Portfolio Optimization based on Stochastic Dominance and Empirical Likelihood*

Thierry Post† and Selçuk Karabatı‡

March 15, 2017

Abstract

We develop and apply a portfolio optimization method based on the Stochastic Dominance (SD) decision criterion and the Empirical Likelihood (EL) estimation method. SD and EL share a distribution-free assumption framework which allows for dynamic and non-Gaussian return distributions. The SD/EL method can be implemented using a two-stage procedure which first elicits implied probabilities using Convex Optimization and subsequently constructs the optimal portfolio using Linear Programming. We apply the method to a range of equity industry momentum strategies. Our moment conditions are based on stylized facts about common risk factors in the stock market. SD/EL yields important ex-ante performance improvements relative to heuristic diversification, Mean-Variance optimization and a simple 'plug-in' approach. Relative to the CRSP all-share index, SD/EL improves average out-of-sample return by more than eight percentage points per annum, with less downside risk, semi-annual rebalancing and no short sales.

Key words: Portfolio optimization, Stochastic Dominance, Empirical Likelihood, Momentum strategies.

JEL Classification: C61, D81, G11

*We are grateful for the comments and suggestions by Yi Fang, Bogdan Grechuk, Campbell Harvey, David Hsieh, Jens Jackwerth, Miloš Kopa, Haim Levy, Meng Meng, Valerio Potì and participants in the Stochastic Dominance Theory and Applications Workshop of the Cambridge-INET Institute (15-16 September 2016) and research seminar at the Fuqua School of Business at Duke University (14 December 2016). Any remaining errors are our own.

†Post is Professor of Finance at the Graduate School of Business of Nazarbayev University; Astana, Kazachstan; e-mail: thierrypost@hotmail.com.

‡Karabatı is Professor of Operations Management at Koç University; 34450 Sarıyer/Istanbul, Turkey; e-mail: skarabati@ku.edu.tr.

1 Introduction

Stochastic Dominance (SD) is a time-honored maxim for investment decision making which avoids the usual Gaussian approximation to the return distribution (Hadar and Russell (1969); Hanoch and Levy (1969); Rothschild and Stiglitz (1970); Bawa (1975); Levy (2016)). This principle is particularly appealing for asset classes and investment strategies with asymmetric risk profiles, for example, small-cap stocks, junk bonds and momentum strategies.

Portfolio choice based on SD is analytically more demanding than mainstream Mean-Variance (MV) analysis. Nevertheless, modern-day computer hardware and optimization software bring this approach within reach of practical application.

For the common second-order SD (SSD) criterion and a discrete or discretized probability distribution, portfolio optimization can be formulated as a Linear Programming (LP) problem (Shalit and Yitzhaki (1994); Rockafellar and Uryasev (2000); Dentcheva and Ruszczynski (2003); Kuosmanen (2004); Roman, Darby-Dowman and Mitra (2006)). The problem is very large, but it remains tractable for realistic applications to security selection and asset allocation (Roman, Mitra and Zverovich (2013); Hodder, Jackwerth and Kolokolova (2015)).

An issue that has received limited attention in this literature is the statistical estimation of the joint return distribution of the base assets. Estimation accuracy is particularly relevant in portfolio optimization, because the optimal portfolio weights can be very sensitive to estimation error and optimized portfolios can perform poorly out-of-sample.

The econometric literature on SD focuses on testing hypotheses of dominance or non-dominance for a given set of choice alternatives (Anderson (1996); Davidson and Duclos (2000); Barrett and Donald (2003); Linton, Maasoumi and Whang (2005); Scaillet and Topaloglou (2010)). These studies unfortunately offer limited guidance for constructing a dominant alternative from elementary prospects.

The existing applications of SD optimization generally analyze the empirical distribution function (EDF). This 'plug-in' approach has sound points in large samples, but it ignores available conditioning information and becomes inaccurate if the available time series is short, especially if the number of base assets is large. The plug-in approach also does not account for common dynamic patterns such as price reversals and volatility jumps in high-frequency data.

This study proposes a framework for SD optimization based on the non-parametric Empirical Likelihood (EL; Owen (2001)) method. EL has several well-known desirable information-theoretic and statistical properties.

In the asset pricing literature, EL and related estimation methods have been proposed by Almeida and Garcia (2012, 2016), Julliard and Ghosh (2012), Garcia, and Vicente (2016) and Ghosh, Julliard and Taylor (2016). The focus of our study is instead on portfolio optimization with SD constraints.

Importantly, EL combines well with SD, due to a shared distribution-free assumption framework and the use of a discrete estimator for the return distribution - the 'EL probabilities'. The complementary relation between EL and SD was recognized earlier by Post and Potì (2017) and Post (2017). Those earlier studies use EL to test whether a passive market index is stochastically dominated, without identifying a superior alternative. By contrast, our study uses the EL probabilities in the construction of an enhanced portfolio.

EL uses the realized return vectors as the support of a multinomial distribution and the probabilities are estimated by minimizing the divergence to the sample distribution subject to a set of moment conditions. This approach preserves the structure of the historical scenarios and allows for a finite, state-dependent representation of the portfolio optimization problem. A block-wise application of EL can account for dynamic patterns in a flexible way (Kitamura (1997)).

The combination of SD and EL can be implemented using a tractable two-stage procedure which first elicits the EL probabilities using standard Convex Optimization (CO) and subsequently constructs the optimal portfolio using LP.

We apply SD/EL to a historical data set of daily returns to equity industry portfolios. Since we form portfolios based on intermediate past returns and rebalance the portfolios after a short-to-intermediate holding period, our investment strategies in effect exploit known industry momentum effects (Moskowitz and Grinblatt (1999)).

Diversification is a core principle for momentum strategies, as a concentrated position in just one or two top-performing industries would be riskier than most investors would be willing to tolerate. The most common momentum strategies rely on heuristic diversification rules without considering the risk levels and dependence structure. Our study aims to improve on heuristic diversification by using a combination of decision theory, mathematical programming and statistical estimation.

Since the objective is active portfolio construction, we avoid the assumption that benchmark index is efficient. The return premium to winner industries seems to defy rational risk-based explanations. More plausible explanations are based on market imperfections such as short sales restrictions and behavioral biases such as the 'disposition effect'.

Instead, our moment conditions are based on general stylized facts about common risk factors in the stock market. The use of risk factor models for improving upon plug-in estimates is well-established for MV optimization. Distinguishing features of our approach are the use of model-free moment conditions which avoid a full-fledged model specification and the estimation of EL probabilities rather than mean and covariance terms only.

2 Methodology

This section presents the SD/EL method. For the sake of brevity, we focus on the common SSD rule. In this case, SD/EL can be implemented using a tractable two-stage procedure which uses standard CO to elicit the EL probabilities and LP to construct the optimal portfolio given the EL probabilities.

The framework can however be applied directly to first-order SD (FSD) and third-order SD (TSD). Stage-Two optimization based on those alternative rules is generally more expensive than SSD; FSD requires Mixed Integer Linear Programming (MILP) and TSD requires Quadratic Programming (QP). However, the probability elicitation is based on a separate optimization problem which does not affect the complexity of the portfolio construction problem.

2.1 Preliminaries

We consider K distinct base assets with random investment returns $\boldsymbol{x} \in \mathcal{X}^K$ with joint Cumulative Distribution Function (CDF) $\mathcal{F}(\boldsymbol{x})$, $\mathcal{X} := [a, b]$, $-\infty < a < b < +\infty$. The portfolio possibility set consists of all convex combinations of the base assets: $\Lambda := \left\{\boldsymbol{\lambda} \in \mathbb{R}^K : \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1}_K = 1; \boldsymbol{\lambda} \geq \mathbf{0}_K\right\}$. Our goal is to enhance a given benchmark portfolio which is characterized by weights $\boldsymbol{\tau} \in \Lambda$ and returns $y := \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}$.

Importantly, the base assets are not restricted to be individual securities. In general, the base assets are defined as the most extreme feasible combinations of individual securities. This formulation allows for general linear portfolio constraints, including bounded short sales constraints, position limits and restrictions on risk factor loadings. To allow for dynamic intertemporal investment problems, these combinations may be periodically rebalanced based on a conditioning information set.

