ORIGINAL ARTICLE

Efficient bias robust regression for time series factor models

R. Douglas Martin[1] · Daniel Z. Xia[2]

Abstract

We introduce a robust regression estimator for time series factor models called the mOpt estimator. This estimator minimizes the maximum bias due to outlier generating distribution deviations from a standard normal errors distribution model, and at the same time has a high normal distribution efficiency. We demonstrate the efficacy of the mOpt estimator in comparison with the non-robust least squares (LS) estimator in applications to both single factor and multifactor time series models. For the case of single factor CAPM models we compared mOpt and LS estimates for cross sections of liquid stocks from the CRSP database in each contiguous two-year interval from 1963 to 1980. The results show that absolute differences between the two estimates greater than 0.3 occur for about 18% of the stocks, and differences greater than 0.5 occur for about 7.5% of the stocks. Our application of the mOpt estimator to multifactor models focuses on fitting the Fama-French 3-factor and the Fama-French-Carhart 4-factor models to weekly stock returns for the year 2008, using both the robust *t*-statistics associated with the mOpt estimates and a new statistical test for differences between the mOpt and LS coefficients. The results demonstrate the efficacy of the mOpt estimator in providing better model fits than the LS estimates, which are adversely influenced by outliers. Finally, since model selection is an important aspect of time series factor model fitting, we introduce a new robust prediction errors based model selection criterion called the Robust Final Prediction Error (RFPE), which makes natural use of the mOpt regression estimator. When applied to the 4-factor model, the RFPE finds as the best subset model the one that contains the Market, SMB and MOM factors, not the three Fama-French factors Market, SMB and HML. We anticipate that RFPE will prove to be quite useful for model selection of time series factor models.

Keywords Robust regression · Time-series factor models · Betas · Factor returns · Bias · Variance · Efficiency · Model selection

Introduction

The purpose of this paper is to introduce and encourage the use of a theoretically motivated and intuitive robust regression estimator for asset factor models, as a standard practice complement to the universal use of least squares to fit factor models. While the robust estimator proposed is equally useful for cross section and time series factor models, we focus in this paper on demonstrating the estimator's value in the context of time series factor models.

In data oriented terms, a robust regression estimator is one that is not much influenced by outliers and that provides a good fit to the bulk of the data. In spite of the fact that least squares (LS) is a best linear unbiased estimator (BLUE) and a maximum-likelihood estimator (MLE) in the case of normally distributed errors, the minimization of a quadratic function of the residuals causes least squares to be quite lacking in such robustness. A small fraction of highly influential outliers often results in least squares regressions analysis conclusions that are quite misleading about the relationship between stock returns and explanatory factors, and on the other hand a robust regression that controls for outliers can lead to a clear understanding of the relationship for the large majority of the data.

Over several decades, a body of knowledge about the statistical properties and computational methods for various robust regression methods has accumulated in the academic statistics research literature, with the books by Hampel et al. (1986), Rousseeuw and Leroy (1987), the Huber and Ronchetti (2009) second edition of Huber (1981),

✉ R. Douglas Martin
doug@amath.washington.edu

[1] Departments of Applied Mathematics and Statistics, University of Washington, Seattle, USA

[2] Independent Researcher, Baltimore, USA

and Maronna et al. (2019) being the primary references. However, this good news is counter-balanced by the bad news that today there are simply too many different types of robust regression estimators, including not only *maximum-likelihood type M-estimators* and *MM-estimators* discussed herein, but other distinct families of estimators such as the Rousseeuw least-trimmed-squares (LTS) discussed in Rousseeuw and Leroy (1987) and used by Knez and Ready (1997), and bounded-influence estimators. Furthermore, even the best known and most frequently used family of regression M-estimator has far too many variants based on different choices of optimization loss function. For example, the SAS software product offers 10 M-estimator variants with names: *Andrews*, *Bisquare*, *Cauchy*, *Fair*, *Hampel*, *Huber*, *Logistic*, *Median*, *Talworth*, *Welsch*.[1] It is therefore not surprising that there is currently no consensus what-so-ever concerning which robust regression estimator one should use, even within the family of regression M-estimators. Correspondingly there is not a common benchmark method for comparison purposes, as there is in the case of least squares. It seems clear that the existence of a large number of robust regression alternatives that are not very well justified theoretically, along with the lack of a common benchmark, has been and continues to be a serious impediment to the routine use of robust regression in many quantitative fields, and in asset management in particular.

As a solution to the conundrum of too many robust regression estimators, we introduce a theoretically well-justified robust regression estimator called an *mOpt* estimator, originally introduced in Chapter 5 of Maronna et al. (2019) and the Konis and Martin (2021) article, and extensively demonstrate the method's efficacy in fitting time series factor models. The mOpt estimator is obtained by replacing the quadratic function of regression residuals with a special bounded function of robustly scaled residuals, and its mathematical foundation was provided in the important technical papers of Yohai and Zamar (1997) and Svarc et al. (2002). Like all good robust regression methods and unlike least squares (LS), the mOpt estimator results in good factor model fits to most of the data and clearly identifies outliers. The mOpt method also results in accurate t-statistics that are representative of a large majority of the data. The mOpt estimator has the property that it minimizes the maximum estimator bias over a family of distributions that contains the normal distribution and an infinite number of non-normal outlier generating distributions, subject to a constraint of specified high normal distribution efficiency. A high normal distribution efficiency constraint means that the robust regression coefficient estimates will have variances that are

not much larger than those of least squares (LS) when the regression model has normally distributed errors. From a data point of view, the mOpt estimator has an appealing intuitive representation as a weighted least squares estimator based on a smooth weighting function that is zero for robustly scaled regression residuals that are greater than 3.0 in magnitude. Such data with scaled residual greater than 3.0 are deemed to be outliers, and in the idealized case of normal distribution data are very infrequently declared to be outliers.

For the applications of robust regression estimators in single-factor time series models there are a number of relevant prior research results. The first two of these were Sharpe (1971) and Cornell and Dietrich (1978), both of whom focused on mean-absolute deviation (MAD) estimators, and based on their empirical studies concluded that the MAD estimator does not yield much improvement over LS. Subsequent papers by Chan and Lakonishok (1992), Mills and Coutts (1996) and Cloete et al. (2002) established superior performance of several regression quantile based beta estimators in the presence of fat-tailed non-normality and outliers, but those estimators have seen little common use. Martin and Simin (2003) introduced robust beta M-estimators similar to the one we discuss herein, and showed that they result in attractive beta point estimates and forecasts. The very interesting paper by Genton and Ronchetti (2008) focuses on robust forecasting of beta that combines Vasicek (1973) shrinkage with a regression M-estimator variant.[2] Finally, Bailer et al. (2011a) extends the work of Martin and Simin (2003) to a large cross section of U.S. stocks, and is a point of departure for the robust single factor models portion of the current paper.

It is curious that in spite of the existing published research on robust estimation of CAPM betas, the literature is quite silent on the topic of robust fitting of multifactor models.[3] In particular, there is no mention what-so-ever of robust regression in any of the many papers on empirical asset pricing articles that use time series factor models, such as Fama and French (1993), Carhart (1997), Fama and French (2015), Hou and Dijk (2018), and Hou et al. (2020). This is also the case in the extensive and growing literature on anomalies, such as Fama and French (2008), Hou et al. (2020), and Feng

---

[1] See Table 104.5 in The ROBUSTREG Procedure at: https://support.sas.com/rnd/app/stat/procedures/robustreg.html.

[2] The historical introductory section of Genton and Ronchetti (2008) is highly recommended.

[3] Factor model fitting based on t-distribution models, e.g., as in Galea et al. (2020) and references therein, fail to qualify as a robust regression method in that they can suffer from arbitrarily large bias under the commonly used mixture distribution model that we discuss in Sect. 2. Furthermore, fitting such distributions well requires very large sample sizes for t-distribution with unknown degrees of freedom.

et al. (2020), and the use of robust time series factor model fitting for research on anomalies is overdue.

The remainder of the paper is organized as follows. Section 2 discusses the mOpt robust regression method, and points out a basic lack of robustness of the well-known Huber robust regression method. Section 3.1 focuses on single factor CAPM beta model estimation for the universe of CRSP database ("Center for Research in Security Prices, LLC") To Editor: the previous reference in quotes would be best as a footnote. RDM liquid stocks from 1963 to 2018, for which extensive empirical analysis of the distribution of LS versus robust mOpt betas reveals frequent substantial adverse influence of outliers on the LS betas. Section 3.2 focuses on mOpt and LS fits of the Fama-French 3-factor (FF3) and the Fama-French-Carhart 4-factor (FFC4) multifactor models to stock returns, using classical t-statistics and adjusted R-squared inference values for the LS fits and robust versions thereof for the mOpt fits. The results show that mOpt robust fit is better than the LS fit for the FF3 model, and for the FF4 the mOpt fit is substantially better than that of LS. Section 3.2 also introduces a new robust test for significant differences between mOpt and LS coefficients and demonstrates its usefulness in comparing robust mOpt fits of the FF3 and FF4 models. Section 4 introduces a robust final prediction error (RFPE) criteria for model selection, and demonstrates its efficacy. Section 5 provides a brief discussion and summary. The Appendices contain supplemental details concerning the mOpt robust regression method.

Efficient bias robust regression for time series factor models

We begin in Sect. 2.1 by defining the time series factor models to which we will apply our efficient bias robust mOpt regression estimator. Then Sect. 2.2 discusses general maximum-likelihood type M-estimators, which are defined in terms of a loss function $\rho(x)$ applied to scaled prediction residuals, and presents the coefficient estimates asymptotic covariance matrix formula, which is used in defining mOpt efficiency relative to the LS and in computing M-estimator coefficient standard errors. Section 2.3 introduces the intuitively appealing bounded loss function $\rho_{\mathrm{mOpt}}(x)$ that defines the mOpt regression M-estimator, shows that the estimator has a weighted least squares representation with data-dependent weights, where the latter are generated by a smooth weight function that is zero for scaled prediction residuals that are larger than 3.0 in absolute value. Section 2.4 discusses the important mOpt estimator bias robustness property of minimizing the maximum asymptotic bias for distributions in a Tukey-Huber family of two-component mixture distributions, one of which is a normal distribution and the other is any distribution function, many of which generate outliers that give rise to M-estimator bias. Finally,

Sect. 2.5 shows that the well-known Huber robust regression estimator, which is defined in terms an unbounded function $\rho_{\mathrm{Huber}}(x)$, lacks bias robustness in comparison to mOpt.

Time series factor models

Throughout this paper we assume that asset returns are generated by a time series factor model of the form

$$
\begin{aligned}
r_t &= \beta_0 + f_{1,t}\beta_1 + f_{2,t}\beta_2 + \cdots + f_{K,t}\beta_K + s\epsilon_t \\
&= \tilde{\mathbf{f}}_t'\boldsymbol{\beta} + s\epsilon_t \\
&= \left(1, \mathbf{f}_t'\right)\boldsymbol{\beta} + s\epsilon_t, \ t = 1, 2, \ldots, T
\end{aligned} \tag{1}
$$

where the r_t are the (typically excess) returns of a specific asset at time t, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)'$ is a $K+1$ dimensional vector of unknown regression coefficients, the ϵ_t are the regression errors, and

$$
\mathbf{f}_t = \left(f_{1,t}, f_{2,t}, \ldots f_{K,t}\right)', \ t = 1, 2, \ldots, T \tag{2}
$$

is a time series of K dimensional random *factor returns*.

