

Discussion Paper No. 05-50

**Is there a Difference?
The Performance Characteristics
of SRI Equity Indexes**

Michael Schröder

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 05-50

**Is there a Difference?
The Performance Characteristics
of SRI Equity Indexes**

Michael Schröder

Download this ZEW Discussion Paper from our ftp server:

<ftp://ftp.zew.de/pub/zew-docs/dp/dp0550.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non-Technical Summary

Socially responsible investments (SRI) in equities are no longer a negligible segment of international capital markets. In the United States about 11% of the funds under management have been invested in socially screened portfolios in 2003. For investors it is interesting to know whether equities selected by an SRI screening process exhibit a different performance than conventional investments in stocks.

In contrast to earlier studies on the performance and risk-return characteristics of SRI equities this study concentrates on SRI indexes and not on investment funds. The advantage of this is that the effects of the SRI screening process can be studied relatively directly without the need to filter out the transaction costs of the funds, the effects of management skills or the timing activities of the fund management. Since the SRI indexes apply SRI screens that are comparable to those of the investment funds, the results of this study are also relevant for the assessment of the performance potential of SRI mutual funds.

Throughout the study different settings to test performance are applied to the 29 SRI stock indexes under consideration. These include (1) single-factor models that use one specific conventional benchmark index that closely approximate the investment universe of each SRI index, (2) three-factor models also using two so called style factors for the analysis of two specific SRI indexes (Naturaktienindex (NAI), Domini 400 index), (3) multi-equation models with the same benchmarks as in (1), and (4), as a check of the robustness of the results, multi-equation models with either one or three world-wide factors.

Using these different models, tests on the parameters are conducted. These tests concern the relative performance and the relative risk of the SRI indexes compared to the benchmarks. The performance tests confirm the results of most of the earlier studies, including the fact that the SRI screens for equities do not lead to a significant performance difference, neither an out- nor an under-performance. The latter is particularly interesting as the SRI screening process decreases the investment universe, which should also lead to a reduction in the risk-adjusted return. This is not the case since an investment in SRI equity indexes seems not to impose additional costs in terms of a performance reduction to the investor.

For some SRI indexes the index series have been calculated backwards from the official inception date on. An important result for performance comparisons is that for these SRI indexes, the backward-looking performance seems to be biased upwards compared to the development from the launch date on. Performance comparisons using the total time series therefore give only an upper bound for the performance of these indexes.

The analysis of the SRI index returns revealed that most of these indexes have a higher risk compared to the benchmarks, which means that changes in the benchmark

are translated into larger changes of the SRI indexes. This is also the reason that for more than half of the SRI indexes, the spanning tests are rejected. For at least 9 out of the 29 SRI indexes the hypothesis of spanning is not rejected in the single-equation/single-factor tests. This means that for these indexes the relevant benchmark index has on average the same risk-return characteristics as the SRI index.

Different multi-equation systems are also used to exploit information from the cross-section of SRI indexes. These tests confirm the results of the single equation tests. Additional robustness checks applying the same set of world-wide diversifying indexes to all SRI indexes – either in a single- or a three-factor framework – also come to the conclusion that SRI indexes exhibit the same performance as the benchmarks and that differences in the risk-return characteristics primarily stem from risk differentials.

Is there a Difference? The Performance Characteristics of SRI Equity Indexes

Michael Schröder

Centre for European Economic Research (ZEW), Mannheim

July 2005

Abstract

Investments in socially responsible investments (SRI) are still a small, but growing segment of international capital markets. This study analyses whether a SRI screening process applied to equities results in a different performance outcome compared to relevant conventional benchmark indexes. In contrast to other studies, the analysis concentrates on SRI indexes and not on investment funds. This has several advantages, which include that the transaction costs of funds, the timing activities and the skill of the fund management do not have to be considered. This leads to a relatively direct measure of the performance effects of SRI screens. The 29 SRI stock indexes are analysed by single-factor models with benchmarks that closely approximate the investment universe of the SRI stock indexes and by multi-equation systems that also exploit the information in the cross-section. The results show that SRI stock indexes do not exhibit a different risk-adjusted return than conventional benchmarks. But many SRI indexes have a higher risk relative to the benchmarks. These findings are robust to the use of different sets of benchmark indexes and apply to all common types of SRI screening.

JEL-Classification: G11, G14, M14

Keywords: Socially responsible investing, equity indexes, performance, risk

Corresponding address:

Michael Schröder
Centre for European Economic Research (ZEW)
P.O. Box 103443
D-68034 Mannheim
Germany

Tel.: +49 621 1235-140
Fax: +49 621 1235-223
Email: schroeder@zew.de
Internet: www.zew.de

1 Introduction

Investments based on social, ethical and environmental criteria are still a small but growing segment of the international capital markets.¹ The development of socially responsible investing (SRI)² is also of interest for academic finance, particularly with respect to performance measurement.

The empirical analysis of SRI funds dates back as early as 1972 to a study of Moskowitz. Since then numerous studies investigated the performance of SRI investments and compared the findings to the performance of conventional assets. The results of many of these studies show that the performance differential of SRI investment funds and traditionally managed funds does not deviate significantly from zero. Some studies (e.g., Derwall et al. (2005)) even find an out-performance of SRI portfolios. This is a puzzling result because SRI funds are restricted to a subset of the total investment universe and should therefore exhibit – at best – the same performance as comparable conventional portfolios.

The aim of this study is the comparison of the main risk-return characteristics of the most important international SRI equity indexes with those of conventional benchmark indexes. The main questions that are investigated in this study include:

1. Do equity indexes based on SRI screening³ exhibit a different performance than conventional benchmark indexes? As most of the earlier studies on the performance of SRI assets focus on mutual funds or some selected indexes such as the Domini 400, this study extends the literature by investigating the 29 most important SRI equity indexes world-wide.
2. Do SRI equity indexes have the same risk as conventional indexes? A common assumption is that the SRI screen should lead to a selection of assets with a relatively low risk. This hypothesis will be tested in our study.
3. Can SRI equity indexes be replicated by conventional benchmark indexes? Technically, this is a question of whether SRI equity indexes are spanned by their benchmark indexes. In that case the information content of SRI indexes would be

¹ According to the Trend Report of the Social Investment Forum (2003) about 2143 bn US dollar have been invested in socially screened portfolios in 2003 in the United States. This is approximately 11% of all investment assets under management in the United States. In June 2004 about 19 bn euro have been invested in this type of asset in Europe, where the largest market for socially screened investments is the United Kingdom with 6,9 bn euro of assets under management (Avanzi (2004)).

² Throughout this study the expression “socially responsible investments“ is used to identify this market segment. This is the common expression in the United States. Other terms which are often used synonymously are “ethical investments” and “sustainable investments”.

³ Social screens usually apply both positive and negative criteria. Negative criteria (as e.g. business activities related to alcohol, firearms, gambling, nuclear power, tobacco) are used to delete specific stocks from the investment universe, whereas positive criteria (e.g. low pollutant emissions, good workplace conditions, equal employment conditions) serve to choose specific stocks for investment.

redundant. This test is stronger and more informative than the pure comparison of the performance.

All mentioned estimations and tests are conducted both for single equations and for groups of SRI indexes. The latter uses the information of the cross-section of several indexes and therefore improves the quality of the tests.

Contrary to the vast majority of research on this topic the analysis is solely based on indexes and not on investment funds. The analysis of SRI indexes (e.g. the Domini 400 Social-Index, the FTSE4Good-indexes or the Dow Jones Sustainability Indexes (DJSI)) has the advantage that the effect of a SRI screen can be measured relatively directly. Through the use of investment funds, the ability of the portfolio management to produce an outstanding performance interferes with the effect of the SRI screening criteria. Particularly market timing and the use of publicly available information as instruments for conditional estimations do not have to be considered when indexes are analysed. In addition, our approach also avoids the difficult task of correctly considering the transaction costs of investment funds.