The marginal CDF for portfolio $\boldsymbol{\lambda} \in \Lambda$ is given by

$$\mathcal{F}_{\boldsymbol{\lambda}}(z) := \int\limits_{\{\boldsymbol{x}\in\mathcal{X}^K : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda}\leq z\}} d\mathcal{F}(\boldsymbol{x}). \tag{1}$$

The first lower partial moment for portfolio $\boldsymbol{\lambda} \in \Lambda$ and threshold return $z \in \mathcal{X}$, amounts to:

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\lambda}}(z) &:= \int_a^z \mathcal{F}_{\boldsymbol{\lambda}}(x)dx \\ &= \mathbb{E}_{\mathcal{F}_{\boldsymbol{\lambda}}}\left[(z-x)\mathcal{I}\left(x \leq z\right)\right], \end{aligned} \tag{2}$$

where $\mathcal{I}(\cdot)$ is an indicator function that takes a value of unity if the condition within parentheses is satisfied and zero otherwise. The first lower partial moment is also known as 'expected shortfall', although the latter term may also refer to Conditional Value at Risk (CVaR).

Definition 2.1.1: *A given portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates the benchmark by SSD, or $\boldsymbol{\lambda} \succeq_{\mathcal{F}} \boldsymbol{\tau}$, if*

$$\mathcal{L}_{\boldsymbol{\lambda}}(z) \leq \mathcal{L}_{\boldsymbol{\tau}}(z), \forall z \in \mathcal{X}. \tag{3}$$

The economic interpretation of SSD is well-established: $\boldsymbol{\lambda} \succeq_{\mathcal{F}} \boldsymbol{\tau}$ if and only if $\mathbb{E}_{\mathcal{F}_{\boldsymbol{\lambda}}}[u(x)] \geq \mathbb{E}_{\mathcal{F}_{\boldsymbol{\tau}}}[u(x)]$ for all monotone and concave utility functions $u(x)$. The equivalence of the two formulations stems from the observation that all relevant utility functions are positive mixtures of elementary Russell and Seo (1989) ramp functions $v_z(x) := \min(x-z, 0)$, $z \in \mathcal{X}$; expected shortfall equals the negative expectation of an elementary utility function: $\mathcal{L}_{\boldsymbol{\lambda}}(z) := -\mathbb{E}_{\mathcal{F}_{\boldsymbol{\lambda}}}[v_z(x)]$.

If the constraints (3) are included in a portfolio optimization problem, then the left-hand side ($\mathcal{L}_{\boldsymbol{\lambda}}(z)$) depends on the portfolio weights ($\boldsymbol{\lambda}$), which are model variables. In general, these constraints are not trivial from a numerical perspective, as they have infinite dimensions and are not linear. However, the conditions can be captured by a finite system of linear constraints for discrete or discretized distributions, as we will see in Section 2.3 below.

2.2 Stage One: Probability Elicitation

The CDF is latent and has to be estimated using a time series of realized return vectors, $\boldsymbol{X}_s := (X_{1,s} \cdots X_{K,s})^{\mathrm{T}}$, $s = 1, \cdots, T$. The EDF amounts to $\mathcal{F}_T(\boldsymbol{x}) := T^{-1} \sum_{t=1}^{T} \mathcal{I}(\boldsymbol{X}_t \leq \boldsymbol{x})$; it assigns an equal probability to each vector, or $\mathbb{P}_{\mathcal{F}_T}[\boldsymbol{x} = \boldsymbol{X}_s] = T^{-1}$, $s = 1, \cdots, T$. Existing empirical applications of SD optimization generally employ this 'plug-in' estimator.

The EDF is a natural starting point, because realized return vectors by their own nature are realistic scenarios. Furthermore, its discrete structure is convenient for numerical optimization. In addition, if the observations are serially Independent and Identically Distributed (IID), then the EDF is a statistically consistent, nonparametric Maximum Likelihood estimator. Unfortunately, the plug-in approach becomes inaccurate if the time series is short or exhibits dynamic patterns.

The EL method seems well suited for improving estimation accuracy, while preserving the aforementioned favorable properties. The method combines the data set and a set of moment conditions, using information theory. Using $\boldsymbol{g}(\boldsymbol{x})$ for a vector-valued moment function on $\mathcal{X}^K$, the moment conditions follow:

$$\mathbb{E}_{\mathcal{F}}\left[\boldsymbol{g}(\boldsymbol{x})\right] = \mathbf{0}_M. \tag{4}$$

The moment function can be used to capture conditioning information, for example, theoretical equilibrium conditions or empirical stylized facts. The moment function generally depends on latent model parameters and/or nuisance parameters that are estimated as part of the procedure, but we suppress those parameters here to simplify the notation. Moment inequalities can be represented by including latent slack variables in the moment function.

It can be convenient to use moment conditions for common factors which are constructed as linear combinations of the returns to the base assets, or $f_j := \boldsymbol{x}^T \boldsymbol{w}_j$, $j = 1, \cdots, F$. This approach can capture prior information about factor risk premiums such as the equity risk premium, term spread and credit spread. If the factors are orthogonalized, then the analysis could focus on restricting the marginal distribution rather than their joint distribution. Section 3.3 illustrates the factor-based approach using the four common factors of Carhart (1997).

Whereas EL was originally developed for testing the moment conditions and inference about the latent model parameters, we use its 'implied probabilities' to improve upon the EDF as an estimator for the CDF, taking the moment conditions to be true. In the same way, the 'EL bootstrap' (Brown and Newey (2002)) uses implied probabilities to simulate the null distribution and 'Bayesian Exponentially Tilted Empirical Likelihood' (BETEL; Schennach (2005)) uses implied probabilities as a nonparametric likelihood function to compute Bayesian posterior probabilities.

EL estimates the distribution function using a multinomial distribution with atoms at the data points. The implied probabilities $p_s := \hat{\mathbb{P}}[\boldsymbol{x} = \boldsymbol{X}_s]$, $s = 1, \cdots, T$, are found by minimizing the Relative Entropy, or the Kullback-Leibler divergence to the sample probabilities $\pi_s := \mathbb{P}_{\mathcal{F}_T}[\boldsymbol{x} = \boldsymbol{X}_s] = T^{-1}$, $s = 1, \cdots, T$,

$$
\begin{aligned}
D_{KL}\left(\{\pi_s\} \parallel \{p_s\}\right) : &= \sum_{s=1}^{T} \pi_s \ln\left(\frac{\pi_s}{p_s}\right) \\
&= \frac{1}{T}\sum_{s=1}^{T} \ln(p_s) - \ln(T),
\end{aligned} \tag{5}
$$

subject to the empirical equivalent of the moment conditions,

$$
\sum_{t=1}^{T} p_t \boldsymbol{g}(\boldsymbol{x}_t) = \mathbf{0}_M. \tag{6}
$$

Block-wise Empirical Likelihood (BEL; Kitamura (1997)) applies this approach to data blocks rather than individual observations. The block-wise application can account for a range of dynamic patterns, including common stationary ARMA, GARCH and stochastic volatility processes. This approach subdivides the original time series into $T^* := (T - B + 1)$ maximally overlapping blocks of B consecutive observations.

Specifying the proper block length B involves a trade-off between the maximal duration of the dynamic effects and the number of independent blocks, or $\lfloor T/B \rfloor$. For serially IID observations, $B = 1$ yields the standard EL method. It is generally recommended that the block length B grows with T but at a lower rate, for example, $(B^{-1} + B^2/T) \to 0$ as $T \to \infty$. The block length also depends also on the aggregation level of the base assets and the return frequency; BEL seems particularly relevant for high-frequency data of individual securities and less relevant for low-frequency data and diversified portfolios.

