Portfolio Optimization based on Stochastic Dominance and Empirical Likelihood*

Thierry Post[†], Selçuk Karabatı[‡] & Stelios Arvanitis[§]

September 26, 2017

Abstract

This study develops 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 multivariate return distributions. The SD/EL method can be implemented using a two-stage procedure which first elicits the implied probabilities using Convex Optimization and subsequently constructs the optimal portfolio using Linear Programming. The solution asymptotically dominates the benchmark and optimizes the goal function in probability, for a class of weakly dependent processes. A Monte Carlo simulation experiment illustrates the improvement in estimation precision using a set of conservative moment conditions about common factors in small samples. In an application to equity industry momentum strategies, SD/EL yields important out-of-sample performance improvements relative to heuristic diversification, Mean-Variance optimization, and a simple ‘plug-in’ approach.

Key words: Stochastic Dominance, Empirical Likelihood, Portfolio optimization, 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 seminars at the Fuqua School of Business at Duke University (14 December 2016), Athens University of Economics and Business (30 March 2017) and the Research Center of Quantitative Economics at Jilin University (7 July 2017). Any remaining errors are our own.

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

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

[§]Arvanitis is Assistant Professor at the Department of Economics of Athens University of Economics and Business; Athens; Greece; email: stelios@aueb.gr.

1 Introduction

Stochastic Dominance (SD) is a time-honored maxim for investment decision making which avoids a parametric specification of the joint return distribution (Hadar and Russell (1969); Hanoch and Levy (1969); Rothschild and Stiglitz (1970); Bawa (1975); Levy (2016)).

The appeal of SD is particularly strong for asset classes and investment strategies with asymmetric risk profiles such as small-cap stocks, junk bonds and momentum strategies. There exists no accepted parametric family of multivariate distributions for these cases; in addition, portfolio analysis is analytically complex for plausible specifications such as the multivariate log-normal or multivariate GARCH models.

For the common second-order SD (SSD) criterion and a discrete return distribution, constructing a portfolio which dominates the benchmark amounts to a Linear Programming (LP) problem (Shalit and Yitzhaki (1994); Dentcheva and Ruszczyński (2003); Kuosmanen (2004); Roman, Darby-Dowman and Mitra (2006)). The optimization problem is much larger than the corresponding Mean-Variance (MV) problem, but it remains tractable for realistic applications (Roman, Mitra and Zverovich (2013); Hodder, Jackwerth and Kolokolova (2015)).

An issue which has received limited attention in this literature is the statistical estimation of the joint return distribution. Estimation precision is particularly relevant for portfolio optimization, because the optimal portfolio weights can be very sensitive to estimation error and optimized portfolios can therefore 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 (Deshpande and Singh (1985); McFadden (1989); Schmid and Trede (1998); Anderson (1996); Davidson and Duclos (2000); Barrett and Donald (2003); Linton, Maasoumi and Whang (2005); Scaillet and Topaloglou (2010); Linton, Post and Whang (2014)). These studies unfortunately offer limited guidance for constructing a dominating portfolio from a set of base assets.

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 side information and becomes inaccurate if the available time series is short and the number of base assets is large. The EDF also does not account for common dynamic patterns such as price reversals and volatility clusters in high-frequency data.

The present study proposes a framework for SD optimization based on the non-parametric Empirical Likelihood (EL; [Owen \(1988, 2001\)](#)) method. EL generates a discrete estimator for the joint distribution, henceforth the ‘implied distribution function’ (IDF), by minimizing the divergence from the ‘implied probabilities’ to the sample probabilities subject to a set of moment conditions.

[Qin and Lawless \(1994\)](#) establish desirable asymptotic properties of this estimator. The ‘EL bootstrap’ ([Brown and Newey \(2002\)](#)) uses the IDF to simulate the null distribution. The ‘Bayesian Exponentially Tilted Empirical Likelihood’ (BETEL; [Schennach \(2005\)](#)) uses the implied probabilities as a nonparametric likelihood function to compute Bayesian posterior probabilities.

In the asset pricing literature, EL and related estimation methods have been proposed by [Almeida and Garcia \(2012, 2016\)](#), [Julliard and Ghosh \(2012\)](#), [Almeida, Ardison, Garcia and Vicente \(2016\)](#) and [Ghosh, Julliard and Taylor \(2016\)](#). Although the misspecification tests have a dual formulation as a specific portfolio problem, they do not produce a portfolio which stochastically dominates the benchmark.

EL combines well with SD, due to a shared distribution-free assumption framework and the discrete nature of the IDF. The combined SD/EL method can be implemented using a tractable two-stage procedure. A blockwise implementation of EL can account for dynamic patterns in a flexible way ([Kitamura \(1997\)](#)), in sufficiently large samples.

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 given portfolio is stochastically dominated, without identifying a superior alternative. By contrast, the present study uses the IDF in the optimization of a dominating portfolio.

In related work, [Rojo-Jiménez and El Barmi \(2003\)](#) and [El Barmi and Marchev \(2009\)](#) propose alternative nonparametric estimators under SSD constraints for a given pair of prospects. Their approach could be used to estimate the distribution of a given portfolio which is known to dominate the benchmark. However, the approach seems less relevant when the optimal dominating portfolio is not known and a general multivariate distribution is analyzed.

In this study, desirable large-sample properties are established for the SD/EL solution portfolio for a class of weakly dependent processes. The limit theory also provides theoretical guidance for the proper specification of the moment conditions.

Unfortunately, a formalization of the finite-sample properties seems elusive

due to the distribution-free assumption framework. The small-sample behavior of the IDF is therefore examined through a Monte Carlo simulation experiment. In addition, the out-of-sample investment performance of the solution portfolio is studied in a representative application, as a historical simulation of the economic goodness of the IDF.

The application studies a historical data set of daily returns to equity industry portfolios, following [Hodder, Jackwerth and Kolokolova \(2015\)](#). Since the portfolios are formed based on an intermediate estimation window and rebalanced after a short-to-intermediate holding period, the investment strategies in effect exploit known industry momentum effects ([Moskowitz and Grinblatt \(1999\)](#)).

In this application, the data dimensions seem unfavorable for the plug-in approach and side information seems a useful addition to the data set. A blockwise implementation however seems too ambitious, because the intermediate estimation window contains an insufficient number of independent data blocks.

2 Methodology

This section presents the SD/EL method. For the sake of brevity, attention is focused on the common SSD rule. In this case, SD/EL can be implemented using a tractable two-stage procedure which first elicits the IDF using standard Convex Optimization and subsequently constructs the optimal portfolio given the IDF using LP.

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 based on SSD: FSD requires Mixed Integer Linear Programming (see [Kopa and Post \(2009\)](#)) and TSD requires Quadratic Programming (see [Post and Kopa \(2017\)](#)). However, the probability elicitation is based on a separate optimization problem which does not affect the complexity of the portfolio optimization problem.

2.1 Preliminaries

Consider K distinct base assets with random investment returns $\mathbf{x} \in \mathcal{X}^K$ with joint Cumulative Distribution Function (DF) $\mathcal{F}(\mathbf{x})$, $\mathcal{X} := [a, b]$, $-\infty < a < b < +\infty$, and means $\boldsymbol{\mu} := \mathbb{E}_{\mathcal{F}}[\mathbf{x}]$. The portfolio possibility set consists of all convex combinations of the base assets: $\Lambda := \left\{ \boldsymbol{\lambda} \in \mathbb{R}^K : \boldsymbol{\lambda}^T \mathbf{1}_K = 1; \boldsymbol{\lambda} \geq \mathbf{0}_K \right\}$.

The research objective is to enhance a given benchmark portfolio which is characterized by weights $\boldsymbol{\tau} \in \Lambda$ and returns $y := \boldsymbol{x}^T \boldsymbol{\tau}$.

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 factor loadings. To allow for dynamic intertemporal investment problems, these combinations may be periodically rebalanced based on a conditioning information set.

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

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

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

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

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 LPM 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*

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

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 $u(x)$. This equivalence can be established using elementary [Russell and Seo \(1989\)](#) ramp functions $v_z(x) := (x - z) \mathcal{I}(x \leq z)$, $z \in \mathcal{X}$. Using $\mathcal{L}_{\boldsymbol{\lambda}}(z) = -\mathbb{E}_{\mathcal{F}_{\boldsymbol{\lambda}}} [v_z(x)]$, (3) amounts to $\mathbb{E}_{\mathcal{F}_{\boldsymbol{\lambda}}} [v_z(x)] \geq \mathbb{E}_{\mathcal{F}_{\boldsymbol{\tau}}} [v_z(x)]$. The condition is sufficient because every admissible utility function is a positive mixture of elementary functions.

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 challenging 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 Section 2.3 shows.

2.2 Stage One: Probability Elicitation

The DF is latent and has to be estimated using a time series of realized return vectors, $\mathbf{X}_s := (X_{1,s} \cdots X_{K,s})^\top$, $s = 1, \dots, T$. The EDF amounts to

$$\mathcal{F}_T(\mathbf{x}) := T^{-1} \sum_{t=1}^T \mathcal{I}(\mathbf{X}_t \leq \mathbf{x}). \quad (4)$$

This estimator assigns an equal probability to each vector, or $\mathbb{P}_{\mathcal{F}_T}[\mathbf{x} = \mathbf{X}_s] = T^{-1}$, $s = 1, \dots, T$. Existing empirical applications of SD optimization generally employ this ‘plug-in’ estimator (Roman, Mitra and Zverovich (2013); Hodder, Jackwerth and Kolokolova (2015)).

