Robust Statistics for Portfolio Construction and Analysis

R. Douglas Martin, Stoyan V. Stoyanov, Kirk Li, and Mahmoud Shammaa

R. Douglas Martin
is a professor emeritus in the Applied Mathematics and Statistics Departments at the University of Washington in Seattle, WA.
martinrd3d@gmail.com

Stoyan V. Stoyanov
is a director at Schwab Equity Ratings for the Schwab Center for Financial Research in Chicago, IL.
stoyan.stoyanov@schwab.com

Kirk Li
is a principal manager at Microsoft in Redmond, WA.
cocokecoli@gmail.com

Mahmoud Shammaa
is a director at Schwab Equity Ratings for the Schwab Center for Financial Research in Chicago, IL.
mahmoud.shammaa@schwab.com

KEY FINDINGS

- The adverse influence of asset returns and factor exposures outliers on least squares and sample covariance matrix estimators, used for portfolio construction and analysis, can be avoided by using modern robust estimators.
- A theoretically justified robust regression estimator is a valuable complement to least squares for cross-sectional and time-series factor model fitting; a highly robust covariance matrix estimate, and the related robust distance measure, are valuable complements to sample covariance matrix estimates and the related distance measure.
- Freely available open source R packages are available for computing the robust estimators and for creating and publishing reproducible research papers.

ABSTRACT

Asset returns and factor exposures frequently exhibit small fractions of extreme outliers, which are often associated with fat-tailed distributions and can have very adverse influence on classical least-squares regression estimators and sample covariance matrices. Over a number of decades, a solid theoretical and computational foundation has been developed for alternative robust estimators that are not much influenced by outliers. Unfortunately, such methods have seen relatively little use in portfolio construction and analysis. An overarching goal of this article is to encourage the use of robust statistics by portfolio managers and analysts, minimally as a complement to classical estimators and in some cases as a replacement. In support of this goal, the authors briefly describe the main data and theoretical foundations of robust statistics, then introduce a best-of-breed robust regression estimator with applications to cross-sectional and time-series factor model data. They go on to describe a highly robust covariance matrix estimator and the closely related robust multidimensional distance measure for outlier detection and shrinkage, applied to stock return and factor exposure data with influential outliers. A unique aspect of the robust estimators and most of the data used in this article is that they are freely available in several open source R packages. Consequently, most of the exhibits are reproducible with R code that may be found at: https://github.com/robustport/PCRA/blob/main/README.md.

There exists a large literature on the theory and methods of robust estimation, which includes the relatively recent theory and methods book by Maronna et al. (2019) and the many articles referenced therein, as well as the classic theoretical foundations books by Huber (1981), along with its second edition Huber and Ronchetti (2009), and Hampel et al. (1986).[1] Furthermore, the open source robust

[1] See also Rousseeuw and Leroy (1987).

statistics R packages Robust and robustbase, have been freely available on CRAN since 2006,[2] and the RobStatTM R package companion to Maronna et al. (2019) has been available on CRAN since 2019.[3] However, there has not yet been much use of robust methods in portfolio construction and analysis—or in empirical asset pricing. Rare exceptions include publicly unpublished but on-the-street awareness of the use of robust regression by Axioma (a Qontigo company); the mean–variance portfolio application research on robust regression described in Guerard, Xu, and Markowitz (2021) and in references therein to earlier work by Guerard and co-authors; and the use of a particular robust regression method in the asset pricing via machine learning article by Gu, Kelly, and Xiu (2020).

The overarching goal of this article is to encourage the following in the context of mean–variance portfolio optimization and risk analysis: 1) the routine use of a well-justified robust regression method, minimally as a diagnostic complement to the use of least squares (LS) regression for cross-sectional and time-series factor model fitting, and possibly as a replacement for LS in some cases; and 2) the routine use of a robust covariance matrix estimate and an associated robust multivariate distance measure as complements to the sample covariance matrix estimates and the related nonrobust Mahalanobis distances.[4] With this purpose in mind, we first provide a brief background on the data-oriented and theoretical foundations of robust estimation and introduce an optimal bias robust regression estimator, which has an intuitive weighted least-squares character, in the context of cross-sectional factor models. In the spirit of empirical asset pricing we provide application examples of the robust regression method for single-factor book-to-price ratio (BP) and earnings-to-price ratio (EP) models as well as for a multifactor model that includes BP and EP along with beta and size factors. As part of that discussion, we point out the inadequacy of Winsorization as it is commonly used in empirical asset pricing research. Then we discuss two applications of the robust regression method to time-series factor models. The first is robust estimation of beta, where we point out an inadequacy of the well-known Huber robust regression, and the second is robust fitting of stock returns to the Fama–French three-factor model (FF3) and four-factor model (FF4). We also introduce and provide an application example of an important robust model selection method, which is ever-more-needed with the explosion of factors being considered for factor investing.

We then introduce an important type of robust covariance matrix estimator, which has an intuitive interpretation as a weighted version of sample estimators of a returns mean vector and covariance matrix and which accurately reflects the correlations character of the bulk of a returns or factor exposures data set with influential outliers down-weighted. The importance of the robust covariance matrix estimators is demonstrated for a small portfolio of small-cap stocks. Furthermore, the robust covariance matrix estimator is used as a key component of a robust multivariate distance measure, which is shown to accurately identify multivariate return and factor exposure outliers and can be used for effective shrinkage of such outliers. This capability is of particular importance for reliable shrinkage cleaning of multifactor exposure outliers in cross-sectional multifactor models.

Creation of the examples in this article made use of the following open source R packages, which are freely available from CRAN (https://cran.r-project.org/): The robust statistics package RobStatTM, the portfolio construction and risk analysis package PCRA, and the cross-sectional factor models package facmodCS. The PCRA

[2] CRAN is the Comprehensive R Archive Network: https://cran.r-project.org.

[3] See https://cran.r-project.org/web/packages/RobStatTM/index.html.

[4] The latter were introduced to the portfolio management community by Chow et al. (1999) and Kritzman and Li (2010).

package is particularly important for instructional use in that it contains: 1) the stocksCRSP data set of stock returns, prices, and other related data for 294 stocks in the CRSP® database from 1993 to 2015;[5] 2) the factorsSPGMI data set of 14 factor exposures from Standard and Poor's Global Market Intelligence (SPGMI), which correspond to the 294 stocks in stocksCRSP. Details concerning these two data sets are available in the document "Introduction to CRSP® Stocks and SPGMI Factors in PCRA," which is downloadable from the PCRA package site at CRAN.

We emphasize that a detailed understanding of the theory of robust statistics is not required for successful application of the methods. In this regard, the code for reproducing most of the exhibits herein based on data in stocksCRSP and/or factorsSPGMI may be found at https://github.com/robustport/PCRA/blob/main/README.md, and the reader can quickly gain considerable immediate experience in the use of robust statistics by using the code on different data sets.

ROBUST STATISTICS OVERVIEW

Data-Oriented Robustness

All classical normal distribution maximum-likelihood estimators, such as the sample mean, least-squares regression estimators and sample covariance matrices, lack robustness toward outliers in the sense that outliers, and even a single outlier, can quite adversely influence the value of the estimate. On the other hand, a robust estimator is one that has the following data-oriented properties. A robust estimator

- is not much influenced by outliers;
- provides a good fit to the bulk of the data;
- results in reliable detection of multivariate outliers;
- provides better forecasts for future outlier-free data.

A simple illustrative example is provided in Exhibit 1, which plots the earnings per share (EPS) of the stock with ticker INVENSYS from 1984 to 2000.[6] It is clear that during 1994–2000, five of the seven robust fit residuals—that is, the difference between the observed EPS and the robust straight line fit—are smaller or considerably smaller than the corresponding LS residuals, while the robust fit residuals for the two obvious outliers in 1997 and 1998 are larger than the LS residuals. This illustrates the general character of robust model fitting in the presence of small fractions of influential outliers: Better fits are obtained for the large majority of the data, and outliers are more clearly exposed.

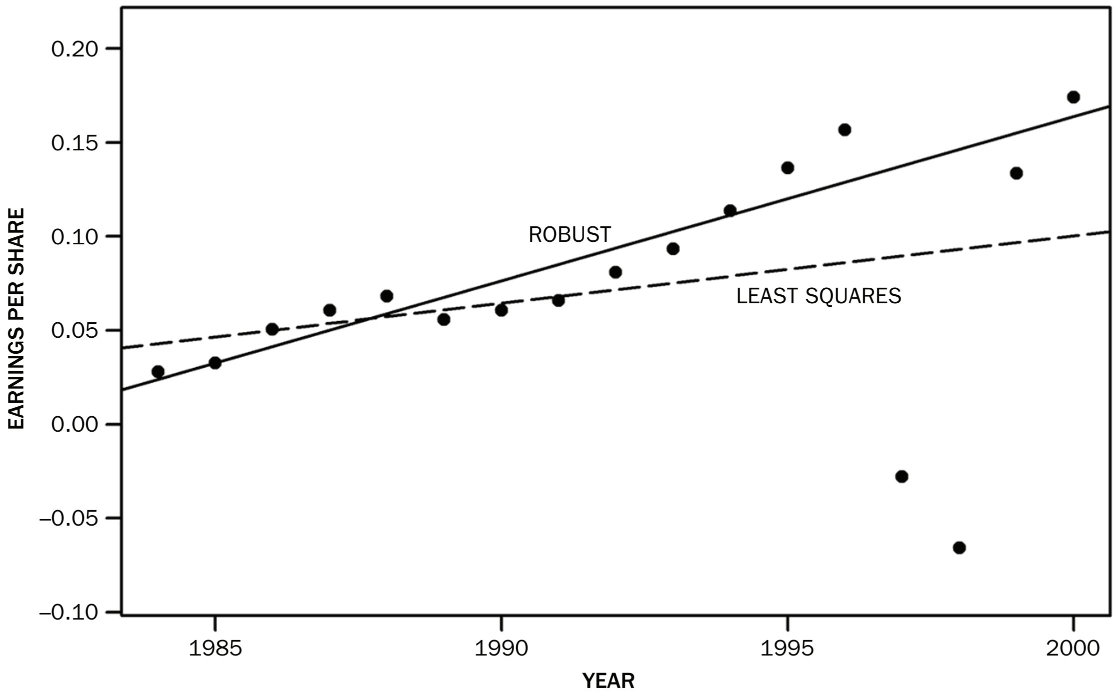
Estimator breakdown point and high breakdown point estimators. The finite sample *breakdown point* $\mathrm{BP}(\hat{\theta})$ of an estimator $\hat{\theta}$ is the largest fraction of the data $\mathbf{r}_n = (r_1, r_2, \ldots, r_n)$ that can be moved to infinity without causing the estimator to take on an infinite value. The sample mean $\bar{r}$ has breakdown point $\mathrm{BP}(\bar{r}) = 0$ because even a single data point $r_i \to \pm\infty$ results in $\bar{r} \to \pm\infty$. The sample median $\mathrm{med}(\mathbf{r}_n)$ has a finite sample breakdown point $\mathrm{BP}(\mathrm{med}(\mathbf{r}_n))$, which is a little less than 0.5 because if less than 1/2 of the data tends to infinity, $\mathrm{med}(\mathbf{r}_n)$ remains finite, and its breakdown point tends to 0.5 as the sample size n tends to infinity. Any estimator with the same breakdown point properties as the sample median is said to be a *high breakdown*

[5] CRSP® stands for the Center for Research in Security Prices, LLC.

[6] The data set used in this study was provided to one of the co-authors by an analyst in the Corporate Finance group at DuPont many years ago. The analyst needed to forecast next year's EPS for hundreds of stocks and was seeking guidance on how to use robust regression to deal with outliers.

EXHIBIT 1

EPS versus Time for Stock with Ticker INVENSYS with LS and Robust Line Fits



point estimator. It turns out that the robust regression and robust covariance matrix estimators we discuss in this article have high breakdown points.

Winsorization not. A γ% *Winsorized* data set is obtained by setting a fraction of smallest original data values equal to the next smallest value and setting a fraction of the largest original data values equal to the next largest data value, thereby *shrinking* a fixed fraction of potential outliers. The sample mean of a Winsorized data set is called a *Winsorized mean*, and a γ% Winsorized mean has a breakdown point $BP = \gamma$. Winsorization was originally introduced by Charles P. Winsor, and it was subsequently used in the often-referenced Tukey (1960) paper. As pointed out in Bali, Engle, and Murray (2016), it is common practice in empirical asset pricing research to Winsorize factor exposures at $\gamma = 0.5\%$ or $\gamma = 1.0\%$.[7]

Winsorization is a quite inadequate method for eliminating adverse outliers influence in least-squares factor model fitting, and we discourage its use in that context for the following reasons: 1) Winsorization is a rigid symmetric outlier shrinkage method that is used without regard to the statistical character of the data. 2) The commonly used 0.5% or 1% Winsorization provides no protection whatsoever against the adverse influence of fractions of outliers in the range 3% to 5%, and sometimes higher, that occur in practice.[8] 3) Most importantly, Winsorization is a one-dimensional method that cannot effectively cope with frequently occurring multivariate outliers that do not fully reveal themselves as one-dimensional coordinate outliers. In this article, we

[7] An alternative to Winsorizing data is to trim the data by deleting the smallest and largest γ% of the data, thereby creating a trimmed data set. Trimming is sometimes called truncating the data, as in Bali, Engle, and Murray (2016). Trimming is seldom used in a factor model context because trimming of either a return or the factor exposure alone results in missing data in the other variable, which results in dropping the entire returns and factor exposures pair for regression purposes.

[8] See the outliers fractions in Table 13 and the Winsorization examples in Figures 14 and 15 of Martin and Xia (2021).

introduce robust distances that are quite capable of identifying multivariate outliers for which one-dimensional outlier shrinkers are quite inadequate.

High breakdown point medians-based outliers shrinkage. It is a fact that factor exposures are often plagued with highly influential outliers that need to be *cleaned*, which usually means shrinking them, for purposes of fitting cross-sectional factor models in the context of mean–variance portfolio optimization and risk analysis. The following high breakdown point medians-based method, which appears to be used in some form in a number of commercial portfolio management systems, has proven to be much more reliable than Winsorization. The first step of the method is to compute the sample median med($\mathbf{x}$) of a set of n data value $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where the sample median is the middle data value for odd n and is the average of the two middle values for even n. The second step is to compute the *calibrated median absolute deviation about the median* robust scale estimate cmadm($\mathbf{r}$) given by

$$\text{cmadm}(\mathbf{x}) = 1.4826 \cdot \text{med}(|\mathbf{x} - \text{med}(\mathbf{x})|) \tag{1}$$

where the constant 1.4826 is used so that the expected value of cmadm($\mathbf{r}$) is approximately equal to the standard deviation σ of normally distributed data, and exactly so asymptotically. Finally, for a user-specified constant k, the robust outlier-shrinkage method consists of computing the lower and upper limit values

$$\begin{aligned} l &= \text{med}(\mathbf{x}) - k \cdot \text{cmad}(\mathbf{x}) \\ u &= \text{med}(\mathbf{x}) + k \cdot \text{cmad}(\mathbf{x}) \end{aligned} \tag{2}$$

and shrinking x_i values smaller than l to the value l, shrinking x_i values larger than u to the value u, and leaving the other x_i unaltered. The choice of k is up to the portfolio manager, and if for example she assumes a nominally normal data distribution and chooses $k = 3$, the probability of shrinking a nonoutlier data (the *false alarm* rate) is the fairly small value $2\Phi(-3) = 0.0027$. Investment context experience will usually lead a portfolio analyst to a reasonable choice of k value.

It is important to note that a medians-based outlier-shrinker could be adapted to skewed returns or factor exposures in one or both of the following ways: 1) by using different upper and lower k values, k_{upper} and k_{lower}, based on empirical analysis, and 2) using separate lower and upper robust scale estimators, $\text{cmad}_{\text{lower}}(\mathbf{r})$ and $\text{cmad}_{\text{upper}}(\mathbf{r})$, where the values of x used in 1 are replaced by those below med($\mathbf{x}$) and those above med($\mathbf{x}$), respectively.

Robust Statistics Theory Overview

An overarching goal of robust estimation is to be highly robust toward outliers by virtue of not being much influenced by them and also to be "close" to optimal in the idealized case of normally distributed data. Correspondingly, research on robust estimators focused on an infinite family of data distributions that contains an idealized normal distribution as a special case, as well as infinitely many non-normal outlier generating distributions. The performance of a robust estimator is then evaluated in terms of its large-sample bias and variance and normal distribution efficiency.