Early studies on the performance of SRI investment funds use time-series regressions based on a one- or two-factor model but without appropriately considering the costs of the SRI investment funds compared to the chosen benchmark. The common performance measure of most studies is Jensen's alpha, i.e. the constant of a regression model that uses one or more factors to explain the fund returns. Luther, Matatko and Corner (1992) analyse British ethical funds, Hamilton, Jo and Statman (1993) investigate U.S. funds and White (1995) researches German and U.S. funds using a simple regression against a market index. Luther and Matatko (1994) is the first study that includes not only a broad market index but also an index for companies with low market capitalisation ("small cap stocks"). The studies find no clear out- or under-performance of SRI funds.

The most advanced studies on the performance of SRI mutual funds apply a so called matching approach, i.e. they compare the performance of SRI and non-SRI investment funds with similar characteristics e.g. concerning investment universe, fund size and fund age. The aim of this approach is to appropriately consider management and transaction costs for both SRI funds and the conventional funds that serve as benchmarks (see e.g. Mallin, Saadouni and Briston (1995), Gregory, Matatko and Luther (1997), Statman (2000), Kreander et al. (2002), and Stone et al. (2001)). The authors of these studies conclude that SRI and conventional funds exhibit a very similar performance.

In a recent study Bauer, Koedijk and Otten (2005) apply a multi-factor model to investigate the performance of British, German and U.S. investment funds which apply a social screening. The authors compare a portfolio of all SRI funds with a portfolio of all conventional funds for each country. They use Carhart's (1997)-four-factor model: The first three factors are those of the Fama-French (1993)-model and the fourth factor captures the effect of momentum. The main findings are, (1) German and U.S. SRI mutual funds under-perform both their relevant indexes and the conventional funds, whereas UK funds slightly out-perform. However these

differences are not significant. (2) SRI funds seem to have investment styles that are different from conventional funds. For example, the funds are tilted towards companies with a low book-to-market value (= growth stocks) and those with low market capitalisation. The latter result is a confirmation of the findings from earlier studies (as e.g. Luther and Matatko (1994)).

Although the matching approach theoretically leads to an unbiased estimate of the performance differential, the practical implementation bears the risk that an important characteristic might not be considered. The use of indexes in our study is a much easier and more direct way to measure the performance contribution of SRI screening.

Derwall et al. (2005) is the only study that finds that SRI screening leads to a highly significant out-performance. They rank equities using an eco-efficiency rating and construct a difference portfolio of the high-ranked and the low-ranked equities. Applying a four-factor model as in Bauer, Koedijk and Otten (2005), the authors estimate a surprisingly high out-performance of the high-ranked companies, which also persists after considering transaction costs. The authors suspect that their result might be an indication of mispricing in the stock market.

On the other hand, Geczy, Stambaugh and Levin (2003) conclude that “investors who allocate their wealth to socially responsible equity mutual funds pay a price”, i.e. they receive a lower risk-adjusted return for their investment. The authors use price data of equity mutual funds and simulate optimal portfolios under the assumption of different pricing models for the stock market. The amount of reduction in the risk-adjusted return depends strongly on the assumed pricing model and the investor behaviour and may in some circumstances amount to only a few basis points per annum.

In contrast to these two studies, which use synthetic portfolios, our study concentrates on SRI equity indexes that are easily accessible to investors. The advantage of using indexes compared to synthetic portfolios is twofold. First, they are used by investors and investment companies as a guideline for their investment in SRI assets, and second, these indexes apply the same (or at least a very similar) SRI screen than most of the SRI equity mutual funds. Thus, the results of the SRI indexes are also applicable for the assessment of the SRI screening procedures of many mutual funds.

So far only a few studies analyse the performance of SRI indexes. These studies are particularly Kurtz and DiBartolomeo (1996), Sauer (1997), DiBartolomeo and Kurtz (1999) and Statman (2000), who investigate the performance of the Domini 400 Social index and Garz, Volk and Gilles (2002) who analyse the Dow Jones Sustainability Index (DJSI) for Europe. The studies on the Domini 400 Social index find a similar performance compared to the benchmark index. According to DiBartolomeo and Kurtz (1999) a small (but not significant) out-performance of the Domini 400-index is attributable to the specific risk exposures of the index, for example, a higher beta compared to the market. For the DJSI STOXX index, Garz et al. (2002) find a slightly significant out-performance compared to the DJ STOXX600 index.

Compared to the aforementioned studies on SRI equity indexes, our study extends the research in two respects. First, we analyse 29 SRI equity indexes and therefore the most important SRI indexes world-wide are covered, and second, we not only use single-equation regressions for the estimation of the performance but also spanning tests and multi-equation tests for specific groups of indexes. The latter also leads to the use of the cross-section of SRI indexes, which improves the quality of the statistical tests.

The 29 SRI equity indexes represent different SRI screening procedures and cover different investment regions (global, Europe, and specific countries such as Australia, Sweden, UK and U.S.). This leads to relatively general conclusions regarding the effects of SRI screening on the performance.

The paper is organised as follows, in section 2 we provide a detailed description of the data, i.e. the SRI equity indexes and the benchmarks. Section 3 is the main part of the study and contains the research methodology and discusses the results of the regression estimates and tests. This section covers the analysis of the performance and risk of the SRI equity indexes, both for single indexes and for index groups. Section 4 concludes. The appendix contains a table which describes the SRI equity indexes and the main characteristics of the specific SRI screens.

2 Description of the Data

The study investigates the characteristics of 29 international SRI equity indexes.⁴ These indexes have been constructed and published by 11 different suppliers. Included are the well-known SRI index families Dow Jones Sustainability Indexes (DJSI), Ethical, FTSE4Good, Humanix and KLD. Table 1 in the appendix provides detailed information on the supplier company, the investment universe, the SRI selection criteria, the available length of the index time series as well as the relevant benchmark indexes.

From the DJSI index family, six indexes are included: the World, STOXX and EUROSTOXX index and the same indexes “ex AGTF”, i.e. without companies engaged in alcohol, gambling, tobacco and firearms. The FTSE4Good index family consists of the following eight indexes, which are all considered in this study, the four tradable indexes Europe50, Global100, US100 and UK50 as well as the broader indexes Europe, Global, US and UK from which the tradable indexes are derived.

Three SRI equity indexes launched by KLD Research & Analytics have been chosen for the analysis, which include the well-known Domini 400 Social index, the Broad Market Social index and the Large Cap Social index. The Italian company E.Partners has constructed a Eurozone and a globally diversified SRI equity index, which form the Ethical index group. The Swedish Humanix Holding has launched the four

⁴ The index time series have been supported in most cases by the suppliers of the indexes themselves, some of the SRI index series and the benchmark indexes have been taken from the databases of Thomson Financial Datastream and Ecwin.

indexes for Europe, Sweden, U.S. and the World.⁵ The other six SRI equity indexes considered in this study are ASPI, Calvert Social Index, Jantzi Social Index, Kempen SNS Smaller European index, Naturaktienindex and Westpac-Monash Eco-Index.

These 29 SRI indexes cover different international investment areas. Seven indexes have a global investment universe and ten cover European stocks, of which four concentrate on the euro area. The other 12 indexes contain stocks of single countries (number of indexes in brackets): Australia (1), Canada (1), Sweden (1), United Kingdom (2) and the United States (7).

A few other SRI equity indexes also exist such as the Ethibel index family, which has not been included in the study. The reason is that the time series of these indexes are still too short to allow sound econometric inference.

Almost all indexes are performance indexes, i.e. they express the total return to the investor and cover therefore not only the changes of the stock prices but also all payments to the investor such as dividend payments. The Naturaktienindex (NAI) is the only price index. All indexes are converted into U.S. dollars.

In most cases, i.e. for 20 out of the 29 SRI indexes, the benchmark indexes are the official benchmarks chosen by the suppliers of the SRI indexes. In the other nine cases, the benchmark index has been selected in such a way that the investment universe of the SRI index is well and closely approximated. Column 6 of Table 1 in the appendix gives an overview of the relevant benchmark indexes.

The so called “small cap bias”, i.e. the relatively high investment weight of stocks with a low market capitalisation, which has been found in several studies⁶, only has minor implications for the analyses in our study. The reason is that most of the SRI equity indexes concentrate on stocks with a large market capitalisation and aim to sufficiently represent the market capitalisation of the stock market.⁷ For this reason and the fact that the composition of the SRI indexes are only infrequently adjusted and do not officially follow specific investment styles, lead to the conclusion that it is also not necessary to use multi-factor models, such as the Fama-French (1993)-model or a four-factor model (Carhart (1997)). Instead, a single-factor model with a benchmark index selected with regard to a close approximation of the investment universe of each SRI index should be sufficient. The relatively high values for the adjusted R^2 of the linear regressions of the index returns against the benchmark returns, which are in most cases above 90% (see table 3 in section 3.1), also support this view.

