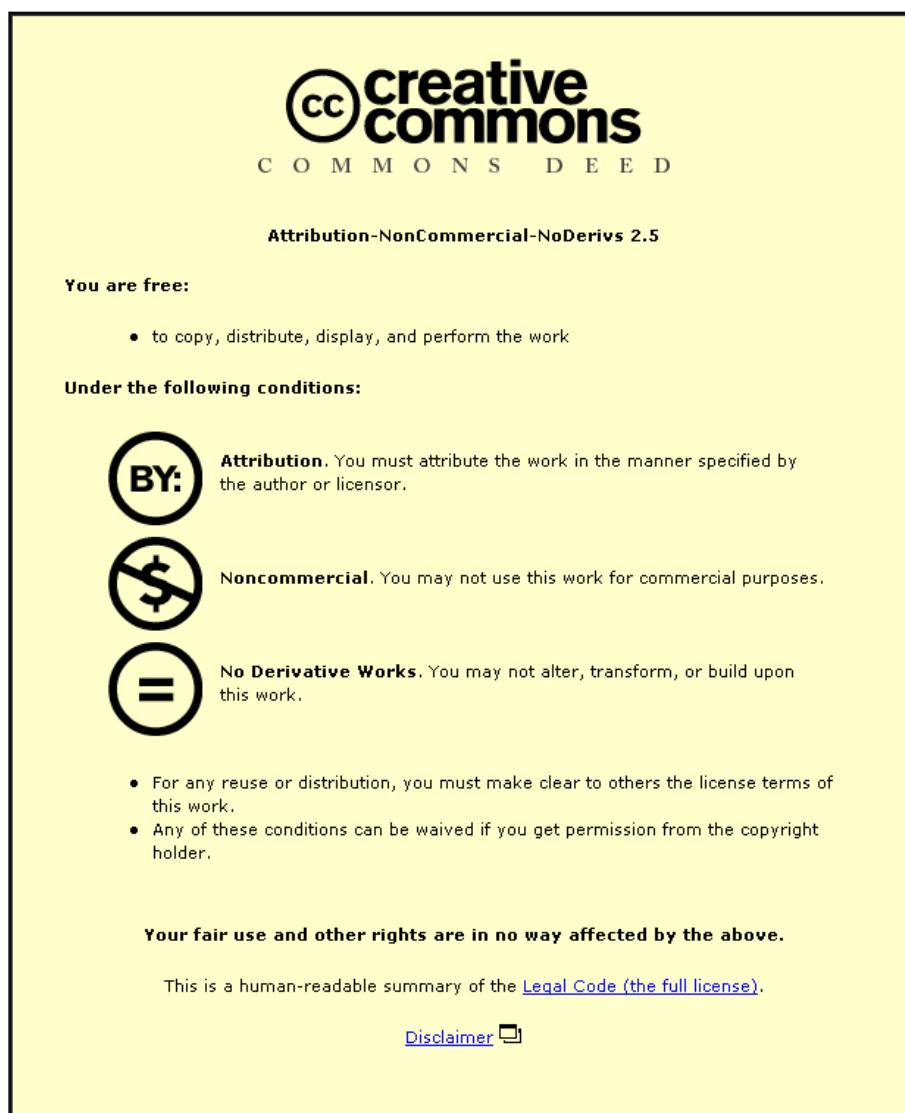




This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Resampling DEA estimates of investment fund performance

John D. Lamb

University of Aberdeen Business School, Aberdeen AB24 3QY

Kai-Hong Tee

Loughborough University Business School, Loughborough LE11 3TU

Data envelopment analysis (DEA) is attractive for comparing investment funds because it handles different characteristics of fund distribution and gives a way to rank funds. There is substantial literature applying DEA to funds, based on the time series of funds' returns. This article looks at the issue of uncertainty in the resulting DEA efficiency estimates, investigating consistency and bias. It uses the bootstrap to develop stochastic DEA models for funds, derive confidence intervals and develop techniques to compare and rank funds and represent the ranking. It investigates how to deal with autocorrelation in the time series and considers models that deal with correlation in the funds' returns.

DATA ENVELOPMENT ANALYSIS, BOOTSTRAP, INVESTMENT FUND, RANK, BIAS

1 Introduction

Consider an investor or fund manager who has data on the returns of n investment funds over T recent time periods and wishes to know if fund o is likely to perform well in future. They have a performance measure ϕ_o . It is not known exactly and can only be estimated from the data as $\phi_{o,T}$. We consider here the case where ϕ_o is a (population) efficiency from a data envelopment analysis (DEA) model whose decision-making units are the n investment funds. But, in general, ϕ_o can be another performance measure such as the Sharpe, Calmar or Sortino (Lhabitant, 2004) ratio. And the methods we describe apply whenever DEA efficiencies are estimated from a time series of observed data.

Our purpose is to investigate the following. First, is $\phi_{o,T}$ a consistent estimator of ϕ_o ? If it is, is it biased? And if it is biased, how can we estimate the bias? Second, how can we obtain a confidence interval for each ϕ_o ? Third, how can we compare funds statistically and rank them according to their performance? We investigate these assuming initially no autocorrelation in the time series and later allowing for autocorrelation. We test the methods we develop on both standard and diversification-consistent (Lamb and Tee, 2011) DEA

models, which are arguably more appropriate for investment funds.

Investment funds (especially hedge funds) are often characterised by strong skewness and kurtosis (Brooks and Kat, 2002) and not just mean and variance. These makes DEA attractive because it can handle multiple properties of the fund distributions and provide a means to rank them. Ranking is important because it helps identify funds that perform well and so, for example, helps form a fund of hedge funds portfolio efficiently. However, the accuracy of the ranking requires the investigation of the statistical properties of the DEA models, the main purpose of the paper.

Various studies (e.g. Murthi et al. (1997); Gregoriou and Zhu (2005); Chen and Lin (2006); Lozano and Gutiérrez (2008a); Kumar et al. (2010)) apply deterministic DEA to the problem of measuring the efficiency of investment funds. Studies have considered both choice of model (e.g. Gregoriou and Zhu (2005); Lozano and Gutiérrez (2008a,b); Briec and Kerstens (2010)) and choice of measure (e.g. Eling (2006)). Our findings are valid for a wide range of models and measures and so we do not discuss these in detail. Lamb and Tee (2011) discuss what models and risk (input) and return (output)

measures are appropriate for modelling funds.