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, the EDF is a statistically consistent, nonparametric Maximum Likelihood estimator, under appropriate stationarity and dependence assumptions. Unfortunately, the plug-in approach can become inaccurate if the time series is short and/or side information on the underlying distribution is available.

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

$$\mathbb{E}_{\mathcal{F}}[\mathbf{g}(\mathbf{x})] = \mathbf{0}_M. \quad (5)$$

The moment function can be used to capture side information. A possible source of information are theoretical equilibrium relations such as the standard orthogonality condition $\mathbb{E}_{\mathcal{F}}[k(\mathbf{x})\mathbf{x}] = \mathbf{0}_M$ based on a pricing kernel function $k(\mathbf{x})$. Alternatively, the specification of $\mathbf{g}(\mathbf{x})$ may be based on empirical stylized

facts about the risk profiles and risk premiums of various asset classes and investment styles.

The function generally depends on latent model parameters and/or nuisance parameters that are estimated as part of the procedure, but those parameters are suppressed here to simplify the notation. Conditioning variables and instrumental variables may also be included. 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 $f_j := \mathbf{x}^T \mathbf{w}_j$, $j = 1, \dots, F$. This approach reduces the number of dimensions and can exploit the existing body of empirical knowledge about factors for various asset classes.

The use of factor models for improving upon plug-in estimates is well-established for MV optimization. In the present SD/EL framework, prior information is used to estimate the entire DF rather than mean and covariance parameters only.

Although EL is more commonly used for testing of the moment conditions and inference about the latent model parameters, the present study uses the IDF to improve upon the EDF as an estimator for the DF, assuming that the moment conditions are correctly specified. For this reason, it is recommended to use conservative, uncontroversial moment conditions which do not require explicit testing.

The distribution function is estimated using a multinomial distribution with atoms at the data points. The ‘implied probabilities’ $p_s := \hat{\mathbb{P}}[\mathbf{x} = \mathbf{X}_s] > 0$, $s = 1, \dots, T$, are found by minimizing the Relative Entropy, or the Kullback-Leibler divergence to the sample probabilities $\pi_s := \mathbb{P}_{\mathcal{F}_T}[\mathbf{x} = \mathbf{X}_s] = T^{-1}$, $s = 1, \dots, T$,

$$\begin{aligned} D_{KL}(\{\pi_t\} \parallel \{p_t\}) : &= \sum_{t=1}^T \pi_t \ln \left(\frac{\pi_t}{p_t} \right) \\ &= -\frac{1}{T} \sum_{t=1}^T \ln(p_t) - \ln(T), \end{aligned} \tag{6}$$

subject to the empirical equivalent of the moment conditions,

$$\sum_{t=1}^T p_t \mathbf{g}(\mathbf{x}_t) = \mathbf{0}_M. \quad (7)$$

Blockwise Empirical Likelihood (BEL; [Kitamura \(1997\)](#)) applies this approach to data blocks rather than individual observations, to account for dynamic patterns. The blockwise implementation subdivides the original time series into $T^* := (T - B + 1)$ maximally overlapping blocks of B consecutive observations.

The optimal block size depends on the context and involves a trade-off between the strength of the dynamic effects and the number of independent blocks, or $\lfloor T/B \rfloor$. It is generally recommended that the block size grows with T but at a lower rate, for example, $(B^{-1} + B^2/T) \rightarrow 0$ as $T \rightarrow \infty$. In practice, the block size can be selected based on prior knowledge about dynamic patterns, estimation of the dependence structure from the data or analyzing the effect on the out-of-sample returns of the SD/EL portfolio.

The BEL procedure assigns probabilities $b_1, \dots, b_{T^*}$ to the data blocks and equal (conditional) probabilities (B^{-1}) to the observations in a given block; the procedure then seeks to minimize $D_{KL}(\{\beta_i\} \parallel \{b_i\})$, $\beta_i := (T^*)^{-1}$, subject to block-level moment conditions $\sum_{i=1}^{T^*} b_i \mathbf{g}_i^* = \mathbf{0}_M$, $\mathbf{g}_i^* := B^{-1} \sum_{j=1}^B \mathbf{g}(\mathbf{x}_{(i-1)B+j})$.

Observation t , $t = 1, \dots, T$, is included in all blocks with indices from $t^- := \max(1, t - B + 1)$ to $t^+ := \min(t, T^*)$. The observation-level probabilities can therefore be recovered from the block-level probabilities in the following way:

$$p_t = B^{-1} \sum_{i=t^-}^{t^+} b_i, \quad t = 1, \dots, T. \quad (8)$$

The block-level probabilities $b_1, \dots, b_{T^*}$ are the solution to the following optimization problem:

$$\begin{aligned}
& \min \left(-\frac{1}{T^*} \sum_{i=1}^{T^*} \ln(b_i) - \ln(T^*) \right) \\
& \text{s.t. } B^{-1} \sum_{i=1}^{T^*} \sum_{j=1}^B b_i \mathbf{g}(\mathbf{x}_{(i-1)B+j}) = \mathbf{0}_M; \\
& \sum_{i=1}^{T^*} b_i = 1; \\
& b_i \geq 0, i = 1, \dots, T^*.
\end{aligned} \tag{9}$$

This problem is a standard Convex Optimization problem, and, for the sample size in this study ($T \approx 250$ trading days in a moving 12-month estimation window), the problem size is small.

The implied probability is a histogram estimator of the joint density. The DF $\mathcal{F}(\mathbf{x})$ is estimated by the following IDF:

$$\hat{\mathcal{F}}(\mathbf{x}) := \sum_{t=1}^T p_t \mathcal{I}(\mathbf{X}_t \leq \mathbf{x}). \tag{10}$$

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

$$\hat{\mathcal{F}}_{\boldsymbol{\lambda}}(z) = \sum_{t=1}^T p_t \mathcal{I}(\mathbf{X}_t^T \boldsymbol{\lambda} \leq z). \tag{11}$$

The IDF can be seen as a constrained nonparametric Maximum Likelihood estimator. [Qin and Lawless \(1994\)](#) establish statistical consistency, asymptotic normality and asymptotic efficiency gains from well-specified moment conditions, under suitable regularity conditions. Section 3 extends their limit theory for the IDF from the original IID framework to a class of stationary and weakly dependent processes, using results from [Kitamura \(1997\)](#).

The proper specification of the moment function is essential for the success of the SD/EL method. Two general requirements for $\mathbf{g}(\mathbf{x})$ are that its population moments are unbiased and its empirical moments are correlated with the empirical LPMs of the base assets. Implementing these guidelines generally requires a good understanding of the relevant base assets and investment strategy.

The present study employs a model-free specification based on lower bounds and upper bounds for the means of a set of common factors:

$$L_j \leq \mathbb{E}_{\mathcal{F}}[f_j] \leq U_j, j = 1, \dots, F. \quad (12)$$

In other words, $M = 2F$ moment inequalities are used and the components of the moment function are given by

$$\begin{aligned} g_{2j-1}(\mathbf{x}) &= f_j - s_j^- - L_j, j = 1, \dots, F; \\ g_{2j}(\mathbf{x}) &= f_j + s_j^+ - U_j, j = 1, \dots, F, \end{aligned} \quad (13)$$

where $s_j^-, s_j^+ \geq 0$ are latent slack variables.

Imposing these moment conditions amounts to ‘winsorizing’ the sample means of the common factors. This approach seems particularly useful to avoid that the estimated distribution of the base assets is dominated by extreme factor moves during the sample period. The simulation and application sections elaborate further on this specification.

2.3 Stage Two: Portfolio Optimization

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

Furthermore, the estimated LPM takes a simple piecewise-linear, increasing and convex shape:

$$\begin{aligned} \hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) &= \mathbb{E}_{\hat{\mathcal{F}}_{\boldsymbol{\lambda}}} [(z - x)\mathcal{I}(x \leq z)] \\ &= \sum_{t=1}^T p_t (z - \mathbf{X}_t^T \boldsymbol{\lambda}) \mathcal{I}(\mathbf{X}_t^T \boldsymbol{\lambda} \leq z). \end{aligned} \quad (14)$$

As a consequence, it suffices to check the dominance condition only at the observed return levels for the benchmark ($z = y_s, s = 1, \dots, T$), 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 IDF, or $\boldsymbol{\lambda} \succeq_{\hat{\mathcal{F}}} \boldsymbol{\tau}$, if and only if*

$$\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_s) \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(y_s), s = 1, \dots, T. \quad (15)$$

(Without further proof.)

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

$$\begin{aligned} \hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) &= \min_{\boldsymbol{\theta}} \sum_{t=1}^T p_t \theta_t \\ \theta_t &\geq z - \mathbf{X}_t^T \boldsymbol{\lambda}, t = 1, \dots, T; \\ \theta_t &\geq 0, t = 1, \dots, T. \end{aligned} \quad (16)$$

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

It follows from (16) that the nonlinear inequality $\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(z) \leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(z)$ can be reformulated using the following linear system:

$$\begin{aligned} \sum_{t=1}^T p_t \theta_t &\leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(z); \\ \theta_t &\geq z - \mathbf{X}_t^T \boldsymbol{\lambda}, t = 1, \dots, T; \\ \theta_t &\geq 0, t = 1, \dots, T. \end{aligned} \quad (17)$$

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

The entire nonlinear system (15) can be captured by applying the linear system (17) to every threshold $z = y_s$, $s = 1, \dots, T$. For this purpose, the model variables $\theta_{s,t}$, $s, t = 1, \dots, T$, are introduced, to capture the terms $(y_s - \mathbf{X}_t^T \boldsymbol{\lambda}) \mathcal{I}(\mathbf{X}_t^T \boldsymbol{\lambda} \leq y_s)$, $s, t = 1, \dots, T$. In addition, the analysis treats the portfolio weights $\boldsymbol{\lambda} \in \Lambda$ as model variables, which does not introduce further complications, because the constraints of problem (16) are linear in the

portfolio weights.