There are only two exceptions: (1) For the Domini 400 index the S&P 500 index only serves as a starting point for the stock selection process. More than half of the

⁵ The Humanix indexes are discontinued by the supplier company from January 2004 on.

⁶ See e.g. Luther, Matatko and Corner (1992), Schroeder (2004) and Bauer, Koedijk and Otten (2005).

⁷ The Kempen SNS Smaller European SRI index concentrates on stocks with low market capitalisation. The HSBC Smaller European index is the official benchmark index and is used in our analysis.

companies included in the Domini 400 index are not part of the S&P 500 index and, consequently, have a smaller market capitalisation than the companies of the S&P 500 index.⁸ A significant weighting on small cap stock indexes has also been found in earlier studies.⁹ Therefore, the performance of the Domini 400 index is also analysed (in section 3.1) using additional U.S. benchmark indexes - a small cap index and two style indexes. These additional indexes are the S&P 600 Small Cap index and the S&P 500 Growth and Value indexes.

(2) The Naturaktienindex (NAI) contains only 25 stocks selected from a world-wide investment universe. Therefore it is difficult to find an appropriate world-wide diversified benchmark index. The MSCI World index is not convincing since the R^2 of a linear regression on the index returns is only 32.6%. But the use of other or additional benchmark indexes does not improve this result significantly. To check the robustness of the results for the single benchmark index, performance regressions are conducted using the Small Cap index and the Growth and Value indexes of MSCI¹⁰ with a world-wide diversification as additional benchmarks (see section 3.1).

In addition, robustness checks using these three MSCI indexes are carried out in section 3.2 for specific index groups that cover most of the SRI indexes.

Table 2 provides an overview of the time series characteristics, i.e. the mean excess returns, standard deviations and the performance of the SRI indexes. All returns are calculated as first differences of the monthly times series in logarithms. The performance is calculated using the Sharpe ratio (SR). The Sharpe ratio measures the return above the risk-free interest rate (= excess return) divided by the total risk of the investment:

$$(1) \quad SR_i = \frac{\mu_i - r_f}{\sigma_i}$$

μ = annualised mean logarithmic return, r_f = risk-free interest rate (U.S. 1-month interbank offered rate), σ = standard deviation of the logarithmic returns. The mean, standard deviation and Sharpe ratio of the indexes are compared to those of an appropriate benchmark index (see Table 1 in the appendix) for the same period. The calculations use the full available data history of each SRI index.

The comparisons show that in 17 out of 29 cases the mean excess return is higher for the SRI index. Since a higher mean return might only be the result of a higher risk exposure, the risk-adjusted returns should also be compared. Using the Sharpe ratio to measure the risk-adjusted return even improves the results for the SRI indexes. For 18 SRI indexes, the Sharpe ratio is higher than for the relevant benchmark index, in ten cases the opposite is true and in one case the Sharpe ratios are equal.

⁸ See http://www.domini.com/Social-Screening/creation_maintenance.doc_cvt.htm

⁹ See e.g. DiBartolomeo and Kurtz (1999)

¹⁰ MSCI is the abbreviation of Morgan Stanley Capital International.

Table 2: Index Performance versus Benchmark Performance: Overview

<i>Abbreviations</i>	<i>Start Date</i> ¹	<i>Mean</i> ²	<i>Mean BM</i> ²	<i>Standard Deviation</i>	<i>Standard Dev. BM</i>	<i>Sharpe Ratio</i> ³	<i>Sharpe Ratio BM</i> ³
Aspi	1/1992	0.064	0.049	0.657	0.635	0.097	0.077
Calv	7/2000	-0.114	-0.087	0.732	0.631	-0.155	-0.139
DJSI1	1/1994	0.044	0.024	0.566	0.523	0.078	0.045
DJSI2	1/1994	0.044	0.024	0.571	0.523	0.077	0.045
DJSI3	1/1999	-0.041	-0.043	0.665	0.644	-0.062	-0.066
DJSI4	1/1999	-0.044	-0.043	0.676	0.644	-0.065	-0.066
DJSI5	1/1999	-0.053	-0.048	0.884	0.780	-0.060	-0.062
DJSI6	1/1999	-0.054	-0.048	0.887	0.780	-0.061	-0.062
Eth1	11/2000	-0.074	-0.046	0.814	0.734	-0.091	-0.062
Eth2	2/2001	-0.067	-0.066	0.645	0.631	-0.104	-0.104
FT1	8/1996	0.053	0.038	0.685	0.615	0.078	0.062
FT2	8/1996	0.044	0.038	0.660	0.615	0.067	0.062
FT3	8/1996	0.048	0.021	0.646	0.569	0.075	0.037
FT4	8/1996	0.043	0.021	0.632	0.569	0.068	0.037
FT5	8/1996	0.049	0.029	0.530	0.520	0.093	0.055
FT6	8/1996	0.026	0.029	0.505	0.520	0.052	0.055
FT7	8/1996	0.054	0.043	0.692	0.606	0.078	0.071
FT8	8/1996	0.049	0.043	0.697	0.606	0.070	0.071
Hu1	7/2001	0.062	0.026	1.013	1.112	0.061	0.023
Hu2	7/2001	-0.080	-0.057	0.763	0.629	-0.105	-0.090
Hu3	7/2001	-0.028	-0.004	0.868	0.730	-0.032	-0.006
Hu4	7/2001	-0.091	-0.036	0.727	0.610	-0.125	-0.060
Jantzi	2/2000	0.010	-0.007	0.740	0.767	0.014	-0.008
Ke	1/1999	0.022	0.015	0.852	0.683	0.026	0.021
KLD1	6/1990	0.074	0.063	0.554	0.522	0.133	0.121
KLD2	2/2001	-0.064	-0.055	0.721	0.645	-0.089	-0.085
KLD3	2/2001	-0.074	-0.062	0.724	0.644	-0.102	-0.097
NAI	5/1997	0.087	-0.012	0.802	0.590	0.108	-0.020
West	2/1999	0.056	0.079	0.719	0.655	0.078	0.121

Notes: ¹Time series begin at the indicated start date and end in December 2003. ²Annualised mean excess return of the SRI index and the benchmark index (*BM*), respectively. ³Sharpe ratios of the SRI index and the benchmark index, respectively. Excess returns and Sharpe ratios are calculated using the 1-month U.S. interbank offered rate as risk-free interest rate. Returns are calculated as logarithmic differences to the month before. All series are denominated in U.S. dollars. The full names of the SRI indexes and the corresponding benchmark indexes can be found in Table 1 in the appendix.

For 15 SRI indexes both the mean excess returns and the Sharpe ratios out-perform the relevant benchmark index. These results indicate that more than half of the SRI indexes have the potential to compete with conventional stock indexes.

In the next section linear regressions are used to test the significance of performance differences and to provide a more general adjustment for risk.

3 Empirical Characteristics of SRI Equity Indexes

The aim of this chapter is to investigate the differences of the time series characteristics between SRI equity indexes and their conventional benchmarks. This includes performance comparisons as well as the measurement of relative risk. In addition, spanning tests are applied to test whether the SRI indexes can be replicated by their benchmarks.

There are different types of benchmarks applied to measure relative performance, which are (1) a single benchmark index specifically selected in order to closely approximate the investment universe of each SRI equity index, (2) an extension of these tests using a small cap as well as two style indexes (growth, value) applied to the Naturaktienindex (NAI) and the Domini 400-index, as for these two SRI indexes a single benchmark might not be sufficient, and (3) additional tests using the same four world-wide diversified MSCI indexes as benchmarks for every SRI index, i.e., the World index, the World Small Cap index and the World Growth and Value indexes.

The tests for out-/under-performance, for differentials in risk exposure and for spanning are separately conducted for all SRI indexes (section 3.1) and for groups of SRI indexes (section 3.2.). The tests for index groups aim to increase the quality of the single-equation tests by using information from the cross-section of SRI indexes.