Examples of the above time series factor model include: (1) the CAPM model 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; (2) the Fama-French 3-factor model (FF3) 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 4-factor model (FFC4) adds the Carhart (1997) Momentum factor to FF3 model, and other multifactor models including the Fama and French (2015) 5-factor model (FF5), and the Hou et al. (2015) q-factor model. When dealing with a cross section of N assets, say stocks, one replaces $r_t, \beta_0, \beta_1, \ldots, \beta_K, \epsilon_t$ with $r_{it}, \beta_{i0}, \beta_{i1}, \ldots, \beta_{iK}, \epsilon_{it}$ for $i = 1, 2, \ldots, N$.
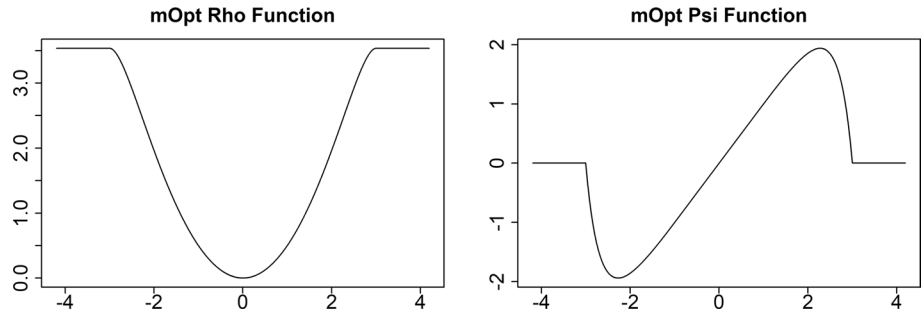
General regression M-estimators

A regression M-estimator $\hat{\boldsymbol{\beta}}$ for the model (1) is defined as a solution to the minimization problem

$$
\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \left[\sum_{t=1}^{T} \rho\left(\frac{r_t - \tilde{\mathbf{f}}_t'\boldsymbol{\beta}}{\hat{s}} \right) \right] \tag{3}
$$

where $\rho(x)$ is a symmetric *rho* function for which $\rho(0) = 0$ and $\rho(x)$ is non-decreasing for $x > 0$. The $\hat{s}$ above is a robust estimate of the scale parameter s, which is computed prior to the above optimization as described in the last paragraph of Sect. 2.3. The corresponding estimating equation obtained by differentiating the above summation expression with respect to $\boldsymbol{\beta}$ is:

Fig. 1 Bounded $\rho_{\mathrm{mOpt}}(x)$ and corresponding $\psi_{\mathrm{mOpt}}(x)$ for which $\psi_{\mathrm{mOpt}}(x) = 0$ and $\rho_{\mathrm{mOpt}}(x)$ is constant for $|x| \geq 3.00$

$$\sum_{t=1}^{T} \tilde{\mathbf{f}}_t \psi\left(\frac{r_t - \tilde{\mathbf{f}}_t' \hat{\boldsymbol{\beta}}}{\hat{s}} \right) = 0 \tag{4}$$

where $\psi(x) = \rho'(x)$ is the corresponding *psi function*, and

$$\hat{\epsilon}_t = \frac{r_t - \tilde{\mathbf{f}}_t' \hat{\boldsymbol{\beta}}}{\hat{s}} \tag{5}$$

are the robustly scaled regression residuals.

Robust regression M-estimators were originally introduced by Huber (1973), and are extensively discussed in Chapters 4 and 5 of Maronna et al. (2019). The least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ is a special non-robust case of an M-estimator obtained with the quadratic rho function $\rho(x) = x^2/2$ and correspondingly $\psi(x) = x$. The *least absolute deviation* (LAD) estimator $\hat{\boldsymbol{\beta}}_{LAD}$ is obtained when $\rho(x) = |x|$.[4]

Regression M-estimator covariance matrix

It can be shown that under rather general conditions an M-estimator $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\hat{\psi})$ based on a sample size T is asymptotically normal with a $(K+1) \times (K+1)$ covariance matrix

$$\mathbf{V}_{\tilde{\mathrm{f}}}(\psi) = \mathrm{v}(\psi, F_0)\mathbf{M}_{\tilde{\mathrm{f}}}^{-1} \tag{6}$$

where

$$\mathbf{M}_{\tilde{\mathrm{f}}} = \mathrm{E}\left(\tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t'\right) = \begin{pmatrix} 1 & \boldsymbol{\mu}_{\mathrm{f}}' \\ \boldsymbol{\mu}_{\mathrm{f}} & \mathbf{M}_{\mathbf{f}} \end{pmatrix} \tag{7}$$

is the *moment matrix* of the $\tilde{\mathbf{f}}_t$ in (1), the formula for $\mathrm{v}(\psi, F_0)$ is given by (8), $\mathbf{M}_{\mathbf{f}}$ is the moment matrix of the factor returns vector $\mathbf{f}_t$, and $\boldsymbol{\mu}_{\mathrm{f}} = E(\mathbf{f}_t)$.

The scalar quantity $\mathrm{v}(\psi, F_0)$ is the variance of an M-estimator of location introduced by Huber (1964), who derived the expression

$$\mathrm{v}(\psi; F_0) = s^2 \frac{E_{F_0} \psi^2(\epsilon_1)}{E_{F_0}^2 \psi'(\epsilon_1)} \tag{8}$$

where F_0 a standard distribution of the ϵ_t in (1), for example a standard normal or a standard t-distribution. A formal statement of the asymptotic normality of a regression M-estimator with asymptotic covariance matrix of the form (6) is provided in Section 5.3.1.3 Maronna et al. (2019), and a proof of the result is provided in Section 10.2.2.

Use of a standard matrix inversion formula for partitioned allows one to express the inverse of the moment matrix (7) in the form

$$\mathbf{M}_{\tilde{\mathrm{f}}}^{-1} = \begin{pmatrix} 1 + \boldsymbol{\mu}_{\mathrm{f}}' \mathbf{C}_{\mathrm{f}}^{-1} \boldsymbol{\mu}_{\mathrm{f}} & -\boldsymbol{\mu}_{\mathrm{f}}' \mathbf{C}_{\mathrm{f}}^{-1} \\ -\mathbf{C}_{\mathrm{f}}^{-1} \boldsymbol{\mu}_{\mathrm{f}} & \mathbf{C}_{\mathrm{f}}^{-1} \end{pmatrix} \tag{9}$$

where $\mathbf{C}_{\mathrm{f}}$ is the covariance matrix of the factor returns $\mathbf{f}_t' = (f_{1,t}, f_{2,t}, \ldots f_{K,t})$.[5] For a time series factor model (1) without an intercept β_0 the right-hand side of the above expression reduces to $\mathbf{C}_{\mathrm{f}}^{-1}$.

It is important to note that the regression M-estimator covariance matrix (6) depends on the choice of rho function $\rho(x)$, and hence the corresponding psi function $\psi(x)$, only through the scalar expression (8). In the special case of the LS estimator where $\psi_{LS}(x) = x$ and F_0 is a standard normal distribution Φ, $\mathrm{v}(\psi_{LS}(x);\Phi) = s^2$ which is the variance of $s\epsilon_1$. Standard errors for the regression M-estimator coefficients are obtained in the usual way by dividing estimates of the diagonal elements of (6) by the sample size T and taking the square root of the results. For details see Appendix A3.

The mOpt robust regression estimator

The robust regression estimator that we use throughout the paper, and recommend in general, is a regression M-estimator based on the rho function $\rho_{\mathrm{mOpt}}(x)$ and psi function $\psi_{\mathrm{mOpt}}(x)$ shown in Fig. 1. We call the estimator

---

[4] A regression M-estimator (3) is a maximum-likelihood estimator (MLE) $\hat{\boldsymbol{\beta}}_{MLE}$ when $\hat{s}$ is replaced by a known s and the rho function is $\rho_{mle}(x) = -log f_0(x)$ where $f_0(x)$ is the probability density of the ϵ_t, and correspondingly the psi function is $\psi_{mle}(x) = -f_0'(x)/f_0(x)$. In applications where s is unknown and is replaced by a consistent estimator $\hat{s}$, the regression coefficient vector estimate is asymptotically an MLE.

[5] We note that $\sigma^2 \mathbf{M}_{\tilde{\mathrm{f}}}^{-1}$ is the known expression for the covariance of the least-squares special case of an M-estimator with σ^2 the variance of $s\epsilon_t$.

an *mOpt* estimator because it is motivated by an *efficiency constrained optimal bias robustness* property described in Sect. 2.4. The rho and psi functions in Fig. 1 have the following intuitive appeal for fitting factor models. The actual distribution of the ϵ_t in (1) is typically approximately normal in the central part of the distribution, but has an unknown fat-tailed and typically skewed distribution outside the central region. Correspondingly, the rho function is quadratic for a large portion of the central region $(-3, +3)$ where it behaves like the least squares loss function, and is constant outside the interval $(-3, +3)$, where it is indifferent to the size of the scaled regression residuals in (3). This behavior represents an agnostic assignment of equal losses to all pairs $(r_t, \tilde{\mathbf{f}}_t')$ for which the absolute value of the robustly scaled residuals is larger than 3.00. Correspondingly, the mOpt psi function has value zero for all such pairs $(r_t, \tilde{\mathbf{f}}_t')$, which therefore have no influence on the regression parameter estimates $\hat{\boldsymbol{\beta}}_{\mathrm{mOpt}}$. An M-estimator psi function that is zero outside some finite interval is called a *redescending* psi function in the robust statistics literature.[6]

A general formula for $\psi_{\mathrm{mOpt}}(x)$ that depends on a *tuning parameter c* is given in Appendix A1, where it is seen that $\psi_{\mathrm{mOpt}}(x) = 0$ for $|x| \geq c$. Since the rho function $\rho_{\mathrm{mOpt}}(x)$ is the integral of the psi function $\psi_{\mathrm{mOpt}}(x)$, it follows that the rho function has a constant value for $|x| \geq c$. The parameter c controls the normal distribution *efficiency* of the mOpt estimator, where efficiency is defined as the ratio of the asymptotic variance of the least squares estimator to that of the mOpt estimator, often expressed as a percent. A formula for the normal distribution efficiency of the mOpt estimator in terms of the tuning constant c is given in Appendix A2. Since an LS estimator has the minimum possible variance for normal distributions, an mOpt estimator has an efficiency less than 100%. The choice of an mOpt value for c controls a trade off between normal distribution efficiency and robustness toward outliers. Increasing the value of c results in the rho function $\rho_{\mathrm{mOpt}}(x)$ becoming closer to the least squares quadratic rho function, with correspondingly higher normal distribution efficiency but less robustness toward outliers, and conversely. Throughout this paper, as in Fig. 1, the mOpt estimator is based on the choice $c = 3.0$, which results in a normal distribution efficiency of 95%.

For the 95% efficient mOpt estimator that we use throughout this paper, and recommend in general, the ratio of the asymptotic mOpt variance to least squares variance is 1.0526

Fig. 2 The mOpt weight function ($c = 3.00$)

(to 4 digits), which means that the ratio of asymptotic standard deviations is 1.026 (to 3 digits). This suggests that in practice the standard errors (SE) of an mOpt estimator will be only about 2.6% larger than that of the least squares estimator. Thinking of the normal distribution excess standard error (SE) as an "insurance premium" to be paid in order to provide protection against bias and inflated variability due to outlier generating non-normal distributions, an SE premium of 2.6% seems quite "inexpensive. We propose that a 95% normal distribution efficiency mOpt estimator be used as a benchmark for evaluating the performance of not only an LS alternative, but also other robust regression estimator alternatives.

