A Stochastic Dominance Approach to Spanning

Thierry Post

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS-2002-01-F&A
Publication	January 2002
Number of pages	16
Email address corresponding author	gtpost@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam P.O. Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	We develop a Stochastic Dominance methodology to analyze if new assets expand the investment possibilities for rational nonsatiable and risk-averse investors. This methodology avoids the simplifying assumptions underlying the traditional mean-variance approach to spanning. The methodology is applied to analyze the stock market behavior of small firms in the month of January. Our findings suggest that the previously observed January effect is remarkably robust with respect to simplifying assumptions regarding the return distribution.	
Library of Congress Classification (LCC)	5001-6182	Business
	4001-4280.7	Finance Management, Business Finance, Corporation Finance
	HG 4529	Investment analysis
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	G 3	Corporate Finance and Governance
	C 19	Econometric and Statistical methods: other
European Business Schools Library Group (EBSLG)	85 A	Business General
	220 A	Financial Management
	250 D	Statistische analyse
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.30	Financieel management, financiering
	85.33	Beleggingsleer
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Financieel management, bedrijfsfinanciering, besliskunde	
	Stochastische methoden, Investeringen, Effectenhandel, Kleinbedrijf	
Free keywords	portfolio selection, portfolio evaluation, stochastic dominance, spanning, linear programming	

A Stochastic Dominance Approach to Spanning

THIERRY POST [*]

[*] Post is with the department of Finance of the Erasmus University Rotterdam in the Netherlands. This study forms part of a research program on stochastic dominance. Details on the program are available at the program homepage*:*

http://www.few.eur.nl/few/people/gtpost/stochastic_dominance.htm.

The author gratefully acknowledges financial support by Tinbergen Institute and Erasmus Research Institute of Management.

ABSTRACT

We develop a Stochastic Dominance methodology to analyze if new assets expand the investment possibilities for rational nonsatiable and risk-averse investors. This methodology avoids the simplifying assumptions underlying the traditional mean-variance approach to spanning. The methodology is applied to analyze the stock market behavior of small firms in the month of January. Our findings suggest that the previously observed January effect is remarkably robust with respect to simplifying assumptions regarding the return distribution.

SPANNING occurs if no investor in a particular class of investors benefits from a particular expansion of the investment possibilities. This concept is useful for numerous problems in financial economics. For example, it is useful for analyzing the impact of the introduction of new assets (e.g. via IPOs) or the relaxation of investment restrictions for existing assets (e.g. liberalization in emerging markets).

Thus far, the literature on spanning predominately focused on mean-variance analysis (MVA); see e.g. Huberman and Kandel (1987). Unfortunately, MVA in many cases is not 'economically meaningful'. For example, it is well known that MVA is consistent with Expected Utility Theory only for restrictive classes of return distributions and investor utility functions. Roughly speaking, the return distribution should be normal or investor utility should be quadratic (see e.g. Bigelow, 1993). A wealth of evidence suggests that both assumptions are highly unrealistic. For example, asset returns exhibit systematic skewness and investors exhibit a preference for positive skewness (see e.g. Friend and Westerfield (1980) and Harvey and Siddique (2001)). One approach to circumvent this limitation is to extend MVA towards a more general framework that also includes higher moments of the return distribution. Unfortunately, economic theory does not forward strong predictions on investor preferences or asset return distributions, and it gives minimal guidance for selecting the appropriate moments.

This paper uses an alternative approach to spanning, using Stochastic Dominance (SD; see e.g. Levy (1998)). SD criteria rely on a minimal set of preference and distribution assumptions, and they effectively consider the entire return distribution rather than a finite set of moments. This approach is useful if there is no prior reason to restrict preferences or distributions, as is generally true for investor behavior and asset returns. Despite its theoretical attractiveness, SD thus far has not seen a strong proliferation in financial economics. (Noteworthy exceptions are Falk and Levy's (1989) study of market reactions to quarterly earnings' announcements and the studies of the January effect by Seyhun (1993) and Larsen and Resnick (1996).) This is presumably caused by several practical problems traditionally associated with SD: (1) a lack of power (=ability to detect inefficient portfolios) in small samples, (2) the absence of tools for statistical inference, and (3) the computational burden for the important case where it is possible to diversify between the choice alternatives. A number of recent developments deals with these problems and provides a strong stimulus towards the further proliferation of SD. First, various approaches have been developed to approximate the sampling distribution of SD results, including bootstrapping (e.g. Nelson and Pope (1990)) and asymptotic distribution theory (see e.g. Davidson and Duclos (2000)). These approaches allow for constructing confidence intervals and for testing hypothesis. Second, Post (2001) presents tractable linear programming (LP) tests for SD efficiency in the case with diversification possibilities. These tests improve computational tractability and power (all diversified portfolios are included in the analysis, which improves the likelihood of detecting inefficient portfolios).

This paper develops a SD methodology to test if new assets are spanned. For simplicity, we focus on the criterion of second-order SD in the presence of a riskless asset (SSDR; e.g. Levy (1998), Section 4.3). The assumptions associated with this criterion (monotonicity and concavity of utility functions) have a good economic interpretation (nonsatiation and risk-aversion). Section I develops empirical test statistics that can be computed using straightforward linear programming (LP). These test statistics are related in a subtle manner to the efficiency test statistics by Post (2001). The latter test statistics ask if a given portfolio is SD efficient (=not dominated). The test statistics in this paper ask if a given asset enters into some portfolio that is SD efficient or if there exist rational investors that invest part of their wealth in the asset. The test statistics are based on the empirical distribution function rather than the true (unknown) distribution function. To account for sampling error, Section II develops a statistical test procedure based on the asymptotic least favorable distribution of the test statistics. Section III presents a simulation experiment that gauges the statistical goodness of this test procedure in small samples. Section IV presents an application for the stock market behavior of small firms in the month of January. Finally, Section V gives concluding remarks and suggestions for further research. The appendix presents formal proofs for our theorems.

