

ROBUST ESTIMATION OF BETA

by

R. Douglas Martin and Tim Simin

TECHNICAL REPORT NO. 350

March 1999

**Department of Statistics
Box 354322
University of Washington
Seattle, WA 98195 USA**

ROBUST ESTIMATION OF BETA

R. Douglas Martin and Tim Simin*

ABSTRACT

The mounting evidence of outlier generating heavy-tailed or mixture distribution functions for equity returns motivates the search for estimation procedures robust toward disparate data points. This paper introduces such an estimator and presents a detailed comparison with the classical ordinary least squares (OLS) estimator using historical weekly and monthly equity returns data. The results show that the two estimates differ significantly for a non-trivial fraction of the equities studied, both statistically and economically. Furthermore, such behavior occurs primarily for firms with small market capitalization. Our comparison also documents the forecasting performance of the new robust beta estimate and relative behavior of the least absolute deviations (LAD) estimator of beta. The results of our study suggest that robust estimates of beta may be of considerable value as a complement to the standard OLS estimates.

* R. Douglas Martin is Professor of Statistics, University of Washington and Chief Scientist of the Data Analysis Products Division of MathSoft, Inc, both in Seattle, WA. (Email: doug@statsci.com). Tim Simin is with the Department of Finance and Business Economics, School of Business, University of Washington, Seattle, WA. (Email: tsimin@u.washington.edu). We are grateful to Wayne Ferson, Victor Yohai and seminar participants at the University of Washington for helpful comments.

1. INTRODUCTION

While academics continue to debate its relevance, the estimated slope coefficient of the classic market model is undoubtedly the best known and most used measure of market risk and return. For example, the widespread use of beta is confirmed in a firm survey concerning the current best practices of cost of capital estimation, by Bruner, Eades, Harris, Higgins (1996). These authors conclude that the CAPM is currently the favored model used for estimating the cost of equity and also note that beta estimates used in practice are drawn primarily from published sources. A review of how these published sources calculate beta reveals that even though many methodological differences exist, the most common estimation technique used to compute the slope coefficient in the market model

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + \varepsilon_{i,t} \quad t = 1, \dots, T \quad (1.1)$$

is ordinary least squares (OLS). Here, $R_{i,t}$ is the time series of the i -th equity's returns and $R_{M,t}$ is the time series of market (index) returns, both in excess of the risk-free interest rate for low frequency data or as raw returns when the frequency of the return data is weekly or higher.¹

The sanctified use of the least squares estimator for statistical modeling and inference is justified by the fact that OLS is the optimal estimate of linear model coefficients and provides a convenient distribution theory for inference when the errors are Gaussian. However, since the seminal work by Mandelbrot (1963) empirical investigations have produced evidence that equity returns may be generated by heavy-tailed or t-distributions, or by Gaussian mixture distributions. Fama (1965) provided evidence in favor of stable distributions, and Fama and Roll (1968, 1971) studied the properties and parameter estimates of symmetric stable distributions. Numerous authors such as Kon (1984), Roll (1988), Connolly (1989), and Richardson and Smith (1994), provide evidence in favor of the use of normal mixture distributions for modeling outliers in financial returns. The stylized facts produced by this literature characterize equity returns as both leptokurtotic and skewed. It is well known in the economics and statistics literature that outliers generated by such distributions can have a substantial influence on the values of least squares estimates.² Therefore, it is important to have alternative *robust* estimators that are not much influenced by outliers.

By now there exist a number of alternatives to least squares that are robust toward outliers according to various statistical criteria. In economics and finance, the *least absolute deviations* (LAD) estimate is perhaps the oldest and most widely known alternative to least squares. In the context of estimating beta, the LAD estimate was employed quite early on by Sharpe (1971) who considered thirty common stocks used to compute the Dow Jones Industrial Average and thirty mutual funds, both in the mid-to-late 1960's. Somewhat later Cornell and Dietrich (1978) also studied the LAD estimate using one hundred companies randomly drawn from the S&P 500 from 1962 to 1975.

These studies were motivated by the knowledge that returns often have outliers caused by non-Gaussian distributions, and that an alternative to least squares might therefore perform better. Both Sharpe and Cornell and Dietrich concluded that the LAD alternative did little to improve the OLS estimate of beta. These negative results are evidently due to the lack of influential outliers in the returns of the relatively large sized firms and mutual funds considered.

Other studies that have relaxed the normality assumption and consider alternative estimators suggest that the classical approach based on least squares and normal theory inference might not be the most preferred method. Koenker and Basset (1978) introduced a class of estimators based on regression-quantiles and established desirable robustness properties of these regression-quantile L-estimates. Connolly (1989) documented the effectiveness of using both M-estimates and regression-quantile type L-estimates as robust alternatives to OLS. The robust estimator results suggested that the weekend effect is much smaller than previously thought. Chan and Lakonishok (1992) showed that the regression quantile method could provide higher efficiency than OLS when estimating beta. Mackinlay and Richardson (1991) use a generalized method of moments approach to calculate and examine the bias in inferences of the unconditional CAPM beta when returns are generated by a (leptokurtotic) multivariate students- t distribution. They showed how deviations from the assumptions of multivariate normality drastically widen rejection regions for standard test statistics. Richardson and Smith (1994) use a similar technique to test a normal mixture distribution model for stock prices, and estimate the unobservable flow of information.

A significantly motivating application of robust regression is presented by Knez and Ready (1997), who used Rousseeuw's (1984) least trimmed squares (LTS) regression method in cross-sectional regressions of equity returns on firm size. These authors show that the OLS estimated negative risk premium on size reported by Fama and French (1992) is caused by a small fraction of outliers (less than 1%) in the data, these outliers typically consisting of exceptionally large returns for small sized firms. Furthermore, eliminating these few outlying firm returns produces a positive relationship between average returns and firm size. In this case interpreting the robust estimate is straightforward. The "typical" relationship between average returns and firm size is positive. Although, the existence of a very small fraction of exceptionally large (outlier) returns results in a "size effect" that is misleading for the majority of the data.

We conjecture that the joint distribution of equity returns and index returns is well modeled by a mixture of distributions with a dominant normal central component and with an outlier-generating component. It is the latter component that sometimes gives rise to severe bias of the OLS estimate (see discussion in Section 2). Motivated by the nature of the empirical evidence along with the successful use of robust estimates in the references above, we propose using a new robust estimate of beta. The new estimator is well suited to coping with the kind of bivariate outliers encountered when estimating beta, and in addition produces robust standard errors, t-statistic tests and a test for outlier-induced bias in the OLS estimate of beta.

A comparative study of more than 5,000 firms from the Center for Research on Stock Prices (CRSP) database reveals that the difference between the OLS and robust estimates of beta sometimes differ by more than one in magnitude, and frequently by more than one-half. Differences of this magnitude are likely to be financially significant in many contexts. Furthermore, robust tests for bias reveal that a large fraction of the differences between the OLS and robust estimates are highly significant in the statistical sense. Examination of the relationship between a firm's market capitalization and the difference between the OLS and robust beta for the firm reveals a very clear small-sized firm effect: Highly influential outliers are associated primarily with small-sized firms, a behavior which is consistent with the Knez and Ready (1987) study. Because of the early historical importance of the LAD estimator, we also compare the performance of LAD with the OLS and robust estimators, and argue why the new robust estimator is preferred to LAD. Finally, we show that the robust estimate is preferred to the OLS estimate as a predictor of future returns using the unconditional single index model. A number of our results are presented in the form of Trellis graphics displays (Becker and Cleveland, 1996), which we believe provide the reader with a more immediate visual grasp of what the results are saying than do conventional tabular displays of numbers.

The remainder of the paper is organized as follows. Section 2 provides several motivating examples of robust beta estimates, introduces a two-component mixture model for the joint distribution of equity returns and market index returns, and discusses the interpretation and use of robust estimates of beta. Section 3 describes our proposed robust beta estimate, including the computation of robust t-statistics. Section 4 presents the results of a simple Monte Carlo study that lends credibility to the statistical efficacy of the robust beta estimate. Section 5 compares the behavior of the robust and OLS beta estimates using two large data sets of returns from the CRSP database. Section 6 compares the well-known LAD estimate with our proposed robust estimate of beta. Section 7 reports the relative forecast performances of the two estimates. Concluding comments are provided in Section 8, and supporting technical details are provided in the Appendix.