As far as we know, stochastic DEA has not been applied before to modelling funds. However, several methods for stochastic DEA are developed in other contexts. In particular we note the methods developed by [Simar and Wilson \(1998\)](#); [Löthgren and Tambour \(1999a,b\)](#); [Löthgren \(2000\)](#); [Kuosmanen and Johnson \(2010\)](#), which have been applied to modelling Italian banking data ([Ferrier and Hirschberg, 1997](#)) and Spanish public services ([González and Miles, 2002](#)). These methods use the bootstrap ([Efron and Tibshirani, 1998](#)) to deal with uncertainty, and [Souza et al. \(2011\)](#) use a similar method for confidence intervals. These methods assume only one observation of each input and output for each decision making unit (DMU), which means the bootstrap data generating processes cannot include every DMU in each replication. By contrast, investment fund data has many observations for each fund (or DMU) and so we can use a more conventional bootstrap method. We know of one example of stochastic DEA with multiple observations for each DMU: [Kao and Liu \(2009\)](#) model Taiwanese banks with five values for each input and output but do not use the bootstrap. [Dyson and Shale \(2010\)](#) discuss more generally the problems that arise in stochastic DEA and how to deal with them.

Investment funds are typically ranked (see, for example, [Morningstar \(2011\)](#)). Ranking is straightforward if we know the values of the performance measure exactly, as in deterministic DEA with precisely known inputs and outputs. In stochastic DEA ranking must be reassessed. The exact rank of a fund cannot be known for sure and may change from one resample to the next. In this case, ranking will be based on the average value of the efficiency scores out of the entire re-sampling. However, using the average value could bias the accuracy of the rank position, an issue which we will address in the paper. Still, the effect of biases could complicate the position ranking of two funds, especially when the average values of their scores are very similar. Therefore, an alternative approach we adopt in this paper is to use partial ranking to ascertain the relative positions of funds. Partial ranking ensures that one fund is ranked above another whenever its performance is statistically significantly better. Partial ranking is a partial ordering \preceq that we can best show in figures like Fig. 2 as a Hasse diagram ([Rutherford, 1965](#)). A Hasse diagram shows an

arrow from i to j if (a) $\phi_i \preceq \phi_j$ and (b) there is no k with $\phi_i \preceq \phi_k$ and $\phi_k \preceq \phi_j$.

Section 2 presents basic DEA models and shows that $\phi_{o,T}$ is, in general, a consistent estimator of ϕ_o . Section 3 describes a bootstrap data generating process and how we can use it to produce an unbiased estimator of ϕ_o . It then shows how we can find bootstrap confidence intervals for DEA efficiency and how we can compare and rank funds statistically. Section 4 discusses how to deal with autocorrelation. Section 5 applies the methods to DEA models that allow the possibility of risk reduction through diversification. Furthermore, Sections 3, 4 and 5 discuss fund ranking and related issues in the resampling process. And Section 6 presents some discussion and conclusions.

2 Consistency of DEA efficiency estimates

A sequence of estimators θ_T ($T = 1, 2, \dots$) for a parameter θ is *consistent* if θ_T converges in probability to θ as $T \rightarrow \infty$. Clearly we want DEA estimates to be consistent. We show they are when T is the number of observations for each fund. Contrast this with [Banker \(1993\)](#), whose result is of the form $\phi_{o,n} \rightarrow \phi_o$ in probability as $n \rightarrow \infty$, where n is the number of DMUs.

We now describe the main DEA models we use. Section 3 discusses the specific measures we use. We assume we have investment funds f_1, \dots, f_n whose returns for time periods $1, \dots, T$ are given as r_{jt} ($j = 1, \dots, n, t = 1, \dots, T$). We assume we have nonnegative risk measures x_1, \dots, x_m and nonnegative return measures y_1, \dots, y_r that we can estimate for any fund. Let $x_{ij} = x_i(f_j)$ ($i = 1, \dots, m, j = 1, \dots, n$) and let $y_{rj} = y_r(f_j)$ ($r = 1, \dots, s, j = 1, \dots, n$). We use the risk and return measures to estimate an efficiency $\phi_{o,T}$ for each fund f_o ($o = 1, \dots, n$).

We use the following input-oriented model. Define $\phi_{o,T}$ ($o = 1, \dots, n$) as follows. If $(x_{1o}, \dots, x_{mo}) \leq \mathbf{0}$ put $\phi_{o,T} = 1$. Otherwise, choose $\phi_{o,T}, \lambda_1, \dots, \lambda_n$ to minimise $\phi_{o,T}$ subject to

$$\sum_{j=1}^n y_{rj} \lambda_j \geq y_{ro}, \quad (r = 1, \dots, s) \quad (1)$$

$$\sum_{j=1}^n x_{ij} \lambda_j \leq x_{io} \phi_{o,T}, \quad (i = 1, \dots, m) \quad (2)$$

$$\sum_{j=1}^n \lambda_j \leq 1, \quad (3)$$

and

$$\lambda_j \geq 0 \quad (j = 1, \dots, n) \quad (4)$$

This is a conventional DEA model in dual form (Gregoriou and Zhu, 2005) with two modifications, explained in detail in Lamb and Tee (2011). First, constraint (3) creates a nonincreasing returns to scale (NIRS) (Cooper et al., 2007; Färe et al., 1994) model. This relaxes the constraint of the more usual variable returns to scale model by allowing some proportion of an investment to be invested in no fund while avoiding the impossible sets of risks and returns that a constant returns to scale model permits. Second, the model allows the risk and return measures to take zero values without giving implausible efficiency scores.

We also use an output-oriented NIRS model that gives an efficiency $\phi'_{o,T}$ in the range $[0, 1]$. It is the model of A simplified by setting $\hat{n} = 0$.

We denote by ϕ_o and ϕ'_o the population efficiencies we would get if we could replace the risk and return measures with the corresponding population statistics.

The following results help establish when $\phi_{o,T}$ and $\phi'_{o,T}$ are consistent. Let X be the $m \times n$ matrix whose (i, j) th entry is x_{ij} . Let Y be the $s \times n$ matrix whose (r, j) th entry is y_{rj} . Call a function $f(X, Y)$ p -continuous if it is continuous with respect to perturbations in X, Y that leave elements that take the value zero unchanged.

