ESTUDIOS DE ECONOMÍA APLICADA VOL. 23-1, 2005. PÁGS. 7-25

A Stochastic Dominance Approach to Spanning. With an Application to the January Effect

THIERRY POST
Erasmus University Rotterdam .P.O. Box 1738, 3000 DR Rotterdam, The Netherlands».
Tfno. (+31) 10 - 408 14 28 Fax. (+31) 10 - 408 91 65. E-mail: gtpost@few.eur.nl

RESUMEN

En este trabajo se desarrolla una metodología tipo Dominio Estocástico para analizar si un inversor racional, insaciable y adverso al riesgo se beneficia de una particular expansión de sus posibilidades de inversión. Mediante el Dominio Estocástico se elimina la asunción simplificadora subyacente a la aproximación tradicional Media Varianza a este fenómeno. En este trabajo se extiende también la aplicación de esta metodología al análisis del comportamiento del mercado de pequeñas firmas en el mes de enero. Los resultados obtenidos sugieren que la explicación de este fenómeno, el Efecto Enero, no es congruente con la asunción simplificadora sobre el comportamiento de los rendimientos.

Palabras Clave: Spanning, Dominio Estocástico, Efecto Enero, Asimetría.

Una aproximación mediante la metodología del dominio estocástico al fenomeno del Spanning. Una aplicación al efecto Enero

ABSTRACT

We develop a Stochastic Dominance methodology to analyze whether rational non-satiable and risk averse investors benefit from a particular expansion of the investment possibilities. 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.

Keywords: Spanning, Stochastic Dominance, January Effect, Asymmetry.

Clasificación JEL: G10, G11, C15.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref.: e-23112

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)

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 elliptical or investor utility should be quadratic (see, e.g., Hanoch and Levy (1969), Section IV). 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., Kraus and Litzenberger (1976) and Harvey and Siddique (2000)). 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 set of 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) the lack of statistical 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 hypotheses. Second, Post (2003) presents tractable linear programming (LP) tests for SD efficiency in the case with diversification possibilities. These tests improve computational tractability and statistical power (all diversified portfolios are included in the analysis, which improves the likelihood of detecting inefficient portfolios).

U. *Since the utility functions of the investors are elements of this set, i.e., for all , we can obtain a weaker spanning condition:*

Definition 2 Asset A is SSDR spanned if and only if no rational non-satiable, risk-averse investor is better off by investing part of his or her wealth in A, i.e.:

$$\max_{\lambda \in \Lambda_1} \int u(\boldsymbol{x}^{\mathrm{T}}\lambda)dG(\boldsymbol{x}) = \max_{\lambda \in \Lambda_2} \int u(\boldsymbol{x}^{\mathrm{T}}\lambda)dG(\boldsymbol{x}) \quad \forall u \in U. \qquad [2]$$

SSDR *spanning* is related in a subtle way to the usual concepts of *efficiency* and *dominance*. Asset *A* is SSDR efficient or not-dominated if it is optimal for some rational non-satiable and risk averse investors. Unfortunately, this concept is relevant only if the equilibrium requires some investors to invest in *A* exclusively. By contrast, SSDR spanning occurs if all portfolios that include *A* are SSDR inefficient. This concept does not require that no investor invests in *A* exclusively. Rather, it requires that no investor invests part of his or her wealth in *A*. Throughout the text, we will test the hypothesis that spanning does not occur, i.e., some rational investors do invest at least part of their wealth in *A*. Rejection of this null gives strong evidence for a capital market 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 with . Since we will not use the timing of the draws, we are free to label the random draws by their ranking with respect to the risky benchmark asset, i.e., .[3] 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) , with , gives a statistically consistent estimator for the CDF.[4] 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:

$$\max_{\lambda \in \Lambda_1} \int u(\boldsymbol{x}^{\mathrm{T}}\lambda)dF(\boldsymbol{x}) = \max_{\lambda \in \Lambda_2} \int u(\boldsymbol{x}^{\mathrm{T}}\lambda)dF(\boldsymbol{x}) \quad \forall u \in U \Leftrightarrow \qquad [3]$$

---

[3] 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).