2. EXAMPLES, INTERPRETATION AND MIXTURE MODELS

2.1 A Few Striking Examples

Table 1 displays summary statistics of monthly equity returns for four of the 6077 firms included in our study of monthly returns in Section 4. The four firms were chosen because their return series contain a few very large outliers. The table exhibits the stylized facts of skewness and kurtosis associated with outlier-generating distributions for equity returns. The outliers have a very substantial influence on the classical mean and standard deviation estimates, as is reflected in the differences between the non-robust estimates and their robust counterparts, the median and the median absolute deviation about the median (MADM), respectively³.

Table I

Summary Statistics for Equity Returns for Four Firms

The return series occur during the 1962 to 1996 period from the NYSE, AMEX, NASDAQ universe found in the monthly CRSP database and are in excess of the one-month nominal rate for the Treasury bill closest to 30 days to maturity also found in the CRSP files. MADM is the median absolute deviation about the median and represents a robust measure of scale.

	A W Computer Systems Inc.	Oil City Petroleum Inc.	Biopharmaceutics Inc.	Metallurgical Industries Inc.
N	184	129	130	258
Mean	0.068	0.017	0.058	0.093
Median	-0.037	-0.002	-0.055	-0.002
Std. Dev.	0.644	0.509	0.645	1.224
MADM	0.179	0.156	0.185	0.149
Skewness	7.174	8.653	6.268	14.58
Kurtosis	60.616	88.176	44.202	224.382

Figure 1 displays the scatter plots of monthly equity excess returns versus excess returns for a market proxy for the four firms of Table 1, with the OLS beta estimate and the proposed robust beta estimates represented by the dashed and solid straight line fits respectively. The outliers have considerable influence on the OLS beta causing it to be a relatively poor fit to the bulk of the data. The robust beta is not much influenced by the outliers and provides a good fit to the bulk of the data, which is an important feature of a good robust estimate. The market proxy used in this figure, and throughout the paper, is the composite NYSE, AMEX, and NASDAQ, with returns being in excess of the one-month T-bill rate from the CRSP database. The robust estimate is described in Section 3.

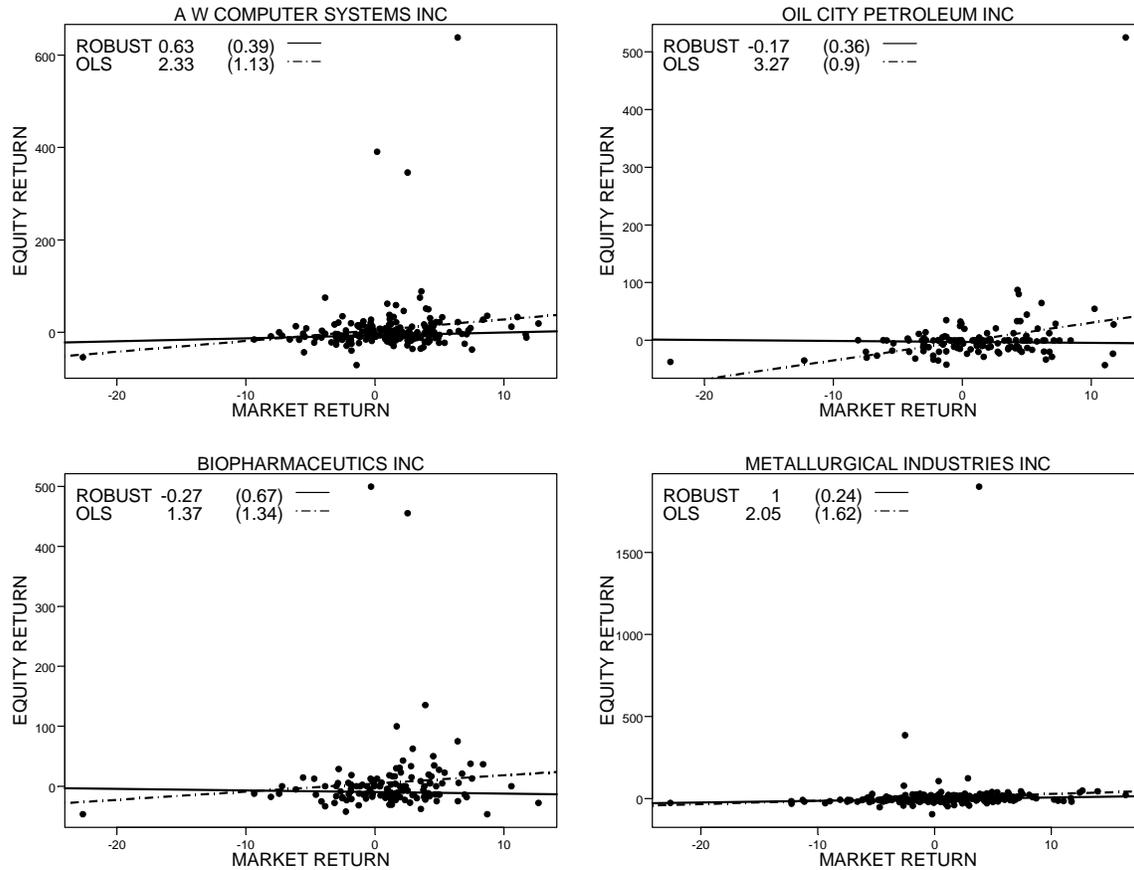


Figure 1. OLS and Robust Beta Estimates for Four Firms provide examples of the bias in the OLS market model regression estimates due to outliers. Excess returns are plotted against the excess returns of the value weighted NYSE, AMEX, NASDAQ market portfolio. The data are monthly excess returns for both the assets and the value-weighted index between January 1962 and December 1996. The dotted line represents the fit from OLS and the solid line represents the robust fit.

The OLS and robust beta estimates of Figure 1 are assembled in Panel A of Table 2. In each of the four examples the OLS beta is greater than one, substantially so in three out of the four cases, and the values are all considerably larger than the robust beta by almost any standard of comparison. This positive *bias* in the OLS estimate is caused by a general tendency for large positive outliers to occur when the market returns are more or less positive.

In all four examples the crash of October 1987 is a leverage point, i.e., an outlier in the excess market returns⁴. One is naturally concerned that this leverage point outlier might substantially contribute to increasing the value of the OLS beta over that of the robust beta. Although, as shown in Panel B of Table 2, the OLS estimate is relatively unaffected by this outlier. It is the other outliers that are causing the OLS beta to be substantially larger than the robust beta.

We emphasize that the behavior exhibited by the above four examples typifies a general tendency observed in other cases where there is a fairly large difference between the OLS

and robust beta. Namely, the outliers often fall in configurations that cause the OLS beta to be biased toward larger values than those of the robust beta. This tendency is particularly strong for the monthly returns studied in Section 4, but is also evident in our study of weekly returns.

Table II

OLS and Robust Beta Estimates and Standard Errors

Firm returns in excess of the one-month T-bill are regressed on the excess returns of the value weighted NYSE, AMEX, NASDAQ market portfolio. The data are monthly excess returns for both the assets and the value-weighted index between January 1962 and December 1996.

Panel A: Beta Estimates, (std. errors)				
	A W Computer Systems Inc.	Oil City Petroleum Inc.	Biopharmaceutics Inc.	Metallurgical Industries Inc.
OLS	2.33 (1.13)	3.27 (0.90)	1.37 (1.34)	2.05 (1.62)
ROBUST	0.63 (0.39)	-0.17 (0.36)	-0.27 (0.67)	1.00 (0.24)
Panel B: Beta Estimates with Oct. 1987 Deleted, (std. errors)				
	A W Computer Systems Inc.	Oil City Petroleum Inc.	Biopharmaceutics Inc.	Metallurgical Industries Inc.
OLS	2.27 (1.24)	3.67 (1.01)	1.10 (1.55)	2.11 (1.71)
ROBUST	0.86 (0.38)	-0.16 (0.36)	-0.28 (0.67)	1.01 (0.27)

It is important to be aware that outliers can substantially influence not only the OLS estimates of the linear model parameters, but also the estimated standard errors of these parameter estimates.⁵ There are two distinct kinds of outlier influences on standard error estimates. First, outliers in the residuals can *increase* the standard error estimate. This is because the standard error estimate uses an error variance estimate based on the sum-of-squared residuals. Second, leverage points can *reduce* the estimated standard error estimate. This occurs because leverage points have a large positive influence on the values of the sums-of-squares and cross products of the predictor variables, and an increase in these values results in an increase in the stated precision of parameter estimation. This second effect is evident in the increase in the standard errors of the OLS estimates for all four firms when the October 1987 leverage point deleted (compare Panels A and B in Table 2).

Because the robust estimate down-weights outliers, it yields a robust estimate of standard errors which is insensitive to outliers, and is influenced by a leverage point only if the residual for the leverage point is small. The fact that the robust coefficient standard error estimates in Table 1 are much smaller than those for OLS reflect the substantial influence of outliers on the latter but not on the former.