3.1 Performance, Risk and Spanning-Tests for Single SRI Indexes

The basic equation used to measure the relative performance of the SRI indexes is the linear regression of the excess returns of the benchmark index ($r_{i,t}^{BM}$) on the excess returns of the SRI index ($r_{i,t}^{SRI}$):

$$(2) \quad r_{i,t}^{SRI} = \alpha_i + \beta_i r_{i,t}^{BM} + \varepsilon_{i,t}$$

The performance of the SRI indexes is estimated by the so called Jensen's alpha (α_i), i.e. the extra-return that is not explained by the risk exposure with respect to the benchmark index. The β_i -coefficient is used to compare the relative risk of the SRI index. As in the Capital Asset Pricing Model (CAPM) a $\beta_i > 1$ indicates that the risk of the SRI index is higher compared to the benchmark because a benchmark return of one would translate into a return of the SRI index, which is larger than one. For $\beta_i < 1$ the SRI index has a lower risk compared to the benchmark.

A test of the joint hypothesis $H_0: (\alpha_i = 0 \text{ and } \beta_i = 1)$ is equal to a so called spanning test.¹¹ If the null hypothesis of spanning is not rejected this means that the SRI index can be replicated by the benchmark index: $r_{i,t}^{SRI} = r_{i,t}^{BM} + \varepsilon_{i,t}$. In this case investing in the benchmark index is on average equivalent to investing in the SRI index, without differences in return or risk.

Compared to the case of performance measurement for investment funds, equation (2) is relatively simple. To estimate the performance of investment funds it is necessary to consider market timing in order to avoid biased estimates of Jensen's alpha and to use a set of different benchmark indexes for the approximation of the specific investment style of each fund. In addition, it would be useful to conduct conditional estimations of the parameters using the publicly available information set of the portfolio management (e.g. concerning macroeconomic developments) as instruments.¹² Such methodological problems hardly exist in the case of SRI indexes as there is no active portfolio management and the investment universe can be in most cases approximated very well by a single benchmark index.

Equation (2) is estimated using ordinary least squares. The variance-covariance matrix of the residuals is corrected for autocorrelation and heteroskedasticity using the Newey-West approach.¹³ Table 3 gives the results for the estimated parameters and the parameter tests.

For some indexes the available time series starts before the official launch date. In these cases the index has been calculated backwards by the supplier of the index. Usually this calculation is done by applying the index composition of the starting date. To consider a so called "backward-looking bias", i.e. a better performance due to information on the future development of stock prices that was not yet available in earlier periods, we additionally estimate the performance using only the period from the official inception date on. These additional results for estimated parameters and tests are given in brackets below the results for the total period.

Column 4 contains the estimated values for parameter α_i . The rejection of the null hypothesis $H_0: \alpha_i = 0$ is indicated by asterisks. The results show that Jensen's alpha, i.e. the relative risk-adjusted performance, is in almost all cases not significantly different from zero. This is a clear indication that the performance of the SRI stock indexes do not deviate systematically from their direct conventional counterparts.

There are only few exceptions. The Westpac-Monash index performed significantly worse than the benchmark index. The ASPI index exhibits a positive out-performance for the total available index time series, which is slightly significant at the 10% level. Interestingly, the performance during the much shorter time period from the official

¹¹ See Huberman and Kandel (1987).

¹² For the consideration of market timing in performance measurement see Treynor and Mazuy (1966) and Admati and Ross (1985). Conditional performance estimation is described in Ferson and Schadt (1996).

¹³ See Newey and West (1987).

launch date on is not significantly different from the benchmark index. The FTSE4Good indexes for Europe show a significantly negative alpha when estimated from the official launch date on, but not for the total time series.

There is clear evidence of a backward-looking bias. In fact, in every case when the available index time series has been calculated backwards the estimated alpha is higher compared to the estimated performance when the data from the official inception date on are used. In 7 out of 15 cases (DJSI1 and DJSI2, FT1, FT2,..., FT5) this difference is significant at least at the 10% level.

Column 5 shows the results for the β -coefficients and for the test of $H_0: \beta_i = 1$. The estimated values can be interpreted as a measure of risk relative to the benchmark index. For many SRI indexes the estimated β is higher than one. When the longest available sample period is considered, 19 out of the 29 SRI indexes can be characterised by a statistically significant high relative risk. There are only few SRI indexes that exhibit a relatively low risk. β -coefficients that are significantly below one are estimated only for the NAI, the Humanix index Europe and the FTSE4Good index UK (FT6).

In the last column of Table 3 the outcome for the spanning tests $H_0: (\alpha_i = 0 \text{ and } \beta_i = 1)$ is documented. When the total available data history is considered, there are eight SRI indexes for which spanning cannot be rejected. These indexes include the DJSI STOXX (including and ex AGTF), the FTSE4Good UK (tradable and broad index), the Ethibel indexes Global and Euro, the Jantzi Social index, and the Kempen SNS Smaller European index. When only the data sample from the official inception date is used then spanning is also not rejected for the ASPI, the DJSI World indexes (including and ex AGTF) and the FTSE4Good US 100 index.

Table 3: Index Performance versus Benchmark Performance: Regression-based Tests

<i>Abbreviation</i>	<i>Start Date</i> ¹	<i>Adjusted R²</i>	<i>Alpha</i> $H_0: \alpha_i = 0$	<i>Beta</i> $H_0: \beta_i = 1$	<i>Spanning Test</i> $H_0: (\alpha_i = 0 \text{ and } \beta_i = 1)$
Aspi	1/1992 (8/2001)	0.977 (0.989)	0.014* (-0.004)	1.024 (1.041)	** (not rejected)
Calv	7/2000	0.973	-0.014	1.144***	***
DJSI1	1/1994 (10/1999)	0.929 (0.940)	0.019 (-0.014)	1.044* (1.062*)	** (not rejected)
DJSI2	1/1994 (10/1999)	0.925 (0.933)	0.019 (-0.016)	1.051* (1.073*)	** (not rejected)
DJSI3	1/1999 (11/2001)	0.966 (0.972)	0.002 (-0.011)	1.014 (1.031)	not rejected (not rejected)
DJSI4	1/1999 (11/2001)	0.964 (0.970)	0.000 (-0.013)	1.029 (1.054*)	not rejected (not rejected)
DJSI5	1/1999 (11/2001)	0.967 (0.982)	0.001 (-0.021)	1.116*** (1.151***)	*** (***)
DJSI6	1/1999 (11/2001)	0.967 (0.982)	-0.000 (-0.023)	1.119*** (1.166***)	*** (***)
Eth1	11/2000	0.944	0.0075	1.072	not rejected
Eth2	2/2001	0.986	-0.0005	1.014	not rejected

Table 3 continued					
<i>Abbreviation</i>	<i>Start Date</i> ¹	<i>Adjusted R²</i>	<i>Alpha</i> <i>H₀: $\alpha_i=0$</i>	<i>Beta</i> <i>H₀: $\beta_i=1$</i>	<i>Spanning Test</i> <i>H₀: ($\alpha_i=0$ and $\beta_i=1$)</i>
FT1	8/1996 (8/2001)	0.941 (0.967)	0.012 (-0.041**)	1.082*** (1.052)	** (**)
FT2	8/1996 (8/2001)	0.972 (0.986)	0.004 (-0.027**)	1.058*** (1.053*)	*** (***)
FT3	8/1996 (8/2001)	0.930 (0.957)	0.026 (-0.031)	1.094*** (1.061)	*** (*)
FT4	8/1996 (8/2001)	0.967 (0.988)	0.020 (-0.010)	1.091*** (1.089***)	*** (***)
FT5	8/1996 (8/2001)	0.902 (0.960)	0.021 (-0.017)	0.969 (0.978)	not rejected (not rejected)
FT6	8/1996 (8/2001)	0.938 (0.979)	-0.000 (-0.013)	0.941** (0.983)	not rejected (not rejected)
FT7	8/1996 (8/2001)	0.893 (0.958)	0.0073 (-0.014)	1.079** (1.057)	* (not rejected)
FT8	8/1996 (8/2001)	0.906 (0.957)	0.0012 (-0.007)	1.095*** (1.086*)	*** (*)
Hu1	7/2001	0.964	0.039	0.894***	***
Hu2	7/2001	0.953	-0.013	1.186***	***
Hu3	7/2001	0.959	-0.023	1.165***	***
Hu4	7/2001	0.936	-0.049	1.154**	**
Jantzi	2/2000	0.807	0.016	0.868	not rejected
Ke	1/1999 (10/2003)	0.781 (not calc.)	0.006 (not calc.)	1.105* (not calc.)	not rejected (not calculated)
KLD1	6/1990	0.967	0.008	1.044***	***
KLD2	2/2001	0.984	-0.003	1.108***	***
KLD3	2/2001	0.982	-0.004	1.115***	***
NAI	5/1997	0.326	0.096	0.786**	*
West	2/1999	0.959	-0.028*	1.075**	***