Iteratively reweighted least squares and MM-estimator

A regression M-estimator psi function $\psi(x)$ defines a *weight function*

$$w(x) = \frac{\psi(x)}{x} \qquad (10)$$

that allows one to express the solution $\hat{\boldsymbol{\beta}}$ of the estimating Eq. (4) in the form of a weighted least squares (WLS) estimate. To see that this is the case, note that (4) can be written in the form

$$\sum_{i=1}^{T} w_t \tilde{\mathbf{f}}_t \left(r_t - \tilde{\mathbf{f}}_t' \hat{\boldsymbol{\beta}} \right) = 0 \qquad (11)$$

where the w_t are the data-dependent weights:

$$w_t = w_t(\hat{\boldsymbol{\beta}}; \hat{s}) = w\left(\frac{r_t - \tilde{\mathbf{f}}_t' \hat{\boldsymbol{\beta}}}{\hat{s}} \right). \qquad (12)$$

Thus the estimate $\hat{\boldsymbol{\beta}}$ may be expressed in the nonlinear weighted least squares (WLS) mathematical form:

---

[6] The mOpt regression estimator is a leading example of an M-estimator that is not a maximum-likelihood estimator. A justification of this claim is the fact that there exists no probability density function $f_0(x)$ that integrates to 1 and has the positive constant value $f_0(x) = exp(-\rho(x))$ for all $|x| \geq c$.
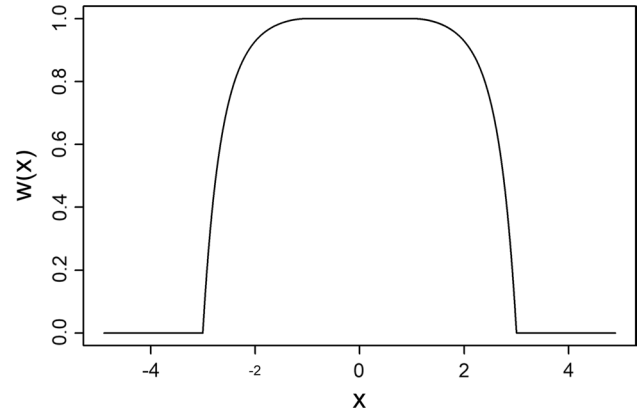
$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{T} w_t\left(\hat{\boldsymbol{\beta}}; \hat{s}\right) \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t' \right)^{-1} \left(\sum_{i=1}^{T} w_t(\hat{\boldsymbol{\beta}}; \hat{s}) \tilde{\mathbf{f}}_t r_t \right). \qquad (13)$$

The weight function $w_{\mathrm{mOpt}}(x) = \psi_{\mathrm{mOpt}}(x)/x$ for our mOpt estimator $\hat{\boldsymbol{\beta}}_{\mathrm{mOpt}}$ is shown in Fig. 2.

The mOpt weight function gives a weight of 1 to all sufficiently small robustly scaled residuals $(r_t - \tilde{\mathbf{f}}_t' \hat{\boldsymbol{\beta}}_{\mathrm{mOpt}})/\hat{s}$, and smoothly transitions to zero weight for scaled residuals whose absolute is greater than 3.00. All asset and factor returns pairs $(r_t, \tilde{\mathbf{f}}_t')$ whose scaled residuals have absolute values larger than 3.0 are said to be *rejected*. In this regard, the mOpt robust regression estimator has the highly intuitive appeal that it is robust smoothed version of non-robust classical *3-sigma edit rule,* according to which all data values in a sample that differ from the sample mean by at least 3 times the sample standard deviation are deleted. Under the assumptions that the residuals in (1) are normally distributed and the estimates $\hat{s}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{mOpt}}$ are equal to the true parameter values, the probability that a pair $(r_t, \tilde{\mathbf{f}}_t')$ is rejected by the mOpt estimator is the probability that a standard normal random variable is larger than 3.00 in magnitude, which is only .27%. Thus, such data pairs are rejected only very rarely in the highly idealized case of normally distributed residuals.

An important aspect of the mOpt WLS version (13) is that it lends itself to the iteratively-reweighted least squares (IRWLS) computation, whose mathematical representation is

$$\hat{\boldsymbol{\beta}}^{k+1} = \left(\sum_{i=1}^{T} w_t(\hat{\boldsymbol{\beta}}^k; \hat{s}) \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t' \right)^{-1} \left(\sum_{i=1}^{T} w_t(\hat{\boldsymbol{\beta}}^k; \hat{s}) \tilde{\mathbf{f}}_t r_t \right), k = 0, 1, 2, \dots$$

$$(14)$$

with an initial estimate $\hat{\boldsymbol{\beta}}^0$. It is shown in Section 9.1 of Maronna et al. (2019) that for M-estimators based on rho functions with a similar shape as $\rho_{\mathrm{mOpt}}(x)$, the IRWLS algorithm converges to a solution of the estimating Eq. (4). However, such a solution may yield only a local minimum of the objective function in (3).

The above IRWLS algorithm needs a highly robust initial estimate $\hat{\boldsymbol{\beta}}^0$ that helps insure that a global minimum of (3) is achieved. An estimator that serves this purpose well is an *S-estimator* that was introduced by Rousseeuw and Yohai (1984), and is described in Section 5.4.1 of Maronna et al. (2019). An S-estimator is a special form of M-estimator that jointly estimates the regression coefficient $\boldsymbol{\beta}$ and the error term scale parameter s, and has a *high breakdown point* but a low *normal distribution efficiency* of about 29%. The S-estimator regression coefficient estimate is used as the initial estimate of the IRWLS algorithm (14) and the S-estimator scale estimate $\hat{s}$ is used in computing the robustly scaled prediction residuals given by (5). The breakdown point (BP) of an estimator is the largest fraction of data outliers whose

value can tend to infinity without taking the estimator to the boundary of the parameter space, which is infinity in the case of regression, and a *high-breakdown point estimator* is one with a breakdown point of 0.5, or approximately so.[7] An M-estimator that uses a high BP but inefficient initial estimator is called an *MM-estimator*, whose theoretical properties were developed by Yohai (1987). The regression estimator used for the examples in this paper and recommended in general, is an *mOpt MM-estimator*, but for simplicity we just refer to it as the mOpt estimator.[8] The mOpt estimator may be computed using the function `lmrobdetMM` in the `RobStatTM` R package that is available at CRAN https://cran.r-project.org/web/packages/RobStatTM/index.html.

Efficient bias robustness of the mOpt regression estimator

Here, we describe an *efficient bias robustness* property of the mOpt regression estimator that does not exist for any other robust regression estimator, and in particular does not exist for any of the 10 SAS M-estimators mentioned in Sect. 1. The bias robustness optimality of the mOpt estimator is obtained with respect to the Tukey-Huber family of distributions

$$F_\gamma(r, \mathbf{f}) = (1 - \gamma) N\left(r - \mathbf{f}'\boldsymbol{\beta}\right) G(\mathbf{f}) + \gamma H(r, \mathbf{f}), 0 \le \gamma < 0.5$$

$$(15)$$

for $(r_t, \mathbf{f}_t)$, where N is a normal distribution for the regression model errors $s\epsilon_t$, $G(\mathbf{f})$ is a fixed multivariate distribution function of factor returns $\mathbf{f}_t$, and $H(r, \mathbf{f})$ is an *unrestricted* joint distribution.[9] When $\gamma = 0$ the distribution $F_0(r, \mathbf{f})$ generates data with normally distributed errors and the MLE of $\boldsymbol{\beta}$ is the least squares estimator, but when $\gamma > 0$ many distributions $H(r, \mathbf{f})$ result in $F_\gamma(r, \mathbf{f})$ generating outliers with probability γ. Correspondingly, when $\gamma > 0$ both LS and M-estimators will be biased for some $H(r, \mathbf{f})$. The optimality problem is to find a regression M-estimator *Opt* that minimizes the maximum asymptotic bias with respect to all possible $H(r, \mathbf{f})$, subject to a constraint of a user specified high normal distribution efficiency. The latter constraint insures that the Opt performance is only slightly worse than

---

[7] For example, the sample mean has a breakdown point BP = 0 and the sample median has a breakdown point BP ≈ 0.5 and exactly 0.5 as the sample size tends to infinity. So the sample median is a high breakdown point estimator. For details concerning estimator breakdown points, see Sections 5.2 and 5.3.1 of Maronna et al. (2019).

[8] Details concerning the mOpt computational method are provided in Sections 5.5 and 5.7.4 of Maronna et al. (2019).

[9] The family of distributions (15) is a regression model generalization of the Tukey–Huber distributions introduced by Tukey (1960) in the context of robust estimation of a mean, and by Huber (1964) in his seminal paper on robust estimation of a location parameter.
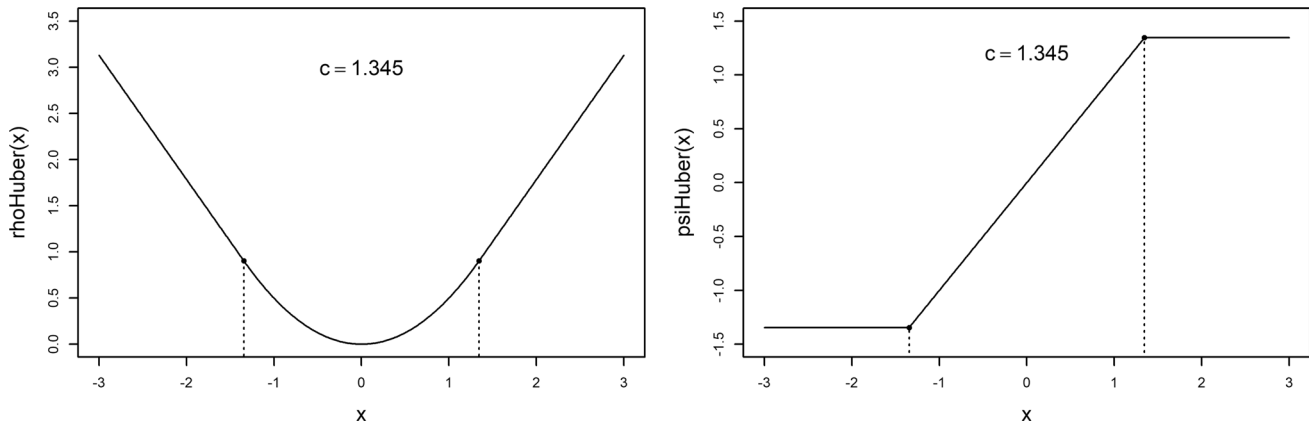
Fig. 3 Huber Rho and Psi function for $c = 1.345$

least squares when $\gamma = 0$ in the Tukey-Huber regression model. Yohai and Zamar (1997) solved this problem locally for small γ, and the mOpt estimator we discuss here is a very slight modification of the Opt estimator that is needed to insure convergence of the IRWLS algorithm in Sect. 2.3.1. Later, Svarc et al. (2002) solved the global optimality problem for all $0 < \gamma < 0.5$, and their show that the Yohai and Zamar Opt local solution is a remarkably good approximation to the global solution.

The huber estimator lack of robustness

It was shown by Martin et al. (1989) that the asymptotic bias of an M-estimator with an unbounded rho function $\rho(x)$ is unbounded with respect to all possible $H(r, \mathbf{f})$ in the Tukey-Huber family 15, but such bias is bounded for an bounded $\rho(x)$. Thus, it is not surprising that the Opt function discovered later by Yohai and Zamar (1997) has a bounded rho function, as does the mOpt variant. On the other hand, the well-known Huber (1973) regression estimator based on the rho function

$$\rho_{\text{Huber}}(x;c) = \begin{cases} 0.5x^2, & |x| \le c \\ c|x| - \frac{1}{2}c^2, & |x| > c \end{cases} \quad (16)$$

is the leading example of a regression M-estimator based on an unbounded rho function. The corresponding Huber psi function is the monotonic function:[10]

$$\psi_{\text{Huber}}(x;c) = \begin{cases} x, & |x| \le c \\ c \cdot \text{sgn}(x), & |x| > c. \end{cases} \quad (17)$$

The monotonicity of the Huber psi function has the undesirable consequence, unlike the mOpt estimator redescending

psi function, that it does not reject outliers. The Huber rho and psi functions are displayed in Fig. 3 for the tuning constant choice $c = 1.345$ that results in an estimator with 95% normal distribution efficiency.

The Huber regression estimator is quite popular, partly because of the seminal min-max variance robustness property established for estimation of a location parameter in Huber (1964) and inherited by the Huber (1973) regression M-estimator, but the more so because unlike the non-convex function $\rho_{\text{mOpt}}(x;c)$, the function $\rho_{\text{Huber}}(x;c)$ is convex. Consequently, the Huber regression M-estimator may be easily computed using readily available convex optimization software.[11] None-the-less, the Huber estimator can have arbitrarily large bias for distributions in the Tukey-Huber family, and this theoretical property has quite practical applications relevance, as is demonstrated by examples in Sect. 3.1 where outliers cause the Huber regression estimator to be as badly biased as least squares (see Fig. 5).