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

$$\begin{aligned}
\sum_{t=1}^T p_t \theta_{s,t} &\leq \hat{\mathcal{L}}_{\boldsymbol{\tau}}(y_s), \quad s = 1, \dots, T; \\
-\theta_{s,t} - \mathbf{X}_t^T \boldsymbol{\lambda} &\leq -y_s, \quad s, t = 1, \dots, T; \\
\mathbf{1}_K^T \boldsymbol{\lambda} &= 1; \\
\theta_{s,t} &\geq 0, \quad s, t = 1, \dots, T; \\
\lambda_k &\geq 0, \quad k = 1, \dots, K.
\end{aligned} \tag{18}$$

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

Using Farkas' lemma, an alternative system of linear constraints can be obtained which can be interpreted in terms of the coefficients of a piece-wise linear utility function (Post (2003), Kopa and Post (2015)).

This alternative formulation is particularly interesting because the marginal utility function represents a possible pricing kernel for the base assets, if the benchmark portfolio is efficient. For this reason, SD efficiency tests typically use this formulation. However, in the present application, it is assumed that the benchmark portfolio is inefficient and the focus is on the composition of superior portfolios.

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

$$G_{\hat{\mathcal{F}}, \boldsymbol{\lambda}} := \mathbb{E}_{\hat{\mathcal{F}}}[\mathbf{x}]^T \boldsymbol{\lambda} - \sum_{t=1}^T w_t \hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_t), \tag{19}$$

where $\mathbb{E}_{\hat{\mathcal{F}}}[\mathbf{x}] = \sum_{t=1}^T p_t \mathbf{X}_t$ is the mean vector and $w_t \geq 0$, $t = 1, \dots, T$, are decision weights. Goal function (19) can be linearized by replacing $\hat{\mathcal{L}}_{\boldsymbol{\lambda}}(y_t)$ with $\sum_{s=1}^T p_s \theta_{t,s}$, $t = 1, \dots, T$, due to the aforementioned Rockafellar-Uryasev relaxation, resulting in an LP problem. Although the optimization problem is much larger than the corresponding MV problem, it is perfectly manageable with standard computer hardware and optimization software in typical applications.

The objective in this study is to maximize the expected portfolio return, or the extreme case with $w_t = 0$, $t = 1, \dots, T$, and $G_{\hat{\mathcal{F}}, \lambda} = \mathbb{E}_{\hat{\mathcal{F}}}[\mathbf{x}]^T \lambda$. This orientation allows for a direct comparison with heuristic portfolio construction rules and with MV optimization based on the same goal 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, alternatively, using multi-objective programming with risk reduction as the secondary goal function.

Another example of a goal 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.

Compared with the existing plug-in approach, SD/EL changes the optimization problem by replacing the equal probabilities with the EL implied probabilities $p_s := \hat{\mathbb{P}}[\mathbf{x} = \mathbf{X}_s]$, $s = 1, \dots, T$. This replacement in turn affects the terms $\hat{\mathcal{L}}_{\boldsymbol{\tau}}(y_s) = \sum_{t=1}^T p_t(y_s - \mathbf{X}_t^T \boldsymbol{\tau}) \mathcal{I}(\mathbf{X}_t^T \boldsymbol{\tau} \leq y_s)$, $s = 1, \dots, T$, in the model constraints (18) and the terms $\mathbb{E}_{\hat{\mathcal{F}}}[\mathbf{x}]^T \lambda = \sum_{t=1}^T p_t \mathbf{X}_t^T \lambda$ and $\hat{\mathcal{L}}_{\lambda}(y_s) = \sum_{t=1}^T p_t(y_s - \mathbf{X}_t^T \lambda) \mathcal{I}(\mathbf{X}_t^T \lambda \leq y_s)$, in the objective function (19).

2.4 Hardware and software

In the below application of SD/EL, the probability estimation problem (9) is modeled in GAMS and solved with CONOPT nonlinear solver ([Drud \(1994\)](#)) on a computer with an Intel Core i7 2.38GHz processor and 8GB memory.

The portfolio optimization problem (maximize objective (19) subject to linear system (18)) is modeled in and solved with IBM ILOG CPLEX Optimization Studio 12.6.3.0 on a computer with 2 x Intel(R) Xeon(R) CPU E5-2695 2.40GHz processors and 512GB memory. This software automatically recognizes problem reduction possibilities in its pre-solution phase, and selects the most effective combination of options for the solution phase.

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 Large-sample Properties

A limit theory for the SD/EL procedure is outlined, building on [Qin and Lawless \(1994\)](#) and [Kitamura \(1997\)](#). The asymptotic properties of the IDF estimator are derived for a class of weakly dependent processes. The first-order limit theory for possible decision errors of the procedure is established using a notion of imperfect mutual dependence. The analysis studies the asymptotic efficiency gains from including side information and, furthermore, provides theoretical guidance for the choice of the moment function. The appendix contains the formal proofs of our results.

3.1 Additional notation and assumptions

Let $\mathcal{D}_\lambda(z) := \mathcal{L}_\tau(z) - \mathcal{L}_\lambda(z)$ be the ‘risk reduction’ for a given threshold and $\mathcal{X}_\lambda^\pm := \{z \in \mathcal{X}_\lambda : \mathcal{D}_\lambda(z) = 0\}$ the ‘contact set’ for λ . The possibility set is partitioned as $\Lambda = \Lambda_{\mathcal{F}}^{\neq} \cup \Lambda_{\mathcal{F}}^{\succ}$ and $\Lambda_{\mathcal{F}}^{\succ} = \Lambda_{\mathcal{F}}^{\succ} \cup \Lambda_{\mathcal{F}}^{\bar{\succ}}$, where

$$\Lambda_{\mathcal{F}}^{\neq} := \{\lambda \in \Lambda : \lambda \not\preceq_{\mathcal{F}} \tau\}; \quad (20)$$

$$\Lambda_{\mathcal{F}}^{\succ} := \{\lambda \in \Lambda : \lambda \succeq_{\mathcal{F}} \tau\}; \quad (21)$$

$$\Lambda_{\mathcal{F}}^{\succ} := \left\{ \lambda \in \Lambda_{\mathcal{F}}^{\succ} : \mathcal{D}_\lambda(z) > 0 \forall z \in \mathcal{X}_\lambda \right\}; \quad (22)$$

$$\Lambda_{\mathcal{F}}^{\bar{\succ}} := \left\{ \lambda \in \Lambda_{\mathcal{F}}^{\succ} : D_\lambda(z) = 0 \text{ for some } z \in \mathcal{X}_\lambda \right\}. \quad (23)$$

Using the goal function $G_{\mathcal{F},\lambda} = \mathbb{E}_{\mathcal{F}}[\mathbf{x}]^T \lambda - \int_{\mathcal{X}} w(y) \mathcal{L}_\lambda(y) dy$, any optimal solution is given by $G_{\mathcal{F}}^* := \sup_{\lambda \in \Lambda_{\mathcal{F}}^{\succ}} G_{\mathcal{F},\lambda}$ and $\lambda_{\mathcal{F}}^* \in \arg \sup_{\lambda \in \Lambda_{\mathcal{F}}^{\succ}} G_{\mathcal{F},\lambda}$.

In what follows, $\|\cdot\|$ denotes the Euclidean norm and $\rightsquigarrow$ convergence in distribution; limits are considered as $T \rightarrow \infty$ unless otherwise stated.

The analysis relies on several standard assumptions concerning the stochastic process of the returns of the basis assets, the moment conditions and the blocking scheme.

Assumption 3.1.1: (a) $(X_t)_{t \in \mathbb{Z}}$ is stationary and strongly mixing with mixing coefficients $(\alpha_k)_{k \in \mathbb{N}}$ which satisfy $\alpha_k = O(k^{-A})$ for $A > 1$; (b) $\mathbf{g}$ is bounded, the moment conditions are well specified, that is, $\mathbb{E}_{\mathcal{F}}[\mathbf{g}(\mathbf{x})] = \mathbf{0}$ and $\text{Var}_{\mathcal{F}} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{g}(X_t) \right] \rightarrow V_g$ which is positive definite; (c) the block size satisfies $B = o\left(T^{\frac{1}{2}}\right)$.

Many stationary models used in financial econometrics obey geometric ergodicity which suffices to establish the assumed mixing properties. Examples are

the strictly stationary versions of multivariate ARMA, GARCH and stochastic volatility models (for example, [Francq and Zacoian \(2011\)](#)). Geometric ergodicity together with the boundedness of the support of X_0 enables the validity of CLTs on stationary mixing processes.

The restriction on the asymptotic long-run covariance matrix V_g can be formulated in terms of restrictions on parameters and properties of the aforementioned models. The restriction on the asymptotic behavior of the block size is standard ([Kitamura \(1997\)](#)).

Let the function $L : \Lambda \times \mathcal{X}^d \times \mathcal{X}^K \rightarrow \mathbb{R}$ be appropriately measurable and bounded. The set of random variables $\{L(\lambda, z, \cdot), (\lambda, z) \in \Lambda \times \mathcal{X}^d\}$ is obviously L^2 -uniformly bounded. For any $\epsilon > 0$, consider the bracketing number of this set $\mathcal{N}_\epsilon$ which is essentially the minimum number of relevant ϵ -brackets needed to cover it; see for example Definition 2.1 in [Andrews and Pollard \(1994\)](#).