Notes: ¹The sample period begins at the indicated start date and ends December 2003. The official launch date is given in brackets if it differs from the start date of the time series. In these cases the available index time series has partly been backcasted by the supplier of the index. All estimations and tests have been done both for the total available time series and for the period from the official launch date on (in brackets). The estimations are based on equation (2) and use excess returns, denominated in U.S. dollars. The excess returns are calculated using the 1-month U.S. interbank offered rate as risk-free interest rate. The full names of the SRI indexes and the corresponding benchmark indexes can be found in Table 1 in the appendix. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

In these cases an investor who is only interested in the financial outcome of the investment and not in the SRI screen can equally invest in the benchmark indexes. Put differently, for an investor primarily interested in socially responsible investments this means that he or she does not have to expect any differences in risk or return compared to the benchmark.

For the other indexes spanning is rejected. In most cases the reason for the rejection is a difference in relative risk compared to the benchmark. These SRI indexes exhibit (with only few exceptions) the same risk-adjusted return but a higher β -coefficient.

The relatively high values for the adjusted R^2 (column 3), which are for most SRI indexes above 90%, show that the majority of SRI indexes can be largely approximated by the benchmark indexes.

Only in the case of the NAI is the R^2 (=32.6%) relatively low. Therefore an extended regression test is conducted for this index using two additional factors, the returns of a small cap stock index ($r_{i,t}^{SC}$) and the return differential between a growth and a value portfolio ($r_{i,t}^{GV}$). As explained in section 2 it is also worthwhile to do this extended analysis for the Domini 400-index. For these two SRI indexes the following equation is estimated:

$$(3) \quad r_{i,t}^{SRI} = \alpha_i + \beta_{1i} r_{i,t}^{BM} + \beta_{2i} \varepsilon_{i,t}^{SC} + \beta_{3i} r_{i,t}^{GV} + \varepsilon_{i,t}$$

Since the returns of the small cap index are highly correlated with those of the single benchmark index¹⁴, the small cap index returns are orthogonalised with respect to the benchmark returns:

$$(3a) \quad r_{i,t}^{SC} = \kappa_i + \delta_i r_{i,t}^{BM} + \varepsilon_{i,t}^{SC}$$

Equation (3) is estimated using the residuals of equation (3a). The residuals represent the part of the small cap index returns which cannot be explained by the benchmark returns. The growth-value factor is constructed as the difference between the returns of a growth- and a value-portfolio:

$$(3b) \quad r_{i,t}^{GV} = r_{i,t}^{Growth} - r_{i,t}^{Value}$$

The growth portfolio contains equities with a low book-to-market value whereas the value portfolio concentrates on stocks with a high book-to-market value. Equation (3) therefore contains, in addition to the market factor, the two factors “size” and “book-to-market value”. These so called style factors are commonly used in performance measurement for investment funds and trace back to the seminal papers of Fama and French (as e.g. from 1993 and 1996).

Table 4 compares the results obtained in the single-factor model (equation (2)) with those of the three-factor model expressed by equation (3). As expected, the adjusted R^2 increases when the two factors are added. But in the case of the NAI, a large fraction of the variance still remains unexplained. For both SRI indexes the estimate of Jensen’s alpha is also in equation (3) not significantly different from zero and even declines slightly compared to the results of equation (2). For the NAI it is a rather astonishing outcome that the estimated annual out-performance of 9.6% (single factor model) or 8.6% (three factor model) is not significant. This is mainly due to the short

¹⁴ In the case of the Domini 400 index, the correlation is 72.6% and for the NAI the correlation amounts to 80%.

data series and strong changes in the relative performance: in 1999 and 2000 the NAI exhibited a clear under-performance compared to the MSCI World index and from mid-2001 on an increasing out-performance.

Table 4: Multifactor Model for the Domini 400-Index and the NAI-Index (Equation (3))

	<i>Domini 400 (KLD1)</i>		<i>NAI</i>	
α , ($H_0: \alpha=0$)	0.008	0.005	0.096	0.086
Market (β_1), ($H_0: \beta_1=1$)	1.044***	1.019	0.786**	0.844*
Small Cap Index (β_2), ($H_0: \beta_2=0$)	--	-0.003	--	0.454***
Growth – Value (β_3), ($H_0: \beta_3=0$)	--	0.123***	--	-0.527***
<i>Adjusted R²</i>	0.967	0.977	0.326	0.429
<i>Spanning-Test</i> , $H_0: \alpha=0$ and $(\beta_1+\beta_2+\beta_3=1)$	***	***	*	not rejected
Start of Sample Period	6/1990	6/1990	5/1997	5/1997

Notes: All estimations and tests have been done for the total available time series from the individual start date until December 2003. The estimations are based on (excess) returns, denominated in U.S. dollars. The excess returns are calculated using the 1-month U.S. interbank offered rate as risk-free interest rate. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

The parameters of the style factors show that the Domini 400 index is tilted towards growth stocks but the small cap returns are not significant. The small cap bias, which has been found for this index in earlier studies, therefore does not seem to prevail. Instead the NAI has a significant weight on value stocks and on small cap stocks. Interestingly, the spanning test for the NAI can no longer be rejected (p-value = 25.9%). Thus, the returns of the NAI can be replicated by the estimated linear combination of the three factors.

In sum, the results of the SRI stock indexes using equations (2) and (3) show that most of the indexes exhibit no out- or under-performance compared to the benchmark. For quite a few SRI indexes even the spanning test is not rejected. Combined with the high estimates for the adjusted R^2 it can be concluded, that most of the SRI stock indexes only slightly deviate from their benchmark. If such a deviation is significant it is in most cases due to a β -coefficient different from one.

Since many of the SRI stock indexes have been created only a few years ago only relatively short data series are available. This reduces the quality of the parameter tests. The resulting wide confidence intervals for the estimated parameters may be one reason for the large number of insignificant alphas. In the next section, therefore, groups of SRI indexes are also analysed. Using the cross-section of indexes in addition to the single time series increases the information content of the estimates and should improve the tests.

3.2 Performance, Risk and Spanning-Tests for Index Groups

In this section the same tests as in section 3.1 are also conducted using the information in the cross-section of the SRI index series by estimating a system of equations. This should lead to an improved performance of the parameter tests and is aimed to overcome - at least to some part - the restrictions that stem from the relatively short data history of most of the SRI indexes.

Instead of equation (2) the following system of n equations is analysed:

$$(4) \quad \begin{aligned} r_{1,t}^{SRI} &= \alpha_1 + \beta_1 r_{1,t}^{BM} + \varepsilon_{1,t} \\ &\dots \\ r_{n,t}^{SRI} &= \alpha_n + \beta_n r_{n,t}^{BM} + \varepsilon_{n,t} \end{aligned}$$

Equations (4) are estimated using the seemingly unrelated regressions (SUR) approach. Since the regressors $r_{i,t}^{BM}$ are different across equations there should be efficiency gains in applying a SUR estimation.¹⁵

Tables 5 and 6 show the results of the tests for different groups of SRI indexes. In table 5 the groups are constructed with regard to the length of the available time series. The following three groups are defined according to three different time intervals, including a long time period (1/1994-12/2003), a medium-term (5/1997-12/2003) and a short-term period (1/1999-12/2003). All SRI indexes, for which the index time series is at least as long as the period under consideration, are included in this group.¹⁶ To form the groups part of the time series before the official launch date are also used for some of the SRI index series. This may lead to an upward bias of the estimates of Jensen's alpha for these indexes. Therefore, from this point of view the estimates and tests of this section refer to an upper bound of the performance of the SRI stock indexes.