2.2 Interpretation of Robust versus OLS Betas

The presence of outliers in the examples of Figure 1 causes significant positive bias in all the OLS betas. In three of these four cases the consequence is that the firms in question are taken to be aggressive assets in the sense that they have a high level of market risk. However, this is not the true nature of these three firms given that the bulk of the data suggests levels of risk at or below that of the market. On the other hand, the returns that give rise to the unusually large values of OLS betas are a very few large positive returns tending to occur when the market returns are more or less positive (which is in itself of some interest), with no corresponding negative returns in evidence. Thus it does not make sense to say, as suggested by the OLS beta, that the firm is more risky than the market. A more accurate characterization would be that for the most part, the firms risk and return characteristics are represented by the robust beta and in addition there are a very small fraction of unusually large returns with no unusually negative returns in evidence. Thus the firm may be worthy of further investigation as a viable asset.

In many other examples we encountered, the outliers exhibit a similar influence to those in Figure 1. Casual inspection of scatter plots of the data for firms where the robust and OLS betas differ by the greatest amount reveal that the difference is caused by outliers associated with tail skewness in the conditional distribution of returns given the market return for certain ranges of the market returns. As the results of Section 5 reveal, the direction of this skewness often makes the asset appear to be more sensitive to market movements than is typical for the firm.

This discussion underscores the fundamental point that neither an OLS nor a robust beta estimate alone suffices to characterize the risk and return characteristics of a firm. When the two estimates agree it signals the absence of influential outliers, and in such cases one can reasonably trust the classical interpretation of the OLS beta. When the two estimates differ significantly according to either a user specified threshold or a test statistic (see Section 5), it signals the presence of influential outliers. In such cases the robust beta characterizes the behavior of the bulk of the data, but additional information needs to be supplied concerning the profit and loss nature of the outliers.

2.3 Relative Difference versus Log Ratio Returns

It is often argued that it makes little difference whether one uses returns based on the relative difference in prices or the log-ratio of prices. The claim is persuasive for sufficiently small returns, based on a Taylor series expansion. However there are substantial differences in the two approaches for large returns, particularly so for very large outliers.

In the case of relative-difference returns, outliers have a very asymmetric character as is evident in Figure 1: positive returns can be arbitrarily large while negative returns are bounded below by -100% . There is concern that this may be a major contributor to the OLS positive bias behavior noted above. On the other hand, it is well known that the logarithmic transformation often symmetrizes the distribution of prices (to log-

normality), and log-ratio returns can have arbitrarily small negative values as well as arbitrarily large positive values. This raises the question of whether or not the effect of outliers is greatly reduced when using log-ratio returns.

We checked a handful of the equities in our study and found that there was still quite large positive bias in the OLS beta relative to the robust beta. Because of this, and because use of relative difference returns is quite commonplace, e.g., as in the CRSP data we use for our study, we use relative difference returns throughout.

2.4 Mixture Models for Outlier Generation

The documented evidence on the viability of mixture distribution models for equity returns, combined with exploratory data analysis of a large number of scatter plots such as those in Figure 1, suggest that a *bivariate* mixture distribution is a good model for the data. For purposes of stating this model, it is convenient to express (1.1) in the form

$$y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

with two-dimensional independent variable vectors x_i , univariate responses y_i , and β representing the slope and intercept in (1.1). A viable mixture model probability density for the data is then:

$$f(x, y) = (1 - \gamma) f_o(x, y) + \gamma h(x, y) \quad (2.2)$$

where h is any bivariate density, and the $f_o(x, y)$ admits to the factored form

$$f_o(x, y) = f_\varepsilon(y - x^T \beta) g(x). \quad (2.3)$$

It is assumed that the error density f_ε is Gaussian but that the density g for the predictor variables may be heavy-tailed to account for leverage points such as the market return for October 1987 in Figure 1. The mixing probability γ is typically fairly small, e.g., in the .01 to .05 range.

Thus the joint returns are generated by the density (2.3) most of the time, in which case the conditional distribution of the equity return given the market return is normal, and corresponds to the model (1.1) – (2.1). For a small fraction of the time the joint returns are generated by the density h . It is to be emphasized that neither nature nor financial theory constrains h to correspond to the classical model (1.1) – (2.1) and in fact rational arguments can be given for the opposite. For example, it is often the case that unusually large returns are associated with some event that causes an atypical price movement. In such cases there is no reason to expect that a linear model which describes the bulk of the data generates the corresponding return. A dummy variable could be used if one knew when some potential cause occurs (e.g., a quarterly earnings report). But this approach may be impractical, and one may not know when a causative event occurs. Hence, representation of such potential movements by the density h and an appropriate probability γ may be quite reasonable.

The general character of a density h which generates returns such as those in Figure 1 is that the density has mass concentrated x and y values for which the market returns are more or less positive and the equity returns tend to be unusually large. Such an outlier-generating density for mixture model (2.2) can give rise to severe *bias* of the OLS estimate. It is important to emphasize that the size of the data sets typically used to estimate beta is seldom sufficient to reliably estimate the mixture density h . There are simply far too few observations from this outlier-generating component of the mixture model. This situation is ideally suited to the use of the robust regression estimate introduced in the next section.

3. THE ROBUST REGRESSION METHOD

The robust regression method we have chosen to use for computing estimates of beta is a regression M-estimate computed in a particular way and using a distinct bounded loss function. We describe the estimation procedure in this section and in the appendices. For a complete description of the overall computational methodology, see Yohai, Stahel, and Zamar (1991).

3.1 Regression M-Estimates

Consider the linear model (1.1) using the notation of (2.1). An M-estimate $\hat{\beta}$ is defined as

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \rho \left(\frac{y_i - x_i^T \beta}{\hat{s}} \right) \quad (3.1)$$

where \hat{s} is a robust scale estimate for the residuals and ρ is a symmetric *loss* function. Dividing the residual by the scale estimate \hat{s} makes the estimator $\hat{\beta}$ invariant with respect to the scale of the error ε_i . The M-estimate class of regression estimates includes the ordinary least squares (OLS) and the least absolute deviation (LAD) estimates as special cases, through the choices $\rho_{OLS}(r) = r^2$ and $\rho_{LAD}(r) = |r|$, respectively.

This class of estimates was first introduced by Huber (1973), as a generalization of his M-estimates of location (Huber, 1964). The corresponding estimating equation, obtained by differentiating (3.1) with respect to β and setting $\psi = \rho'$, is

$$\sum_{i=1}^n x_i \psi \left(\frac{y_i - x_i^T \hat{\beta}}{\hat{s}} \right) = 0. \quad (3.2)$$

Huber's favorite choice of ρ in (3.1) was the *unbounded* convex function $\rho = \rho_H$ that is quadratic in a central interval $(-c, c)$ and linear outside that interval. The corresponding psi-function $\psi = \psi_H$ is linear with unity slope on the central interval $(-c, c)$ and has constant value $\psi = c \cdot \text{sgn}(r)$ outside that interval.

When an arbitrary contamination density h does not conform to the model (2.1), an M-estimate will be biased⁶. Since the empirical evidence indicates that the OLS estimate of beta suffers from such bias, a desirable estimation goal is to control bias for mixture models of the form (2.2) – (2.3) for arbitrary unknown h . At the same time, it is desirable to have a reasonably high efficiency when the errors in the model (2.1) are normal, e.g., in the vicinity of 90% efficiency.

In the special case when the M-estimate is the OLS estimate, the bias is obviously unbounded over all possible choices for h . What may be less obvious is that the bias is also unbounded for any M-estimate based on an *unbounded* ρ , such as ρ_H or ρ_{LAD} ⁷. See Martin, Yohai and Zamar (1989), who show that one can obtain bias-robust regression M-estimate by using a *bounded* ρ .⁸

3.2 An Optimal Bias-Robust Regression M-Estimate

We use a bounded function ρ contained in a family having the piece-wise polynomial shapes, three of which are shown in the left panel of Figure 2. The “cutoff” values $c = .868, .944, \text{ and } 1.060$ correspond to Gaussian model efficiencies of $.85, .90 \text{ and } .95$, i.e., when the joint distribution of x and y is generated according to the linear regression model (2.1) with a normal error distribution. The corresponding ψ functions are shown in the right-hand panel. The cutoff values of $-c$ and $+c$ are the values at which the ψ function becomes zero, and correspondingly the ρ function becomes constant. As c increases the Gaussian model efficiency increases, and in the limit as c tends to infinity the estimator becomes the fully efficient least squares estimator. The formulas for the ρ and ψ functions are given in Appendix A.1.

