JEL Classification: C61, D81, G11
Keywords: Data Envelopment Analysis, risk measures, index efficiency, stochastic dominance

DEA-Risk Efficiency and Stochastic Dominance Efficiency of Stock Indices*

Martin BRANDA—Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics (branda@karlin.mff.cuni.cz)
Miloš KOPA—Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic (kopa@karlin.mff.cuni.cz)

Abstract
In this article, we deal with the efficiency of world stock indices. Basically, we compare three approaches: mean-risk, data envelopment analysis (DEA), and stochastic dominance (SD) efficiency. In the DEA methodology, efficiency is defined as a weighted sum of outputs compared to a weighted sum of inputs when optimal weights are used. In DEA-risk efficiency, several risk measures and functionals which quantify the risk of the indices (var, VaR, CVaR, etc.) as DEA inputs are used. Mean gross return is considered as the only DEA output. When only one risk measure as the input and mean gross return as the output are considered, the DEA-risk efficiency is related to the mean-risk efficiency. We test the DEA-risk efficiency of 25 indices and we analyze the sensitivity of our results with respect to the selected inputs. Using stochastic dominance criteria, we test pairwise efficiency as well as portfolio efficiency, allowing full diversification across assets. While SD pairwise efficiency testing is performed for first-order stochastic dominance (FSD) as well as for second-order stochastic dominance (SSD), the SD portfolio efficiency test is considered only for the SSD case. Our numerical analysis compares the results using two sample datasets: before- and during-crisis. The results show that SSD portfolio efficiency is the most powerful efficiency criterion, that is, it classifies only one index as efficient, while FSD (SSD) pairwise efficiency tends to be very weak. The proposed DEA-risk efficiency approach represents a compromise offering a reasonable set of efficient indices.

1. Introduction

In the theory of decision-making, questions about how to maximize profit and how to diversify risk have been around for centuries; however, both of these questions took another dimension with the work of Markowitz (1952). In his work, Markowitz identified the two main components of portfolio performance—mean reward and risk represented by variance—and by applying a simple parametric optimization model he found the optimal trade-off between these two components. If the parameter is known one can easily find the optimal portfolio. If not, at least the set of efficient portfolios (the efficient frontier) can be identified. In this case, the portfolio is seen as efficient if there is no better portfolio, i.e., a portfolio with a higher mean and smaller variance. In the last 60 years, the theory of mean-risk models has been enriched by using other risk measures instead of variance, for example, semi-variance (see Markowitz, 1959), Value at Risk (VaR), and Conditional Value at Risk (CVaR) (see Rockafellar and Uryasev, 2000, 2002).

* Acknowledgements: The research was partly supported by the Czech Science Foundation under grants P402/10/1610, P402/12/0558.

Alternatively, one can adopt utility functions for modeling an investor's risk attitude, especially in the approach maximizing expected utility. Again, if the utility function is perfectly known, one can find the optimal decision. If that is not the case, one can identify at least the set of efficient portfolios with respect to a chosen set of utility functions. Considering all utility functions, that is, assuming only non-satiation for the investor's preferences, leads to the first-order stochastic dominance (FSD) relation (see Levy, 2006, and references therein). Adding the very common assumption of risk aversion, one can obtain the second-order stochastic dominance (SSD) approach. Using pairwise comparisons, an alternative (asset) is classified as FSD (SSD) efficient if there is no other alternative that dominates the former alternative with respect to FSD (SSD). If investors may combine assets in portfolios, tests for portfolio efficiency allowing full diversification across the indices can be of interest. Because of computational considerations, we limit our attention to the SSD portfolio efficiency tests developed in Post (2003), Kuosmanen (2004), Kopa and Chovanec (2008), and Kopa (2010). These tests classify an asset as SSD efficient if there is no portfolio created from the assets that SSD dominates. Alternatively, one can apply FSD efficiency tests as in Kuosmanen (2004) and Kopa and Post (2009), or statistical SSD efficiency tests as in Scaillet and Topaloglou (2010).

A third approach to efficiency is based on Data Envelopment Analysis (DEA) –see, for example, Charnes et al. (1978) and Banker et al. (1984). For example, Basso and Funari (2001, 2003), Murthi et al. (1997), Daraio and Simar (2006), and Galagadera and Silvapulle (2002) applied DEA models to mutual fund performance analysis. A fund is classified as DEA efficient if the inputs of the fund are accurate to its outputs. The mean (gross) return is usually considered as the output, and other fund characteristics, such as transaction costs and risk measures, serve as the inputs. DEA models provide a popular tool for efficiency measurement even in other recent applications, e.g. Cook and Zhu (2010), Roháčová (2011), and Průša (2012).

In this paper we apply all of the above-mentioned approaches to efficiency analysis of stock indices. We consider 25 world stock indices as the basic assets. We compare efficient indices selected according to different criteria: mean-variance efficiency, DEA efficiency, and FSD and SSD efficiency using pairwise comparisons, as well as SSD portfolio efficiency. An index that is classified as efficient with respect to one of the criteria can be attractive for any investor who uses the corresponding criterion for modeling his risk attitude. Moreover, if some index is efficient with respect to more than one criterion, a larger group of decision-makers is interested in it. In the ideal case, we would like to find at least some indices that are classified as efficient when using all of the criteria considered. However, this requirement proved to be very strict. We empirically examine the power of the efficiency approaches considered: the smaller an efficiency set is, the more powerful is the criterion considered. We consider two datasets of weekly returns: before-crisis (September 2006––September 2008) and during-crisis (September 2008–September 2010).