Column 3 of table 5 shows the values of the Chisquare-test statistic for the joint tests of Jensen's alpha, $H_0 : (\alpha_i = 0, \forall i : i = 1, \dots, n)$. The results for the joint test $H_0 : (\beta_i = 1, \forall i : i = 1, \dots, n)$ are given in column 4 and the last column refers to the joint spanning tests. For all tests Newey-West (1987)-corrected standard errors are used.

The results confirm the findings from the single-equations, which state that the estimates of alpha are not significantly different from zero, whereas the β -coefficients deviate significantly from one. The joint spanning tests are rejected for the short and the long period but not for the medium-term estimates.

¹⁵ See e.g. Greene (2000, section 15.4.1).

¹⁶ Only those SRI indexes are excluded which are part of the same index family and have a very similar investment universe compared to another index in the same index family. Therefore, for example, only the tradable FTSE4Good US (FT7) is considered but not the broader and very similar US-index (FT8). The same rule is applied to the DJSI indexes where the ex AGTF indexes are not included in the groups. This avoids that these two index families get an excessively strong weight in the tests.

Table 5: Joint Coefficient-Tests for Different Time Intervals using a Single Benchmark Index

<i>Period</i>	<i>Included Indexes</i>	$H_0: \text{all } \alpha_i=0$	$H_0: \text{all } \beta_i=1$	<i>Spanning</i> $H_0: \text{all } \alpha_i=0$ <i>and all } \beta_i=1</i>
1/1999 – 12/2003	12 Indexes: FT1, FT3, FT5, FT7, Ke, KLD1, DJSI1, DJSI3, DJSI5, Aspi, Westpac, NAI	14.32 (0.281)	137.85*** (0.000)	152.11*** (0.000)
5/1997 – 12/2003	8 Indexes: FT1, FT3, FT5, FT7, KLD1, DJSI1, Aspi, NAI	3.70 (0.883)	19.06** (0.015)	22.82 (0.119)
1/1994 – 12/2003	3 Indexes: KLD1, DJSI1, Aspi	4.02 (0.259)	7.50* (0.057)	12.23* (0.057)

Notes: The H_0 are tested by Chisquared-tests. Degrees of freedom (*dof*) are equal to the number of included indexes in the case of a single parameter test, and for the joint hypothesis *dof* are equal to 2*number of indexes. The figures show the value of the test statistics, p-values are given in brackets. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

The three groups are relatively heterogeneous and include indexes that differ in at least three respects, the SRI screening process, the investment universe and the time period. Thus, these tests give representative results for a wide range of different SRI equity indexes.

In Table 6 the groups are formed by index families. In contrast to the results shown in Table 5 now the groups are relatively homogeneous. The indexes within a group only differ regarding the investment universe, but the SRI screening process and the time intervals are equal. Thus, this analysis concentrates on the effects of a specific SRI screening process represented by the SRI indexes of the same index family. The estimations refer to the longest possible common period for which data of the included indexes are available.

Table 6: Joint Coefficient-Tests for Index Families Using the a Single Benchmark Index

<i>Index Group</i>	<i>Included Indexes</i>	<i>Period</i>	$H_0: \text{all } \alpha_i=0$	$H_0: \text{all } \beta_i=1$	<i>Spanning</i> $H_0: \text{all } \alpha_i=0$ <i>and all } \beta_i=1</i>
FTSE4Good	8 Indexes: FT1, FT2, FT3, FT4, FT5, FT6, FT7, FT8	8/1996-12/2003	8.08 (0.426)	51.16*** (0.000)	60.15*** (0.000)
Dow Jones Sustainability	6 Indexes: DJSI1, DJSI2, DJSI3, DJSI4, DJSI5, DJSI6	1/1999-12/2003	1.50 (0.960)	102.03*** (0.000)	104.9*** (0.000)
Humanix	4 Indexes: Hu1, Hu2, Hu3, Hu4	7/2001-12/2003	7.15 (0.128)	37.81*** (0.000)	45.17*** (0.000)
KLD	3 Indexes: KLD1, KLD2, KLD3	1/2001-12/2003	1.81 (0.613)	99.67*** (0.000)	107.39*** (0.000)
Ethical	2 Indexes: Eth1, Eth2	2/2001-12/2003	0.48 (0.789)	3.71 (0.157)	4.39 (0.356)

Notes: The H_0 are tested by Chisquared-tests. Degrees of freedom (*dof*) are equal to the number of included indexes in the case of a single parameter test, and for the joint hypothesis *dof* are equal to 2*number of indexes. The figures show the value of the test statistics, p-values are given in brackets. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

For all five index families the estimated α_i are not significantly different from zero in the joint tests. The test on $H_0: \beta_i=1$ and the spanning tests are not rejected for the two Ethical indexes only. This means that the different SRI screening processes are comparable insofar as they do not lead to a different performance compared with the relevant benchmark index. Since spanning is rejected for most of the index families these SRI indexes cannot be directly replicated by their benchmarks.

To summarize, the joint significance tests confirm the results of the single-equation tests. The results show that the SRI stock indexes exhibit on average no out- or under-performance compared to their benchmarks and deviations can be attributed to differences in relative risk.

Since the question of an out- or under-performance is very important for the assessment of the SRI stock indexes and the underlying SRI screening process, additional tests of the robustness of the results are conducted. In the first setting equation system (4) is used again but the MSCI World equity index now serves as the single benchmark for all SRI indexes.

In the second setting two additional factors are included in the equations, a small cap factor and a growth-value factor. These two factors are constructed using the World Small Cap index as well as the World Growth and Value indexes of MSCI. The following system of equations is estimated:

$$(5) \quad \begin{aligned} r_{1,t}^{SRI} &= \alpha_1 + \beta_{11}r_t^{BM} + \beta_{21}\varepsilon_t^{SC} + \beta_{31}r_t^{GV} + \varepsilon_{1,t} \\ \dots \\ r_{n,t}^{SRI} &= \alpha_n + \beta_{1n}r_t^{BM} + \beta_{2n}\varepsilon_t^{SC} + \beta_{3n}r_t^{GV} + \varepsilon_{n,t} \end{aligned}$$

The procedure is the same as described for the equations (3), (3a) and (3b) in section 3.1. BM indicates the returns of the MSCI World index, the small cap factor (SC) is orthogonalised with respect to the World index (5a) and the growth-value factor is the difference between the returns of the Growth and the Value index (5b).

$$(5a) \quad r_t^{SC} = \kappa_i + \delta_i r_t^{BM} + \varepsilon_t^{SC}$$

$$(5b) \quad r_t^{GV} = r_t^{Growth} - r_t^{Value}$$

Table 7 shows the test results for the same index groups as in Table 5. Interestingly, the estimates for the group that consists of the indexes, Domini 400 (KLD1), DJSI World (DJSI1) and ASPI, exhibit a significant alpha at the 10% level. A look at the single equations reveals that this stems from an out-performance of the Domini 400-index. But when three factors are used (equations (5)) no significant performance difference can be found for all three groups.

Table 8 provides the estimates of Jensen's alpha for the same index family groups as Table 6. A significant alpha is found for the Humanix indexes when only a single factor is used. The estimates of the single equations for this group show that this performance differential is due to a significant under-performance of the Humanix

World index (Hu4). But when the three-factor model is estimated no significant performance difference exists for any of the five groups.

