Improved Portfolio Choice Using Second-Order Stochastic Dominance

by

James E. Hodder

Jens Carsten Jackwerth

Olga Kolokolova

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 particular choices based on Kuosmanen (2004) plus Kopa and Post (2011), comparing their performance to other SSD-related strategies and to standard portfolio choice approaches. These SSD-related choices outperform portfolios chosen based on their Sharpe ratio, information ratio, or using equal weights. 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,* Securities Exchange Commission, jhodder@bus.wisc.edu

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

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

> ***** *The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues on the staff of the Commission.*

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 out-perform competing portfolio choice methodologies out-of-sample.

The concept of SSD is theoretically appealing: 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), Kopa and Post (2011) 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. Additionally, the test procedures of Kuosmanen (2004) as well as Kopa and Post (2011) identify an efficient portfolio that dominates the evaluated portfolio if the latter is not efficient itself.

The main limitation of all these works is that they only analyze in-sample performance. For practical portfolio allocation problems, it is important 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 or did not rely on the SSD criterion for estimating portfolio weights themselves when considering out-of-sample performance.

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 algorithms developed in Kuosmanen (2004) as well as Kopa and Post (2011) to find optimal portfolio weights. We next test whether these SSD-based portfolios dominate 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 allocation procedures going forward.

We also compare the performance of our SSD-based portfolios 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 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 value-weighted 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 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 eleven approaches mentioned above. The light grey area indicates the set of portfolios that have positive test 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. For this illustrative example, we use annual returns on five Fama-French industry portfolios over the complete sample period from 1927 to 2012.[1] In this particular example of *in-sample* analysis, MinVar, MaxSharpe, Kuosmanen, KP2011Power3, and KP2011Min are located on the mean-variance efficient frontier and dominate

[1] The five Fama-French industry portfolios were obtained from the Kenneth French Data Library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). They include Consumer products, Manufacturing, Hi-Tech, Health, and Other industries. With this limited number of assets, we can exhaustively test all possible portfolios (with weight increments of 0.05) using the Davidson (2009) non-dominance test to find the set that have positive and significant (at the 10% level) test statistics.

Bench.[2] Equal and InformationRatio also dominate Bench, but they are located well inside the mean-variance efficient frontier. Other portfolios lie in the same dominance class with Bench. Note the position of the MinVarBench portfolio. Bench happens to have a rather low mean return and plots below the mean-variance frontier obtained using just the five industry portfolios as assets. In other words, the location of Bench is not directly attainable using those five industry portfolios. To match the benchmark return and be on the mean-variance frontier, MinVarBench winds up situated to the right of Bench and just on the lower part of the efficient frontier.

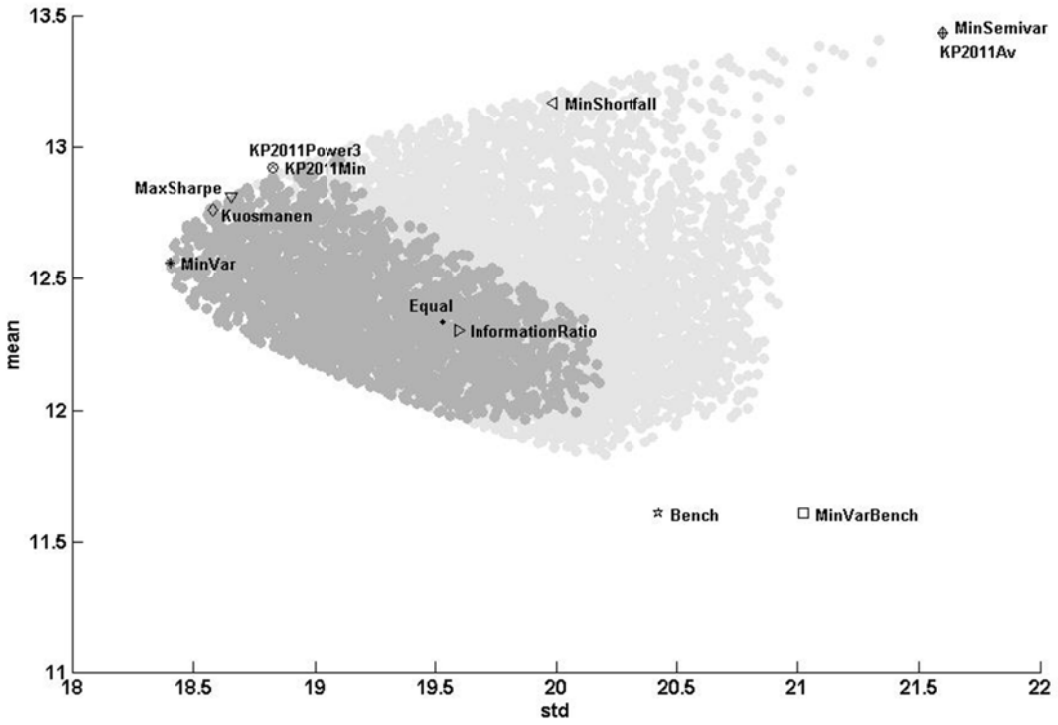
To preview our main results, we find that SSD-related approaches typically perform well out-of-sample using the SSD criterion. In other words, the typical market portfolio as proxied by the benchmark is often not at all well-structured from a SSD perspective. Prior research indicated the market portfolio 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 par with the benchmark. The portfolios based on the information ratio and the Sharpe ratio also tend to perform poorly. Out-of-sample, SSD-based portfolios generally perform well not only in terms of SSD but also on several more traditional performance measures including Sharpe ratio, Sortino ratio, certainty equivalent value, and turnover.