While mean-variance or FSD (SSD) efficiency tests are given precisely, the DEA efficiency model can be constructed in various ways. Contrary to Basso and Funari (2001) and Murthi, Choi, and Desai (1997), we do not consider the transaction costs connected with buying or selling the indices, because we rely only on the mean and risk characteristics of the indices' returns. We choose various risk measures as inputs, starting from the standard deviation up to the modern ones: VaR, CVaR, and corresponding drawdown measures; moreover, the probability risk measures–i.e., VaR, CVaR, and drawdown measures–are used at several probability levels. The only output considered is mean gross return. These DEA-risk models classify an index as efficient if the "total" risk is accurate to its mean gross return, where the total risk is described by a linear convex combination of the risk measures considered. The idea of a combination of risk measures corresponds to risk shaping (see, for example, Cheklov et al., 2005), where weighted values of a given risk measure at different levels are considered. The weights are given by the decision-maker. On the other hand, in the DEA-risk model, more than one risk measure is considered and the weights are specified by the optimization problems, as is usual in DEA models. Moreover, if only one input is considered, then DEA-risk efficiency implies mean-risk efficiency with respect to the same risk measure. For example, if the variance of an index is used as the input, then the DEA-risk efficient index is always mean-variance efficient, too. Therefore, DEA-risk models can be seen as a generalization of mean-risk models.

In the models proposed in Basso and Funari (2001) and Murthi et al. (1997), the weights of the risk measures had to be higher than a small constant, i.e., they always influence fund efficiency. In our DEA-risk models, any risk measure can be eliminated implicitly if the optimization model determines zero weight as optimal. This can give us information about which measures have the greatest or least effect on the DEA efficiency of world stock indices.

In the empirical part of the paper, the DEA efficient indices are compared to those when using mean-risk and stochastic dominance criteria. In addition, we study the sensitivity of the DEA-risk results to including or excluding a single risk measure or a whole group of measures. Moreover, the stochastic dominance analysis includes four different efficiency approaches: the two most popular orders (FSD and SSD) and the two most frequent choices of comparisons (pairwise and portfolio efficiency). Finally, two different periods are considered in the empirical study. An interesting question is how the set of efficient indices differs for pre-crisis data from that for during-crisis data using various efficiency approaches.

The remainder of this paper is structured as follows. Section 2 defines the risk measures considered. It is followed (Section 3) by DEA-risk efficiency formulations using risk measures as inputs and mean gross return as the output. Section 4 recalls the basic ideas of the FSD and SSD approaches and presents the pairwise efficiency tests. Section 5 generalizes the previous SSD results, allowing portfolio efficiency testing with full diversification across the indices. Section 6 presents an empirical application comparing several types of efficiency: mean-variance efficiency, DEA-risk efficiency, FSD and SSD pairwise efficiency, and SSD portfolio efficiency. Section 7 concludes.

2. Risk Measures

In this section, we will show how risk measures can be computed based on discretely distributed returns. We consider a random vector $\mathbf{r} = (r_1, r_2, ..., r_N)'$ of returns of N indices with a discrete probability distribution described by T equiprobable scenarios. The returns of the indices for the various scenarios are given by:

$$X = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^T \end{pmatrix}$$

where $\mathbf{x}^t = \left(x_1^t, x_2^t, ..., x_N^t\right)$ is the t-th row of matrix **X**, representing the index returns along the t-th scenario. That is, x_i^t, $t = 1,...,T$, $i = 1,...,N$, is the t-th realization of the i-th index return. It can be computed as $x_i^t = \frac{P_i^t}{P_i^{t-1}} - 1$, where P_i^t, P_i^{t-1} are the prices of the i-th index at the end of time periods t and $(t-1)$, respectively. Then the *lower semideviation* of order p for the i-th index return can be computed as

$$lsd_i(p) = \left(\frac{1}{T} \sum_{t=1}^{T} \left[x_i^t - \overline{x}_i \right]_-^p \right)^{1/p}$$

where $\overline{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_i^t$ and $[\cdot]_- = -\min\{\cdot, 0\}$.

Following Rockafellar and Uryasev (2002), we will define Value at Risk and Conditional Value at Risk. Firstly, we can sort the realizations of the i-th index return in descending order, that is, $x_i^{[1]} > x_i^{[2]} > ... > x_i^{[T]}$. Secondly, for $\alpha \in (0,1)$, we find the unique index t_α satisfying

$$\frac{t_\alpha - 1}{T} < \alpha \le \frac{t_\alpha}{T}$$

Then the *Value at Risk* (VaR) of the i-th index is defined as

$$VaR_i(\alpha) = -x_i^{[t_\alpha]}$$

There exist several possible definitions of Conditional Value at Risk–see Pflug (2000) and Rockafellar and Uryasev (2000, 2002). If $\alpha > 1 - 1/T$, then the *Conditional Value at Risk* of the i-th index is equal to

$$CVaR_i(\alpha) = VaR_i(\alpha) = -x_i^{[T]}$$

or else it can be computed as a weighted average

$$CVaR_i(\alpha) = \frac{-1}{1-\alpha} \left[\left(\frac{t_\alpha}{T} - \alpha \right) x_i^{[t_\alpha]} + \frac{1}{T} \sum_{t=t_\alpha+1}^{T} x_i^{[t]} \right]$$

Conditional Value at Risk can be also computed using the minimization formula introduced by Pflug (2000) and Rockafellar and Uryasev (2000, 2002). We propose only the formula for the discrete distribution considered:

$$CVaR_i(\alpha) = \min_{y \in R} y + \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} \left[-x_i^t - y \right]_+$$

where $[\cdot]_+ = \max\{\cdot, 0\}$. As the left point of the compact interval of optimal solutions we obtain the Value at Risk at level α.

Denote by $c_i^t, t = 1,\ldots,T$ the uncompounded cumulative rate of the i-th index return:

$$c_i^t = \sum_{k=1}^{t} x_i^k, t = 1,\ldots,T$$

The *absolute drawdown* at the end of period t is defined as

$$AD_i^t = \max_{1\le k\le t} c_i^k - c_i^t$$

The absolute drawdown function compares the current rate of return with the up-to-date minimum value of the returns. When the return drops below the minimum value, the drawdown function is the mirror image of the return until it returns to the minimum value.

Finally, the *Drawdown at Risk* (DaR) of the i-th index can be computed as the VaR on equiprobable absolute drawdowns of the i-th index, and the *Conditional Drawdown at Risk* (CDaR) of the i-th index as the CVaR defined on equiprobable absolute drawdowns of the i-th index. Precise definitions can be found in Cheklov et al. (2003).