I. SSDR SPANNING

We consider a competitive capital market with K investors and three financial assets: a riskless benchmark asset (F) and a risky benchmark asset (M), and a new risky asset (A). (The extensions towards multiple risky benchmark assets and multiple new assets are discussed in Section V.) Throughout the text, we will use the index sets $J \equiv \{1,\cdots,K\}$ and $I \equiv \{F, M, A\}$ to denote the different investors and assets respectively. Investors may construct portfolios as convex combinations of the three assets. Throughout the text, we will denote the portfolio weights by the vector $l \in \Re^3$ and the portfolio possibilities by the set $\Lambda \equiv \left\{ l \in \Re^3_+ : l^T\boldsymbol{e} = 1 \right\}$.[1,2] The investors are nonsatiable and risk averse, and strictly increasing and concave von Neuman-Morgenstern utility functions u_j, $j \in J$, represent their preferences. In addition, the returns $\boldsymbol{x} \equiv \left(\boldsymbol{x}_F \ \boldsymbol{x}_M \ \boldsymbol{x}_A \right)$ are random variables with a continuous joint cumulative distribution function (CDF), $G(\boldsymbol{x})$. Each investor $j \in J$ takes the return distribution as given and chooses a portfolio to maximize his expected utility, i.e. the optimal solution to $\max_{l \in \Lambda} \int u_j(\boldsymbol{x} l) \partial G(\boldsymbol{x})$.

This model imposes minimal structure on investor preferences and return distributions. Unfortunately, we cannot fully characterize the market equilibrium (the

---

[1] Throughout the text, we will use $\Re^m$ for an *m*-dimensional Euclidean space, and $\Re^m_+$ denotes the positive orthant. Further, to distinguish between vectors and scalars, we use a bold font for vectors and a regular font for scalars. Finally, we use $\boldsymbol{e}$ for a unity vector with dimensions conforming to the rules of matrix algebra.

[2] The set Λ assumes that short sales are not allowed and that no additional restrictions are imposed on the portfolio weights. Our analysis is based on the optimality conditions for optimizing a concave utility function over a convex set (see the proof to Theorem 1). In principle, the analysis can be extended towards a general polyhedral portfolio possibilities set, and hence it is possible to introduce (bounded) short selling and to impose additional investment restrictions. We basically have to check if all hyperplanes that support the extreme points of the original portfolio possibilities set also support the extreme points of the extended possibilities set.

prices of the different assets and the amounts invested by the different investors) without imposing further structure. Still, we do know that we can rationalize investment in *A* only if at least one investor invests at least part of his wealth in *A*. This naturally introduces the concept of spanning:

DEFINITION 1 *Asset A is spanned if and only if no investor $j \in J$ is better off by investing part of his wealth in A, i.e.:*

$$(1) \qquad \max_{l \in \Lambda} \int u_j(\boldsymbol{x}\, l)\partial G(\boldsymbol{x}) = \max_{l \in \Lambda^*} \int u_j(\boldsymbol{x}\, l)\partial G(\boldsymbol{x}) \quad \forall j \in J\ ,$$

with $\Lambda^* \equiv \{l \in \Lambda : l_A = 0\}$ for the investment possibilities excluding A.

In practical applications, full information about investor preferences typically is not available, and one generally cannot directly test spanning. This provides the rationale for using the SSDR criterion, which considers the entire set of strictly increasing and concave utility functions, say *U*. Since the utility functions of the investors are elements of this set, i.e. $u_j \in U$ for all $j \in J$, we can obtain a weaker spanning condition:

DEFINITION 2 *Asset A is SSDR spanned if and only if no rational nonsatiable, risk-averse investor is better off by investing part of his wealth in A, i.e.:*

$$(2) \qquad \max_{l \in \Lambda} \int u(\boldsymbol{x}\, l)\partial G(\boldsymbol{x}) = \max_{l \in \Lambda^*} \int u(\boldsymbol{x}\, l)\partial G(\boldsymbol{x}) \quad \forall u \in U\ .$$

SSDR *spanning* is related in a subtle way to the usual concepts of *efficiency* and *dominance*. A portfolio is SSDR inefficient or dominated if all rational nonsatiable and risk averse investors prefer a second portfolio to the former portfolio. Unfortunately, this concept is relevant only if all investors hold the same portfolio and if the composition of the portfolio is known. For example, testing whether asset *A* is SSDR efficient is relevant only if all investors invest in *A* exclusively. By contrast, SSDR spanning occurs if *all* portfolios including *A* are SSDR inefficient. This concept is also relevant if different portfolios hold different portfolios and if the composition of those portfolios is unknown.

Throughout the text, our null hypothesis will be that spanning does *not* occur, i.e. some rational investors would invest at least part of their wealth in *A*. Rejection of this null gives strong evidence in favor of an imbalance, because the concept of SSDR spanning is based on minimal prior assumptions.

Apart from investor preferences, the CDF generally is not known, and hence we cannot directly test SSDR spanning. Rather, information typically is limited to a discrete set of time series observations, say $\mathrm{X} \equiv (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T)^T$ with $\boldsymbol{x}_t \equiv (\boldsymbol{x}_F\ \boldsymbol{x}_{Mt}\ \boldsymbol{x}_{At})$. For simplicity, we assume that the observations are serially independent and identically distributed (IID) random drawings from the CDF. Under this assumption, the empirical distribution function (EDF) $F(\boldsymbol{x}) \equiv \mathrm{card}\{t \in \Theta : \boldsymbol{x}_t \leq \boldsymbol{x}\}/T$, with

$\Theta \equiv \{1, \cdots, T\}$, gives a statistically consistent estimator for the CDF.[3] By focusing on the EDF rather than the CDF, we can obtain an empirical spanning condition:

DEFINITION 3 *Asset A is empirically SSDR spanned if and only if:*

(3) $$\max_{l \in \Lambda} \int u(\boldsymbol{x}\, l)\partial F(\boldsymbol{x}) = \max_{l \in \Lambda^*} \int u(\boldsymbol{x}\, l)\partial F(\boldsymbol{x}) \quad \forall u \in U \Leftrightarrow$$
$$\max_{l \in \Lambda} \sum_{t \in \Theta} u(\boldsymbol{x}_t\, l)/T = \max_{l \in \Lambda^*} \sum_{t \in \Theta} u(\boldsymbol{x}_t\, l)/T \quad \forall u \in U\,.$$

