Improved Portfolio Choice Using Second-Order Stochastic Dominance

by

James E. Hodder

Jens Carsten Jackwerth

Olga Kolokolova

September 19, 2013

Abstract

Constructing portfolios based on second-order stochastic dominance (SSD) is theoretically attractive since all risk-averse investors would prefer a dominating portfolio. However, choosing among SSD efficient portfolios is a challenge without an obvious ranking metric. We explore a particular choice based on Kuosmanen (2004) and compare its performance to other SSD-related strategies and to standard portfolio choice approaches. The SSD-related choices (including the Kuosmanen approach) outperform portfolios based on the Sharpe ratio, equal weights, and the information ratio. Portfolios based on minimum variance that also match the benchmark's mean return perform on a par with the SSD-related choices.

JEL: G12, G13, G14

Keywords: Second Order Stochastic Dominance, Portfolio Optimization, Performance Measurement

James E. Hodder, 5 Lakes Financial Research, jhodder@bus.wisc.edu

Jens Jackwerth, University of Konstanz, jens.jackwerth@uni-konstanz.de

Olga Kolokolova, Manchester Business School, olga.kolokolova@mbs.ac.uk

1. Introduction

In this paper, we use the concept of second-order stochastic dominance (SSD) to construct portfolios. We shall see that using SSD in-sample allows us to construct portfolios that often dominate competing portfolio choice methodologies out-of-sample.

Frequently, portfolio construction has based weights on mean-variance optimization; however, the basic mean-variance criterion has well-known limitations. It is symmetric, and its theoretical justification requires either a quadratic utility function or multivariate normality of returns. It thus considers only the first two moments of the return distribution; and this can be a substantial problem when returns are not, at least approximately, normally distributed. In situations where an actively managed investment fund has heterogeneous investors, portfolio selection procedures based (perhaps implicitly) on a single, specific utility function will invariably lead to undesirable allocations for some of these investors. This problem can be exacerbated when returns are substantially non-normal since the mean-variance approximation may turn out to be quite poor for many of those investors. Furthermore, optimization procedures used to construct mean-variance portfolios in practice often result in extreme portfolio weights when using historical inputs that contain estimation errors relative to the true underlying return distributions.

We propose to rank portfolio return distributions based on second-order stochastic dominance as a comparison criterion. If a return distribution "A" second-order stochastically dominates another distribution "B", then all risk-averse investors with increasing and concave utility functions will prefer A to B. Thus, SSD can be used to identify dominating return distributions that would be preferred to the dominated distribution by all risk-averse investors in a large fund without needing to have detailed knowledge of their individual preferences. Importantly, the SSD criterion also does not focus on a limited number of moments but accounts for the complete return distribution, considering both gains and losses. Furthermore, tests for SSD do not need any distributional assumptions for their implementation.

SSD is a powerful tool for ranking distributions. It has been used, for example, to analyze aggregated investor preferences and beliefs by Post and Levy (2005). De Giorgi (2005) as well as Russell and Seo (1980) have applied the SSD concept to a theoretical portfolio choice problem and discuss the properties of the SSD criterion compared to the mean-variance approach. They show that the sets of mean-variance efficient portfolios and SSD efficient portfolios overlap but do not coincide.

The concept of stochastic dominance has been empirically applied to the portfolio choice problem by Post (2003), Kuosmanen (2004), as well as Post and Kopa (2013). These authors test for stochastic dominance of a specified portfolio (the market portfolio) with respect to all other portfolios that can be constructed in a given asset span. Only the test procedure of Kuosmanen (2004) additionally identifies an efficient portfolio that dominates the evaluated portfolio if the latter is not efficient. Going one step further, Scaillet and Topaloglou (2010) augment the testing procedures of Post (2003) and Kuosmanen (2004) to allow for time-varying return distributions and test for the SSD efficiency of the market portfolio. The main limitation of all these works is that they only analyze in-sample performance. For practical portfolio allocation problems, it is essential to establish the out-of-sample properties of SSD-efficient portfolios.

Out-of-sample stochastic dominance analysis was conducted by Meyer, Li, and Rose (2005). These authors consider the benefits of international portfolio diversification compared with a New Zealand-only portfolio. They use the concept of third-order stochastic dominance, arguing that their second-order stochastic dominance tests lacked power. Their in-sample portfolio choice, however, is still conducted using the mean-variance approach with a fixed target return. Thus, prior empirical work on portfolio allocation using the SSD concept has been largely restricted to in-sample analysis and, in any case, did not rely on the SSD criterion for estimating portfolio weights themselves.

In this paper, we extend the above work in several ways. We propose to determine the optimal in-sample portfolio based on the SSD criterion and use the algorithm developed in Kuosmanen (2004) to find optimal portfolio weights. We next test whether this SSD-based portfolio dominates the benchmark portfolio out-of-sample, where we use the non-dominance test of Davidson (2009). Such out-of-sample assessment allows us to properly judge the performance of different portfolio allocations going forward.

We also compare the performance of our SSD-based portfolio with several other competing portfolio choice approaches. The comparison alternatives include three mean-variance-related portfolios: maximum Sharpe ratio (MaxSharpe), maximum Information ratio (InformationRatio), and a portfolio with the minimum possible variance given the same in-sample mean return as the benchmark (MinVarBench). One could think of this latter portfolio as improving on the benchmark by shifting it onto the mean-variance efficient frontier. We also use three alternative portfolios that we describe as SSD-related since their focus on minimizing risk is conceptually related to the SSD approach. Those SSD-related portfolios use minimum-variance (MinVar), minimum semi-variance (MinSemivar), and minimum-shortfall (MinShortfall) as their portfolio construction criteria. Another comparison is an equally-weighted portfolio (Equal), which DeMiguel, Garlappi, and Uppal (2009) found to perform on a par with several more complex portfolio choice mechanisms.

One feature of second order stochastic dominance is that portfolios can at times not be uniquely ranked according to SSD. Thus, the investor needs to decide on how to pick portfolios out of the set of SSD efficient portfolios. A further contribution of our paper is to compare some possible ways of picking one portfolio out of such an efficient set.

We evaluate performance of all portfolios with respect to the market benchmark portfolio, where we proxy market performance by the returns on the CRSP all-share index. The analysis is conducted using non-overlapping yearly windows of daily returns on Fama-French 49 industry portfolios from January 1927 to December 2012. In the robustness section, we also investigate using monthly returns over this same sample period.

To illustrate some implications of differences between these portfolio construction approaches, we plot in Figure 1 an example of the mean-variance location for portfolios chosen according to the eight approaches mentioned above.[1] The light grey area indicates the set of portfolios that have positive test

[1] This illustration uses only 4 assets – portfolio indices for daily returns of stocks, bonds, commodities, and REITs. With this limited number of assets, we can exhaustively test all possible portfolios (with weight increments of 0.05)

statistics using Davidson (2009) with respect to the benchmark (Bench). Thus, those portfolios can potentially dominate Bench. The darker gray area indicates portfolios for which the hypothesis of non-dominance over Bench is rejected at the 10% significance level. In this particular example, MinVar, MinSemivar and MinShortfall lie close together and do not dominate Bench. Instead, they are in the same dominance class as Bench. Kuosmanen is located on the mean-variance efficient frontier; and so are MaxSharpe, MinVarBench, and MinVar. The benchmark (Bench), equally-weighted (Equal), and InformationRatio portfolios are well inside the mean-variance efficient frontier.

We find that SSD-related approaches perform well out-of-sample using the SSD criterion, and often dominate the benchmark according to the third order stochastic dominance (TSD) criterion. In other words, the typical market portfolio as proxied by the benchmark is often not at all well-structured from a SSD/TSD perspective. Prior research indicated the market portfolio could frequently be dominated in-sample, but our results also suggest it is frequently dominated out-of-sample. The minimum variance portfolio with the same mean as the benchmark also performs well. The equally-weighted portfolio performs worse and roughly on a par with the benchmark. The portfolios based on the information ratio and the Sharpe ratio also tend to perform poorly. Our qualitative results are robust to numerous checks including variation in the benchmark.

We thus feel that SSD-based portfolio construction offers a superior approach for actively managed funds with heterogeneous investors. It identifies portfolios which will be acceptable to any risk-averse investor. Out-of-sample, SSD-based portfolios generally perform well not only in terms of SSD and TSD but also on several more traditional performance measures including Sharpe ratio, certainty equivalent value, and turnover.