The Tukey–Huber model. The leading such family is a two-component mixture family of distributions, with a normal distribution central component and a second component that may be any one of an infinite number of distributions that includes outlier-generating distributions. Models of this type were originally introduced by Tukey (1960) and Huber (1964) in the context of estimating the mean parameter

μ of a nominal normal distribution. The Tukey version is a mixture of two normal distributions, and the Huber version is the following, more-general, two-component mixture model:

$$F_\mu(r) = (1-\gamma)N(r;\mu,\sigma^2) + \gamma H(r), \quad 0 \le \gamma < 0.5 \tag{3}$$

where σ^2 is a nuisance parameter and $H(r)$ is any one of a more or less unrestricted infinite family of distributions that includes outlier-generating distributions.[9] The unknown mixing parameter γ is an indicator of the closeness of the actual distribution function $F_\mu(r)$ to the idealized normal distribution function $N(r;\ \mu,\ \sigma^2)$. While robust statistics researchers have obtained theoretically attractive robust estimators for values of the mixing parameter over the full range $0 \le \gamma < 0.5$, for applications of robust estimators in quantitative finance research, the values of γ are typically in the range of 3% to 5% and rarely greater than 10%. See, for example, Martin and Xia (2021, Section 3.3.3).

Example 1: A Two-Component Normal Mixture Model

This example in the spirit of Tukey (1960) illustrates how a two-component normal mixture distribution that is close to a normal distribution can result in arbitrarily large bias and variance of the sample mean estimator. The density form of the mixture model is

$$f_{\text{mix}}(r;\gamma,\mu,\mu_{\text{mix}},\sigma^2,\sigma^2_{\text{mix}}) = (1-\gamma)f_N(r;\mu,\sigma^2) + \gamma f_N(r;\mu_{\text{mix}},\sigma^2_{\text{mix}}), \quad 0 < \gamma < 0.5 \tag{4}$$

where $f_N(r;\ \mu,\ \sigma^2)$ is a normal probability density function with mean μ and variance σ^2. It is an easy exercise to check that the density f_{mix} has the following mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$:

$$\tilde{\mu} = (1-\gamma)\cdot\mu + \gamma\cdot\mu_{\text{mix}} \tag{5}$$

$$\tilde{\sigma}^2 = (1-\gamma)\cdot\left[\sigma^2 + \gamma\cdot(\mu-\mu_{\text{mix}})^2\right] + \gamma\cdot\sigma^2_{\text{mix}} \tag{6}$$

The expected value and variance of the sample mean $\bar{r}$ for the distribution with density in Equation (4) are $\tilde{\mu}$ and $\tilde{\sigma}^2/n$, respectively, and the bias of the sample mean is $\gamma\cdot(\mu_{\text{mix}} - \mu)$. No matter how small a positive value γ is, the bias and variance of the sample mean can be arbitrarily large by virtue of μ_{mix} and σ^2_{mix} being arbitrarily large, the latter of which gives rise to outliers. Thus, the sample mean is quite nonrobust with regard to both variance and bias, and a robust alternative estimator is needed.

Robust estimator large-sample bias and variance. The theoretical properties of finite sample size robust estimators have essentially all been obtained in the context of the *large sample*, equivalently *asymptotic* values of *bias* and *variance* of the finite-sample estimator as the sample size tends to infinity. Such theoretical results are then supported to various degrees by finite-sample Monte Carlo studies. Here we describe the large-sample bias and variance of a sequence $\hat{\mu}_n$ of estimators of the unknown parameter μ in the Tukey–Huber model (3), but the essential concepts extend to other scalar- and vector-valued estimators.

[9] Of course sometimes σ is the parameter of interest, with μ a nuisance parameter, in which case the robust estimator is a little more complicated. See, for example, Chapter 2.5 and related sections of Maronna et al. (2019).

The large-sample bias $B(\hat{\mu}_n; F_\mu)$ is defined as the difference between the asymptotic value, $\mu(F_\mu)$ to which $\hat{\mu}_n$ converges (in probability), and the true parameter value μ:[10]

$$B(\hat{\mu}_n; F_\mu) = \hat{\mu}(F_\mu) - \mu. \tag{7}$$

Example 1 provides the simple explicit form shown in Equation (5) for $\mu(F_\mu)$ in the case of the sample mean estimator.

Unlike the generally nonzero large-sample bias of an estimator such as $\hat{\mu}_n$, the estimator finite sample variance goes to zero at a rate $1/n$. Consequently, the large-sample estimator variance is based on the limiting distribution of the standardized estimator $\sqrt{n}(\hat{\mu}_n - \tilde{\mu}(F_\mu))$, which is normal, with a mathematical expression for the large-sample variance, which depends on the form of the estimator and the underlying distribution function F_μ for the data.

Normal distribution efficiency. The degree to which a robust estimator is nearly optimal at a normal distribution is measured by a robust estimator's *normal distribution efficiency*, which is defined as

$$\mathrm{EFF}_\Phi(\hat{\theta}_{\mathrm{ROB}}) = \frac{V_\Phi(\hat{\theta}_{\mathrm{MLE}})}{V_\Phi(\hat{\theta}_{\mathrm{ROB}})}, \tag{8}$$

where $V_\Phi(\hat{\theta}_{\mathrm{MLE}})$ is the asymptotic variance of the maximum-likelihood estimator (MLE) for data with a standard normal distribution Φ and $V_\Phi(\hat{\theta}_{\mathrm{ROB}})$ is the corresponding robust estimator asymptotic variance.[11] For linear regression and covariance matrix estimators treated herein, the normal distribution MLEs are the least-squares estimator and the sample covariance matrix, respectively. Because we are using asymptotic variance and an MLE has the minimum achievable asymptotic variance, a robust estimator's efficiency is less than 1, or expressed as a percentage, is less than 100%. An overarching goal for robust estimation is to be "highly robust" toward outliers but also have a "high" normal distribution efficiency, which is commonly taken to be 90% or 95%. These goals compete with each other, and robust estimators typically have a tuning parameter that a user can specify to control a trade-off between robustness toward outliers and normal distribution efficiency. For example, in the case of the Huber min-max variance estimator of Example 2, c is the tuning constant, and the value $c = 1.345$ results in 95% normal distribution efficiency.

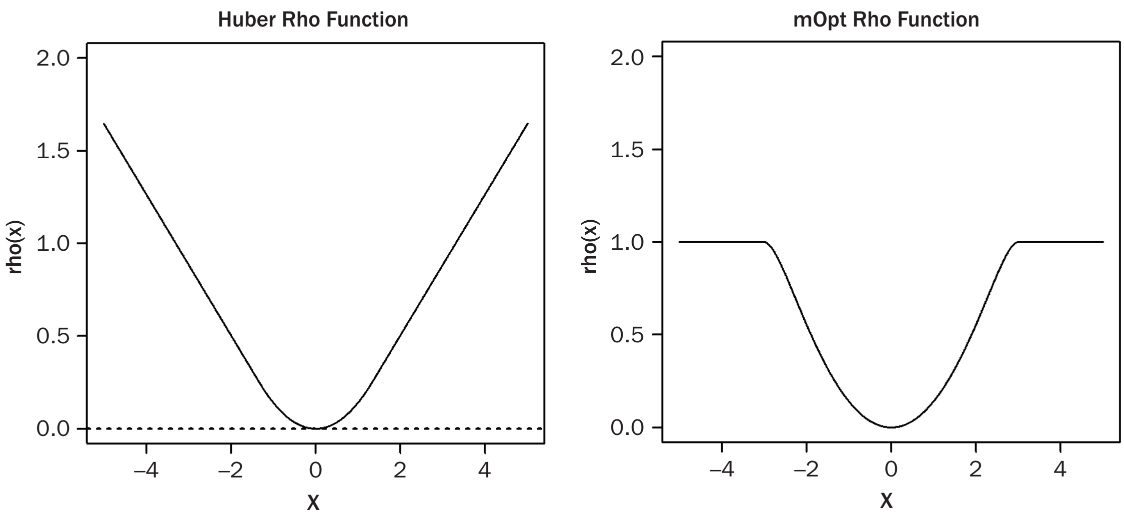
Example 2: The Huber Min-Max Variance Robust Mean Estimator

The mean of the central normal distribution in the Tukey–Huber model Equation (3) is called a "location" parameter in the robust statistics literature, for example, as in the seminal article by Huber (1964) on robust estimation of a location parameter. Huber's robust estimator is obtained by replacing the sample mean least-squares quadratic loss function with a loss function $\rho_{\mathrm{HUB}}(x)$ that is quadratic in a central region $[-c, +c]$ and is an absolute value function outside that interval, where c is a tuning parameter, which we will discuss shortly. A scaled version of $\rho_{\mathrm{HUB}}(x)$ is shown in Exhibit 2, Panel A. Huber established the remarkable result that this estimator

[10] Usually. in statistics one uses an unbiased estimator, or an estimator that has only a finite sample size bias that goes to zero as the sample sizes goes to infinity. This turns out to be impossible in robust statistics theory unless one assumes an unrealistic symmetry feature of the data distribution. Instead, one needs to design a robust estimator whose maximum large-sample bias is "small."

[11] This definition is also used for finite-sample estimators, where $\hat{\theta}_{MLE}$ is sometimes replaced with a "best known" estimator.

EXHIBIT 2
Huber and mOpt Rho Loss Functions



minimizes the maximum asymptotic variance over the distribution family (3) with $\sigma = 1$ and where $H(r)$ is *any* distribution function that is *symmetric* about μ.

Example 3: The Sample Median is a Min-Max Bias Estimator

When the $H(r)$ symmetry assumption of Example 2 is dropped, the Huber min-max variance estimator is asymptotically biased for some asymmetric $H(r)$. To deal with the bias problem caused by dropping the $H(r)$ symmetry assumption, Huber focused on the model shown in Equation (3) with $H(r)$ totally unrestricted, and showed for that case the sample median minimizes the maximum asymptotic bias. This attractive min-max bias property of the sample median deserves to be much better known than it is, even among statisticians.

Example 4: A Highly Robust Bounded Loss Function

While the min-max variance property of the Huber loss function $\rho_{HUB}(x)$ is attractive, the linearly unbounded character of $\rho_{HUB}(x)$ is of concern for data with very large outliers, and thus even for the problem of robustly estimating the mean, one may prefer a bounded loss function. Exhibit 2, Panel B, displays one such function, the so-called *mOpt* loss function $\rho_{mOpt}(x)$. This mOpt rho function turns out to have an attractive regression min-max bias optimality property, and so robust regression with the mOpt rho function is what we focus on extensively in the next section.

Extension of the Tukey–Huber model to robust regression and covariance matrix estimation. In theoretical discussion of robust covariance matrix estimation and robust regression, simple vector and matrix extensions of the simple Tukey–Huber location parameter model Equation (3) are in common use. In the case of robust regression studies, the normal distribution $N(r; \mu, \sigma^2)$ is replaced by a joint distribution for a return r and factor exposure variables $f_1, f_2, \ldots, f_K$ with normally distributed regression residuals. In the case of robust covariance matrix estimation studies, the normal distribution $N(r; \mu, \sigma^2)$ in Equation (3) is replaced by a multivariate normal distribution with unknown mean vector and covariance matrix.

ROBUST FACTOR MODELS

In this section, we first describe a highly robust method of fitting cross-sectional factor model fitting, with application examples, and then briefly do likewise for time-series factor models.

Robust Cross-Sectional Factor Models and the Fama–MacBeth Method

Factor model regression M-estimators. The cross-sectional K-factor model that relates a cross-section of returns, $r_{i,t}$, to K regressions slopes, f_{kt}, at each of the times $t = 1, 2, \ldots, T$ is given by

$$r_{i,t} = \alpha_t + \sum_{k=1}^{K} b_{ik,t-1} f_{kt} + \epsilon_{i,t}, \quad i = 1, 2, \cdots, N \tag{9}$$

where α_t is an intercept, $b_{ik,t-1}$ are the lagged factor exposures, and the errors $\epsilon_{i,t}$ have zero mean and are uncorrelated with f_{kt}. Viewed as time series, the f_{kt} are the *factor returns*. Collecting 1 and $b_{ik,t-1}$ into the vector $\mathbf{x}_i$, collecting α_t and the exposures $b_{ik,t-1}$ into the vector $\boldsymbol{\theta}_t$, and dropping the subscript t for notational convenience but keeping in mind that the exposures b_{ik} are measured at the previous time period, the model is

$$r_i = \mathbf{x}_i' \boldsymbol{\theta}_0 + \epsilon_{i,}, \quad i = 1, 2, \cdots, N \tag{10}$$

where $\boldsymbol{\theta}_0$ is the true but unknown value of $\boldsymbol{\theta}$. The previous cross-sectional regression models have two well-known important uses, the first of which is for constructing Markowitz mean–variance optimal portfolios and analyzing their risk, and the second is for empirical asset pricing studies as exemplified by Fama and French (1992) and many subsequent articles over many years. Here our application examples will focus on the latter type of application, using CRSP stock data and S&P Global factor data, using both least squares and the robust regression estimator that we now introduce.

For an arbitrary value of $\boldsymbol{\theta}$, the regression residuals are

$$\hat{\epsilon}_i(\boldsymbol{\theta}) = r_i - \mathbf{x}_i' \boldsymbol{\theta} \tag{11}$$

and a *regression M-estimator* $\hat{\boldsymbol{\theta}}_M$ minimizes the objective function

$$\sum_{i=1}^{N} \rho\left(\frac{\hat{\epsilon}_i(\boldsymbol{\theta})}{\hat{s}}\right), \tag{12}$$

where the *rho function* $\rho(x)$ is a symmetric loss function and $\hat{s}$ is a robust scale estimate computed prior to the minimization. The least squares and least absolute deviation (LAD) estimators are special cases of an M-estimator based on the loss function choices $\rho_{LS}(x) = x^2$ and $\rho_{LAD}(x) = |x|$, respectively, and the Huber estimator $\rho_{HUB}(x)$ in Example 2 is a blend between the LS and LAD rho functions.

A value $\hat{\boldsymbol{\theta}}_M$ that minimizes the objective (12) satisfies the equation

$$\sum_{i=1}^{N} \mathbf{x}_i \psi\left(\frac{\hat{\epsilon}_i(\hat{\boldsymbol{\theta}}_M)}{\hat{s}}\right) = \mathbf{0}, \tag{13}$$

where $\psi(x) = \rho'(x)$ is referred to as a *psi function*. The psi function defines a corresponding nonlinear *weight function*

$$w(x) = \frac{\psi(x)}{x} \tag{14}$$

and multiplying and dividing Equation 14 by $\hat{\epsilon}_i(\hat{\theta}_M)/\hat{s}$ shows that $\hat{\theta}_M$ is a solution of the nonlinear weighted least-squares (WLS) equation:

$$\sum_{i=1}^{N} w\left(\frac{\hat{\epsilon}_i(\hat{\theta}_M)}{\hat{s}}\right) \cdot \mathbf{x}_i \left(r_i - \mathbf{x}_i \hat{\theta}_M\right) = \mathbf{0}. \tag{15}$$

The nonlinear WLS equation may be conveniently and effectively solved using K steps of an *iteratively reweighted least-squares* (IRWLS) algorithm

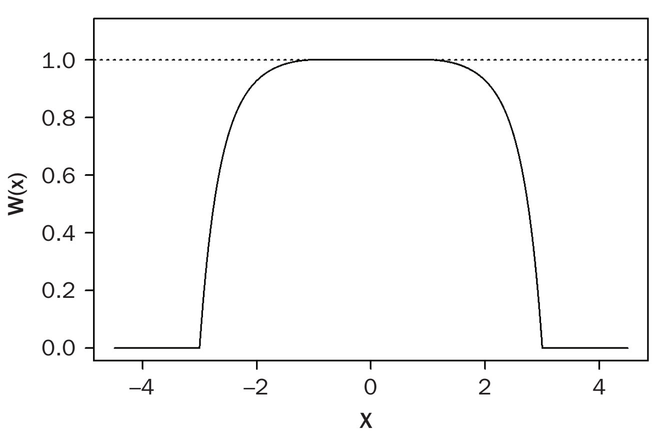
$$\sum_{i=1}^{N} w\left(\frac{\hat{\epsilon}_i\left(\hat{\theta}^{k-1}\right)}{\hat{s}}\right) \cdot \mathbf{x}_i \left(r_i - \mathbf{x}_i \hat{\theta}^{k}\right) = \mathbf{0}, \quad k = 1, 2, \cdots, K \tag{16}$$

where $\hat{\theta}^0$ is a robust initial estimate. Further details concerning the mOpt regression estimator are available in the context of cross-sectional empirical asset pricing in Martin and Xia (2021) and for time-series modeling in Martin and Xia (2022).

The mOpt regression estimator. The *mOpt* regression estimator is a solution to Equation (15) using the weight function $w_{\mathrm{mOpt}}(x)$ whose formula is given by Equation (A2) and whose constants are chosen so that the mOpt estimator has 95% normal distribution efficiency. This weight function is plotted in Exhibit 3. The weight function $w_{\mathrm{mOpt}}(x)$ is a very small modification of weight function $w_{\mathrm{Opt}}(x)$ of the Yohai and Zamar (1997) optimal robust regression M-estimator, which minimizes the maximum bias over a regression version of the Tukey–Huber family of distributions shown in Equation (3), where the family of distributions $H(r)$ is *totally unrestricted*.[12] The modification was made to ensure that the IRWLS algorithm converges. It follows from early work by Yohai (1987) that with a high breakdown point initial estimator $\hat{\theta}^0$, the mOpt regression estimator has a high breakdown point.