PROPOSITION 1 *The efficiencies ϕ_o and ϕ'_o are p -continuous for almost every value of X and Y .*

Proof. Note that if we add a fund all of whose risk and return measures are zero to either model, we get a variable returns to scale model. Then the result for ϕ_o follows *a fortiori* from Scheel and Scholtes (2003, Proposition 3). The result of Scheel and Scholtes (2003) requires a linear programme with a bounded feasible set and so cannot be used directly for ϕ'_o . But if $(y_{1o}, \dots, y_{so}) = \mathbf{0}$ then ϕ'_o is p -continuous with probability 1 by definition. Otherwise, define ψ_{ro} ($r = 1, \dots, s$) by the linear programme: choose $\lambda_1, \dots, \lambda_n$ to

maximise

$$\psi_{ro} = \sum_{j=1}^n \lambda_j y_{rj} \quad (5)$$

subject to constraints (9) (with $\hat{n} = 0$) and (4). (Banker (1993) uses this programme to estimate frontier points when $s = 1$.) Put $\eta_{ro} = y_{ro}/\psi_{ro}$ if $\psi_{ro} > 0$ and 0 otherwise. Then $\eta_o = \max_{r=1, \dots, s} \eta_{ro}$ is the solution to the output-oriented NIRS model. The linear programme has a bounded feasible set and so (see Scheel and Scholtes (2003); Robinson (1977)) each ψ_{ro} is continuous at X and Y . Since the functions \max and η_{ro} are p -continuous almost everywhere it follows that $\phi'_o = 1/\eta_o$ is also p -continuous almost everywhere. ■

The following well-known result of Mann and Wald (1943) is called the continuous mapping theorem.

PROPOSITION 2 *Let $g : S \rightarrow T$ be continuous almost everywhere and let \mathbf{w}_n be a sequence of random vectors on S such that \mathbf{w}_n converges in probability to \mathbf{w} . Then $g(\mathbf{w}_n)$ converges in probability to $g(\mathbf{w})$.* ■

PROPOSITION 3 *Suppose $x_1, \dots, x_m, y_1, \dots, y_s$ are consistent measures. Suppose that each estimate of $x_i(f)$ ($i = 1, \dots, m$) is either always zero or zero with probability 0 and that each estimate of $y_r(f)$ ($r = 1, \dots, s$) is either always zero or zero with probability 0. Then $\phi_{o,T}$ and $\phi'_{o,T}$ are consistent estimators of ϕ_o and ϕ'_o for $o = 1, \dots, n$.*

Proof. We can consider ϕ_o and ϕ'_o as functions of the elements of X, Y that are not always zero. Then they are continuous almost everywhere by Proposition 1 and so

$\phi_{o,T}$ and $\phi'_{o,T}$ are consistent estimators by Proposition 2. ■

Although Proposition 3 is proved for NIRS models, it is straightforward to check that it holds also for VRS models.

3 Bootstrap, bias, confidence intervals and comparisons

We have shown $\phi_{o,T}$ is consistent but wish to know more about its properties. The bootstrap is a method for investigating the properties of a sample statistic (for example, estimating bias and standard error) such as $\phi_{o,T}$ when the distribution of the sample statistic is unknown. It is particularly useful when parametric methods are inappropriate. Clearly the distribution of $\phi_{o,T}$ is bounded below by zero and above by one and plausibly may take either with nonzero probability. We know of no parametric method that can be used in such circumstances.

3.1 Bootstrap data generating process

The bootstrap needs a data generating process (Efron and Tibshirani, 1998). Our data generating process is as follows. We generate a number of replications. These are data sets of the form $r_{it}^* = r_{i\rho(t)}$ ($i = 1, \dots, n$, $t = 1, \dots, T$) where $\rho(1), \dots, \rho(t)$ are generated uniformly at random with replacement from $\{1, \dots, T\}$. This is a standard approach for generating data sets of vectors and preserves correlations between returns of different funds. We use 2000 replications in each test we carry out because Efron and Tibshirani (1998) indicate this is a reasonable number for each bootstrap method we use. We estimate the DEA efficiencies for each replication and use them in various bootstrap procedures.

3.2 Specific data and measures

We use data comprising 60 monthly returns from 30 hedge funds for the period 2000–2004. The funds come from Center for International Securities and Derivatives Markets (2010) and are broadly classified as market neutral (MN), long/short strategy (LS) and global macro (GM).

Although we may use any risk or return measures, we

illustrate our method with measures that satisfy a property called *convexity consistency* (Lamb and Tee, 2011) because we need such measures in Section 5. Convexity consistency is closely related to the well-known notion of a *coherent measure of risk* (Artzner et al., 1999; Rockafellar and Uryasev, 2002). We use measures base on two particular coherent measures of risk. The first is conditional value at risk: cvar_α is the expected loss on an investment conditional on that loss not exceeding var_α , where var_α , the value at risk at α , is the negative of the $100\alpha\%$ percentile of the distribution of returns. Both cvar and var are widely used in the finance literature (Jorion, 2007). The second measure is the lower semideviation minus the mean (Ogryczak and Ruszczyński, 1999). We use the same four specific risk and return measures for every test we carry out. The risk measures are $\max(\text{cvar}_{0.2}, 0)$ and $\max(\text{sd}, 0)$, where sd is the lower semideviation minus the mean. The return measures are the mean return and $\max(-\text{cvar}_{0.1}, 0)$.

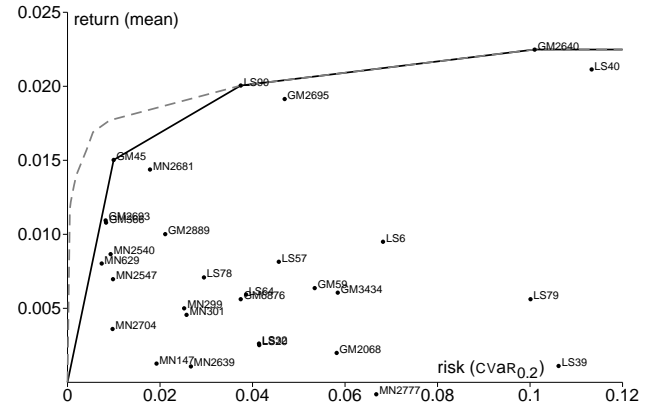


Figure 1: Efficient frontiers of simplified NIRS and diversification-consistent NIRS models

Fig. 1 plots mean against $\max(\text{cvar}_{0.2}, 0)$ for each fund and the solid line shows the efficient frontier from the simplified NIRS DEA model that uses only these two measures. Note that the frontier does not depend on whether the model is input-oriented or output-oriented.

3.3 Bias

We first estimate the bias and derive bias-corrected estimates of the DEA efficiencies. We use the bootstrap estimate of bias of Efron and Tibshirani (1998, Sec-

tion 10.2). Although there are better bootstrap methods to estimate bias (Efron, 1990), we cannot adapt them for DEA because the efficiencies are given by linear programmes rather than an explicit formula.

Columns 2–4 of Table 1 show, for each fund, the DEA efficiency (labelled ϕ_T , raw efficiency for short) bootstrap bias-corrected DEA efficiency (labelled $\phi_{bc,T}$) and bootstrap standard error (labelled SE) for the input-oriented model. The bootstrap estimate of bias can be found as

$$\widehat{\text{bias}} = \phi_T - \phi_{bc,T}. \quad (6)$$

We estimate also the bootstrap standard error using Efron and Tibshirani (1998, Algorithm 6.1). The table is ordered by bias-corrected efficiency and we note that the ordering is noticeably different from that given by the raw efficiencies.