Our choice of these ρ and ψ functions is based on the recent results of Yohai and Zamar (1997), moreover the psi-function is a close least-squares polynomial approximation to the optimal function found in Yohai and Zamar (1987). These functions yield *efficiency-constrained bias-robust* estimates: that is ρ and ψ minimize the maximum asymptotic bias over arbitrary h in the mixture model (2.2) – (2.3), subject to a user-specified efficiency at the Gaussian model. This result is rather remarkable in combining the following features:

- (a) It is the first theoretical justification for a non-monotonic ψ -function based on the concept of controlling bias while achieving a desired Gaussian model efficiency⁹
- (b) It is the first theoretical justification for a ψ -function that redescends to zero at all argument values outside a finite interval¹⁰
- (c) The resulting shape is intuitively satisfying.

Elaborating on this last claim, first note that the ψ -functions of Figure 2 have the following intuitive interpretation. They are smooth approximations to the so-called *hard-*

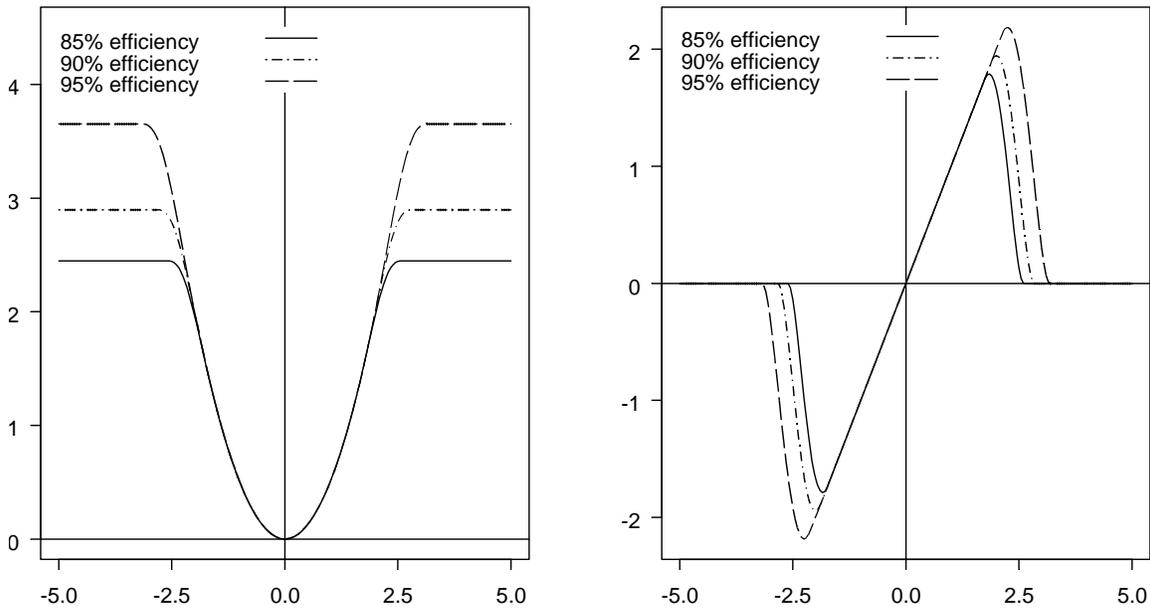


Figure 2. The Optimal ρ and ψ Functions for Robust Estimation of Beta. Both functions are plotted for three cutoff values; $c = 0.868, 0.944,$ and 1.060 corresponding to Gaussian efficiencies of 0.85, 0.90, and 0.95.

rejection ψ -function, which is linear with slope unity on the interval $(-c, c)$ and has value zero outside that interval. The smooth transition of ψ from extreme values to zero is intuitively appealing and preferable to the discontinuous behavior of the hard-rejection ψ -function. The effect of the latter is to regard all sufficiently small observations to be “completely” good, and all others to be “completely” bad, which is patently unnatural. On the other hand, the region of smooth transition of ψ in Figure 2 occurs where most natural, namely in the “flanks” of the distribution where it is most difficult to decide whether an observation is “good” or “bad”.

Another advantage of the smooth ψ in Figure 2 over using a discontinuous hard-rejection ψ -function is that it results in better behavior of the variance expression for the estimated regression coefficients and in estimation of this expression from the data.¹¹

It should be pointed out that, whereas robust regression based on Huber’s favorite convex function ρ_H involves non-linear optimization with an essentially unique solution, robust regression based on the bounded non-convex loss ρ of Figure 2 involves non-linear optimization with multiple local minima. Fortunately, a reliable computational strategy is available as described in Appendix A.2.

For the remainder of this paper we use the value $c = 0.868$ in the rho and psi-functions, which this yields an 85% variance efficiency for normal errors. This corresponds to 92%

efficiency on a standard deviation basis, i.e., the standard deviations of the robust parameter estimates is about 8% greater than that of OLS when we use $c = .868$. This is in return for better bias control than when using the more common variance efficiency choice of 90%

3.3 Approximate Standard Errors

The lack of inference for robust regression parameter estimates has undoubtedly hindered the widespread use of such estimators. A feature of the robust regression coefficient estimates we use is that they are asymptotically multivariate normal, with an asymptotic covariance matrix which has a relatively simple form that admits to robust estimation from the data. The expression for this covariance matrix is

$$C_{\hat{\beta}} = \frac{1}{n} (X^T X)^{-1} \cdot v \quad (3.3)$$

where X is the n -by- p matrix of independent variables, and the scalar quantity v is given by

$$v = \frac{s^2 \cdot E\psi^2(\varepsilon/s)}{E^2\psi'(\varepsilon/s)}. \quad (3.4)$$

Here ε is the error term in (2.1) and s is the asymptotic value of the robust scale estimate. See for example Huber (1981) and Yohai (1988).

We estimate $C_{\hat{\beta}}$ by replacing v in (3.4) with a natural estimate \hat{v} and X with a weighted version \tilde{X} yielding

$$C_{\hat{\beta}} = \frac{1}{n} (\tilde{X}^T \tilde{X})^{-1} \cdot \hat{v}. \quad (3.5)$$

See appendix A.3 for details concerning computation of \hat{v} and \tilde{X} .

The estimate (3.5) provides a consistent, non-parametric, robustly estimated covariance matrix. The square roots of the diagonal elements of this matrix are used as standard errors of the coefficient estimates.

4. A SMALL MONTE CARLO COMPARISON

In order to compare the behavior of the distributions of the OLS and proposed robust estimator of beta, we performed Monte Carlo simulations for the bivariate model

$$y_t = \alpha + \beta x_t + \varepsilon_t, \quad i = 1, \dots, T \quad (4.1)$$

where the true α and β are equal to zero. We compare the distributions of the OLS and the robust parameter estimates from a simulation using 1000 Monte Carlo replicates of 100 observations of (x_i, ε_i) for each of two situations. In the first, we use for every replicate a fixed set of 100 x_i 's drawn once from a $N(0,1)$ distribution, and 1000 sets of 100 independent ε_i 's drawn from a $N(0,1)$ for each replicate. In the second, we use the same 1000 replicate samples of length 100 of x_i and ε_i , but with probability $\gamma = .01$ we replace (x_i, ε_i) with an independent $x_i \sim N(2, 0.25)$ and $\varepsilon_i \sim N(10, .0.25)$ pair. The results of the Monte Carlo simulation are presented in Table 3 and in Figure 5.

Table III

Monte Carlo Comparison of the OLS and Robust Estimates

In the first simulation, Panel A, the parameters of the market model are estimated using 1000 samples of 100 observations of draws of (x, ε) from independent standard normal distributions. The y are constructed by $y = \beta x + \varepsilon$ where $\beta = 0$. In the second simulation, Panel B, the same data is used with the exception that there is a 1% chance of an (x, ε) outlier. The outlier is formed by drawing x from a $N(2, 0.25)$ and ε from a $N(10, 0.25)$. The summary statistics and empirical percentiles are for the distributions of the 1000 values of the estimated β .

	Panel A: X and Y ~ N(0,1)				Panel B: X and Y ~ .99N(0,1)+.01N(10,1)			
	Summary Statistics:				Summary Statistics:			
	Mean	Std. Error	Skewness		Mean	Std. Error	Skewness	
OLS	-0.004	0.084	0.067		0.058	0.117	-0.127	
ROB	-0.002	0.115	-0.102		-0.001	0.105	-0.140	
	Monte Carlo Percentiles:				Monte Carlo Percentiles:			
	2.5%	5%	95%	97.5%	2.5%	5%	95%	97.5%
OLS	-0.231	-0.189	0.177	0.215	-0.231	-0.190	0.178	0.227
ROB	-0.159	-0.138	0.137	0.159	-0.138	-0.102	0.236	0.268

Conditional on the values of the fixed set of independent variables x_i , the distribution of $\hat{\beta}$ is $N\left(0, (X^T X)^{-1}_{2,2}\right) = N(0, 0.0072)$ when the errors ε_i are $N(0,1)$. This normal density is overlaid as a reference in all four panels in Figure 3. The solid vertical lines are located at the 2.5 and 97.5 percentiles of this reference distribution.