EXHIBIT 3

The Weight Function $w_{\mathrm{mOpt}}(x)$ for the 95% Normal Distribution Efficiency mOpt Regression Estimator



The mOpt weight function gives a weight of 1 to all robustly scaled residuals $\hat{\epsilon}_t(\hat{\theta}) = (r_i - \mathbf{x}_i'\theta)/\hat{s}$ that are less than 1 in magnitude and smoothly transitions to a weight of 0 for robustly scaled residuals whose absolute value is greater than 3.00. Data pairs $(r_t, \mathbf{x}_{t,i})$ that result in 0 weights are said to be *rejected*. For normally distributed data and true parameter values, the probability that such a pair is rejected is only 0.27%, and the estimator is essentially equivalent to the LS estimator. A considerable intuitive appeal of the mOpt estimator is that it is a robust smooth version of the nonrobust *classical three-sigma edit rule* of deleting any data point $(r_i, \mathbf{x}_i)$ for which the magnitude of the least-squares residual $r_i - \mathbf{x}_i'\hat{\theta}_{LS}$ is greater or equal to three times the residuals standard deviation $\hat{\sigma}$.

Fama–MacBeth regressions. The Fama–MacBeth method, commonly used in empirical asset pricing research, is to first compute LS regression estimates

[12] This is in sharp contrast to the Huber min-max variance estimator, described in the context of robust estimation of the mean in Example 2, which assumes that H(r) is symmetric about the regression mean.

$\hat{\alpha}_t, \hat{f}_{1t}, \hat{f}_{2t}, \cdots, \hat{f}_{Kt}$ for each $t = 1, 2, \ldots, T$, and then use each of these time series of factor returns estimates to compute *t*-statistics for the intercept and the *K* factors. We expand the standard Fama–MacBeth method based on LS regression by also using the mOpt regression estimator introduced earlier. As is standard in empirical asset pricing research, we use HAC adjusted *t*-statistics that are corrected for autocorrelation and heteroscedasticity in the time series of intercepts and factor returns. It is quite common to use the well-known Newey–West correction, but there is evidence[13] that an alternative HAC correction due to Andrews is better, so we use the Andrews version.[14] For assessing significance of the robust and LS regression results, we use an absolute *t*-statistics threshold of 3.0 to declare a factor to be significant and take factors with *t*-statistic absolute values between 2.0 and 3.0 to be weakly significant.[15]

From a practical point of view, the cross-sectional regression is used as standard statistical machinery to identify factors that drive the expected stock return in the cross section. Outliers may have a substantial influence on which factors are identified as significant because of the high sensitivity to them of the classical least-squares method, as has been substantiated in Martin and Xia (2021). In this section, we provide empirical examples that illustrate the adverse impact of outliers on LS cross-sectional regressions, and the lack of influence of such outliers on the mOpt robust regression method.

Our empirical examples are created using the R packages PCRA, RobStatTM, and facmodCS. Apart from providing tools for cross-sectional and time-series regression model fitting and analysis, the PCRA package contains freely available CRSP stock data and S&P Global factor exposure data. The time period covered in the data sets is from January 1993 to December 2015, and the size of the cross section is 294 companies. On one hand, the limited data availability prevents us from drawing firm empirical conclusions, but on the other hand, all examples are reproducible. If the researcher has access to a bigger data set, all experiments can be repeated easily in R.

Our examples concern the beta, size, and BP factors of the Fama–French cross-sectional three-factor model introduced in Fama and French (1992), along with the earnings-to-price ratio factor for a couple of reasons. First, it is instructive to illustrate the difference between the commonly used least-squares method and robust regression method on a standard and well-known model. Second, there is an interesting discussion in the academic literature on the relative strength of BP and EP when they are included jointly in a cross-sectional regression. Fama and French (1992) conclude that EP is not significant in FF3, but Martin and Xia (2021) show, using all non-financial firms listed on New York Stock Exchange (NYSE), American Stock Exchange (AMEX), or NASDAQ in the CRSP database for the 1993–2015 period, that this conclusion is essentially driven by small fractions of outliers ranging from 3% to 5% and that a robust cross-sectional regression in fact demonstrates the opposite—that BP is insignificant in the presence of EP. In this section, we provide illustrations of these relationships with the smaller returns and factors cross section size 294 contained in the PCRA package. We begin with a discussion of univariate asset pricing tests for BP and EP, and we conclude with a multivariate example.

We ran univariate cross-sectional regressions from January 1993 to December 2015 for both BP and EP factors. The time series of the BP and EP factor returns

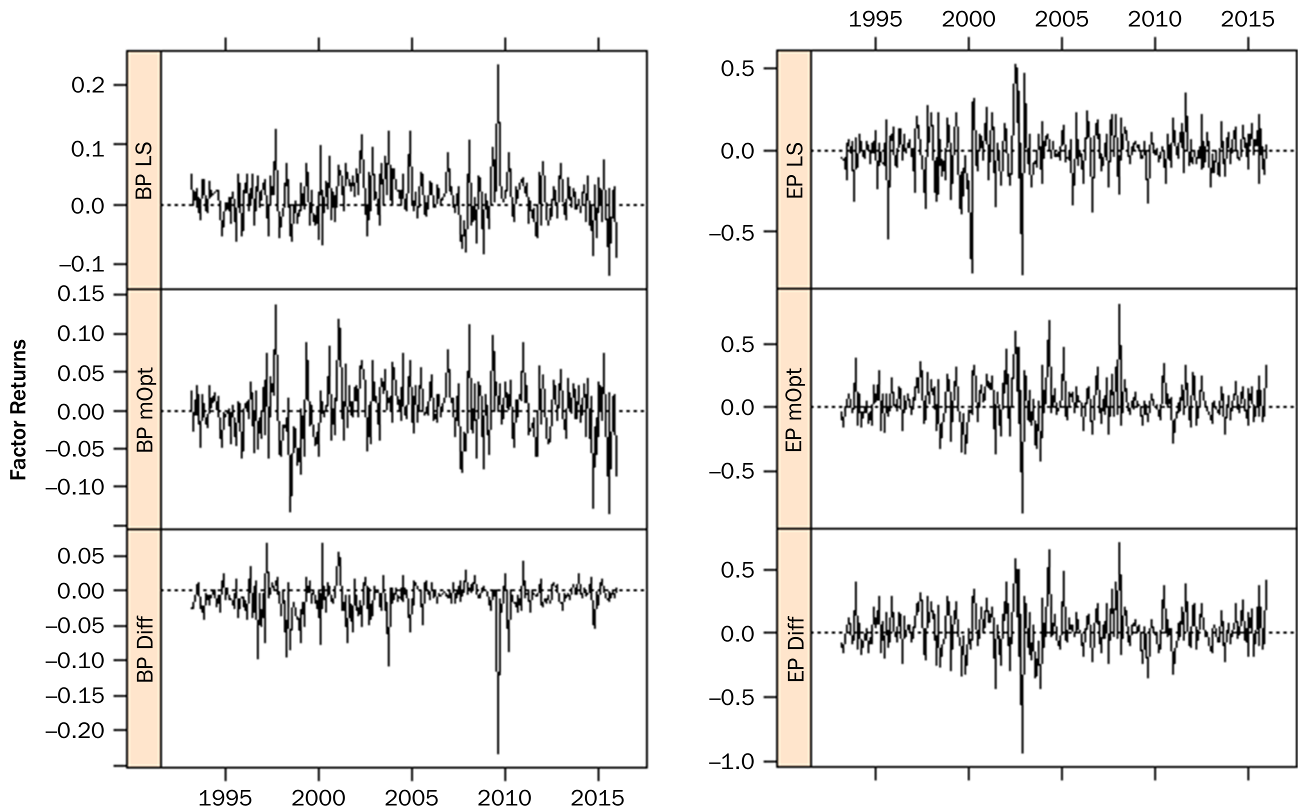
[13] See, for example, Section 5.2 of Chen and Martin (2020).

[14] These HAC corrections are done using the `lrvar` function in the `sandwich` package.

[15] Harvey, Liu, and Zhu (2016) recommended use of a threshold of 3.0 for a newly discovered factor. We use 3.0 here because robust regression is a quite new approach to empirical asset pricing, for which it is prudent to be conservative in declaring significance.

EXHIBIT 4

BP and EP Returns Computed with the Least-Squares and Robust (mOpt) Methods and the Difference (mOpt – LS)



NOTE: Univariate cross-sectional regression run at the end of each month, January 1993 to December 2015.

using LS and the robust mOpt method are provided in the first two rows of Exhibit 4, and the third row of the exhibit shows the differences between the mOpt and LS factor returns.

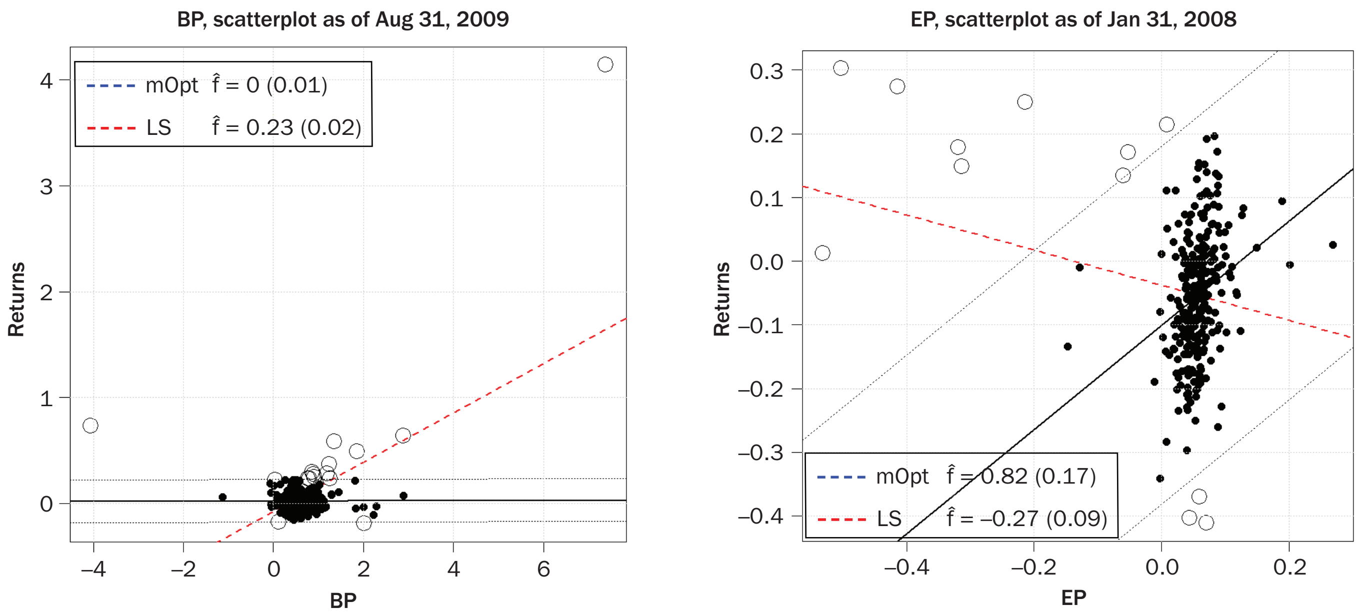
Both the BP and EP factor return differences plots exhibit time-varying volatility and substantial differences over time, the largest of which not surprisingly occurs during the 2008–2009 Global Financial Crisis. It is also to be noted that while the time-series average value of the BP factor returns appears to be close to zero, the average value of the EP factor returns appears to be positive.

To understand why the LS and mOpt factors returns can be so different in some months, we focus on the months in which the disagreement is most pronounced: August 2009 for BP and January 2008 for EP. Scatterplots of the data with superimposed LS and mOpt regression lines are provided in Exhibit 5. Both plots illustrate clearly that the reason for the different behavior of the two regression lines is the presence of outliers, which are identified by the circle symbols in the plot. These data points are the ones for which the mOpt weight function in Equation (15) is equal to zero. These results demonstrate that a theoretically justified robust regression can provide a scientific way of identifying and rejecting only those outliers that may adversely affect the estimated regression relationship.

The effect of the commonly used 1% Winsorization on the LS regression is illustrated in Exhibit 6 for the case of EP from Exhibit 5. The data points in the extreme 1% of each tail are denoted by the circle symbol, and the modified data points are denoted

EXHIBIT 5

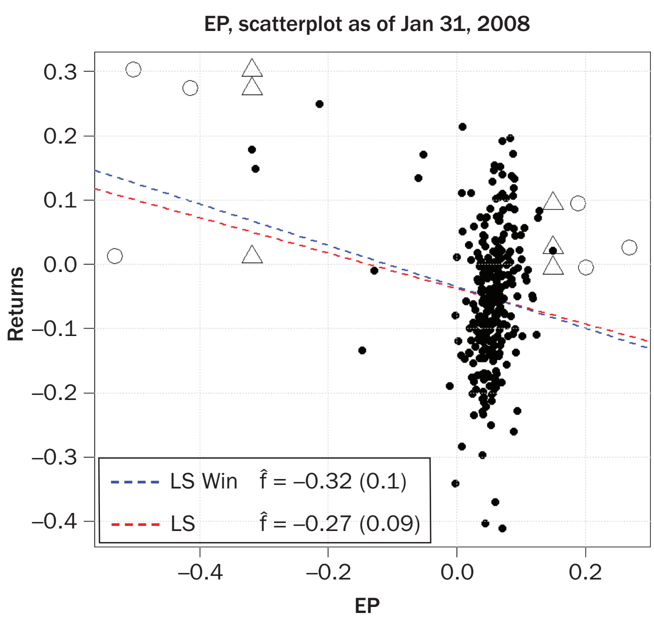
Stock Returns, BP, EP, and Estimated Regressions with the Least-Squares (LS) and Robust (mOpt) Methods



NOTES: The circle symbols indicate the outliers identified and rejected by the mOpt robust regression. The two dotted lines parallel to the regression lines define the region outside of which data points are rejected due to the mOpt weight function in Equation (15) having value 0. The estimated factor returns together with their standard errors are provided in the legends.

EXHIBIT 6

EP and Estimated LS Regression with and without Winsorization at 1%



NOTE: The circle symbols indicate 1% of the observations in each EP tail, and the triangle symbols identify the Winsorized data points.

by the triangle symbol. First, the plot demonstrates that Winsorization cannot address asymmetric outliers in an adequate manner. Second, the commonly used 1% Winsorization has a negligible impact on the estimated regression line.

We present the BP and EP HAC *t*-statistics for the mOpt and LS regressions in Exhibit 7. Panel A is based on the entire sample from 1993 to 2015, while Panels B and C are based on two subsamples—covering the 1993–2006 and the 2007–2015 periods. The reason for this separation is that the BP value factor is widely known to have stopped performing well in the years after the financial crisis of 2007–2008.

We draw several important conclusions from the results in Exhibit 7. Firstly, Panel A demonstrates that the differences between the mOpt and LS factor returns in Exhibit 4 translate into different averages over the whole time period, which lead to very different economic significance conclusions about the two factors. The LS regression leads to the conclusion that BP is a significant factor for the cross section of stock returns while EP is not, and the mOpt robust regression conclusion is the exact opposite, namely BP is not at all significant. These results for BP and EP collectively suggest that the BP LS regression is

EXHIBIT 7

HAC *t*-Statistics of the Average Returns of BP and EP Computed for January 1993 to December 2015 and Two Subperiods

	mOpt	LS	LS-Win-1%
Panel A: Full Period (1993–2015)			
BP	0.139	4.030	3.737
EP	3.597	–1.023	–0.217
Panel B: 1993–2006			
BP	0.708	5.005	4.489
EP	2.599	–1.140	–0.268
Panel C: 2007–2015			
BP	–0.774	0.963	0.978
EP	2.776	0.107	0.138

NOTES: The robust (mOpt) method and the least-squares (LS) method are shown with and without Winsorization. Strongly significant *t*-statistics are highlighted in green, while weakly significant ones are highlighted in yellow.

influenced by outliers in a way, such as in the left-hand plot of Exhibit 5, which leads to a conclusion that BP is significant, while the mOpt robust regression rejects or down-weights such outliers. The Panel B 1993–2006 period results are quite similar to those of Panel A, except that EP is now only weakly significant. The mOpt and LS results for BP in Panel C are both consistent with the well-known fact that BP was no longer a significant value factor post 2007. However, the EP *t*-statistic of 2.78 for mOpt in panel C indicates that the value character of EP continued post 2007, while in panel C the small EP *t*-statistic of –0.107 for LS is evidently due to outliers influence on the LS regression, as is suggested by the right-hand plot in Exhibit 5. The "LS Win 1%" column in the exhibit contains the HAC *t*-statistics for the case where the factor exposures for BP and EP are Winsorized at 1%, as is common in asset pricing research. The numerical results indicate that the 1% Winsorization is not at all effective in mitigating the influence of factor exposure outliers on LS regressions, as already implied by Exhibit 6.[16]

Although the single-factor model results presented so far already illustrate the extent to which the LS and mOpt factor returns may differ, in practice multivariate factor models are needed. A historically important multivariate cross-sectional model is the 4-factor model based on the beta, logmktcap (size), value (BP) and earnings-to-price ratio (EP) factors studied in the seminal Fama–French study (1992; hereafter, FF92). We recall that FF92 found that beta does not explain the cross section of returns, which gave rise to considerable controversy and subsequent studies. Furthermore, FF92 found that EP is not significant when included in a three-factor model with size and BP, which led to extensive further studies of BP as a value factor and, to a large extent, the demise of EP as a value factor. Here, we will fit that four-factor model with an intercept, which is the "alpha," using both mOpt and LS to see what emerges. In doing so, we follow the usual practice of estimating beta with a 60-month rolling window. At the end of each month, the mOpt and LS cross-sectional regressions produce an alpha and a factor return for each of the four factors, estimated jointly to account for the correlations between the corresponding factor exposures.