The bootstrap bias we observe is invariably less than the standard error, but is large enough to suggest it may be better to use bias-corrected efficiency estimates. The bias is negative for funds that have efficiency close to 1 though some funds have positive bias. This is expected. The basic DEA model must produce some funds with efficiency 1 but it is unlikely that these funds will achieve efficiency 1 in every replication.

Section 5 discusses columns 5–7 of Table 1.

3.4 Confidence intervals

There are several bootstrap methods for confidence intervals (Efron and Tibshirani, 1998). We cannot use the parametric bootstrap because it assumes we know the form of the distribution of the DEA efficiencies. So we seek a nonparametric bootstrap method. Suppose a method gives a $1 - 2\alpha$ confidence interval $(\phi_\alpha, \phi_{1-\alpha})$ satisfying

$$\begin{aligned} P(\phi < \phi_\alpha) - \alpha &= O(T^{-r}) \quad \text{and} \\ P(\phi > \phi_{1-\alpha}) - \alpha &= O(T^{-r}), \end{aligned} \quad (7)$$

for some r , where T is the sample size and O is the Landau symbol. The method is *first-order accurate* if it satisfies Eq. (7) with $r = 1$ and *second-order accurate* if it satisfies it with $r = 2$. Ideally we want a second-order accurate method. The bias-corrected accelerated (BC_a)

method (Efron and Tibshirani, 1998) is second-order accurate depending on the assumption (Efron, 1987) that there is a strictly increasing function g and constants z_0 and a such that

$$g(\phi_T) = g(\phi) + (1 + ag(\phi))(Z - z_0), \quad Z \sim N(0, 1), \quad (8)$$

where ϕ is the population efficiency and ϕ_T is sample efficiency. However, $\phi_{o,T}$ is bounded below by zero and above by one, and we find that in practice it may take either value with positive probability. So the assumption of a strictly-increasing g satisfying Eq. (8) is not plausibly even approximately true for DEA efficiency estimates.

The bootstrap percentile interval (Efron and Tibshirani, 1998) is first order accurate and does not depend on assumptions about the distribution of efficiency estimates, and so we report percentile confidence intervals for the efficiencies. In practice, we find the percentile and BC_a intervals are very similar.

Table 2 presents the funds in the same order as Table 1. Columns 2–3 show the 95% percentile bootstrap confidence intervals for each fund using the input-oriented model. Columns 2–3 of Table 3 show the 95% percentile bootstrap confidence intervals using the output-oriented model. Sections 4 and 5 discuss the remaining columns of both tables.

3.5 Comparisons

The confidence intervals of Table 2 are wide. Nonetheless, we may reasonably hope to observe significant differences between many pairs of funds if we can use a matched pairs test. Such a test is possible because we have matching observations for each fund in each time period. We can test individual differences in efficiency between funds using a one-sample bootstrap hypothesis test (Efron and Tibshirani, 1998, Chapter 16) based on the difference between efficiencies. As in Section 3.4, we use bootstrap percentiles for the differences in efficiency to construct individual hypothesis tests.

As usual, when we carry out multiple matched-pairs tests, some tests may appear significant by chance alone (our example has 435 tests). The usual adjustments are inappropriate because a transitive comparison rela-

FUND	<i>basic</i>			<i>diversification consistent</i>		
	ϕ_T	$\phi_{bc,T}$	SE	ϕ_T	$\phi_{bc,T}$	SE
GM45	1.000	0.931	0.195	0.391	0.196	0.334
GM2693	0.881	0.727	0.313	0.031	0.052	0.130
GM366	0.860	0.695	0.306	0.030	0.046	0.112
LS90	1.000	0.659	0.502	0.773	0.442	0.519
MN629	0.720	0.584	0.322	0.025	0.028	0.063
MN2681	0.535	0.564	0.325	0.173	0.106	0.200
GM2695	0.691	0.520	0.406	0.456	0.331	0.394
GM2640	1.000	0.492	0.662	1.000	0.440	0.712
MN2540	0.618	0.491	0.305	0.022	0.024	0.055
LS40	0.582	0.416	0.347	0.362	0.430	0.411
MN2547	0.470	0.365	0.234	0.017	0.020	0.044
GM2889	0.315	0.305	0.254	0.011	0.027	0.085
MN2704	0.245	0.180	0.133	0.009	0.010	0.032
LS6	0.092	0.131	0.194	0.003	0.049	0.156
LS78	0.160	0.125	0.112	0.006	0.008	0.021
MN299	0.131	0.107	0.104	0.005	0.007	0.033
LS57	0.118	0.105	0.113	0.004	0.011	0.046
LS39	0.007	0.101	0.261	0.000	0.069	0.239
GM3434	0.069	0.098	0.174	0.002	0.029	0.123
LS79	0.037	0.098	0.216	0.001	0.050	0.185
MN301	0.117	0.097	0.099	0.004	0.007	0.040
LS64	0.102	0.097	0.133	0.004	0.014	0.071
GM6876	0.100	0.094	0.107	0.004	0.007	0.031
GM59	0.079	0.086	0.148	0.003	0.022	0.114
MN2777	0.000	0.083	0.253	0.000	0.062	0.237
LS20	0.040	0.073	0.169	0.001	0.029	0.149
LS32	0.042	0.070	0.150	0.001	0.025	0.129
MN2639	0.027	0.052	0.139	0.001	0.021	0.134
MNI47	0.043	0.050	0.002	0.010	0.097	0.085
GM2068	0.023	0.041	0.092	0.001	0.007	0.068

Table 1: DEA efficiencies and bootstrap bias-corrected values

tion (see Section 1, Shaffer (1986)) implies the tests are not all independent of each other. Shaffer (1986) describes a procedure for such a situation. It modifies for matched pairs the well-known Bonferonni correction that estimates experimentwise significances for individual comparisons from a familywise significance level. We modify the procedure slightly by using the more accurate Šidák (1967) equation rather than the Bonferonni approximation for the experimentwise significance level. We also consider in Section 5 a version of Shaffer's procedure, which is called *least significant differences*

(LSD). It uses a single significance level and identifies all individually significant differences that do not violate the transitivity assumption. In all the comparisons we use a familywise significance level of 5%.

Fig. 2 (first diagram) should be compared with Table 1. It shows as a Hasse diagram (See Section 1) the ranking of the funds in the input-oriented model using Shaffer's procedure. The same comparison procedure with the output-oriented model shows far fewer significant differences.