The Monte Carlo histogram of the OLS estimate for normal errors is quite close to the theoretical normal distribution. The Monte Carlo histogram of the robust estimate for normal errors also behaves much as expected, being reasonably normal in shape, well centered on zero, and with a standard deviation 0.11 which is somewhat larger than the OLS standard deviation .084. The increased standard error represents the price paid, in terms of lower efficiency, when the errors are normal in return for bias robustness toward outliers.¹²

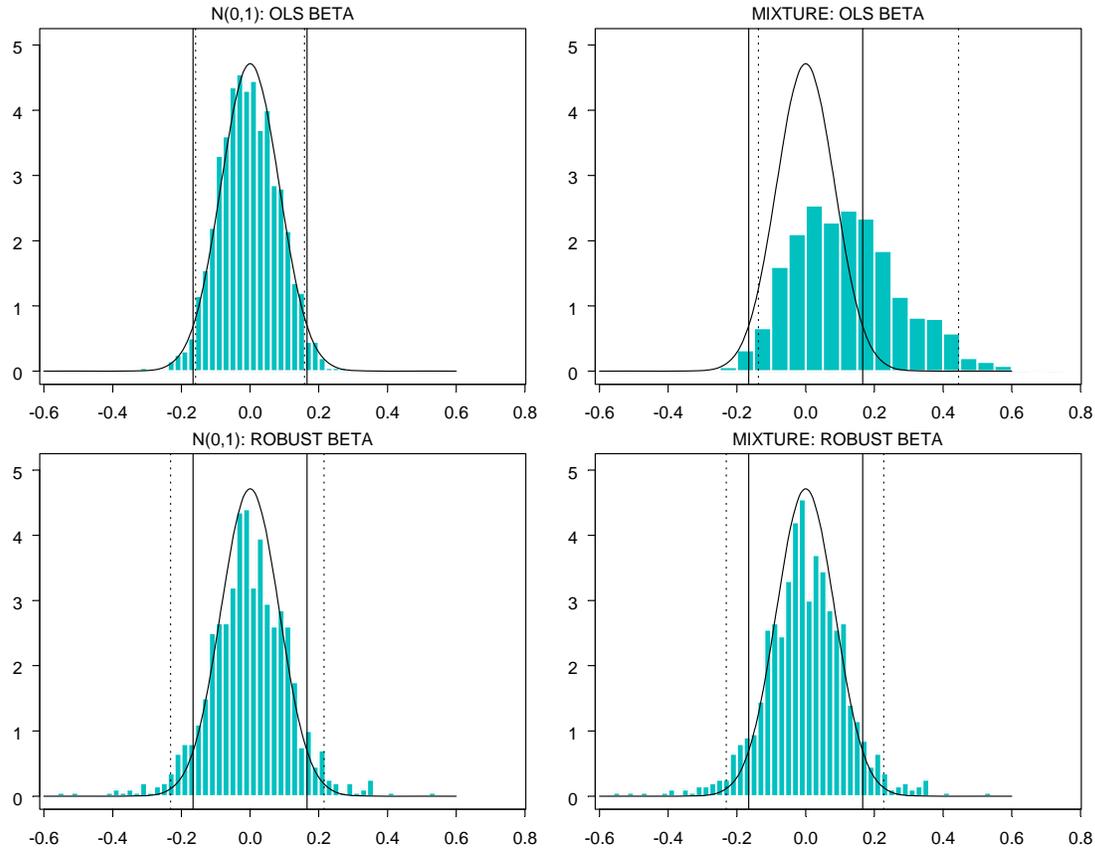


Figure 3. Monte Carlo Performance of OLS and Robust Beta Estimates. The left two panels display histograms of the distribution of the OLS and the robust estimated beta coefficient when both the independent variable and the error term are generated from independent standard normal distributions. The right two panels display the distribution of the beta estimates when both the same data is used with the exception that there is a 1% chance of an (x, ε) outlier. The outlier is formed by drawing x from a normal distribution with mean 2 and standard deviation .25 and ε from a normal distribution with mean 10 and standard deviation .25. The dashed vertical lines indicate the empirical 2.5 and 97.5 percentiles. The overlaid density is the “true” distribution, a $N\left(0, \left(X^T X\right)_{2,2}^{-1}\right) = N(0, 0.0072)$. The solid vertical lines are the 2.5 and 97.5 percentiles of the “true” distribution.

When a bias producing normal mixture distribution is used, the distribution of the robust estimate is virtually identical to that obtained when the linear model errors are normal.

The distribution of OLS estimate under the bias producing normal mixture distribution is radically shifted in location and shape from that for a linear model with normally distributed errors. The OLS distribution is still unimodal, but exhibits a substantial increase in mean from -0.004 when the data does not contain any outliers to 0.058 with bias producing outliers. The OLS distribution also exhibits a substantial increase in standard error from 0.084 to 0.117. The shift of the histogram to the right is due to the high probability (0.67) of having outliers in a sample of size 100 when the mixing probability is $\gamma = .01$. Note also that the confidence interval associated with the robust estimate of beta is quite robust in terms of both the stability of interval length and the

centering across normal and non-normal distribution situations. The confidence interval associated with OLS is very misleading when the distribution generates large bias producing outliers.

5. THE CRSP DATA SET COMPARISONS

5.1 The Data

We examine two data sets including monthly and weekly returns for firms listed on the NYSE, AMEX, and NASDAQ exchanges from the CRSP database. The first data set is comprised of all monthly returns available for firms having at least 10-years data within the period 1962 to 1996. The second data set contains all weekly returns available for firms having five years of data for the period 1992 to 1996. The weekly returns represent aggregated daily returns. The three exchanges list 20,272 equities over the longer time period, with 6,077 having at least 10 years of data at some time during the 1962 to 1996 period. The weekly 5-year data set contains 4,440 firms having complete weekly series from 1992 to 1996. The equity returns $r_{i,t}$ are holding period returns found in the time series record file from CRSP. The market proxy returns $r_{m,t}$ are the value weighted returns including all distributions excluding ADR's for the composite of firms listed on the NYSE AMEX, and NASDAQ. Excess returns are formed for the monthly data using a risk-free rate (r_f) which is the one-month nominal average of the bid and ask for the Treasury bill closest to 90 days to maturity. As is common practice when calculating beta, raw returns are used for the weekly data.

5.2 The Distribution of the Estimates and their Differences

One of our primary interests is to characterize how often the robust estimator produces a beta that differs substantially from the corresponding OLS beta. To this end, we present the results in Table 4 that displays binned counts of the absolute and raw differences between the OLS and corresponding robust beta. The count for each bin is the number of firms for which the difference between the OLS and robust beta estimates fall in that cell. Most of the betas for the robust estimator are within ± 0.25 of the corresponding OLS beta. In particular, 64% of the monthly robust betas and 50% of the weekly robust betas are within this range. But it is noteworthy that between 14% of the monthly and 23% weekly robust betas differ from the corresponding OLS betas by more than 0.50. And a fractionally small, but still nontrivial number of the differences exceed 1.0 in magnitude - 130 in the case of the ten year data and 261 in the case of the five year data (about 2% and 5%, respectively).

Table IV

Binned Counts of Differences Between OLS and Robust Beta's

The robust betas for each data set are sorted by the absolute value (Panel A) of their difference and the difference (Panel B) from the matching OLS betas. The counts in each table cell are the number of betas that fall in a given Δ bin.

Panel A: $\Delta = \beta_{OLS} - \beta_{ROB} $			Panel B: $\Delta = \beta_{OLS} - \beta_{ROB}$		
Δ	10-years Monthly	5-years Weekly	Δ	10-years Monthly	5-years Weekly
0.00+ to 0.25	3879	2662	< -1.00	7	27
0.25+ to 0.50	1325	1124	-1.00+ to -0.50	58	154
0.50+ to 0.75	552	417	-0.50+ to -0.25	164	353
0.75+ to 1.00	191	130	-0.25+ to 0.00	1229	992
1.00+ to 1.25	84	67	0.00+ to 0.25	2650	1670
1.25+ to 1.50	25	21	0.25+ to 0.50	1161	771
1.50+ to 1.75	10	11	0.50+ to 1.00	685	393
1.75+ to 2.00	5	2	1.00+ to 2.00	117	75
> 2.00+	6	6	> 2.00+	6	5
N =	6077	4440	N =	6077	4440

Figure 4 displays three different views of the distributions of the difference between robust and OLS estimates. The first row displays histograms corresponding to the bins of Table 4, the second row contains kernel density estimates, and the third normal qq-plots of the differences.