[4] However, there is substantial evidence that the distribution of assets returns (e.g. interest rates, risk premiums, 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.

$$\max_{\lambda \in \Lambda_1} \sum_{t \in \Theta} u(\boldsymbol{x}_t^{\mathrm{T}} \lambda)/T = \max_{\lambda \in \Lambda_2} \sum_{t \in \Theta} u(\boldsymbol{x}_t^{\mathrm{T}} \lambda)/T \quad \forall u \in U \, . \qquad [4]$$

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 (2003) to develop a more tractable approach. For simplicity, 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.,* $\min_{t \in \Theta} x_{it} < x_F < \sum_{t \in \Theta} x_{it}/T$ *for . Under this assumption, some investors will invest part of their wealth in the riskless asset, but no investor will invest all of his or her 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).*

The following two test statistics can apply Definition 3 to empirical data:

$$\psi_P \equiv \inf_{\beta \in \mathrm{B}} \left\{ \sum_{t \in \Theta} \beta_t (x_{Mt} - x_{At})/T : \sum_{t \in \Theta} \beta_t (x_{Mt} - x_F)/T \geq 0 \right\}, \qquad [5]$$

with $\mathrm{B} \equiv \left\{ \boldsymbol{\beta} \in \Re_+^T : \beta_1 \geq \beta_2 \geq \cdots \geq \beta_T = 1 \right\}$, and

$$\psi_D \equiv \sup_{\theta} \left\{ \xi_T(\theta) : \xi_t(\theta) \geq 0 \quad \forall t \in \Theta \right\}, \qquad [6]$$

with $\xi_t(\theta) \equiv \sum_{s=1}^{t} ((1-\theta)x_{Ms} - x_{As} + \theta x_F)/T$.

These test statistics are derived from the optimality conditions for convex optimization problems (see the proof to Theorem 1). The statistic basically checks if these conditions hold for some utility function . The variables $\beta \in \mathrm{B}$ represent a supergradient vector $\partial u(\boldsymbol{x}_M) \equiv (\partial u(x_{M1}) \cdots \partial u(x_{MT}))^{\mathrm{T}}$ for a utility function ; represents the restrictions that follow from the assumptions of nonsatiation and risk aversion. The statistic ψ_D follows from the linear programming dual of ψ_P.

Theorem 1 Asset A is empirically SSDR spanned if and only if $\psi_P = \psi_D \geq 0$.

The test statistics ψ_P and ψ_D can be computed by straightforward linear programming (LP); full LP formulations are included as (P) and (D) in the proof in appendix. 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 $\sum_{t\in\Theta} x_{At}/T > \sum_{t\in\Theta} x_{Mt}/T$. 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.)

To develop some intuition for the theorem, Figure 1 shows a simple example with two periods, i.e., $\Theta = \{1,2\}$. The initial portfolio possibilities set is *MF*, and the efficient set is *MF* (excluding *F*; recall Arrow's theorem). Introducing the additional asset can change the possibilities set and the efficient set. The figure displays two different cases, labeled *A* and *A'*, and the extended possibilities sets *MFA* and *MFA'*. In the first case, with new asset *A*, spanning occurs; *MF* (excluding *F*) remains the efficient set. In fact, spanning occurs for all new assets included in the gray area $S \equiv \{\boldsymbol{x} \in \Re^2 : x_1 + x_2 \leq (1-\theta)(x_{M1} + x_{M2}) + 2x_F\theta, x_1 \leq (1-\theta)x_{M1} + x_F\theta \quad \theta \geq 0\}$ In the second case, with new asset *A'*, spanning does not occur. The efficient set changes from *MF* (excluding *F*) to *F'A'* (excluding the riskless *F'*) and all investors invest at least part of their wealth in the new asset.

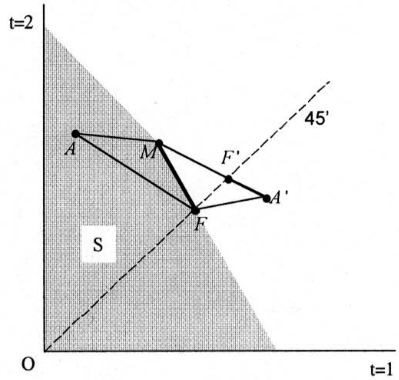

Figure 1
Two Period, Three Assets Example

Adding a third asset to *M* and *F* expands the investment possibilities. Introducing *A* does not affect the efficient set *MF* (excluding *F*), and hence spanning occurs. By contrast, introducing *A'* changes the efficient set to *F'A'* (excluding *F'*) and spanning does not occur.