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

[2] Here, Kuosmanen is the optimal portfolio obtained by the Kuosmanen (2004) algorithm; KP2011Power3, KP2011Min, and KP2011Av are optimal portfolios obtained using the Kopa and Post (2011) algorithm with different weighting schemes, as will be detailed below.

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. It is based on annual returns from 1927 to 2012 for five Fama-French industry portfolios. The benchmark is the value weighted CRSP all-share index. The light grey area covers all the portfolios that have a positive test statistic with respect to Bench. The 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 $t_0 - \Delta t$ to t_0. Our main analysis uses a yearly time period with daily returns. For the same (in-sample) time period, we construct several alternative SSD-based portfolios using the Kuosmanen (2004) linear programming (LP) approach (labeled Kuosmanen) as well as the Kopa and Post (2011) LP approach

(labeled KP2011), detailed below. In Kuosmanen (2004), the procedure is based on the necessary condition for SSD efficiency; and it measures the degree of inefficiency for the benchmark portfolio in terms of SSD. 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. The LP of Kopa and Post (2011) is based on both necessary and sufficient conditions for SSD efficiency. This procedure takes as an additional input a vector of weights, which characterize the importance of different partial cumulative returns for an investor. This allows different optimal SSD efficient portfolios that dominate the benchmark. Following Kopa and Post (2011), we use three specifications of the weight vector: (i) focusing on risk reduction (improving minimum return), (ii) focusing on increases in mean return, and (iii) focusing on improvement relative to some utility function, where we choose a power utility function with a risk aversion parameter of 3. We label these portfolios KP2011Min, KP2011Av, and KP2011Power3 respectively.

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.[3] 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).

Another 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

[3] 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.

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 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 below.

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.

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

In constructing our SSD-based portfolio, we adopt LP algorithms developed in Kuosmanen (2004) plus Kopa and Post (2011). These procedures allow for testing if a benchmark portfolio return distribution is SSD efficient relative to a given asset span. If the benchmark is not efficient, the solution also delivers a vector of portfolio weights corresponding to a well-diversified SSD-efficient portfolio that second-order stochastically dominates the benchmark. That portfolio is optimal given the structure for choosing among the SSD-efficient portfolios specified by Kuosmanen (2004) or by Kopa and Post (2011), respectively. We use these in-sample optimal portfolios as SSD-based competing portfolios.

We considered several other alternative approaches for constructing SSD-based portfolios, but they seemed to be inferior to Kuosmanen (2004) as well as Kopa and Post (2011). 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. 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. 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 approaches of Kuosmanen (2004) plus Kopa and Post (2011).

Below we formally present the LP formulations of Kuosmanen (2004) as well as Kopa and Post (2011), with notation changed to conform to our paper and constraints specified to match our requirement of non-negative weights.[4] In both specifications, the asset span consists of N assets with T daily return observations y_{it} each. The vector of portfolio weights to be optimized is denoted by λ.

We implement the following LP procedure of Kuosmanen (2004):

$$
\begin{aligned}
&\theta_2(y_{Bench}) = \max_{\lambda, W}\left(\sum_{t=1}^{T}\sum_{i=1}^{N} y_{i,t}\lambda_i - \sum_{t=1}^{T} y_{Bench,t}\right)\frac{1}{T} \\
&s.t.\ \sum_{i=1}^{N} y_{i,t}\lambda_i \geq \sum_{j=1}^{T} W_{t,j}\, y_{Bench,t} \ \text{for}\ t = 1,...,T \\
&\quad W \in \left\{[W_{i,j}]_{T\times T} \mid 0 \leq W_{i,j} \leq 1; \sum_{i=1}^{T} W_{i,j} = \sum_{j=1}^{T} W_{i,j} = 1\ \forall i, j = 1,...,T\right\} \qquad (2) \\
&\quad 0 \leq \lambda_i \leq 1, \forall i = 1,...,N \\
&\quad \sum_{i=1}^{N} \lambda_i = 1
\end{aligned}
$$

Where we use $y_{Bench,t}$ to denote the benchmark portfolio return at time t, 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 Kuosmanen.[5] Note that the procedure of Kuosmanen (2004) does not require the benchmark returns to be attainable in the asset span of the given N assets.