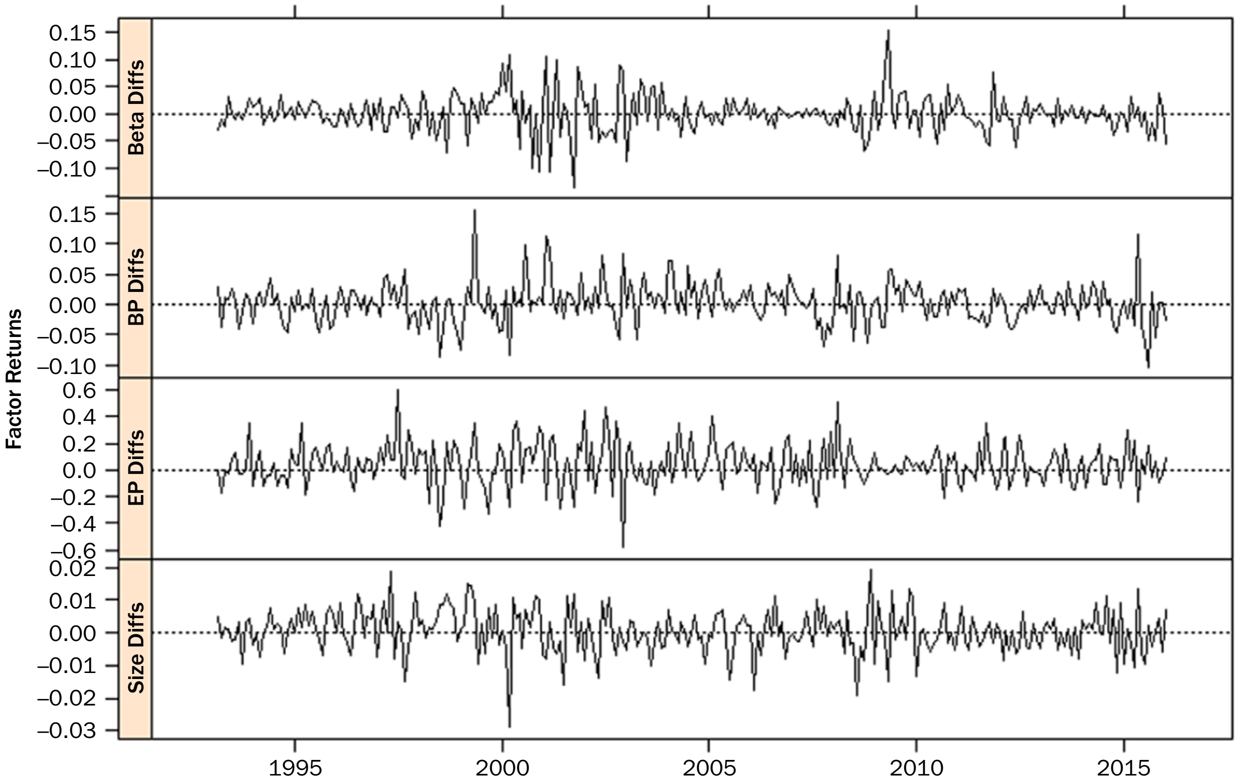
Time series of the factor-return differences—mOpt factor returns subtracted from LS factor returns—are presented in Exhibit 8. The time-series average values of beta and BP differences between the mOpt and LS factor returns appear to be close to zero, but those for EP and size appear to be positive. Our significance test results confirm these appearances, but we shall see that the reasons for the EP and size positive time-series average factor returns are different.

The HAC *t*-statistics for the mOpt, LS and LS-Win-1% results are provided in Exhibit 9. The LS and LS-Win-1% results for the 1993–2006 first period are quite consistent with the FF92 1963–1990 study results for all four factors, with beta and EP quite insignificant, a significantly positive BP premium, and a significantly negative

[16] We also Winsorized the data at 3% and 5% to see if this would lead to more effective outlier control when followed by LS, but these larger Winsorization percentages resulted in only very small differences in the *t*-statistic values relative to the values in the LS-Win-1% column of Exhibit 7.

EXHIBIT 8

Factor Return Difference Estimated with the Beta, Size, BP, and EP Multivariate Four-Factor Model, Using mOpt and LS Regressions Monthly (February 28, 1993–December 31, 2015)



NOTE: The differences are computed by subtracting the LS factor returns from the mOpt factor returns.

EXHIBIT 9

HAC *t*-Statistics of the Average Returns of the Four Factors behind FF3 Augmented with EP, Full Sample and Two Subperiods

	Full Sample			1993–2006			2007–2015		
	mOpt	LS	LS-Win-1%	mOpt	LS	LS-Win-1%	mOpt	LS	LS-Win-1%
Alpha	–0.286	3.082	3.092	–0.965	2.470	2.425	1.023	1.916	2.038
Beta	0.413	1.317	1.315	0.720	1.422	1.467	–0.232	0.318	0.220
BP	1.002	3.321	2.870	1.859	4.350	3.752	–0.785	0.400	0.374
EP	4.280	–0.291	0.243	3.443	–0.735	–0.070	2.614	0.933	1.038
Size	1.107	–2.712	–2.773	1.487	–2.220	–2.219	–0.397	–1.617	–1.790

NOTE: Computed in the full sample and two subsamples with the mOpt method and the least-squares method, with and without Winsorization at 1%.

size premium.[17] And once again, the 1% Winsorization has little effect on the results. By way of sharp contrast, for the mOpt regressions, only EP has a significant positive premium, strongly so for the first period and the entire period and weakly so for the

[17] The same is true for the entire 1963–2015 period, evidently due to the fact that the first 14-year period has more influence than the shorter 9-year second period, but the LS factor premiums all disappear in the 2007–2015 period.

second period. This is consistent with the apparent positive average of the size mOpt versus LS factor returns in Exhibit 8. However, the apparent positive time-series average of the mOpt minus LS factor returns is because size is not at all significant for mOpt while for LS it has a significant negative premium.

The robust mOpt factor returns provide a much different perspective than the LS results. In contrast to the LS case and similar to the single-factor model analysis, the EP factor is significant in the full sample and the two subsamples while BP is consistently insignificant. Even though our data set is quite limited, Martin and Xia (2021) reached similar conclusions using the much larger time-varying cross sections of all US stocks (except financials) in the CRSP database. Furthermore, similar to the univariate case in Exhibit 7, Winsorizing the factor data at 1% for the four-factor model does not change the least-squares conclusions.

In summary, the empirical examples in this section demonstrate that least-squares cross-sectional analysis can be significantly adversely influenced by outliers, which are known to be present in cross-sectional data. Instead of employing ad hoc methods to mitigate the issue, such as excluding micro-caps or using Winsorization, researchers can adopt robust regression techniques, which are based on solid statistical theory. Furthermore, they are readily available in different modeling environments and are easy to use. For example, the open source R packages PCRA, RobStatTM, and factmodCS used for the computations leading to the exhibits in this article provide a wide range of tools for robust modeling and analysis of time-series and cross-sectional regression data and are freely available online at CRAN (https://cran.r-project.org/).

With regard to applications of robust cross-sectional regression in the context of mean–variance portfolio optimization and performance analysis, pioneering research on the topic was initiated early on in Bloch et al. (1993) and was continued in more recent years by Guerard, Markowitz, and Xu (2015), Guerard (2016), and Guerard, Xu, and Markowitz (2021). These studies focused on a particular type of robust regression "bi-square" weight function proposed by Beaton and Tukey (1974). Consistent with our results here and in Martin and Xia (2021), these articles show that EP is a statistically significant factor whose contribution to portfolio performance is greater than that of BP. The important point about the Tukey bi-square weight function is that it is derived from a bounded rho function whose shape is similar to that of the mOpt rho function in Exhibit 2. Current research is under way to investigate the relative performance of mean–variance optimal portfolios based on robust cross-sectional factor model fits with mOpt versus the bi-square weight functions.

Robust Time-Series Factor Models[18]

A time-series factor model for a single asset with returns r_t has the form

$$r_t = \beta_0 + f_{1,t}\beta_1 + f_{2,t}\beta_2 + \cdots + f_{K,t}\beta_K + s\epsilon_t, \quad t = 1,2,\cdots,T \tag{17}$$

where the $f_{k,t}$ are *known* random factor returns, the β_k are *unknown* regression coefficients, the ϵ_t are zero mean regression errors that are uncorrelated with the factor returns, and s is an unknown scale value.[19] Examples of such time-series

[18] The material in this section is based on parts of Martin and Xia (2022), which was published under the Open Source Creative Commons License and is freely available online at https://link.springer.com/content/pdf/10.1057/s41260-%20022-00258-0.pdf.

[19] With $\tilde{\mathbf{f}}_t = (1, \mathbf{f}_t)$, where $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots f_{K,t})$, t = 1, 2, …, T is the vector of factor returns, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)'$ is a $K + 1$, the time-series factor model has the compact form $r_t = \tilde{\mathbf{f}}_t\boldsymbol{\beta} + s\epsilon_t$.

factor models include the capital asset pricing model (CAPM), where $K = 1, r_t = r_t^e$ is the asset return in excess of a risk-free rate, and $f_{1,t} = r_{M,t}^e$ is the market return in excess of a risk–free rate; the Fama–French three-factor model, where $K = 3$ with $f_{1,t} = r_{M,t}^e$, $f_{2,t}$ the *small-minus-big* (SMB) factor return and $f_{3,t}$ the *high-minus-low* (HML) factor return. The Fama–French–Carhart four-factor model (FFC4) adds the Carhart (1997) momentum factor to the FF3 model, and other multifactor models including the Fama and French (2015) five-factor model (FF5) and the Hou, Xue, and Zhang (2015) *q*-factor model.

Robust single-factor model fitting with mOpt. Single-factor models are used extensively by practitioners in the context of active management relative to a benchmark and by financial data service providers to compute least-squares CAPM beta estimates. We first show with several striking examples, using two-year intervals of weekly returns and market returns, that LS beta estimates are often considerably biased by outliers that have little influence on the mOpt estimates. Then, we describe the extent to which LS and mOpt beta estimates differ in absolute value and the ratio of positive differences to negative differences for the cross section of liquid stocks in the CRSP database, over two-year contiguous intervals of weekly returns from 1963 to 2018. Finally, we show that the well-known Huber robust regression estimator is subject to outlier-induced bias that is sometimes as bad as that of LS, and thus it cannot be recommended as a replacement for the mOpt estimator.

Exhibit 10 displays the LS and robust mOpt straight-line fits of weekly stock returns to market returns for the four stocks with tickers EDS, OFG, WTS, and DD, using for each the two-year interval displayed in the titles of each of the four panels. For the stocks with tickers EDS, OFG, and WTS, there are quite substantial absolute differences of 0.77, 2.26, and –0.60 between the mOpt and LS slopes, whose values are shown in the legends along with their standard errors. The significance of these differences are also conveniently described by the multiplication factors $(\hat{\beta}_{LS} - \hat{\beta}_{mOpt}) / SE(\hat{\beta}_{mOpt})$, where $SE(\hat{\beta}_{mOpt})$ is the standard error of the mOpt slope estimate, and these factors for EDS, OFG, and WTS have the very large values 5.50, 9.40, and –5.45, respectively.

Conversely, for the stock with ticker DD in the lower right of Exhibit 10, the mOpt and LS slopes of 1.21 and 1.19, respectively, are virtually identical in spite of the huge outlier in both the market and the OFG returns in the lower left of the plot. The joint DD return and market return in the lower left of the plot occurred on October 20, 1987, the famous Black Monday market crash. The fact that the mOpt and LS slopes are in close agreement is indicative of the following attractive property of the mOpt estimator: When a joint stock return and market return outlier are consistent with an LS straight line fit to rest of the data, they are not down-weighted by the mOpt estimator, and the LS and mOpt slopes agree quite closely.

The distribution of LS and mOpt differences in micro-cap, small-cap, and big-cap groups and in the market overall. In order to partially characterize the cross-sectional distribution of LS and robust mOpt Betas for the cross sections of liquid stocks in the CRSP database over the 28 contiguous two-year segments from 1963 to 2018, we tabulated the percentages of stocks in each segment for which the LS beta minus the mOpt beta is greater in absolute value than 0.3 and 0.5. Here we are assuming that absolute differences of at least 0.3 will be of concern to many investors, and absolute differences of 0.5 will be of concern for all investors. The Cap Group Percentage columns in Exhibit 11 give the average of the percentages across the 28 two-year segments for the micro-cap, small-cap, and big-cap groups and the market, where big-caps consist of mid-caps, large-caps, and mega-caps. Following Fama and French (2008) and others, the two breakpoints for the three cap groups are the 20%

EXHIBIT 10

mOpt and LS Single-Factor Model Fits of Weekly Returns to Market Returns for Stocks with Tickers EDS, OFG, WTS, and DD

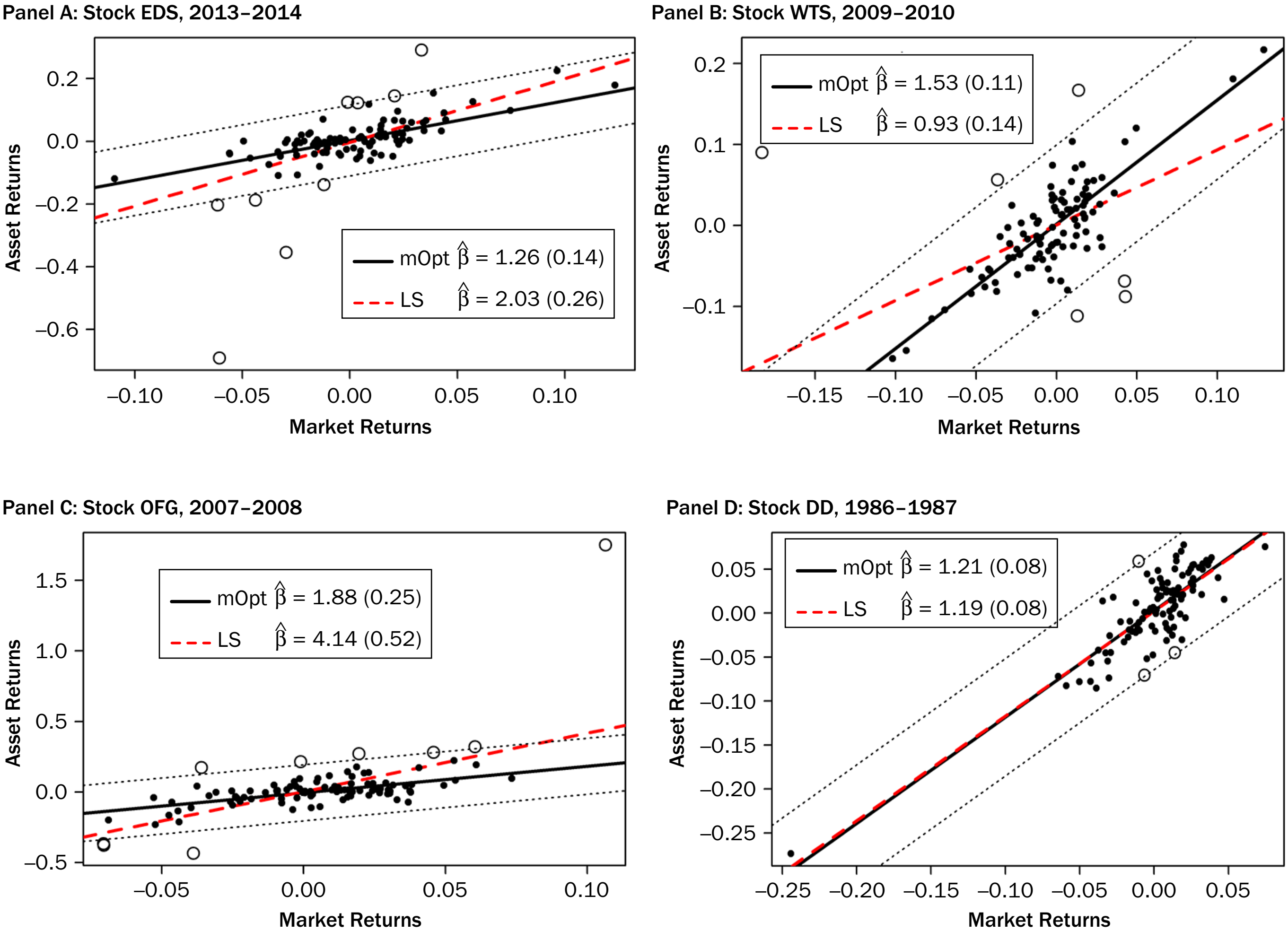


EXHIBIT 11

Average Percentage of Stocks in Market-Cap Groups and in the Market for Which the Absolute Differences of LS and mOpt Betas Exceed Thresholds of 0.3 and 0.5, along with Ratios of Positive to Negative Exceedances, 1963–2018