The BEL procedure assigns probabilities $b_1, \cdots, b_{T^*}$ to the data blocks and equal (conditional) probabilities (B^{-1}) to the observations in a given block; the procedure then seeks to minimize $D_{KL}\left(\{\beta_i\} \parallel \{b_i\}\right)$, where $\beta_i := (T^*)^{-1}$. Observation $t = 1, \cdots, T$ is included in all blocks with indices from $t^- := \max(1, t - B + 1)$ to $t^+ := \min(t, T^*)$. The implied probabilities therefore amount to

$$p_t \propto (t^+ - t^- + 1)^{-1} \left(\sum_{i=t^-}^{t^+} b_i \right),\ t = 1, \cdots, T. \tag{7}$$

The block-level probabilities $b_1, \cdots, b_{T^*}$ are the solution to the following Minimum Relative Entropy (MRE) problem::

$$\begin{aligned} \min & \left(-\frac{1}{T^*} \sum_{i=1}^{T^*} \ln(b_i) - \ln(T^*) \right) \\ \text{s.t.} & \sum_{t=1}^{T} \sum_{i=t^-}^{t^+} b_i (t^+ - t^- + 1)^{-1} \boldsymbol{g}(\boldsymbol{X}_t) = \mathbf{0}_M; \\ & \sum_{i=1}^{T^*} b_i = 1; \\ & b_i \geq 0,\ i = 1, \cdots, T^*. \end{aligned} \tag{8}$$

This problem is a standard CO problem, and, for sample size in our study ($T \approx 250$ trading days in a moving 12-month estimation window), the problem size is small. The problem however can become computationally demanding in applications which analyze intraday data or solve the problem repeatedly in simulation or re-sampling routines. For such applications, there exists an alternative, more tractable representation of the problem in terms of the Lagrange multipliers of the original minimization problem.

Using the implied probabilities, the joint CDF is estimated by

$$\hat{\mathcal{F}}(\boldsymbol{x}) := \sum_{t=1}^{T} p_t \mathcal{I}\left(\boldsymbol{X}_t \leq \boldsymbol{x}\right). \tag{9}$$

Similarly, the EL estimator for the marginal CDF for portfolio $\boldsymbol{\lambda} \in \Lambda$ is

$$\hat{\mathcal{F}}_{\boldsymbol{\lambda}}(z)=\sum_{t=1}^{T} p_t \mathcal{I}\left(\boldsymbol{X}_s^{\mathrm{T}}\boldsymbol{\lambda} \leq z\right). \tag{10}$$

As the sample size increases, the effect of the moment conditions vanishes asymptotically and p_t, $t = 1, \cdots, T$, converges to the value of T^{-1}, under suitable regularity conditions. Hence, as $T \to \infty$, the EL estimator $\hat{\mathcal{F}}(\boldsymbol{x})$ converges to the EDF $\mathcal{F}_T(\boldsymbol{x})$, which itself converges in probability to the true CDF $\mathcal{F}(\boldsymbol{x})$. However, in small samples, $\hat{\mathcal{F}}(\boldsymbol{x})$ in contrast to $\mathcal{F}_T(\boldsymbol{x})$ has the advantage of obeying the moment conditions, which represent known properties of $\mathcal{F}(\boldsymbol{x})$.

2.3 Stage Two: Portfolio Construction

The EL estimator $\hat{\mathcal{F}}(\boldsymbol{x})$ is also analytically convenient for portfolio analysis. The state-dependent structure is advantageous, because the state probabilities (p_s, $s = 1, \cdots, T$) are independent of the portfolio weights ($\boldsymbol{\lambda}$) which simplifies the statistical calculus for portfolio construction.

Furthermore, the estimated lower partial moment takes a simple piecewise-linear, increasing and convex shape:

$$\begin{aligned}\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) &= \mathbb{E}_{\hat{\mathcal{F}}_{\boldsymbol{\lambda}}}\left[(z-x)\mathcal{I}\left(x \leq z\right)\right] \\ &= \sum_{t=1}^{T} p_t (z - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda})\mathcal{I}\left(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq z\right).\end{aligned} \tag{11}$$

As a consequence, we have to check the dominance condition only at the observed return levels for the benchmark ($z = y_s$, $s = 1, \cdots, T$), as in Bawa, Bodurtha, Rao and Suri (1985, Section I.C):

Proposition 2.3.1: *A given portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates the benchmark by SSD under the implied probability distribution, or $\boldsymbol{\lambda} \succeq_{\hat{\mathcal{F}}} \boldsymbol{\tau}$, if and only if*

$$\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_s) \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(y_s),\ s = 1, \cdots, T. \tag{12}$$

Although the number of constraints is finite (T), the left-hand sides, $\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_s)$, $s = 1, \cdots, T$, are non-linear functions of the portfolio weights. At first sight, the binary variables $\mathcal{I}\left(\boldsymbol{X}_s^{\mathrm{T}}\boldsymbol{\lambda} \leq z\right)$, $s = 1, \cdots, T$, seem to require integer programming. However, we may avoid integer programming by using a linear relaxation the spirit of Rockafellar and Uryasev (2000, Section 3):

$$
\begin{aligned}
\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) = \min_{\theta} \sum_{t=1}^{T} p_t \theta_t \qquad (13) \\
\theta_t \geq z - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda},\ t = 1, \cdots, T; \\
\theta_t \geq 0,\ t = 1, \cdots, T.
\end{aligned}
$$

This problem is designed such that $\theta_t^* = (z - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda}) \mathcal{I}\left(\boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda} \leq z\right),\ t = 1, \cdots, T$, is the optimal solution, which removes the need to use binary variables.

It follows from (13) that we can reformulate the inequality $\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(z)$ using the following linear system:

$$
\begin{aligned}
\sum_{t=1}^{T} p_t \theta_t \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(z); \qquad (14) \\
\theta_t \geq z - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda},\ t = 1, \cdots, T; \\
\theta_t \geq 0,\ t = 1, \cdots, T.
\end{aligned}
$$

Specifically, if there exists any feasible solution θ_t^{**}, $t = 1, \cdots, T$, to this system, then the minimizer $\theta_t^* = (z - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda}) \mathcal{I}\left(\boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda} \leq z\right)$, $t = 1, \cdots, T$, is also an feasible solution, which implies $\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(z)$.

We can apply linear system (14) to every threshold $z = y_s$, $s = 1, \cdots, T$, in the SSD conditions (12). For this purpose, we introduce the model variables $\theta_{s,t}$, $s, t = 1, \cdots, T$, to capture the terms $(y_s - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda}) \mathcal{I}\left(\boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda} \leq y_s\right)$, $s, t = 1, \cdots, T$. In addition, we treat the portfolio weights $\boldsymbol{\lambda} \in \Lambda$ as model variables, which does not introduce further complications, because the constraints of problem (13) are linear in the portfolio weights.

Combining these insights, we can identify SSD enhanced portfolios as solutions to the following system of linear constraints:

$$\begin{aligned}\sum_{t=1}^{T} p_t \theta_{s,t} \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(y_s),\ s = 1, \cdots, T; \\ -\theta_{s,t} - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\lambda} \leq -y_s,\ s, t = 1, \cdots, T; \\ \mathbf{1}_K^{\mathrm{T}} \boldsymbol{\lambda} = 1; \\ \theta_{s,t} \geq 0,\ s, t = 1, \cdots, T; \\ \lambda_k \geq 0,\ k = 1, \cdots, K.\end{aligned} \tag{15}$$

Any feasible solution $\boldsymbol{\lambda}^*$ to this system dominates the benchmark portfolio $\boldsymbol{\tau}$ by SSD.

To find an SSD enhanced portfolio, we can develop mathematical programs that optimize an objective function given the above system of constraints. A convenient specification for the objective function is

$$O_{\boldsymbol{\lambda}} := \hat{\boldsymbol{\mu}}^{\mathrm{T}} \boldsymbol{\lambda} - \sum_{t=1}^{T} w_t \hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_t), \tag{16}$$

where $\hat{\boldsymbol{\mu}} := \sum_{t=1}^{T} p_t \boldsymbol{X}_t$ is the mean vector and $w_t \geq 0$, $t = 1, \cdots, T$, are decision weights. Objective function (16) is a linear function of the portfolio weights, and hence we end up with an LP problem. Although the optimization problem is much larger than a standard MV problem, it is perfectly manageable with standard computer hardware and optimization software in typical applications.

The goal in our application section is to maximize the expected portfolio return, or the extreme case with $w_t = 0$, $t = 1, \cdots, T$. This orientation allows for a direct comparison with heuristic portfolio construction rules and with MV optimization based on the same objective function. From a theoretical perspective, this specification introduces the possibility that the solution portfolio is not fully efficient. However, unreported results show no material performance improvements from using small non-zero decision weights or multi-objective programming with risk reduction as the secondary objective function.