While theoretical demonstration of unbounded bias for regression M-estimators with unbounded rho functions in Martin et al. (1989) is highly technical, the following simple heuristic argument shows why the unbounded rho and monotonic psi function of the Huber regression estimator is vulnerable to arbitrarily large bias due to outliers of unbounded size. Consider the single factor model (18) with intercept $\beta_0 = 0$, arbitrary β_1, and residual scale parameter $s = 1$. In that case the Huber M-estimator equation is:

$$\sum_{t=1}^{T} f_t \psi_{\text{huber}}\left(r_t - \hat{\beta}_1 f_t\right) = 0.$$

Suppose that the factor return at any given time, say for example f_1 at $t = 1$, takes on an arbitrarily large absolute

---

[10] The function sgn(x) is the *sign* function which is equal to -1 for $x < 0$, +1 for $x > 0$, and 0 for $x = 0$.

[11] For example, see the CVXR code at https://cvxr.rbind.io/cvxr_examples/cvxr_huber-regression/.

value with a corresponding arbitrary asset return value r_1. For any pair of values r_1 and f_1 with $r_1 \neq \hat{\beta}_1 f_1$ and f_1 arbitrarily large, the term $f_1 \psi_{\text{huber}}(r_1 - \hat{\beta}_1 f_1)$ will be non-zero and arbitrarily large, thereby dominating $\sum_{t=2}^{T} f_t \psi_{\text{huber}}(r_t - \hat{\beta}_1 f_t)$. Consequently, the solution of the above estimating equation will in the limit, as f_1 gets arbitrarily large, be the solution of

$$f_1 \psi_{\text{huber}}(r_1 - \hat{\beta}_1 f_1) = 0$$

which is $\hat{\beta}_1 = r_1 / f_1$. Since this estimate can take on any value determined by the ratio r_1 / f_1, $\hat{\beta}_1$ can have an unbounded bias $B = r_1 / f_1 - \beta_0$. Note that this does not happen in the case of the redescending psi function $\psi(x) = \psi_{\text{mOpt}}(x)$ because for fixed r_1 the term $f_1 \psi_{\text{huber}}(r_1 - \hat{\beta}_1 f_1)$ will be zero for arbitrarily large $|f_1|$.

Robust time series factor models

The mOpt robust regression fitting method can be applied in a wide variety of time series factor models, including single-factor index models and multi-factor models such as the macro-economic models of Chen et al. (1986), the style analysis models introduced by Sharpe (1988) and discussed in Sharpe (1992), and an increasingly wide variety of empirical asset pricing factor models. Early examples of the latter are the Fama and French (1993) three-factor model (FF3) and the Fama-French-Carhart four-factor model (FFC4) that adds the Carhart (1997) momentum factor to the FF3 model. More recently introduced empirical asset pricing models include the Fama and French (2015) five-factor model (FF5), and the q-factor and augmented q-factor models discussed in Hou et al. (2015) and Hou et al. (2020). See also Feng et al. (2020) who refer to the wide variety of factor models as "The Factor Zoo". Given the large number of time series factor models that currently exist in the literature, and the likelihood of more in the future, the potential value of mOpt robust fitting of these models is considerable.

We illustrate mOpt versus LS fitting first for the single factor model that is used extensively by practitioners and financial data service providers to compute least squares CAPM beta estimates. For the CAPM betas we show first by a few striking examples, and then by extensive empirical analysis of the cross sections of mOpt and LS beta estimates based on two-year intervals of weekly returns, that LS beta estimates are often considerably biased by outliers that have little influence on the mOpt estimates. In particular, we show that LS and mOpt betas differ in absolute value by at least 0.3 for roughly 26% of microcap stocks, 14% of smallcaps and 7% of bigcaps, and differ by at least 0.5 for roughly 12% of microcap stocks, 5% of smallcaps and 2% of bigcaps Then we present striking examples of multifactor models applications of the mOpt versus LS fits of stock returns for

the FF3 model and the FFC4 model. We also introduce a test for a significant difference between mOpt and LS coefficient estimates, and apply it to both the single factor and the multifactor models. Finally, we introduce a new robust model selection criteria called RFPE as an alternative to the non-robust AIC based method, and illustrate the use of RFPE versus AIC in selecting a best subset of the FFC4 model factors.

Single factor time series models

A single factor model has the form

$$r_t = \alpha + \beta f_t + s \epsilon_t \quad t = 1, 2, \ldots, T \tag{18}$$

where r_t is an equity return and f_t a factor return that is typically a market proxy or an active manager's index benchmark such as the S&P500, Russell 1000, Russell 2000, or Russell 3000, among many others. When f_t is a market return $r_{M,t}$, and both the equity return and market return are excess returns relative to a risk-free rate, this is the time series version of the CAPM.

Here, we extend the robust beta M-estimators study of Bailer et al. (2011b) in the following ways: (a) we use the improved mOpt estimator based on the analytic form of $\psi_{\text{mOpt}}(x)$ and its integral $\rho_{\text{mOpt}}(x)$, rather than a polynomial approximation of these functions; (b) we demonstrate by example a robustness inadequacy of the Huber M-estimator rho and psi functions discussed in Sect. 2.5; (c) we extend the study period from 1963–2009 to 1963–2018, and compute mOpt and LS betas using weekly return for cross sections of *liquid* stocks on 28 contiguous two-year intervals. We define liquid stocks as those that have at least 100 non-zero returns in a two-year interval. The numbers of such stocks in each two-year window vary from a minimum of 1,725 for 1963-1964 to a maximum of 7,126 for 1996-1998, with an average of 4,207 stocks over all two-year intervals.

We begin with the two examples in Fig. 4 that display the results of computing robust mOpt and LS beta estimates using two years of weekly returns for the stocks with tickers OFG and DD. The legend gives the beta estimate values with their standard errors in parentheses. The dotted lines, parallel to the solid line mOpt fit, define strips outside of which the stock and market returns pairs are rejected by the mOpt estimator, and as such have no influence on the fit. The rejected data points, which we define to be outliers, are indicated by the open circle symbols. For the OFG stock results in the left-hand plot, the mOpt estimator rejects 8 outliers, 6 of them barely so. In this example the mOpt robust method provides a very good fit to the bulk of the data in spite of the huge outlier in the upper-right part of the figure, but the LS estimator is quite adversely influenced by that outlier and fails to
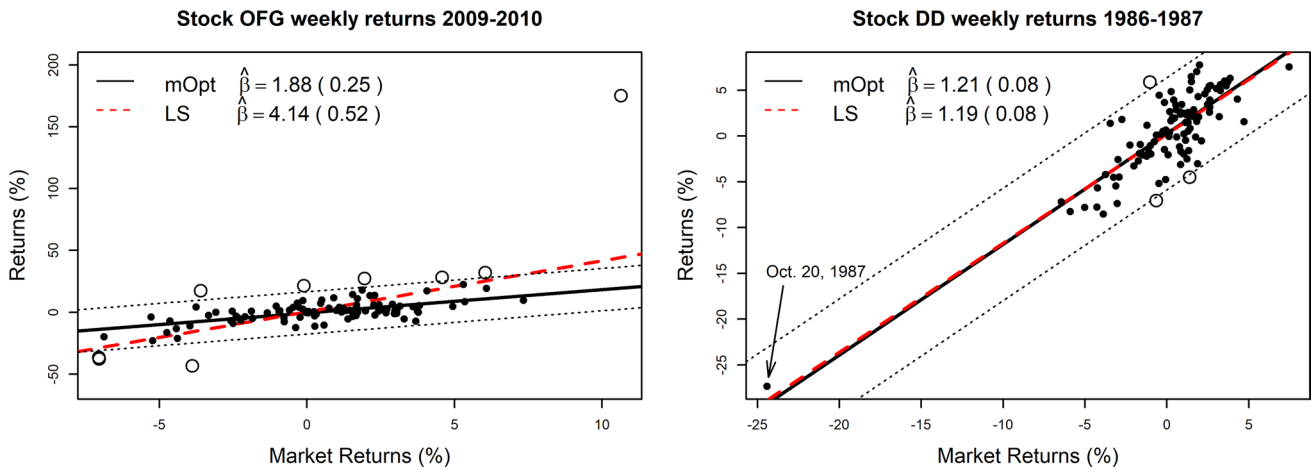
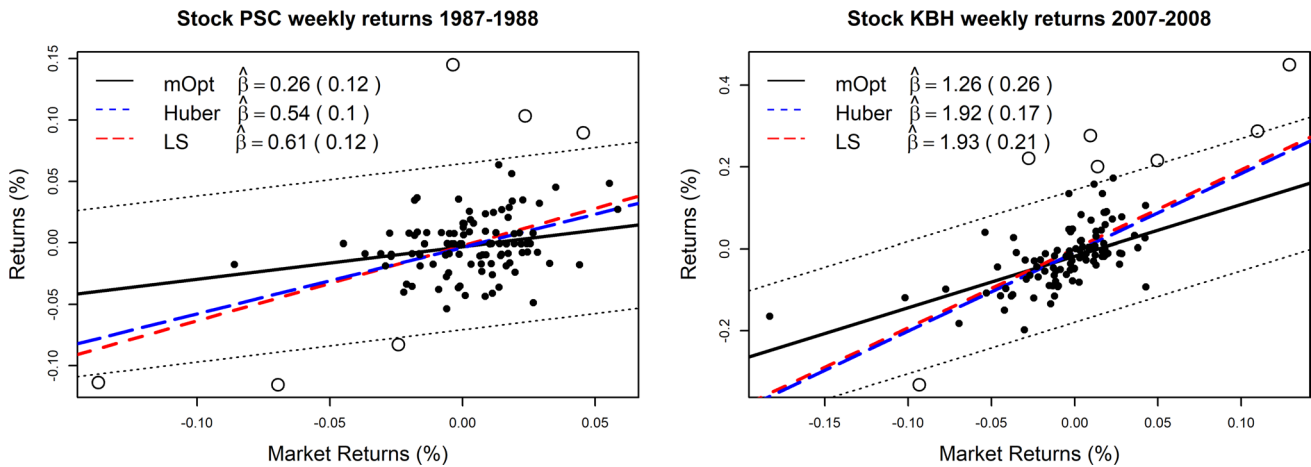**Fig. 4** mOpt and LS fits for OFG and DD

Fig. 5 mOpt, Huber and LS fits

provide a good fit to the bulk of the data. The difference of 2.26 between the LS and mOpt beta estimates of OFG would be of concern to any financial analyst or portfolio manager. Furthermore, the 2.26 difference in the OFG LS and mOpt slopes is a littler over 9 times the mOpt robust standard error value of 0.25, indicating that the difference between the LS and mOpt beta estimates is very highly significant for OFG.

The results for the stock with ticker DD in the right-hand plot of Fig. 4, in which there is a huge outlier in lower left corner of the plot, is dramatically different from the left-hand plot in that the mOpt and LS fits are essentially identical for the stock DD. That huge outlier is due to the October 19, 1987 ("Black Monday") U.S. market crash, when the Dow Jones fell 22.6%. The mOpt robust estimator does not reject that outlier because it is consistent with the mOpt fit of DD returns to the market returns for the rest of the

non-outliers data. This is a general property of the mOpt estimator that makes good sense.

Huber estimator lack of robustness examples

The lack of robustness of the Huber regression estimator discussed in Sect. 2.5 manifests itself in Huber beta estimates having slopes that are sometimes quite close to those of the LS estimate model fits. This behavior is illustrated in the two plots of Fig. 5, where the Huber estimators are almost the same as the LS estimators, and therefore provide no protection against outliers bias. It is the vulnerability to *leverage outliers* that cause Huber estimate to be so close to the LS estimate in the figure, where a leverage outlier in the present context is loosely defined to be a market return and stock return pair for which the market return is extreme and the data pair is not consistent with the a robust fit to the bulk of
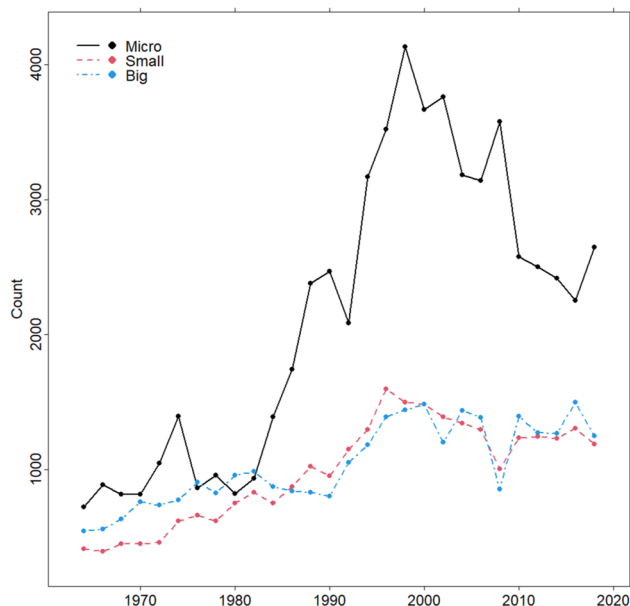
Fig. 6 Biweekly liquid CRSP stocks market capitalization group counts and percentages

the data.[12] In this regard, the October 20, 1987 market return in the lower-left of the right-hand plot of Fig. 4 is a market return outlier but it is not a leverage outlier since the market return and stock return pair is consistent with the robust fit to the bulk of the data.