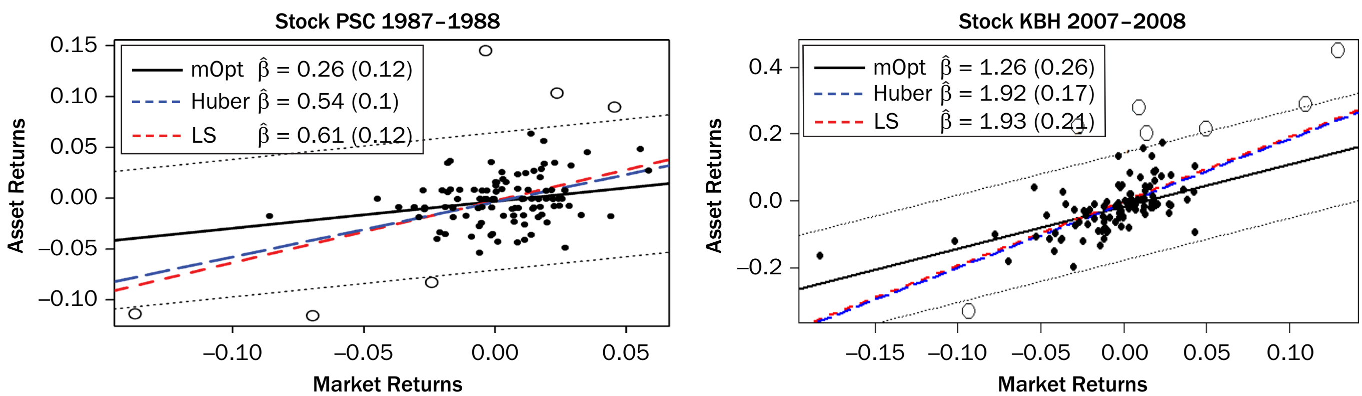
	Cap Group Percentage				Positive/Negative Exceedances Ratio			
Threshold	Micro	Small	Big	Market	MicroPN	SmallPN	BigPN	MarketPN
0.3	26.2	14.1	6.8	18.1	3.3	3.2	2.7	3.8
0.5	11.9	4.7	1.8	7.5	3.8	4.2	3.5	3.2

and 50% breakpoints of market-cap values for the NYSE. While the percentage of differences between LS and mOpt increases with decreasing market capitalization, the percentages for small cap, and even possibly big cap, should be of concern to many portfolio managers and investors.

The Positive/Negative Exceedances Ratio columns reveal that positive differences of LS betas and mOpt betas occur about three to four times as often as negative differences. This is likely due in part to using arithmetic returns, and it would be interesting to see how these ratios change with logarithmic returns, which are generally better for reporting purposes.

EXHIBIT 12

mOpt, LS, and Huber Single-Factor Model Fits of Weekly Returns to Market Returns for Stocks with Tickers PSC and KBH



The Huber regression estimator lack of bias robustness. The Huber robust regression estimator, whose loss function rho is shown in the left-hand plot of Exhibit 2, is quite well known and has been gaining popularity in some of the machine-learning literature. This is partly because of the seminal min-max variance robustness property established for estimation of a location parameter by Huber (1964) and which was inherited by the Huber (1973) regression M-estimator. It is also due in part to the fact that the Huber regression optimization is a convex optimization, which the mOpt regression is not. However, it turns out that the Huber regression estimator can suffer from arbitrarily large biases for distributions *H* in a regression version of the Tukey–Huber model given by Equation (3) that generates joint stock and market returns outliers.

This fact is illustrated in Exhibit 12, which illustrates that for some data sets with market outliers the Huber regression estimator can be as bad as the LS estimator. It is the vulnerability to *leverage outliers* that causes the Huber estimate to be so close to the LS estimate in the exhibit, where a leverage outlier is taken to be a market return and stock return pair for which the market return is extreme and the data pair is not consistent with a robust fit to the bulk of the data. In this regard, the October 20, 1987, market return in the lower left of Panel D of Exhibit 10 is a market return outlier, but it is not a leverage outlier because the market return and stock return pair is consistent with the robust fit to the bulk of the data.

Of course, there are also cases where the Huber estimator beta will be in between that of the mOpt and LS betas, and sometimes closer to the mOpt beta than the LS beta. Being curious to better understand the behavior of the Huber beta relative to the mOpt and LS betas, Martin and Xia (2022) determined the percentage of liquid stocks for which the Huber beta is closer to the LS beta than the mOpt beta in each of 28 contiguous two-week intervals of weekly returns from 1963 to 2018.[20] The percentage ranges from a minimum of 34% to a maximum of 47%, which we

[20] We took liquid stocks to be all those in the CRSP database that have at least 100 nonzero returns in a two-year interval. The number of stocks in the cross section ranged from roughly 1,500 in 1963 to a maximum of nearly 7,000 in 1999, and decreased to about 5,000 in 2018. Most of the increase before 1999 and decrease thereafter was due to the growth and decline in the number of micro-caps.

regard as a substantial lack of robustness of the Huber estimator. Consequently, the Huber estimator should only be used with full recognition of this limitation, and we strongly recommend the mOpt estimator instead when considering a robust regression estimator in practice.[21]

Multivariate time-series factor models robust fits and robust model selection. Here we will focus on robust mOpt versus LS fitting of the FF3 and FFC4 factor models, the latter of which has the form:

$$r_t^e = \alpha + f_{\mathrm{MKT},t}^e \beta_1 + f_{\mathrm{SMB},t} \beta_2 + f_{\mathrm{HML},t} \beta_3 + f_{\mathrm{MOM},t} \beta_4 + \epsilon_t, \tag{18}$$

where r_t^e is a time series of the asset excess returns relative to a risk-free rate, $f_{\mathrm{MKT},t}^e$ is a time series of market excess returns, $f_{\mathrm{SMB},t}$ are the returns of the Fama–French "small minus big" (SMB) size factor portfolio, $f_{\mathrm{HML},t}$ are the returns on the "high minus low" (HML) value factor portfolio, $f_{\mathrm{MOM},t}$ are the returns on the Carhart (1997) momentum factor portfolio, and α is an intercept. The FF3 model is obtained by dropping the term $f_{\mathrm{MOM},t}\beta_4$.[22]

It is common practice to use time-series factor models such as the FF3 and FFC4 models for analysis where the r_t represents the returns of a fund or other portfolio of interest. However, these models can in principle be used for individual stocks with returns r_t, where for example the goal might be to select stocks that have desired exposures, or lack thereof, to one or more of the factors. Here, we examine the mOpt and LS fits of the FF3 and FFC4 models to the weekly excess returns of the stock with ticker FNB for the year 2008. Exhibit 13 displays the pairwise scatterplots of the FNB, MKT, SMB, HML, and MOM data. Given the turbulence of the 2007–2008 market crisis, it is not surprising to see bivariate outliers, some of which may also be 3- or 4-dimensional outliers.

The coefficients of the LS and mOpt fits of the FF3 and FFC4 models to the FNB returns are displayed in Exhibit 14, along with the (t-statistics), and the adjusted R-squared values.[23] Here, we use the usual absolute t-value threshold of 1.96 for a t-statistic to be significant, and such t-stat values are highlighted in green. For both models, the intercept estimates of both the LS and mOpt fits are very close to zero, and the t-statistic values are all quite insignificant. The main differences in FF3 LS and mOpt model fits are that the mOpt fit has a slightly higher robust adjusted R-squared, and for the HML factor the mOpt slope is a little over twice that of the LS slope and, correspondingly, the mOpt t-statistic value is considerably larger than that of the LS estimate.

For the FFC4 model, the LS fit t-statistics for the HML and MOM factors are now both insignificant, leaving only the MKT and SMB as significant factors, but the mOpt fit t-statistics are significant for all but the HML factor, whose t-statistic is quite insignificant, and its MOM slope is negative. This latter result is not surprising in view of the fact, illustrated in Exhibit 13, that the HML and MOM factors are negatively correlated and MOM replaces HML in the FFC4 model. Exhibit 14 also reveals that the FFC4 mOpt fit has the highest adjusted R-squared among the LS and mOpt fits of both the FF3 and FFC4 models, and overall, the results in the table lead to the sense

[21] Further details concerning the lack of bias robustness of the Huber estimator may be found in Martin and Xia (2022). Also, the lack of bias robustness of any regression M-estimator with an unbounded rho function was established in Martin, Yohai, and Zamar (1989).

[22] The time series of the SMB, HML and MOM factors in the model Equation (18) are available at Professor Ken French's Data Library: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[23] These quantities are all robust for the mOpt estimator. For the robust adjusted R-squared, see Section 5.12.3 of Maronna et al. (2019).

EXHIBIT 13

Pairwise Scatterplots of FNB 2008 Weekly Returns and Corresponding MKT, SMB, HML, MOM Factors

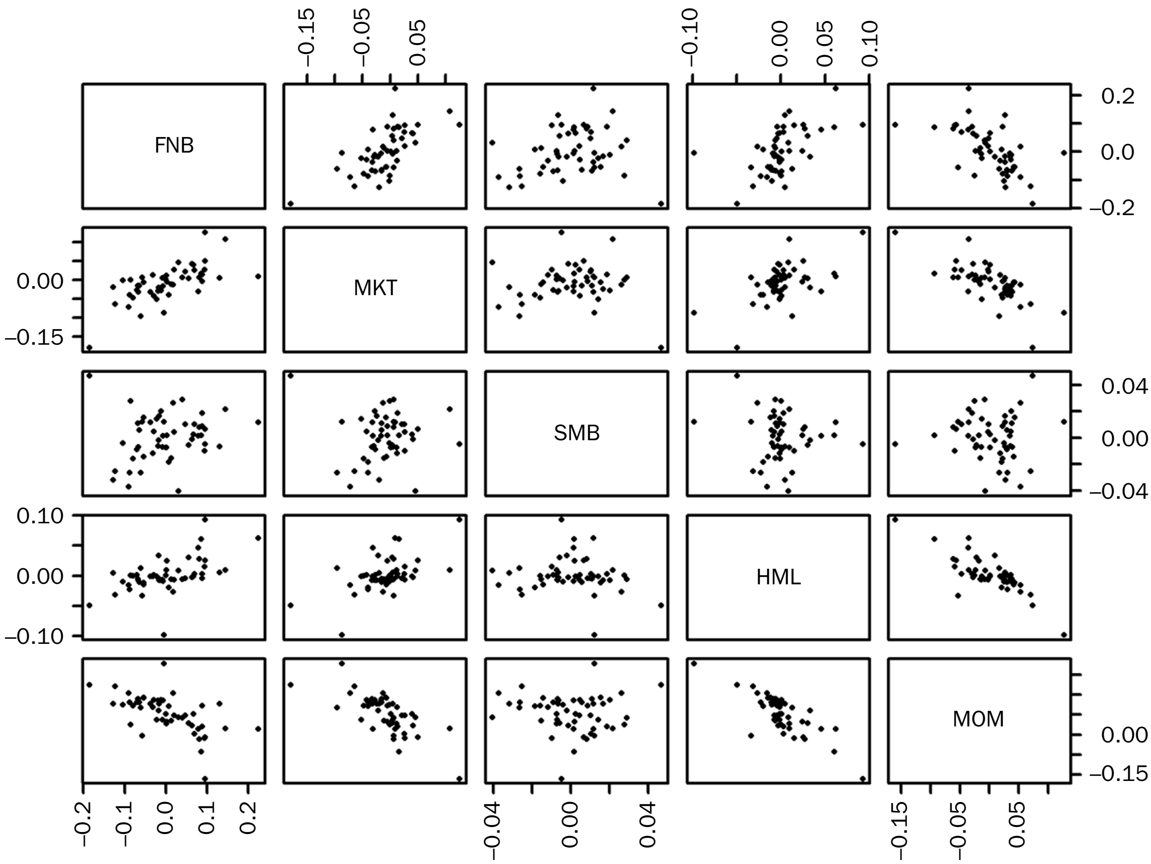


EXHIBIT 14

mOpt and LS Fits of FF3 and FFC4 Models to FNB 2008 Weekly Returns

	Alpha	MKT	SMB	HML	MOM	Adj. R^2
FF3-LS	0.01 (0.9)	0.83 (4.2)	0.96 (2.1)	0.83 (2.5)		0.5
FF3-mOpt	0.01 (0.7)	0.91 (4.3)	1.01 (2.1)	1.71 (4)		0.53
FFC4-LS	0.01 (0.9)	0.85 (3.5)	0.97 (2.1)	0.86 (1.8)	0.03 (0.1)	0.49
FFC4-mOpt	0.01 (1.6)	0.7 (4.5)	0.81 (2.6)	0.2 (0.4)	−0.91 (−3.3)	0.57

that the FFC4 model is preferable to the FF3 model. However, a proper robust model selection method is needed for this purpose, and we now discuss such a method.

A section of Martin and Xia (2022) is devoted to describing a robust regression model subset selection method using a *robust final prediction error* (RFPE) criterion, which was originally described by Maronna et al. (2019). The RFPE robust model selection method is defined as follows, using the robust mOpt regression estimator. For a subset of k factors in a maximal model $C_{\max}$ with K predictor variables $\tilde{\mathbf{f}}_{C_{\max},t}$, let $\hat{\beta}_{C_k}$ be the mOpt regression coefficient for a subset of $k \le K$ factors obtained as

$$\hat{\beta}_{C_k} = \arg\min_{\beta} \left[\sum_{t=1}^{T} \rho_{\text{mOpt}} \left(\frac{r_t - \tilde{\mathbf{f}}'_{C_k,t}\beta}{\hat{s}} \right) \right], \tag{19}$$

where the robust residuals scale estimate $\hat{s}$ is computed for the maximal model C_{max} with K predictor variables. This serves to provide a smallest robust scale that serves as an equitable scale reference across models with different numbers of predictor variables. With robust residuals for the subset model C_k defined as

$$\hat{\epsilon}_{C_k,t} = r_t - \tilde{\mathbf{f}}'_{C_k,t}\beta_{C_k}, \quad t = 1, 2, \cdots, T$$

RFPE is defined as

$$\mathrm{RFPE}(C_k) = \frac{1}{T}\sum_{t=1}^{T}\rho_{\mathrm{mOpt}}\left(\frac{\hat{\epsilon}_{C_k,t}}{\hat{s}}\right) + \frac{k}{T}\frac{A}{B}, \tag{20}$$

where the formulas for A and B in the penalty term $k\,A/T\,B$ are given by Martin and Xia (2022).

Ideally, one would like to choose a best model by evaluating RFPE(C_k) for all subsets of the full model C_K. But this is often infeasible due to the computing time required by a very large number of subsets unless K is rather small. However, a backward step-wise RFPE criterion selection method is feasible for the mOpt estimator and is available in the R package RobStatTM.[24] We illustrate use of the method for the problem of finding a best subset of the FFC4 factors for modeling the FNB stock returns, and the results are shown in Exhibit 15.

The Full Model columns of Exhibit 15 show the RFPE value 0.221 of the full FFC4 model in the All row, and subsequent rows show the RFPE values for each of the three-factor models obtained by deleting each of the MKT, SMB, HML, and MOM factors one at a time. Because the RFPE value 0.217 obtained by deleting the HML factor is the smallest RFPE among all three deletions and it is smaller than the full four-factor model RFPE value 0.221, this factor is deleted and results in the new All row value in the MKT+SMB+MOM columns of the table. Because none of the single-factor deletions of MKT, SMB, and MOM result in a smaller value of RFPE than 0.221, the process stops with those three factors as the best subset of the four FFC4 factors. This result is quite consistent with FFC4-mOpt row of Exhibit 14, where the HML factor is quite insignificant.

Martin and Xia (2022) also used the well-known least-squares based Akaike information criterion (AIC) model selection for the FFC4 factors, and the results are shown in Exhibit 16. The AIC method results in deleting the MOM factor, leaving the three factors MKT, SMB, and HML as the best model. This is in reasonable agreement with the LS fit of the full FFC4 set of factors to FNB that finds MOM quite insignificant and HML close to significant with a t-statistic value 1.8, but in disagreement with the RFPE best set of factors where MOM is retained instead of HML. This

EXHIBIT 15

RFPE Backward Step-Wise Model Selection from FNB ~ MKT + SMB + HML + MOM

FULL MODEL		MKT + SMB + MOM	
Factor	**RFPE**	**Factor**	**RFPE**
ALL	0.221	ALL	0.217
MKT	0.235	MKT	0.248
SMB	0.234	SMB	0.23
HML	0.217	MOM	0.26
MOM	0.23		

EXHIBIT 16

AIC Backward Step-Wise Model Selection from FNB ~ MKT + SMB + HML + MOM

FULL MODEL		MKT + SMB + HML	
Factor	**AIC**	**Factor**	**AIC**
ALL	–294.6	ALL	–296.5
MKT	–284.5	MKT	–282.6
SMB	–292.1	SMB	–293.9
HML	–293	HML	–292.4
MOM	–296.5		

[24] In particular, the function step.lmrobdetMM implements the method using the lmrobdetMM fitted model as input.

shows that AIC, like LS, suffers from the adverse influence of outliers and cannot serve as a reliable method for model selection.