Table 7: Joint Coefficient-Tests for Different Time Intervals Using World-Wide Benchmarks

<i>Period</i>	<i>Included Indexes</i>	<i>Single World-Index: H₀: all $\alpha_i=0$</i>	<i>Three World-Indexes: H₀: all $\alpha_i=0$</i>
1/1999 – 12/2003	12 Indexes: FT1, FT3, FT5, FT7, Ke, KLD1, DJSI1, DJSI3, DJSI5, Aspi, Westpac, NAI	7.866 (0.796)	9.28 (0.678)
5/1997 – 12/2003	8 Indexes: FT1, FT3, FT5, FT7, KLD1, DJSI1, Aspi, NAI	6.71 (0.568)	7.80 (0.453)
1/1994 – 12/2003	3 Indexes: KLD1, DJSI1, Aspi	6.46* (0.091)	5.17 (0.160)

Notes: The H_0 are tested by Chisquared-tests. Degrees of freedom (*dof*) are equal to the number of included indexes in the case of a single parameter test, and for the joint hypothesis *dof* are equal to 2*number of indexes. The figures show the value of the test statistics, p-values are given in brackets. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

Table 8: Joint Coefficients-Tests for Index Families Using World-Wide Benchmarks

<i>Index Group</i>	<i>Included Indexes</i>	<i>Period</i>	<i>Single World-Index: H₀: all $\alpha_i=0$</i>	<i>Three World-Indexes: H₀: all $\alpha_i=0$</i>
FTSE4Good	8 Indexes: FT1, FT2, FT3, FT4, FT5, FT6, FT7, FT8	8/1996-12/2003	12.26 (0.140)	8.19 (0.415)
Dow Jones Sustainability	6 Indexes: DJSI1, DJSI2, DJSI3, DJSI4, DJSI5, DJSI6	1/1999-12/2003	1.08 (0.982)	2.21 (0.900)
Humanix	4 Indexes: Hu1, Hu2, Hu3, Hu4	7/2001-12/2003	8.91* (0.064)	6.43 (0.169)
KLD	3 Indexes: KLD1, KLD2, KLD3	1/2001-12/2003	3.74 (0.290)	2.64 (0.451)
Ethical	2 Indexes: Eth1, Eth2	2/2001-12/2003	0.01 (0.997)	0.799 (0.671)

Notes: The H_0 are tested by Chisquared-tests. Degrees of freedom (*dof*) are equal to the number of included indexes in the case of a single parameter test, and for the joint hypothesis *dof* are equal to 2*number of indexes. The figures show the value of the test statistics, p-values are given in brackets. *, **, *** = H_0 rejected at a significance level of 10%, 5% or 1%, respectively. Newey/West (1987)-corrected standard errors have been applied.

To sum up, the robustness checks confirm the results of the estimations using single equations or equations systems. SRI stock indexes exhibit no out- or under-performance compared to conventional benchmark indexes. Nevertheless, differences in the risk-return characteristics exist. But these differences mainly stem from a higher relative risk.

4 Conclusion

Socially responsible investments (SRI) in equities are no longer a negligible segment of international capital markets. For investors it is interesting to know whether equities selected by an SRI screening process exhibit a different performance than conventional investments in stocks.

In contrast to other studies on the performance of SRI equities, this study concentrates on SRI indexes and not on investment funds. This has the advantage that the effects of the SRI screening process on the risk-return characteristics can be studied directly without the need to filter out the transaction costs of the funds, the management skills or the timing activities of the fund management. Since the SRI stock indexes apply SRI screening processes that are comparable to those of the investment funds, the results of this study are also relevant for the assessment of the performance potential of SRI mutual funds.

Throughout the study different settings for performance tests are applied to the 29 SRI stock indexes, including (1) single-factor models with conventional benchmark indexes that closely approximate the investment universe of the SRI indexes, (2) three-factor models for two specific SRI indexes (Naturaktienindex, Domini 400 index), (3) multi-equation models with the same benchmarks as in (1), and (4) multi-equation models with either one world-wide benchmark index or three World factors, which are used for robustness checks.

Using these models, different tests on the parameters are conducted. These tests concern the relative performance and the relative risk of the SRI indexes compared to the benchmarks.

The performance tests confirm the results of most of the earlier studies, the SRI screens for equities neither lead to a significant out-performance nor an under-performance compared to the benchmarks. The latter is particularly interesting as the SRI screening process reduces the investment universe which should, according to optimal portfolio theory, lead to a reduction in the risk-adjusted return. As this is not the case, an investment in SRI equity indexes does not impose additional costs in terms of lower returns to the investor.

For some SRI indexes the index series has been calculated backwards from the official inception date on. An important result for performance comparisons is that for these SRI indexes the backward-looking performance seems to be biased upwards compared to the development from the launch date on. Performance estimates using the total time series therefore give an upper bound in these cases.

The analysis of the SRI index returns revealed that most of the indexes have a higher risk compared to the benchmarks. This means that changes in the benchmarks are translated into larger changes of the SRI indexes. This is also the reason that for more than half of the SRI indexes, the spanning tests are rejected. For at least 9 out of the

29 SRI indexes the hypothesis “spanning” is not rejected in the single-equation/single-factor tests. This means that for these indexes the relevant benchmark index has on average the same risk and return characteristics.

Multi-equation tests are also used to exploit information from the cross-section of SRI indexes. These tests confirm the results of the single equation tests. Additional robustness checks applying the same set of world-wide diversified indexes – either in a single- or a three-factor model – also come to the conclusion that SRI indexes exhibit the same performance as the benchmarks and that differences in the risk-return characteristics primarily stem from risk differentials.

References

- Admati, A. and S. Ross (1985), Measuring Investment Performance in a Rational Expectations Model, *Journal of Business* 58, pp. 1-26.
- Avanzi (2004), Green, Social and Ethical Funds in Europe, http://www.avanzi-sri.org/pdf/Complete_report_2004_final.pdf
- Bauer, R., K. Koedijk and R. Otten (2005), International Evidence on Ethical Mutual Fund Performance and Investment Style, *Journal of Banking and Finance* 29, 1751-1767.
- Carhart, M. (1997), On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.
- Cochrane, J. (2001), *Asset Pricing*, Princeton University Press, Princeton.
- Derwall, J., N. Guenster, R. Bauer and K. Koedijk (2005), The Eco-Efficiency Premium Puzzle, *Financial Analysts Journal* 61, No. 2, 51-63.
- DiBartolomeo, D. and L. Kurtz (1999), *Managing Risk Exposures of Socially Screened Portfolios*, Northfield Information Services, Boston, www.northinfo.com.
- Fama, E. and K. French (1993), Common risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.
- Fama, E. and K. French (1996), Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55-84.
- Ferson, W. and R. Schadt (1996), Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance* 51, 425-461.
- Garz, H., C. Volk and M. Gilles (2002), More Gain than Pain - SRI: Sustainability Pays Off, WestLB Panmure, http://www.westlbpanmure.com/sri/pdf/sri_nov2002.pdf
- Geczy, C., R. Stambaugh and D. Levin (2004), Investing in Socially Responsible Mutual Funds, Wharton School Working Paper, <http://finance.wharton.upenn.edu/~rlwctr/papers/0402.pdf>

- Greene, W. (2000), *Econometric Analysis*, Prentice Hall, New Jersey.
- Gregory, A., J. Matatko and R. Luther (1997), Ethical Unit Trust Financial Performance: Small Company Effects and Fund Size Effects, *Journal of Business Finance & Accounting* 24, 705-725.
- Hamilton, S., H. Jo and M. Statman (1993), Doing Well While Doing Good? The Investment Performance of Socially Responsible Mutual Funds, *Financial Analysts Journal*, November/December 1993, 62-66.
- Huberman, G. and S. Kandel (1987), Mean-Variance Spanning, *Journal of Finance* 42, 873-888.
- Kreander, N., R. Gray, D. Power and C. Sinclair (2002), The Financial Performance of European Ethical Funds 1996-1998, *Journal of Accounting and Finance* 1, 3-22.
- Kurtz, L. and D. DiBartolomeo (1996), Socially Screened Portfolios: An Attribution Analysis of Relative Performance, *Journal of Investing*, Fall 1996, 35-41.
- Luther, R. and J. Matatko (1994), The Performance of Ethical Unit Trusts: Choosing an Appropriate Benchmark, *British Accounting Review* 26, 77-89.
- Luther, R., J. Matatko and D. Corner (1992), The Investment Performance of UK "Ethical" Unit Trusts, *Accounting Auditing & Accountability Journal* 5, 57-70.
- Mallin, C., B. Saadouni and R. Briston (1995), The Financial Performance of Ethical Investment Trusts, *Journal of Business Finance & Accounting* 22, 483-496.
- Moskowitz, M. (1972), Choosing Socially Responsible Stocks, *Business and Society*, vol. 1 (Spring), 71-75.
- Newey, W. and K. West (1987), A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, vol. 55, 703-708.
- Sauer, D. (1997), The Impact of Social-Responsibility Screens on Investment Performance: Evidence from the Domini 400 Social Index and Domini Equity Fund, *Review of Financial Economics* 6, 23-35.
- Schroeder, M. (2004), The Performance of Socially Responsible Investments: Investment Funds and Indexes, *Financial Markets and Portfolio Management* 18, 122-142.
- Social Investment Forum (2003), Report on Socially Responsible Investing Trends in the United States,
http://www.socialinvest.org/areas/research/trends/sri_trends_report_2003.pdf
- Statman, M. (2000), Socially Responsible Mutual Funds, *Financial Analysts Journal*, May/June 2000, 30-39.