A straightforward approach to testing empirical SSDR spanning is to check if every portfolio that includes *A* is empirically SSDR inefficient. Unfortunately, computational burden prohibits this approach, as there are infinitely many portfolios that include *A*. However, we can extend the analysis by Post (2001) to develop a more tractable approach. To simplify notation, we assume that the data are ranked in ascending order by the return of *M*, i.e. $\boldsymbol{x}_{M1} < \boldsymbol{x}_{M2} < \cdots < \boldsymbol{x}_{MT}$.[4] Further, we assume that the risk free return exceeds the minimum return for the risky assets, and that it falls below the average return for the risky assets, i.e. $\min_{t \in \Theta} \boldsymbol{x}_{it} < \boldsymbol{x}_F < \sum_{t \in \Theta} \boldsymbol{x}_{it}/T$ for $i \in \{M, A\}$. Under this assumption, some investors will invest part of their wealth in the riskless asset, but no investor will invest all of his wealth in the riskless asset, reflecting Arrow's theorem - '*A risk averter takes no part of an unfavorable or barely fair game; on the other hand, he always takes some part of a favorable gamble*' (Arrow, 1971, p. 100). Using these simplifications, we can obtain the following result:

THEOREM 1 *Empirical SSDR spanning can be tested using the primal test statistic*

(4) $$y_P \equiv \inf_{\boldsymbol{\beta} \in B} \left\{ \sum_{t \in \Theta} \boldsymbol{\beta}_t (\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T : \sum_{t \in \Theta} \boldsymbol{\beta}_t (\boldsymbol{x}_{Mt} - \boldsymbol{x}_F)/T \geq 0 \right\},$$

with $B \equiv \{\boldsymbol{\beta} \in \Re_+^T : \boldsymbol{\beta}_1 \geq \boldsymbol{\beta}_2 \geq \cdots \geq \boldsymbol{\beta}_T = 1\}$, *or alternatively using the dual test statistic*

(5) $$y_D \equiv \sup_q \{x_T(q) : x_t(q) \quad \forall t \in \Theta \setminus T\},$$

with $x_t(q) \equiv \sum_{s=1}^{t} ((1-q)\boldsymbol{x}_{Ms} - \boldsymbol{x}_{As} + q\boldsymbol{x}_F)/T$. *Specifically, asset A is empirically SSDR spanned if and only if* $y_P = y_D \geq 0$.

The test statistics y_P and y_D can be computed by straightforward linear programming (LP); full LP formulations are included as (P) and (D) in the proof in

[3] However, there is substantial evidence that the distribution of assets returns (e.g. interest rates, risk premia, volatilities and correlation coefficients) varies through time. This problem is especially relevant for applications that use data of long time periods. One possible approach to account for time variation is to use econometric time series estimation techniques to estimate a conditional CDF. The empirical test developed below could then be applied to random samples from the estimated CDF.

[4] Since we assume a continuous return distribution, ties do not occur and the ranking is unique. Still, the analysis can be extended in a straightforward way to cases where ties do occur e.g. due to a discrete return distribution or due to measurement problems or rounding (see Post, 2001).

Appendix A. The problem involves only T variables and $T+1$ constraints. For small data sets up to hundreds of observations, this problem can be solved with minimal computational burden, even with desktop PCs and standard solver software (like LP solvers included in spreadsheets). Still, the computational complexity, as measured by the required number of arithmetic operations, and hence the run time and memory space requirement, increases progressively with the number of model variables. Therefore, specialized LP solver software is recommended for large-scale problems involving thousands of observations.[5] Note that the primal problem may be unbounded (and the dual infeasible) if A is *not* empirically SSDR spanned. For example, this occurs if $\sum_{t\in\Theta} \boldsymbol{x}_{At}/T > \sum_{t\in\Theta} \boldsymbol{x}_{Mt}/T$. In these cases, the test statistics take the value minus infinity and spanning does not occur. (The application in Section IV includes such cases; see Table 2.)

II. SAMPLING ERROR

The test statistics y_P and y_D are based on the EDF rather than the CDF, and the test results are likely to be affected by sampling error. The applied researcher must have knowledge of the sampling distribution in order to make inferences about the true classification (SSDR spanned or not spanned). Post (2001) derived an analytical characterization of the asymptotic sampling distribution of his efficiency tests. This section extends the Post results towards the SSDR test statistic y_P. (Duality implies that the results apply with equal strength to y_D.)

There are various hypotheses that could serve as the null hypothesis in a test procedure. In SD analysis, a typical null hypothesis is that the risky choice alternatives are independent random variables with the same population distribution, or alternatively the choice alternatives are contemporaneously IID. We adopt this null for our spanning tests and we assume that $\boldsymbol{x}_A$ and $\boldsymbol{x}_M$ are contemporaneously IID random variables with univariate CDF $H:\Re \to [0,1]$ with variance $s^2 < \infty$. The shape of the distribution of y_P under the null generally depends on the shape of $H(x)$. Our approach will be to focus on the least favorable distribution, i.e. the distribution that maximizes the size or relative frequency of Type I error (rejecting the null when it is true). This approach stems from the desire to be protected against Type I error. For each $H(x)$, the size is always smaller than the size for the least favorable distribution. Interestingly, the least favorable distribution is relatively simple and known results can derive the asymptotic probability of exceedance or *p*-value for y_P. The use of the most favorable distribution implies that we accept a high frequency of Type II error (accepting the null when it is not true) or a low power (1- the relative frequency of Type II error). Future research could focus on tests that minimize Type II error.

THEOREM 2 *For the he asymptotic least favorable distribution,* y_P *behaves as a normal random variable with mean zero and variance* $2s^2/T$.