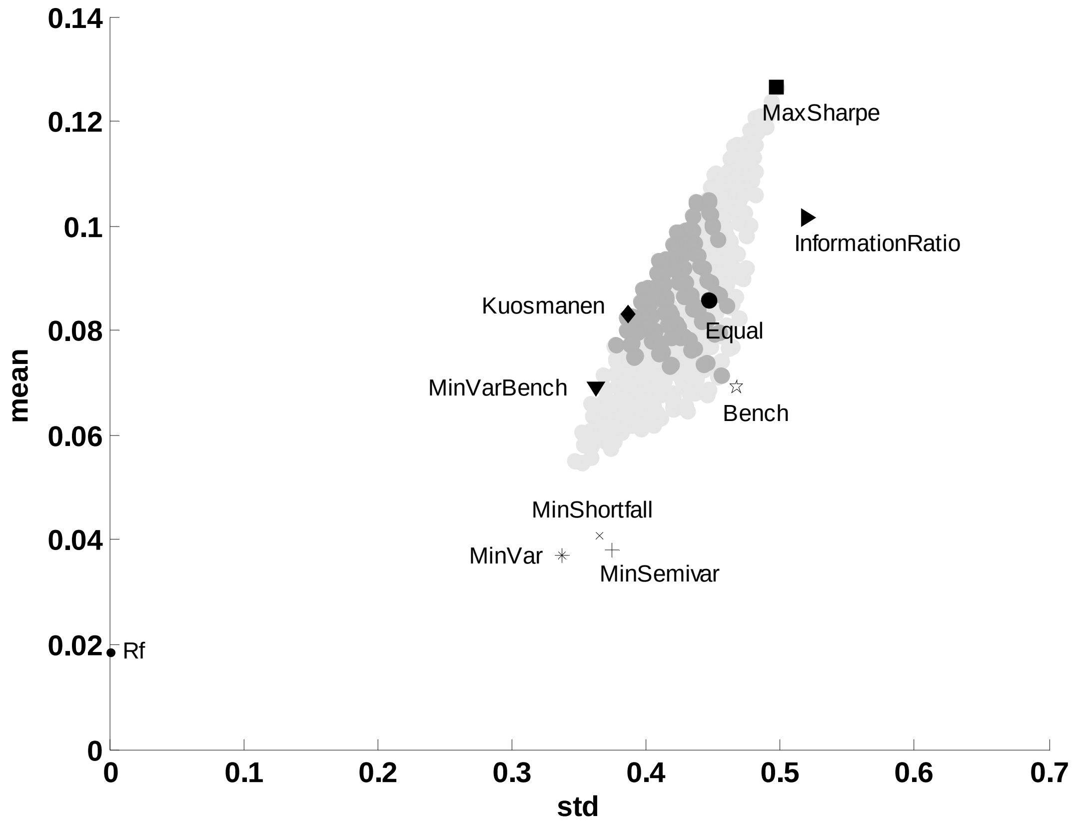
---

using the Davidson (2009) non-dominance test to find the set that have positive and significant (at the 10% level) test statistics. The example year is 1996.

We discuss the methodology in Section 2, while data are described in Section 3. Empirical results follow in Section 4. Numerous robustness checks are provided in Section 5 and in an Online Appendix. Section 6 concludes.

Figure 1. Mean-Variance Location of Competing Portfolios

The figure depicts an example of the mean-variance location for several portfolios constructed using competing approaches. The data stem from year 1996. The light grey area covers all the portfolios that have a positive test statistic with respect to Bench, and darker grey area covers the portfolios that statistically dominate the benchmark in the SSD sense at the 10% significance level, when 5% tail trimming is used.



2. Methodology

Consider a specified benchmark-portfolio (Bench) of *s* assets which is held for a time period from t_0 - Δt to t_0. Our main analysis uses a yearly time period with daily returns. For the same (in-sample) time period, we construct the SSD-based portfolio (labeled Kuosmanen) which is the optimal portfolio obtained using the Kuosmanen (2004) LP approach, detailed below. That procedure measures the degree of inefficiency for the benchmark portfolio in terms of SSD; and the objective function can be interpreted as the maximum increase of mean return that could be obtained by choosing a portfolio from the subset that dominates the benchmark portfolio in the SSD sense.

We also create several competing portfolios using the same in-sample data. The first group of alternative portfolios, labeled SSD-related, consists of approaches minimizing some SSD-consistent risk measures.[2] This group includes: a) the global minimum variance portfolio (MinVar), b) the global minimum left semi-variance portfolio (MinSemivar), and c) the minimum expected shortfall portfolio (MinShortfall).

The second group of competing portfolios we label mean-variance related. This group includes three mean-variance type portfolios: a) the portfolio with the highest in-sample Sharpe ratio (MaxSharpe), b) the portfolio with the highest information ratio (expected excess return above the benchmark divided by the standard deviation of this excess return) with respect to Bench (InformationRatio), and c) the minimum-variance portfolio which has the same mean return as Bench (MinVarBench). A practical problem with these portfolios is their tendency to have unstable and sometimes extreme weights on individual securities due to the characteristics of mean-variance optimization coupled with estimation error in the parameter inputs -- see for example, Michaud (1989), Jorion (1992), as well as DeMiguel, Garlappi, and Uppal (2009). As a result, the mean-variance related

---

[2] SSD-consistent risk measures rank portfolios the same way as the SSD criterion. That is, if A dominates B from a SSD perspective, A will have smaller variance, semi-variance, etc. (so, these measures are SSD consistent). On the other hand, measures based on the Sharpe ratio or CEV are examples which are not SSD consistent.

methods frequently exhibit poor out-of-sample performance. In response to this problem of weight instability, we include in our comparison group an equally weighted portfolio (Equal) which DeMiguel, Garlappi, and Uppal (2009) found to perform relatively well in their analysis.

The optimal weights for all these portfolios are determined using the in-sample data from $t_0 - \Delta t$ to t_0. Using those portfolio weights, we then compute the out-of-sample returns of those portfolios for the period t_0 to $t_0+\Delta t$. The performance of the portfolios is compared with the benchmark's out-of-sample return based on both traditional and stochastic-dominance performance measures, all of which we discuss later.

We repeat the analysis using T non-overlapping windows. The former out-of-sample period becomes the new in-sample period for portfolio weight estimation, and the performance is then measured for the next out-of-sample period from $t_0+\Delta t$ to $t_0+2\Delta t$. Rolling this procedure forward results in T-1 out-of-sample periods.

We impose short-sale constraints in the portfolio selection process which ensures that all our constructed portfolios are feasible choices for delegated money management structures, where shorting is frequently not allowed. Thus, portfolio weights are restricted to be non-negative and sum to one for each of the considered portfolios. The following sub-sections address the above steps in more detail.

2.1. In-sample portfolio optimization

2.1.1. Constructing portfolios using SSD

Graphically, second-order stochastic dominance (SSD) implies that two cumulative distribution functions cross; but the area under the dominating distribution is always smaller or equal to that of the dominated distribution for each threshold level z. If those cumulative distribution functions do not cross, first-order stochastic dominance is observed.

Formally, distribution A with cumulative distribution function $F_A(y)$ is said to second-order stochastically dominate another distribution B with cumulative distribution function $F_B(y)$ if, for all possible threshold levels z, the expected losses with respect to this threshold in distribution A are not larger than that in distribution B with at least one strict inequality for some level of z.

$$\int_{-\infty}^{z} (z-y)dF_A(y) \leq \int_{-\infty}^{z} (z-y)dF_B(y),\ \forall \mathrm{z} \in \mathrm{R} \tag{1}$$

In constructing our SSD-based portfolio, we adopt an LP algorithm developed in Kuosmanen (2004). This procedure allows for testing if a benchmark portfolio return distribution is SSD efficient relative to a given asset span. If the benchmark is not efficient, the optimal solution also delivers a vector of portfolio weights corresponding to a well-diversified portfolio that second-order stochastically dominates the benchmark. We use this in-sample optimal portfolio as our SSD-based competing portfolio.

We considered several other alternative approaches, but they seem to be inferior to Kuosmanen (2004). For example, the algorithm developed in Post (2008) was generating too little change in the optimal portfolio weights relative to the benchmark portfolio in our sample. This algorithm is a simpler version of Kuosmanen (2004) and thus more limiting. Shalit and Yitzhaki (1994) only allowed the assessment of marginal investments in a stock, given a particular benchmark portfolio. We would like to investigate portfolio changes which are larger than purely marginal, and Kuosmanen (2004) allows for such changes. Clark, Jokung, and Kassimatis (2011) provide an algorithm which uses the marginal improvements analyzed by Shalit and Yitzhaki (1994) and then moves in 0.1% portfolio weight reallocations towards a portfolio where no further marginal improvement exists. With a substantial number of assets, their iterative approach is tedious; and we prefer the more elegant LP based approach of Kuosmanen (2004).