While this example is quite simple, it suggests the potential usefulness of RFPE in selecting time-series factors from the large "zoo" of such factors. This is a topic for research.

ROBUST COVARIANCE MATRICES AND DISTANCES

Robust Covariance Matrices

The joint estimators $\hat{\mu}$ and $\hat{\mathbf{C}}$ of an $N \times 1$ mean vector μ and an $N \times N$ covariance matrix $\mathbf{C}$, which we discuss here and are available in the RobStatTM package, are solutions of the following equations:

$$\begin{aligned} &\sum_{t=1}^{T} \mathrm{W}\left(\frac{d^2(\mathbf{r}_t)}{c \cdot s}\right)(\mathbf{r}_t - \mu) = \mathbf{0} \\ &\frac{1}{T}\sum_{t=1}^{T} \mathrm{W}\left(\frac{d^2(\mathbf{r}_t)}{c \cdot s}\right)(\mathbf{r}_t - \mu)(\mathbf{r}_t - \mu)' = \mathbf{C} \end{aligned} \tag{21}$$

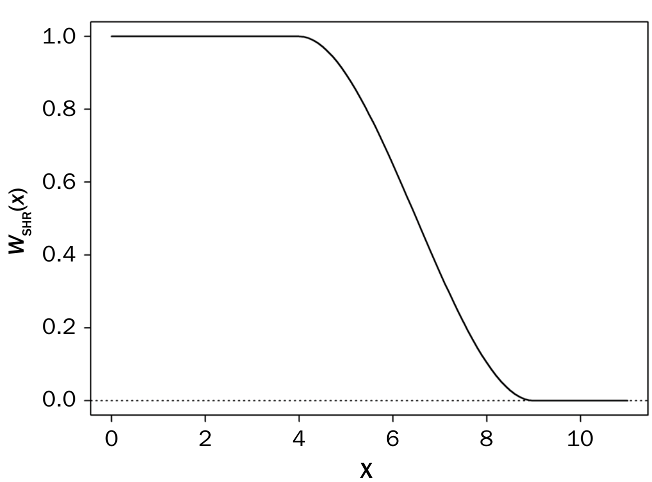
where $d(\mathbf{r}_t)$ is a *distance function* that measures the distance of the t-th return vector $\mathbf{r}_t$ from its mean $\boldsymbol{\mu}$, s is a scaling parameter for standardizing the $d^2(\mathbf{r}_t)$ values, c is a tuning constant that controls the extent of robustness toward outliers, and $\mathrm{W}(x)$, $x \geq 0$ is a scaler-valued weighting function. Thus, the estimates $\hat{\mu}$ and $\hat{\mathbf{C}}$ have the simple interpretation of being a weighted version of the usual equations for the sample mean and sample covariance matrix, to which they reduce by setting $\mathrm{W}(x) = 1$ for all x. The form of $\mathrm{W}(x)$ is chosen to increasingly down-weight increasingly large return vectors, such that $\mathbf{r}_t$ outliers are given weight 0, in which case they are said to be *rejected*. We describe the form of distance function $d(\mathbf{r}_t)$ shortly.

The tuning constant c control is a trade-off between robustness toward outliers and the normal distribution variability of the robust covariance matrix estimator. The ratio of variability of the normal distribution covariance matrix MLE to the normal distribution variability of the robust covariance matrix estimator is called the estimator *normal distribution efficiency*. The tuning constant c for the robust covariance matrix estimators depends upon the number of assets N and the sample size T, and it is computed automatically so that the estimators have a normal distribution efficiency of 90%.

The type of estimators $\hat{\mathbf{C}}$, $\hat{\mu}$, $\hat{s}$ obtained as solutions of Equation (21) that we focus on here are called *MM-Estimators*, which are designed for portfolios with less than 10 assets and which are computed with the function covRobMM in RobStatTM. Details concerning the estimate $\hat{s}$ are provided by Maronna et al. (2019). The weight function for the covRobMM function is called a *smoothed hard rejection* (SHR) function, and its shape is shown in Exhibit 17.[25]

EXHIBIT 17

Smoothed Hard Rejection Weight Function, $W_{SHR}(x)$



[25] Its formula is given in the appendix.

The following example illustrates the importance of robust covariance matrix estimation in the context of a small portfolio of four stocks, where a small fraction of outliers adversely influences classical covariance and correlation estimates, and a robust covariance matrix estimate and its associated robust correlation estimates more accurately represent the correlations of the majority of the returns.

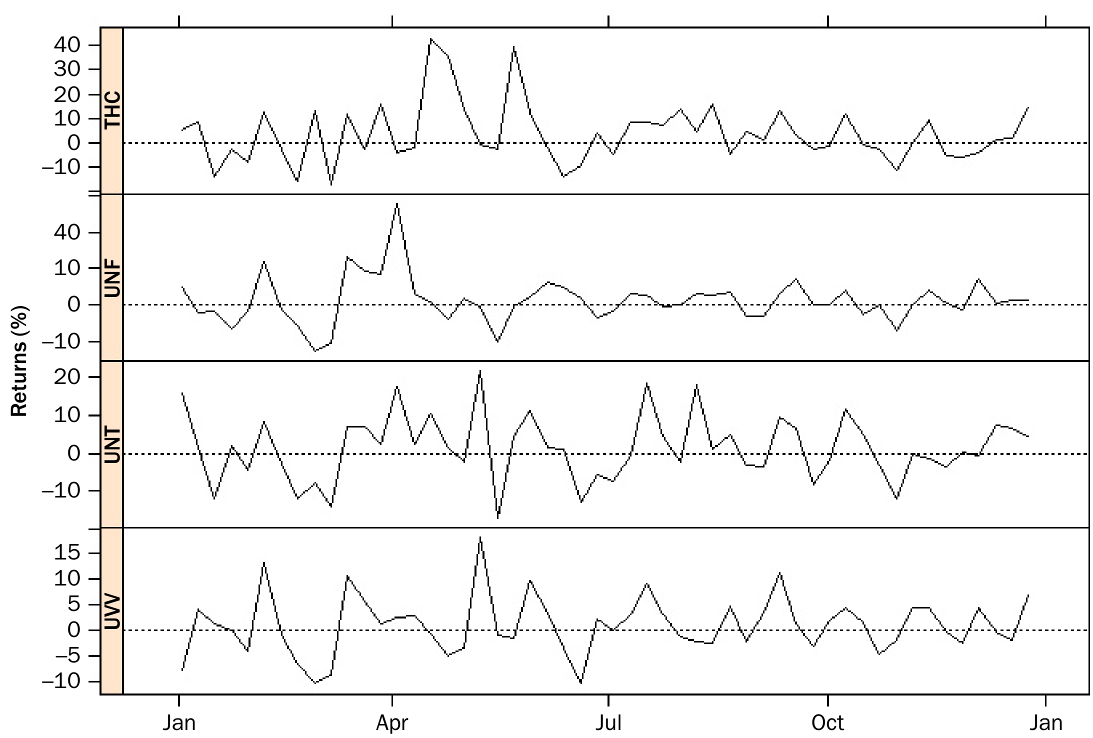
Example 5: Weekly Returns of Four Small-Cap Stocks with Influential Outliers during 2009

Exhibit 18 shows the time series of weekly returns of the four small-cap stocks with tickers THC, UNF, UNT, and UVV, from January 2, 2009, to December 25, 2009. Exhibit 19 displays the pairwise scatterplots of the returns of the four stocks, which reveals a variety of potentially influential outliers, not all of which are univariate outliers.

The lower left portion of Exhibit 20 shows the classical sample correlation estimate values 0.10, 0.39, 0.58, 0.16, 0.42, 0.55 and the corresponding robust correlation estimate values 0.47, 0.59, 0.62, 0.54, 0.50, 0.52, for the fund pairs UNF-THC, UNT-THC, UNT-UNF, UVV-THC, UVV-UNF, UVV-UNT, respectively, as extracted from the corresponding covariance matrix estimates. The robust correlation estimate values are not much influenced by the outliers in the scatterplots, and as such they represent the correlations for the bulk of the data. Conversely, the considerably smaller classical sample correlation estimates for the UNF-THC, UNT-THC, UVV-THC pairs are negatively biased by the outliers. An informative visual display of the differences between the classical and robust correlations is provided

EXHIBIT 18

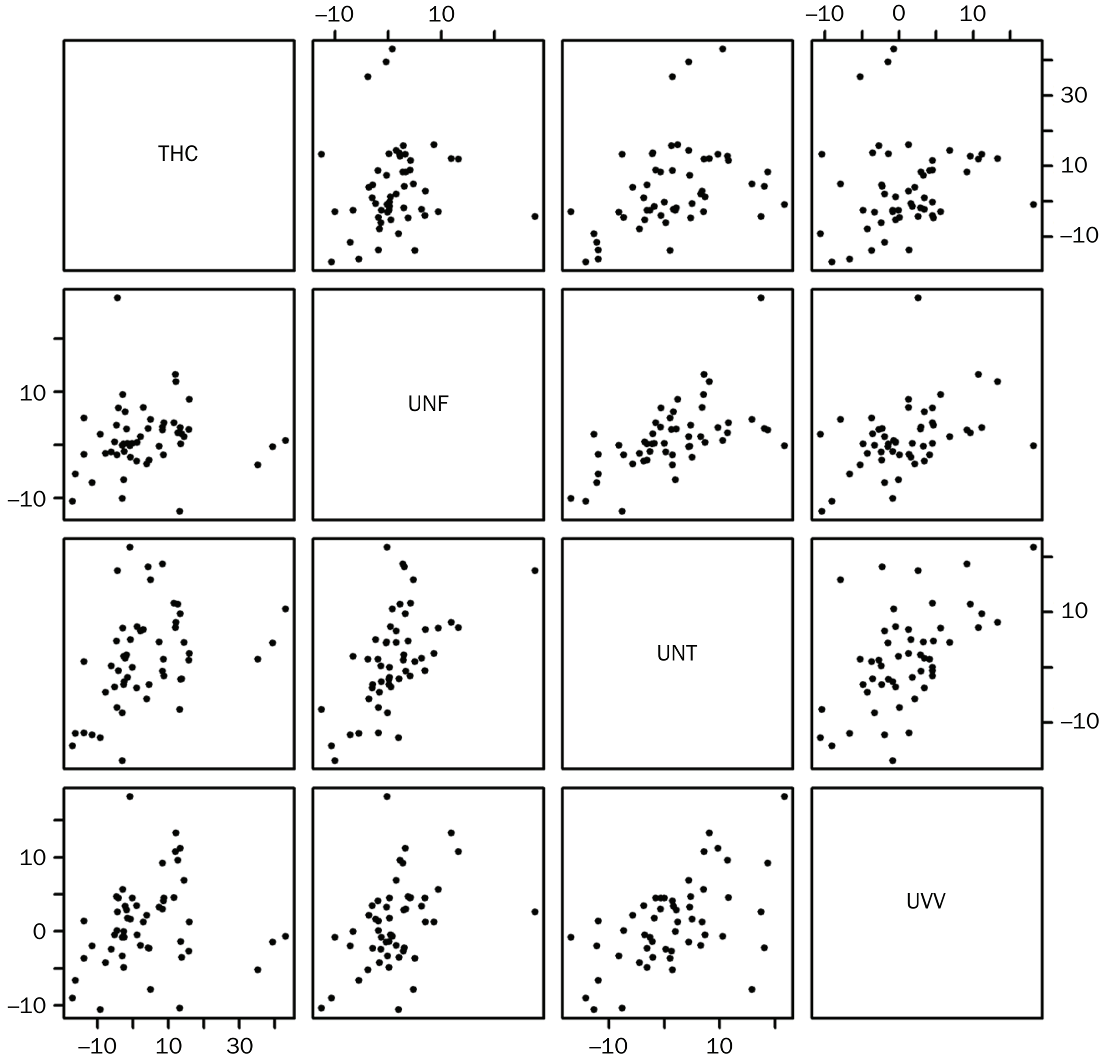
Weekly Returns of Stocks with Tickers THC, UNF, UNT, and UVV, January 2, 2009–December 25, 2009



SOURCE: Center for Research in Security Prices, LLC.

EXHIBIT 19

Pairwise Scatterplot of the Weekly Returns of Stocks with Tickers THC, UNF, UNT, and UVV, January 2, 2009–December 25, 2009



SOURCE: Center for Research in Security Prices, LLC.

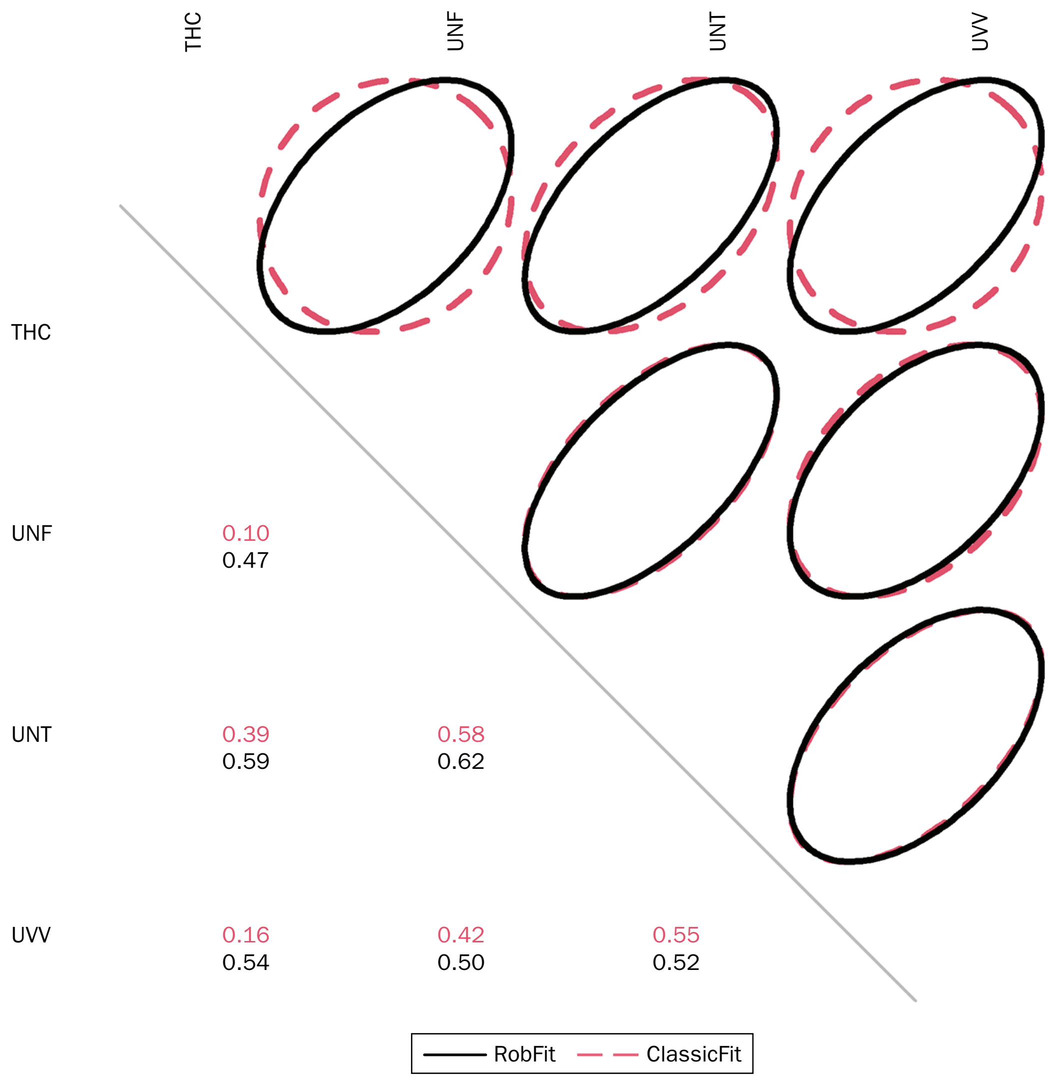
in the upper right part of Exhibit 20, which displays overlaid bivariate standard normal density ellipses corresponding to the sample correlation estimates and robust correlation estimates.[26]

In the context of mean–variance optimal portfolios, with all other things being equal, smaller correlations imply more portfolio diversification benefit than larger correlations. But in this case, the smaller classical correlation estimates for the UNF-THC, UNT-THC, UVV-THC pairs are due to the influence of a relatively small fraction of outliers that are down-weighted by the robust covariance matrix estimate. As such, the classical correlation estimates do not represent the bulk of the returns well and lead to an overly optimistic impression of the portfolio diversification benefit.