---

[5] For an elaborate introduction in LP, we refer to Chvatal (1983). In practice, very large LPs can be solved efficiently by both the simplex method and interior-point methods. An elaborate guide to LP solver software can be found at the homepage of the Institute for Operations Research and Management Science (INFORMS); http://www.informs.org/.

The theorem implies that p-values $P(y_P \geq y | H_0)$ may be found as $\Phi\left(\frac{-y}{\sqrt{2/T}\sigma}\right)$, with $\Phi(\cdot)$ for the cumulative standard normal distribution function. These p-values converge to zero as the length of the time series (T) grows. This makes intuitive sense, because the EDF is a statistically consistent estimator for the CDF under our maintained assumptions (see Section I). Still, for small time series, the p-values can be very large and a naïve approach to the test statistic (reject efficiency if $y_P > 0$) is unlikely to yield anything but noise. A more sound approach is to compare the p-value for the observed value of y_P with a predefined level of significance α; we may reject the null if the p-value is smaller than or equal to the significance level. Alternatively, we may reject the null if the test statistic y_P exceeds the critical value $\Phi^{-1}(1-\alpha)\sqrt{2/T}\sigma$. Computing p-values or critical values requires the unknown population variance σ^2. We may estimate this parameter in a distribution-free and consistent manner using the sample equivalent:

$$(6) \qquad \hat{\sigma}^2 \equiv \sum_{i \in \{M,A\}} \sum_{t \in \Theta} (\mathbf{x}_{it} - \sum_{t \in \Theta} \mathbf{x}_{it} / T)^2 / 2T .$$ [6]

III. SIMULATION EXPERIMENT

To assess the goodness of the test procedure outlined in Section II, we extend the simulation experiment used by Kroll and Levy (1980) and Nelson and Pope (1991). Assume x_F equals 0.06, and x_M and x_A obey a bivariate normal distribution with means $\mu_M = 0.20$ and $\mu_A = 0.15$, standard deviations $\sigma_M = \sigma_A = 0.20$, and correlation coefficient ρ.[7] Asset A is SSDR dominated by M, because M achieves a higher mean and a lower standard deviation than A. However, rational investors may still invest in A in the context of a diversified portfolio. The diversification benefits from A depend on ρ. It is easy to verify that M and F SSDR span A if and only if $\rho \geq 0.50$. However, sampling errors complicate the empirical determination of the known classification. The results based on a sample from the population may give one of the outcomes listed below.

	No spanning In population ($\rho < 0.50$)	Spanning in population ($\rho \geq 0.50$)
No spanning in sample ($y_P < \Phi^{-1}(1-\alpha)\sqrt{2/T}\hat{\sigma}$)	No Error	Type II error

[6] This is simply the equally weighed average of the sample variance of $\mathbf{x}_A$ and $\mathbf{x}_M$; under the null, both choice alternatives have the same variance.

[7] Our experiment differs from the original Kroll-Levy experiment in three respects. First, Kroll and Levy focus on efficiency, while we analyse spanning. Second, Kroll and Levy use data sets of 100 observations and set the correlation coefficient at zero. By contrast, we consider various different sample sizes and correlation coefficients. Finally, the original experiment did not include a riskless asset, while we use the SSDR criterion that does use a riskless asset.

Spanning in sample ($y_P \geq \Phi^{-1}(1-a)\sqrt{2/T}\hat{s}$)	Type I error	No Error

We draw through Monte-Carlo simulation 1000 random samples of N observations from the bivariate normal distribution, and apply our test procedure to each sample. We follow this procedure for correlation coefficients of $r \in \{0,0.05,\cdots,1\}$ and for samples of size $N \in \{100,500,2000\}$. The nominal level of significance a (or the size for the asymptotic least favorable distribution) is set at 5 percent.

[INSERT FIGURE 1 ABOUT HERE]

For all sample sizes and correlation coefficients, the size of the test procedure approximates zero, which reflects the conservative nature of our test. The size of the test comes at the cost of a low power in small samples. The test procedure is powerful only if the sample is large or if spanning is 'strong' i.e. r is well above 0.5. The lack of power in small samples makes intuitive sense for two reasons. First, the CDF needs to satisfy a series of conditions in order to establish spanning. If the EDF violates a single condition, then spanning will not be detected. Second, the procedure to account for sampling error builds on the least favorable distribution that minimizes Type I error at the cost of Type II error.

Fortunately, large data sets are available for many applications in financial economics. Further, we could apply econometric time series techniques to obtain an estimate for the CDF that is more efficient than the EDF. We could then apply our test to a large random sample from the estimated CDF rather than the raw data. This approach effectively uses prior distribution information to generate artificial return observations. Still, it is desirable to develop a more powerful test, e.g. a test that explicitly minimizes the probability of Type II error rather than Type I error, or a test that is based on a particular class of return distributions.

IV. THE JANUARY EFFECT

A wealth of empirical evidence suggests that the stock market returns of small firms systematically outperform the returns of large firms during the month of January (see e.g. Keim (1983)). Several explanations have been forwarded for this phenomenon, including 'window dressing' by institutional investors (see e.g. Haugen and Lakonishok (1988)) and 'tax-loss selling' by individual investors (see e.g. Reinganum (1983). Another explantion is the mismeasurement of risk. The returns of small firms may be more risky than than the returns of large firms, and a higher average return may serve as a compensation for the additional risk.

The potential of using SD to account for risk was recognized by Seyhun (1993). He studied the January effect by examining whether different decile portfolios are SD efficient in January. The results suggest that the January effect can not be explained by mismeasurement of risk; all portfolios except the smallest decile portfolios are inefficient in January. Larsen and Resnick (1996) extended this study by means of bootstrapping, so as to assess the sensitivity of the results to sampling variation. Their results confirm the Seyhun results, although the pattern is somewhat different; only the six largest decile portfolios are inefficient to a statistically significant degree.