The following lemma facilitates the derivation of asymptotic properties of scaled partial sum processes below.

Lemma 3.1.2: *Suppose that Assumption 3.1.1 holds and that $\int_0^1 \sqrt{\frac{\mathcal{N}_\epsilon}{\epsilon}} d\epsilon < \infty$. Then,*

$$\sqrt{T} \left(\sum_{t=1}^T p_t L(\lambda, z, X_t) - \mathbb{E}(L(\lambda, z, X_0)) \right) \rightsquigarrow G_L(\lambda, z), \quad (24)$$

in the space of bounded real valued functions defined on $\Lambda \times \mathcal{X}^d$, where $G_L(\cdot, \cdot)$ is a zero mean Gaussian process with covariance kernel given by

$$\begin{aligned} \mathcal{C}_L((\lambda_m, z_n), (\lambda_{m^*}, z_{n^*})) &:= \sum_{i=0}^{\infty} \kappa_i \text{Cov}_{\mathcal{F}}(L(\lambda_m, z_n, X_0), L(\lambda_{m^*}, z_{n^*}, X_i)) \\ &- \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g'(X_0) L(\lambda_m, z_n, X_i)] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g(X_0) L(\lambda_{m^*}, z_{n^*}, X_i)] \end{aligned} \quad (25)$$

$$\text{and } \lambda_m, \lambda_{m^*} \in \Lambda, z_n, z_{n^*} \in \mathcal{X}, \kappa_i = \begin{cases} 1, & i = 0 \\ 2, & i > 0 \end{cases}.$$

Due to Assumption 3.1.1, the condition $\int_0^1 \sqrt{\frac{\mathcal{N}_\epsilon}{\epsilon}} d\epsilon < \infty$ holds if $\mathcal{N}_\epsilon = O(\epsilon^{-B})$ for $B > 1$, which is true in the applications of the result below.

The assumption framework is completed by introducing a specific notion of imperfect mutual dependence. Let $\lambda(\gamma) := (1 - \gamma)\lambda + \gamma\tau$, $\gamma \in [0, 1]$, be a mixture portfolio. Since the LPM function [\(2\)](#) is convex, $\mathcal{L}_{\lambda(\gamma)}(z) \leq (1 - \gamma)\mathcal{L}_\lambda(z) + \gamma\mathcal{L}_\tau(z)$ for all $z \in \mathcal{X}$ and $\gamma \in [0, 1]$. Strict convexity can arise locally in case of imperfect dependence:

Definition 3.1.3: Portfolios $\lambda \in \Lambda$ and $\tau \in \Lambda$ are weakly independent if

$$\begin{aligned} \mathcal{L}_{\lambda(\gamma)}(z) &< (1 - \gamma)\mathcal{L}_{\lambda}(z) + \gamma\mathcal{L}_{\tau}(z) \\ \Leftrightarrow \mathcal{D}_{\lambda(\epsilon)}(z) &> (1 - \epsilon)\mathcal{D}_{\lambda}(z), \end{aligned} \quad (26)$$

for all $z \in \mathcal{X}_{\lambda}^{\leq}$ and small enough $\epsilon > 0$.

Assumption 3.1.4: Any optimal solution $\lambda^*_{\mathcal{F}}$ and the benchmark τ are weakly independent.

3.2 Limit Theory for the IDF

Using Lemma 3.1.2 for $d = K$ and $L(\lambda, z, X) := 1_{X \leq z}$ and Example 19.6 in [Van Der Vaart \(2000\)](#), a limit theory for the IDF is obtained, where the inequality relation $\leq$ is interpreted in a pointwise way.

Proposition 3.2.1: Under Assumption 3.1.1,

$$\sqrt{T} \left(\hat{\mathcal{F}}(z) - \mathcal{F}(z) \right) \rightsquigarrow G_{\mathcal{F}}(z), \quad (27)$$

in the space of bounded real valued functions defined on $\mathcal{X}$, where $G_{\mathcal{F}}(\cdot)$ is a zero mean Gaussian process with covariance kernel given by

$$\begin{aligned} \mathcal{C}_{\mathcal{F}}(z_n, z_{n^*}) &:= \sum_{i=0}^{\infty} \kappa_i \text{Cov}_{\mathcal{F}}(1_{X_0 \leq z_n} 1_{X_i \leq z_{n^*}}) \\ &- \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g'(X_0) 1_{X_0 \leq z_n}] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g(X_0) 1_{X_i \leq z_{n^*}}], \end{aligned} \quad (28)$$

$$\text{and } z_n, z_{n^*} \in \mathcal{X}, \kappa_i = \begin{cases} 1, & i = 0 \\ 2, & i > 0 \end{cases}.$$

This proposition generalizes the analysis in ‘Example 3’ of [Qin and Lawless \(1994\)](#) to the present context of weak dependence. The result naturally implies that

$$\sup_{\mathbf{x} \in \mathcal{X}^K} \left| \hat{\mathcal{F}}(\mathbf{x}) - \mathcal{F}(\mathbf{x}) \right| = o_p(1). \quad (29)$$

It is straightforward to obtain analogous results at the portfolio level for $\hat{\mathcal{F}}_{\lambda}$.

3.3 Decision errors

Two types of decision error may occur in SD/EL: (i) the solution portfolio may not dominate the benchmark (violation of the model constraints) and (ii)

the solution may not maximize the goal score (sub-optimality). Using the uniform convergence property and the notion of weak independence, it is shown that these decision errors vanish in large samples.

The solution portfolio dominates the benchmark under the latent DF, with probability converging to one:

Proposition 3.3.1: *Under Assumption 3.1.1,*

$$\mathbb{P} [\boldsymbol{\lambda}^*_{\hat{\mathcal{F}}} \succeq_{\mathcal{F}} \boldsymbol{\tau}] \rightarrow 1. \quad (30)$$

Furthermore, the solution maximizes the latent optimal goal score, with probability converging to one:

Proposition 3.3.2: *Under Assumption 3.1.1 and Assumption 3.1.4, for arbitrary small $\epsilon > 0$,*

$$\mathbb{P} \left[\left(G_{\mathcal{F}}^* - G_{\mathcal{F}, \boldsymbol{\lambda}^*_{\hat{\mathcal{F}}}} \right) \leq \epsilon (G_{\mathcal{F}}^* - G_{\mathcal{F}, \boldsymbol{\tau}}) \right] \rightarrow 1. \quad (31)$$

3.4 Illustration

Figure 1 illustrates the pivotal role of weak independence in the analysis. The DF $\mathcal{F}(\boldsymbol{x})$ in this example is an IDF from the empirical application in Section 5. Shown is the risk reduction $\mathcal{D}_{\boldsymbol{\lambda}}(z)$ achieved by the optimal portfolio $\boldsymbol{\lambda}_{\mathcal{F}}^*$ and mixture portfolios $\boldsymbol{\lambda}_{\mathcal{F}}^*(w)$, $w = 0.02, 0.05, 0.10$, for every relevant return threshold.

The figure shows that $\mathcal{D}_{\boldsymbol{\lambda}_{\mathcal{F}}^*}(z) = 0$ for thresholds around $z = -1.8$ and $z = -0.5$. For these ‘contact areas’, $\mathbb{P}[\hat{\mathcal{D}}_{\boldsymbol{\lambda}_{\mathcal{F}}^*}(z) < 0]$ does not converge to zero in large random samples drawn from the DF and hence the dominance relation is not asymptotically almost sure.

However, the solution is weakly independent of the benchmark. The graphs for the mixtures show that strict convexity (26) occurs for all relevant thresholds, including the ‘contact areas’. The mixtures have no binding dominance constraints and they will asymptotically almost surely dominate the benchmark and thus represent feasible solutions which are approximately optimal.

[Insert Figure 1 about here.]

3.5 Asymptotic Efficiency

An asymptotic distribution for the empirical risk reduction process $\hat{\mathcal{D}}$ can also be obtained, allowing for an analysis of relative asymptotic efficiency of the SD/EL procedure compared with the simpler ‘plug-in’ approach.

Using Lemma 3.1.2 for $d = 1$ and $L(\boldsymbol{\lambda}, z, X) = M(\boldsymbol{\lambda}, z, X) := \max(0, z - \tau'X) - \max(0, z - \boldsymbol{\lambda}'X)$ and ‘Example 19.7’ in [Van Der Vaart \(2000\)](#), the following result is obtained:

Proposition 3.5.1 *Under Assumption 3.1.1,*

$$\sqrt{T} \left(\hat{\mathcal{D}}_{\boldsymbol{\lambda}}(z) - \mathcal{D}_{\boldsymbol{\lambda}}(z) \right) \rightsquigarrow G_{\mathcal{D}}(\boldsymbol{\lambda}, z), \quad (32)$$

in the space of bounded real valued functions defined on $\Lambda \times \mathcal{X}$, where $G_{\mathcal{D}}(\cdot, \cdot)$ is a zero mean Gaussian process with covariance kernel given by

$$\begin{aligned} \mathcal{C}_{\mathcal{D}}((\boldsymbol{\lambda}_m, z_n), (\boldsymbol{\lambda}_{m^*}, z_{n^*})) &:= \sum_{i=0}^{\infty} \kappa_i \text{Cov}_{\mathcal{F}}(M(\boldsymbol{\lambda}_m, z_n, X_0), M(\boldsymbol{\lambda}_{m^*}, z_{n^*}, X_i)) \\ &- \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}'(X_0)M(\boldsymbol{\lambda}_m, z_n, X_i)] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}(X_0)M(\boldsymbol{\lambda}_{m^*}, z_{n^*}, X_i)] \end{aligned} \quad (33)$$

$$\text{and } \boldsymbol{\lambda}_m, \boldsymbol{\lambda}_{m^*} \in \Lambda, z_n, z_{n^*} \in \mathcal{X} \text{ and } \kappa_i = \begin{cases} 1, & i = 0 \\ 2, & i > 0 \end{cases}.$$

Analogous results are obtained for the empirical risk reduction process based on the EDF, by letting the moment function $\mathbf{g}$ to be constant at zero in the relevant formulas. These results imply the following result on relative asymptotic efficiency of $\hat{\mathcal{D}}$, using $\mathcal{C}_0$ for the covariance kernel of the limiting Gaussian process in the EDF case.