2. SAMPLING ERROR

The test statistics ψ_P and ψ_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 (2003) derived an analytical characterization of the asymptotic sampling distribution of his efficiency tests. This section extends the Post results towards the SSDR test statistic . (Duality implies that the results apply with equal strength to ψ_D.)

There are various hypotheses that could serve as the null hypothesis in a test procedure. In SD analysis, a typical null hypothesis (H_0) 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 $H : \Re \rightarrow [0,1]$ with variance $\sigma^2 < \infty$. The shape of the distribution of ψ_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 ψ_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 asymptotic least favorable distribution,* ψ_P *behaves as a normal random variable with mean zero and variance* $2\sigma^2/T$.

The theorem implies that p-values $P(\psi_P \geq y | H_0)$ may be found as $\Phi\left(-y/(\sqrt{2/T}\sigma)\right)$, 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

[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/.

p-values can be very large and a naïve approach to the test statistic (reject efficiency if $\psi_P > 0$) is unlikely to yield anything but noise. A more sound approach is to compare the p-value for the observed value of ψ_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 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:

$$\hat{\sigma}^2 \equiv \sum_{i\in\{M,A\}} \sum_{t\in\Theta} (x_{it} - \sum_{t\in\Theta} x_{it}/T)^2 / 2T \quad [7]$$

[6]

3. SIMULATION EXPERIMENT

To assess the goodness of the above test procedure, 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 . 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.

---

[6] This is simply the equally weighed average of the sample variance of and ; 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.

	No spanning in population ($\rho < 0.50$)	Spanning in population ($\rho \geq 0.50$)
No spanning in sample ($\psi_P < \Phi^{-1}(1-\alpha)\sqrt{2/T}\hat{\sigma}$)	No Error	Type II error
Spanning in sample ($\psi_P \geq \Phi^{-1}(1-\alpha)\sqrt{2/T}\hat{\sigma}$)	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 $\rho \in \{0, 0.05, \cdots, 1\}$ and for samples of size $N \in \{100, 500, 2000\}$. The nominal level of significance α (or the size for the asymptotic least favorable distribution) is set at 5 percent.

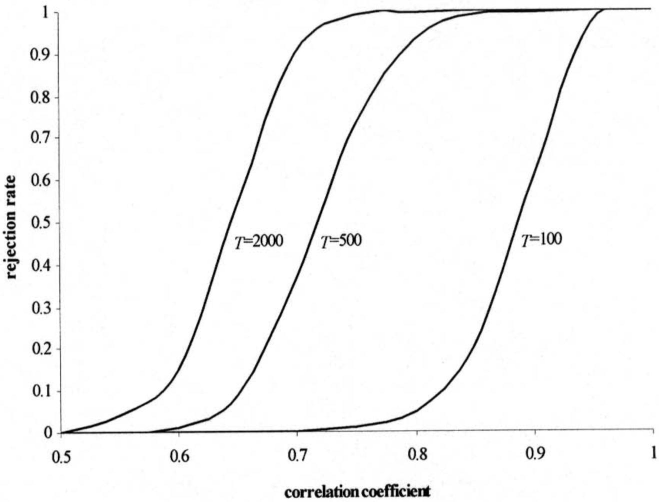

Figure 2
Rejection Rates for the Extended Kroll-Levy Experiment

The figure gives the rejection rates for the null hypothesis (no spanning) based on 10,000 random samples of $T \in \{100, 500, 2000\}$ observations, from a bivariate normal distribution with means $\mu_M = 0.20, \mu_A = 0.15$, standard deviations $\sigma_M = \sigma_A = 0.20$, and correlation coefficient $\rho \in [0.5, 1]$. For each sample, the null hypothesis of no spanning is rejected if and only if $\psi_P \geq \Phi^{-1}(1-\alpha)\sqrt{2/T}\hat{\sigma}$, using a significance level $\alpha = 0.05$.

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., ρ 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.

4. THE JANUARY EFFECT

Empirical evidence suggests that the stock market returns of small firms are abnormally high 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 explanation 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.[8] 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. Table 1 gives some descriptive statistics for the data set.[9]

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

---

[8] Typically, previous papers that have examined stochastic dominance have not only included size decile portfolios but also equally- and value-weighted indices, so as to account for diversification possibilities. This approach makes sense if investor preferences are homogeneous. In this case, the market portfolio must be efficient relative to portfolios formed on size. The standard crossing algorithms can then be used to test if the market is dominated by any of the ten size deciles. Similarly, the Post (2003) test can be used to test for efficiency of the market

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 test statistic ψ_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 $\hat{\sigma}^2$ 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 decile the full sample, spanning occurs for none of the 9 higher 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.[10] 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 2).