We use the following dual formulation of the LP from Kopa and Post (2011):

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

[5] 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.

$$
\begin{aligned}
&\xi(\tau \mid \mathrm{w}) = \max_{d_s, v_{t,s}, \lambda} \sum_{s=1}^{T} w_s d_s \\
&s.t. -T^{-1} \sum_{i=1}^{N+1} y_{i,t} \lambda_i + \frac{1}{s} d_s - v_{t,s} + \frac{1}{s} \sum_{k=1}^{T} v_{k,s} \le -T^{-1} \frac{1}{s} \sum_{k=1}^{s} \sum_{i=1}^{N+1} y_{i,s} \tau_i \text{ for } t, s = 1,...,T \\
&\quad 0 \le \lambda_i \le 1, \forall i = 1,...,N \\
&\quad \sum_{i=1}^{N} \lambda_i = 1 \\
&\quad \lambda_{\mathrm{N+1}} = 0 \\
&\quad v_{t,s} \ge 0, \text{ for } t, s = 1,...,T \\
&\quad d_s \ge 0, \text{ for } s = 1,...,T
\end{aligned}
\tag{3}
$$

Compared to Kuosmanen (2004), the procedure of Kopa and Post (2011) technically implies that the benchmark portfolio is attainable using the given asset span. It takes as an input a vector of benchmark portfolio weights τ. In our current setting, the benchmark portfolio (CRSP all-share index) may not be directly attainable within our asset span (Fama-French 49 industry portfolios). To address this situation, we include the benchmark return ($y_{Bench,t}$) as an additional asset (N+1) in the LP procedure. The vector of the weights τ now has the length of (N+1), and we restrict the weight on the benchmark asset $\lambda_{\mathrm{N+1}}$ to be zero in the optimal portfolio. Finally, the matrix of returns on all assets is sorted such that the benchmark returns are in ascending order ($y_{Bench,s} \le y_{Bench,s+1}$).

In Equation (3), w = $\{w_s\}$, s = 1,..,T is a vector of the weights assigned to each cumulative return while constructing the optimal portfolio.

We consider three choices of the w vector, such that the optimal portfolio is:

(i) focusing on risk reduction; denoted KP2011Min

$$
w_s = \begin{cases} 0.001, \text{ for } s = 2,...,T \\ 1 - \sum_{i=2}^{T} w_i, \text{ for } s = 1 \end{cases}
\tag{4}
$$

(ii) focusing on an increase in average return; denoted KP2011Av

$$w_s = \begin{cases} 0.001, \text{for } s = 1,...,T-1 \\ 1 - \sum_{i=1}^{T-1} w_i, \text{for } s = T \end{cases} \tag{5}$$

(iii) focusing on improvement relative to the power utility function with a risk aversion parameter of 3; denoted KP2011Power3

$$w_s = \begin{cases} \left(u'(1+y_{Bench,s}) - u'(1+y_{Bench,s+1})\right) / u'(1+y_{Bench,1}), \text{for } s = 1,...,T-1 \\ u'(1+y_{Bench,s}) / u'(1+y_{Bench,1}), \text{for } s = T \end{cases} \tag{6}$$

where $u(1+y_t) = \dfrac{(1+y_t)^{1-\gamma}}{1-\gamma}$ and $u'(1+y_t) = (1+y_t)^{-\gamma}$.

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 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. We will refer to $D_K^S(z)$ as a dominance function:

$$D_K^S(z) = \frac{1}{(S-1)N_K} \sum_{i=1}^{N_K} (\max(z - y_{i,K}, 0))^{S-1}, \tag{7}$$

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.[6] For each level of z, the standardized difference between the two dominance functions is computed:

$$t(z) = \frac{D_B^S(z) - D_A^S(z)}{\left(\hat{Var}(D_A^S(z)) + \hat{Var}(D_B^S(z)) - 2\hat{Cov}(D_A^S(z), D_B^S(z))\right)^{1/2}} \quad . \tag{8}$$

The final test statistic is obtained as:

$$t^* = \min_z t(z) \tag{9}$$

which is asymptotically normally distributed. For small samples, Davidson (2009) describes a bootstrap algorithm to obtain the p-values.[7]

[6] One needs to trim the set of thresholds in order to achieve higher power of the test. The test of 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.

[7] 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

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 criterion. 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. To provide a sense for the distribution of those T-1 test statistics, we report quartile p-values from that distribution, with the underlying tests using the null hypothesis that the competing portfolio does not SSD the benchmark.[8]

2.2.2. Other portfolio performance criteria

In order to assess 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 several additional performance measures: certainty equivalents with various risk-aversion parameters (CEV1, CEV3, and CEV5), the Sortino ratio, and turnover (Turnover).

---

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.

[8] We also report in an Online Appendix compute three additional summary characteristics regarding out-of-sample stochastic dominance: (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^-).

CEV1, CEV3, and CEV5 are 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 1, 3, and 5 respectively:

$$CEV = u^{-1}\left(\frac{1}{T_r}\sum_{t=1}^{T_r} u(1+r_t)\right) \tag{10}$$

$$u(1+r_t) = \frac{(1+r_t)^{1-\gamma}}{1-\gamma} \tag{11}$$

where T_r in equation (10) is the total number of return observations in all out-of-sample periods.

The Sortino ratio is the portfolio mean excess return above the risk-free rate divided by the left semi-deviation of the returns.

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:

$$Turnover = \frac{0.5}{T-1}\sum_{t=2}^{T}\sum_{i=1}^{N}\left|w_{it} - w_{it-1}\right|, \tag{12}$$

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.[9] 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 years of data (22,731 daily return observations), which results in 85 annual out-of-sample periods for our main analysis.[10] The average daily return across industry portfolios ranges 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) across industries 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 tests.