Corollary 3.5.2: *For any λ , the efficiency gain from the inclusion of the side information in the derivation of $\hat{\mathcal{D}}_{\lambda}$ is represented by the collection over every non empty and finite $\mathcal{X}^* \subset \mathcal{X}$ of the semi-positive definite Gramian matrices with elements*

$$\begin{aligned} \mathcal{C}_0((\boldsymbol{\lambda}, z_n), (\boldsymbol{\lambda}, z_{n^*})) - \mathcal{C}_{\mathcal{D}}((\boldsymbol{\lambda}, z_n), (\boldsymbol{\lambda}, z_{n^*})) &:= \\ \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}'(X_0)M(\boldsymbol{\lambda}_m, z_n, X_i)] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}(X_0)M(\boldsymbol{\lambda}_{m^*}, z_{n^*}, X_i)], \end{aligned}$$

for any $z_n, z_{n^*} \in \mathcal{X}^*$.

The difference between the two covariance kernels represents the asymptotic gain from the inclusion of side information. The form of the elements of the matrix results from the blocking scheme which is asymptotically equivalent to optimal weighting (see [Kitamura \(1997\)](#)).

The formulation sheds further light on the proper specification of the moment function. Informational gains stem from the ‘long-run covariance’ between the empirical shortfall moments and the empirical side moments, as measured by $\sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}} [\mathbf{g}'(X_0) M(\boldsymbol{\lambda}_m, z_n, X_i)]$. Hence, the side moments should generally be chosen so as to be unbiased and highly ‘long-run correlated’ with the empirical shortfall moments of the relevant elements of Λ and $\mathcal{X}$.

Implementing these theoretical guidelines naturally requires a good understanding of the relevant base assets and investment strategy.

3.6 Further research

It stands to reason that the relative efficiency gains for given portfolios carries over to the solution portfolios and the opportunity loss, given the results about vanishing decision errors in Section 3.2.

It is conjectured that Proposition 3.5.1 and Corollary 3.5.2 imply standard convergence rates and asymptotic normality, as well as relative asymptotic efficiency for the solution portfolios, under appropriate assumptions about neighborhoods and contact sets of optimal portfolios. It seems particularly useful to establish the invertible linearization with respect to Hadamard derivatives of the asymptotic moment conditions on $\mathcal{X}_{\boldsymbol{\lambda}^* \mathcal{F}}^=$, in the spirit of [Fang and Santos \(2014\)](#), which would imply that the solution portfolios are linear transformations of the Gaussian process described in Proposition 3.5.1.

Since the focus of the present study is on the effect of including side information through EL in small samples, the authors plan to present a full-blown limit theory for portfolio optimization with SD constraints along these lines in a separate study.

4 Monte Carlo Simulation

A simulation experiment is developed to illustrate the potential improvements in estimation precision in finite samples from incorporating side information through the IDF. The experiment also provides guidance for the specification of

the SD/EL investment strategy in the next section.

The experiment creates 10,000 random samples of T consecutive daily observations from the 1927-2015 time series of the four common stock-market 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').

The sample size is $T = 50, 250, 1000$. Of particular interest seems $T = 250$, because momentum strategies typically use a one-year estimation window and a typical year includes roughly 250 trading days. By drawing data blocks rather than individual observations, the samples display realistic dynamic patterns.

The objective in the experiment is to estimate the latent factor means based on random samples, using either the EDF (4) or the IDF (10).

To implement EL, the experiment employs the model-free moment function (13) for 'winsorizing' factor means. Three alternative specifications are considered: 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 ($F = 4$).

The lower bound, L_j , $j = 1, \dots, 4$, is equated with the C -th percentile of historical distribution of the annual sample mean of the relevant factor, using $C = 5, 10, 25$; similarly, for the upper bound, U_j , $j = 1, \dots, 4$, the $(100 - C)$ -th percentile is used.

For example, using $C = 10$, the lower bound for RMRF equals $L_1 \approx -0.065$, which amounts to the uncontroversial assumption that the market risk premium is larger than minus 6.50 basis points per day (or minus 15.25 percent per annum).

The IDF is based on a block size of $B = 1, 3, 10$ trading days. A larger block size ($B > 10$) seems too demanding given the sample size ($T \leq 1000$).

The benefits of the blockwise approach are expected to be limited in this experiment, due to the specific structure of the moment conditions (13). If the sample factor mean violates a given bound (say, $\mathbb{E}_{\mathcal{F}_T}[f_1] < L_1$), then the (B)EL estimate generally equals that bound ($\mathbb{E}_{\hat{\mathcal{F}}}[f_1] = L_1$), regardless of the block size.

Nevertheless, increasing the block size is expected to increase the divergence $D_{KL}(\{\pi_t\} \parallel \{p_t\})$ in small samples, which in turn can introduce estimation error for factors which are not restricted by the moment conditions.

Table I shows the simulated (normalized) Root Mean Squared Error (RMSE) for the estimation of the mean of each factor, for a selection of relevant combi-

nations of T , F , C and B .

Not surprisingly, the RMSE decreases with the sample size, consistent with the known convergence properties of the EDF and IDF. The advantage of the IDF is most pronounced for small samples, which by their own nature show relatively frequent and large violations of the moment conditions. The IDF more than halves the RMSE for every factor in small samples ($T = 50$). As the sample size increases, the effect of the moment conditions vanishes and the results of the two methods converge. The RMSE reduction is one quarter to one third for intermediate samples ($T = 250$) and only a few percent for large samples ($T = 1000$).

The estimation precision increases as moment conditions for additional factors are introduced. Since the factors have a low mutual correlation, the moment conditions for a given factor have a limited effect on the estimation precision of the other factors. If relevant factors are omitted, then the divergence falls, and the RMSE increases.

Tightening the moment conditions by raising the value of C is another way to improve estimation precision. Of course, tightening the moment conditions introduces specification risk and may introduce a need for hypothesis tests or strategy backtesting in practice.

As expected, the benefits of the blockwise approach remain limited in this experiment. Although the block size has no material effect on the RMSE for the factor means, it does introduce excessive divergence $D_{KL}(\{\pi_t\} \parallel \{p_t\})$ in small samples, which increases estimation error for factors which are not explicitly constrained by the moment conditions.

For example, for $T = 50$, $F = 3$ and $C = 10$, increasing the block size from $B = 1$ to $B = 10$ increases the divergence by a factor of 13 (!) and increases of the RMSE for the omitted MOM factor.

[Insert Table I about here.]

5 Application

This section applies SD/EL and other methods to equity industry momentum strategies. Diversification is relevant in this area, because a concentrated po-

sition in just one or two top-performing industries would be riskier than most investors would be willing to tolerate.

The most common strategies rely on heuristic diversification rules without considering the risk levels and dependence structure. The objective here is to improve on heuristic diversification by using a combination of decision theory, statistical estimation and mathematical programming.

5.1 Data

The 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.

The analysis was repeated 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 effect and it can easily be enhanced using estimation and optimization in the present study. Better performance can be achieved using a conditional $1/K$ rule, which is discussed below. However, also that heuristic appears suboptimal in this study.

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 investment 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)).

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

5.2 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 (Moskowitz and Grinblatt (1999, Table III)). The focus here is therefore 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, five different enhanced portfolios are constructed from the base assets.

The first, heuristic portfolio (‘EW5’) is an equal-weighted average of the five industries with the highest realized return in the past 12 months. The number of five corresponds to roughly ten percent of the base assets, which is a common choice in the momentum literature. This specification also ensures that the heuristic portfolio has a similar risk level as the index.

The other four enhanced portfolios are constructed through estimation and optimization. The objective is to maximize the mean ($G_{\hat{\mathcal{F}}, \boldsymbol{\lambda}} = \mathbb{E}_{\hat{\mathcal{F}}} [\boldsymbol{x}]^T \boldsymbol{\lambda}$) subject to the restriction that the enhanced portfolio dominates the benchmark by MV or SSD, using the EDF or the IDF.

The IDF uses a small block size of $B = 1, 2, 3$ trading days. The typical estimation window contains only roughly 250 trading days, and using a larger block size would lead to the type of excessive divergence seen in the simulation experiment.

As in the simulation experiment, the analysis uses the model-free moment function (13) for ‘winsorizing’ the factor means of the market model ($F = 1$),

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

the three-factor model ($F = 3$) or the four-factor model ($F = 4$).

The bounds are the C -th and $(100 - C)$ -th percentiles of the historical distribution of the annual sample mean, where $C = 5, 10, 25$. The entire sample period is used to compute the bounds. Using backward-looking bounds based on past years leads to very similar results but this approach is not feasible for the first part of the sample period.

In the simulation experiment, the moment conditions are known features of the data generating process. In this application, the same specification is motivated by stylized facts about common factors and industry portfolios.