[26] The contours of constant value c of a bivariate normal density vector $\mathbf{y} = (y_1, y_2)'$ with mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{C}$ are the values $\{\mathbf{y}: (\mathbf{y} - \boldsymbol{\mu})'\mathbf{C}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = c\}$.

EXHIBIT 20

Correlation Ellipses Plots Based on Classic and Robust Covariance Matrix Estimates for Stocks with Tickers THC, UNF, UNT, and UVV, January 2, 2009–December 25, 2009



SOURCE: Center for Research in Security Prices, LLC.

By way of contrast, the larger robust correlation estimates provide a more realistic view of the available portfolio diversification benefit.

For portfolios with 10 or more assets, an alternative robust covariance matrix estimator based on a different weight function is recommended. This alternative estimator is implemented in the RobStatTM package function covRobRocke. Details are provided in Section 6.4.4 of Maronna et al. (2019).

Classic and Robust Distances

Robust covariance matrices and correlations provide clear indications on whether or not outliers are substantially influencing the classical estimates. However, they provide no information concerning the temporal locations of such outliers. In this section, we describe a *robust distance* (RD) method that clearly indicates the time location of influential multivariate returns outliers. We first review the mathematical character of an *ideal Mahalanobis distance* (idealMD) based on a known mean vector and covariance matrix, and the *classic MD* obtained from it by replacing the unknown mean vector and covariance matrix with their sample-based estimates. The classic MD was introduced in the context of measuring turbulence in financial markets by Chow et al. (1999) and, subsequently, for measuring and managing risk and to improve investment performance by Kritzman and Li (2010). The robust distance is obtained by replacing the sample mean vector and covariance matrix of the classical MD with the robust covariance matrix MM-estimator. In addition to their use in detecting the temporal location and sizes of multivariate outliers, the RD plays a key intertwined role in the computation of the robust mean vector and covariance matrix solution to Equation (21).

An *ideal Mahalanobis Distance* (MD) is defined by the square root of

$$d^2_{\text{idealMD}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})'\mathbf{C}^{-1}(\mathbf{r} - \boldsymbol{\mu}), \tag{22}$$

where the N-dimensional $\boldsymbol{\mu}$ and the $N \times N$ positive definite covariance matrix $\mathbf{C}$ are known. The ideal MD is a generalization of the ordinary N-dimensional squared Euclidean distance $d^2_{\text{Eucl}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})'(\mathbf{r} - \boldsymbol{\mu})$ of $\mathbf{r}$ from $\boldsymbol{\mu}$, such that the MD accounts for the covariance (and hence correlation) of the components r_i and r_j of $\mathbf{r}$ in a natural way.[27]

A useful way to understand this geometry is as follows. A positive definite matrix $\mathbf{C}$ can be factored into the product of its square roots: $\mathbf{C} = \mathbf{C}^{1/2}\mathbf{C}^{1/2}$, and correspondingly $\mathbf{C}^{-1} = \mathbf{C}^{-1/2} \cdot \mathbf{C}^{-1/2}$. Let

$$\mathbf{z} = \mathbf{C}^{-1/2}(\mathbf{r} - \boldsymbol{\mu}) \tag{23}$$

be an affine transformation from $\mathbf{r}$ to $\mathbf{z} = (z_1, z_2, \ldots, z_N)'$, which we refer to as the *MD transformation*. The z_i obviously have mean zero, and their covariance matrix is the identity matrix $\mathbf{I}_N$. The ideal squared MD can be written in terms of $\mathbf{z}$ as

$$d^2_{\text{idealMD}}(\mathbf{r}) = \mathbf{z}'\mathbf{z}, \tag{24}$$

which is equal to $\sum_{i=1}^{N} z_i^2$. For multivariate normally distributed r_i, the z_i are standard normal random variables, and the ideal MD has a chi-squared distribution with N degrees of freedom.

An *in-sample MD* value is obtained as the square root of the Mahalanobis squared distance given by

$$d^2_{\text{MD}}(\mathbf{r}_t) = \left(\mathbf{r}_t - \bar{\mathbf{r}}\right)' \hat{\mathbf{C}}^{-1} \left(\mathbf{r}_t - \bar{\mathbf{r}}\right), \quad t = 1, 2, \cdots, T, \tag{25}$$

where use of the subscript *MD*, rather than *idealMD*, indicates that it is the in-sample version. It is natural to also use the term *classic MD* to refer to this "in-sample MD."

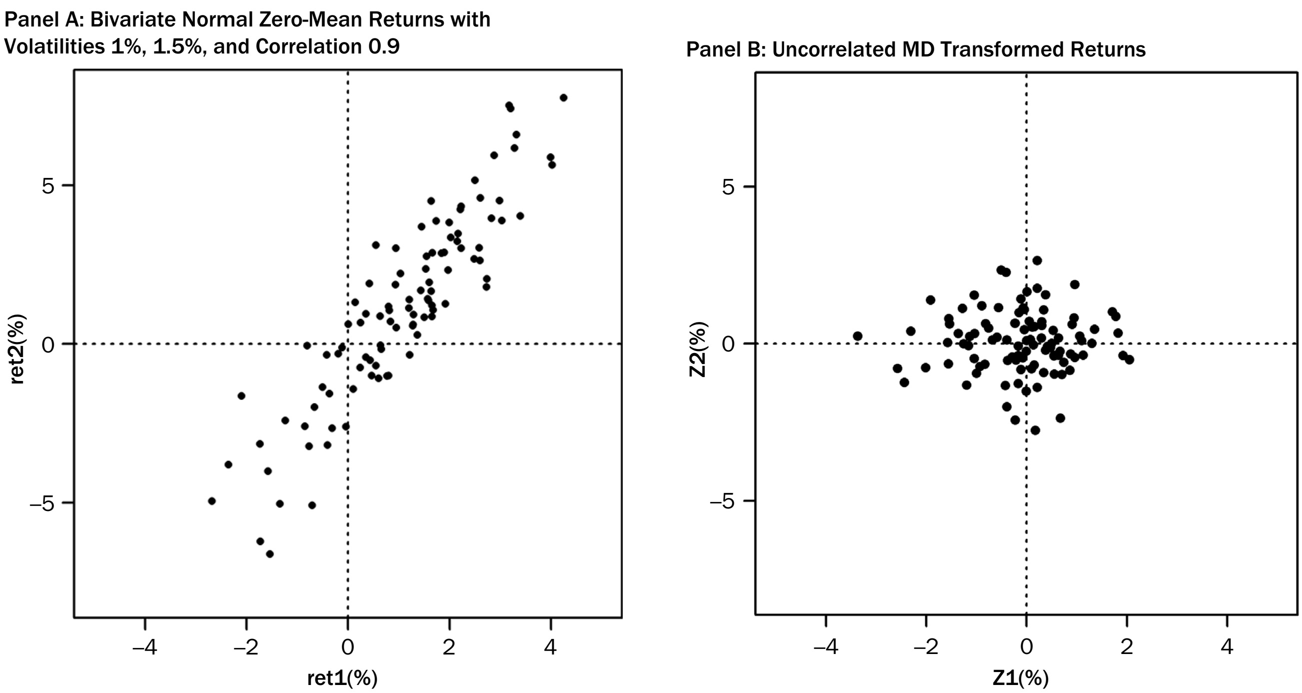
Example 6: Classic MD Transformation of Correlated Normally Distributed Returns to Uncorrelated Returns

Exhibit 21, Panel A shows a return sample of size 100 from a bivariate normal distribution with mean vector $\boldsymbol{\mu} = (1\%, 1\%)'$, standard deviation vector $\boldsymbol{\sigma} = (1.5\%, 3.0\%)'$, and correlation coefficient $\rho = 0.9$. The sample mean vector is $\hat{\boldsymbol{\mu}} = \bar{\mathbf{r}} = (1.10, 1.21)$, the sample correlation coefficient is 0.91, and the sample variances are 2.26 and 0.62.

[27] All vectors $\mathbf{r}$ that lie on the N-dimensional ellipsoid defined by the equality $(\mathbf{r} - \boldsymbol{\mu})'\mathbf{C}^{-1}(\mathbf{r} - \boldsymbol{\mu}) = d^2$ have the same distance d from $\boldsymbol{\mu}$.

EXHIBIT 21

MD Transformation of Correlated Returns to Uncorrelated Returns



The MD transformed data, $\mathbf{z}_t = (z_{1t}, z_{2t})$, shown in Panel B are obtained from $\mathbf{r}_t$ using Equation (23) with $\mu = \hat{\mu}$ and $\mathbf{C}^{-1/2} = \hat{\mathbf{C}}^{-1/2}$.[28] The sample correlation coefficient for the MD transformed data has the nearly zero value of 0.006, as one expects. The main features of the MD transformed data are that they are centered very close to the origin and have a circular shape consistent with the sample correlation value of 0.006. These features extend to the higher dimensional case, where the circular shape is nearly an N-dimensional spheroid.

We have seen earlier in this section that the classical sample covariance matrix estimator can be adversely influenced by outliers. It follows that the same is true of the Mahalanobis classic MD, which consequently will not reliably measure the extent to which one or more returns are outlying. The solution to this problem is to calculate a robust MD whose squared distances

$$d^2_{\mathrm{ROB}}(\mathbf{r}_t) = \left(\mathbf{r}_t - \hat{\mu}\right)^T \hat{\mathbf{C}}^{-1} \left(\mathbf{r}_t - \hat{\mu}\right), \quad t = 1, 2, \cdots, T \tag{26}$$

are based on the robust estimates $\hat{\mu}$ and $\mathbf{C}$ obtained as solutions to Equation (21). As the following example vividly illustrates, the classic MD can substantially fail to detect multivariate outliers, which the robust distance reliably detects.

Example 7: Robust Distances versus Classical Distances for the Four Stocks in Exhibit 18

Exhibit 22 displays the classic MDs in the left-hand panel, and the RDs in the right-hand panel, for the portfolio of the four stocks whose returns are displayed in

[28] The R function solve is used to compute $\hat{\mathbf{C}}^{-1}$, and then the tensr package function mhalf is used to compute the inverse square root matrix.

EXHIBIT 22

Classic and Robust MDs for the Four Stocks with Tickers THC, UNF, UNT, and UVV

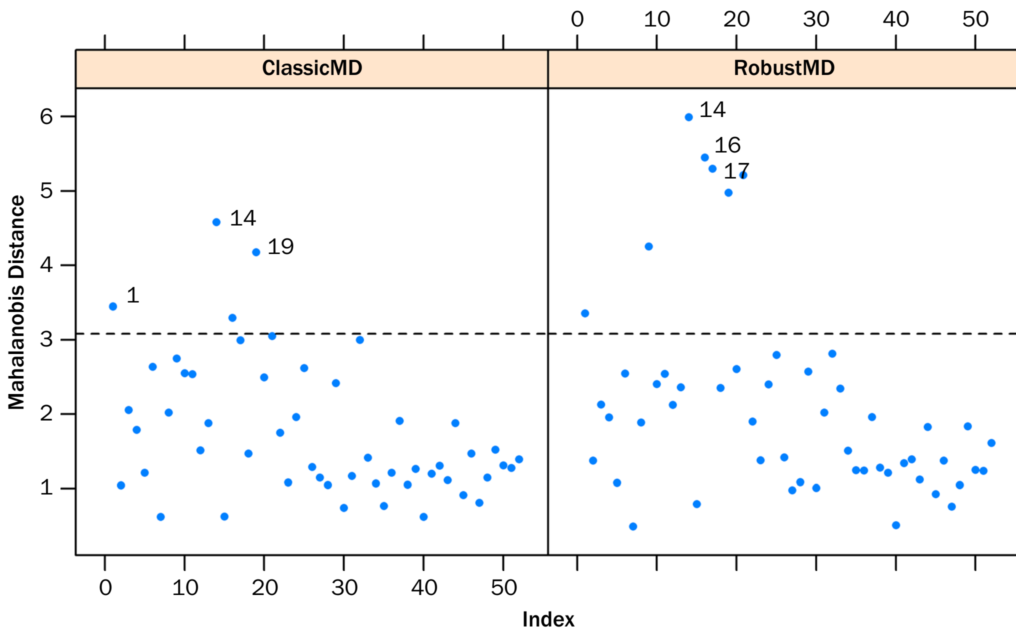


Exhibit 18. The dashed horizontal lines in the two panels are located at the square root of the upper 1% tail probability $q_{chisq,4}(0.99)$ of a chi-squared distribution with $\nu = 4$ degrees of freedom. Due to the outlier distortion of the sample covariance matrix evident in Exhibit 20, the classic MDs in Exhibit 22 detect only four outliers (two of them marginally so), while the RDs in the right-hand panel detect seven outliers (one of them marginally so). Furthermore, a careful comparison of Exhibit 22 and Exhibit 18 shows that the five largest robust distances comprise a time cluster of outliers during five of the six consecutive weeks in April and the first two weeks of May, which could be examined more carefully to see if any economic meaning can be attributed.

Note that in Exhibit 22 we have computed horizontal dotted-line thresholds based on an incorrect assumption that classical MD and RD both have an exact chi-squared distribution, which is not correct. However, it is common practice to use the chi-squared distribution approximation.

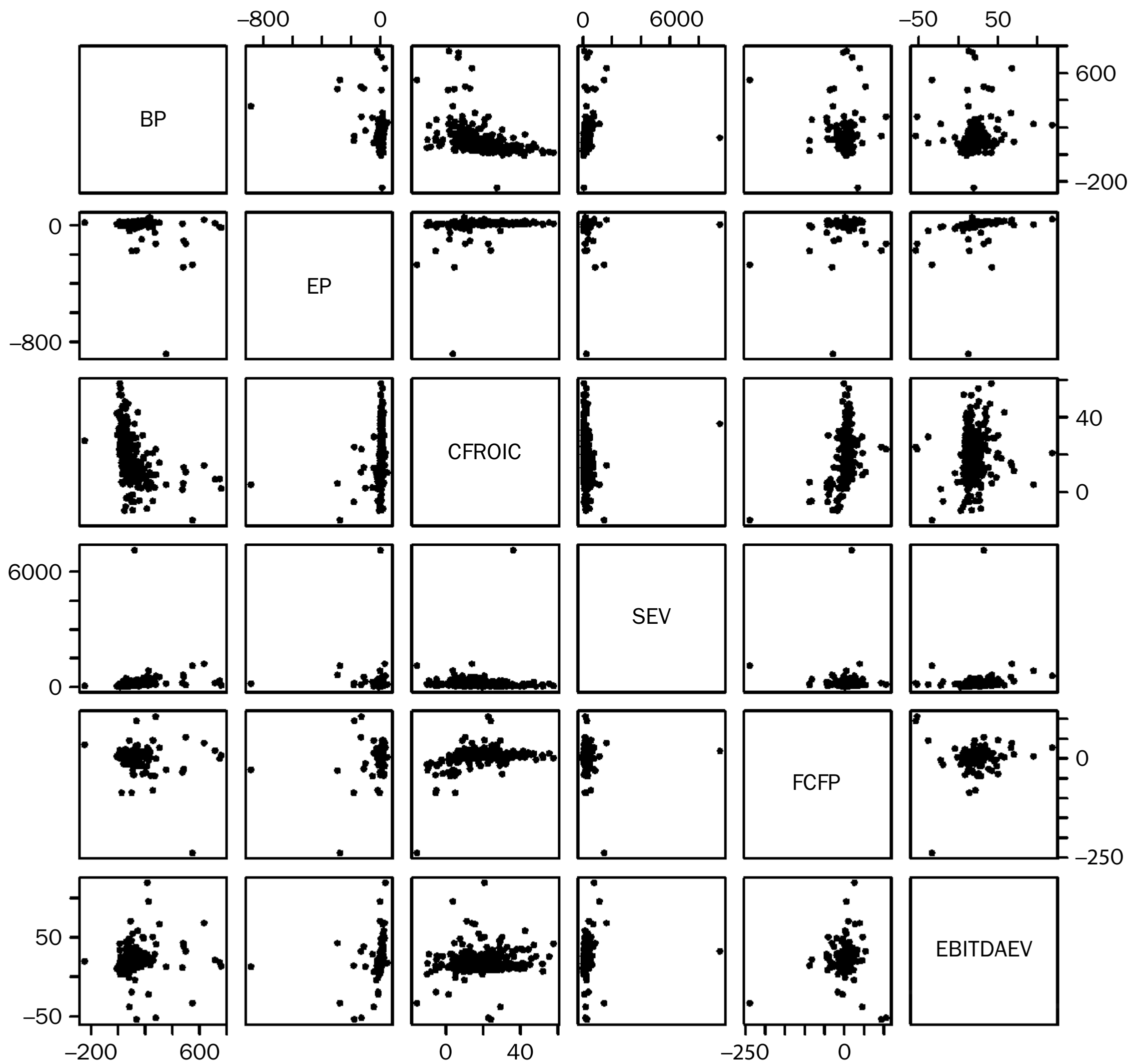
Example 8: Six-Dimensional SPGMI Factor Exposures Outliers

Factor exposures data are notorious for being "dirty" in the sense of being non-stationary and having missing data and outliers, some of which are quite huge, and thereby result in a challenging factor data cleaning problem. Here, we illustrate the outliers problem for the six factors BP, EP, CFROIC, SEV, FCFP, and EBITDAEV in the factorsSPGMI data set for the month of January 2009, using the same kind of pair-wise scatterplots, ellipses plots, and distance plots used previously for the portfolio of four stocks. The definitions of these factors are provided in the factorsSPGMI Help file in the PCRA package.