While the Huber beta estimate is sometimes close to the LS beta estimate, as in the extreme case of the nearly equal betas in Fig. 5, it turns out the Huber beta estimate is also sometimes close to or nearly equal to the mOpt beta estimate. In order to quantify the extent to which the Huber beta estimate is close to the LS beta estimate, we computed the percent of the stocks in each of the 28 cross sections studied for which the Huber beta is closer to the LS beta than the mOpt beta. The percent ranges from a minimum of 34% and a maximum of 47%, which we regard as a substantial lack of robustness of the Huber estimators. Consequently, the Huber estimator should only be used with full recognition of this limitation, and we strongly recommend instead the mOpt estimator when considering a robust regression estimator in practice.

The cross section distribution of robust mOpt and LS betas

The existence of differences between robust mOpt betas and LS betas, such as those discussed above, motivated us

to carry out a detailed empirical analysis of the behavior of these two beta estimates for large cross sections of U.S. stocks from the CRSP database. Specifically, we computed mOpt and LS betas using two years of weekly returns of "liquid" stocks during each of the 28 contiguous two-year intervals during 1963 to 2018. Here, we take liquid stocks to be all those in the CRSP database that have at least 100 non-zero returns in a two-year interval. Then, we split the mOpt and LS beta estimates into microcap, smallcap and bigcap groups, where we follow Fama and French (2008) and Hou et al. (2020) in using the 20th and 50th percentiles of the NYSE capitalization data as breakpoints.

Figure 6 shows the time series of the resulting biennial microcap, smallcap and bigcap group counts. The rapid growth of microcap stocks from the early 1980's to the late 1990's is striking, as is the abrupt subsequent reversal to declining numbers of microcaps that was precipitated by the dot-com bubble collapse. From the early 1990's until just before the onset of the financial crisis in 2007, the relative growth of microcap, smallcap and bigcap stocks was such that they had approximately constant percentages of the market during that time period. Then, after the momentary decrease in the number of small and bigcap stocks and increase in the number of microcap stocks during the 2007-2008 crisis, the counts of the three cap groups were fairly constant. The dominant numbers of microcap stocks from the mid 1980's onward are of interest here because microcap stock returns are more volatile and prone to outliers than smallcap and bigcap stocks, and this leads one to anticipate that the adverse influence of outliers on LS beta estimates will be most severe and frequent for microcap stocks, somewhat less frequent for midcaps, and even less frequent for bigcaps.

With the preceding anticipated behavior in mind, we turn to Fig. 7 display of the boxplot representation of the cross section distribution of mOpt robust betas for each contiguous two-year interval between 1963 and 2018. The solid dot in each boxplot represents the cross section median of the betas for each two-year interval, the empty circle symbol represents the average of the betas, and the ends of the boxes are located at the lower and upper quartiles of the data.

The most striking features of the robust mOpt betas boxplots in Fig. 7 are: (a) They reveal interactions between beta and firm size, with almost all mOpt median betas less than 1 for microcaps, a little more than one half of mOpt median betas less than 1 for smallcaps, and most betas quite close to 1 for bigcaps; (b) the temporal patterns of the distributions of mOpt betas for the microcap and smallcap groups are strikingly similar from 1963 through end of the collapse of the dotcom bubble in 2002, and over that time interval both groups have most median betas less than 1 and often considerably so. Then for 2002 through 2018 the microcap and smallcap mOpt betas have similar increasing temporal

---

[12] Our usage of the term *leverage outlier* here is non-standard. For a discussion of leverage data in regression diagnostics, see Belsley et al. (1980)
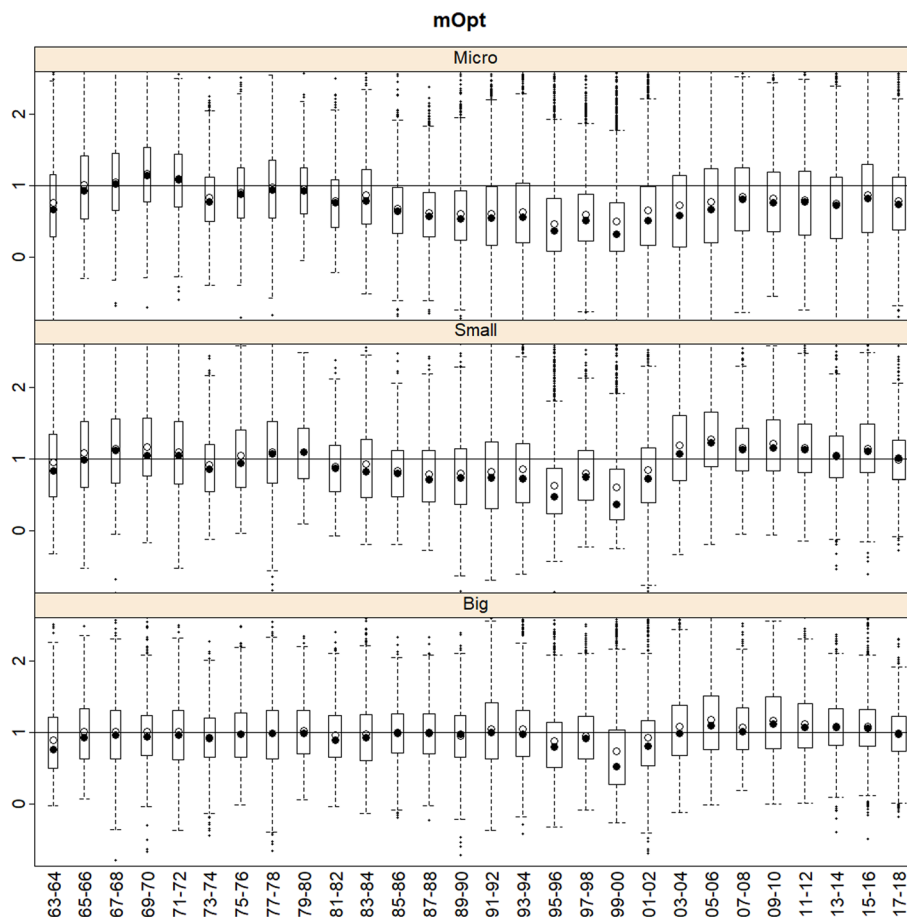
patterns, with the medians of the microcaps remaining below
one and settling around 0.75 toward the end of that interval,
while during that time period the medians of the smallcaps
drift to slightly above 1; (c) the mOpt beta medians of the
bigcap group only deviate substantially from being close to
one during 1963–1964, 1995–1996, and most substantially
during the time period 1999–2002 when the dotcom bubble
collapsed and subsequent rebound initiation. The temporal
pattern of the cross section distributions of the bigcap and
smallcap stocks are strikingly similar from 1999 through
2018.

Figure 8 displays the boxplots of the paired differences
between LS and mOpt betas, i.e., LS beta minus mOpt beta
for each stock, with horizontal dotted lines at +0.3 and -0.3.
These lines were chosen for the following reasons: (a) Data
values that fall outside the "whiskers" of a boxplot are by
convention deemed outliers, which are plotted as individual
points. For the bigcap boxplots in the lower panel, the aver-
age locations of the whiskers are at about $+ 0.3$ and $- 0.3$,
and so beta differences larger than +0.3 and smaller than
$- 0.3$ for bigcaps are considered to be LS minus mOpt beta
outliers. We then use +0.3 and $- 0.3$ as boundary lines for
the smallcap and microcap as well because those choices
clearly reveal the increased dispersion and skewness of the

LS minus mOpt betas distributions and outliers, relative to
the bigcaps; (b) We believe that a broad range of consum-
ers of beta estimates, who have the market beta value of
1 as a reference point, will find that LS minus mOpt beta
differences with absolute values greater than 0.3 will be a
useful alert that outliers are likely to be influencing the LS
estimate, and that the asset returns data should be examined
for influential outliers, and possible firm financial conditions
or market conditions that may give rise to such outliers.

It is clear in Fig. 8 that the cross section distributions of
the paired differences between the LS and mOpt betas tend
to be concentrated on positive values to various degrees,
with this effect most substantial for microcaps for all 28
two-year intervals during 1963 to 2018, and most striking
for the 1975–1976, 1987–1988, 2007–2008 intervals asso-
ciated with major market events, where well over 25% of
LS betas are larger than the mOpt robust betas by at least
0.3.[13] For smallcaps, the clear tendency for LS and mOpt

---

[13] The 1975–1976 interval was in the aftermath of the 1973–1974
market crash and recession that lasted until early 1975, the 1987–
1988 interval contained the Black Monday market crash of October
19, 1987, and of course 2007–2008 contained the financial crisis.

Fig. 8 LS minus mOpt beta differences cross section distribution boxplots using the same returns data as for Fig. 7
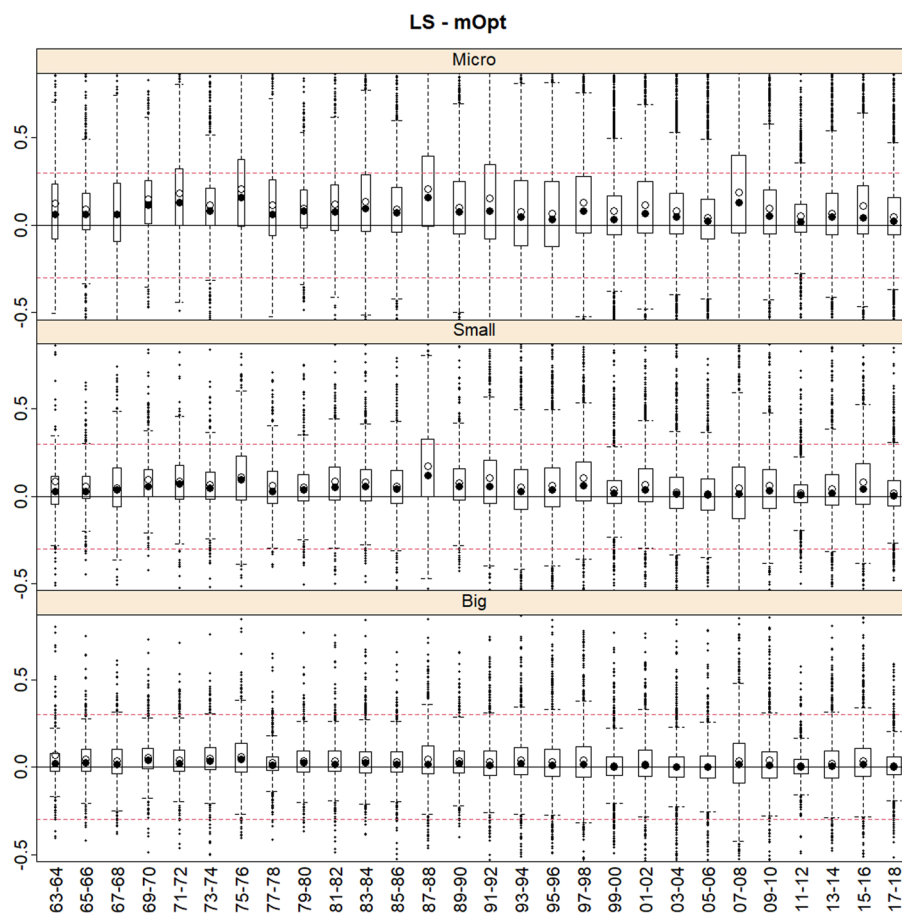

Table 1 1963–2018 average percent of stocks in marketcap groups and in the market for which the absolute differences of LS and mOpt betas exceed thresholds of 0.3 and 0.5, along with ratios of positive to negative exceedances

Threshold	Cap group percent				Positive/negative excedances ratio			
	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

beta differences to be concentrated on positive values in the early years diminishes over time, and is essentially absent after 2002, and a similar comment applies to the bigcap beta differences.