The Seyhun (1993) and Larsen and Resnick (1996) approach implicitly assumes that investors have to choose one of the decile portfolios. Hence, this approach ignores the possibility to diversify between the decile portfolios and to invest in a riskless asset. To test whether the January effect is robust with respect to the inclusion of diversification possibilities and a riskless asset, we apply our SSDR spanning test. We analyze ten value-weighted decile portfolios of NYSE, AMEX, and NASDAQ stocks, and the one-month US Treasury bill (the riskless asset). We use data on monthly dividend-adjusted returns from July 1926 to December 2000 (894 observations) obtained from the data library on the homepage of Kenneth French (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/). Table 1 gives some descriptive statistics for the data set.[8]

[INSERT TABLE 2 ABOUT HERE]

We test whether the smallest decile portfolio and the Treasury bill span the larger decile portfolios. Specifically, for every decile portfolio, we compute the value of the primal test statistic y_P, using the smallest decile portfolio and the Treasury bills as benchmark assets. Next, we compute the asymptotic least favorable *p*-value, with the sample variance $\hat{s}^2$ to proxy the unknown population variance. If this *p*-value is smaller than or equal to the significance level, then we may conclude that SSDR spanning occurs. Table 2 gives the results. For the full sample, spanning occurs for none of the 9 higher decile portfolios. Hence, there exist rational, risk-averse investors that invest at least part of their wealth in the higher decile portfolios, and we cannot conclude that the lowest decile portfolio exhibits abnormal performance. The results change remarkably if we consider the January returns only. The smallest decile portfolio and the T-bill span all of the 9 higher decile portfolios. For the 8 highest decile portfolios, the classification is statistically significant at a level of confidence of about 95 percent. These results support the results by Seyhun and Larsen and Resnick; the January effect is not explained away by the mismeasurement of risk. The robustness of the January effect is remarkable, especially because our test is based on the asymptotic least favorable distribution and it typically involves low power for samples as small as 74 observations (see Figure 1).

[INSERT TABLE 3 ABOUT HERE]

V. CONCLUDING REMARKS

1. We stress that the SD tests are not intended to replace the MVA tests. SD uses minimal prior preference and distribution assumptions and it therefore involves less Type I error (wrongly classifying an efficient portfolio as inefficient) than MVA does. However, by imposing prior structure on the data MVA involves more power (or less Type II error; wrongly classifying an inefficient portfolio as efficient) than SD does. Therefore, the SD tests are natural complements rather than substitutes for the existing MVA tests.

2. Our spanning tests effectively test if the risky asset A improves the investment possibilities available from two benchmark assets: the riskless asset F and the risky asset M. This approach is useful if we can aggregate in a meaningful way all

[8] To account for the variation over time of the return distribution, the raw returns in month t are corrected for the difference between the riskless rate at time t and the riskless rate for December 2000.

benchmark assets and all new assets. Still, it would be interesting to extend our analysis to the case with multiple risky benchmark assets and multiple new assets. Our test is based on checking whether all hyperplanes that support F and M also support A (see the Proof to Theorem 1). Introducing multiple new assets is relatively simple: we can simply check if the hyperplanes support all new assets. This boils down to simply applying our test for all new assets. (Section IV effectively uses this approach to analyze if the smallest decile portfolio and the T-bills span the 9 higher decile portfolios.) By contrast, introducing multiple risky benchmark assets substantially increases computational complexity. In our model, all portfolios of M and F involve the same ranking for the returns (recall that the test statistics y_P and y_D use ordered return observations). In case of multiple risky benchmark assets, many different rankings generally occur. Determining all different rankings is not easy and enumerating all possible rankings involves substantial computational burden. Finding a more tractable approach is an interesting route for further research.

3. We have focused on obtaining an analytical characterization of the sampling distribution of our test statistics. Bootstrapping is another approach to sampling error. The bootstrap, first introduced by Efron (1979) and Efron and Gong (1983), is a well-established statistical tool to analyze the sensitivity of empirical estimators to sampling variation in situations where the sampling distribution is difficult or impossible to obtain analytically. Nelson and Pope (1991) demonstrated in a convincing way that this approach can quantify the sensitivity of the EDF to sampling variation, and that SD analysis based on the bootstrapped EDF is more powerful than comparison based on the original EDF. The tractable LP structure of our tests suggests that it is possible also for SSDR spanning to substitute brute computational force to overcome the analytical intractability of SD.

APPENDIX A

Proof to Theorem 1 We first consider the sufficient condition. If spanning does not occur, then

(i) $\exists u \in U : \sum_{t\in\Theta} u(\boldsymbol{x}_t t(u,\Lambda))/T > \sum_{t\in\Theta} u(\boldsymbol{x}_t t(u,\Lambda^*))/T$,

and

(ii) $t_A(u,\Lambda) > 0$,

with $t(u,\Lambda) \equiv \arg\max_{l\in\Lambda} \sum_{t\in\Theta} u(\boldsymbol{x}_t l)/T$. If we optimize $\sum_{t\in\Theta} u(\boldsymbol{x}_t l)/T$, $u \in U$, over Λ^*, then the optimality conditions for convex problems (see e.g. Hiriart-Urruty and Lemaréchal (1993), Thm. VII:1.1.1 and Cond. VII: 1.1.3) require that there exists an increasing hyperplane that is tangent at $\mathrm{X}t(u,\Lambda^*)$ and that supports $\boldsymbol{x}_M$ and $\boldsymbol{x}_F$ from above. (This optimality condition generalizes the well-known Kuhn-Tucker conditions for continuously differentiable utility functions towards superdifferentiable utility functions, including the piecewise-linear utility functions used below). Arrow's (1971) theorem (see Section I) implies that M is always included in the optimal portfolio relative to Λ^*, i.e. $t_M(u,\Lambda^*) > 0$, and the optimality condition therefore implies:

(iii) $\sum_{t\in\Theta} u(\boldsymbol{x}_t t(u,\Lambda^*))/T = \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda^*))\boldsymbol{x}_{Mt}/T \geq \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda^*))\boldsymbol{x}_F/T$,

with $\partial u(t(u,\Lambda^*)) \equiv \left(\partial u(\boldsymbol{x}_1 t(u,\Lambda^*))\cdots\partial u(\boldsymbol{x}_T t(u,\Lambda^*))\right)$ for a supergradient at $\mathrm{X}t(u,\Lambda^*)$. Concavity of u implies

(iv) $\sum_{t\in\Theta} u(\boldsymbol{x}_t t(u,\Lambda))/T \leq \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda^*))\boldsymbol{x}_t t(u,\Lambda)/T \leq \max_{i\in \mathrm{I}:t_i(u,\Lambda)>0} \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda))\boldsymbol{x}_{it}/T$.