Another example of an objective function which is consistent with the SSD criterion is minimizing the portfolio expected shortfall for a given threshold level or a positive linear combination of expected shortfall for multiple threshold levels. Kopa and Post (2015, Section 5) show how to specify decision weights based on a pre-specified utility function.

2.4 Hardware and software

In our application of SD/EL below, the probability estimation problem (8) is modeled in GAMS and solved with CONOPT nonlinear solver (Drud (1994)) on a machine with an Intel Core i7 2.38GHz processor and 8GB memory. The portfolio optimization problem (maximize objective (16) subject to linear system (15)) is modeled in and solved with IBM ILOG CPLEX Optimization Studio 12.6.3.0 on a machine with 2 x Intel(R) Xeon(R) CPU E5-2695 2.40GHz processors and 512GB memory. The combined computation time for the probability estimation problem and portfolio optimization problem was around half a minute for the typical data set ($K = 49$ base assets, $T \approx 250$ observations, $M = 8$ moment conditions).

3 Application

We implement various equity industry momentum strategies and analyze the performance improvements from using the SD choice criterion and the EL estimation method. In this application, the data dimensions seem unfavorable for the plug-in approach and conditioning information seems a useful addition to the data set. We propose a set of model-free moment conditions based on stylized factors of common risk factors for stock returns.

3.1 Data

Our benchmark is the all-share index from the Center for Research in Security Prices (CRSP), a value-weighted average of common stocks listed on the NYSE, AMEX and NASDAQ stock exchanges.

We repeated our analysis with the equal-weighted average of the base assets as the benchmark. 'Naive diversification' is a simple and effective way to achieve robust outperformance (DeMiguel, Garlappi and Uppal (2009)). However, the equal-weighted portfolio does not capture the momentum effects and it can easily be enhanced using optimization in this study. Better performance can be achieved using a conditional $1/K$ rule, which is discussed below. However, also that heuristic appears suboptimal in our analysis.

The base assets consist of a set of $K = 10, 17, 49$ value-weighted industry portfolios which are formed by grouping individual stocks based on their four-digit Standard Industrial Classification (SIC) codes.

Since the base assets are diversified industry portfolios, the analysis does not allow for concentrated positions in individual stocks. The analysis also does not allow for short sales, because the base assets include only long positions in individual stocks. These portfolio-weight restrictions limit the effects of estimation error (Frost and Savarino (1988), Board and Sutcliffe (1994), Jagannathan and Ma (2003), DeMiguel *et al.* (2009)). In addition, the resulting strategies can be implemented at lower transactions costs than a typical stock-level long-short strategy. A possible, cost-effective implementation method would buy exchange-traded funds (ETFs) that track specific sector indices. Furthermore, excluding short sales limits the effect of momentum crashes, which are partly attributable to the rebound of loser stocks after market sell-offs (Daniel and Moskowitz (2016)).

Our data set consists of daily total returns to the industry portfolios and daily values of risk factors from Kenneth French's online data library.[1] The sample period is from January 3, 1927, through December 31, 2015.

3.2 Investment strategies

Earlier studies of industry momentum show that buying winners is most profitable when using an intermediate estimation window and a short-to-intermediate holding period. We therefore focus on an estimation window of 12 months and a holding period of $H = 3, 6, 12$ months. Annual rebalancing occurs on the first day of the month of January of every year; semi-annual rebalancing in the months of January and July; quarterly rebalancing in January, April, July and October.

At every formation date, we construct five different enhanced portfolios from the base assets. The first, heuristic portfolio is an equal-weighted average of the five industries with the highest realized return in the past 12 months. This 'EW5' portfolio uses a conditional version of 'naive diversification'. In contrast to the equal-weighted average of all industries, this strategy does capture the momentum effects. The other four enhanced portfolios are constructed through optimization. The objective is to maximize the mean ($O_{\boldsymbol{\lambda}} = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\lambda}$) subject to the restriction that the enhanced portfolio dominates the benchmark by MV or SSD, using naive sample probabilities or EL implied probabilities.

Our EL analysis uses a short block length of $B = 1, 2, 3$ trading days. A short length is used, because the base assets are diversified industry portfolios, rather

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library

than individual stocks, and, consequently, we don't observe strong dynamic patterns. In addition, our typical estimation window contains only roughly 250 trading days, and a long block length would leave us with only few non-overlapping blocks compared with the number of base assets.

3.3 Moment conditions

Since our estimation window is short, it seems useful to incorporate conditioning information using implied probabilities. We deliberately use conservative, uncontroversial moment conditions that do not require explicit testing, although the investment performance from using the resulting implied probabilities arguably is an implicit test.

Our specification is based on general stylized facts about stock returns. The bulk of the joint variation of returns can be explained with a small number of common factors. The predictability of these factors is limited, as expected in an competitive capital market.

As a case in point, the daily excess return to a general stock market index, which is a popular choice as a common factor of individual stocks returns, shows only weak predictability using past returns and fundamentals compared with the daily returns to individual stocks. Therefore, the market return strongly regresses towards the mean after large downswings or upswings.

In order to capture this regression effect in a model-free way, we 'winsorize' the factor means using moment inequalities. Our moment conditions are

$$L_j \leq \mathbb{E}_{\mathcal{F}}[f_j] \leq U_j,\ j = 1, \cdots, F, \tag{17}$$

where L_j and U_j, $j = 1, \cdots, F$, are cut points for the left and right tails, respectively. Put differently, we use $M = 2F$ moment conditions and the elements of the moment function $\boldsymbol{g}(\boldsymbol{x})$ are given by $g_{2j-1}(\boldsymbol{x}) = f_j - s_j^- - L_j$ and $g_{2j}(\boldsymbol{x}) = f_j + s_j^+ - U_j$, $j = 1, \cdots, F$, where $s_j^-, s_j^+ \geq 0$ are latent slack variables.

We use the four factors of Carhart (1997): (1) general stock market return minus T-bill return ('RMRF'), (2) small-caps return minus big-caps return ('SMB'), (3) value-stocks return minus growth-stocks return ('HML') and (4) winner-stocks return minus loser-stocks return ('MOM'). For the typical industry portfolio, these factors explain more than 60% of the daily return variation.

We consider $F = 1, 3, 4$ factors: the single-factor 'market model' based on RMRF ($F = 1$); the Fama and French (1993) three-factor model based on RMRF, SMB and HML ($F = 3$); the Carhart (1997) four-factor model based on all four factors ($F = 4$).

Let $C = 5, 10, 25$. For the lower cut point, L_j, $j = 1, \cdots, 4$, we use the C-th percentile of the historical distribution, across all estimation windows, of the factor sample mean; similarly, for the upper upper cut point, U_j, $j = 1, \cdots, 4$, we use the $(100 - C)$-th percentile.

To address possible concerns about forward-looking bias, the analysis for the second half of the sample period was repeated using cut points based on the estimation windows from the first half of the sample period only. Encouragingly, the results are not materially affected by this modification, consistent with the below robustness for the specification of C.

If the moment conditions are violated in a given estimation window, then the implied probabilities diverge from the EDF to adjust the factor means. For our base case with $F = 4$ and $C = 10$, almost half of the samples are adjusted in this way. The adjustments are relatively modest, in the sense that regression-based forecasts of the factors will generally be closer to the mean than our cut points and therefore require stronger divergence from the EDF.

In practice, the effect of these adjustments is to avoid large directional bets on common factors, and to improve the focus on industry-specific events. For example, the adjustments reduce the bias of the optimal portfolio towards defensive industries after a general market sell-off and hence reduce the exposure to momentum crashes.

Figure 1 illustrates the regression effect and our moment conditions. Each panel plots the in-sample mean against the out-of-sample mean for one of the four factors, using a 6-month holding period ($H = 6$). The regression effect is particularly strong after large negative values, such as the market crash of 2008 and the momentum crash of 2009. The dotted 45° line represents naive extrapolation of the in-sample values. The solid, piecewise line represents our moment conditions, using winsorizing at the tenth and 90-th percentile. For all four factors, the latter approach gives a better fit to the data.

[Insert Figure 1 about here.]