We also notice from Fig. 8 that for each two-year interval there exist at least some and often many stocks for which the absolute difference between LS and mOpt beta estimates exceed not only the value 0.3, but also the value 0.5. In order to quantify the frequency of such differences, we tabulated for each two-year interval the percent of stocks in each market cap group and and the market for which the absolute difference of LS versus mOpt betas exceeds level thresholds of both 0.3 and 0.5. We then computed the average of these percentages over all the two-year intervals between 1963 and 2018, and display the results in the

Cap Group Percent columns of Table 1. The results clearly reveal that the problem of outlier bias influence on LS betas is most frequent for microcaps, less so for smallcaps, relatively infrequent for bigcaps, and significant for the overall market. The market wide exceedance frequencies of 18.1% and 7.5% for thresholds 0.3 and 0.5, respectively, are results that practitioners should be aware of. The Positive/Negative Exceedances ratios in Table 1, that range from 2.7 to 4.2, reflect the positive skewness of the distribution of LS minus mOpt betas.

One may argue that relative differences in LS and mOpt betas are more meaningful than absolute differences, which is reasonable enough when both betas are larger than 1. However, when both betas are small, relative differences are not so meaningful, e.g., for LS and mOpt betas are .2

Table 2 1963–2018 average percent of stocks in market cap groups and the market for which an absolute difference of LS and mOpt betas relative to the mOpt beta exceed thresholds of 0.3 and 0.5, along with ratios of positive to negative exceedances

Threshold	Cap group percent				Positive/negative exceedances ratio			
	Micro	Small	Big	Market	MicroPN	SmallPN	BigPN	MarketPN
All mOpt Betas								
0.3	41.0	21.9	11.4	28.8	3.1	5.3	4.8	3.4
0.5	27.8	12.3	5.0	18.4	3.3	6.1	7.6	3.6
|mOpt Beta|>0.5								
0.3	27.3	13.6	6.8	17.6	4.0	6.9	5.5	4.2
0.5	13.7	5.1	1.8	7.9	5.7	10.2	6.4	6.1

and .05 the relative difference is 3.0. So in Table 2 we display exceedance ratios of the absolute difference between LS and mOpt betas relative to mOpt betas for not only all mOpt estimates, but also only for mOpt betas whose absolute values are at least 0.5. It is not surprising that the cap group ratios are substantially smaller when we only look at mOpt betas with absolute value larger than 0.5. We also find the interesting fact that the values in the "|mOpt Beta|> 0.5" rows of Table 2 are rather similar overall to those in Table 1.

Thin trading consideration

The downward bias of the median behavior of the mOpt beta estimates, which is also the case for LS estimates, raises the question of the extent to which this is due to thin trading bias. We studied this issue using the Dimson (1979) method, corrected as in Davidson and Josev (2005), for both mOpt and LS beta estimators. It turns out that this correction indeed reduces the thin trading bias, but it also inflates the variability of the betas, and correspondingly results in larger exceedance percentages than those in Table 1. Our study of a robust Dimson method needs to be extended to include the Scholes and Williams (1977) correction method.

Multifactor time series models

Our examples in this section are focused on fitting the FF3 and FFC4 time series models to and asset returns time series r_t. The FFC4 model has the form

$$r_t^e = \alpha + f_{MKT,t}^e \beta_1 + f_{SMB,t} \beta_2 + f_{HML,t} \beta_3 + f_{MOM,t} \beta_4 + \epsilon_t \quad (19)$$

where r_t^e is a time series of the asset excess returns relative to a risk-free rate, $f_{MKT,t}^e$ is a time series of market excess returns, $f_{SMB,t}$ are the returns of the Fama-French "small minus big" (SMB) size factor portfolio, $f_{HML,t}$ are the returns on the "high minus low" (HML) value factor portfolio, $f_{MOM,t}$ are the returns on the Carhart (1997) momentum

factor portfolio, and α is an intercept. The FF3 model is obtained by dropping the term $f_{MOM,t} \beta_4$.[14]

It is common practice to use time series factor models such as the FF3 and FFC4 models for analysis where r_t is a portfolio return. However, the models can also 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. Figure 9 displays the pairwise scatter plots 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 results of the LS and mOpt fits of the FF3 and FFC4 models to the FNB returns are displayed in Table 3 coefficient estimates, standard and robust t-statistics, and standard and robust adjusted R-squared values.[15] Since we are no longer in the context of new factor significance discovery, we use the usual absolute *t*-value threshold of 1.96 for a t-statistic to be significant. For both models the intercept estimates of both the LS and mOpt fits are equal to zero to two significant digits, and correspondingly the t-statistics 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.RF and SMB as significant factors, but the mOpt fit t-statistics are significant for all but the HML factor, whose

---

[14] The time series of the four factors in the model (19) are available at Professor Ken French's website.https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

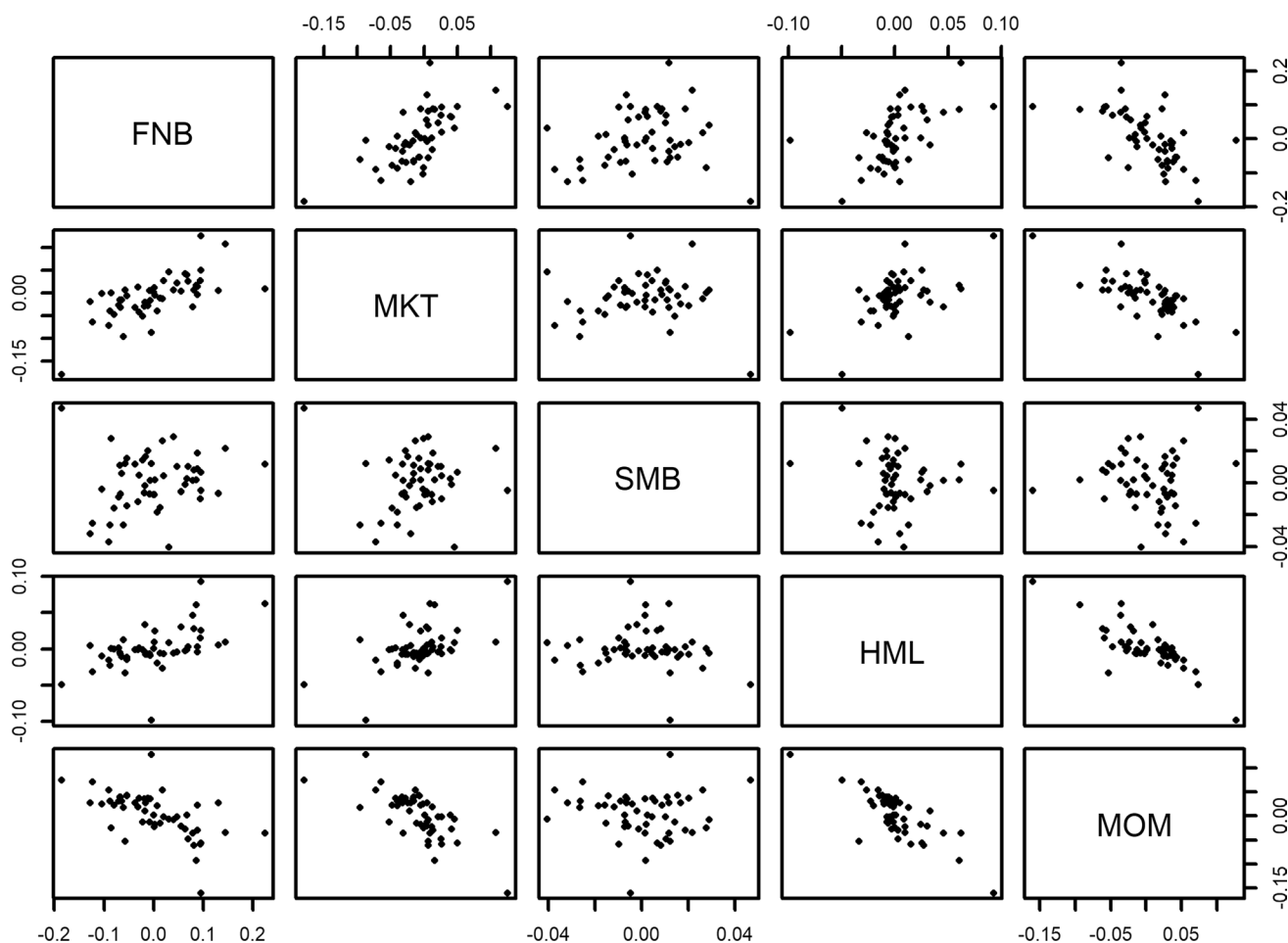[15] For the robust adjusted R-squared, see Section 5.12.3 of Maronna et al. (2019).

Fig. 9 Pairwise scatter plots of FNB 2008 weekly returns and corresponding MKT, SMB, HML, MOM factors

Table 3 mOpt and LS fits of FF3 and FFC4 models to FNB 2008 weekly returns

	Alpha	MKT	SMB	HML	MOM	AdjRSQ
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

t-statistic is quite insignificant, and its MOM slope is negative. This latter result is not surprising in view of the fact, illustrated in Fig. 9, that the HML and MOM factors are negatively correlated, and MOM replaces HML in the FFC4 model. Table 3 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 that the FFC4 model is preferred to the FF3 model. However, a proper robust model selection method is needed for this purpose, and we discuss this in the next section.

It is useful to compare the mOpt and LS fits of the FFC4 model using Fig. 10 model fit residuals normal *QQ plots* (quantile-quantile plots), where the one on the left is for the LS fit and the one on the right is for the mOpt fit. The dotted lines in the figure are the point-wise 95% confidence intervals. The QQ plot for the LS fit in the left panel has a number data points in the central region that deviate from a straight
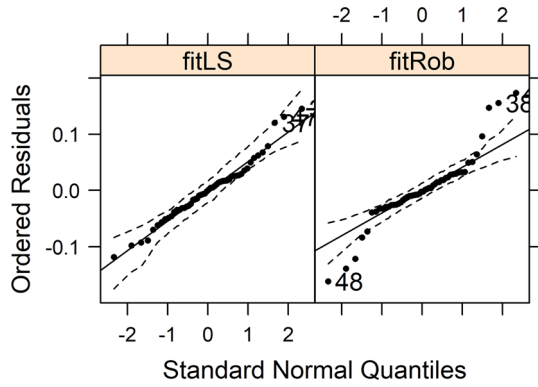
Fig. 10 Normal qqplots of residuals from robust LS and mOpt fits of the FFC4 model to FNB 2008 weekly returns

Table 4 Tests for differences between the mOpt and LS overall model and individual coefficients

	p values				
	MODEL	MKT	SMB	HML	MOM
FF3 Fit	0.000	0.476	0.822	0.000	
FFC4 Fit	0.011	0.432	0.656	0.199	0.003

line and are near or outside the dotted line for the bulk of the data, which indicates that the model fit is not very good, and there is no indication of residuals outliers. On the other hand, the majority of the data points in the mOpt QQ plot in the right panel fall along a straight line that is well within the dotted lines, thereby indicating a good model fit for most of the data, and the extreme regions of the qqplot reveal a number of positive and negative residuals outliers that are well outside the dotted lines. This is the typical behavior of the residuals from an mOpt robust estimate relative to an LS estimate, i.e., the robust regression fit is very good for the bulk of the data in the middle of the distribution and outliers are clearly revealed, but this is not the case for a LS regression.