---

relative to all possible portfolios constructed from the ten size deciles. However, these tests are no longer relevant if investors have heterogeneous preferences. Dybvig and Ross (1982) have demonstrated that the SSD efficient set generally is not convex, and hence, there is no guarantee that the market portfolio is efficient if different investors hold different portfolios. (In this case, there is no 'size anomaly' if, e.g., the market portfolio and the large decile portfolios are inefficient, but different investors hold different portfolios that include all ten deciles.) A test for spanning effectively tests if all assets are included in some efficient portfolio (not necessarily with a weight that equals the relative market capitalization). This test is also relevant if different investors hold different portfolios.

[9] 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.

[10] Each p-value corresponds to the hypothesis that a single decile portfolio is not spanned. If we test the joint hypothesis that all decile portfolios are not spanned, then we should adjust the significance level, so as to avoid a 'fishing expedition'; increasing the number of portfolios increases the likelihood of finding small p-values. For example, a Bonferroni correction (see, e.g., Miller (1981, pages 6-8)) uses a significance level of /N for each of N individual tests, which guarantees that the overall significance level is less than . Using this approach for our study, we can reject thejoint hypothesis that no decile portfolio is spanned with at least 99 percent confidence.

Table 2
Test Results

The table gives the observed value for the primal test statistic ψ_P, as well as the asymptotic least favorable *p*-value $1-\Phi\left(\psi_P/(\sqrt{2/T})\hat{\sigma}\right)$. 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 occur (see Section II).

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

5. 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 risky benchmark assets and all new assets (e.g., using a two-fund separation theorem). 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 check if the hyperplanes support all new assets. This boils down to simply applying our test for all new assets. (Section V 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 ψ_P and ψ_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 asymptotic 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. The bootstrap potentially offers more power than the analytical characterization in Theorem 2, as the theorem uses the least favorable distribution that minimizes Type I error at the cost of Type II errors. Of course, this benefit has to be balanced against the computational burden associated with bootstrapping.

APPENDIX

Proof to Theorem 1 We first consider the sufficient condition. The problem $\max_{\lambda\in\Lambda_2}\sum_{t\in\Theta}u(\boldsymbol{x}_t^{\mathrm{T}}\lambda)/T$, $u\in U$, maximizes a concave objective function over a convex set. Using, $\lambda_2^*(u)$ for the optimal portfolio relative to and, the (necessary and sufficient) 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 $\mathrm{X}^{\mathrm{T}}\lambda_2^*(u)$ 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 function used below). The assets that are included in the optimal portfolio should lie on the tangent hyperplane and the assets that are not included should lie below it. Arrow's (1971) theorem (see Section II) implies that *M* is always included, i.e., $\lambda_{2M}^*(u)>0$, and the optimality conditions therefore amount to:

$$\sum_{t\in\Theta}\partial u(\boldsymbol{x}_t^{\mathrm{T}}\lambda_2^*(u))(x_{Mt}-x_F)/T\geq 0, \qquad \textbf{[A1]}$$

If *A* is empirically SSDR spanned, then $\boldsymbol{x}^{\mathrm{T}}\lambda_1^*(u)=\boldsymbol{x}^{\mathrm{T}}\lambda_2^*(u)$ and $\lambda_{1A}^*(u)=0$, and the optimality conditions also imply

$$\sum_{t\in\Theta}\partial u(\boldsymbol{x}_t^{\mathrm{T}}\lambda_2^*(u))(x_{Mt}-x_{At})/T\geq 0, \qquad \textbf{[A2]}$$

for all $u\in U$. By contrast, if *A* is not spanned, then $\lambda_{1A}^*(u)>0$ and A must lie above the tangent hyperplane through $\mathrm{X}^{\mathrm{T}}\lambda_2^*(u)$, i.e.:

$$\sum_{t\in\Theta}\partial u(\boldsymbol{x}_t^{\mathrm{T}}\lambda_2^*(u))(x_{Mt}-x_{At})/T< 0, \qquad \textbf{[A3]}$$

for some $u\in U$. By construction, $\partial u(\boldsymbol{x}^{\mathrm{T}}\lambda_2^*(u))$ is a feasible solution, i.e., $\partial u(\boldsymbol{x}^{\mathrm{T}}\lambda_2^*(u))\in\mathrm{B}$ (recall that all portfolios of M and F have the same ranking as M). The inequalities (A1) and (A3) imply that this solution is associated with a strictly negative solution value. Hence, spanning does not occur only if $\psi_P<0$, or, alternatively, spanning occurs if $\psi_P\geq 0$.