FUND	STANDARD BOOTSTRAP				BLOCK (SIZE 3) BOOTSTRAP			
	basic		DC		basic		DC	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
GM45	0.058	1.000	0.000	1.000	0.000	1.000	0.000	1.000
GM2693	0.000	1.000	0.000	0.442	0.000	1.000	0.000	0.449
GM366	0.000	1.000	0.000	0.391	0.000	1.000	0.000	0.395
LS90	0.000	1.000	0.000	1.000	0.018	1.000	0.000	1.000
MN629	0.177	1.000	0.000	0.198	0.175	1.000	0.000	0.206
MN2681	0.177	1.000	0.000	0.712	0.157	1.000	0.000	1.000
GM2695	0.079	1.000	0.000	1.000	0.097	1.000	0.000	1.000
GM2640	0.019	1.000	0.000	1.000	0.004	1.000	0.000	1.000
MN2540	0.183	1.000	0.000	0.181	0.145	1.000	0.000	0.170
LS40	0.019	1.000	0.000	1.000	0.023	1.000	0.000	1.000
MN2547	0.142	1.000	0.000	0.143	0.137	1.000	0.000	0.166
GM2889	0.091	1.000	0.000	0.209	0.083	1.000	0.000	0.202
MN2704	0.052	1.000	0.000	0.079	0.029	1.000	0.000	0.076
LS6	0.034	1.000	0.000	0.480	0.027	1.000	0.000	0.276
LS78	0.031	1.000	0.000	0.059	0.014	1.000	0.000	0.049
MN299	0.019	1.000	0.000	0.055	0.009	0.668	0.000	0.038
LS57	0.029	1.000	0.000	0.079	0.022	1.000	0.000	0.048
LS39	0.024	1.000	0.000	1.000	0.019	1.000	0.000	1.000
GM3434	0.019	1.000	0.000	0.342	0.011	1.000	0.000	0.227
LS79	0.022	1.000	0.000	1.000	0.020	1.000	0.000	0.575
MN301	0.012	1.000	0.000	0.048	0.006	1.000	0.000	0.036
LS64	0.019	1.000	0.000	0.113	0.000	0.892	0.000	0.027
GM6876	0.016	1.000	0.000	0.051	0.020	1.000	0.000	0.100
GM59	0.021	1.000	0.000	0.195	0.000	1.000	0.000	0.037
MN2777	0.033	1.000	0.000	1.000	0.015	1.000	0.000	0.907
LS20	0.011	1.000	0.000	0.557	0.021	1.000	0.000	0.829
LS32	0.008	1.000	0.000	0.385	0.015	0.989	0.000	0.713
MN2639	0.003	1.000	0.000	0.053	0.004	1.000	0.000	0.042
MN147	0.000	1.000	0.000	0.032	0.000	1.000	0.000	0.033
GM2068	0.000	1.000	0.000	0.024	0.000	1.000	0.000	0.032

Table 2: Bootstrap percentile confidence intervals for basic and diversification-consistent (DC) input-oriented models

4 Dealing with autocorrelation

The standard bootstrap assumes that we have a set of *independent identically distributed* (iid) observations. The observations may be vectors. But they should not form a time series with trend or autocorrelation.

We may assume that returns from investment funds are available as a time series of vectors $\mathbf{r}_t = (r_{1t}, \dots, r_{nt})$

($t = 1, \dots, T$). In general we might reasonably expect this time series to be stationary but to show some autocorrelation. Exploratory data analysis in the form of sequence plots and Ljung–Box (Ljung and Box, 1978) tests suggests that our data have these properties. Several bootstrap methods are available (Bühlmann, 2002) for stationary vector time series. These deal with autocorrelation either by resampling blocks of data or by resampling the (assumed iid) residuals of a model of the data. We consider three: the block bootstrap (Kün-

model. We note that the method is problematic for the data we have. The matrices A_1, \dots, A_k are sparse for both $k = 2$ and $k = 3$. And we have no theoretical grounds for supposing a particular structure for them.

We note that the ranking of funds is broadly similar to the original ranking for all three methods we consider and suggest the block bootstrap with small block size is appropriate when there is reason to doubt the assumptions needed for the other methods.

5 Models that deal with diversification

Consider Fig. 1. The solid line shows the standard DEA frontier. The grey dashed line shows another frontier. It shows the maximum return that an investor wealthy enough to hold a portfolio of hedge funds could have achieved at any given risk level. The reason these frontiers are different is that standard DEA does not allow for the very real possibility of reducing risk by diversification: that is, choosing a portfolio of imperfectly correlated investments. Several DEA models allow for diversification (Morey and Morey, 1999; Joro and Na, 2006; Lozano and Gutiérrez, 2008a,b; Briec and Kerstens, 2009, 2010). Most use nonlinear programmes. Lamb and Tee (2011) discuss them and introduce the *diversification-consistent model*, which we use here because it uses a linear programme and does not prescribe specific risk and return measures.

Lamb and Tee (2011) describe the input-oriented diversification-consistent model, and A describes the output-oriented diversification-consistent (NIRS) model and the iterative procedure we use to compute efficiencies using the models.

Columns 5–7 of Table 1 show the raw and bootstrap bias-corrected efficiency estimates and bootstrap standard errors for our funds with the input-oriented diversification-consistent model. As before, we observe substantial bias in many raw efficiency estimates. Columns 4–5 of Table 2 show the 95% bootstrap percentile confidence intervals for our funds with the input-oriented diversification-consistent model. Columns 8–9 show the corresponding intervals with block (size 3) bootstrap. The main features to note are that both upper and lower bounds are invariably reduced rela-

tive to the corresponding confidence interval from the standard DEA model, and the lower bounds are zero in every case in our example. Columns 4–5 of Table 3 show the 95% bootstrap percentile confidence intervals for our funds with the output-oriented diversification-consistent model, and columns 8–9 show the corresponding intervals with block (size 3) bootstrap. Again, the upper and lower bounds are invariably reduced relative to those from the corresponding standard model. But the reduction is much less and very few lower bounds are zero.

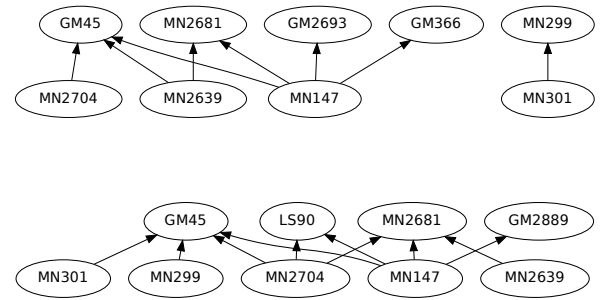


Figure 3: Output-oriented diversification-consistent DEA efficiency rankings using standard (top) and block size 3 (bottom) bootstrap, and Shaffer's procedure

The Shaffer procedure that is applied to our data in the input-oriented diversification-consistent model shows no significant differences, though there are a small number of significant differences with LSD. However, we observe some significant differences in the output-oriented diversification-consistent model. Fig. 3 shows as a Hasse diagrams the rankings of the funds found by Shaffer's procedure, excluding the 21 funds that are not significantly different from any other fund. The top diagram shows the standard bootstrap ranking and the bottom the block (size 3) ranking. Fig. 4 shows least significant differences: that is, it shows the pairs of funds that are experimentwise significantly different. Four funds are not significantly different in LSD from any other.