We implement the following LP procedure of Kuosmanen (2004), with changed notation conforming to our paper and constraints specified to match our requirement of non-negative weights[3]:

$$\theta_2(y_{Bench}) = \max_{\lambda,W}\left(\sum_{t=1}^{\tau}\sum_{i=1}^{N} y_{it}\lambda_i - \sum_{t=1}^{\tau} y_{Bench,t}\right)\frac{1}{\tau}$$

$$s.t.\ \sum_{i=1}^{N} y_{it}\lambda_i \geq \sum_{j=1}^{\tau} W_{tj}\, y_{Bench,t} \quad \text{for } t = 1,...,\tau$$

$$W \in \left\{ [W_{ij}]_{T\times T} \middle| 0 \leq W_{ij} \leq 1; \sum_{i=1}^{\tau} W_{ij} = \sum_{j=1}^{\tau} W_{ij} = 1\ \forall i, j = 1,...,\tau \right\} \tag{2}$$

$$0 \leq \lambda_i \leq 1, \forall i = 1,...,N$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

Here the asset span consists of N assets, with τ daily return observations y_{it} each. We use $y_{Bench,t}$ to denote the benchmark portfolio return at time t. The vector of portfolio weights to be optimized is denoted λ, and W is a doubly stochastic matrix. The vector of optimal portfolio weights λ^* from the above procedure is used to construct the SSD-based portfolio.[4]

2.1.2. Constructing competing portfolios

In constructing competing portfolios, we start with a group of portfolios based on risk measures consistent with SSD, such as left semi-variance (MinSemivar) and expected shortfall (MinShortfall) -- see for example Porter (1974), Fishburn (1977), and Ogryczak and Ruszczynski (1999). For MinSemivar, the

[3] Kuosmanen (2004) also provides a simplified version of the LP problem which uses only positive weights. We report here, however, the general version, in which various linear constraints on the weight vector can be incorporated.

[4] The above LP algorithm, generally, tests only for the necessary condition of SSD. We also implement the sufficient test for SSD as in Kuosmanen (2004), equation (3), and obtain identical results for all cases in our sample.

portfolio weights are chosen to minimize the in-sample left semi-variance subject to the short-sale constraint. MinShortfall chooses weights (subject to the short-sale constraint) that minimize the expected shortfall below the 5% quantile of the in-sample portfolio returns. The minimum variance portfolio is always SSD efficient and cannot be dominated in-sample, see Russell and Seo (1980). Consequently, we also include the global minimum-variance portfolio with short-sale constraints (MinVar) in our set of alternative portfolios.

We next consider the mean-variance related group of portfolio choices. We compute the maximum Sharpe ratio portfolio (MaxSharpe), where we proxy for the risk-free rate by using returns on the 90-day Treasury bill. The maximum Information ratio portfolio (InformationRatio) is computed by choosing weights that maximize the difference between in-sample mean returns for that portfolio and Bench, scaled by standard deviation of the tracking error between this portfolio and Bench. When finding the optimal weights for these portfolios, we include short-sale constraints, which has the added benefit of reducing the sensitivity of mean-variance optimization to estimation errors, outliers, and mistakes in the data – see Jagannathan and Ma (2003) who use short-sale constraints in combination with a minimum-variance portfolio.

To stabilize estimated weights, different approaches have been used by various authors. Kan and Zhou (2007), for example, use a mixture of mean-variance and minimum-variance portfolios. Following this path, we construct another alternative portfolio (MinVarBench), in which the variance is minimized while the mean is restricted to equal the in-sample mean for Bench.

Other techniques to improve mean-variance portfolio construction exist; and DeMiguel, Garlappi and Uppal (2009) compare the performance of 14 different models with the naive equally-weighted scheme. They find that none of the advanced models consistently outperforms the simple equally-weighted strategy out-of-sample based on three comparison criteria: the out-of-sample Sharpe ratio, the certainty-equivalent return for a mean-variance investor, and turnover measured as trading volume. This is in line with Martellini and Ziemann (2010), who argue that estimation errors often offset the benefits of rather complicated optimal portfolio choice approaches. Moreover, the equally-weighted portfolio allocation strategy is preference free, delivers a reasonable level of diversification, and does not rely on any estimation (thus, it does not incorporate estimation errors). This led us to include the equally-weighted portfolio (Equal) as a competing portfolio in our analysis.

2.2. Out-of-sample portfolio performance assessment

2.2.1. Stochastic-dominance related criteria

A number of statistical tests for stochastic dominance have been developed -- see for example, Anderson (1996), Kaur, Prakasa Rao, and Singh (1994), Davidson and Duclos (2000), Barrett and Donald (2003), Linton, Maasoumi, and Whang (2003), and Davidson (2009). The main differences among these tests are the way the null hypothesis is formulated, the type of test statistic employed, the ability of the test to handle correlated samples, and the approach to computing p-values.

For the purpose of this paper, the most appealing test specification is the one of Davidson (2009). First, the test allows for correlated samples. This is an important limitation for most existing tests of stochastic dominance, which can deal only with uncorrelated samples. When comparing portfolios that consist of the same assets (but in different proportions), we have to consider correlated samples. The test of Davidson and Duclos (2000) can also handle correlated samples but can only be evaluated at a fairly low number of returns (about 20) and not over the complete return distribution.

Second, the Davidson (2009) test starts with the null hypothesis of non-dominance for one distribution over another, whereas the majority of other tests have as their null hypothesis dominance -- see, e.g., Anderson (1996), Davidson and Duclos (2000), plus Barrett and Donald (2003). Rejecting the null of dominance by the first distribution does not then imply dominance by the second distribution, since it can also happen that the test fails to rank those distributions. However, rejecting the null of non-dominance delivers an unambiguous result of dominance. This formulation of the null hypothesis is also used by Kaur, Prakasa Rao, and Singh (1994); however, their approach cannot cope with correlated samples. We thus rely on the Davidson (2009) test to establish the dominance relation between different portfolio return distributions in our out-of-sample tests.

As the true return generating process is not known, one cannot directly compute and compare the integrals from Equation (1). Rather, one has to use their sample counterparts. Following Davidson (2009), we label the sample counterparts of the integrals from Equation (1) as $D_K^S(z)$, where K denotes the two sample distributions (A or B) that are being compared, and S denotes the degree of stochastic dominance, with S=2 for SSD and S=3 for TSD. We will refer to $D_K^S(z)$ as a dominance function:

$$D_K^S(z) = \frac{1}{(S-1)N_K} \sum_{i=1}^{N_K} (\max(z - y_{i,K}, 0))^{S-1}, \tag{3}$$

where N_K is a number of observations in distribution sample K, $y_{i,K}$ is the i-th observation in this sample, and z is the threshold of interest.

The set of thresholds {z} includes all unique observations from both samples $\{y_{i,A}\}$ and $\{y_{i,B}\}$ lying in the joint support of those samples, where we trim the 5% highest and lowest observations.[5] For each level of z, the standardized difference between the two dominance functions is computed:

$$t(z) = \frac{D_B^S(z) - D_A^S(z)}{\left(V\hat{a}r(D_A^S(z)) + V\hat{a}r(D_B^S(z)) - 2C\hat{o}v(D_A^S(z), D_B^S(z))\right)^{1/2}} \quad . \tag{4}$$

The final test statistic is obtained as:

$$t^* = \min_z t(z) \tag{5}$$

[5] One needs to trim the set of thresholds in order to achieve higher power of the test. The test of the stochastic dominance is then restricted to the chosen interior interval of the joint support. We investigate smaller and larger levels of trimming in the robustness section.

which is asymptotically normally distributed. For small samples, Davidson (2009) describes a bootstrap algorithm to obtain the p-values.[6]

SSD requirements are rather strict and often it is not possible to discriminate between portfolios based on the SSD criterion – i.e. the portfolios lie in the same dominance class. A generally more powerful test of portfolio dominance uses the concept of the third-order stochastic dominance (TSD). TSD requires one additional assumption about the investor utility function – that its third derivative is positive, which is described as "prudence". When return distribution $F_A(y)$ dominates distribution $F_B(y)$ by TSD, it indicates that all risk-averse and prudent investors will prefer $F_A(y)$ to $F_B(y)$. When implementing the Davidson (2009) procedure, the TSD test uses the value S=3 in equation (3) and then proceeds with calculations leading to equation (5).

The results described below entail T-1 yearly periods of in-sample fitting for all portfolios of interest with corresponding out-of-sample performance comparisons based on the SSD and TSD criteria. There is no established way to aggregate T-1 values of the Davidson (2009) test statistics in order to obtain a unique measure of portfolio quality. We propose to use three relevant summary characteristics regarding out-of-sample performance: (1) the number of cases in which a given portfolio choice approach provides portfolios that dominate the benchmark out-of-sample (N^+), (2) the number of cases in which those portfolios belong to the same dominance class as the benchmark (N^0), and (3) the number of cases in which those portfolios are dominated by the benchmark (N^-). We will also use median p-values of the

---

[6] Applying dominance tests to time series data, one needs to be concerned about test performance if there is time dependence in the data, such as autocorrelation in returns or GARCH effects in volatility. Unfortunately, no test so far explicitly accounts for such time-series effects. Nolte (2008) shows that the Davidson SSD test loses power if the data are strongly serially correlated. As we will document below, serial correlation is not pronounced in the data used for the current study. Nolte also shows that the Davidson test performs well in the presence of GARCH effects. Thus, we feel comfortable using the Davidson approach.

null hypothesis that the competing portfolio does not SSD (TSD) the benchmark as an additional performance characteristic.