Combining (i) to (iv), we find that spanning does not occur only if

(v) $\exists u\in U: \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda))\boldsymbol{x}_{At}/T > \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda))\boldsymbol{x}_{Mt}/T \geq \sum_{t\in\Theta} \partial u(\boldsymbol{x}_t t(u,\Lambda^*))\boldsymbol{x}_F/T$.

If these inequalities apply for $u\in U$, then they also apply for the standardized utility function $v \equiv u/\partial u(\boldsymbol{x}_T t(u,\Lambda^*))\in U$. By construction, $\partial v(t(u,\Lambda^*))$ is a feasible solution, i.e. $\partial v(t(u,\Lambda^*))\in \mathrm{B}$ (recall that all portfolios of M and F have the same ranking as M). The inequalities (iv) imply that this solution is associated with a strictly negative solution value. Hence, spanning does not occur only if $y_P < 0$, or alternatively spanning occurs if $y_P \geq 0$.

We next consider the necessary condition. If $y_P < 0$, then

(vi) $\sum_{t\in\Theta} b_t^*(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T < 0; \sum_{t\in\Theta} b_t^*(\boldsymbol{x}_{Mt} - \boldsymbol{x}_F)/T \geq 0$,

with $b^* \in B$ for the optimal solution. We can then always find $k \in \Lambda : k_A > 0$ such that the ranking of $\boldsymbol{x}_M$ is preserved, i.e. $\boldsymbol{x}_1 k < \boldsymbol{x}_2 k < \cdots < \boldsymbol{x}_T k$, and

(vii) $\sum_{t\in\Theta} b_t^*(\boldsymbol{x}_{Mt} - \boldsymbol{x}_t k)/T < 0; \sum_{t\in\Theta} b_t^*(\boldsymbol{x}_{Mt} - \boldsymbol{x}_F)/T \geq 0$.

Now consider the piecewise linear utility function $p(x) \equiv \min_{t\in\Theta}(a_t + b_t^* x)$, with $a_t \equiv 0.5\sum_{s=t}^{T-1}(b_{s+1}^* - b_s^*)(\boldsymbol{x}_t k + \boldsymbol{x}_{t+1} k)$. By construction, this function is monotone increasing and concave and hence $p(x)\in U$. It is easy to verify that $\sum_{t\in\Theta} p(\boldsymbol{x}_t k)/T = \sum_{t\in\Theta}(a_t + b_t^* \boldsymbol{x}_t k)/T$ and $\sum_{t\in\Theta} p(\boldsymbol{x}_t t(p,\Lambda^*))/T \leq \sum_{t\in\Theta}(a_t + b_t^* \boldsymbol{x}_t t(p,\Lambda^*))/T$. Combining this with (vii), we find that $y_P < 0$ implies that

(viii) $\sum_{t\in\Theta} p(\boldsymbol{x}_t k)/T > p(\boldsymbol{x}_t t(p,\Lambda^*))/T$.

Hence, if $y_P < 0$, then (i) is satisfied and spanning does not occur. Alternatively, spanning occurs only if $y_P \geq 0$.

The alternative formulation y_D is obtained by applying linear duality theory to y_P. Specifically, the following is a full LP formulation for y_P:

(P)
$$
\begin{aligned}
&\inf \sum_{t=1}^{T} b_t(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T \\
&\text{s.t.} \sum_{t=1}^{T} b_t(\boldsymbol{x}_{Mt} - \boldsymbol{x}_F)/T \geq 0 \\
&b_t - b_{t+1} \geq 0 \quad t = 1,\cdots,T-1 \\
&b_T = 1 \\
&b_t \text{ free} \quad t = 1,\cdots,T
\end{aligned}
$$

The LP dual of (P) is:

(D) $\sup s_T$

$$\text{s.t.}\ q(\boldsymbol{x}_{M1} - \boldsymbol{x}_F)/T + s_1 = (\boldsymbol{x}_{M1} - \boldsymbol{x}_{A1})/T$$

$$\text{s.t.}\ q\sum_{s=1}^{t}(\boldsymbol{x}_{Ms} - \boldsymbol{x}_F) + s_t - s_{t-1} = \sum_{s=1}^{t}(\boldsymbol{x}_{Ms} - \boldsymbol{x}_{As})/T \quad t = 2,\cdots,T$$

$$s_t \geq 0 \quad t = 1,\cdots,T$$

$$q \ \text{free}$$

The equality restrictions can be satisfied only by setting $s_t = x_t(q)$. Substituting $x_t(q)$ for s_t in (D) gives y_D. If spanning does not occur, then (P) is unbounded and (D) is infeasible. However, if spanning does occur, then The Duality Theorem for Linear Programming implies that (P) and (D) have the same solution value and hence $y_P = y_D \leq 0$. *Q.E.D.*

Proof of Theorem 2: Known results can derive the exact asymptotic sampling distribution of $\sum_{t\in\Theta}(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T$. Under the null, $\boldsymbol{x}_{it}$, $i \in \{M, A\}$, $t \in \Theta$, are serially and contemporaneously IID random variables with variance $s^2 < \infty$. Hence, the central limit theorem implies that $\sum_{t\in\Theta}\boldsymbol{x}_{it}/T$, $i \in \{M, A\}$, obey an asymptotically IID normal distribution with variance s^2/T, and $\sum_{t\in\Theta}(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T$, obeys an asymptotically normal distribution with zero mean and variance $2s^2/T$. Since the unity vector is a feasible solution to the primal problem, i.e. $\boldsymbol{e} \in \mathrm{B}$, we know that $y_P \leq \sum_{t\in\Theta}(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T$ for all return distributions $H(x)$. Moreover, there exist $H(x)$ for which $\sum_{t\in\Theta}(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T$ approximates y_P (see e.g. Post, 2001, Theorem 3), and therefore the asymptotic distribution of $\sum_{t\in\Theta}(\boldsymbol{x}_{Mt} - \boldsymbol{x}_{At})/T$ also represents the asymptotic least favorable distribution for y_P. *Q.E.D.*

REFERENCES

Arrow, K. (1971): '*Essays in the Theory of Risk-Bearing*', North-Holland Publishing Company, Amsterdam.