3. Data Envelopment Analysis

Data Envelopment Analysis (DEA) was introduced by Charnes et al. (1978) as a way to state the efficiency of a decision-making unit over all other decision-making units. Let $Z_{1i},\ldots,Z_{Ki}$ denote the inputs and $Y_{1i},\ldots,Y_{Li}$ denote the outputs of unit i from the N units considered. The DEA efficiency of unit I is then evaluated using the optimal value of the following program, where the weighted inputs are compared with the weighted outputs:

$$\max \frac{\sum_{l=1}^{L} y_{lI} Y_{lI}}{\sum_{k=1}^{K} w_{kI} Z_{kI}}$$

$$s.t.$$

$$\frac{\sum_{l=1}^{L} y_{lI} Y_{li}}{\sum_{k=1}^{K} w_{kI} Z_{ki}} \le 1, i = 1,\ldots,N$$

$$w_{kI} \ge 0, k = 1,\ldots,K$$

$$y_{lI} \ge 0, l = 1,\ldots,L$$

Unit I is then *DEA efficient* if the optimal value is equal to 1, otherwise it is *DEA inefficient*. The program can be rewritten as a linear program based on fractional programming reformulation (see Charnes et al., 1978):

$$\max \sum_{l=1}^{L} y_{lI} Y_{lI}$$

$$s.t.$$

$$\sum_{k=1}^{K} w_{kI} Z_{kI} = 1$$

$$-\sum_{k=1}^{K} w_{kI} Z_{ki} + \sum_{l=1}^{L} y_{lI} Y_{li} \leq 0, i = 1,...,N$$

$$w_{kI} \geq 0, k = 1,...,K$$

$$y_{lI} \geq 0, l = 1,...,L$$

In the numerical comparison of world stock indices we will consider several risk measures as inputs and mean gross return as the output. This special case of DEA efficiency can be called DEA-risk efficiency. The combination of mean, risk measures, and DEA models can also be found in Lozano and Gutierrez (2008). Contrary to that approach, we consider several risk measures jointly in one DEA model.

3.1 DEA-Risk Efficiency

Let us consider *K* values of risk measures $\rho_{1i},...,\rho_{Ki}$ as inputs $Z_{1i},...,Z_{Ki}$ and the mean gross return $\mu_i = 1 + \overline{x}_i$ as output Y_i of the *i*-th index. We say that index *I* is *DEA-risk efficient* if the optimal objective value of the following problem:

$$\max y_I \mu_I$$
$$s.t.$$
$$\sum_{k=1}^{K} w_{kI} \rho_{kI} = 1, \ -\sum_{k=1}^{K} w_{kI} \rho_{ki} + y_I \mu_i \leq 0, i = 1,...,N \qquad (1)$$
$$w_{kI} \geq 0, k = 1,...,K$$
$$y_I \geq 0$$

is equal to 1. In the special case where only one type of risk measure (for example, CVaR) but at different levels α is considered, DEA-risk modeling is related to the risk-shaping approach (see, for example, Rockafellar and Uryasev, 2002, or Cheklov et al., 2005). The main difference is in the choice of weights. Our model allows another weight scheme, w_{kI}, for each index. The choice is always done in the optimal way, which leads to measurement of the risk of index *I* using the best combination of the risk measures considered.

Example 1

In this example we consider only one risk measure. We will show the basic property of a DEA-risk efficient index. Moreover, we will prove that the DEA-risk efficient index is always mean-risk efficient, too. Let ρ_i be the value of the risk measure considered and μ_i be the mean return of the *i*-th index for $i = 1,...,N.$ Then we say that index *I* is *mean-risk* efficient if there is no index *i* such that $\rho_i \leq \rho_I$ and $\mu_i \geq \mu_I$ with at least one strict inequality. In this setting, the definition of the linear problem (1) simplifies to:

$$
\begin{aligned}
&\max y_I \mu_I \\
&\quad s.t. \\
&w_I \rho_I = 1 \\
&-w_I \rho_i + y_I \mu_i \leq 0, i = 1, \ldots, N \\
&w_I \geq 0 \\
&y_I \geq 0
\end{aligned}
$$

which can be easily rewritten as:

$$
\begin{aligned}
&\max y_I \mu_I \\
&\quad s.t. \\
&y_I \leq \frac{\rho_i}{\mu_i \rho_I}, i = 1, \ldots, N \\
&y_I \geq 0
\end{aligned}
\tag{2}
$$

Since the optimal solution of (2) is

$$
y_I^* = \frac{1}{\rho_I} \min_i \frac{\rho_i}{\mu_i}
$$

the index I is DEA-efficient if it has a minimal *risk-mean ratio*, that is, the ratio of the value of the considered risk measure and mean return is minimal. Assume now that index I is not mean-risk efficient. Then there exists another index with a lower risk-mean ratio. Hence the optimal objective value of (2) is strictly smaller than 1, that is, index I is not DEA-efficient. Summarizing, we have shown that DEA-risk efficiency with only one input always implies mean-risk efficiency.

4. Stochastic Dominance Relations

Stochastic dominance is an appealing approach for comparing random variables. In our case we apply it to compare random stock index returns. Following Levy (2006) and references therein, the i-th index *dominates* the j-th index *with respect to the first-order stochastic dominance* (FSD) if $Eu(r_i) \geq Eu(r_j)$ for all utility functions with strict inequality for at least one utility function. If the same property is satisfied for all concave utility functions with strict inequality for at least one such utility function, then the i-th index *dominates* the j-th index *with respect to second-order stochastic dominance* (SSD). It is clear that an FSD relation implies an SSD relation. Let $F_i(x)$ denote the cumulative probability distribution function of the returns of the i-th index. The twice cumulative probability distribution function of the returns of the i-th index is defined as:

$$
F_i^{(2)}(t) = \int_{-\infty}^{t} F_i(x)\, \mathrm{d}x
$$

Then the FSD and SSD relations can be alternatively defined as follows:

– the i-th index dominates the j-th index with respect to FSD if

$$F_i(t) \le F_j(t) \forall t \in R$$

with strict inequality for at least one $t \in R$;

– the i-th index dominates the j-th index with respect to SSD if

$$F_i^{(2)}(t) \le F_j^{(2)}(t) \forall t \in R$$

with strict inequality for at least one $t \in R$.