Another way to assess the out-of-sample performance is to consider portfolio returns over the complete set of (T-1) out-of-sample periods. With this approach, each year's out-of-sample performance was generated by a portfolio estimated using a different in-sample period but the same portfolio construction procedure (Kuosmanen, MinVar, etc.).

With 85 years of daily data, the number of observations is enormous. To make the testing more manageable, we first aggregate the returns to a monthly level and then test for their non-dominance over Bench based on SSD and TSD criteria as described above. We always check the reverse hypothesis for non-dominance of Bench over the alternative portfolios and report the corresponding t-statistics and p-values.

2.2.2. Other portfolio performance criteria

In order to access the performance of the competing portfolios along other dimensions, we compute several standard descriptive statistics of portfolio performance. In particular, we compute mean and median returns, return standard deviation, skewness, and kurtosis, as well as minimum and maximum returns over the sample period. We report the Sharpe ratio for each of the portfolios, using the 3-month T-bill rate as the risk-free rate. Two measures capture the left tail risk: the sample 5% value at risk, VaR(5%); and the sample expected shortfall, Shortfall(5%), which is measured as average return conditional on it being below VaR(5%). We also include two additional performance measures: certainty equivalent (CEV3) and turnover (Turnover).

CEV3 is defined as the inverse of the expected utility function, where we proxy for expected utility using the average of realized values. The utility function is constant relative risk aversion with a risk aversion parameter γ of 3:

$$CEV = u^{-1}\left(\frac{1}{T_r}\sum_{t=1}^{T_r} u(1+r_t)\right) \quad (8)$$

$$u(1+r_t) = \frac{(1+r_t)^{1-\gamma}}{1-\gamma} \quad (9)$$

where T_r in equation (8) is the total number of return observations in all out-of-sample periods.

Turnover serves as a proxy for exposure to transaction costs associated with portfolio rebalancing. It is computed as the average absolute change summed across all N portfolio weights:

$$Turnover = \frac{0.5}{T-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left|w_{it} - w_{it-1}\right|, \quad (10)$$

where w_{it} is the optimal portfolio weight of the asset i in year *t*. Our measure of turnover is based on DeMiguel, Garlappi, and Uppal (2009) where the main difference is that we scale the measure by 0.5. This insures that selling the complete portfolio and buying a new one results in a turnover of 100%.

The analysis is performed using the daily observations for the out-of-sample portfolios over all T-1 years. We also aggregate the daily returns within each year to obtain one-year return estimates and re-compute all the performance measures based on those T-1 yearly returns.

3. The data

We utilize daily return data for the Fama and French 49 industry portfolios.[7] The daily returns start as early as June 1926, and we choose the first observation from January 1927 as the starting day of our sample. The last observed return in December 2012 is the end date of the sample. Thus, we have 86

[7] The data were downloaded from the data library of Kenneth French at:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html years of data (22,731 daily return observations), which results in 85 annual out-of-sample periods for our main analysis.[8] The average daily returns range from 0.03% to 0.08%. The daily returns on all the industry portfolios exhibit excess kurtosis (with mean/median being 49.03/21.19) and are thus not normally distributed. This is a potential issue for Mean-Variance-Related approaches but not for the SSD-related alternatives, which consider the entire distribution and not just the first two moments. The serial correlation in the returns is not strongly pronounced with the mean (median) being 3.52% (4.55%). The maximum serial correlation of 17.52% is documented for the Healthcare industry, whereas the minimum of -18.46% corresponds to the Business Services industry. Thus, serial correlation should not introduce any problems in our out-of-sample SSD and TSD tests.

The benchmark portfolio in our basic runs represents the stock market, and we use daily returns on CRSP value-weighted all-share index as its proxy. It has a 0.04% mean daily return and a 1.07% daily standard deviation over the entire period.

4. Empirical results

In our main tests, we use one-year estimation and forecast windows. With 86 years of data and the first year used for the initial estimation, we obtain 85 non-overlapping estimates for out-of-sample portfolio performance.[9]

4.1. Out-of-sample portfolio performance according to stochastic dominance criteria

Results for the out-of-sample portfolio analysis using SSD and TSD are summarized in Table 1. We use "Win" to indicate that a given portfolio strategy dominated the benchmark out-of-sample at the

[8] The risk-free rate for this period is the sequence of yields on 3-month Treasury bills, which are also obtained from the Kenneth French data library.

[9] There is an implicit assumption here that target portfolio weights are adjusted annually and that portfolios are rebalanced on a daily frequency back to the fixed weights.

10% significance level. "Loss" indicates that a portfolio was dominated by the benchmark, and "Tie" indicates that both portfolios were in the same dominance class. Panel A uses the SSD criterion, whereas Panel B uses the TSD criterion. The last column of each panel reports the median p-values across 85 years for the null hypothesis that the portfolio in question does not dominate the benchmark.

Table 1. Out-of-Sample Performance of the Alternative Portfolios

This table reports the number out of 85 forecast windows from 1928 – 2012, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

Panel A: Out-of-Sample SSD

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
SSD-Related							
Kuosmanen	27	32	56	66	2	2	0.28
MinShortfall	27	32	58	68	0	0	0.30
MinVar	30	35	55	65	0	0	0.23
MinSemivar	29	34	56	66	0	0	0.30
Equal	8	9	70	82	7	8	0.93
Mean-Variance-Related							
MinVarBench	29	34	56	66	0	0	0.20
MaxSharpe	11	13	66	78	8	9	0.95
InformationRatio	0	0	80	94	5	6	0.97

Panel B: Out-of-Sample TSD

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
SSD-Related							
Kuosmanen	55	65	26	31	4	5	0.05
MinShortfall	64	75	21	25	0	0	0.03
MinVar	71	84	14	16	0	0	0.02
MinSemivar	63	74	21	25	1	1	0.02
Equal	14	16	52	61	19	22	0.91
Mean-Variance-Related							
MinVarBench	68	80	17	20	0	0	0.02
MaxSharpe	21	25	35	41	29	34	0.92
InformationRatio	3	4	57	67	25	29	0.92

We see in Table 1 Panel A that the SSD-related group of portfolios performs well. Kuosmanen and MinShortfall won against the benchmark in 27 periods out of 85, MinVar won in 30 periods, and MinSemivar won in 29 periods. Kuosmanen was dominated by Bench out-of-sample in only 2 periods. However, all these portfolios had ties in more than 65% of periods, where those portfolios ended up in the same dominance class as the benchmark. The median p-values are between 23% and 30%.

The equally-weighted portfolio does not perform well in the SSD sense. It wins against Bench in 8 periods out of 85 and loses in 7 periods. MaxSharpe also has a relatively weak performance: it dominates Bench in 11 periods and is dominated by Bench in 8 other periods. InformationRatio is the worst performer: it never dominates Bench and is dominated by Bench 5 times. The generally weaker performance of the mean-variance-related strategies appears to be due to unstable and extreme weights generated by the mean-variance optimization approach.

MinVarBench is an exception to the generally weak performance within the mean-variance-related group, performing at a level similar to the SSD-related portfolio approaches. It dominated Bench in 29 periods and experienced no losses. The reason is that MinVarBench is often located in the SSD efficient set and thus behaves in much the same way.

The TSD related results (Table 1 Panel B) reveal that the TSD criterion is substantially more powerful than SSD in ranking portfolios. For example, within the SSD-related group, ties now occur in only 16 - 31% of periods instead of the 65 – 68% when using the SSD criterion. Generally, the qualitative ranking of performance by alternative portfolio construction approaches remains similar to that observed using the SSD criterion. That is, the SSD-related portfolios plus MinVarBench perform much better than the other alternative portfolios.

The same conclusion is supported when we consider all 85 out-of-sample years of portfolio returns as a single set. As mentioned previously, we aggregate the massive number of daily returns to a monthly level. Table 2 reports the corresponding t-statistics and p-values for SSD and TSD non-dominance tests of Davidson (2009).

Table 2. Out-of-Sample Performance of the Alternative Portfolios over Complete Sample

This table reports t-statistics and corresponding p-values for the SSD and TSD tests of non-dominance by the alternative portfolios over Bench, and Bench over the alternative portfolios using monthly returns during 1928 – 2012. The alternative portfolios are based on Fama-French 49 industry portfolios. We use ***, **, and * to indicate significance at the 1%, 5%, and 10% levels, respectively.