The benchmark portfolio represents the stock market, and we use daily returns on the 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.

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

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[10] 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.

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.[11]

4.1. Out-of-sample portfolios: descriptive statistics of the returns

Table 1 reports descriptive statistics for the returns delivered by the alternative portfolio strategies as well as Bench.

Table 1. Descriptive Statistics of Portfolio Performance

This table reports descriptive statistics for the portfolios chosen by competing strategies based on annual percentage returns. The statistics are computed using 85 yearly returns from 1928 thru 2012.

	Mean	Median	STD	Min	Max	Skewness	Kurtosis
Bench	11.38	14.36	20.40	- 44.14	56.89	- 0.41	2.98
SSD-Related							
Kuosmanen	12.16	12.20	19.59	- 41.50	88.67	0.14	5.32
KP2011Av	14.77	16.07	22.05	- 37.60	89.70	0.25	4.03
KP2011Min	14.93	16.75	22.34	- 36.91	92.21	0.31	4.13
KP2011Power3	14.96	17.52	22.32	- 37.59	90.42	0.25	4.04
MinShortfall	10.60	10.02	17.97	- 40.92	68.81	0.03	3.94
MinVar	10.67	12.91	17.61	- 42.19	70.59	- 0.12	4.30
MinSemivar	10.52	11.67	17.76	- 40.02	63.70	- 0.03	3.68
Equal	13.58	14.97	22.88	- 41.08	90.46	0.07	3.77
Mean-Variance-Related							
MinVarBench	11.64	12.56	18.41	- 47.21	77.76	- 0.10	4.86
MaxSharpe	17.15	16.88	27.02	- 47.34	134.31	0.88	6.41
InformationRatio	11.90	15.16	20.04	- 41.53	51.84	- 0.41	2.89

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 also allowing large gains (maximum return of 88.67%). This occurs because the Kuosmanen

[11] 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.

approach tries to structure a portfolio that will avoid large negative returns and tend to have positive skewness. MinVar, MinSemivar, and MinShortfall reduce portfolio return standard deviation even further; however, this also lowers their mean returns somewhat.

The three KP2011 portfolios have higher mean returns than Bench (14.77% to 14.96% vs. 11.38%), but also higher return standard deviations (22.05% to 22.34% vs. 20.40%). They have positive skewness and achieve the best minimum returns (-36.91% to -37.60%).

The simple equally-weighted portfolio in our sample delivered a higher mean and standard deviation as compared to Bench, and its minimum return is comparable with the one of Kuosmanen.

Turning to the mean-variance related strategies, we can see MinVarBench performs well along multiple dimensions. It has similar mean return and lower standard deviation than Bench. It has, however, a worse minimum return of -47.21% per year compared to -44.14% for Bench, -41.50% for Kuosmanen, and -36.91% for KP2011Min.

MaxSharpe delivers the highest mean return (17.15%), but the associated costs are the highest STD (27.02%) as well as the worst minimum return (-47.34%). The InformationRatio performs slightly better than Bench with somewhat higher mean and lower STD. The differences are, however, marginal.

4.2. Out-of-sample portfolios: portfolio performance and risk measures

Columns (1) to (3) of Table 2 report the quartiles from the distribution of p-values across 85 years for the null hypothesis that the portfolio in question does not dominate the benchmark.[12,13] Columns (4) to (11) of Table 2 report several other measures of portfolio performance: the Sharpe ratio, the Sortino ratio, VaR(5%), Shortfall(5%), certainty equivalents (CEV1, CEV3, CEV5), and Turnover.

[12] The 85 p-values are considered (one for each forecast window) from 1928 – 2012 using daily returns when reporting the percentiles. The interested reader can also find results on the number of periods during which the alternative portfolios dominate (are dominated by) the benchmark in the Online Appendix Table A1.

[13] Note, that the p-value distribution reflects both the actual performance of the portfolios according to the SSD criterion as well as the power of the Davidson (2009) test.

We see in Columns (1) to (3) of Table 2 that the SSD-related portfolios demonstrate a substantial range of performance. MinVar has the best median p-value of 0.231. The three KP2011 portfolios have the highest median p-values within the SSD-related group of portfolios, ranging from 0.608 to 0.686. Other SSD-related portfolios have median p-values around 0.30. Similarly, the 25^{th} percentile of the distribution of the p-values lies below 0.10 for all SSD-related strategies except for the KP2011 portfolios. This indicates that the hypothesis of non-dominance over Bench can be rejected for at least 25% of yearly tests for all the SSD-related portfolio choice techniques except the three KP2011 strategies. The high standard deviations of KP2011 portfolios that we saw in Table 1 are a major cause for their relative weak out-of-sample performance with respect to the SSD criteria.

The equally-weighted portfolio performs poorly in the SSD sense and has a median p-value of 0.934 with the 25^{th} percentile being 0.647. MaxSharpe and InformationRatio also have weak performances with the median p-values being 0.952 and 0.967 with 25^{th} percentiles being 0.357 and 0.860, respectively. 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.