We noted that the diversification-consistent models give smaller values for both upper and lower confidence interval bounds. We also noted the input-oriented diversification-consistent model produces fewer significant differences than the diversification-consistent

FUND	STANDARD BOOTSTRAP				BLOCK (SIZE 3) BOOTSTRAP			
	basic		DC		basic		DC	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
GM45	0.647	1.000	0.460	1.000	0.578	1.000	0.366	1.000
GM2693	0.460	1.000	0.346	0.861	0.453	1.000	0.336	0.861
GM366	0.454	1.000	0.339	0.843	0.454	1.000	0.334	0.848
LS90	0.439	1.000	0.418	1.000	0.496	1.000	0.444	1.000
MN629	0.340	1.000	0.273	0.643	0.316	1.000	0.243	0.618
MN2681	0.458	1.000	0.382	0.977	0.462	1.000	0.373	1.000
GM2695	0.379	1.000	0.360	1.000	0.363	1.000	0.356	1.000
GM2640	0.503	1.000	0.503	1.000	0.307	1.000	0.307	1.000
MN2540	0.344	1.000	0.260	0.720	0.291	1.000	0.215	0.721
LS40	0.489	1.000	0.489	1.000	0.377	1.000	0.377	1.000
MN2547	0.269	0.972	0.207	0.565	0.268	0.889	0.184	0.559
GM2889	0.282	1.000	0.260	0.860	0.267	1.000	0.243	0.825
MN2704	0.117	0.536	0.087	0.371	0.074	0.516	0.053	0.366
LS6	0.111	0.979	0.109	0.944	0.098	0.876	0.098	0.845
LS78	0.140	0.754	0.132	0.650	0.093	0.699	0.088	0.612
MN299	0.095	0.733	0.085	0.649	0.100	0.557	0.088	0.505
LS57	0.140	0.767	0.135	0.711	0.170	0.704	0.160	0.665
LS39	0.000	0.776	0.000	0.773	0.000	0.662	0.000	0.659
GM3434	0.056	0.897	0.055	0.870	0.066	0.900	0.066	0.884
LS79	0.091	1.000	0.091	1.000	0.077	1.000	0.077	1.000
MN301	0.081	0.710	0.072	0.628	0.086	0.533	0.076	0.487
LS64	0.075	0.833	0.073	0.749	0.084	0.616	0.078	0.562
GM6876	0.125	0.767	0.118	0.672	0.089	0.838	0.087	0.752
GM59	0.095	0.887	0.090	0.848	0.109	0.674	0.102	0.661
MN2777	0.000	0.472	0.000	0.436	0.000	0.449	0.000	0.419
LS20	0.000	0.598	0.000	0.527	0.000	0.466	0.000	0.410
LS32	0.000	0.603	0.000	0.530	0.000	0.469	0.000	0.414
MN2639	0.010	0.434	0.009	0.394	0.021	0.437	0.018	0.407
MN147	0.023	0.353	0.017	0.325	0.015	0.296	0.012	0.273
GM2068	0.058	0.597	0.058	0.536	0.059	0.570	0.059	0.524

Table 3: Bootstrap percentile confidence intervals for basic and diversification-consistent (DC) output-oriented models

output-oriented model. And we noted in Section 3.5 that the input-oriented standard model produced more significant differences than the output-oriented model. Fig. 1 helps illustrates why we might observe these effects. Given a point on the standard frontier, we can always find a point on the diversification-consistent frontier with at least as much return on each return measure and at most as much risk on each risk measure. So, any solution of a standard model gives a feasible solution for a corresponding diversification-consistent model and so

in each bootstrap resample the diversification-consistent model efficiency of a fund must be at most the standard model efficiency. Thus both the lower and upper bounds of a percentile bootstrap efficiency confidence interval from the diversification-consistent model must be at most those of the standard model.

Consider a fund o . Any solution $\lambda_1, \dots, \lambda_n, \hat{\lambda}_1, \dots, \hat{\lambda}_{\hat{n}}$ (see constraint (10)) of the diversification-consistent output-oriented model must have at least as much re-

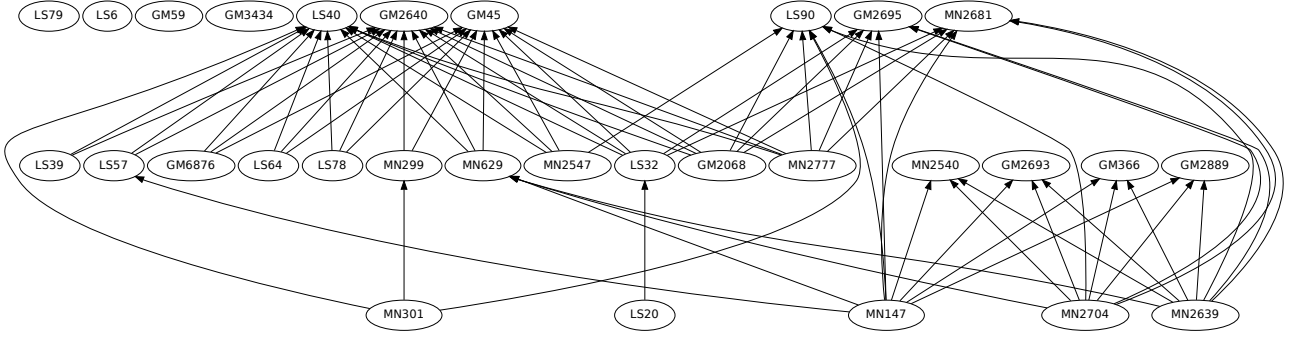


Figure 4: Output-oriented diversification-consistent DEA efficiency rankings using block (size 3) bootstrap and LSD

turn in each measure as the solution $\lambda_1, \dots, \lambda_n$ of the standard model. But the return is not infinite and so, provided o produces positive return in some measure, $\phi'_o > 0$ in the diversification-consistent model. By contrast a solution of the diversification-consistent input-oriented model has at most as much risk as the corresponding solution in the standard model and may have zero risk and so zero efficiency provided o has positive risk. In practice, we find this happens rather often in bootstrap replications for low-return funds, making it hard to distinguish statistically among them.