The tendency for the robust estimator to produce betas that are greater than the OLS estimate is evident in the plots of Figure 4. For the ten-year data set, 76% of the differences between OLS and ROB are positive. For the 5-year weekly data set, 66% of the beta differences are positive. The tendency for the weekly data to produce more negative differences than the monthly data is quite apparent in the counts of Table 4 and in the plots of Figure 4. The qq-plots show that the distribution of the differences across all firms is itself a heavy-tailed deviation from normality.

Overall, the foregoing provides a substantial indication that when there are significant outliers in the data, the OLS beta estimate will tend to be biased upward and hence overestimate market risk and return.

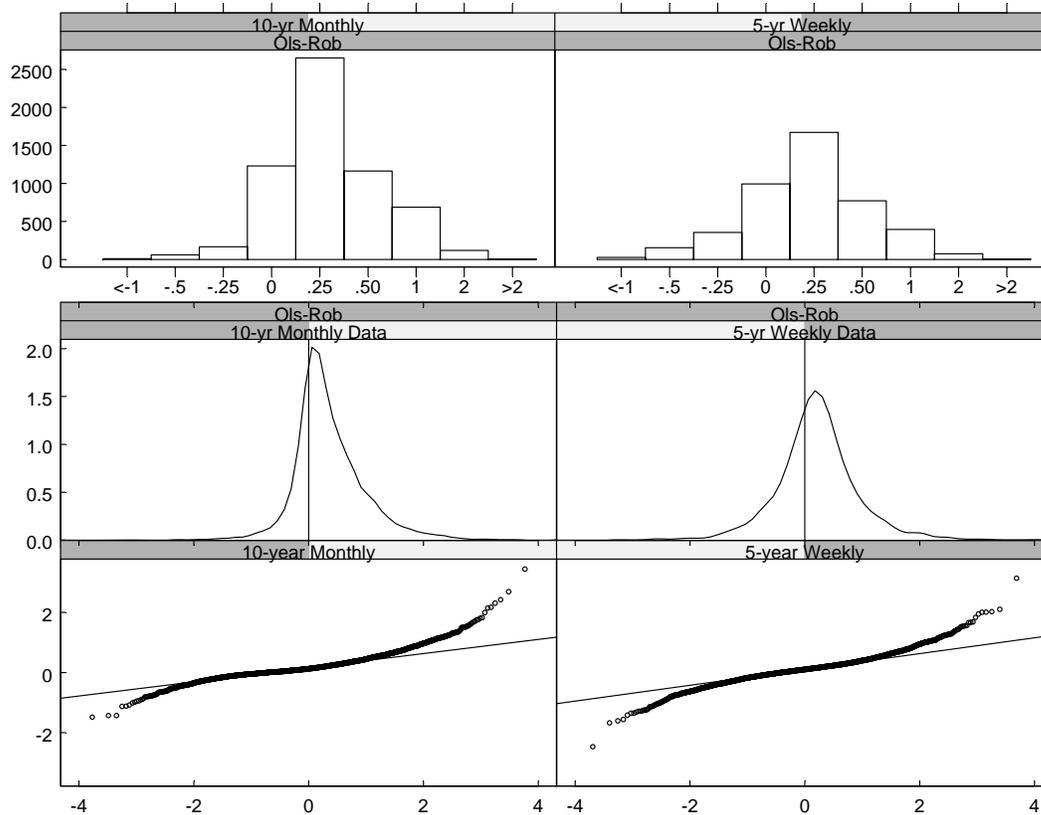


Figure 4. Distribution of Differences between OLS and Robust Beta's. In the upper two panels, the robust betas for each data set are sorted by their difference from the matching OLS betas. The counts in each bar are the number of betas that fall in a given Δ bin. The center panels are kernel estimates of the probability densities of the difference between OLS and ROB for the 10-year monthly and 5-year weekly data sets. The bottom panels of the figure contain normal quantile-quantile plots of the difference between the two estimates. The bottom four panels do not plot several large points used in the density and quantile estimates so that the scale of the plot is useful.

5.3 Conditional Distributions of the Estimates

We also examine the individual conditional distributions of the OLS and robust estimates of beta, conditioned on the difference between the two estimates. Figure 5 contains Trellis displays of conditional box plots of the OLS and robust beta estimates. The vertical axis labels indicate which difference interval the estimate is associated with. The two sets of strip labels for each panel in the Trellis display indicate which estimate and which data set the box plots are based on. For example, the strip labels indicate that the bottom two panels are for the five-year weekly data and the top two panels are for the ten-year monthly data. They also indicate that the first and third panels are for the OLS beta while the second and fourth are for the robust beta. For general information on Trellis displays, see Becker and Cleveland (1998), and for applications to analyzing equity returns, see Martin (1998).

The Trellis display in Figure 5 immediately reveals the following behavior. For firms where the difference between robust and OLS betas is small, the distributions of the estimates look quite similar. This is consistent with the relative behavior of the OLS and

Distribution of Betas: 10-year Monthly and 5-year Weekly Data

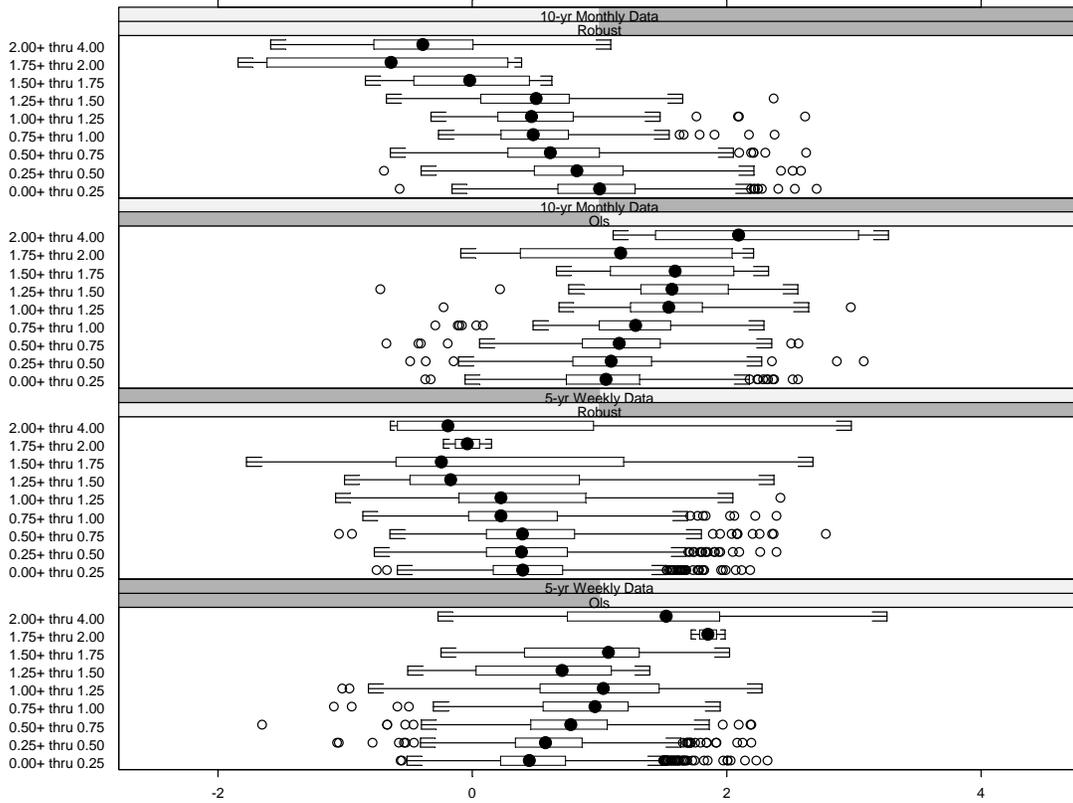


Figure 5. Conditional Distributions of OLS and Robust Beta's. The box plots of the estimates produced by OLS and ROB for the 10-year monthly and 5-year weekly data conditional on being in a given absolute value bin. Betas are sorted based on the absolute value of the difference between OLS and ROB and the distribution is then plotted. The solid dot represents the median of the distribution. The box indicates the inter-quartile range (IQR: the middle half of the data), the whiskers cover the range $1.5 \times$ IQR and points beyond the whiskers may be considered outliers. One robust estimate is not plotted to make the scale of the plot useful.

the robust estimator for Gaussian data where the robust estimator has high efficiency and is therefore highly correlated with the OLS estimate. For firms where the difference between the OLS and robust betas is large, the OLS beta distributions are shifted toward larger values and the robust beta distributions are shifted toward small values. Notice how this effect increases with increasing conditioning interval as is evidenced in the “v” configurations of the box plots which is clearly revealed in the Trellis display.