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

Columns (1) – (3) report quartile p-values from the distribution of p-values across 85 out-of-sample periods with the underlying tests having the null hypothesis that the alternative portfolio does not SSD Bench. Columns (4) – (11) report other portfolio performance measures as indicated. Those statistics are computed using 85 yearly returns from 1928 to 2012.

	(1) 25% p-value	(2) Median p-value	(3) 75% p-value	(4) Sharpe ratio	(5) Sortino ratio	(6) VaR(5%)	(7) Shortfall (5%)	(8) CEV1	(9) CEV3	(10) CEV5	(11) Turnover
Bench	--	--	--	0.38	0.54	-28.18	-36.43	9.34	4.51	-1.34	0.00
SSD-Related											
Kuosmanen	0.062	0.282	0.708	0.44	0.58	-23.59	-33.27	10.38	6.26	1.12	0.65
KP2011Av	0.163	0.609	0.988	0.51	0.65	-22.56	-30.81	12.61	7.98	2.91	0.81
KP2011Min	0.216	0.686	0.990	0.51	0.64	-23.15	-30.67	12.74	8.06	2.99	0.82
KP2011Power3	0.208	0.608	0.985	0.51	0.65	-22.56	-31.28	12.76	8.00	2.78	0.81
MinShortfall	0.050	0.304	0.703	0.39	0.52	-18.68	-28.74	9.05	5.75	1.94	0.50
MinVar	0.052	0.231	0.617	0.40	0.52	-20.01	-29.64	9.09	5.82	1.80	0.41
MinSemivar	0.037	0.295	0.581	0.39	0.55	-19.93	-29.07	9.19	5.77	1.66	0.53
Equal	0.647	0.934	0.999	0.44	0.57	-28.54	-36.68	11.17	5.71	-0.57	0.00
Mean-Variance-Related											
MinVarBench	0.024	0.199	0.555	0.44	0.59	-21.17	-32.40	10.02	6.14	0.89	0.55
MaxSharpe	0.357	0.952	1.000	0.50	0.61	-26.93	-37.51	14.12	7.59	-0.05	0.87
InformationRatio	0.860	0.967	1.000	0.42	0.57	-27.27	-34.93	9.96	5.40	-0.03	0.29

Although within the mean-variance-related group, MinVarBench performs very well in the SSD sense. Its median p-value of 0.199 and 25th percentile of 0.024 are the smallest among all the strategies we tested. MinVarBench is often located in the SSD efficient set and thus behaves in much the same way as the SSD-related strategies. Consequently, its good performance on SSD tests is not surprising.

We now turn to the other, more traditional performance measures displayed in Columns (4) to (11) of Table 2. We see that, compared to Bench, Kuosmanen has higher realized values for the Sharpe ratio and Sortino ratio, lower VaR(5%) and Shortfall(5%), as well as better CEV values. Only for Turnover, where Bench by definition has none, does Kuosmanen fail to out-perform Bench. Due to their basic character of risk avoidance, MinVar, MinSemivar, and MinShortfall perform better than Kuosmanen based on left tail risk measures (VaR and expected Shortfall) as well as CEV5. Also, their Turnover values are somewhat lower than Kuosmanen. However, these approaches do not perform as well as Kuosmanen in terms of Sharpe and Sortino ratios or CEV1 and CEV3. The three KP2011 portfolio strategies out-perform Kuosmanen (and Bench) on all the traditional performance measures we examined except Turnover. They also out-perform MinVar, MinSemivar, and MinShortfall in terms of Sharpe and Sortino ratios as well as CEV values but not Var(5%), Shortfall(5%), or Turnover.

Compared with Bench, the equally-weighted portfolio in our sample delivered higher Sharpe and Sortino ratios (similar to Kuosmanen) as well as better CEV values. However, the equal-weighted portfolio had quite similar Var(5%) and Shortfall(5%) values to Bench as well as the same zero Turnover.

Turning to the mean-variance related strategies, we again see that MinVarBench performs similarly to the SSD-related group of portfolios, with results that are fairly close to those of Kuosmanen. MaxSharpe delivers relatively high out-of-sample Sharpe and Sortino ratios plus the highest CEV1 value (14.12). However, MaxSharpe does not perform well in terms of left tail risk measures (VaR and expected Shortfall) or CEV5. It also has the largest Turnover measure, indicating that the portfolio weights are not stable. The InformationRatio strategy performs similarly to Equal but with somewhat better performance on Var(5%), Shortfall(5%), and CEV5 while somewhat poorer values on Sharpe ratio, CEV1 and CEV3.

As indicated above, the SSD-related approaches as well as MinVarBench are generally performing well in terms of SSD as well as traditional performance measures. There is definitely variation in performance, with for example, the three Kopa2011 strategies not doing as well on SSD tests but better on Sharpe and Sortino ratios as well as CEV values. MinVar, MinSemivar, and MinShortfall tend to perform best on measures that are sensitive to minimizing left-tail risk but not as well on measures where higher mean returns receive more weight. In contrast, Max Sharpe, InformationRatio, and Equal tend to perform poorly on the SSD tests as well as measures that are sensitive to minimizing left-tail risk but do much better on measures where higher mean returns receive more weight.