3.4 Evaluation method

We evaluate out-of-sample performance using annual, January-December returns from 1928 through 2015, a total of $N = 88$ evaluation periods. For short holding periods ($H = 3, 6$), this approach means that rebalancing takes place not only at the beginning of the evaluation period, but also during the evaluation period.

We report the mean, standard deviation, semi-deviation and skewness of the annual out-of-sample returns. We also report the mean spreads between the competing strategies and the associated t-statistics, to gauge the level of statistical significance of the performance improvements.

As distribution-free risk measures, we include the Value-at-Risk (VaR) and CVaR for a confidence level of 95%. As a risk-adjusted performance measure, we use the certainty equivalent return for a standard logarithmic utility function, $u(x) = \ln(100 + x)$.

We also report the Sharpe ratio and, to account for asymmetry, the Sortino ratio. However, the usual interpretation of these ratios does not apply here, as it is not possible to 'scale' the enhanced portfolio without violating the constraints on short sales and benchmark risk. It seems more relevant to evaluate the strategies by their ability to simultaneously enhance the mean and reduce downside risk.

Corrections for exposures to common risk factors have a limited effect on the relative performance of the strategies in our application. As a result of the portfolio-weight restrictions, all enhanced portfolios have a similar market beta. The exposures to non-market risk factors are limited, due to the diversified nature of the industry portfolios, the absence of short positions and the frequency of portfolio rebalancing. Even the exposures to stock-level momentum factors are limited, because the strategies rely on industry momentum rather than stock-level momentum and on buying winners rather than selling losers.

3.5 The base case

We start our analysis with the combination of $K = 49$ industries; a six-month holding period ($H = 6$); $F = 4$ risk factors (the Carhart (1997) model); cut points for the factor means based on $C = 10$; a block length of $B = 1$ (the IID EL method). This combination seems favorable given the aforementioned industry momentum studies and stylized facts.

Table I summarizes the out-of-sample performance of the competing portfolios. The benchmark index ('Bench') on average yields 11.54% per annum with a standard deviation of 20.27% in this sample period. The negative skewness reflects the elevated correlation between stocks during market downswings, which limits the potential for reducing downside risk through broad diversification.

The heuristic strategy (EW5) enhances average return by 6.70% to 18.24%. The EW5 portfolio has a higher standard deviation than the benchmark, due to the lower level of diversification. It is possible to reduce the standard deviation by including more winner industries (for example, 'EW10'), but this approach will also lower the mean. A more promising route is to explicitly account for the risk and dependence structure of the industries, using estimation and optimization.

Using the plug-in approach, the performance enhancement from MV optimization remains limited. The MV approach improves average return by 6.82% to 18.36%, only 0.12% above the EW5 strategy. Using implied probabilities clearly improves matters by adding another 0.56% per annum. The standard deviation of the portfolio exceeds that of the benchmark, but the variability stems mostly from upside rather than downside deviations. Unfortunately, the MV rule does not recognize this asymmetry and it therefore penalizes the upside potential of concentrated positions in past winner industries.

SD optimization leads to important further performance improvements. Using the plug-in approach, the improvements are again limited, but the full potential of SD comes to light when we use EL estimation. Using implied probabilities, SD enhances average return by 8.38% to 19.92%, an improvement of 1.68% compared with EW5, 1.55% compared with MV and 1.00% compared with MV/EL. The improvements from using implied probabilities are statistically significant at every conventional significance level. EL estimation is particularly beneficial for the SD approach, because the SD portfolio tends to be less diversified and more exposed to regression effects than the MV portfolio.

Due to the asymmetry of the return profile, the improvements are robust to risk corrections, despite the higher standard deviation. Specifically, the SD/EL portfolio has substantially lower VaR and CVaR than the benchmark and it maximizes the certainty equivalent and Sortino ratio.

[Insert Table I about here.]

Figure 2 illustrates the cumulative performance of the competing portfolios for the entire sample period. Shown is the dollar value of the benchmark and the five enhanced portfolios, assuming an initial investment of one dollar in January

1928. Not surprisingly, the differences in investment return of tens or hundreds of basis points per annum translate into exponentially growing differences in portfolio value over time. The SD/EL portfolio leads all other portfolios from 1931 until 2015. In 2015, after 88 years, it reaches a value of $830,487, compared with only $3,020 for the benchmark and $561,869 for the second-best portfolio (MV/EL). The limited downside risk of the SD/EL portfolio is reflected by a maximum drawdown of 41%, which compares favorably with 69% for the benchmark and 46% for the MV/EL portfolio.

[Insert Figure 2 about here.]

The above performance improvements seems impressive given that they are achieved with the 'minimal' adjustment of winsorizing the factor means. The moment inequalities are binding only in the samples in which some of the factors take extreme values. Conditional on extreme events, the performance improvements from using implied probabilities is even more impressive.

Table II gives more details of the performance following the most severe 'factor crashes'. Whereas the heuristic strategy fails to improve on the benchmark, SD/EL enhances average return by as much as 15.49%. Roughly half of these improvements stem from the use of implied probabilities. If a factor crashes, then industries with low exposure to that factor will appear attractive, based on sample probabilities. Including these industries in the momentum portfolio will then lower the portfolio's expected return, due to the regression effect. Using implied probabilities mitigates this problem and hence offers 'crash protection'.

[Insert Table II about here.]

Figure 3 plots the relation between the implied probabilities and the factor values during the factor crashes. In these periods, the implied probabilities must diverge from the sample probabilities in order to raise the expected value of the factor above the lower cut point. Hence, the implied probability tends to increase with the factor value. If the other factors do not take extreme values, then a perfect bivariate relation arises. By contrast, if multiple moment conditions are violated, then the bivariate relation becomes noisy. If BEL is used, then the noise level tends to increase at the level of individual trading days but it decreases at the level of data blocks.

[Insert Figure 3 about here.]

3.6 Alternative specifications

We will now analyze the effects of varying the parameters of the investment strategy: the number of base assets (K), the length of the holding period (H), the number of risk factors (F), the choice of the cut points (C) and the block length (B). Table III summarizes the results. To allow for a compact presentation, we vary only one parameter (K, H, F, C or B) at the time, using the strategy with $K = 49$, $H = 6$, $F = 4$, $C = 10$ and $B = 1$ as the base case. In addition, we now tabulate only the average return and suppress all other output.

The performance of all momentum strategies improves as we widen the investment universe, consistent with the results of previous momentum studies. Given that estimation risk increases with the number of base assets, it seems not surprising that the contribution of the EL method is largest for the broadest cross-section ($K = 49$). A broad cross-section is also beneficial for SD optimization, as it allows for engineering positions with limited downside risk.

The general performance level also improves as we shorten the holding period. However, the improvements seem not enough to warrant the additional portfolio turnover and transaction costs, assuming that every round trip will cost at least tens of basis points. In addition, the use of implied probabilities is less effective, because it is more difficult to predict the out-of-sample values of the factors, for short holding periods. Nevertheless, SD/EL remains the most appealing strategy also for a short horizon.

As we increase the number of risk factors, and hence the number of moment conditions, the performance improvements from using implied probabilities increases. It seems vital to the success of our approach that the factors exhibit low mutual correlation and that the factor exposures show high between-industry variation. In this situation, every additional factor leads to further adjustments of the implied probabilities and the relative returns of the industries.

In contrast to the number of factors, the precise value of the cut points seems of limited importance for investment performance. Using $C = 5$ or $C = 25$ leads to similar results as $C = 10$ in our analysis. Cut points outside the range $[5, 25]$ seem irrelevant at forehand, as they would lead to either minimal and inconsequential deviations from the sample probabilities ($C < 5$) or large and counter-factual adjustments that contradict the observed drift of the risk factors ($C > 25$).

As expected, the investment performance deteriorates as we increase the block length. The adverse effect in our application presumably reflects our combination of diversified base assets, which exhibit relatively weak dynamic patterns, and an intermediate estimation window, which implies that the number of non-overlapping blocks is small compared with the number of base assets. In this situation, BEL introduces excessive variation of the implied probabilities which is not related to the risk factors and which distorts the industry-level drift signal.

[Insert Table III about here.]