	SSD				TSD			
	Alt. does not dominate Bench		Bench does not dominate Alt.		Alt. does not dominate Bench		Bench does not dominate Alt.	
SSD-Related								
Kuosmanen	1.58*	0.06	- 5.75	1.00	2.39***	0.01	- 4.60	1.00
MinShortfall	0.49	0.31	- 7.81	1.00	3.14***	0.00	- 6.07	1.00
MinVar	0.64	0.26	- 8.54	1.00	3.30***	0.00	- 6.61	1.00
MinSemivar	0.38	0.35	- 7.57	1.00	3.02***	0.00	- 5.90	1.00
Equal	- 4.44	1.00	- 1.86	0.97	- 4.36	1.00	2.45***	0.01
Mean-Variance-Related								
MinVarBench	1.26	0.10	- 8.44	1.00	3.22***	0.00	- 6.73	1.00
MaxSharpe	- 3.70	1.00	- 1.19	0.88	- 3.24	1.00	1.88**	0.03
InformationRatio	- 1.53	0.94	- 3.56	1.00	- 1.17	0.88	- 1.40	0.92

Consistent with the results from Table 1, Equal, MaxSharpe, and InformationRatio are rather weak performers. Equal and MaxSharpe are even dominated by Bench using the TSD criterion, and InformationRatio lies in the same dominance class. The SSD-related portfolios together with MinVarBench perform well, dominating Bench out-of-sample according to the TSD criterion, with p-values all being below 1%. Kuosmanen is the only portfolio that also significantly dominates Bench using the SSD criterion over the complete 1928 – 2012 period, with the corresponding p-value being 6%.

4.2. Out-of-sample portfolio performance according to other criteria

Table 3 reports descriptive statistics for the returns delivered by the alternative portfolio strategies, as well as certainty equivalent (CEV3) and turnover (Turnover) measures. Our discussion focuses on Panel A, which is based on annual returns. However, similar patterns can be seen in Panel B with daily returns.

Table 3. Descriptive Statistics of Portfolio Performance

This table reports descriptive statistics for the competing portfolio-choice strategies. Panel A is based on annual percentage returns. The statistics are computed using 85 yearly returns from 1928 to 2012. Panel B uses daily percentage returns across the entire time period.

	Mean	Median	STD	Min	Max	Skewness	Kurtosis	Sharpe ratio	VaR(5%)	Shortfall(5%)	CEV3	Turnover
Panel A: Annual Returns												
Bench	11.38	14.36	20.40	-44.14	56.89	-0.41	2.98	0.38	-28.18	-36.43	4.51	0.00
SSD-Related												
Kuosmanen	12.16	12.20	19.59	-41.50	88.67	0.14	5.32	0.44	-23.59	-33.27	6.26	0.65
MinShortfall	10.60	10.02	17.97	-40.92	68.81	0.03	3.94	0.39	-18.68	-28.74	5.75	0.50
MinVar	10.67	12.91	17.61	-42.19	70.59	-0.12	4.30	0.40	-20.01	-29.64	5.82	0.41
MinSemivar	10.52	11.67	17.76	-40.02	63.70	-0.03	3.68	0.39	-19.93	-29.07	5.77	0.53
Equal	13.58	14.97	22.88	-41.08	90.46	0.07	3.77	0.44	-28.54	-36.68	5.71	0.00
Mean-Variance-Related												
MinVarBench	11.64	12.56	18.41	-47.21	77.76	-0.10	4.86	0.44	-21.17	-32.40	6.14	0.55
MaxSharpe	17.15	16.88	27.02	-47.34	134.31	0.88	6.41	0.50	-26.93	-37.51	7.59	0.87
InformationRatio	11.90	15.16	20.04	-41.53	51.84	-0.41	2.89	0.42	-27.27	-34.93	5.40	0.29
Panel B: Daily Returns												
Bench	0.040	0.074	1.076	-17.135	15.505	-0.139	19.403	0.025	-1.554	-2.581	0.022	
SSD-Related												
Kuosmanen	0.041	0.071	0.879	-12.913	13.946	-0.276	28.494	0.032	-1.217	-2.116	0.030	
MinShortfall	0.036	0.051	0.754	-10.225	12.451	-0.093	26.476	0.030	-1.032	-1.784	0.027	
MinVar	0.036	0.056	0.729	-11.346	11.148	-0.289	30.547	0.031	-0.987	-1.722	0.028	
MinSemivar	0.036	0.047	0.760	- 8.774	12.756	-0.106	24.434	0.030	-1.046	-1.786	0.027	
Equal	0.046	0.089	1.073	-16.544	16.260	-0.235	20.491	0.031	-1.558	-2.588	0.029	
Mean-Variance-Related												
MinVarBench	0.039	0.061	0.796	-12.208	14.818	-0.317	32.359	0.033	-1.081	-1.883	0.030	
MaxSharpe	0.060	0.087	1.375	-25.615	23.857	0.125	34.150	0.034	-1.810	-3.257	0.031	
InformationRatio	0.042	0.075	1.092	-17.123	15.465	-0.108	19.206	0.026	-1.576	-2.622	0.024	

Compared to Bench, the Kuosmanen portfolio performs nicely along multiple dimensions. It improves the mean return somewhat (12.16% vs. 11.38%) and reduces the standard deviation (19.59% vs. 20.40%). Kuosmanen also managed to reduce the maximum loss (minimum annual return of -41.50%) while still allowing large gains (maximum return of 88.67%). This occurs because the Kuosmanen approach tries to structure a portfolio that will avoid large negative returns and tend to have positive skewness. Also, Kuosmanen had higher realized values for the Sharpe Ratio (0.44 vs. 0.38 for Bench) and CEV3 (6.26 vs. 4.51 for Bench).

The other SSD-related strategies reduce portfolio return standard deviation even further; however, this also lowers mean returns somewhat. Due to their basic character of risk avoidance, it is not surprising that MinVar, MinSemiVar, and MinShortfall perform somewhat better than Kuosmanen based on left tail risk measures. That is, they have smaller (in absolute value) VaR and expected Shortfall than Kuosmanen. At the same time, these approaches cannot achieve as high Sharpe ratios and CEV3 as Kuosmanen.

Compared with Bench, the equally-weighted portfolio in our sample delivered a higher mean and standard deviation as well as a higher Sharpe ratio and CEV3. The CEV3 of Equal is somewhat smaller than that of Kuosmanen but similar to the other SSD-related approaches.

Turning to the mean-variance related strategies, we can see MinVarBench performs rather well along multiple dimensions, although not quite to the same level as its performance in terms of SSD and TSD. It has similar mean return and lower standard deviation than Bench, while delivering a Sharpe ratio and CEV3 comparable to Kuosmanen. It has, however, a worse minimum return of -47.21% per year compared to -44.14% for Bench and -41.50% for Kuosmanen.

MaxSharpe delivered the highest out-of-sample Sharpe ratio (0.50) and CEV3 (7.59). This performance results from the highest mean (17.15%), but the associated costs are the highest STD (27.02%) as well as the worst minimum return (-47.34%) and shortfall (-37.51%). It also has the largest Turnover measure, indicating that the portfolio weights are not stable.

The InformationRatio performs slightly better than Bench on several dimensions in Panel A of Table 3. In Panel B, this pattern is sometimes reversed. For example, InformationRatio has a higher STD as well as more negative VaR and Shortfall than does Bench in Table B. This is consistent with the results from Table 1 where InformationRatio was often in the same dominance class as Bench (94% of periods using the SSD criterion in panel A).

As indicated in Table 3, the SSD-related approaches are not only performing well in terms of SSD and TSD but also in terms of more traditional performance measures such as STD, Minimum Return, CEV3, VaR(5%) and Shortfall(5%). It is instructive to compare performance of the SSD-related approaches with that of the equally-weighted portfolio, which does not have estimation error problems and is also not specifically designed for a mean-variance environment. The equally-weighted portfolio has a substantially higher STD, VaR(5%), and Shortfall(5%). MinVar, MinSemivar, and MinShortfall do best on these measures because their focus on risk minimization exhibits persistence out-of-sample. Kuosmanen is accepting of somewhat more risk in exchange for more upside, based on looking at the whole distribution.

5. Robustness

In this section, we discuss several robustness checks regarding methodological changes. Further, we investigate performance of the eight alternative portfolio construction approaches using monthly returns. Having many fewer data points in an annual period, leads us to lengthen the in-sample estimation period and to adjust the out-of-sample performance assessment procedure as discussed below.