Another informative view using Trellis displays is obtained by examining box plots of the difference between the robust and OLS estimates of a firm’s beta, conditioned on the magnitude of the OLS beta. Figure 6 provides such a display, with one strip label indicating the values of the conditioning variable (the absolute value of the OLS beta), which increase going from the bottom panel to the top panel. The other strip label indicates ten year monthly versus five-year weekly data (left column and right column of panels, respectively).

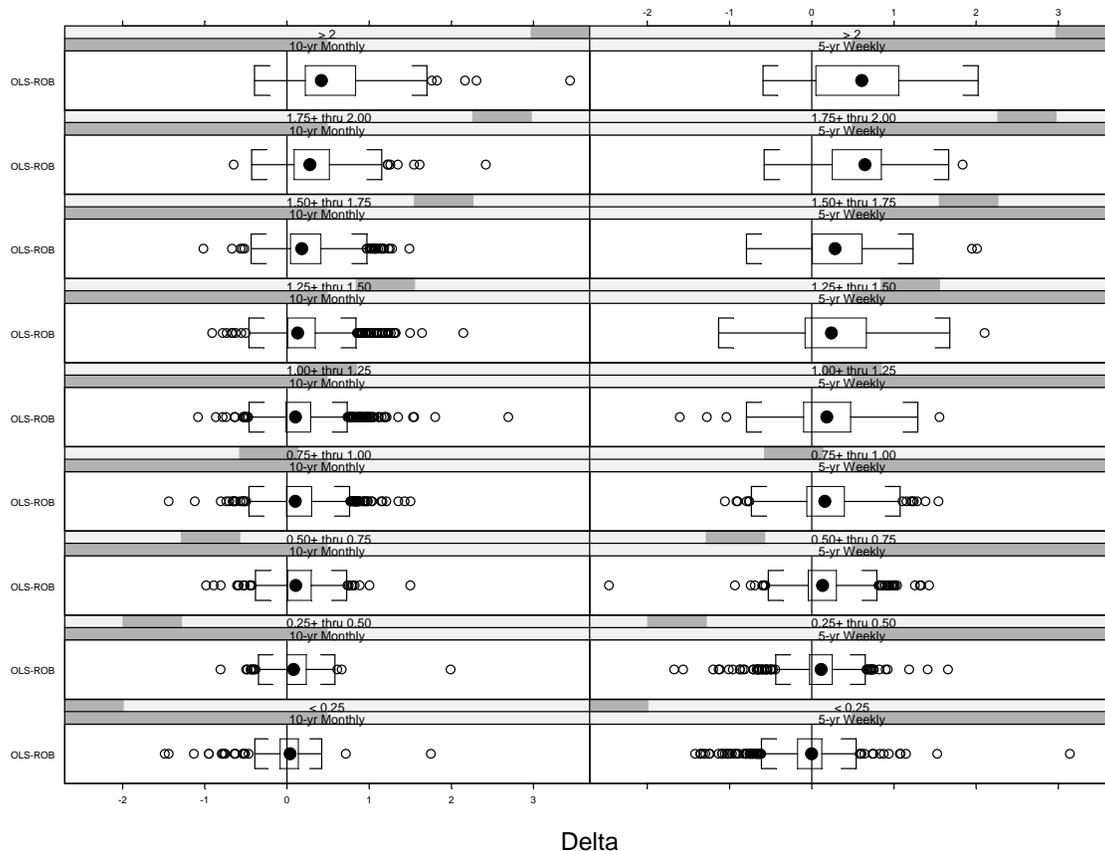


Figure 6. OLS and Robust Beta Differences Conditioned on the OLS. Box plots of the difference between the estimates produced by OLS and ROB for the 10-year monthly and 5-year weekly data conditional on the OLS beta being in a given bin. The differences between OLS and robust betas are sorted based on the size of the OLS beta and the distribution of the differences is then plotted. The solid dot represents the median of the distribution. The box indicates the inter-quartile range (IQR: the middle half of the data), the whiskers cover the range $1.5 \times \text{IQR}$ and points beyond the whiskers may be considered outliers.

For the ten-year monthly data, when the OLS beta is small, the difference between OLS and robust is also small with little variation. For larger OLS betas, the distribution of the differences shifts to the right and spreads out. Similar but more pronounced behavior is evident in the 5-year weekly data. Under the reasonable assumption that the robust beta accurately represents the bulk of the data, these results suggest that the larger OLS betas may be biased by a small fraction of the data.

5.4 Testing for Bias in the OLS Beta

While the differences between the OLS and robust beta's seem substantial indeed, it is of interest to determine to what extent they are statistically significant. To this end we use a modification of a test for bias proposed by Yohai, Stahel and Zamar (1991). The test statistic we use is based on a studentized version of the difference between a robust scale estimate \hat{s} for the residuals and the usual sample standard deviation estimate $\hat{\sigma}$. Namely

$$T = \frac{2n \cdot (\hat{\sigma} - \hat{s})}{a \cdot d^2 \cdot \hat{s}^2} \quad (5.1)$$

$$a = \frac{(1/n) \sum \psi'(\tilde{r}_i)}{(\hat{s}/n) \sum \psi(\tilde{r}_i) \cdot \tilde{r}_i} \quad (5.2)$$

$$d^2 = (1/n) \sum (r_i - \psi_i)^2. \quad (5.3)$$

where $\tilde{r}_i = (y_i - x_i^T \cdot \hat{\beta}) / \hat{s}$, $\hat{\beta}$ is a robust estimate of β , and

$$\psi_i = \frac{\psi(\tilde{r}_i)}{(1/n) \sum \psi'(\tilde{r}_i)}, \quad i = 1, \dots, n \quad (5.4)$$

Here ψ is the psi-function type shown in Figure 2. It can be shown that T is asymptotically equivalent to the natural quadratic form for assessing the size of the difference $\hat{\beta} - \hat{\beta}_{OLS}$, using the estimated inverse covariance matrix of the difference in estimates. Correspondingly, T has an asymptotic χ^2 distribution with p degrees of freedom (see Yohai, Stahel and Zamar, 1991).

Table 5 presents the number of times the p-value for modified statistic to test for OLS bias falls below a given level. Clearly, a substantial fraction of the OLS betas exhibit statistically significant bias; Close to 20% of the betas for both the weekly and monthly data exhibit statistically significant bias at the 0.01 level. The range of the absolute value of the differences between OLS and robust weekly betas that are statistically different at the 0.01 level is between 0.08 and 2.48 with a mean of 0.52. For the monthly betas with significant differences at the 0.01 level, the differences range from 0.05 to 3.45 with a mean of 0.52. While it may be difficult to ascertain what an investor believes is an economically significant difference between OLS and robust beta, the statistical significance is striking.

Table V
Results of Test for Bias of OLS Estimate

The table presents the number (and percentage of total) of p-values that are below the given significance level. The test statistics were calculated for each firm in both data sets and the p-values were then accumulated.		
Level	≤ 0.01	≤ 0.05
10-year Monthly Data	1364 (22.5%)	2749 (43.3%)
5-year Weekly Data	756 (17.0%)	2204 (49.6%)

5.5 Outliers and Firm Size

Using the difference between the OLS and robust estimate as an indicator of outliers in a given firm data set, we examine the distribution of the average market capitalization (price times shares outstanding) over the sample period conditional on the absolute difference in the point estimates. In Table VI the market capitalization for any given firm is calculated as the average over the sample period of the price per share multiplied by the number of shares outstanding. Both the price per share and the shares outstanding are the values closest to the end of the year.

Table VI

Market Capitalization (in millions \$)

The time series average over the sample period for each firm of the year-end price/share times the number of shares is used as the firm's market capitalization rate. The summary statistics are for the distribution of these means in a given bin determined by the difference between the OLS and robust beta estimates. MADM is the median absolute deviation about the median and represents a robust measure of scale.