4 Conclusions

The plug-in approach hampers the investment performance of optimized portfolios, if the estimation window is short. EL estimation can improve the estimation accuracy by incorporating conditioning information and accounting for dynamic patterns, while preserving the discrete structure and realistic scenarios of the EDF. SD/EL can be implemented with a two-stage procedure using standard mathematical programming methods.

In the application, SD/EL leads to important ex-ante performance improvements, often in excess of a full percentage point per annum, compared with heuristic diversification and optimization based on the MV rule and/or the plug-in method. Heuristic diversification overlooks the risk and dependence structure of the base assets; the MV rule does not recognize the favorable upside potential of concentrated winner portfolios; relying on sample probabilities overlooks the regression effect of common risk factors.

The sensitivity to estimation error remains limited in our analysis, due to the use of diversified base assets and portfolio-weight restrictions. Without those benign features, improving the estimation accuracy seems even more important than in our study.

We used moment conditions for winsorizing common risk factors, in order to account for the regression effect and reduce the adverse effects of factor crashes. These moment conditions are 'loose' in the sense that any reasonable regression model will predict that the factors move closer to the mean than our cut points and, therefore, require stronger divergence from the EDF. Further research could focus on exploring tighter moment conditions for momentum strategies, as well as other investment strategies.

Using the block-wise approach has an adverse effect in our study, because our base assets are diversified industry portfolios, which exhibit relatively weak dynamic patterns, and, furthermore, the number of non-overlapping blocks is small compared with the number of base assets, as a result of using an intermediate estimation window. The BEL seems more appropriate when the base assets are individual securities or dynamic portfolios or when intraday returns are studied.

Our analysis can be generalized in a straightforward way to Generalized Empirical Likelihood (GEL), which allows for 'exponential tilting' and 'continuous updating' estimators. EL is closely related to the Generalized Method of Moments (GMM) and non-parametric Bayesian inference. Not surprisingly, we find very similar results in our application using GMM implied probabilities or BETEL probabilities based on the same moment function.

We hope that our analysis contributes to the portfolio optimization literature by developing a framework to incorporate conditioning information and dynamic patterns in SD optimization and by illustrating its potential profitability using a set of model-free moment conditions; we also hope to contribute to the empirical momentum literature by showing that optimization, when properly combined with decision theory and statistical estimation, can improve significantly upon heuristic investment strategies.

References

[1] Almeida C. and R. Garcia, 2012, Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators, *Journal of Econometrics* 170, 519-537.

[2] Almeida C. and R. Garcia, 2016, Economic Implications of Nonlinear Pricing Kernels, *Management Science*, forthcoming.

[3] Almeida C., K. Ardison, R. Garcia and J. Vicente, 2016, Nonparametric Tail Risk, Stock Returns and the Macroeconomy, *Journal of Financial Econometrics*, forthcoming.

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

[5] Barrett, G. and S. Donald, 2003, Consistent tests for stochastic dominance, *Econometrica* 71, 71-104.

[6] Bawa, V.S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.

[7] Bawa, V.S., J.N. Bodurtha Jr., M.R. Rao and H.L. Suri, 1985, On Determination of Stochastic Dominance Optimal Sets, *Journal of Finance* 40, 417–431.

[8] Board, J.L.G. and C.M.S. Sutcliffe, 1994, Estimation Methods in Portfolio Selection and the Effectiveness of Short Sales Restrictions: UK Evidence, *Management Science* 40, 516–534.

[9] Brown, B. and W. Newey, 2002, Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference, *Journal of Business & Economic Statistics* 20, 507–517.

[10] Carhart, M.M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

[11] Daniel, K. and T. Moskowitz, 2016, Momentum Crashes, forthcoming in *Journal of Financial Economics.*

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

[13] Dentcheva, D. and A. Ruszczynski, 2003, Optimization with Stochastic Dominance Constraints, *SIAM Journal on Optimization* 14, 548-566.

[14] DeMiguel, V., L. Garlappi, F.J. Nogales and R. Uppal, 2009, A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms, *Management Science* 55, 798-812.

[15] DeMiguel, V., L. Garlappi and R. Uppal, 2009, Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?, *Review of Financial Studies* 22, 1915-1953.

[16] Drud, A.S., 1994, CONOPT—A Large-Scale GRG Code, *ORSA Journal on Computing* 6, 207-216.

[17] Fama, E.F. and K.R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33 (1), 3-56.

[18] Frost, P.A. and J.E. Savarino, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 15, 29-34.

[19] Ghosh A., C. Julliard and A. Taylor, 2016, What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models, *Review of Financial Studies*, forthcoming.

[20] Julliard, C. and A. Ghosh, 2012, Can Rare Events Explain the Equity Premium Puzzle?, *Review of Financial Studies* 25, 3037-3076.

[21] Hadar, J. and W.R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2–34.

[22] Hanoch, G. and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335–346.

[23] Hodder, E., J.C. Jackwerth and O. Kolokolova, 2015, Improved Portfolio Choice Using Second-Order Stochastic Dominance, *Review of Finance* 19, 1623-1647.

[24] Jagannathan, R. and T.S. Ma., 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1683.

[25] Kitamura, Yuichi, 1997, Empirical likelihood methods with weakly dependent processes, *The Annals of Statistics* 25, 2084-2102.

[26] Kopa, M. and Th. Post, 2015, A General Test for SSD Portfolio Efficiency, *OR Spectrum* 37, 703-734.

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

[28] Levy, H., 2016, *Stochastic Dominance: Investment Decision Making Under Uncertainty*, Springer, Third Edition.

[29] Linton, O.B., E. Maasoumi and Y.-J. Whang. 2005, Consistent Testing for Stochastic Dominance under General Sampling Schemes, *Review of Economic Studies* 72, 735-765.

[30] Moskowitz, T.J. and M. Grinblatt, 1999, Do Industries Explain Momentum?, *Journal of Finance* 54, 1249-90.

[31] Owen, A.B., 2001, *Empirical Likelihood*, New York: CRC Press.

[32] Post, Th., 2017, Empirical Tests for Stochastic Dominance Optimality, forthcoming in *Review of Finance*.

[33] Post, Th. and V. Potì, 2017, Portfolio Analysis using Stochastic Dominance, Relative Entropy and Empirical Likelihood, *Management Science* 63, 153-165.

[34] Rockafellar, R. T. and S. Uryasev, 2000, Optimization of Conditional Value-at-Risk, *Journal of Risk* 2, 21-41.

[35] Roman, D., Darby-Dowman, K. and Mitra, G., 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.

[36] Roman, D., G. Mitra and V. Zverovich, 2013, Enhanced indexation based on second-order stochastic dominance, *European Journal of Operational Research* 228, 273-281.

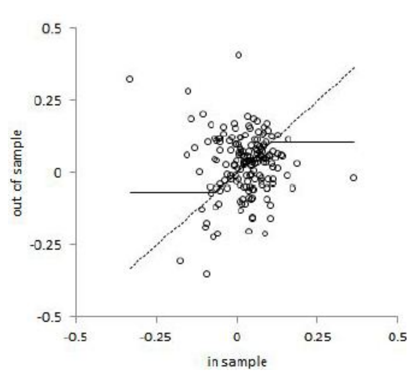
[37] Rothschild, M. and J. E. Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.

[38] Russell, W. R. and T. K. Seo, 1989, Representative sets for stochastic dominance rules, in: TB Fomby and TK Seo (eds.), Studies in the economics of uncertainty: In honor of Josef Hadar, New York: Springer Verlag, 59-76.

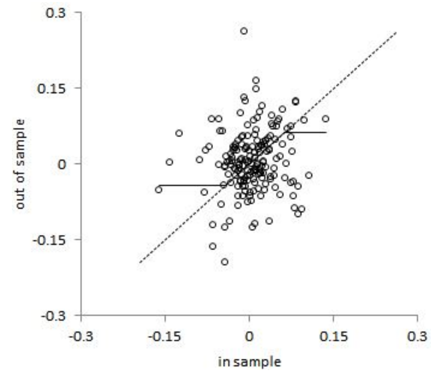
[39] Scaillet, O., and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169-180.

[40] Schennach, S. M., 2005, Bayesian exponentially tilted empirical likelihood, *Biometrika* 92, 31-46.