6 Discussion

The ranking of investment funds is important in supporting fund selection, especially for hedge funds. And DEA allows us to compare and rank funds based on the characteristics of funds' return distributions. But, the accuracy of the ranking also depends on the statistical properties of DEA. We investigate these properties and conclude the following.

First, the DEA efficiency estimates of investment fund performance typically have substantial bias. Worse, this bias may be markedly different for different funds. In particular, high-risk high-return funds may show greater bias than low-risk low return funds. Bootstrap methods help us deal with this bias.

Second, the bootstrap allows us to estimate the accuracy of DEA efficiency estimates. In particular, it gives us confidence intervals for efficiency scores.

Third, resampling gives us a better way than using simple DEA scores to approximately rank investment funds. The ranking should nearly always be a partial ordering rather than a total ordering, and we find Hasse diagrams indicate the ranking parsimoniously.

Fourth, the substantial bias, wide confidence intervals and partial orderings whose precise structure depends on the bootstrap method used suggest we should be modest in how we interpret DEA efficiencies. This is especially true for the model suggested by [Lamb and Tee \(2011\)](#) that accounts for correlation between the performance of different funds. We can identify approximately how well we expect an individual fund to perform in future compared to its peers. But it is difficult to predict with confidence precisely what its future efficiency will be.

We note that the methods we use can be applied to many different DEA models or even to comparing funds using more conventional methods such as Sharpe ratios. We anticipate the same issues of bias, accuracy and ranking will arise.

A Output-oriented diversification-consistent DEA model and iterative procedure

[Lamb and Tee \(2011\)](#) describe an input-oriented diversification-consistent DEA model. The corresponding output-oriented model is as follows. A *portfolio* is

a combination of funds $\sum_{j=1}^n \mu_j f_j$ with $\sum_{j=1}^n \mu_j \leq 1$ and $\mu_j \geq 0$ for $j = 1, \dots, n$. Suppose we have portfolios $\hat{f}_1, \dots, \hat{f}_{\hat{n}}$, which we call *notional units*. Let \hat{x}_{ij} be the risk measure estimate using measure x_j for \hat{f}_j ($i = 1, \dots, m, j = 1, \dots, \hat{n}$). Let \hat{y}_{rj} be the risk measure estimate using measure y_r for \hat{f}_j ($r = 1, \dots, s, j = 1, \dots, \hat{n}$). Define $\phi'_{o,T}$ ($o = 1, \dots, n$) as follows. If $(y_{1o}, \dots, y_{so}) = \mathbf{0}$ put $\phi'_{o,T} = 1$ or 0 according as $(x_{1o}, \dots, x_{mo}) = \mathbf{0}$ or $(x_{1o}, \dots, x_{mo}) > \mathbf{0}$. Otherwise, choose $\eta_o, \lambda_1, \dots, \lambda_n, \hat{\lambda}_1, \dots, \hat{\lambda}_{\hat{n}}$ to

maximise η_o subject to

$$\sum_{j=1}^n x_{ij} \lambda_j + \sum_{j=1}^{\hat{n}} \hat{x}_{ij} \hat{\lambda}_j \leq x_{io}, \quad (i = 1, \dots, m) \quad (9)$$

$$\sum_{j=1}^n y_{rj} \lambda_j + \sum_{j=1}^{\hat{n}} \hat{y}_{rj} \hat{\lambda}_j \geq y_{ro} \eta_o, \quad (r = 1, \dots, s) \quad (10)$$

$$\sum_{j=1}^n \lambda_j + \sum_{j=1}^{\hat{n}} \hat{\lambda}_j \leq 1, \quad (11)$$

and

$$\begin{aligned} \lambda_j &\geq 0 \quad (j = 1, \dots, n), \\ \hat{\lambda}_j &\geq 0 \quad (j = 1, \dots, \hat{n}), \end{aligned} \quad (12)$$

and put $\phi'_{o,T} = \eta_o$.

This model and the input-oriented diversification-consistent model are not generally diversification consistent if they include no notional units (Lamb and Tee, 2011): that is, its efficient frontier does not take into account the possibility of improving performance by diversification. So we use the iterative procedure of Fig. 1 to add notional units and approximate a diversification-consistent model arbitrarily accurately provided the risk and return measures are convexity consistent (see Section 3.2 and Lamb and Tee (2011)).

Lamb and Tee (2011) suggest choosing $\alpha = 0.5$. However, here we choose α using 10 iterations of golden

```

put  $F = \{f_1, \dots, f_n\}$ ;
put  $\hat{F} = \emptyset$ ;
put  $P = \{\{f_i, f_j\} : i, j \in \{1, \dots, n\}, i < j\}$ ;
while  $P \neq \emptyset$  and no other stopping condition is met:
  choose  $\{f, g\} \in P$  and put  $P = P \setminus \{f, g\}$ ;
  choose a pair of distinct funds  $f$  and  $g$  in  $S$  not
  considered before;
  for  $e \in \{f, g\}$ :
    if  $e \in \hat{F}$  and  $e$  is inefficient:
      put  $\hat{F} = \hat{F} \setminus \{e\}$ ;
      for  $p \in P$ : if  $e \in p$ : put  $P = P \setminus \{p\}$ ;
  if neither  $f$  nor  $g$  was removed from  $\hat{F}$ :
    put  $h = \alpha f + (1 - \alpha)g$ ;
    if  $h$  is efficient and  $x_i(h) \neq \alpha x_i(f) + (1 - \alpha)x_i(g)$ 
    for each  $i = 1, \dots, m$  and  $y_r(h) \neq \alpha y_r(f) + (1 - \alpha)y_r(g)$ 
    for each  $r = 1, \dots, s$ :
      for  $e \in F \cup \hat{F}$ : put  $P = P \cup \{e, h\}$ ;
      put  $\hat{F} = \hat{F} \cup \{h\}$ ;

```

Figure 1: Iterative procedure to generate diversification-consistent models

section search (see Winston and Albright (2001)) because we find it gives more accurate approximations without slowing the iterative process too much. We need many computations and to solve large numbers of linear programmes during the iterative procedure. We coded C++ programmes and libraries for the computation and use CPLEX 12.1 (IBM, 2010) to solve the linear programmes. The code is available from <http://www.abdn.ac.uk/~cms127/code.html> together with documentation. We carried out all the computations on a gnu/Linux operating system with a 2.4GHz Intel quad core processor. We stop the iterative procedure when the average reduction in efficiency over ten improving iterations falls below 10^{-4} . In practice, it takes about two hours to calculate efficiencies from 2000 bootstrap replications.