5.1. Methodological robustness checks with daily data

We increased trimming levels for the z-interval to 10% and lowered them to 1%, used quarterly as well as two-year period lengths for estimation and forecast windows, lowered significance levels to 5% and even 1% for determining dominance, and used the equally-weighted (instead of value-weighted) CRSP all-share index as a benchmark. A brief summary is provided below, and the full results tables can be found in the online appendix. Results using both the SSD and TSD criteria are provided in our internet appendix tables; however, our discussion below focuses on the more powerful TSD results.

In Online Appendix Table A1, we see that lowering tail trimming from 5% to 1% substantially reduces the number of out-of-sample dominating portfolios for all strategies (except InformationRatio, which was already at 0). For example, the number of winning periods for Kuosmanen drops from 55 (with 5% trimming in Table 1) to 27 with 1% trimming. Similarly, MinVarBench now dominates Bench in 34 periods (compared to 68 with 5% trimming). Using 10% trimming increases the number of dominating periods across all strategies, see Online Appendix Table A2. For example, Kuosmanen increases to 58 wins out of 85 periods; and MinVarBench, now has 74 wins. In both Tables A1 and A2, the basic rankings of various portfolio choice approaches remain quite similar.

Using two-year estimation and forecast windows instead of one-year windows cuts the number of out-of-sample periods in half but has little effect on portfolio ranking, see Online Appendix Table A3. Here, all the SSD-related portfolios together with the MinVarBench are strong performers; dominating Bench between 32 to 35 two-year periods out of 42. With quarterly estimation and forecast windows[10], however, winning percentage declines dramatically for all strategies except InformationRatio, whose performance remains abysmal but improves to 5% Wins. See Online Appendix Table A4. Kuosmanen is particularly hard hit with its Win percentage plummeting to 3%. MinVar and MinVarBench hold up better than other approaches, but their winning percentages both drop by 33 and 32 points respectively. Even so, Kuosmanen outperforms all other SSD-related portfolios, MinVarBench, and Bench in terms of Sharpe ratio and CEV3, which in non-tabulated results are 0.45 and 6.65 respectively. Only the MaxSharpe strategy delivers a somewhat stronger performance along these dimensions (0.46 and 7.78).

Tightening the significance level for determining dominance reduces the number of wins for all strategies, but their ranking remains almost unchanged. See Online Appendix Table A5 for results using a 5% significance level and Online Appendix Table A6 for results with a 1% significance level. For

---

[10] As the number of observations declines to about 62 per period for this table, we computed bootstrapped p-values to determine stochastic dominance as described in Davidson (2009).

example, MinVar has the largest win percentage in both those tables; however, that percentage drops from 84% in Table 1 (10% significance level) to 66% in Table A5 and 34% in Table A6.

Normally, we use the value-weighted CRSP index as our market benchmark. In Online Appendix Table A7, we report results using the equally-weighted CRSP index. That index assigns disproportionally higher weights to small stocks which historically have often delivered higher returns. That makes it more difficult for any other portfolio to win against such a benchmark in the SSD or TSD sense. For example, Kuosmanen now wins only 28% of the time compared with a 65% winning percentage in Table 1. MinVar now wins 60% of the time (down from 84% in Table 1), while MinVarBench has its win percentage drop from 80% in Table 1 to 36% in Table A7.

5.2. Monthly returns

With daily returns, the data (49 industry portfolios) exhibited substantial excess kurtosis and non-normality. With monthly returns, those industry portfolios have closer to normal distributions but still exhibit substantial excess kurtosis. For the complete 86 year period, mean/median kurtosis of monthly returns for the 49 industry portfolios is 14.58/9.54 compared to 49.03/21.19 with daily returns. Using monthly returns, we also have many fewer data points in an annual estimation period (12 months versus roughly 260 daily returns). Clearly, 12 data points is not going to provide a decent estimate of the underlying return distribution. We tried a 5-year estimation period (60 monthly observations) but discovered that the Kuosmanen procedure does better with more data.[11] We ultimately opted to use 20

[11] The Kuosmanen procedure attempts to maximize the mean return among the set of portfolios that SSD dominate the benchmark in-sample. When the number of observations is relatively small, the estimated mean is not very reliable and the Kuosmanen procedure tends to find weights which do not perform well out-of-sample and are not very stable across estimation periods. In exploring this issue, we also conducted a simulation study in which the mean returns of all assets matched the average risk-free rate over the corresponding period. In this setting, Kuosmanen delivers similar SSD and TSD performances to those of other SSD-related portfolios that do not require estimation of a mean. We conclude that difficulty obtaining precise mean estimates in short samples and

years (240 monthly observations) as the in-sample estimation period when working with monthly data. With 240 observations, we have a similar number of in-sample data points as when we used one year of daily data (roughly 260 observations).

For our analysis with monthly return data, we used a rolling-window procedure. The first 20-year period was used to estimate the weights for Kuosmanen as well as the alternative portfolio construction approaches, and the performance of each approach was observed during the following year. We then moved forward one year, re-estimated the weights using 20 years of data (19 from the previous in-sample period plus one new year) and observed performance during the following year. This process continued through all the available data. This resulted in 66 years of out-of-sample performance measurements for each approach. We used the Davidson (2009) SSD and TSD tests for dominance using the full 66-year period. Results are in Table 4.

We see in Table 4 that using the SSD criterion most approaches are in the same dominance class with Bench. The one exception is MaxSharpe, which is dominated by Bench with a p-value of 0.02. Using the more powerful TSD test, Kuosmanen as well as the other SSD-based approaches and MinVarBench all dominate Bench with p-values at or below 0.02. MaxSharpe and Equal are both dominated by Bench using the TSD test. InformationRatio is in the same dominance class with Bench using both the SSD and TSD criteria.

---

corresponding over-fitting are relevant issues for Kuosmanen. To mitigate this problem, a sufficiently long time series of data is needed.

Table 4. Out-of-Sample Performance of the Alternative Portfolios with Monthly Returns

This table reports t-statistics and corresponding p-values for the SSD and TSD tests of non-dominance by the alternative portfolios over Bench, and Bench over the alternative portfolios using monthly returns during 1947 – 2012,. The alternative portfolios are based on monthly returns on the Fama-French 49 industry portfolios. We use 20-year rolling estimation windows with 1 year out-of-sample forecasts. We use ***, **, and * to indicate significance at the 1%, 5%, and 10% levels, respectively.

	SSD				TSD			
	Alt. does not dominate Bench		Bench does not dominate Alt.		Alt. does not dominate Bench		Bench does not dominate Alt.	
SSD-Related								
Kuosmanen	- 0.65	0.74	- 5.42	1.00	2.30***	0.01	- 4.91	1.00
MinShortfall	0.98	0.16	- 5.27	1.00	2.26***	0.01	- 4.57	1.00
MinVar	0.41	0.34	- 7.29	1.00	2.38***	0.01	- 5.83	1.00
MinSemivar	0.45	0.33	- 4.28	1.00	2.06**	0.02	- 3.90	1.00
Equal	- 2.79	1.00	- 0.79	0.78	- 2.68	1.00	1.79**	0.04
Mean-Variance-Related								
MinVarBench	0.21	0.42	- 6.87	1.00	2.15**	0.02	- 5.38	1.00
MaxSharpe	- 3.12	1.00	2.09**	0.02	- 3.10	1.00	1.76**	0.04
InformationRatio	- 0.53	0.70	- 1.09	0.86	- 0.67	0.75	- 0.37	0.64

Table 5 contains results with monthly data for the other performance measures over the 66 out-of-sample years. Kuosmanen and the other SSD-related approaches plus MinVarBench all did a good job of reducing STD relative to Bench. Not surprisingly, MinVar had the lowest STD and MinShortfall had the best shortfall performance out-of-sample. Interestingly, MinShortfall had the best CEV3 and Sharpe ratio performance, with MinVar being second best in those categories. MaxSharpe generally did not perform well out-of-sample, with a lower mean return and higher STD than Bench. In fact, MaxSharpe had the lowest Sharpe ratio out-of-sample. InformationRatio and Equal performed roughly on a par with Bench.

Table 5. Descriptive Statistics of Portfolio Performance, Monthly Returns

This table reports descriptive statistics for the competing portfolio-choice strategies. Panel A is based on annual percentage returns. The statistics are computed using 66 yearly returns from 1947 to 2012. Panel B uses monthly percentage returns across the entire time period.