Exhibit 23 reveals a variety of scatterplot shapes and outliers configurations, and some truly huge outliers—for example, one somewhat less than –800 for EP and one greater than 6,000 for SEV. In standard industry practice, such outliers would be shrunk in a one-dimensional way, for example, using the median-based outlier shrinker defined by Equations (1) and (2). However, multidimensional outliers that do not reveal themselves as one-dimensional outliers also need to be down-weighted, and because

EXHIBIT 23

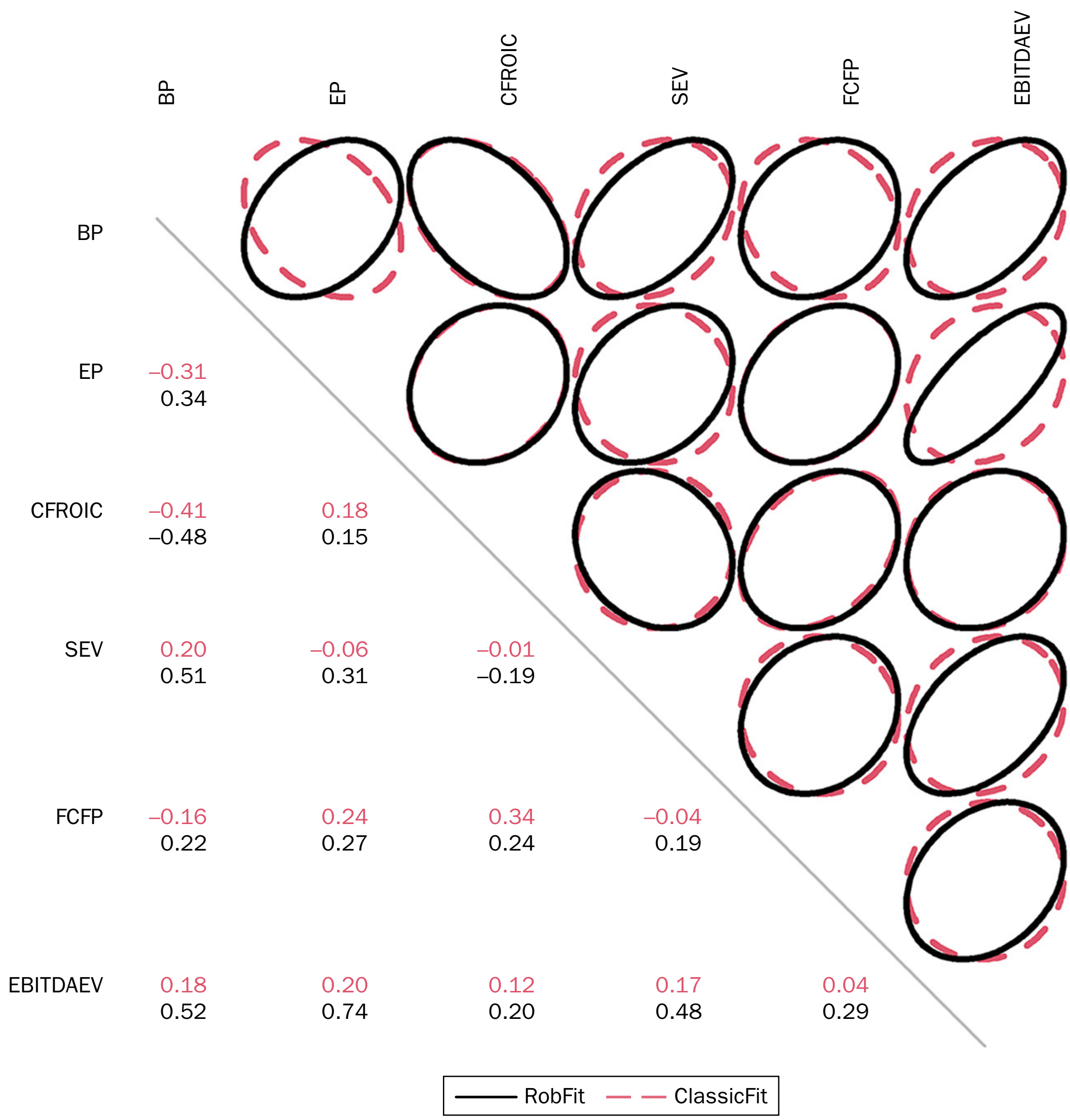
Pairwise Scatterplots of the SPGMI Factors BP, EP, CFROIC, SEV, FCFP, and EBITDAEV for January 2009



SOURCE: Standard & Poor's Global Market Intelligence.

EXHIBIT 24

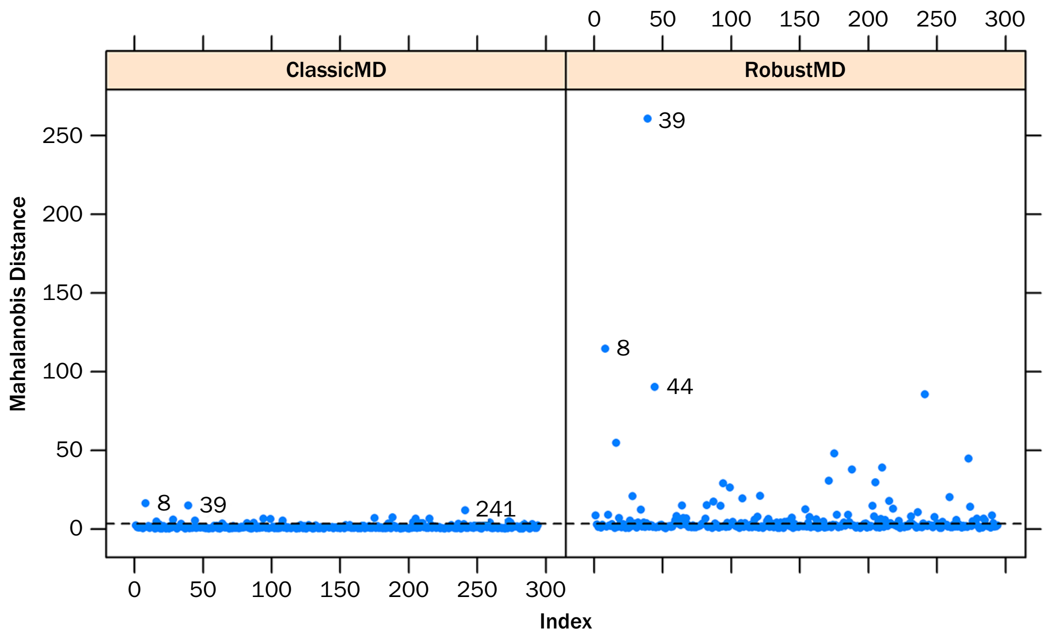
Correlation Ellipses Plots Based on Classic and Robust Covariance Matrix Estimates for the SPGMI Factors BP, EP, CFROIC, SEV, FCFP, and EBITDAEV for January 2009



SOURCE: Standard & Poor's Global Market Intelligence.

EXHIBIT 25

Classical and Robust MDs for the SPGMI Factors BP, EP, CFROIC, SEV, FCFP, and EBITDAEV, January 2009



SOURCE: Standard & Poor's Global Market Intelligence.

our robust covariance matrix-based robust distances are based on down-weighting all influential outliers, whether or not they are one-dimensional outliers, we recommend using the robust distances to shrink outliers in a manner to be described shortly.

Exhibit 24 reveals substantial differences between the correlations based on the sample covariance matrix and the covRobMM robust covariance matrix estimator, and these differences result in the startling differences between the classic MDs and robust MDs in Exhibit 25.

SUMMARY

Financial data sets often exhibit a small fraction of outliers that can have an adverse effect on the classical least-squares regression and sample covariance matrix estimators—the workhorses of empirical finance. This article provides a review of best available robust methods developed over several decades, with a reliable implementation in the R RobStatTM package. Our objective is to demonstrate that it is straightforward to use these packages without a detailed understanding of the corresponding statistical theory. At the same time, if one is more curious about the details of these methods, we offer an approachable introduction with intuitive interpretations. The estimators considered in this article are simply weighted versions of the classical ones in which the weight depends on the degree to which the corresponding data point is classified as an outlier by the method.

With this article, our goal is to encourage the routine use of robust methods at least as a diagnostic complement to their classical counterparts because of the following compelling reason—the empirical results may differ and may lead to very different conclusions. Knowing that the classical outcomes are sensitive to outliers that may not reflect relationships in normal market conditions, one is then left with the important task to reflect upon the differences. In particular, we showed the following:

1. Different factors may be identified as significant drivers of the cross-sectional returns in an asset-pricing context.
2. Time-series factor portfolio loadings sometimes vary significantly when estimated with robust regressions as compared with least squares.
3. Different factors may get selected in a robust machine-learning type of step-wise model selection versus normal theory selection.
4. Stocks may appear less correlated than they actually are because of outliers with significant implications about the potential for diversification.

For instructional purposes, and in an effort to facilitate a wider adoption of robust methods, the code used to create most examples in this article is available at https://github.com/robustport/PCRA/blob/main/README.md.

APPENDIX

mOpt Psi AND WEIGHT FUNCTIONS

The general formula for the mOpt robust regression psi function $\psi(x)$ is

$$\psi_{\text{mOpt}}(x) = \begin{cases} x, & |x| \leq 1 \\ \dfrac{\phi(1)}{\phi(1)-a}\left(x - \text{SGN}(x)\dfrac{a}{\phi(x)}\right)U(c-|x|), & |x| > 1 \end{cases} \tag{A1}$$

where $\phi(x)$ is the standard normal density function; $\phi(x)$ is the standard normal density function; SGN(x) is the "sign" function, which is equal to +1 for positive x, –1 for negative x, and 0 for $x = 0$; and $U(y)$ is the unit step function whose value is 0 for $y < 0$ and 1 for $y \geq 0$. The constant a determines the normal distribution efficiency of the mOpt regression estimator, and a determines the value of c. The choices $a = 0.0132$ and $c = 3.00$ result in a 95% normal distribution efficiency of the mOpt estimator. Details are provided in Konis and Martin (2021), who also discuss the rho function $\rho_{\text{mOpt}}(x)$ obtained as the integral of $\psi_{\text{mOpt}}(x)$.

In order to speed up the computation of robust regression, the current implementation of the RobStatTM R package function lmrobdetMM, which computes the mOpt regression coefficients, uses an accurate polynomial version of $\rho_{\text{mOpt}}(x)$.

The weight function $w_{\text{mOpt}}(x) = \psi_{\text{mOpt}}(x)/x$ corresponding to the $\psi_{\text{mOpt}}(x)$ with $a = 0.0132$ and $c = 3.00$ is

$$w_{\text{mOpt}}(x) = \begin{cases} 1, & |x| \leq 1 \\ \dfrac{\phi(1)}{\phi(1)-0.0132}\left(1 - \text{SGN}(x)\dfrac{0.0132}{x\phi(x)}\right)U(3.00-|x|), & |x| > 1 \end{cases} \tag{A2}$$

and its shape is displayed in Exhibit (3).

covRobMM WEIGHT FUNCTION

The formula for the smoothed hard rejection (SHR) function is

$$W_{SHR}(x) = \begin{cases} 1, & 0 < x \le 4 \\ q(x), & 4 < x \le 9 \\ 0, & x > 9 \end{cases} \tag{A3}$$

where

$$q(x) = -1.944 + 1.728d - 0.312d^2 + 0.016d^3$$

ACKNOWLEDGMENTS

All opinions and analysis are those of the authors and do not reflect those of Charles Schwab Corporation.

REFERENCES

Bali, T. G., R. F. Engle, and S. Murray. 2016. *Empirical Asset Pricing: The Cross Section of Stock Returns*. John Wiley & Sons.

Beaton, A. E., and J. W. Tukey. 1974. "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data." *Technometrics* 16: 147–185. doi: 10.1080/00401706.1974.10489171.

Bloch, M., J. Guerard, H. Markowitz, P. Todd, and G. Xu. 1993. "A Comparison of Some Aspects of the U.S. and Japanese Equity Markets." *Japan and the World Economy* 5: 3–26. doi: 10.1016/0922-1425(93)90025-y.

Carhart, M. M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52: 57–82. doi: 10.1111/j.1540-6261.1997.tb03808.x.

Chen, X., and R. D. Martin. 2020. "Standard Errors of Risk and Performance Estimators with Serially Correlated Returns." *Journal of Risk* 22 (7). doi: 10.2139/ssrn.3085672.

Chow, G., E. Jacquier, M. Kritzman, and K. Lowry. 1999. "Optimal Portfolios in Good Times and Bad." *Financial Analysts Journal* 55: 65–73. doi: 10.2469/faj.v55.n3.2273.

Fama, E. F., and K. R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427–465.

——. 2008. "Dissecting Anomalies." *The Journal of Finance* 63 (4): 1653–1678.

——. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116: 1–22.

Gu, S., B. Kelly, and D. Xiu. 2020. "Empirical Asset Pricing via Machine Learning." *The Review of Financial Studies* 33 (5): 2223–2273.

Guerard, J. B. 2016. "Investing in Global Markets: Big Data and Applications of Robust Regression." *Frontiers in Applied Mathematics and Statistics*: 1–16. issn: 2297-4687. doi: 10.3389/fams.2015.00014.

Guerard, J. B., H. Markowitz, and G. Xu. 2015. "Earnings Forecasting in a Global Stock Selection Model and Efficient Portfolio Construction and Management." *International Journal of Forecasting* 31: 550–560. doi: 10.1016/j.ijforecast.2014.10.003.

Guerard, J. B., G. Xu, and H. Markowitz. 2021. "A Further Analysis of Robust Regression Modeling and Data Mining Corrections Testing in Global Stocks." *Annals of Operations Research* 303: 175–195. doi: 10.1007/s10479-020-03521-y.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. "… and the Cross-Section of Expected Returns." *Review of Financial Studies* 29: 5–68. doi: 10.1093/rfs/hhv059.

Hou, K., C. Xue, and L. Zhang. 2015. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies* 28 (3): 650–705.

Huber, P. 1964. "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics* 35: 73–101.

——. 1981. *Robust Statistics*. New York: John Wiley & Sons.

Huber, P. J. 1973. "Robust Regression: Asymptotics, Conjectures and Monte Carlo." *Annals of Statistics* 1: 799–821.

Huber, P. J., and E. M. Ronchetti. 2009. *Robust Statistics*, 2nd edition. John Wiley. doi: 10.1002/9780470434697.

Konis, K., and R. D. Martin. 2021. *Optimal Bias Robust Psi and Rho*. RobStatTM package Vignette available at CRAN (https://cran.r-project.org/web/packages/RobStatTM/index.html).

Kritzman, M., and Y. Li. 2010. "Skulls, Financial Turbulence, and Risk Management." *Financial Analysts Journal* 66: 30–4. doi: 10.2469/faj.v66.n5.3.

Maronna, R. A., R. D. Martin, V. J. Yohai, and M. Salibian-Barrera. 2019. *Robust Statistics: Theory and Methods (with R)*, 2nd edition. John Wiley & Sons.

Martin, R. D., and D. Xia. 2021. "Efficient Bias Robust Cross Section Factor Models." *SSRN*. doi: 10.2139/ssrn.3921175. url: https://ssrn.com/abstract=3921175.

Martin, R. D., and D. Z. Xia. 2022. "Efficient Bias Robust Regression for Time Series Factor Models." *Journal of Asset Management* (March): 1–20. doi: 10.1057/s41260-022-00258-0. url: https://link.springer.com/content/pdf/10.1057/s41260-%20022-00258-0.pdf.

Martin, R. D., V. J. Yohai, and R. H. Zamar. 1989. "Min-Max Bias Robust Regression." *Annals of Statistics* 17: 1608–1630.

Rousseeuw, P. J., and A. M. Leroy. 1987. *Robust Regression and Outlier Detection*. John Wiley. doi: 10.1002/0471725382.

Tukey, J. W. 1960. "A Survey of Sampling from Contaminated Distributions." I. Olkin, ed., *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford, CA.

Yohai, V. J. 1987. "High Breakdown-Point and High Efficiency Robust Estimates for Regression." *Annals of Statistics* 15 (2): 642–656.

Yohai, V. J., and R. H. Zamar. 1997. "Optimally Local Robust M-Estimates of Regression." *Journal of Statistical Planning and Inference* 64 (2): 309–323.