In addition to above visual assessment of comparative LS and mOpt robust fits, and standard statistical assessment as provided in Table 3, one can use a significance test statistic to test for a non-zero difference between and LS and mOpt fits. For estimators $\hat{\boldsymbol{\beta}}_{LS}$ and $\hat{\boldsymbol{\beta}}_{mOpt}$ that converge in probability to the true parameter vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$, it was shown by Maravina and Martin (2012) that the difference $\triangle\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{LS} - \hat{\boldsymbol{\beta}}_{mOpt}$ converges in distribution to a zero mean multivariate normal distribution with covariance matrix

$$\mathbf{V}_{\triangle\hat{\boldsymbol{\beta}}} = \delta^2 \mathbf{C_f}^{-1} \qquad (20)$$

where $\mathbf{C_f}$ is the covariance matrix of factor returns appearing in (9), and

$$\delta^2 = \mathrm{E}\left(\frac{s\psi_{\mathrm{mOpt}}(u/s)}{\mathrm{E}\psi'_{\mathrm{mOpt}}(u/s)} - u \right)^2 \qquad (21)$$

where $u = s\epsilon_1$ in (18).[16] It follows that the quadratic form $\triangle\hat{\boldsymbol{\beta}}' \delta^2 \mathbf{C_f}^{-1} \triangle \hat{\boldsymbol{\beta}}$ has an asymptotic chi-squared distribution with K degrees of freedom, and the significance test statistic

$$\mathrm{D}_K = \triangle\hat{\boldsymbol{\beta}}' \left(\frac{\hat{\delta^2}}{T} \hat{\mathbf{C}}_{\mathbf{f}}^{-1} \right) \triangle \hat{\boldsymbol{\beta}} \qquad (22)$$

has an approximate chi-squared distribution with K degrees of freedom. For our applications here where we use two years of weekly returns, a normal distribution approximation will suffice for Δ_{beta}, and hence a chi-squared approximation for D_K will suffice for computing p values.

The p values obtained by applying the above test statistic for the FF3 and FFC4 models are displayed in Table 4. For the FF3 model, the $\triangle\hat{\boldsymbol{\beta}}_K$ difference and the HML coefficient difference are both significant with common p values of 0.000 to three digits. This is not too surprising in view of the large difference between the LS and mOpt HML coefficients in Table 3, but neither of the MKT and SMB factor LS and mOpt coefficient differences are significant. On the other hand for the FFC4 model the $\triangle\hat{\boldsymbol{\beta}}_K$ difference is is quite significant with a p value 0.011, and the MOM factor difference is highly significant with a p value 0.003, but none of the differences in the MKT, SMB and HML factors are significant.

Robust time series model selection

While the robust regression results in Table 3 suggest that the FFC4 is a better factor model than the FF3, use of a proper model selection method that penalizes for over-fitting is needed when searching for a "best" model. The normal distribution theory based *Akaike information criterion* (AIC) criterion introduced by Akaike (1973) is commonly used for this purpose. In the case of time series linear regression models with sample size T and number of variables $p \leq p_{\max}$, the AIC model selection method chooses the model order p that minimizes $T\log(\hat{\sigma}) + p$, where $\hat{\sigma}^2$ is the

---

[16] The expression for δ^2 takes into account the covariance between the LS and mOpt estimators, and the higher the correlation between LS and mOpt estimators the smaller the δ^2 will be. In the estimator version of the above expression, where the expectations are replaced by sample averages, the u_i in the first term will be the mOpt fit residuals, the u_i in the second term will be the LS fit residuals, and the estimate $\hat{s}$ of s will be the robust residuals scale estimate computed by the mOpt estimator.

Table 5 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		

Table 6 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	MOM	− 292.4
MOM	− 296.5		

average of the squared regression residuals. The AIC criterion lacks robustness in two ways, the first is the lack of robustness of the least squares coefficient estimates used to compute the residuals, and the second is that outliers can adversely influence the average of the squared residuals.[17] We propose to replace AIC with a *robust final prediction error* (RFPE) method that makes natural use of the mOpt robust regression method.

The RFPE method was motivated by the *final prediction error* (FPE) method introduced by Akaike ([1969](#)) in the context of estimating the order of an autoregressive time series. The FPE method chooses the autoregression order k that minimizes

$$\text{FPE}(k) = \hat{\sigma}_k^2 \frac{T + k}{T - k} \approx \hat{\sigma}_k^2 \left(1 + 2 \frac{k}{T} \right) \text{ for } k \ll T \tag{23}$$

where $\hat{\sigma}_k^2$ is the average of the squared residuals from the least-squares fit. For linear regression models the FPE criterion is used by considering a maximal linear model that has K predictor variables, computing FPE(k) for subset models indexed by C_k with $k \leq K$ predictor variables, and finding the subset C_k^* that minimizes FPE(k). However, the FPE criterion lacks robustness for the same reasons that the AIC lacks robustness

The RFPE robust model selection method is defined as follows using the mOpt regression estimator in an intuitive manner.[18] Let

$$\hat{\boldsymbol{\beta}}_{C_k} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left[\sum_{t=1}^{T} \rho_{\text{mOpt}} \left(\frac{r_t - \tilde{\mathbf{f}}'_{C_k,t} \boldsymbol{\beta}}{\hat{s}} \right) \right] \tag{24}$$

where the robust residuals scale estimate $\hat{s}$ is computed for the maximal model C_{max} with K predictor variables $\tilde{\mathbf{f}}_{C_{\text{max}},t}$.

This serves to provide a smallest robust scale that serves as an equitable scale reference across models with different predictor variables. With robust residuals for the subset model C_k defined as

$$\hat{\epsilon}_{C_k,t} = r_t - \tilde{\mathbf{f}}'_{C_k,t} \hat{\boldsymbol{\beta}}_{C_k}, \, t = 1, 2, \ldots, T.$$

the RFPE is defined as

$$\text{RFPE}(C_k) = \frac{1}{T} \sum_{t=1}^{T} \rho_{\text{mOpt}} \left(\frac{\hat{\epsilon}_{C_k,t}}{\hat{s}} \right) + \frac{k}{T} \frac{A}{B} \tag{25}$$

where

$$A = \frac{1}{T} \sum_{t=1}^{T} \psi_{\text{mOpt}}^2 \left(\frac{\hat{\epsilon}_{C_k,t}}{\hat{s}} \right),$$
$$B = \frac{1}{T} \sum_{t=1}^{T} \psi'_{\text{mOpt}} \left(\frac{\hat{\epsilon}_{C_k,t}}{\hat{s}} \right). \tag{26}$$

Note that if $\rho_{\text{mOpt}}(r)$ is replaced with the least squares rho function $\rho(r) = r^2$, then $\psi(r) = 2r, \psi'(r) = 2$, and RFPE($C_k$) is the same as the approximate version of ([23](#)).[19]

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 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 the R package RobStatTM.[20] We illustrate the use of the method for the problem of finding a best subset of the FFC4 factors for modeling the FNB stock returns, with the results shown in Table [5](#).

The Full Model columns of the table 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 3-factor

---

[17] Similar comments apply to the *Bayesian information criterion* of Schwarz ([1978](#)) and Akaike ([1978](#)).

[18] The same definition applies to other bounded rho function robust regression estimators.

[19] Details concerning the derivation of ([25](#)) and ([26](#)) are available in Sections 5.6.2 and 5.13.7 of Maronna et al. ([2019](#)).

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

models obtained by deleting each of the MKT, SMB, HML and MOM factors one at a time. Since 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. Since 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 Table 3, where the HML factor is quite insignificant.

We also used least-squares based AIC model selection for the FFC4 factors, and the results are shown in Table 6. 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 above RFPE best set of factors where MOM is retained instead of HML. This shows that AIC, like LS, suffer from the adverse influence of outliers and can not serve as a reliable method for model selection.

While the above example is quite simple it suggests possible considerable usefulness of routine selection of time series factor model factors with RFPE. Among the many possible applications, one would be to model portfolio or asset style drift over time using RFPE model selection on a moving time window.

Summary and discussion

We introduced a theoretically well-justified mOpt robust regression method in the context of time series factor models. The efficacy of the mOpt regression in providing a good fit to the bulk of the data is demonstrated for both single factor and multifactor models where LS fits are adversely influenced by outliers. We also showed that the well-known Huber robust regression estimator is lacking in bias robustness, both theoretically and in applications.

For the case of single factor model CAPM betas, we compared the mOpt and LS estimates for two-year intervals of weekly returns of liquid stocks from the CRSP database for 1963 to 2018. The lack of robustness toward outlier of the LS estimator is manifested in absolute beta value differences between mOpt and LS estimates of at least 0.3 for about 18% of the liquid stocks in the market, and at least 0.5 for about 7.5% of those stocks. Our analysis showed that the lack of robustness of the LS estimator is revealed most frequently

for microcap stocks, less frequently for smallcaps, and considerably less frequently for bigcaps. Such differences can be of concern to a wide range of users of beta estimates.

For the Fama-French three-factor model (FF3) and the Fama-French-Carhart four-factor model (FFC4) fits to weekly stock returns that we studied, the mOpt estimator continues to produce good fits to the bulk of the data that differ significantly from LS fits. An important feature of mOpt is that in addition to robust coefficients, it computes robust t-statistics. For the FF3 model, the robust t-statistics lead to same significant Market, SMB and HML factors as the LS fit classical t-statistics. But for the FFC4 model the robust t-statistics result in significant Market, SMB and MOM factors, whereas the LS classical t-statistics only lead to significant Market and SMB factors. We also introduced a new test of significant difference between mOpt and LS estimates, and illustrated its usefulness in the context of FF3 and FFC4 model fitting. Finally, the new Robust Final Prediction Error (RFPE) model selection criterion produces better fitting factor models than the well-known normal distribution based AIC method, which is quite non-robust toward outliers.

The above results are only first steps in understanding the use of mOpt robust regression estimators, robust t-statistics, and RFPE model selection for time series multifactor models. Additional directions for further research include the following. The first is to carry out an in-depth empirical study of the relative performance of the mOpt versus LS fits of the FF3 and FFC4 models, and the RFPE versus AIC model selection, for cross sections of liquid stocks, similar to what we have done for single factor models herein. A second step will be to do a similar empirical study for other increasingly popular multifactor models, e.g., the Fama and French (2015) 5-factor model, and the Hou et al. (2020) augmented q-factor model. Furthermore, it will be important to assess the efficacy of mOpt robust versus non-robust LS fits of multifactor models in evaluating the character and performance of funds, where individual stock returns outliers influence are mitigated by the portfolio nature of the funds, but the fund returns as well as the factors can none-the-less contain outliers. Finally, we have developed a vary natural way to robustify the Vasicek (1973) Bayesian beta estimates using the mOpt estimator, and it remains to study its empirical efficacy for cross sections of returns.

The results presented in this paper lead us to strongly recommend standard use of the mOpt robust regression estimator as a complement to least-squares for time series factor models research and applications, as well as a standard against which to evaluate alternative proposed robust regression methods.

Appendix

The mOpt estimator psi function

The formula for the mOpt psi function is

$$\psi_{mOpt,c}(x) = \begin{cases} x & |x| \le 1 \\ \dfrac{\phi(1)}{\phi(1) - a}\left(x - \text{SGN}(x)\dfrac{a}{\phi(x)}\right) & 1 < |x| < c \\ 0 & |x| \ge c . \end{cases} \tag{27}$$

where $\phi(x)$ is the standard normal density function, and the constants a and c depend on the desired normal distribution efficiency. The corresponding rho function $\rho_{mOpt,c}(x)$ is obtained from the psi function $\psi_{mOpt,c}(x)$ by integration, and it is clear from the definition of the psi function that the rho function is constant outside the interval $(-c, +c)$. Details concerning the rho and psi functions, including how to determine the constants a and c to achieve a desired normal distribution efficiency are provided in Konis and Martin (2021). A 95% normal distribution efficiency is obtained with $a = 0.0132$ and $c = 3.00$.