4.1 SD Pairwise Efficiency

In this section, we first formulate a computationally attractive algorithm of FSD pairwise efficiency testing. Then we present a modification for SSD pairwise efficiency. Following Levy (2006), let v_i^t , $t = 1,2,..,T$ denote the ordered returns of the i-th index in ascending order, that is, $v_i^1 \le v_i^2 \le ... \le v_i^T$. Then the i-th index dominates the j-th index with respect to first-order stochastic dominance if and only if

$$v_i^t \ge v_j^t, t = 1,2,...,T \tag{3}$$

with at least one strict inequality. Moreover, we classify the j-th index as *FSD pairwise inefficient* if there exists some i-th index satisfying (3). Otherwise, the j-th index is *FSD pairwise efficient*. Therefore, the algorithm for testing the FSD pairwise efficiency of the j-th index consists of two steps. Firstly, we order the returns in ascending order v_i^t for all $i = 1,2,...,N$, $s = 1,2,..,T$. Secondly, we try to find some i satisfying (3). If such i exists then the j-th index is SSD pairwise inefficient. If not, then the j-th index is FSD pairwise efficient.

Testing of SSD pairwise efficiency is performed in a similar way to the previous algorithm using criterion

$$\sum_{t=1}^{s} v_i^t \ge \sum_{t=1}^{s} v_j^t, s = 1,2,...,T$$

instead of (3). Moreover, we use the simple fact that FSD pairwise inefficiency implies SSD pairwise inefficiency. Or, equivalently, an SSD pairwise efficient index is always FSD pairwise efficient, too.

5. Stochastic Dominance Portfolio Efficiency

Contrary to the previous case, an investor may combine indices into his portfolio. We will use $\lambda = \left(\lambda_1, \lambda_2, ..., \lambda_N\right)'$ for a vector of portfolio weights, and the portfolio possibilities are given by

$$\Lambda = \left\{\lambda \in R^N \mid 1'\lambda = 1, \lambda_n \ge 0, n = 1,2,...,N\right\}$$

For computational reasons, we limit our attention only to the SSD case, that is, we define SSD portfolio efficiency with respect to all possible portfolios that can be created from the indices. The j-th index is *SSD portfolio inefficient* if there exists portfolio $\lambda \in \Lambda$ such that $\mathbf{r}'\lambda$ dominates r_j by SSD. Otherwise, the j-th index is *SSD portfolio efficient*. Similarly to how we defined the SSD portfolio efficiency for the j-th index, we can define the SSD portfolio efficiency of any portfolio $\tau \in \Lambda$, using $\mathbf{r}'\tau$ instead of r_j. The following SSD portfolio efficiency tests are formulated for this more general setting. However, one could easily consider $\tau = (1,0,...,0)$, $\tau = (0,1,...,0)$,..., $\tau = (0,0,...,1)$ to test the efficiency of the indices.

5.1 SSD Portfolio Efficiency Test

The following SSD portfolio efficiency test was introduced by Kopa and Chovanec (2008). Alternatively, one can use the SSD portfolio efficiency tests developed in Post (2003) or Kuosmanen (2004). Let $\alpha_k = k/T$, $k \in K = \{0,1,...,T-1\}$. Let

$$D^*(\tau) = \max_{D_k,\lambda_n} \sum_{k=0}^{T-1} D_k$$
$$s.t.$$
$$CVaR_{\alpha_k}(-\mathbf{r}'\tau) - CVaR_{\alpha_k}(-\mathbf{r}'\lambda) \geq D_k, k \in K$$
$$D_k \geq 0,\ k \in K$$
$$\lambda \in \Lambda$$

Using linear programming reformulation of CVaR (see Rockafellar and Uryasev, 2002), we can compute the measure of inefficiency $D^*(\tau)$ as follows:

$$D^*(\tau) = \max_{D_k,\lambda_n,b_k,w_k^t} \sum_{k=1}^{T} D_k$$
$$s.t.$$
$$CVaR_{\frac{k-1}{T}}(-\mathbf{r}'\tau) - b_k - \frac{1}{\left(1-\frac{k-1}{T}T\right)} \sum_{t=1}^{T} w_k^t \geq D_k,\ k \in K$$
$$w_k^t + \mathbf{x}^t\lambda \geq -b_k,\ t,k \in K$$
$$w_k^t \geq 0,\ t,k \in K$$
$$D_k \geq 0, k \in K$$
$$\lambda \in \Lambda$$

If $D^*(\tau) > 0$, then τ is SSD portfolio inefficient and $\mathbf{r}'\lambda^* \succ_{SSD} \mathbf{r}'\tau$. Otherwise, $D^*(\tau) = 0$ and τ is SSD portfolio efficient.

Testing of FSD portfolio efficiency is much more computationally demanding than in the SSD case and therefore we were not able to apply it in the following empirical study. However, formulations of FSD portfolio efficiency tests can be found in Kuosmanen (2004) and Kopa and Post (2009).

6. Stock Index Efficiency–Empirical Study

We consider the following *25 world financial (stock) indices* listed on Yahoo Finance:

– *America* (5): MERVAL BUENOS AIRES, IBOVESPA, S&P TSX Composite index, S&P 500 INDEX RTH, IPC,

Table 1 Descriptive Statistics of Index Returns (before crisis)