References

- Artzner, P., Delbaen, F., Eber, J. M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Banker, R. D., 1993. Maximum likelihood, consistency

- and data envelopment analysis: A statistical foundation. *Management Science* 39, 1265–1273.
- Briec, W., Kerstens, K., 2009. Multi-horizon Markowitz portfolio performance appraisals: A general approach. *Omega* 37, 50–62.
- Briec, W., Kerstens, K., 2010. Portfolio selection in multidimensional and partial moment space. *Journal of Economic Dynamics & Control* 34, 636–656.
- Brooks, C., Kat, H. M., 2002. The statistical properties of hedge fund index returns and their implications for investors. *The Journal of Alternative Investments* 5 (2), 26–44.
- Bühlmann, P., 1997. Sieve bootstrap for time series. *Bernoulli* 32 (2), 123–148.
- Bühlmann, P., 2002. Bootstraps for time series. *Statistical Science* 17 (1), 52–72.
- Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T., Künsch, H. R., 1998. Matched-block bootstrap for dependent data. *Bernoulli* 4 (3), 305–328.
- Center for International Securities and Derivatives Markets, 2010. Accessed March 2010.
URL <http://www.isenberg.umass.edu/CISDM/>
- Chen, Z., Lin, R., 2006. Mutual fund performance evaluation using data envelopment analysis with new risk measures. *OR Spectrum* 28, 375–398.
- Cooper, W. W., Seiford, L. M., Tone, K., 2007. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd Edition. Springer.
- Dyson, R. G., Shale, E., 2010. Data envelopment analysis, operational research and uncertainty. *Journal of the Operational Research Society* 61 (1), 25–34.
- Efron, B., 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82 (397), 171–185.
- Efron, B., 1990. More efficient bootstrap computations. *Journal of the American Statistical Association* 85 (409), 79–89.
- Efron, B., Tibshirani, R. J., 1998. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- Eling, M., 2006. Performance management of hedge funds using data envelopment analysis. *Financial Markets and Portfolio Management* 20 (4), 442–472.
- Färe, R., Grosskopf, S., Lovell, C., 1994. *Production Frontiers*. Cambridge University Press.
- Ferrier, G. D., Hirschberg, J. G., 1997. Bootstrapping confidence intervals for linear programming efficiency scores: with an illustration using Italian banking data. *Journal of Productivity Analysis* 8, 19–33.
- González, X. M., Miles, D., 2002. Statistical precision of DEA: a bootstrap application to Spanish public services. *Applied Economics Letters* 9, 127–132.
- Gregoriou, G. N., Zhu, J., 2005. *Evaluating Hedge Fund and CTA Performance: Data Envelopment Analysis Approach*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Hall, P., Horowitz, J. R., Jing, B.-Y., 1995. On blocking rules for the bootstrap with dependent data. *Biometrika* 82 (3), 561–574.
- IBM, 2010. CPLEX. <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>, accessed November 2010.
- Inoue, A., Kilian, L., 2002. Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. *International Economic Review* 43 (2), 309–331.
- Jorion, P., 2007. *Value at Risk*, 3rd Edition. McGraw-Hill.
- Joro, T., Na, P., 2006. Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research* 175, 446–461.
- Kao, C., Liu, S.-T., 2009. Stochastic data envelopment analysis in measuring the efficiency of taiwan commercial banks. *European Journal of Operational Research* 196, 312–322.
- Kumar, U. D., Roy, A. B., Saranga, H., Singal, K., 2010. Analysis of hedge fund strategies using slack-based dea models. *Journal of the Operational Research Society* 61, 1746–1760.

- Künsch, H. R., 1989. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17 (3), 1217–1241.
- Kuosmanen, T., Johnson, A. L., 2010. Data envelopment analysis as nonparametric least-squares regression. *Operations Research* 58 (1), 149–160.
- Lamb, J. D., Tee, K.-H., 2011. Data envelopment analysis models of investment funds. *European Journal of Operational Research*, To appear.
- Lhabitant, F.-S., 2004. *Hedge Funds: Quantitative Insights*. John Wiley & Sons, Chichester.
- Ljung, G. M., Box, G. E., 1978. On a measure of lack of fit in time series models. *Biometrika* 65 (2), 297–303.
- Löthgren, M., 2000. On the consistency of the dea-based average efficiency bootstrap. *Applied Economics Letters* 7, 53–57.
- Löthgren, M., Tambour, M., 1999a. Bootstrapping the data envelopment analysis Malmquist index. *Applied Economics* 31, 417–425.
- Löthgren, M., Tambour, M., 1999b. Testing scale efficiency in dea models: a bootstrapping approach. *Applied Economics* 31, 1231–1237.
- Lozano, S., Gutiérrez, E., 2008a. Data envelopment analysis of mutual funds based on second-order stochastic dominance. *European Journal of Operational Research* 189, 230–244.
- Lozano, S., Gutiérrez, E., 2008b. TSD-consistent performance assessment of mutual funds. *Journal of the Operational Research Society* 59, 1352–1362.
- Lütkepohl, H., 2007. *New Introduction to Multiple Time Series Analysis*. Springer.
- Mann, H. B., Wald, A., 1943. On stochastic limit and order relationships. *The Annals of Mathematical Statistics* 14 (3), 217–226.
- Morey, M. R., Morey, R. C., 1999. Mutual fund performance appraisals: a multi-horizon perspective with endogenous benchmarking. *Omega* 27, 241–288.
- Morningstar, 2011. Accessed August 2011. URL <http://www.morningstar.com/>
- Murthi, B. P. S., Choi, Y. K., Desai, P., 1997. Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. *European Journal of Operational Research* 98, 408–418.
- Ogryczak, W., Ruszczyński, A., 1999. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research* 116, 33–50.
- Robinson, S. M., 1977. A characterization of stability in linear programming. *Operations Research* 25 (3), 435–447.
- Rockafellar, R. T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* 26, 1443–1472.
- Rutherford, D. E., 1965. *Introduction to Lattice Theory*. Oliver and Boyd, Edinburgh, UK.
- Scheel, H., Scholtes, S., 2003. Continuity of DEA efficiency measures. *Operations Research* 51 (1), 149–159.
- Shaffer, J. P., 1986. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association* 81 (395), 826–831.
- Simar, L., Wilson, P. W., 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* 44, 49–61.
- Souza, G., Souza, M., Gomes, E., October 2011. Computing confidence intervals for output-oriented dea models: an application to agricultural research in brazil. *Journal of the Operational Research Society* 62 (10), 1844–1850.
- Šidák, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62 (318), 626–633.
- Winston, W. L., Albright, S. C., 2001. *Practical Management Science*, 2nd Edition. Thomson Learning, Pacific Grove, CA.