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

$$\sum_{t\in\Theta}\beta_t^*(x_{Mt}-x_{At})/T<0;\sum_{t\in\Theta}\beta_t^*(x_{Mt}-x_F)/T\geq 0, \qquad \textbf{[A4]}$$

with $\beta^*\in B$ for the optimal solution. From β^*, we may construct two piecewise linear utility functions: (1) $p(x)\equiv\min_{t\in\Theta}(\alpha_t+\beta_t^*x)$, with

$\alpha_t \equiv 0.5\sum_{s=t}^{T-1}(\beta_{s+1}^* - \beta_s^*)(x_{Mt} + x_{Mt+1})$, *and (2)* $p^*(x) \equiv p((x - \lambda_{1F}^*(p)x_F)/\lambda_{1M}^*(p))$.

By construction, both functions are monotone increasing and concave and hence $p(x) \in U$ and $p^*(x) \in U$. For the second function, we have $\partial p^*(\mathbf{x}^{\mathrm{T}}\lambda_1^*(p)) = \partial p(x_M) = \beta^*$. Hence, (A4) implies that A is included in the optimal portfolio for $p^*(x)$. Consequently, if $\psi_P < 0$, then spanning does not occur, or, alternatively, spanning occurs only if $\psi_P \geq 0$.

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

$$\inf \sum_{t=1}^{T} \beta_t (x_{Mt} - x_{At})/T \qquad \textbf{[P]}$$

$$\text{s.t.} \sum_{t=1}^{T} \beta_t (x_{Mt} - x_F)/T \geq 0$$

$$\beta_t - \beta_{t+1} \geq 0 \quad t = 1,\cdots,T-1$$

$$\beta_T = 1$$

$$\beta_t \text{ free} \quad t = 1,\cdots,T$$

The LP dual of (P) is:

$$\sup s_T \qquad \textbf{[D]}$$

$$\text{s.t. } \theta(x_{M1} - x_F)/T + s_1 = (x_{M1} - x_{A1})/T$$

$$\theta \sum_{s=1}^{t}(x_{Ms} - x_F) + s_t - s_{t-1} = \sum_{s=1}^{t}(x_{Ms} - x_{As})/T \quad t = 2,\cdots,T$$

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

$$\theta \text{ free}$$

The equality restrictions can be satisfied only by setting $s_t = \xi_t(\theta)$. Substituting $\xi_t(\theta)$ for s_t in (D) gives ψ_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 $\psi_P = \psi_D \leq 0$.Q.E.D.

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

REFERENCES

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

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.

DYBVIG, PHILLIP AND STEPHEN ROSS, 1982, *Portfolio efficient sets,* Econometrica *50, 1525-1546.*

EFRON, BRADLEY, 1979, *Bootstrap methods: Another look at the jackknife,* Annals of Statistics *7, 1-26.*

EFRON, BRADLEY AND GAIL GONG, 1983, *A leisurely look at the bootstrap, the jackknife, and cross validation,* The American Statistician *37, 36-48.*

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.*

HANOCH, G., AND H. LEVY (1969): *The Efficiency Analysis of Choices Involving Risk,* Review of Economic Studies *36, 335-346.*

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.*

HIRIART-URRUTY, J.-B. AND C. LEMARÉCHAL (1993), 'Convex Analysis and Minimization Algorithms I & II', *Springer Verlag.*

HUBERMAN, G. AND S. KANDEL (1987): *«Mean-Variance Spanning»*, Journal of Finance *42 (4), 873-888.*

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

KRAUS, ALAN AND ROBERT LITZENBERGER, 1976, *Skewness preference and the valuation of risk*

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.*

MILLER, RUPERT G., 1981*,* Simultaneous Statistical Inference, Second Edition, *New York, Springer-Verlag.*

NELSON, R. D. AND R. POPE (1991) *'Bootstrapped insights into empirical applications of stochastic dominance ',* Management Science *37, 1182-1194.*

POST, G. T., 2003, *'Empirical Tests for Stochastic Dominance Efficiency',* Journal of Finance *58, 1905-1932.*

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.*