Index	mean	std	Skewness	Kurtosis	min	max
MERVAL BUENOS AIRES	0.0006	0.0285	-0.3796	3.4926	-0.0930	0.0790
IBOVESPA	0.0043	0.0354	-0.3984	2.9726	-0.0792	0.0914
S&PTSX Composite index	0.0012	0.0207	-0.8538	4.2427	-0.0693	0.0405
S&P 500 INDEX.RTH	-0.0003	0.0206	-0.3688	3.0898	-0.0541	0.0487
IPC	0.0021	0.0275	-0.2260	3.1223	-0.0766	0.0749
ALL ORDINARIES	0.0000	0.0226	-0.1447	3.6553	-0.0539	0.0735
SSE Composite Index	0.0030	0.0492	0.0064	3.6819	-0.1384	0.1496
HANG SENG INDEX	0.0017	0.0350	0.2907	3.9350	-0.0752	0.1243
BSE SENSEX	0.0021	0.0353	-0.1442	3.0517	-0.0912	0.0959
Jakarta Composite Index	0.0031	0.0376	-0.8524	5.5700	-0.1353	0.1229
FTSE Bursa Malaysia KLCI	0.0010	0.0262	-0.8195	4.6614	-0.0926	0.0688
NIKKEI 225	-0.0024	0.0261	-0.3031	3.6745	-0.0889	0.0639
NZX 50 INDEX GROSS	-0.0008	0.0186	-0.4676	4.0477	-0.0537	0.0533
STRAITS TIMES INDEX	0.0005	0.0282	-0.2139	3.7221	-0.0810	0.0763
KOSPI Composite Index	0.0011	0.0308	-0.1832	4.2746	-0.1041	0.0936
TSEC weighted index	-0.0006	0.0307	-0.7637	4.2176	-0.1049	0.0741
ATX	-0.0010	0.0298	0.1956	2.8467	-0.0728	0.0724
CAC 4	-0.0013	0.0249	-0.3099	2.5887	-0.0638	0.0485
DAX	0.0006	0.0238	-0.4518	3.1551	-0.0680	0.0580
AEX	-0.0018	0.0251	-0.2151	3.1011	-0.0659	0.0657
SMSI	-0.0004	0.0234	-0.4680	2.5474	-0.0564	0.0432
OMX Stockholm PI	-0.0016	0.0267	-0.2470	3.0273	-0.0715	0.0675
SMI	-0.0013	0.0224	-0.2825	2.8023	-0.0573	0.0545
FTSE 100	-0.0007	0.0217	-0.4911	3.1398	-0.0702	0.0447
TEL AVIV TA-100 IND	-0.0002	0.0283	-1.3721	6.7134	-0.1319	0.0655

– *Asia/Pacific* (11): ALL ORDINARIES, SSE Composite Index, HANG SENG INDEX, BSE SENSEX, Jakarta Composite Index, FTSE Bursa Malaysia KLCI, NIKKEI 225, NZX 50 INDEX GROSS, STRAITS TIMES INDEX, KOSPI Composite Index, TSEC weighted index,
– *Europe* (8): ATX, CAC 4, DAX, AEX, SMSI, OMX Stockholm PI, SMI, FTSE 100,
– *Middle East* (1): TEL AVIV TA-100 IND.

In our analysis we describe each index by its weekly rates of return. We divided the returns into two periods:

– before-crisis (B): September 11, 2006–September 15, 2008,
– during-crisis (D): September 16, 2008–September 20, 2010.

Of course, the exact starting date of the crisis is not known and can be debated at length. We chose September 16, 2008 because all stock indices plummeted in the week starting that day. The descriptive statistics of the returns are summarized in *Tables 1* and *2*. Almost all the returns are negatively skewed. Moreover, comparing the before-crisis data with the during-crisis data, we find that the during-crisis returns usually have a higher standard deviation and kurtosis. Focusing on the mean and variance of the returns, we can start our efficiency analysis by identifying the mean-variance efficient indices. Following Markowitz (1952, 1959) we say that the *i*-th

Table 2 Descriptive Statistics of Index Returns (during crisis)

Index	mean	std	Skewness	Kurtosis	min	max
MERVAL BUENOS AIRES	0.0062	0.0631	-0.8847	7.5042	-0.2679	0.1990
IBOVESPA	0.0037	0.0519	-0.1997	6.7006	-0.2001	0.1834
S&PTSX Composite index	0.0004	0.0419	-0.5470	6.1356	-0.1609	0.1368
S&P 500 INDEX.RTH	0.0000	0.0420	-0.5036	6.0514	-0.1820	0.1203
IPC	0.0036	0.0480	0.3712	7.0059	-0.1641	0.2042
ALL ORDINARIES	0.0003	0.0360	-1.0859	6.0336	-0.1623	0.0844
SSE Composite Index	0.0029	0.0409	0.2583	3.1658	-0.0827	0.1366
HANG SENG INDEX	0.0023	0.0449	-0.3925	4.5676	-0.1632	0.1204
BSE SENSEX	0.0045	0.0464	-0.2953	4.8833	-0.1595	0.1408
Jakarta Composite Index	0.0066	0.0437	-1.2644	8.5629	-0.2137	0.1153
FTSE Bursa Malaysia KLCI	0.0035	0.0202	-0.9039	5.2007	-0.0813	0.0498
NIKKEI 225	-0.0011	0.0454	-1.3699	9.9061	-0.2433	0.1213
NZX 50 INDEX GROSS	0.0003	0.0234	-1.0423	7.7343	-0.1099	0.0563
STRAITS TIMES INDEX	0.0027	0.0415	-0.1928	7.3056	-0.1518	0.1656
KOSPI Composite Index	0.0032	0.0435	-0.7363	9.9194	-0.2049	0.1857
TSEC weighted index	0.0036	0.0352	-0.3891	3.4918	-0.1065	0.0987
ATX	-0.0003	0.0612	-1.0259	7.5609	-0.2892	0.1880
CAC 4	-0.0001	0.0474	-0.9893	6.9359	-0.2216	0.1324
DAX	0.0014	0.0481	-0.6176	7.2379	-0.2161	0.1612
AEX	0.0001	0.0494	-1.1847	8.4153	-0.2499	0.1329
SMSI	0.0002	0.0488	-1.0729	6.0520	-0.2099	0.1131
OMX Stockholm PI	0.0033	0.0436	-1.0648	7.5521	-0.2059	0.1161
SMI	-0.0001	0.0416	-0.9636	11.8030	-0.2228	0.1407
FTSE 100	0.0014	0.0423	-0.8733	9.1228	-0.2105	0.1341
TEL AVIV TA-100 IND	0.0043	0.0348	-0.6951	4.3946	-0.1140	0.0903

Table 3 Mean-Variance Efficient Indices (B–before crisis, D–during crisis)

Period	B	D
IBOVESPA	X	
S&PTSX Composite index	X	
S&P 500 INDEX,RTH	X	
IPC	X	
Jakarta Composite Index		X
FTSE Bursa Malaysia KLCI		X
NZX 50 INDEX GROSS	X	
TEL AVIV TA-100 IND		X

index is mean-variance efficient if there is no other index having returns with higher or equal mean and smaller or equal variance (with at least one strict inequality). It is no surprise that four out of the five American indices considered are classified as mean-variance efficient in the before-crisis case. Perhaps surprisingly, all of them become mean-variance inefficient in the during-crisis period. Moreover, all the European indices are mean-variance inefficient in both periods. The list of mean-variance efficient indices is presented in *Table 3*. In the last two columns, the mean-variance indices are marked by X.