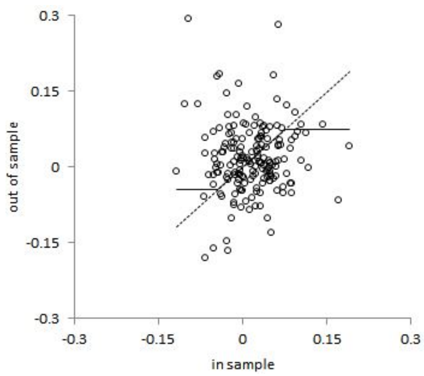
[41] Shalit, H. and S. Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670–684.



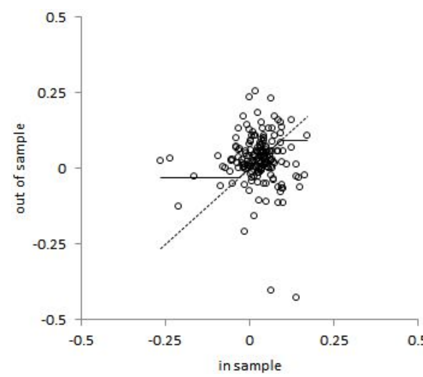
Factor 1: RMRF



Factor 2: SMB



Factor 3: HML



Factor 4: MOM

Figure 1: Regression Effects and Moment Conditions

Each panel plots the in-sample mean against the out-of-sample mean for one of the four factors of Carhart (1997): 'RMRF', 'SMB', 'HML' and 'MOM'. The in-sample mean is the sample mean of the daily values in a 12-month estimation window; the out-of-sample mean is the sample mean in the subsequent 6-month holding period. The first estimation window is from January 1927 to December 1927 and the first holding period is from January 1928 to June 1928; last estimation window is from July 2014 to June 2015 and the last holding period is from July 2015 to December 2015. The dotted 45° line represents historical extrapolation; it equates the in-sample and out-of-sample means. The solid, piecewise line represents our moment conditions (9). The lower cut point L_j, $j = 1, \cdots, 4$, equals the tenth percentile of the distribution of the sample mean across all estimation windows; the upper cut point U_j, $j = 1, \cdots, 4$, equals the 90-th percentile.

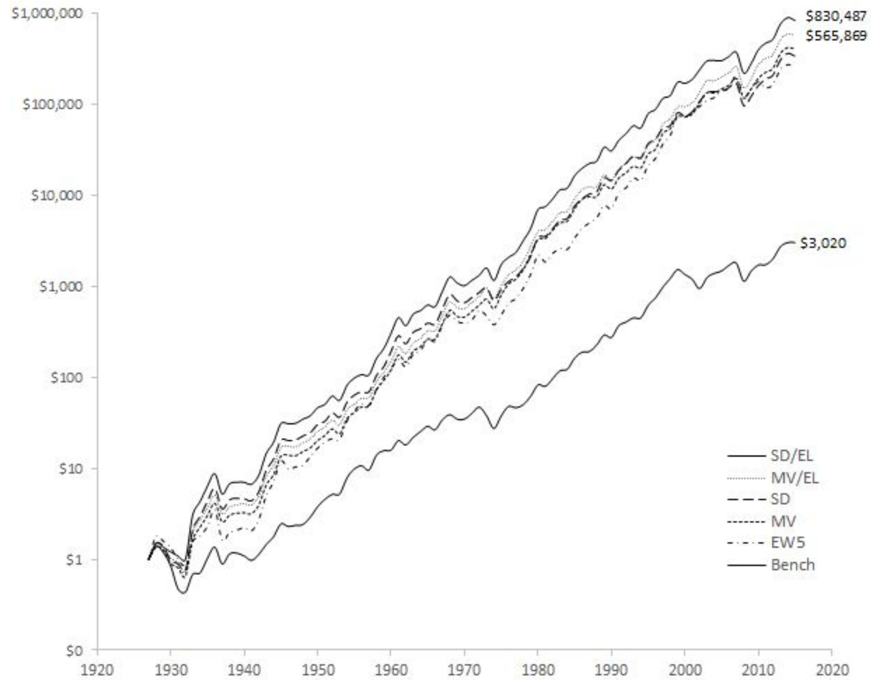


Figure 2: Cumulative Performance

This figure shows the development of the dollar value of six portfolios over the entire sample period from 1928 through 2015: the passive benchmark ('Bench'); the heuristic portfolio ('EW5'); the MV portfolio based on plug-in estimates; the MV/EL portfolio; the SD portfolio based on sample probabilities; the SD/EL portfolio. The initial investment in January 1928 in every portfolio is one dollar. The portfolios are formed and rebalanced at the beginning of a 6-month holding period based on a trailing 12-month estimation window of daily returns. The first estimation window is January 1927 - December 1927 and the first holding period is January 1928 - June 1928; the last estimation window is July 2014 - June 2015 and the last holding period is July 2015 - December 2015. The implied probabilities for MV/EL and SD/EL are based on the moment conditions (9); we use the $F = 4$ factors of Carhart (1997); the lower cut point L_j, $j = 1, \cdots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper cut point U_j, $j = 1, \cdots, 4$, equals the 90-th percentile. We use the IID EL method, or block size $B = 1$. The graph uses a logarithmic scale.

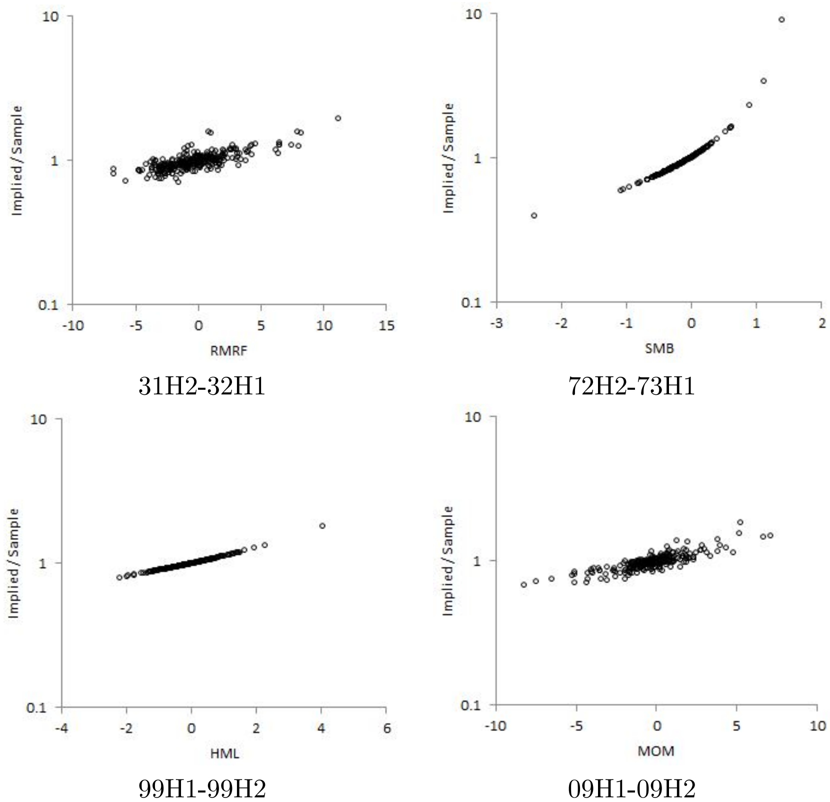


Figure 3: Close-up of Implied Probabilities during Factor Crashes

Each panel shows the implied probabilities p_t, $t = 1, \cdots, T$, for a 12-month estimation window during which one of the risk factors achieves its lowest sample mean across all estimation windows. The implied probabilities are based on the moment conditions (9); we use the $F = 4$ factors of Carhart (1997); the lower cut point L_j, $j = 1, \cdots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper cut point U_j, $j = 1, \cdots, 4$, equals the 90-th percentile. The RMRF factor crashes in the period July 1932 - June 1932 ($T = 301$); SMB in July 1972 - June 1973 ($T = 249$); HML in January 1999 - December 1999 ($T = 252$); MOM in January 2009 - December 2009 ($T = 252$). We use the IID EL method, or block size $B = 1$. The implied probabilities are normalized by dividing by the sample probability T^{-1}. The graph uses a logarithmic scale.