	Mean	Median	STD	Min	Max	Skewness	Kurtosis	Sharpe ratio	VaR(5%)	Shortfall(5%)	CEV3	Turnover
Panel A: Annual Returns												
Bench	9.42	7.31	16.31	-38.21	50.18	-0.01	3.07	0.35	-13.51	-26.44	5.49	0.00
SSD-Related												
Kuosmanen	8.82	8.71	15.06	-28.81	55.51	0.36	3.77	0.34	-14.80	-22.53	5.71	0.28
MinShortfall	9.88	9.86	15.01	-31.01	55.51	0.30	3.49	0.42	-13.91	-18.96	6.80	0.14
MinVar	9.24	8.62	13.92	-30.62	44.95	0.08	3.28	0.40	-13.76	-20.94	6.49	0.09
MinSemivar	9.62	9.27	15.35	-38.32	50.77	0.06	3.32	0.39	-14.23	-20.75	6.20	0.20
Equal	10.22	7.61	17.17	-39.05	53.53	0.18	3.40	0.38	-15.19	-26.81	6.00	0.00
Mean-Variance-Related												
MinVarBench	9.07	8.58	14.33	-30.10	59.53	0.33	4.17	0.38	-14.76	-21.23	6.23	0.16
MaxSharpe	8.47	4.18	17.57	-30.41	67.72	0.64	3.85	0.28	-19.12	-23.30	4.47	0.28
InformationRatio	9.58	8.07	16.23	-38.22	52.40	0.09	3.31	0.37	-14.44	-25.63	5.75	0.10
Panel B: Monthly Returns												
Bench	0.94	1.30	4.31	-22.54	16.56	-0.52	4.90	0.14	-6.36	-9.31	0.66	
SSD-Related												
Kuosmanen	0.89	1.13	3.86	-21.80	20.45	-0.25	6.41	0.14	-5.21	-8.34	0.66	
MinShortfall	0.98	1.17	3.65	-16.01	15.79	-0.35	4.39	0.17	-5.30	-7.62	0.78	
MinVar	0.93	1.03	3.50	-16.55	14.56	-0.26	4.89	0.16	-4.86	-7.31	0.74	
MinSemivar	0.96	1.08	3.77	-18.60	15.62	-0.36	4.90	0.16	-5.12	-7.77	0.74	
Equal	1.03	1.28	4.64	-25.95	18.11	-0.50	5.69	0.14	-6.69	-10.05	0.69	
Mean-Variance-Related												
MinVarBench	0.91	1.10	3.56	-16.29	16.52	-0.27	4.93	0.15	-5.22	-7.55	0.72	
MaxSharpe	0.85	0.94	4.69	-23.77	23.74	-0.20	5.30	0.10	-7.13	-10.04	0.52	
InformationRatio	0.96	1.19	4.32	-22.86	16.30	-0.51	5.00	0.14	-6.57	-9.34	0.67	

6. Concluding comments

There is much to be said in favor of using SSD to measure portfolio performance and also to choose portfolios. This is particularly relevant for delegated fund management with heterogeneous investors. All risk-averse investors would prefer such a portfolio to a dominated benchmark. An important question is how to choose among SSD efficient portfolios. We compare the method of Kuosmanen (2004) with other SSD-related choices such as the minimum-variance portfolio. Kuosmanen (2004) finds the portfolio with the highest in-sample mean which dominates a benchmark in the SSD sense, if that is possible.

The SSD-related portfolio strategies (Kuosmanen, MinShortfall, MinVar, and MinSemivar) perform well out-of-sample in terms of not only dominating the benchmark (using SSD and TSD criteria) but also in terms of traditional performance measures with a risk-reduction orientation such as volatility, minimum returns, Var(5%), and Shortfall(5%). When calibration is based on daily data, Kuosmanen also does quite well in terms of Sharpe ratio and CEV3. With monthly data, MinShortfall and MinVar perform very well in terms of Sharpe ratio and CEV3. A strategy that minimized variance while being constrained to matching the mean benchmark return (calibrated in-sample) performs roughly on a par with the SSD-related strategies.

The benchmark in our main analysis is the value-weighted market portfolio. It is frequently dominated out-of-sample by the SSD-related portfolio strategies. Moreover, those strategies also frequently out-perform that benchmark on a variety of other dimensions. One interesting measure is CEV3. In Table 3 using our main runs, the SSD-related approaches have CEV3 results which improve over that of the benchmark by a range of 28% to 39%.

The other portfolio strategies we tested (MaxSharpe, Information Ratio, and Equal) typically had a poor out-of-sample performance on dominance tests. Calibrated with daily return data, MaxSharpe was the best performer out-of-sample in terms of Sharpe Ratio and CEV3; however calibrated with monthly return data, it was the worst performer on those dimensions. InformationRatio tended to perform on traditional measures in a manner similar to the benchmark. The equal-weighted portfolio typically performed better than InformationRatio on traditional measures but frequently did less well than the SSD-related strategies.

The SSD-related procedures perform well because their emphasis on risk-reduction tends to generate weights that continue to do a good job of risk control out-of-sample. In a sense, there is performance persistence regarding risk reduction with these portfolio approaches. MinVarBench seems to also perform rather well for essentially the same basic reason. That portfolio minimizes variance (in-sample) while matching the benchmark's mean return. As long as that benchmark mean return is not very extreme, MinVarBench will also have a substantial emphasis on risk reduction and behave much like the SSD-related approaches.

References

Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64, 1183-1193.

Barrett, G.F., Donald, S.G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71-104.

Clark, E., Jokung, O., and Kassimatis K., 2011. Making inefficient market indices efficient. European Journal of Operational Research, 209, 83-93.

Davidson, R., 2009. Testing for restricted stochastic dominance: Some further results. Working paper, McGill University.

Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68, 1435-1464.

De Giorgi, E., 2005. Reward-risk portfolio selection and stochastic dominance. Journal of Banking and Finance 29, 895-926

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22, 1915-1953.

Fishburn, P.C., 1977. Mean-risk analysis with risk associated with below-target returns. The American Economic Review 67, 116-126.

Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance 58, 1651-1683.

Jorion, P., 1992. Portfolio optimization in practice. Financial Analysts Journal 48, 68-74.

Kan, R., Zhou, G., 2007. Optimal portfolio choice with parameter uncertainty. Journal of Financial & Quantitative Analysis 42, 621-656.

Kaur, A., Prakasa Rao, B.L.S., Singh, H., 1994. Testing for second-order stochastic dominance of two distributions. Econometric Theory 10, 849-866.

Kopa, M., 2009. An efficient LP test for SSD portfolio efficiency. Working paper, Charles University in Prague.

Kuosmanen, T., 2004. Efficient diversification according to stochastic dominance criteria. Management Science 50, 1390-1406.

Linton, O.B., Maasoumi, E., Whang, Y.-J., 2003. Consistent testing for stochastic dominance under general sampling schemes. Review of Economic Studies 72, 735-765.

Martellini, L., Ziemann, V., 2010. Improved estimates of higher-order comoments and implications for portfolio selection. Review of Financial Studies 23, 1467-1502.

Meyer, T.O., Li, X., Rose, L.C., 2005. Comparing mean variance tests with stochastic dominance when assessing international portfolio diversification benefits. Financial Services Review 14, 149–168.

Michaud, R.O., 1989. The Markowitz optimization enigma: Is 'optimized' optimal? Financial Analysts Journal 45, 31-42.

Nolte, I., 2008. Stochastic dominance tests under test. Working paper, University of Konstanz.

Ogryczak, W., Ruszczynski, A., 1999. From stochastic dominance to mean-risk models: Semideviations as risk measures. European Journal of Operational Research 116, 33-50.

Porter, R.B., 1974. Semivariance and stochastic dominance: A comparison. The American Economic Review 64, 200-204.

Post, T., 2003. Empirical tests for stochastic dominance efficiency. Journal of Finance 58, 1905-1931.

Post, T., 2008. On the dual test for SSD efficiency with an application to momentum investment strategies. European Journal of Operational Research 185, 1564-1573.

Post, T., Kopa, M., 2013. General liner formulations of stochastic dominance criteria, European Journal of Operational Research 230, 321-332.

Post, T., Levy, H., 2005. Does risk seeking drive stock prices? A stochastic dominance analysis of aggregate investor preferences and beliefs. Review of Financial Studies 18, 925-953.

Russell, W.R., Seo, T.K., 1980. Efficient portfolios by stochastic dominance. Southern Economic Journal 46, 877-882.

Scaillet, O., Topaloglou, N.L., 2010. Testing for stochastic dominance efficiency. Journal of Business and Economic Statistics, 28, 169-180.

Shalit, H., Yitzhaki, S., 1994. Marginal conditional stochastic dominance. Management Science, 40, 670-684.

Online Appendix for

Improved Portfolio Choice Using Second-Order Stochastic Dominance

by

James E. Hodder

Jens Carsten Jackwerth

Olga Kolokolova

In this appendix, we present the complete collection of tables with results discussed in the robustness section 5.1 of the main paper.

Table A1. Out-of-Sample Performance of the Alternative Portfolios, 1% Tail Trimming

This table reports the number out of 85 forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark. We trim 1% of highest and 1% of lowest observations when conducting the out-of-sample SSD and TSD tests.