Normal distribution efficiency

Since the asymptotic covariance matrices of regression M-estimators all have the form (6), one can define the *normal distribution efficiency* EFF(ψ, Φ) of any regression M-estimator relative to the LS estimator as the ratio of $V_{loc}(\psi_{LS}; \Phi) = s^2$ to $V_{loc}(\psi; \Phi)$, and in particular to $V_{loc}(\psi_{mOpt}; \Phi)$ in the case of the mOpt estimator. Using the formula (8) with $F_0 = \Phi$ and $\psi = \psi_{mOpt,c}$ where c is the tuning constant gives the following expression for the normal distribution efficiency of the mOpt estimator:

$$\text{EFF}(\psi_{mOpt,c}, \Phi) = \left(\frac{E_\Phi \psi^2_{mOpt,c}(\epsilon_1)}{E^2_\Phi \psi'_{mOpt,c}(\epsilon_1)}\right)^{-1}. \tag{28}$$

The value $c = 3.00$ used for the 95% normal distribution efficiency mOpt throughout this paper is easily obtained via numerical integration and root finding using the $\psi_{mOpt,c}$ formula (27) as described in Konis and Martin (2021).

The expression (8) is the formula for a location M-estimator asymptotic variance derived in Huber (1964).

mOpt coefficient estimates standard errors

The standard errors of regression M-estimator coefficient estimates $\hat{\beta}_{k,n}$ of $\hat{\boldsymbol{\beta}}_n$ are computed using the formulas

$$SE(\hat{\beta}_{k-1}) = \left[\frac{1}{T}\hat{v}_{loc}\left(\hat{\mathbf{M}}^{-1}_{\tilde{f}}\right)_{kk}\right]^{1/2}, \ k = 1, 2, \ldots, K+1 \tag{29}$$

where $\hat{v}_{loc}(\psi)$ is an estimate of $v_{loc}(\psi; F_0)$ and $\hat{\mathbf{M}}_{\tilde{f}}$ is an estimator of the moment matrix $\mathbf{M}_{\tilde{f}}$ given by (7). The estimate $\hat{v}_{loc}(\psi_{mOpt})$ of the expression (8) with $\psi = \psi_{mOpt}$ is:

$$\hat{v}_{loc}(\psi_{mOpt}) = \hat{s}^2 \frac{\dfrac{1}{T-(K+1)}\sum_{t=1}^{T}\psi^2_{mOpt}\left(\dfrac{r_i - \tilde{\mathbf{f}}'_t\hat{\boldsymbol{\beta}}}{\hat{s}}\right)}{\left[\dfrac{1}{T}\sum_{t=1}^{T}\psi'_{mOpt}\left(\dfrac{r_i - \tilde{\mathbf{f}}'_t\hat{\boldsymbol{\beta}}}{\hat{s}}\right)\right]^2}. \tag{30}$$

Note that for the LS estimator obtained when $\psi(x) = x$, the above expression reduces to the usual least squares unbiased estimate of the error variance with divisor equal to the sample size minus the number of parameters. The above standard errors are used in the usual way to compute *t*-statistics and *p* values.

In the RobStatTM package, the robust regression function `lmrobdetMM` for computing the mOpt regression estimate uses the moment matrix estimate

$$\hat{\mathbf{M}}_{\tilde{f}} = \frac{1}{\sum_{t=1}^{n} w_t(\hat{\boldsymbol{\beta}};\hat{s})}\sum_{t=1}^{n} w_t(\hat{\boldsymbol{\beta}};\hat{s})\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}'_t \tag{31}$$

where the weights $w_t(\hat{\boldsymbol{\beta}};\hat{s})$ are given by (12). The motivation for the weights in the above expression is to down-weight outlier vectors $\tilde{\mathbf{f}}_t$ for which the corresponding regression residual $r_t - \tilde{\mathbf{f}}'_t\hat{\boldsymbol{\beta}}$ is large, and thereby avoid adverse influence of such $\tilde{\mathbf{f}}_t$ on the estimate $\hat{\mathbf{M}}_{\tilde{f}}$. Note that if $\tilde{\mathbf{f}}_t$ is an outlier vector for which the regression residual is $r_t - \tilde{\mathbf{f}}'_t\hat{\boldsymbol{\beta}}$ sufficiently small then we will have $w_t(\hat{\boldsymbol{\beta}};\hat{s}) = 1$, and correspondingly $\tilde{\mathbf{f}}_t$ is not altered in the estimate $\hat{\mathbf{M}}_{\tilde{f}}$. It is argued in Section 5.6 of Maronna et al. (2019) that under the normal distribution model $\hat{\mathbf{M}}_{\tilde{f}}$ is a consistent estimator of $\mathbf{M}_{\tilde{f}}$.

References

Akaike, H. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21: 243–247. https://doi.org/10.1007/bf02532251.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second international symposium on information theory*, ed. B. Petrov and F. Csaki, 267–281. Budapest: Akademiai Kiado.

Akaike, H. 1978. A new look at the Bayes procedure. *Biometrika* 65: 53–59. https://doi.org/10.1093/biomet/65.1.53.

Bailer, H.M., T.A. Maravina, and R.D. Martin. 2011a. Robust betas in asset. *Management*. https://doi.org/10.1093/oxfordhb/9780199553433.013.0011.

Bailer, H., T.A. Maravina, and R.D. Martin. 2011b. Robust betas in asset management. In *Oxford handbook of quantitative asset management*, ed. B. Scherer and K. Winston, 203–242. Oxford: Oxford University Press.

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression diagnostics. Identifying influential data and sources of collinearity*. New York: Wiley. https://doi.org/10.1002/0471725153.

Carhart, M.M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52: 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Chan, L.K.C., and J. Lakonishok. 1992. Robust measurement of beta risk. *Journal of Financial and Quantitative Analysis* 27: 265. https://doi.org/10.2307/2331371.

Chen, N.-F., R. Roll, and S.A. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59: 383. https://doi.org/10.1086/296344.

Cloete, G.S., P.J. de Jonah, and T. de Wet. 2002. Combining Vasicek and robust estimators for forecasting systematic risk. *Investment Analysts Journal* 31: 37–44. https://doi.org/10.1080/10293523.2002.11082435.

Cornell, B., and J.K. Dietrich. 1978. Mean-absolute-deviation versus least-squares regression estimation of beta coefficients. *Journal of Financial and Quantitative Analysis* 13: 123. https://doi.org/10.2307/2330527.

Davidson, S., and T. Josev. 2005. The impact of thin trading adjustments on Australian beta estimates. *Accounting Research Journal* 18: 111–117. https://doi.org/10.1108/10309610580000679.

Dimson, E. 1979. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics* 7 (2): 197–226.

Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.

Fama, E.F., and K.R. French. 2008. Dissecting anomalies. *The Journal of Finance* 63 (4): 1653–1678.

Fama, E.F., and K.R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the factor zoo: A test of new factors. *Journal of Finance* 75 (3): 1327–1370. https://doi.org/10.3386/w25481.

Galea, M., et al. 2020. Robust inference in the capital asset pricing model using the multivariate t-distribution. *Journal of Financial and Risk Management* 13: 123. https://doi.org/10.3390/jrfm13060123.

Genton, M.G., and E. Ronchetti. 2008. Robust prediction of beta. In *Computational methods in financial engineering, essays in Honour of manfred Gilli*, ed. E.J. Kontoghiorghes, B. Rustem, and P. Winker, 147–161. Berlin: Springer. https://doi.org/10.1007/978-3-540-77958-2_8.

Hampel, F.R., et al. 1986. *Robust statistics: The approach based on influence functions*. New York: Wiley.

Hou, K., and M.A.V. Dijk. 2018. Resurrecting the size effect: Firm size, profitability shocks, and expected stock returns. *The Review of Financial Studies* 32 (7): 2851–2889. https://doi.org/10.2139/ssrn.1005664.

Hou, K., Mo, H. et al. 2020. An augmented q-factor model with expected growth. *Review of Finance* 24: 1–41.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* 28 (3): 650–705.

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *The Review of Financial Studies* 33 (5): 2019–2133. https://doi.org/10.1093/rfs/hhy131.

Huber, P. 1981. *Robust statistics*. New York: Wiley.

Huber, P.J. 1964. Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35: 73–101.

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 ed. New York: Wiley. https://doi.org/10.1002/9780470434697.

Knez, P., and Ready. 1997. On the robustness of size and book-to-market in cross-sectional regressions. *Journal of Finance* 52 (4): 1355–1382.

Konis, K., and Martin, R. D. 2021. Optimal bias robust psi and rho revisited. Available at SSRN: https://ssrn.com/abstract=3902862..

Maravina, T., and Martin, R. D. 2012. Tests for differences between least squares and robust regression coefficients. SSRN: https://ssrn.com/abstract=3902878

Maronna, R.A., et al. 2019. *Robust statistics: Theory and methods (with R)*, 2nd ed. New York: Wiley.

Martin, R.D., and T. Simin. 2003. Outlier resistant estimates of beta. *Financial Analysts Journal* 59 (5): 56–69.

Martin, R.D., V.J. Yohai, and R.H. Zamar. 1989. Min-max bias robust regression. *The Annals of Statistics* 17: 1608–1630.

Mills, T.C., and J.A. Coutts. 1996. Misspecification testing and robust estimation of the market model: Estimating betas for the FT-SE industry baskets. *The European Journal of Finance* 2: 319–331. https://doi.org/10.1080/13518479600000012.

Rousseeuw, P.J., and A.M. Leroy. 1987. *Robust regression and outlier detection*. New York: Wiley. https://doi.org/10.1002/0471725382.

Rousseeuw, P. J., and Yohai, V. J. 1984. Robust regression by means of S-estimators. Robust and nonlinear time series analysis, ed J. Franke, W. Härdie and D. Martin. Vol. 26. Lecture Notes in Statistics, 256–272. New York: Springer.

Scholes, M., and J. Williams. 1977. Estimating betas from nonsynchronous data. *Journal of Financial Economics* 5: 309–327. https://doi.org/10.1016/0304-405x(77)90041-1.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*. https://doi.org/10.1214/aos/1176344136.

Sharpe, W.F. 1971. Mean-absolute-deviation characteristic lines for securities and portfolios. *Management Science* 18: B-1-B-13. https://doi.org/10.1287/mnsc.18.2.b1.

Sharpe, W.F. 1988. Determining a funds effective asset mix. *Investment Management Review* 2 (6): 59–69.

Sharpe, W.F. 1992. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management* 18 (2): 7–19.

Svarc, M., Yohai, V. J. and Zamar, R. 2002. Optimal bias-robust M-estimates of regression. In *Statistical data analysis based on the l1 norm and related methods*, eds. Y. Dodge, 191–200. Birkhäuser

Tukey, J. W. 1960. A survey of sampling from contaminated distributions. In *Contributions to probability and statistics: Essays in honor of harold hotelling*, ed I. Olkin. Stanford. ISBN: 0804705968

Vasicek, O. A. 1973. A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas. The Journal of Finance 28,

pp. 1233-1239. issn: 0022-1082. https://doi.org/10.1111/j.1540-6261.1973.tb01452.x

Yohai, V.J. 1987. High breakdown-point and high efficiency robust estimates for regression. *The 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.

R. Douglas Martin Professor Martin began his academic career in Electrical Engineering at the University of Washington and subsequently moved to the Statistics Department, where he was the second Chair of Statistics. In 2012 he moved to the Applied Mathematics Department to develop a new self-sustaining MS degree program in Computational Finance and Risk Management. A consultant in the Mathematics and Statistics Research Center at Bell Laboratories from 1973 to 1983, Martin later founded Statistical Sciences to commercialize the S language for data analysis and statistical modelling in the form of S-PLUS. Subsequently he was a co-founder and Chairman of FinAnalytica, Inc., developer of the Cognity portfolio construction and risk management system, serving as CEO from 2006 to 2008. Martin has authored numerous publications on time series and robust statistical methods, and is co-author of two books: Modern Portfolio Optimization (2005) and the second edition of Robust Statistics: Theory and Methods (2018). His research is on applications of modern statistical methods in finance and investment. He holds a Ph.D. in Electrical Engineering from Princeton University. Martin retired from the Applied Mathematics and Statistics Departments in June 2016, and remains active in basic research projects with students, and former PhD students and other colleagues in the finance industry. He is also active in mentoring students in Google Summer of Code projects focused on quantitative finance open-source software.

Daniel Z. Xia is an independent quantitative finance researcher who received the M.S. Degree in Computational Finance at the University of Washington. Currently he works as a financial data scientist in the U.S. His research interests include empirical asset pricing, down-side risk portfolio optimization, and machine learning in finance. He is author and maintainer of the skewt Dist R package: https://github.com/dan9401/skewtDist.