Table 4 Mean-Variance Efficient and SSD Pairwise Efficient Indices (B–before crisis, D–during crisis)

Period	mean-variance		SSD pairwise	
	B	D	B	D
IBOVESPA	X		X	
S&PTSX Composite index	X		X	
S&P 500 INDEX,RTH	X		X	
IPC	X		X	
ALL ORDINARIES			X	
HANG SENG INDEX			X	
BSE SENSEX			X	X
Jakarta Composite Index		X	X	X
FTSE Bursa Malaysia KLCI		X		X
NZX 50 INDEX GROSS	X		X	
TSEC weighted index				X
DAX			X	
TEL AVIV TA-100 IND		X		X

6.1 Efficiency with Respect to Stochastic Dominance

Using methods introduced in Sections 4 and 5, we analyze index efficiency with respect to stochastic dominance criteria. We start with FSD pairwise efficiency testing, because it identifies the largest number of efficient indices. In this case, an index is efficient if the expected utility of its returns is maximal for at least one utility function. This means that the FSD pairwise efficient index is the best choice (out of all the 25 indices considered) for at least one decision-maker, because it maximizes his expected utility. Using the algorithm described in Section 4.1, we classify all indices in both periods as FSD pairwise efficient. Therefore, we want to observe how the set of efficient indices is reduced when only risk-averse decision-makers are considered, which leads to SSD pairwise efficiency testing. The results of these tests are compared to the mean-variance efficiency in *Table 4*.

It is well known that mean-variance efficiency does not imply stochastic dominance efficiency (see Levy, 2006). However, *Table 4* shows that every mean-variance efficient index is SSD pairwise efficient, too. Hence, in this study, SSD pairwise efficiency can be seen as a generalization of the mean-variance efficiency approach. Moreover, there are only six indices that are classified differently—five of them for the before-crisis data and two of them for the during-crisis data. Finally, we allow full diversification among the indices and apply the Kopa and Chovanec (2008) test to identify SSD portfolio efficient indices. This property is much stronger than SSD pairwise efficiency because a given index is SSD portfolio efficient if there exists no linear convex combination of the considered indices which SSD-dominates the underlying index. Therefore, every SSD portfolio efficient index is also SSD pairwise efficient. The comparison of these two approaches can be seen in *Table 5*.

As we expected, the SSD portfolio efficiency classification is very strong. Only the index with the highest mean is SSD portfolio efficient (in both periods). All other indices are SSD portfolio inefficient, that is, all risk-averse decision-makers prefer some combination of other indices to the index.

Table 5 SSD Portfolio Efficient and SSD Pairwise Efficient Indices (B–before crisis, D–during crisis)

Period	mean-variance		SSD pairwise	
	B	D		B
IBOVESPA	X		X	
S&PTSX Composite index			X	
S&P 500 INDEX,RTH			X	
IPC			X	
ALL ORDINARIES			X	
HANG SENG INDEX			X	
BSE SENSEX			X	X
Jakarta Composite Index		X	X	X
FTSE Bursa Malaysia KLCI				X
NZX 50 INDEX GROSS			X	
TSEC weighted index				X
DAX			X	
TEL AVIV TA-100 IND				X

Table 6 DEA-risk (B–before crisis, D–during crisis)

Period	B	D
S&PTSX Composite index	X	
S&P 500 INDEX,RTH	X	
ALL ORDINARIES	X	
FTSE Bursa Malaysia KLCI		X
NZX 50 INDEX GROSS	X	X
FTSE 100	X	

6.2 DEA-Risk Efficiency

In both periods, we compute the risk measures (at given levels) that are used as the inputs to the DEA-risk model:

– *Inputs* (28): std; lsd for powers 1,2,3; VaR, CVaR, DaR and CDaR at levels 0.75, 0.9, 0.95, 0.99, 0.995.
– *Output:* mean gross return.

In general, one can choose arbitrary risk measures at arbitrary levels as the inputs of a DEA-risk model. The only limitation is that all inputs of all indices must be non-negative. We employ the measures that have tended to be the most popular in recent years. On the other hand, the output of the DEA-risk model is precisely specified. Since the outputs of all indices must be non-negative, the mean return cannot generally be used. Therefore, we modified it to the mean gross return. The results of DEA-efficiency testing can be found in *Table 6*.

Similar to the case of mean-variance efficiency, only a few indices are classified as efficient and the crisis almost completely changed the set of efficient indices. However, contrary to the mean-variance case, an index exists that is DEA-efficient in both periods.

As a by-product of DEA-efficiency testing, we can analyze the optimal weights of the inputs. Specifically, we can distinguish between zero and non-zero weights. For a given index, omitting the input causes no harm if the optimal weight

Table 7 Number of Indices with Positive Weight of Selected Risk Measure in DEA-Risk Model

			VaR		
Level	0.75	0.9	0.95	0.99	0.995
Before crisis	4	7	4	15	5
During crisis	2	4	5	2	3
			CVaR		
Level	0.75	0.9	0.95	0.99	0.995
Before crisis	0	0	4	0	0
During crisis	0	0	0	0	1
			DaR		
Level	0.75	0.9	0.95	0.99	0.995
Before crisis	9	0	0	1	0
During crisis	2	0	0	3	1
			CDaR		
Level	0.75	0.9	0.95	0.99	0.995
Before crisis	0	0	0	0	0
During crisis	0	0	1	0	1
		lsd		**std**	
Power	1	2	3		
Before crisis	0	0	1	0	
During crisis	2	0	0	1	

of an input is zero; that is, the input has no relevant impact on DEA-risk efficiency. In *Table 7*, we present the number of indices having positive optimal weights of particular risk measures in the DEA-risk model.