The performance of Equal seems particularly instructive since it is not subject to estimation error problems and is also not specifically designed for a mean-variance environment. It does not perform well on SSD tests or measures that are sensitive to minimizing left-tail risk, presumably because it does not give any weight to reducing variance or left-tail risk. The equally-weighted portfolio performs better, but not spectacularly well, on measures such as the Sharpe and Sortino ratios that give more weight to mean returns. In effect, using equal weights is an attempt to avoid estimation errors and optimization procedures that may put extreme weights on some securities. However, that approach ignores information in the return sample that may allow construction of portfolios that exhibit desirable persistence in performance on at least some dimensions. In our tests, the SSD-related strategies as well as MinVarBench look promising for creating portfolios that perform well out-of-sample.

5. Robustness

In this section, we discuss robustness checks regarding methodological changes. We also investigate performance of the eleven 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

To explore robustness, we increased trimming levels for the z-interval to 10% and lowered them to 1% as well as using quarterly and two-year period lengths for estimation and forecast windows. A brief summary is provided below, with full results in the Online Appendix.

In Online Appendix Table A2 Column (1), we see that lowering tail trimming from 5% to 1% substantially increases the median p-values relative to those in Table 2 of the main paper. For example, the p-value for Kuosmanen increases from 0.282 (with 5% trimming in Table 2) to 0.611 here with 1% trimming. All the portfolio choice strategies have such increases because the reduced tail trimming decreases the power of the Davidson (2009) test. In Column (2) of Table A2, we see the opposite effect (decreased median p-values) when the trimming is increased to 10%. Note that results for MinVarBench and MinVar are now indicating rejection of non-dominance for those strategies over Bench in at least half the yearly test periods at the 5% and 6% levels, respectively. The rejection of non-dominance over Bench is only slightly weaker for Kuosmanen, Minshortfall, and MinSemivar.

Using two-year estimation and forecast windows instead of one-year windows cuts the number of out-of-sample periods in half but lengthens each test period. Strategies with relative low median p-values in Table 2 exhibit still lower median p-values with the two-year window, see Online Appendix Table A2 Column (3). In that column, we also see increases in the median p-values of strategies that had relatively high values in Table 2; however, there is little effect on ranking across the alternative strategies.

With quarterly estimation and forecast windows, median p-values generally increase with the notable exception of KP2011Min, see Online Appendix Table A2 Column (4).[14] The shorter window seems to hamper the strategies' performance as well as power of the Davidson test. Kuosmanen is particularly hard hit with its median p-value surging to 0.971. MinVar and MinVarBench hold up better, but their p-values both increase to approximately 0.35.

[14] 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).

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 monthly returns versus roughly 260 daily returns). Clearly, 12 data points is not going to provide a decent estimate of the underlying return distribution. Also, some of the approaches (e.g. Kuosmanen and KP2011Av) are effectively focused on maximizing 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 procedures tend to find weights which do not perform well out-of-sample and are not very stable across estimation periods.[15] We ultimately opted to use 20 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 each of 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

[15] In exploring this issue, we 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, KP2011Av and KP2011Power3 deliver similar SSD performances to those of other SSD-related portfolio strategies that do not require estimation of a mean. We conclude that difficulties in obtaining precise mean estimates in short samples and the corresponding over-fitting are relevant issues for these strategies. To mitigate this problem, a sufficiently long time series of data is needed.

available data. This resulted in 66 years of out-of-sample performance measurements for each approach.[16] We conducted the Davidson (2009) SSD test for non-dominance using the full 66-year period. Results can be found in Online Appendix Table A3, with most approaches in the same dominance class as Bench. The one exception is MaxSharpe, which is dominated by Bench with a p-value of 0.02. The lowest p-value for the null of non-dominance of an alternative portfolio over Bench is 0.16, which is achieved by MinShortfall. Being unable to detect stochastic dominance in this setting is not surprising. Since the optimal portfolio weights are estimated using 20 years of history, they do not focus on recent return dynamics and are likely to include structural breaks. Consequently, using monthly returns with a 20-year calibration period does not seem to be a good way to implement the SSD models we examined.

Table A4 in the Online Appendix contains results with monthly data for the other performance measures over the 66 out-of-sample years. Kuosmanen, MinShortfall, MinVar, MinSemivar, and MinVarBench all reduced return standard deviation relative to Bench. Not surprisingly, MinVar had the lowest standard deviation; 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 and having the highest Sortino ratio. The three KP2011 portfolios have slightly higher mean returns and somewhat higher standard deviation values relative to Bench (similar to what we saw previously in Table 2); however, those portfolios do not exhibit the marked improvement over Bench in terms of Sharpe ratios, Sortino ratios, and CEV values that we saw previously. 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, Sortino ratio, and CEV values out-of-sample. InformationRatio and Equal performed roughly on a par with Bench.

[16] Note, this approach is somewhat different from the one used in the main body of the paper with daily returns. Here we are not able to reliably apply our usual non-dominance test for every year, as we have only 12 monthly return observations per year.

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 a dominating portfolio to a dominated benchmark. An important question is how to choose among SSD efficient portfolios. We compare the methods of Kuosmanen (2004) as well as Kopa and Post (2011) 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 procedure of Kopa and Post (2011) allows for different SSD efficient portfolios depending on the specified preferences.