Table I: Performance of Base Case Industry Momentum Strategy

The table summarizes the out-of-sample performance of several portfolios. The passive benchmark ('Bench') is the CRSP all-share index. The heuristic portfolio ('EW5') is the equal-weighted average of the 5 industries with the highest realized return. The MV portfolio maximizes the mean subject to a lower variance than the benchmark using plug-in estimates; the MV/EL portfolio solves the same problem under implied probabilities. The enhanced portfolios are formed at the beginning of a 6-month holding period based on a trailing 12-month estimation window of daily returns. The first estimation window is from January 1927 to December 1927 and the first holding period is from January 1928 to June 1928; the last estimation window is from July 2014 to June 2015 and the last holding period is from July 2015 to December 2015. The implied probabilities for MV/EL and SD/EL are based on the moment conditions (9); we use the $F = 4$ factors of Carhart (1997); the lower cut point L_j, $j = 1, \cdots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper cut point U_j, $j = 1, \cdots, 4$, equals the 90-th percentile. We use the IID EL method, or block size $B = 1$. We evaluate the annual percentage out-of-sample returns in $N = 88$ years from 1928 through 2015, using returns from January to December. Shown are the mean, standard deviation, semi-deviation and skewness of annual returns as well as the Value-at-Risk and Conditional Value-at-Risk for a 95% confidence level (VaR95 and CVaR95) and the Certainty Equivalent Return (CER) for log utility $u(x) = \ln(100 + x)$. Also shown are the Sharpe ratio, Sortino ratio and, finally, the t-statistic for testing whether the average return spread equals zero. We assume a riskless rate of 3.40%, for the computation of the semi-variance, Sharpe ratio and Sortino ratio.

	Bench	EW5	MV	MV/EL	SD	SD/EL	SD/EL -Bench	SD/EL -EW5	SD/EL -MV	SD/EL -MV/EL	SD/EL -SD
mean	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.55	1.00	1.36
st. dev.	20.27	26.24	25.55	26.57	28.35	30.54	10.27	4.29	4.99	3.96	2.19
semi-dev.	18.29	21.12	17.50	17.59	18.66	17.59	-0.70	-3.52	0.09	0.00	-1.07
skew.	-0.40	0.15	1.36	1.79	1.96	2.81	3.20	2.66	1.44	1.02	0.85
VaR95	28.49	18.93	20.79	19.29	23.39	16.26	-12.23	-2.67	-4.53	-3.03	-7.13
CVaR95	35.28	33.49	30.59	29.36	33.69	29.99	-5.30	-3.50	-0.60	0.63	-3.71
CER	9.53	15.20	15.79	16.24	15.56	16.75	7.22	1.55	0.96	0.51	1.19
Sharpe	0.40	0.57	0.59	0.58	0.53	0.54					
Sortino	0.44	0.70	0.85	0.88	0.81	0.94					
t-stat							3.73	0.84	1.63	1.22	2.65

Table II: Close-up of Returns After Factor Crashes

The table summarizes the out-of-sample returns of several portfolios following severe 'factor crashes'. The passive benchmark ('Bench') is the CRSP all-share index. The heuristic portfolio ('EW5') is the equal-weighted average of the 5 industries with the highest realized return. The MV portfolio maximizes the mean subject to a lower variance than the benchmark using plug-in estimates; the MV/EL portfolio solves the same problem under implied probabilities. Similarly, the SD portfolio is formed maximizes the mean subject to stochastic dominance of the benchmark using sample probabilities; the SD/EL portfolio solves the same problem under implied probabilities. The enhanced portfolios are formed at the beginning of a 6-month holding period based on a trailing 12-month estimation window of daily returns. The implied probabilities for MV/EL and SD/EL are based on the moment conditions (9); we use the $F = 4$ factors of Carhart (1997); the lower cut point L_j, $j = 1, \cdots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper cut point U_j, $j = 1, \cdots, 4$, equals the 90-th percentile. We use the IID EL method, or block size $B = 1$. For each of the four factors, the table shows the sample mean and implied mean in the estimation window; also shown are the sample mean of the factor and the investment returns to the six portfolios in the subsequent holding period.

	Estimation Period	Sample Mean	Implied Mean	Holding Period	Sample Mean	Bench	EW5	MV	MV/EL	SD	SD/EL
RMRF	31H2-32H1	-0.331	-0.065	32H2	0.322	52.52	22.37	14.30	27.68	21.22	32.48
SMB	72H2-73H1	-0.160	-0.043	73H2	-0.052	-4.58	14.55	25.08	32.07	25.38	34.45
HML	99H1-99H2	-0.118	-0.045	00H1	-0.009	-1.32	11.84	18.25	19.32	19.69	19.94
MOM	09H1-09H2	-0.263	-0.038	10H1	0.023	-5.74	-8.24	7.14	17.02	6.92	15.95
Mean						10.22	10.13	16.19	24.02	18.30	25.71

Table III: Alternative Specifications

Shown are the out-of-sample means for various portfolios formed from the $K = 10, 17, 49$ base assets: the passive benchmark ('Bench'); the heuristic portfolio ('EW5'); the MV portfolio based on plug-in estimates; the MV portfolio based on the implied distribution ('MV/EL'); the SD portfolio based on sample probabilities; the SD portfolio based on the implied distribution ('SD/EL'). We consider holding periods of $H = 3, 6, 12$ months of daily returns. The implied probabilities for MV/EL and SD/EL are based on the moment conditions (9). We use $F = 1, 3, 4$ factors: the market model ($F = 1$); the Fama and French (1993) three-factor model ($F = 3$) and the Carhart (1997) four-factor model ($F = 4$). The lower cut point L_j, $j = 1, \cdots, 4$, equals the C-th percentile of the historical distribution of the factor's sample mean across all estimation windows; the upper cut point U_j, $j = 1, \cdots, 4$, equals the $(100-C)$-th percentile. We consider $C = 5, 10, 25$. We use a BEL with maximally overlapping blocks of $B = 1, 2, 3$ consecutive observations. We evaluate annual percentage out-of-sample returns to the portfolios in $N = 88$ years from 1928 through 2015, using returns from January to December. To reduce the output, we vary only one parameter at the time, using the strategy with $K = 49$, $H = 6$, $F = 4$, $C = 10$ and $B = 1$ as the base case.

K	H	F	C	B	Bench	EW5	MV	MV/EL	SD	SD/EL	SD/EL -Bench	SD/EL -EW5	SD/EL -MV	SD/EL -MV/EL	SD/EL -SD
10	6	4	10	1	11.54	13.22	13.50	13.80	13.39	13.64	2.10	0.42	0.14	-0.16	0.25
17	6	4	10	1	11.54	14.19	14.11	14.54	14.62	15.23	3.69	1.04	1.12	0.69	0.61
49	6	4	10	1	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.56	1.00	1.36
49	3	4	10	1	11.54	19.26	19.37	19.79	20.02	20.33	8.79	1.07	0.96	0.54	0.31
49	6	4	10	1	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.56	1.00	1.36
49	12	4	10	1	11.54	17.14	16.74	17.17	17.84	18.67	7.13	1.53	1.93	1.50	0.83
49	6	1	10	1	11.54	18.24	18.36	18.25	18.56	18.74	7.20	0.50	0.38	0.49	0.18
49	6	3	10	1	11.54	18.24	18.36	18.60	18.56	19.38	7.84	1.14	1.02	0.78	0.82
49	6	4	10	1	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.56	1.00	1.36
49	6	4	5	1	11.54	18.24	18.36	18.87	18.56	19.66	8.12	1.42	1.30	0.79	1.10
49	6	4	10	1	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.56	1.00	1.36
49	6	4	25	1	11.54	18.24	18.36	19.17	18.56	19.73	8.19	1.49	1.37	0.56	1.17
49	6	4	10	1	11.54	18.24	18.36	18.92	18.56	19.92	8.38	1.68	1.56	1.00	1.36
49	6	4	10	2	11.54	18.24	18.36	18.64	18.56	19.28	7.74	1.04	0.92	0.64	0.72
49	6	4	10	3	11.54	18.24	18.36	18.16	18.56	18.78	7.25	0.54	0.42	0.62	0.22