Bigelow, J. A. (1993): 'Consistency of mean-variance analysis and expected utility analysis: A complete characterisation', *Economic Letters* 43: 187-192.

Chvatal, V. (1983), *'Linear Programming'*, Freeman.

Davidson, R. and J.-Y. Duclos (2000), 'Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality', *Econometrica* 68, 1435-1464.

Falk, H. and H. Levy (1989): ‘Market Reaction to Quarterly Earnings’ Announcements: A Stochastic Dominance Based Test of Market Efficiency’, *Management Science* 35 (4), 425-446.

Friend, I. and Westerfield, R. (1980): Co-skewness and Capital Asset Pricing, *Journal of Finance* 35, 897-913.

Harvey, C. R. and A. Siddique (2000), 'Conditional Skewness in Asset Pricing Tests', *Journal of Finance* 55 (3), 1263-1295.

Haugen, R. A. and J. Lakonishok (1988): *'The Incredible January Effect'*, Dow Jones-Irwin, Homewood.

Huberman, G. and S. Kandel (1987): "Mean-Variance Spanning", *Journal of Finance* 42 (4), 873-888.

Hiriart-Urruty, J.-B. and C. Lemaréchal (1993), *'Convex Analysis and Minimization Algorithms I & II',* Springer Verlag.

Keim, D. B. (1983): 'Size-related anomalies and stock return seasonality: Further empirical evidence', *Journal of Financial Economics* 12, 13-32.

Kroll, Y. and H. Levy (1980): Sampling Errors and Portfolio Efficient Analysis, *Journal of Financial and Quantitative Analysis* 15, 655-688.

Larsen, G. A. and B. G. Resnick (1996): "Refining the Bootstrap Method of Stochastic Dominance Analysis: The Case of the January Effect," *Review of Quantitative Finance and Accounting* 7 (1), 65-79.

Levy, H. (1998): *Stochastic Dominance,* Kluwer Academic Publishers, Norwell, MA.

Nelson, R. D. and R. Pope (1991) 'Bootstrapped insights into empirical applications of stochastic dominance ', *Management Science* 37, 1182-1194.

Post, G. T. (2001): 'Stochastic Dominance in case of Portfolio Diversification: Linear Programming Tests', Erasmus Research Institute of Management (ERIM) Report ERS-2001-38-F&A, forthcoming in *The Journal of Finance*.

Reinganum, M. R. (1983): 'The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects', *Journal of Financial Economics* 12, 89-104.

Seyhun, H. N. (1993): ‘Can Omitted Risk Factors Explain the January Effect? A Stochastic Dominance Approach’, *Journal of Financial an Quantitative Analysis* 28, 195-212.

Figure 1: Rejection Rates for the Extended Kroll-Levy Experiment.

The figure gives the rejection rates for the null hypothesis (no spanning) based on 1000 random samples of $N \in \{100, 500, 2000\}$ observations, from a bivariate normal distribution with means $m_M = 0.20, m_A = 0.15$, standard deviations $s_M = s_A = 0.20$, and correlation coefficient $r \in \{0, 0.05, \cdots, 1\}$. For each sample, the null hypothesis of no spanning is rejected if and only if $y_P \geq \Phi^{-1}(1-a)\sqrt{2/T}\hat{s}$, using a significance level a =0.05.

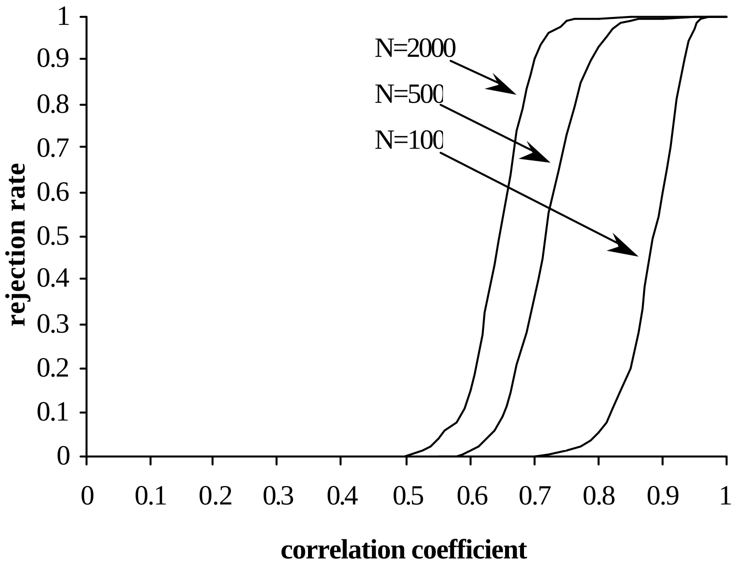

Table 1: Descriptive Statistics

Monthly dividend-adjusted returns from 1927 to 2000 for the ten value-weighted decile portfolios of NYSE, AMEX, and NASDAQ stocks. Panel A gives the descriptives for the full sample (894 observations). Panel B focuses on the observations for the month of January returns (74 observations). Source: Kenneth French data library at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