A large part of the general return variation of the industry portfolios is explained by the four factors. However, the industry momentum effect is not driven by continuation of the factors. In fact, the year-to-year predictability of the factors is limited and the sample means show strong regression to the mean after large positive or negative realizations.

Using the information about this regression effect helps to avoid large directional bets on common factors and improves the focus on industry-specific events. For example, the IDF reduces the bias of the optimal portfolio towards defensive industries after a general market sell-off and hence reduces the exposure to momentum crashes.

Figure 2 illustrates the regression effect and the 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 the 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 2 about here.]

5.3 Evaluation method

Out-of-sample performance is evaluated 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.

The analysis considers the mean, standard deviation, semi-deviation and skewness of the annual out-of-sample returns. Also reported are 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, the Value-at-Risk (VaR) and CVaR for a confidence level of 95% are used. As a risk-adjusted performance measure, the certainty equivalent return for a standard logarithmic utility function, $u(x) = \ln(100 + x)$, is reported.

The analysis also reports 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 risk levels. 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 factors have a limited effect on the relative performance of the strategies in this application. As a result of the investment restrictions, all enhanced portfolios have a similar market beta. The exposures to non-market 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.

5.4 The base case

The analysis starts with the combination of $K = 49$ industries; a six-month holding period ($H = 6$); $F = 4$ factors (the [Carhart \(1997\)](#) model); bounds for the factor means based on $C = 10$; a block size of $B = 1$ (the IID specification). This combination seems favorable given the aforementioned simulation results and stylized facts from industry momentum studies.

Table II 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 from 11.54% to 18.24%, an increase of 6.70%. 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 achieves an average return of 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 which prevents more 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 EL estimation is used. Using implied probabilities, SD achieves an average return of 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.

Due to the asymmetry of the risk 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.

The implied probabilities seem 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. Specifically, the median number of included industries is 5 for the SD portfolio versus 6 for the MV portfolio. Nevertheless, the SD/EL portfolio has a similar median turnover as the MV/EL portfolio (around 70 percent), due to a higher success in selecting the industries which are top-ranked out of sample and hence remain in the portfolio after rebalancing.

[Insert Table II about here.]

Figure 3 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, 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 3 about here.]

As the simulated RMSE reductions, the above out-of-sample performance improvements seem impressive given that they are achieved with the conservative moment inequalities (13). The 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.

Panel A of Table III gives more details about the investment 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% in these periods. 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 a momentum portfolio would then lower the portfolio’s expected return, due to the regression effect for the factors. Using implied probabilities mitigates this problem and hence offers ‘crash protection’.

Panel B further illustrates the pattern using the out-of-sample rank of the five industries with the highest in-sample rank based on either the sample probabilities or the implied probabilities.

The IDF better predicts the out-of-sample ranking than the EDF does. For example, following the ‘SMB crash’ of 72H2-73H1, the three top-ranked industries by IDF also occupy the three top spots in the out-of-sample holding period of 73H2. By contrast, using the EDF, only one top-three industry remains in the top three out of sample.

[Insert Table III about here.]

Figure 4 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 factor mean above the lower bound. 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 observations but it decreases at the level of data blocks.

[Insert Figure 4 about here.]

5.5 Alternative specifications

This section analyzes 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 factors (F), the choice of the bounds (C) and the block size (B). Table IV summarizes the results. To allow for a compact presentation, only one parameter is changed at the time, using the strategy with $K = 49$, $H = 6$, $F = 4$, $C = 10$ and $B = 1$ as the base case. In addition, only the average return is reported and all other output is suppressed.

The performance of all momentum strategies improves as the investment universe is widened, supporting the choice of $K = 49$ in [Hodder, Jackwerth and Kolokolova \(2015\)](#). 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.

The general performance level also improves as the holding period is shortened. 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 becomes less effective, because it is more difficult to predict the out-of-sample values of the factors, for short holding periods. Nevertheless, MV/EL and SD/EL remains the most appealing strategy also for a short horizon.

As the number of factors (and thus moment conditions) increases, the performance improvements from using implied probabilities increases, consistent with the RMSE reduction in the simulation experiment. The EL method presumably benefits from the low mutual correlation of the factors and the high between-industry variation of the factor exposures. In this situation, every additional factor leads to further adjustments of the implied probabilities and the relative returns of the industries.

The precise value of the bounds seems of limited importance for investment performance. This robustness contrasts with the reported improvements in estimation precision in the simulation experiment. Relatively small adjustments of the EDF suffice to reduce the appeal of industries with low factor loadings after factor crashes, as illustrated in Table III. Furthermore, high values of C contradict the factor drift which occurs in the sample but not in the simulations.

As expected based on the simulation experiment, the investment performance deteriorates as the block size increases. The number of non-overlapping blocks in an intermediate estimation window is relatively small compared with the number of base assets. In this situation, the blockwise approach introduces excessive divergence which is not related to the factors and which distorts the industry-level drift signal.

[Insert Table IV about here.]

6 Concluding Remarks

The Monte Carlo simulation illustrates the potential improvements in estimation precision in finite samples from incorporating side information through EL, even if the prior is limited to conservative lower and upper bounds for a small number of common risk factors.

In the application, the SD/EL combination achieves important out-of-sample 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.

In contrast to heuristic diversification, SD/EL accounts for the risk level and dependence structure of the base assets. Contrary to MV optimization, the method recognizes the favorable upside potential of concentrated winner portfolios. In contrast to the plug-in approach, the use of side information about the regression effect of common factors reduces the adverse effects of factor crashes.

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

Using the blockwise approach has an adverse effect in the application, because the number of non-overlapping blocks is relatively small compared with the number of base assets, as a result of using an intermediate estimation window. The BEL seems more appropriate for investment strategies which employ a larger estimation window, such as long-term price reversal strategies.

The relevant set of moment conditions naturally depends on the set of base assets and the investment strategy. Winsorizing the factor means is more effective if the estimation window is short and the sample means show high variability. In addition, the approach is more effective if the base assets show more cross-sectional variation in their factor exposures. The specification of the optimal moment conditions for various asset classes and investment styles seems an interesting topic for future research.

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] Andrews D. W. K. and D. Pollard, 1994, An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes, *International Statistical Review*, 62, 119-132.
- [6] Barrett, G. and S. Donald, 2003, Consistent tests for stochastic dominance, *Econometrica* 71, 71-104.
- [7] Bawa, V.S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.
- [8] 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.
- [9] 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.
- [10] Brown, B. and W. Newey, 2002, Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference, *Journal of Business & Economic Statistics* 20, 507-517.
- [11] Carhart, M.M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
- [12] Daniel, K. and T. Moskowitz, 2016, Momentum Crashes, forthcoming in *Journal of Financial Economics*.
- [13] Davidson, R. and J.-Y. Duclos, 2000, Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica* 68, 1435-1464.
- [14] Dentcheva, D. and A. Ruszczyński, 2003, Optimization with Stochastic Dominance Constraints, *SIAM Journal on Optimization* 14, 548-566.

- [15] 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.
- [16] 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.
- [17] Deshpande, J.V. and H. Singh, 1985, Testing for second order stochastic dominance, *Comm. Statist. Theory Method* 14, 887-893.
- [18] Drud, A.S., 1994, CONOPT—A Large-Scale GRG Code, *ORSA Journal on Computing* 6, 207-216.
- [19] El Barmi, H. and D. Marchev, 2009, New and Improved Estimators of Distributions Functions under Second Order Stochastic Ordering, *Journal of Nonparametric Statistics* 21, 2, 143-153.
- [20] 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.
- [21] Fang, Z and A Santos, 2014, Inference on directionally differentiable functions, arXiv preprint arXiv:1404.3763.
- [22] Francq, C and JM Zakoian, 2011. *GARCH models: structure, statistical inference and financial applications*, John Wiley & Sons.
- [23] Frost, P.A. and J.E. Savarino, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 15, 29-34.
- [24] 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.
- [25] Julliard, C. and A. Ghosh, 2012, Can Rare Events Explain the Equity Premium Puzzle?, *Review of Financial Studies* 25, 3037-3076.
- [26] Hadar, J. and W.R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2-34.
- [27] Hanoch, G. and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335-346.
- [28] Hodder, E., J.C. Jackwerth and O. Kolokolova, 2015, Improved Portfolio Choice Using Second-Order Stochastic Dominance, *Review of Finance* 19, 1623-1647.
- [29] Jagannathan, R. and T.S. Ma., 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651-1683.

- [30] Kitamura, Yuichi, 1997, Empirical likelihood methods with weakly dependent processes, *Annals of Statistics* 25, 2084-2102.
- [31] Kopa, M. and Th. Post, 2009, A Portfolio Efficiency Test Based on FSD, *Journal of Financial and Quantitative Analysis* 44, 1103-1124.
- [32] Kopa, M. and Th. Post, 2015, A General Test for SSD Portfolio Efficiency, *OR Spectrum* 37, 703-734.
- [33] Kuosmanen, T., 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390-1406.
- [34] Levy, H., 2016, *Stochastic Dominance: Investment Decision Making Under Uncertainty*, Springer, Third Edition.
- [35] 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.
- [36] Linton, O.B., Th. Post & Y.-J. Whang, 2014, Testing for the Stochastic Dominance Efficiency of a Given Portfolio, *Econometrics Journal* 17, 59-74.
- [37] McFadden, D., 1989, Testing for stochastic dominance. In *Studies in Economics of Uncertainty* (Edited by T. Fomby and T. K. Seo). New York.
- [38] Moskowitz, T.J. and M. Grinblatt, 1999, Do Industries Explain Momentum?, *Journal of Finance* 54, 1249-90.
- [39] Owen, A.B., 1988, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika* 75, 237-249.
- [40] Owen, A.B., 2001, *Empirical Likelihood*, New York: CRC Press.
- [41] Post, Th., 2003, Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance* 58, 1905-1932.
- [42] Post, Th., 2017, Empirical Tests for Stochastic Dominance Optimality, *Review of Finance* 21, 793-810.
- [43] Post, Th. and M. Kopa, 2016, Portfolio Choice based on Third-degree Stochastic Dominance, forthcoming in *Management Science*.
- [44] Post, Th. and V. Potì, 2017, Portfolio Analysis using Stochastic Dominance, Relative Entropy and Empirical Likelihood, *Management Science* 63, 153-165.
- [45] Qin, J. and J. Lawless, 1994, Empirical likelihood and general estimating equations, *Annals of Statistics* 22, 300-325.
- [46] Rio, E, 2013, Inequalities and limit theorems for weakly dependent sequences, working paper.