As we can see, Value at Risk at all considered levels plays the crucial role in our DEA-risk analysis. Particularly in the before-crisis period, VaR at level 0.99 is an important measure for 15 indices. In Section 6.1 we found that first-order stochastic dominance pairwise efficiency testing identified the largest set of efficient indices. Since the FSD criterion can be expressed using VaR (see Ogryczak and Ruszczynski, 2002), the important role of VaR is in accordance with our FSD testing results. On the other hand, omitting the standard deviation or CDaR measures from DEA-risk testing causes no harm for nearly all the indices.

6.3 Sensitivity of DEA-Risk Efficiency

We also studied the sensitivity of our results with respect to the inputs. We start with stability analysis with respect to selected types of risk measures. *Tables 8* and *9* present the efficient indices when only one type of risk measure (at all considered levels) is considered.

In both periods, we find that Value at Risk identifies the largest set of efficient indices. If another type of risk measure is considered, then always only one index is DEA-risk efficient. Similarly, we can test DEA-efficiency if one type of risk measure is omitted. The results are summarized in *Tables 10* and *11*.

Again, we can see the crucial role of VaR in both periods. If it is kept in the model with reduced inputs, the results obtained are the same as those in the full

Table 8 Sensitivity of Results–with One Type of Risk Measure (before crisis)

	Only with				
	VaR	CVaR	DaR	CDaR	lsd and std
S&PTSX Composite index	X		X	X	
S&P 500 INDEX,RTH	X				
NZX 50 INDEX GROSS	X	X			X
FTSE 100	X				

Table 9 Sensitivity of Results–with One Type of Risk Measure (during crisis)

	Only with				
	VaR	CVaR	DaR	CDaR	lsd and std
FTSE Bursa Malaysia KLCI	X	X	X	X	X
NZX 50 INDEX GROSS	X				

Table 10 Sensitivity of Results–with One Type of Risk Measure Omitted (before crisis)

	Without				
	VaR	CVaR	DaR	CDaR	lsd and std
S&PTSX Composite index	X	X	X	X	X
S&P 500 INDEX,RTH	X	X	X	X	X
ALL ORDINARIES	X	X		X	X
NZX 50 INDEX GROSS	X	X	X	X	X
FTSE 100		X	X	X	X

Table 11 Sensitivity of Results–with One Type of Risk Measure Omitted (during crisis)

	Without				
	VaR	CVaR	DaR	CDaR	lsd and std
FTSE Bursa Malaysia KLCI	X	X	X	X	X
NZX 50 INDEX GROSS		X	X	X	X

model (see *Table 6*). On the other hand, if VaR is not considered as an input, the number of efficient indices decreases. The same can be concluded for DaR, but only in the case of the before-crisis data.

Similar to the previous analysis, we will examine the sensitivity of the result with respect to the levels considered. In *Tables 12*, *13*, *14*, and *15* we can see results for models in which only risk measures at particular levels are either kept as inputs or dropped. Dropping a level hardly changed the set of efficient indices at all.

Finally, we tested the sensitivity of the results to dropping exactly one input. If we drop almost any input, we get the same results as in the case of the full model (*Table 6*). The only exceptions are $DaR_{0.75}$, which causes the All Ordinaries index to become inefficient in the before-crisis period, and $VaR_{0.9}$, which causes the NZX 50 Index Gross to be identified as inefficient based on during-crisis data.

Table 12 Sensitivity of Results–with Only One Level (before crisis)

Index	With level				
	0.75	0.9	0.95	0.99	0.995
S&PTSX Composite index	X	X	X	X	X
S&P 500 INDEX,RTH			X	X	X
NZX 50 INDEX GROSS	X	X	X	X	X
FTSE 100			X		

Table 13 Sensitivity of Results–with Only One Level (during crisis)

Index	With level				
	0.75	0.9	0.95	0.99	0.995
FTSE Bursa Malaysia KLCI	X	X	X	X	X
NZX 50 INDEX GROSS		X			

Table 14 Sensitivity of Results–without Only One Level (before crisis)

Index	Without level				
	0.75	0.9	0.95	0.99	0.995
S&PTSX Composite index	X	X	X	X	X
S&P 500 INDEX,RTH	X	X	X	X	X
ALL ORDINARIES		X	X	X	
NZX 50 INDEX GROSS	X	X	X	X	X
FTSE 100	X	X	X	X	X

Table 15 Sensitivity of Results–withut Only One Level (during crisis)

Index	Without level				
	0.75	0.9	0.95	0.99	0.995
FTSE Bursa Malaysia KLCI	X	X	X	X	X
NZX 50 INDEX GROSS	X		X	X	X

6.4 Comparison of Different Efficiency Approaches

In this section we compare all four efficiency approaches considered in this paper:

– mean-variance efficiency,
– DEA-risk efficiency with all inputs,
– SSD pairwise efficiency,
– SSD portfolio efficiency.

We do not include results of FSD pairwise efficiency testing, because the FSD criterion is too weak to classify any index as inefficient. The comparison is presented in *Table 16*.

Unfortunately, no index is classified as efficient using all four methods. Since the SSD portfolio efficiency tests identified only one efficient index in both periods, we limit our attention to the other three approaches. In the before-crisis case, we can find three indices (S&PTSX Comp. index; S&P 500 INDEX, RTH; NZX 50 INDEX GROSS) that are mean-variance efficient and DEA-risk efficient as well as SSD pairwise efficient, and 14 indices are classified as inefficient in all three cases. Using

Table 16 All Four Types of Efficiency (B–before crisis, D–during crisis)

Period	Mean-variance		DEA-risk		SSD Pairwise		SSD Portfolio	
	B	D	B	D	B	D	B	D
IBOVESPA	X				X		X	
S&PTSX Comp. index	X		X		X			
S&P 500 INDEX,RTH	X		X		X			
IPC	X				X			
ALL ORDINARIES			X		X			
HANG SENG INDEX					X			
BSE SENSEX					X	X		
Jakarta Composite Index		X			X	X		X
FTSE Bursa Malaysia		X		X		X		
NZX 50 INDEX GROSS	X		X	X	X			
TSEC weighted index						X		
DAX					X			
FTSE 100			X					
TEL AVIV TA-100 IND		X				X		

during-crisis data, only one index (FTSE Bursa Malaysia) is mean-variance efficient and DEA-risk efficient as well as SSD pairwise efficient, and 21 indices are classified as inefficient in all three cases.