Panel A: Full sample						
	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
1st decile	0.014	0.105	3.396	30.703	-0.346	1.157
2nd decile	0.013	0.091	2.373	22.718	-0.329	0.946
3rd decile	0.013	0.082	1.785	17.704	-0.328	0.755
4th decile	0.012	0.076	1.554	15.321	-0.317	0.658
5th decile	0.012	0.074	1.313	14.294	-0.309	0.629
6th decile	0.012	0.070	1.015	11.834	-0.314	0.547
7th decile	0.012	0.067	0.938	11.972	-0.295	0.545
8th decile	0.011	0.063	0.787	10.957	-0.308	0.516
9th decile	0.011	0.060	0.697	11.237	-0.324	0.485
10th decile	0.010	0.052	0.072	6.725	-0.272	0.335
Panel B: January observations						
	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
1st decile	0.085	0.101	1.692	3.531	-0.066	0.431
2nd decile	0.060	0.085	1.642	5.619	-0.094	0.455
3rd decile	0.049	0.073	1.057	2.306	-0.105	0.318
4th decile	0.041	0.073	1.412	4.132	-0.092	0.354
5th decile	0.036	0.065	0.759	1.679	-0.096	0.248
6th decile	0.031	0.064	1.071	2.786	-0.089	0.286
7th decile	0.025	0.059	1.069	2.285	-0.081	0.227
8th decile	0.021	0.053	0.502	0.759	-0.083	0.187
9th decile	0.020	0.050	0.306	0.114	-0.084	0.157
10th decile	0.012	0.046	0.213	-0.209	-0.079	0.134

Table 2: Test Results

The table gives the observed value for the primal test statistic y_P, as well as the asymptotic least favorable p-value $1-\Phi\left(\frac{y_P}{\sqrt{2/T}\hat{s}}\right)$. Panel A gives results for the full sample (894 observations); Panel B gives the results for the month of January returns (74 observations). If the test statistic takes the value $-\infty$, then the primal problem (P) is unbounded and the dual (D) infeasible, and spanning does not occurs (see Section I).

Panel A: Full sample		
	Statistic	p-value
1st decile	0.000	0.500
2nd decile	0.001	0.429
3rd decile	$-\infty$	1.000
4th decile	$-\infty$	1.000
5th decile	$-\infty$	1.000
6th decile	$-\infty$	1.000
7th decile	$-\infty$	1.000
8th decile	$-\infty$	1.000
9th decile	$-\infty$	1.000
10th decile	$-\infty$	1.000
Panel B: January observations		
	Statistic	p-value
1st decile	0.000	0.500
2nd decile	0.004	0.413
3rd decile	0.035	0.016
4th decile	0.030	0.033
5th decile	0.049	0.002
6th decile	0.025	0.062
7th decile	0.060	0.000
8th decile	0.049	0.001
9th decile	0.059	0.000
10th decile	0.048	0.002

Publications in the Report Series Research* in Management

ERIM Research Program: "Finance and Accounting"

2002

A Stochastic Dominance Approach to Spanning
Thierry Post
ERS-2002-01-F&A

Testing for Third-Order Stochastic Dominance with Diversification Possibilities
Thierry Post
ERS-2002-02-F&A

Modeling the Conditional Covariance between Stock and Bond Returns: A Multivariate GARCH Approach
Peter De Goeij & Wessel Marquering
ERS-2002-11-F&A

2001

Non-Parametric Tests for Firm Efficiency in Case of Errors-In-Variables: Efficiency Depth
Timo Kuosmanen & Thierry Post
ERS-2001-06-F&A

Testing for Productive Efficiency with Errors-in-Variables. With an Application to the Dutch Electricity Sector
Timo Kuosmanen, Thierry Post & Stefan Scholtes
ERS-2001-22-F&A

Nonparametric Efficiency Estimation in Stochastic Environments (II): Noise-to-Signal Estimation, Finite Sample Performance and Hypothesis Testing
Laurens Cherchye & Thierry Post
ERS-2001-26-F&A

Unpredictable after all? A short note on exchange rate predictability
Gerard A. Moerman
ERS-2001-29-F&A

Testing for Stochastic Dominance with Diversification Possibilities
Thierry Post
ERS-2001-38-F&A

Regime Jumps in Electricity Prices
Ronald Huisman & Ronald Mahieu
ERS-2001-48-F&A

Methodological Advances in DEA: A Survey and an Application for the Dutch Electricity Sector
Laurens Cherchye & Thierry Post
ERS-2001-52-F&A

---

* A complete overview of the ERIM Report Series Research in Management:
http://www.ers.erim.eur.nl

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship

Informed Option Trading Strategies: The Impact on the Underlying Price Process
ERS-2001-55-F&A
Cyriel de Jong

Spanning and Intersection: A Stochastic Dominance Approach
ERS-2001-63-F&A
Thierry Post

LP tests for MV efficiency
ERS-2001-66-F&A
Thierry Post

The Economic Value of Predicting Stock Index Returns and Volatility
Wessel Marquering & Marno Verbeek
ERS-2001-75-F&A

Investment and Internal Finance: Asymmetric Information or Managerial Discretion?
Hans Degryse & Abe de Jong
ERS-2001-86-F&A

The Role of Self-Regulation in Corporate Governance: Evidence from The Netherlands
Abe de Jong, Douglas V. DeJong, Gerard Mertens & Charles Wasley
ERS-2001-87-F&A

The Dividend and Share Repurchase Policies of Canadian Firms: Empirical Evidence based on a New Research Design
Abe de Jong, Ronald van Dijk & Chris Veld
ERS-2001-88-F&A

The Impact of Institutional Differences on Derivatives Usage: A Comparative Study of US and Dutch Firms
Gordon M. Bodnar, Abe de Jong & Victor Macrae
ERS-2001-89-F&A

2000

Evaluating Style Analysis
Frans A. De Roon, Theo E. Nijman & Jenke R. Ter Horst
ERS-2000-11-F&A

From Skews to a Skewed-t: Modelling option-implied returns by a skewed Student-t
Cyriel de Jong & Ronald Huisman
ERS-2000-12-F&A

Currency Hedging for International Stock Portfolios: A General Approach
Frans A. de Roon, Theo E. Nijman & Bas J.M. Werker
ERS-2000-21-F&A