- Stone, B., J. Guerard, Jr., M. Gultekin and G. Adams (2001), Socially Responsible Investment Screening: Strong Evidence of no Significant Cost for Actively Managed Portfolios,
<http://www.institutionalshareowner.com/presentations/Socially-Responsible-Investment-actively-managed-value.pdf>
- Treynor, J. and F. Mazuy (1966), Can Mutual Funds Outguess the Market?, Harvard Business Review 45, 131-136.
- White, M. (1995), The Performance of Environmental Mutual Funds in the United States and Germany: Is there Economic Hope for “Green” Investors?, Research in Corporate Social Performance and Policy, Supplement 1, 323-344.

Appendix

Table 1: Description of the SRI-Equity Indexes

<i>Index Name</i>	<i>Abbrev</i>	<i>Supplier</i>	<i>Region</i>	<i>Start Date¹</i>	<i>Benchmark²</i>	<i>Characteristics</i>
ASPI	Aspi	Vigeo	Eurozone	12/1991 (7/2001)	DJ Euro Stoxx*	Consists of 120 companies of the DJ Euro Stoxx. Applies positive criteria related to e.g. social commitment, stakeholder relationships, health, environment, labor conditions.
Calvert Social Index (Calvin)	Calv	Calvert Group, Ltd.	U.S.	6/2000	Russell 1000	Consist of large, U.S.-based socially responsible companies. Positive criteria: environment, workplace issues and employee relations, good community relations. Negative criteria: weapons, tobacco, alcohol, gambling, pornography. About 630 companies in December 2003.
DJSI World	DJSI1	Dow Jones Indexes, STOXX Ltd. and SAM Group	World	12/1993 (9/1999)	DJ World*	The Dow Jones Sustainability Indexes (DJSI) are based on the Dow Jones Global Indexes and the STOXX index family, respectively. The selected companies represent the top 10% of the leading sustainability companies in the case of the global index and the top 20% of companies for the European sustainability indexes. The DJSI World indexes consist of more than 300 companies, the European indexes contain about 180 companies. The “ex AGTF”-indexes exclude companies that are engaged in: alcohol, gambling, tobacco, firearms.
DJSI World ex AGTF	DJSI2		World	12/1993 (9/1999)	DJ World*	
DJSI STOXX	DJSI3		Europe	12/1998 (10/2001)	DJ Stoxx*	
DJSI STOXX ex AGTF	DJSI4		Europe	12/1998 (10/2001)	DJ Stoxx*	
DJSI EUROSTOXX	DJSI5		Eurozone	12/1998 (10/2001)	DJ Eurostoxx*	
DJSI EUROSTOXX ex AGTF	DJSI6		Eurozone	12/1998 (10/2001)	DJ Eurostoxx*	
Ethical Index Euro	Eth1	E. Capital Partners	Eurozone	10/2000	DJ Stoxx	Negative criteria: weapons, nuclear power, alcohol, gambling, tobacco, pornography, violation of human rights, etc. Positive criteria are, e.g., good relations with communities, no racial or sexual discrimination, good environmental strategies, eco friendly products. Application of a “best-in-class” approach. The indexes concentrate on companies with large market capitalisation: 150 in the European index, 750 in the global index.
Ethical Index Global	Eth2	S.p.A.	World	1/2001	MSCI World	

FTSE4Good Europe 50	FT1	FTSE Group	Europe	7/1996 (7/2001)	FTSE AW Europe*	The FTSE4Good indexes are based on a large set of social, environmental and human rights criteria. Exclusionary criteria are: tobacco, nuclear power, weapons, uranium. The index series consists of the tradable indexes Europe50, Global100, UK50 and US100 as well as a broader class of indexes from which the tradable indexes are constructed (Europe, Global, UK, US).
FTSE4Good Europe	FT2		Europe	7/1996 (7/2001)	FTSE AW Europe*	
FTSE4Good Global 100	FT3		World	7/1996 (7/2001)	FTSE Developed World*	
FTSE4Good Global	FT4		World	7/1996 (7/2001)	FTSE Developed World*	
FTSE4Good UK 50	FT5		UK	7/1996 (7/2001)	FT All Share*	
FTSE4Good UK	FT6		UK	7/1996 (7/2001)	FT All Share*	
FTSE4Good US 100	FT7		U.S.	7/1996 (7/2001)	FTSE Local USA*	
FTSE4Good US	FT8		U.S.	7/1996 (7/2001)	FTSE Local USA*	
Humanix 175 Europe	Hu1	Humanix Holding AB	Europe	6/2001	DJ Stoxx	The Humanix indexes consist of the largest companies in terms of market capitalisation. The number of companies is 50 for Sweden, 175 for Europe and the U.S., and 200 for the World. Criteria for the selection of companies: environmental risks (-), human rights (+), firearms (-), alcohol (-), tobacco (-). All Humanix indexes are price indexes. The Humanix indexes are no longer calculated and end in December 2003.
Humanix 175 U.S.	Hu2		U.S.	6/2001	S&P 500	
Humanix 200 Global	Hu3		World	6/2001	MSCI World	
Humanix 50 Sweden	Hu4		Sweden	6/2001	MSCI Sweden	
Jantzi Social Index	Jantzi	Michael Jantzi Research Associates, Inc.	Canada	1/2000	S&P/TSE60*	Consists of 60 Canadian companies with a large market capitalisation that meet social and environmental criteria: good environmental records, diversity in workplace, good community relationships. Negative criteria are, e.g., nuclear power, tobacco, weapons, poor relationships with aboriginal communities, fraudulent business practices, manufacture unsafe products, etc.

Kempen SNS Smaller European SRI Index	Ke	Kempen Asst Mgmt, SNS Capital Mgmt.	Europe	12/1998 (10/2003)	HSBC Smaller European Index*	Selection criteria: environmental performance, social performance, business ethics. Index consists of about 70 companies with small market capitalisation. The target for the number of companies included is 150. About 1700 companies are under investigation.
KLD Domini 400 Social Index	KLD1	KLD Research & Analytics, Inc.	U.S.	5/1990	S&P 500	The index consists of 400 U.S. companies that passed the KLD screening criteria. Negative criteria: alcohol, tobacco, gambling nuclear power, firearms, military weapons, etc. Positive criteria: environmental impact, citizenship, employee relations and diversity.
KLD BMSI	KLD2		U.S.	1/2001	Russell 3000*	The Broad Market Social Index consists of all companies in the Russel 3000-index that pass the SRI screening of KLD.
KLD LCSi	KLD3		U.S.	1/2001	Russell 1000*	The Large Cap Social Index is a socially screened subset of the Russel 1000-index. The LCSi represents about 92% of the U.S. market capitalisation.
Naturaktien-index (NAI)	NAI	Securvita	World	4/1997	MSCI World	Consisted of originally of 20 companies, now 25. Environmental, sustainability and social criteria. Exclusionary criteria: weapons, discrimination of women, nuclear energy, tobacco, etc. The NAI is a price index.
Westpac-Monash Eco-Index	West	Westpac Investment Mgmt.	Australia	1/1999	S&P/ASX200*	Consists of 75 Australian companies, which cover 24 sectors. The SRI evaluation concentrates on the following fields: environmental strategy, environmental management, stakeholder relationships, production and products.

Notes: ¹ Start date of the time series. Figures in brackets indicate the official launch date of the index. Values of the index that are available in periods before the official start date have been calculated backwards by the supplier of the index. ²An asterisk (*) indicates that the benchmark index is the official benchmark chosen by the supplier company.