Finally, we compare the efficiency classifications of the before-crisis data with those of the during-crisis data. Perhaps surprisingly, there are only three indices (BSE SENSEX; Jakarta Composite Index; NZX 50 INDEX GROSS) that are classified as efficient in both periods using at least one approach.

7. Conclusions

In this paper we analyzed the efficiency of 25 world stock indices using three different approaches: mean-risk efficiency, stochastic dominance efficiency, and DEA-risk efficiency. We identified efficient portfolios with respect to mean-variance, DEA-risk efficiency, FSD pairwise, SSD pairwise, and SSD portfolio criteria. We considered two periods: before-crisis and during-crisis. We started with the classical mean-variance method. We found that four out of the five American indices considered were classified as mean-variance efficient in the before-crisis period, but none were so in the during-crisis period. Following, for example, Basso and Funari (2001) and Murthi et al. (1997), we proposed a new DEA-risk efficiency test based on the classical DEA model of Charnes et al. (1978), where several risk measures and functionals which quantify the risk are used as inputs, and mean gross return is used as the output. Using the stochastic dominance approach, we found the FSD pairwise test too weak because it classified all indices as efficient (in both periods). Moreover, the SSD portfolio efficiency test tended to be too strong–only one index was SSD portfolio efficient (in each period). Therefore, we focused mainly on comparing the mean-variance, DEA-risk, and SSD pairwise results. We found four indices that were classified as efficient using all three methods: S&PTSX Comp., S&P 500 INDEX, RTH, NZX 50 INDEX GROSS (before the crisis), and FTSE Bursa Malaysia (during the crisis). Comparing the number of efficient indices, we can order the efficiency criteria considered with respect to their power of inefficiency identification from the weakest to the strongest: FSD pairwise efficiency < SSD pairwise efficiency < mean-variance efficiency < DEA-risk efficiency < SSD portfolio efficiency. Although the comparison of power is the same for both periods considered, we discovered that the crisis almost completely changed the set of efficient indices, no matter which approach we used. Finally, we analyzed the sensitivity of the DEA-risk efficiency classification with respect to changes in inputs. The study showed that Value at Risk played the most important role among the risk measures considered. On the other hand, the impact of standard deviation and CDaR measures proved to be negligible.

REFERENCES

Banker RD, Charnes A, Cooper W (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9):1078–1092.

Basso A, Funari S (2001): A data envelopment analysis approach to measure the mutual fund performance. *European Journal of Operations Research*, 135(3):477–492.

Basso A, Funari S (2003): Measuring the performance of ethical mutual funds: A DEA approach. *Journal of the Operational Research Society*, 54(5):521–531.

Charnes A, Cooper W, Rhodes E (1978): Measuring the efficiency of decision-making units. *European Journal of Operations Research*, 2(6):429–444.

Cheklov A, Uryasev S, Zabarankin M (2003): Portfolio Optimization with Drawdown Constraint. In: Scherer B (Ed.): *Asset and Liability Management Tools, Risk Books*. London.

Cheklov A, Uryasev S, Zabarankin M (2005): Drawdown Measure in Portfolio Optimization. *International Journal of Theoretical and Applied Finance*, 8(1):13–58.

Cook WD, Zhu J (2010): Context-dependent performance standards in DEA. *Annals of Operations Research*, 173(1):163–175.

Daraio C, Simar L (2006): A robust nonparametric approach to evaluate and explain the performance of mutual funds. *European Journal of Operational Research*, 175(1):516–542.

Galagadera UA, Silvapulle P (2002): Australian mutual fund performance appraisal using data envelopment analysis. *Managerial Finance*, 28(9):60–73.

Kopa M (2010): Measuring of second-order stochastic dominance portfolio efficiency. *Kybernetika*, 46(3):488–500.

Kopa M, Chovanec P (2008): A Second-Order Stochastic Dominance Portfolio Efficiency Measure. *Kybernetika,* 44(2):243–258.

Kopa M, Post T (2009): A portfolio efficiency test based on FSD optimality. *Journal of Financial and Quantitative Analysis*, 44(5):1103–1124.

Kuosmannen T (2004): Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10):1390–1406.

Levy H (2006): *Stochastic Dominance: Investment Decision Making under Uncertainty*. Second edition. Springer Science, New York.

Lozano S, Gutierrez E (2008): Data envelopment analysis of mutual funds based on second-order stochastic dominance. *European Journal of Operational Research*, 189(1):230–244.

Markowitz HM (1952): Portfolio Selection. *Journal of Finance*, 7(1):77–91.

Markowitz HM (1959): *Portfolio Selection: Efficient Diversification in Investments.* John Wiley & Sons, Inc., New York.

Murthi BPS, Choi YK, Desai P (1997): Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach. *European Journal of Operational Research*, 98(2): 408–418.

Ogryczak W, Ruszczyński A (2002): Dual stochastic dominance and related mean-risk models. *SIAM J. Optim*., 13(1):60–78.

Pflug G (2000): Some Remarks on the value-at-risk and the Conditional value-at-risk. In: (Uryasev S, Ed.): *Probabilistic Constrained Optimization*. Kluwer, Dordrecht, pp. 272–281.

Post T (2003): Empirical tests for stochastic dominance efficiency. *Journal of Finance*, 58(5): 1905–1932.

Průša J (2012): The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis. *Finance a úvěr-Czech Journal of Economics and Finance*, 62(1):44–65.

Rockafellar RT, Uryasev S (2000): Optimization of Conditional Value-At-Risk. *Journal of Risk*, 2(3):21–41.

Rockafellar RT, Uryasev S (2002): Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26(7):1443–1471.

Roháčová V (2011): Cost efficiency of selected transport companies of the Czech Republic–an application of data envelopment analysis (in Slovak). *Ekonomická revue-Central European Review of Economic Issues*, 14(3):175–182.

Scaillet O, Topaloglou N (2010): Testing for Stochastic Dominance Efficiency. *Journal of Business and Economic Statistics*, 28(1):169–180.