Panel A: Out-of-Sample SSD

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
SSD-Related							
Kuosmanen	12	14	73	86	0	0	0.611
MinShortfall	10	12	75	88	0	0	0.556
MinVar	11	13	74	87	0	0	0.516
MinSemivar	9	11	76	89	0	0	0.514
Equal	2	2	79	93	4	5	0.965
Mean-Variance-Related							
MinVarBench	14	16	71	84	0	0	0.501
MaxSharpe	5	6	79	93	1	1	0.970
InformationRatio	0	0	84	99	1	1	0.975

Panel B: Out-of-Sample TSD

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
SSD-Related							
Kuosmanen	27	32	57	67	1	1	0.132
MinShortfall	29	34	56	66	0	0	0.130
MinVar	37	44	48	56	0	0	0.110
MinSemivar	33	39	52	61	0	0	0.123
Equal	5	6	74	87	6	7	0.919
Mean-Variance-Related							
MinVarBench	34	40	51	60	0	0	0.115
MaxSharpe	11	13	59	69	15	18	0.922
InformationRatio	1	1	78	92	6	7	0.930

Table A2. Out-of-Sample Performance of the Alternative Portfolios, 10% Tail Trimming

This table reports the number out of 85 forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark. We trim 10% of highest and 10% of lowest observations when conducting the out-of-sample SSD and TSD tests.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	42	49	40	47	3	4	0.103
MinShortfall	42	49	43	51	0	0	0.103
MinVar	50	59	35	41	0	0	0.057
MinSemivar	42	49	43	51	0	0	0.112
Equal	12	14	64	75	9	11	0.887
Mean-Variance-Related							
MinVarBench	53	62	32	38	0	0	0.052
MaxSharpe	17	20	52	61	16	19	0.913
InformationRatio	2	2	72	85	11	13	0.967
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	58	68	21	25	6	7	0.014
MinShortfall	73	86	12	14	0	0	0.007
MinVar	78	92	7	8	0	0	0.003
MinSemivar	69	81	15	18	1	1	0.008
Equal	18	21	44	52	23	27	0.888
Mean-Variance-Related							
MinVarBench	74	87	11	13	0	0	0.006
MaxSharpe	22	26	30	35	33	39	0.922
InformationRatio	6	7	48	56	31	36	0.923

Table A3. Out-of-Sample Performance of the Alternative Portfolios, Two-Year Periods

This table reports the number out of 42 two-year forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	17	40	26	60	0	0	0.198
MinShortfall	18	42	25	58	0	0	0.171
MinVar	20	47	23	53	0	0	0.131
MinSemivar	17	40	26	60	0	0	0.180
Equal	4	9	34	79	5	12	0.966
Mean-Variance-Related							
MinVarBench	20	47	23	53	0	0	0.112
MaxSharpe	6	14	27	63	10	23	1.000
InformationRatio	2	5	32	74	9	21	0.930
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	34	79	8	19	1	2	0.007
MinShortfall	34	79	7	16	2	5	0.006
MinVar	36	84	6	14	1	2	0.006
MinSemivar	32	74	9	21	2	5	0.009
Equal	11	26	21	49	11	26	0.956
Mean-Variance-Related							
MinVarBench	35	81	8	19	0	0	0.008
MaxSharpe	8	19	13	30	22	51	0.996
InformationRatio	11	26	16	37	16	37	0.892

Table A4. Out-of-Sample Performance of the Alternative Portfolios, Quarterly Periods

This table reports the number out of 343 quarterly forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	9	3	288	84	46	13	0.971
MinShortfall	63	18	274	80	6	2	0.482
MinVar	75	22	267	78	1	0	0.349
MinSemivar	64	19	276	80	3	1	0.472
Equal	30	9	288	84	25	7	0.890
Mean-Variance-Related							
MinVarBench	77	22	265	77	1	0	0.350
MaxSharpe	32	9	282	82	29	8	0.864
InformationRatio	13	4	310	90	20	6	0.929
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	10	3	260	76	73	21	0.951
MinShortfall	135	39	208	61	0	0	0.130
MinVar	174	51	168	49	1	0	0.099
MinSemivar	123	36	217	63	3	1	0.144
Equal	41	12	263	77	39	11	0.840
Mean-Variance-Related							
MinVarBench	163	48	179	52	1	0	0.109
MaxSharpe	42	12	253	74	48	14	0.825
InformationRatio	17	5	306	89	20	6	0.867

Table A5. Out-of-Sample Performance of the Alternative Portfolios, 5% Significance Level

This table reports the number out of 85 forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). We use a 5% significance level to define dominance. The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	**20**	**24**	**64**	**75**	**1**	**1**	**0.282**
MinShortfall	**20**	**24**	**65**	**76**	**0**	**0**	**0.304**
MinVar	**25**	**29**	**60**	**71**	**0**	**0**	**0.231**
MinSemivar	**21**	**25**	**64**	**75**	**0**	**0**	**0.295**
Equal	**7**	**8**	**72**	**85**	**6**	**7**	**0.934**
Mean-Variance-Related							
MinVarBench	**25**	**29**	**60**	**71**	**0**	**0**	**0.199**
MaxSharpe	**7**	**8**	**75**	**88**	**3**	**4**	**0.952**
InformationRatio	**0**	**0**	**84**	**99**	**1**	**1**	**0.967**
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	**42**	**49**	**40**	**47**	**3**	**4**	**0.054**
MinShortfall	**51**	**60**	**34**	**40**	**0**	**0**	**0.028**
MinVar	**56**	**66**	**29**	**34**	**0**	**0**	**0.016**
MinSemivar	**47**	**55**	**38**	**45**	**0**	**0**	**0.023**
Equal	**11**	**13**	**59**	**69**	**15**	**18**	**0.908**
Mean-Variance-Related							
MinVarBench	**55**	**65**	**30**	**35**	**0**	**0**	**0.018**
MaxSharpe	**19**	**22**	**41**	**48**	**25**	**29**	**0.922**
InformationRatio	**3**	**4**	**64**	**75**	**18**	**21**	**0.923**

Table A6. Out-of-Sample Performance of the Alternative Portfolios, 1% Significance Level

This table reports the number out of 85 forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). We use a 1% significance level to define dominance. The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	12	14	72	85	1	1	0.282
MinShortfall	12	14	73	86	0	0	0.304
MinVar	13	15	72	85	0	0	0.231
MinSemivar	6	7	79	93	0	0	0.295
Equal	4	5	76	89	5	6	0.934
Mean-Variance-Related							
MinVarBench	12	14	73	86	0	0	0.199
MaxSharpe	5	6	79	93	1	1	0.952
InformationRatio	0	0	84	99	1	1	0.967
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	20	24	64	75	1	1	0.054
MinShortfall	29	34	56	66	0	0	0.028
MinVar	29	34	56	66	0	0	0.016
MinSemivar	26	31	59	69	0	0	0.023
Equal	3	4	78	92	4	5	0.908
Mean-Variance-Related							
MinVarBench	25	29	60	71	0	0	0.018
MaxSharpe	8	9	69	81	8	9	0.922
InformationRatio	0	0	82	96	3	4	0.923

Table A7. Out-of-Sample Performance of the Alternative Portfolios, Equally-Weighted CRSP Index

This table reports the number out of 85 forecast windows, where the considered portfolios dominate the benchmark (Win), are dominated by the benchmark (Loss), or lie in the same dominance class (Tie). The alternative portfolios are based on Fama-French 49 industry portfolios. Panel A is based on the SSD criterion whereas Panel B is based on the TSD criterion. The last column of each panel reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark. The benchmark is the returns on the equally-weighted CRSP all-share index.

	Win		Tie		Loss		Median p-Value
	#	%	#	%	#	%	
Panel A: Out-of-Sample SSD							
SSD-Related							
Kuosmanen	9	11	65	76	11	13	0.940
MinShortfall	15	18	66	78	4	5	0.896
MinVar	17	20	66	78	2	2	0.862
MinSemivar	15	18	68	80	2	2	0.902
Equal	6	7	52	61	27	32	1.000
Mean-Variance-Related							
MinVarBench	14	16	57	67	14	16	0.925
MaxSharpe	6	7	58	68	21	25	1.000
InformationRatio	0	0	59	69	26	31	1.000
Panel B: Out-of-Sample TSD							
SSD-Related							
Kuosmanen	24	28	48	56	13	15	0.488
MinShortfall	48	56	33	39	4	5	0.062
MinVar	51	60	31	36	3	4	0.052
MinSemivar	43	51	37	44	5	6	0.089
Equal	19	22	33	39	33	39	0.998
Mean-Variance-Related							
MinVarBench	31	36	41	48	13	15	0.440
MaxSharpe	18	21	30	35	37	44	0.996
InformationRatio	0	0	31	36	54	64	1.000