- [47] Rockafellar, R. T. and S. Uryasev, 2000, Optimization of Conditional Value-at-Risk, *Journal of Risk* 2, 21-41.
- [48] Rojo-Jiménez, J. and H. El Barmi, 2003, Estimation of Distribution Functions under Second Order Stochastic Orders Constraints, *Statistica Sinica* 13, 903-926.
- [49] Roman, D., Darby-Dowman, K. and Mitra, G., 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.
- [50] Roman, D., G. Mitra and V. Zverovich, 2013, Enhanced indexation based on second-order stochastic dominance, *European Journal of Operational Research* 228, 273-281.
- [51] Rothschild, M. and J.E. Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.
- [52] 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.
- [53] Scaillet, O., and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169-180.
- [54] Schennach, S.M., 2005, Bayesian exponentially tilted empirical likelihood, *Biometrika* 92, 31-46.
- [55] Shalit, H. and S. Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670-684.
- [56] Schmid, F. and Trede, M., 1998, A Kolmogorov-type test for second order stochastic dominance, *Statist. Probab. Lett.* 37, 183-193.
- [57] Van der Vaart, AW, 2000, *Asymptotic statistics*, Cambridge University Press.

Appendix

Proof for Lemma 3.1.2: Using Euler's Theorem, we have that

$$\sum_{t=1}^T p_t (L(\boldsymbol{\lambda}, z, X_t) - \mathbb{E}(L(\boldsymbol{\lambda}, z, X_0))) = \sum_{j=1}^{T^*} b_j L_j(\boldsymbol{\lambda}, z),$$

with $L_j(\boldsymbol{\lambda}, z) := \frac{1}{B} \sum_{i=1}^B (L(\boldsymbol{\lambda}, z, X_{(i-1)B+j}) - \mathbb{E}(L(\boldsymbol{\lambda}, z, X_0)))$. Furthermore, standard results in the EL theory imply that $b_j = \frac{1}{T^*} \frac{1}{1+\gamma_{T,B}^T \mathbf{g}_j}$, with $\mathbf{g}_j := \frac{1}{B} \sum_{i=1}^B \mathbf{g}(X_{(i-1)B+j})$. The arguments in the proof of Theorem 1 in Kitamura (1997) imply for the vector of Lagrange multipliers $\gamma_{T,B}$ that

$$\begin{aligned} \frac{\sqrt{T}}{B} \gamma_{T,B} &= \left(\frac{1}{T^*} \sum_{j=1}^{T^*} B \mathbf{g}_j \mathbf{g}_j' \right)^{-1} \frac{\sqrt{T}}{T^*} \sum_{j=1}^{T^*} \mathbf{g}_j + o_p(1) = O_p(1); \\ \gamma_{T,B} \max_{j=1, \dots, B} \|\mathbf{g}_j\| &= o_p(1). \end{aligned}$$

These results, together with Taylor expansions of $\frac{1}{1+x}$ around zero, imply that

$$\sqrt{T} \sum_{j=1}^{T^*} b_j L_j(\boldsymbol{\lambda}, z) = \frac{\sqrt{T}}{T^*} \sum_{j=1}^{T^*} L_j(\boldsymbol{\lambda}, z) - \frac{1+o_p(1)}{T^*} \sum_{j=1}^{T^*} B \mathbf{g}_j L_j(\boldsymbol{\lambda}, z) \frac{\sqrt{T}}{B} \gamma_{T,B}.$$

Using this result, together with stationarity and ergodicity, boundedness of $\mathbf{g}$ and L and Corollary 4.1 of Rio (2013), we obtain asymptotic normality for any finite dimensional distribution. Now, endow $\Lambda \times \mathcal{X}$ with the restriction of the relevant Euclidean metric w.r.t. which it is totally bounded. For arbitrary $v := (\boldsymbol{\lambda}, z)$ and arbitrary stochastic sequences $u_T, u_T^* \rightarrow u$ almost surely, we have that

$$\frac{1}{T^*} \sum_{j=1}^{T^*} B \mathbf{g}_j (L_j(\boldsymbol{\lambda}_T, z_T) - L_j(\boldsymbol{\lambda}_T^*, z_T^*)) \frac{\sqrt{T}}{B} \gamma_{T,B} \rightarrow 0, \text{ a.s.}$$

due to the previous, the boundedness of $\mathbf{g}$ and L , and the Uniform Law of Large Numbers for stationary and ergodic processes. Hence the $\frac{1+o_p(1)}{T^*} \sum_{j=1}^{T^*} B \mathbf{g}_j L_j(\boldsymbol{\lambda}, z) \frac{\sqrt{T}}{B} \gamma_{T,B}$ term in the previous expansion is asymptotically tight. Given Assumption 3.1.1 and the finiteness of $\int_0^1 \sqrt{\frac{N_\epsilon}{\epsilon}} d\epsilon$, asymptotic tightness for $\frac{\sqrt{T}}{T^*} \sum_{j=1}^{T^*} L_j(\boldsymbol{\lambda}, z)$ follows by Theorem 2.2 of Andrews and Pollard (1994). Hence, $\sqrt{T} \sum_{j=1}^{T^*} b_j L_j(\boldsymbol{\lambda}, z)$ is asymptotically tight and thereby Theorem 18.14 of Van Der Vaart (2000) is applicable, establishing the needed result. ■

Proof for Proposition 3.3.1: Consider the event $\mathcal{E} := \left(\boldsymbol{\lambda} \not\prec_{\hat{\mathcal{F}}} \boldsymbol{\tau} \forall \boldsymbol{\lambda} \in \Lambda_{\hat{\mathcal{F}}}^{\neq} \right)$.

For any $\boldsymbol{\lambda} \in \Lambda_{\hat{\mathcal{F}}}^{\neq}$, a strict inequality $\mathcal{D}_{\boldsymbol{\lambda}}(z) < 0$ occurs for some $z \in \mathcal{X}_{\boldsymbol{\lambda}}$.

Consequently, the uniform convergence property of $\hat{D}_\lambda(z)$ implies that $\mathcal{E} = \left(\left(\hat{D}_\lambda(z) < 0, z \in \mathcal{X}_\lambda \right) \forall \lambda \in \Lambda_{\mathcal{F}}^{\neq} \right)$ is asymptotically almost sure. Since $\lambda_{\mathcal{F}}^* \succeq_{\hat{\mathcal{F}}} \tau$ (by construction), it follows that $\mathcal{E} \Rightarrow \lambda_{\mathcal{F}}^* \in \Lambda_{\mathcal{F}}^{\succeq}$. Hence, $\mathbb{P}[\lambda_{\mathcal{F}}^* \succeq_{\mathcal{F}} \tau] \geq \mathbb{P}[\mathcal{E}] \rightarrow 1$, as $T \rightarrow \infty$. ■

Proof for Proposition 3.3.2: Consider the following three asymptotically almost sure events:

$$\begin{aligned} \mathcal{E}_1 &:= (\lambda \succeq_{\hat{\mathcal{F}}} \tau \forall \lambda \in \Lambda_{\mathcal{F}}^{\succeq}); \\ \mathcal{E}_2 &:= \left(G_{\hat{\mathcal{F}}, \lambda} \leq G_{\hat{\mathcal{F}}, \lambda_{\mathcal{F}}^*(\epsilon)} \forall \lambda \in \Lambda : G_{\mathcal{F}, \lambda} < G_{\mathcal{F}, \lambda_{\mathcal{F}}^*(\epsilon)} \right); \\ \mathcal{E}_3 &:= \left(G_{\hat{\mathcal{F}}, \lambda_{\mathcal{F}}^*} \neq G_{\hat{\mathcal{F}}, \lambda_{\mathcal{F}}^*(\epsilon)}; \lambda_{\mathcal{F}}^* \neq \lambda_{\mathcal{F}}^*(\epsilon) \right) \cup (\lambda_{\mathcal{F}}^* = \lambda_{\mathcal{F}}^*(\epsilon)). \end{aligned}$$