Several SSD-related portfolio strategies (Kuosmanen, MinShortfall, MinVar, and MinSemivar) perform well out-of-sample in terms of not only dominating the benchmark (using the SSD criterion) but also in terms of traditional performance measures. MinShortfall, MinVar, and MinSemivar do particularly well on measures focusing on risk reduction such as VaR(5%) and Shortfall(5%). The three KP2011 portfolio strategies do not perform as strongly in the SSD tests but do very well in terms of Sharpe and Sortino ratios as well as CEV values. 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 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 2 using our main runs, the SSD-related approaches have CEV3 results which improve over that of the benchmark by a range of 28% to 79%. With a greater level of risk aversion (CEV5), the situation is even more striking as the CEV5 of Bench is negative, indicating an individual with that utility function would pay not to invest in Bench.

The other portfolio strategies we tested (MaxSharpe, Information Ratio, and Equal) typically had a poor out-of-sample performance on dominance tests. Using traditional performance measures,

MaxSharpe did reasonably well out-of-sample in terms of Sharpe Ratio and CEV3 but still slightly underperformed the three KP2011 portfolios on those measures. On the other hand, MaxSharpe had the highest STD as well as the worst Min and Shortfall(5%). Intuitively, MaxSharpe has a propensity to find weights that accept more risk while seeking higher returns. When those weights do not generate high returns, the results can be quite poor. InformationRatio and the equal-weighted portfolio tended to perform on traditional measures slightly better than the benchmark. However, those two strategies frequently did less well than the SSD-related strategies.

Our robustness results indicate that power of the Davidson (2009) test does depend importantly on the trimming level. Also, using a quarterly estimation window does not seem to work well and harms the out-of-sample performance of the SSD-related approaches, with the notable exception of KP2011min. Using a two-year widow seemed to improve performance on the out-of-sample dominance tests for portfolio choice strategies that were doing well with the one-year window.

The concept of SSD-based portfolio choice is appealing theoretically, and there are several procedures allow for testing the in-sample SSD efficiency of a benchmark. Kuosmanen (2004) as well as Kopa and Post (2011) also deliver an optimal portfolio that dominates a benchmark portfolio in-sample. Interestingly, the out-of-sample performance of these portfolios is similar to relatively simple portfolio choice techniques that focus on risk reduction – in particular, MinVar and MinVarBench. The latter portfolio is constructed to minimize variance (in-sample) while matching the benchmark's mean return. As long as that benchmark mean return is not too extreme, MinVarBench will have a substantial emphasis on risk reduction, behaving much like MinVar and the other SSD-related approaches. It appears there is performance persistence regarding risk reduction captured by all these portfolios.

One reason why the more complex SSD-based approaches we tested do not outperform simpler ones out-of-sample may be that they are not designed to find a robust SSD efficient portfolio. Their primary goal is to test for in-sample efficiency of a benchmark. The optimal dominating portfolio is a "by-product" of that test. Developing advanced techniques of finding robust SSD-efficient portfolios seems to be an important avenue for future research.

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.

Kopa, M., Post, T., 2011. A general test for 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.

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 tables with the additional results mentioned in the section 4.2 and the robustness section 5.1 of the main paper.

Table A1. Out-of-Sample Performance of the Alternative Portfolio Choice Strategies

This table reports the number out of 85 forecast windows from 1928 – 2012, where the considered portfolios second-order stochastically dominate the benchmark (N^+), are dominated by the benchmark (N^-), or lie in the same dominance class (N^0). The alternative portfolios are constructed using the Fama-French 49 industry portfolios as assets. The last column reports the median p-values for the null hypothesis that the portfolio in question does not dominate the benchmark.

	N^+		N^-		N^0		Median p-Value
	#	%	#	%	#	%	
SSD-Related							
Kuosmanen	**27**	**32**	**56**	**66**	**2**	**2**	**0.282**
KP2011Av	**18**	**21**	**65**	**76**	**2**	**2**	**0.609**
KP2011Min	**14**	**16**	**69**	**81**	**2**	**2**	**0.686**
KP2011Power3	**16**	**19**	**67**	**79**	**2**	**2**	**0.608**
MinShortfall	**27**	**32**	**58**	**68**	**0**	**0**	**0.304**
MinVar	**30**	**35**	**55**	**65**	**0**	**0**	**0.231**
MinSemivar	**29**	**34**	**56**	**66**	**0**	**0**	**0.295**
Equal	**8**	**9**	**70**	**82**	**7**	**8**	**0.934**
Mean-Variance-Related							
MinVarBench	**29**	**34**	**56**	**66**	**0**	**0**	**0.199**
MaxSharpe	**11**	**13**	**66**	**78**	**8**	**9**	**0.952**
InformationRatio	**0**	**0**	**80**	**94**	**5**	**6**	**0.967**

Table A2. Out-of-Sample Performance of the Alternative Portfolios: Robustness

The table reports the median from the distribution of p-values obtained testing the null hypothesis that the portfolio in question does not dominate the benchmark over forecast windows from 1928 – 2012 using daily returns. In columns (1) and (2) there are 85 annual forecast windows. In column (3), there are 42 biannual forecast windows; and column (4) is based on 343 quarterly forecast windows. The alternative portfolios are constructed using the Fama-French 49 industry portfolios as assets. In Column (1), we trim the 1% highest and 1% lowest observations when conducting the out-of-sample SSD tests. In column (2), we trim the 10% highest and 10% lowest observations when conducting these tests. In column (3), we use two-year estimation and forecast periods; and in column (4), we use quarterly estimation and forecast periods.

	(1) 1% Tail Trimming	(2) 10% Tail Trimming	(3) 2-year estimation	(4) Quarterly estimation
SSD-Related				
Kuosmanen	**0.611**	**0.103**	**0.198**	**0.971**
KP2011Av	**0.802**	**0.517**	**0.767**	**0.972**
KP2011Min	**0.822**	**0.583**	**0.761**	**0.495**
KP2011Power3	**0.802**	**0.523**	**0.839**	**0.969**
MinShortfall	**0.556**	**0.103**	**0.171**	**0.482**
MinVar	**0.516**	**0.057**	**0.131**	**0.349**
MinSemivar	**0.514**	**0.112**	**0.180**	**0.472**
Equal	**0.965**	**0.887**	**0.966**	**0.890**
Mean-Variance-Related				
MinVarBench	**0.501**	**0.052**	**0.112**	**0.350**
MaxSharpe	**0.970**	**0.913**	**1.000**	**0.864**
InformationRatio	**0.975**	**0.967**	**0.930**	**0.929**

Table A3. Out-of-Sample Performance of the Alternative Portfolios with Monthly Returns

This table reports p-values for the SSD 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 constructed using the Fama-French 49 industry portfolios as assets. We use 20-year rolling estimation windows with 1 year out-of-sample forecasts.

	Alt. does not dominate Bench	Bench does not dominate Alt.
SSD-Related		
Kuosmanen	**0.74**	**1.00**
KP2011Av	**0.89**	**0.60**
KP2011Min	**0.93**	**0.49**
KP2011Power3	**0.84**	**0.74**
MinShortfall	**0.16**	**1.00**
MinVar	**0.34**	**1.00**
MinSemivar	**0.33**	**1.00**
Equal	**1.00**	**0.78**
Mean-Variance-Related		
MinVarBench	**0.42**	**1.00**
MaxSharpe	**1.00**	**0.02**
InformationRatio	**0.70**	**0.86**

Table A4. Descriptive Statistics of Portfolio Performance, Monthly Returns

This table reports descriptive statistics and traditional performance measures for the competing portfolio-choice strategies based on annual percentage returns. The alternative portfolios are constructed using the Fama-French 49 industry portfolios as assets. We use 20-year rolling estimation windows with 1 year out-of-sample forecasts. The statistics are computed using 66 yearly returns from 1947 to 2012.

	Mean	Median	STD	Min	Max	Skew-ness	Kurto-sis	Sharpe ratio	Sortino ratio	VaR (5%)	Shortfall (5%)	CEV1	CEV3	CEV5	Turnover
Bench	9.42	7.31	16.31	-38.21	50.18	-0.01	3.07	0.35	0.44	-13.51	-26.44	8.17	5.49	2.42	0.00
SSD-Related															
Kuosmanen	8.82	8.71	15.06	-28.81	55.51	0.36	3.77	0.34	0.42	-14.80	-22.53	7.79	5.71	3.53	0.28
KP2011Av	9.73	7.61	17.65	-33.94	71.57	0.69	4.65	0.35	0.41	-17.50	-25.14	8.37	5.64	2.81	0.26
KP2011Min	9.96	7.51	17.93	-35.03	71.39	0.63	4.46	0.35	0.42	-17.39	-25.84	8.55	5.71	2.72	0.27
KP2011Power3	9.59	8.02	17.37	-33.85	71.48	0.65	4.65	0.34	0.41	-17.65	-25.10	8.27	5.60	2.81	0.26
MinShortfall	9.88	9.86	15.01	-31.01	55.51	0.30	3.49	0.42	0.47	-13.91	-18.96	8.53	6.80	3.49	0.14
MinVar	9.24	8.62	13.92	-30.62	44.95	0.08	3.28	0.40	0.50	-13.76	-20.94	8.86	6.49	4.64	0.09
MinSemivar	9.62	9.27	15.35	-38.32	50.77	0.06	3.32	0.39	0.48	-14.23	-20.75	8.35	6.20	4.49	0.20
Equal	10.22	7.61	17.17	-39.05	53.53	0.18	3.40	0.38	0.46	-15.19	-26.81	8.88	6.00	2.70	0.00
Mean-Variance-Related															
MinVarBench	9.07	8.58	14.33	-30.10	59.53	0.33	4.17	0.38	0.45	-14.76	-21.23	8.14	6.23	4.21	0.16
MaxSharpe	8.47	4.18	17.57	-30.41	67.72	0.64	3.85	0.28	0.33	-19.12	-23.30	7.11	4.47	1.90	0.28
InformationRatio	9.58	8.07	16.23	-38.22	52.40	0.09	3.31	0.37	0.45	-14.44	-25.63	8.36	5.75	2.75	0.10