Since the individual events are asymptotically almost sure, their intersection is also asymptotically almost sure: $\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \rightarrow 1$, as $T \rightarrow \infty$. Furthermore, it can be established that $\lambda_{\mathcal{F}}^*(\epsilon) \in \Lambda_{\mathcal{F}}^{\succeq}$, using $\lambda_{\mathcal{F}}^* \in \Lambda_{\mathcal{F}}^{\succeq}$ and weak independence. Consequently, $\mathcal{E}_1 \Rightarrow (\lambda_{\mathcal{F}}^*(\epsilon) \succeq_{\hat{\mathcal{F}}} \tau) \Rightarrow (G_{\hat{\mathcal{F}}, \lambda_{\mathcal{F}}^*} \geq G_{\hat{\mathcal{F}}, \lambda_{\mathcal{F}}^*(\epsilon)})$ and hence $(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3) \Rightarrow (G_{\mathcal{F}, \lambda_{\mathcal{F}}^*} \geq G_{\mathcal{F}, \lambda_{\mathcal{F}}^*(\epsilon)})$. Finally, $G_{\mathcal{F}, \lambda_{\mathcal{F}}^*(\epsilon)} \geq (1 - \epsilon)G_{\mathcal{F}}^* + \epsilon G_{\mathcal{F}, \tau}$, using the convexity of the underlying LPM function. Using these arguments, the proof can be completed in the following way: As $T \rightarrow \infty$,

$$\begin{aligned} \mathbb{P} \left[(G_{\mathcal{F}}^* - G_{\mathcal{F}, \lambda_{\mathcal{F}}^*}) \leq \epsilon (G_{\mathcal{F}}^* - G_{\mathcal{F}, \tau}) \right] &= \mathbb{P} \left[G_{\mathcal{F}, \lambda_{\mathcal{F}}^*} \geq (1 - \epsilon)G_{\mathcal{F}}^* + \epsilon G_{\mathcal{F}, \tau} \right] \\ &\geq \mathbb{P} \left[G_{\mathcal{F}, \lambda_{\mathcal{F}}^*} \geq G_{\mathcal{F}, \lambda_{\mathcal{F}}^*(\epsilon)} \right] \geq \mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \rightarrow 1. \blacksquare \end{aligned}$$

Figure 1: Weak mutual dependence

The IDF for the 2013 estimation window from the empirical application is treated as the latent DF. Shown is the risk reduction $\mathcal{D}_{\lambda}(z)$ achieved by the SD/EL optimal portfolio $\lambda_{\mathcal{F}}^*$ and mixture portfolios $\lambda_{\mathcal{F}}^*(w)$, $w = 0.02, 0.05, 0.10$, for every relevant return threshold. Section 5 provides details about the underlying data set and investment strategy.

Figure 2: 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 the moment conditions (13). The lower bound L_j , $j = 1, \dots, 4$, equals the tenth percentile of the distribution of the sample mean across all estimation windows; the upper bound U_j , $j = 1, \dots, 4$, equals the 90-th percentile.

Figure 3: 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 (13); the analysis uses the $F = 4$ factors of Carhart (1997); the lower bound L_j , $j = 1, \dots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper bound U_j , $j = 1, \dots, 4$, equals the 90-th percentile. The IID EL method, or block size $B = 1$, is used. The graph uses a logarithmic scale.

Figure 4: Close-up of Implied Probabilities during Factor Crashes

Each panel shows the implied probabilities p_t , $t = 1, \dots, T$, for a 12-month estimation window during which one of the factors achieves its lowest sample mean across all estimation windows. The implied probabilities are based on the moment conditions (13); the analysis uses the $F = 4$ factors of Carhart (1997); the lower bound L_j , $j = 1, \dots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper bound U_j , $j = 1, \dots, 4$, 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 ($T = 252$). The IID EL method, or block size $B = 1$, is used. The implied probabilities are normalized by dividing by the sample probability T^{-1} . The graph uses a logarithmic scale.

Table I: Monte Carlo Simulation

The experiment creates 10,000 random samples of $T = 50, 250, 1000$ consecutive daily observations from the 1927-2015 time series of the four common stock-market factors of [Carhart \(1997\)](#). The latent mean vector $\mathbb{E}_{\mathcal{F}}[\mathbf{f}]$ is estimated using either the EDF ($\mathbb{E}_{\mathcal{F}_T}[\mathbf{f}]$) or the IDF ($\mathbb{E}_{\mathcal{F}}[\mathbf{f}]$). The EL method is implemented using the moment inequalities (13). The analysis uses $F = 1, 3, 4$ 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 bound L_j , $j = 1, \dots, 4$, equals the C -th percentile of the historical distribution of the factor's sample mean; the upper bound U_j , $j = 1, \dots, 4$, equals the $(100 - C)$ -th percentile, where $C = 5, 10, 25$. The analysis uses a BEL with maximally overlapping blocks of $B = 1, 3, 10$ consecutive observations. The table shows the (normalized) Root Mean Squared Error (RMSE) of the IDF-based estimator and its ratio to the RMSE of the EDF-based estimator, for a selection of parameter specifications (T, F, C, B). The RMSE is normalized by division by the standard deviation of the relevant factor. The final column shows the average value of the divergence to the sample probabilities, $D_{KL}(\{\pi_t\} \parallel \{p_t\})$.

T	F	C	B	$RMSE_{IDF}$				$RMSE_{IDF}/RMSE_{EDF}$				D_{KL}
				RMRF	SMB	SMB	MOM	RMRF	SMB	SMB	MOM	
50	4	10	1	0.0631	0.0776	0.0784	0.0667	0.3989	0.4773	0.4209	0.3629	0.0230
250	4	10	1	0.0514	0.0584	0.0650	0.0512	0.7047	0.7673	0.7350	0.6273	0.0014
1000	4	10	1	0.0307	0.0438	0.0340	0.0386	0.9260	1.0032	1.0015	0.9607	0.0000
250	1	10	1	0.0531	0.0799	0.0814	0.0800	0.7277	1.0505	0.9203	0.9796	0.0002
250	3	10	1	0.0516	0.0590	0.0648	0.0762	0.7075	0.7750	0.7326	0.9330	0.0011
250	4	10	1	0.0514	0.0584	0.0650	0.0512	0.7047	0.7673	0.7350	0.6273	0.0014
250	4	5	1	0.0610	0.0659	0.0728	0.0578	0.8365	0.8670	0.8232	0.7075	0.0006
250	4	10	1	0.0514	0.0584	0.0650	0.0512	0.7047	0.7673	0.7350	0.6273	0.0014
250	4	25	1	0.0351	0.0349	0.0417	0.0307	0.4815	0.4583	0.4710	0.3762	0.0054
250	4	10	1	0.0514	0.0584	0.0650	0.0512	0.7047	0.7673	0.7350	0.6273	0.0014
250	4	10	3	0.0511	0.0576	0.0646	0.0511	0.7001	0.7570	0.7299	0.6261	0.0026
250	4	10	10	0.0501	0.0563	0.0634	0.0510	0.6861	0.7403	0.7170	0.6252	0.0084
50	3	10	1	0.0639	0.0780	0.0796	0.1355	0.4041	0.4800	0.4273	0.7373	0.0179
50	3	10	3	0.0618	0.0755	0.0769	0.1360	0.3909	0.4645	0.4132	0.7401	0.0379
50	3	10	10	0.0567	0.0732	0.0739	0.1411	0.3584	0.4500	0.3970	0.7676	0.2385

Table II: Performance of Base Case 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 (13); the analysis uses the $F = 4$ factors of Carhart (1997); the lower bound L_j , $j = 1, \dots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper bound U_j , $j = 1, \dots, 4$, equals the 90-th percentile. The IID EL method, or block size $B = 1$, is used. The analysis evaluates 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. A riskless rate of 3.40% is assumed, 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 III: Close-up of Returns After Factor Crashes

Panel A summarizes the out-of-sample returns of several portfolios following severe ‘factor crashes’. Portfolio ‘Bench’ is the CRSP all-share index. EW5 is the equal-weighted average of the 5 industries with the highest realized return. MV maximizes the mean subject to a lower variance than the benchmark using plug-in estimates; MV/EL solves the same problem under implied probabilities. Similarly, SD maximizes the mean subject to stochastic dominance of the benchmark using sample probabilities; SD/EL solves the same problem under implied probabilities. 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 moment conditions (13); the analysis uses the $F = 4$ factors of Carhart (1997); the lower bound L_j , $j = 1, \dots, 4$, equals the tenth percentile of the historical distribution of the sample mean; the upper bound U_j , $j = 1, \dots, 4$, equals the 90-th percentile. The IID EL method, or block size $B = 1$, is used. 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. Panel B shows the out-of-sample rank of the five industries with the highest in-sample rank based on either sample probabilities or implied probabilities.

Panel A: Common Factors and Solution Portfolios												
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		
Panel B: Out-of-Sample Rank of Winner Industries												
Estimation Period	Holding Period	Sample Rank	Implied Rank									
RMRF	31H2-32H1	32H2	1	2	3	4	5	1	2	3	4	5
			23	43	37	40	41	23	41	22	37	40
SMB	72H2-73H1	73H2	2	11	19	36	1	2	1	3	36	44
HML	99H1-99H2	00H1	7	2	5	36	10	10	2	1	32	7
MOM	09H1-09H2	10H1	47	24	45	26	20	20	44	3	1	24

Table IV: 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 IDF ('MV/EL'); the SD portfolio based on sample probabilities; the SD portfolio based on the IDF ('SD/EL'). Holding periods of $H = 3, 6, 12$ months of daily returns are considered. The implied probabilities for MV/EL and SD/EL are based on the moment conditions (13). The analysis uses $F = 1, 3, 4$ 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 bound $L_j, j = 1, \dots, 4$, equals the C -th percentile of the historical distribution of the factor's sample mean across all estimation windows; the upper bound $U_j, j = 1, \dots, 4$, equals the $(100 - C)$ -th percentile, where $C = 5, 10, 25$. The analysis uses a BEL with maximally overlapping blocks of $B = 1, 2, 3$ consecutive observations. The analysis evaluates 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 amount of output, only one parameter is changed 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