Bin	Panel A: 10-yr Monthly Data					Panel B: 5-yr Weekly Data				
	N	Mean	Median	Std. Dev.	MADM	N	Mean	Median	Std. Dev.	MADM
0.00+ to 0.25	3796	442.266	78.167	1464.724	99.104	2662	1372.593	165.021	5199.570	216.343
0.25+ to 0.50	1307	159.088	33.136	651.360	38.699	1124	1017.008	113.333	4172.456	147.835
0.50+ to 0.75	546	113.978	24.818	700.322	27.215	417	747.853	83.178	3270.095	102.966
0.75+ to 1.00	189	44.144	14.777	113.581	13.640	130	204.462	54.058	371.990	62.600
1.00+ to 1.25	84	51.730	20.297	103.092	18.918	67	184.496	56.404	436.764	70.547
1.25+ to 1.50	25	18.265	10.701	21.884	7.308	21	89.952	18.557	175.896	19.157
1.50+ to 1.75	10	13.938	12.884	7.460	10.349	11	54.361	26.193	57.257	30.998
1.75+ to 2.00	4	30.726	24.533	23.297	17.372	2	13.003	13.003	17.019	17.842
2.00+ to 4.00	6	12.169	10.028	7.075	6.127	6	44.639	29.953	54.353	16.581

Consistent with the findings of Knez and Ready (1998), we find that the firms with the largest differences between the OLS and robust estimator (firms whose data most likely contain outliers) are on average the smallest firms in terms of market capitalization. In both panel A and panel B the average firm size and deviation around that average decreases nearly monotonically as the difference between the OLS and robust beta increases. The results do not change when the median and MADM estimates of location and scale are used indicating this result is robust toward outliers in market capitalization.

In Figure 7 the distributions of the time series averages of the natural logarithm of year-end market capitalization (price per share times shares outstanding) for each firm are plotted. The box plots provide a clear picture of the distribution of market capitalization conditional on the difference between OLS and the robust estimator. The median size of the firm decreases as the difference between OLS and the robust estimator increases.

Also, the distributions become more tightly clustered around the median as the presence of outliers becomes more prevalent.

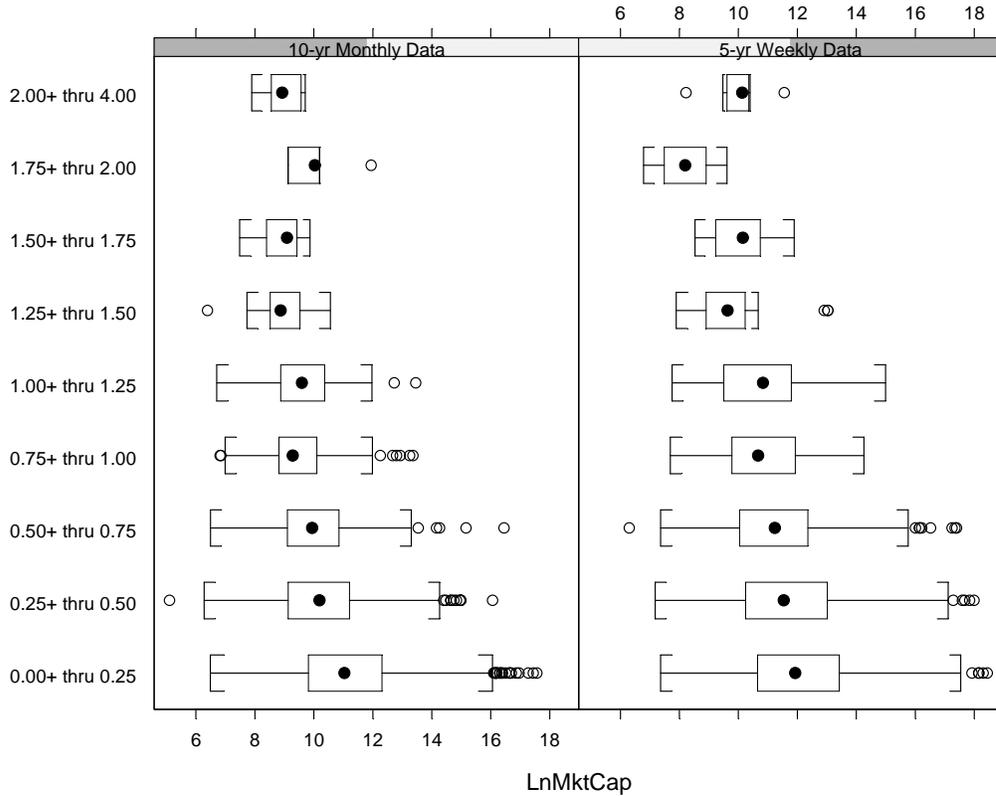


Figure 7. Conditional Distributions of Ln(Market Capitalization). The box plots represent the distributions of the time series averages of the natural logarithm of year-end market capitalization (price per share times shares outstanding). The Ln(Mkt. Cap.) are sorted based on the absolute value of the difference between OLS and ROB and the distribution is then plotted. The solid dot represents the median of the distribution. The box indicates the inter-quartile range (IQR: the middle half of the data), the whiskers cover the range 1.5* IQR and points beyond the whiskers may be considered outliers. One robust estimate is not plotted to make the scale of the plot useful.

6. A COMPARISON WITH THE LEAST ABSOLUTE DEVIATION (LAD)

The susceptibility of OLS to outliers and leverage points has been known for some time and has led to many different more outlier resistant estimation techniques. The first attempt to find a more robust regression method came in 1887 when Edgeworth proposed the least absolute deviation (LAD) regression estimator. Rather than minimizing the sum-of-squared residuals, a loss function that is very sensitive to the influence of outliers, the LAD estimate $\hat{\beta}_{LAD}$ minimizes the sum of absolute residuals:

$$\hat{\beta}_{LAD} = \arg \min_{\beta} \sum_{i=1}^n |y_i - x_i^T \beta|. \quad (6.1)$$

This estimate is also known as the minimum absolute deviation (MAD) estimate and as the L1 estimator. This is the estimator used by Sharpe (1971) and Cornell and Dietrich (1978) in two of the earliest attempts to estimate financial beta using a robust regression technique.

While the LAD estimator does a much better job than OLS in dealing with outliers in the error term of the linear model, it does not cope well with leverage points, i.e., data points with large values of the independent variables and which do not fit the linear model well. This is also a weakness of the Huber type M-estimate. Furthermore, it is known that M-estimates such as these with unbounded loss functions ρ can have arbitrarily large bias under contamination models of the type described in Section 2 (see Martin, Yohai and Zamar, 1989). None-the-less, it is worthwhile to compare the behavior of the classic LAD estimate of beta with that of the new robust estimate. Table 6 displays the binned counts of the absolute and raw differences between the OLS and LAD for the 10-year monthly and 5-year weekly data sets.

Table VII

Binned Counts of Differences Between LAD and Robust Beta's

The LAD betas for each data set are sorted by the absolute value (Panel A) of their difference and the difference (Panel B) from the matching OLS betas. The counts in each table cell are the number of betas that fall in a given Δ bin.

$\Delta = \beta_{OLS} - \beta_{LAD} $			$\Delta = \beta_{OLS} - \beta_{LAD}$		
Δ	10-years Monthly	5-years Weekly	Δ	10-years Monthly	5-years Weekly
0.00+ to 0.25	4617	3075	< -1.00	1	4
0.25+ to 0.50	1103	935	-1.00+ to -0.50	22	34
0.50+ to 0.75	257	294	-0.50+ to -0.25	132	164
0.75+ to 1.00	72	84	-0.25+ to 0.00	1441	1024
1.00+ to 1.25	19	35	0.00+ to 0.25	3176	2051
1.25+ to 1.50	3	13	0.25+ to 0.50	971	771
1.50+ to 1.75	3	3	0.50+ to 1.00	307	344
1.75+ to 2.00	1	1	1.00+ to 2.00	25	48
> 2.00+	2	0	> 2.00+	2	0
N =	6077	4440	N =	6077	4440

We note from these results that the differences between the OLS and LAD estimates have a propensity to be clustered in the smaller bins, relative to the differences between the OLS and new robust estimate. For example, comparing Tables 1 and 4 for the ten-year data set, 64% of the robust estimates are within ± 0.25 and 86% are within ± 0.50 of the OLS betas compared to 76% of the LAD estimates within ± 0.25 and 94% are within ± 0.50 of the OLS betas. This pattern is evident in the five-year weekly data set also. Here, 48% of the robust estimates are within ± 0.25 and 75% are within ± 0.50 of the OLS betas as compared to 59% of the LAD estimates being within ± 0.25 and 82% falling

within ± 0.50 of the OLS betas. For both data sets, as the bin size increases, the binned counts off more quickly for the LAD estimates than for the robust estimates.

Figure 8 presents histograms, density estimates and normal qq-plots of both the differences between OLS and LAD and the differences between OLS and ROB (the new robust estimate).

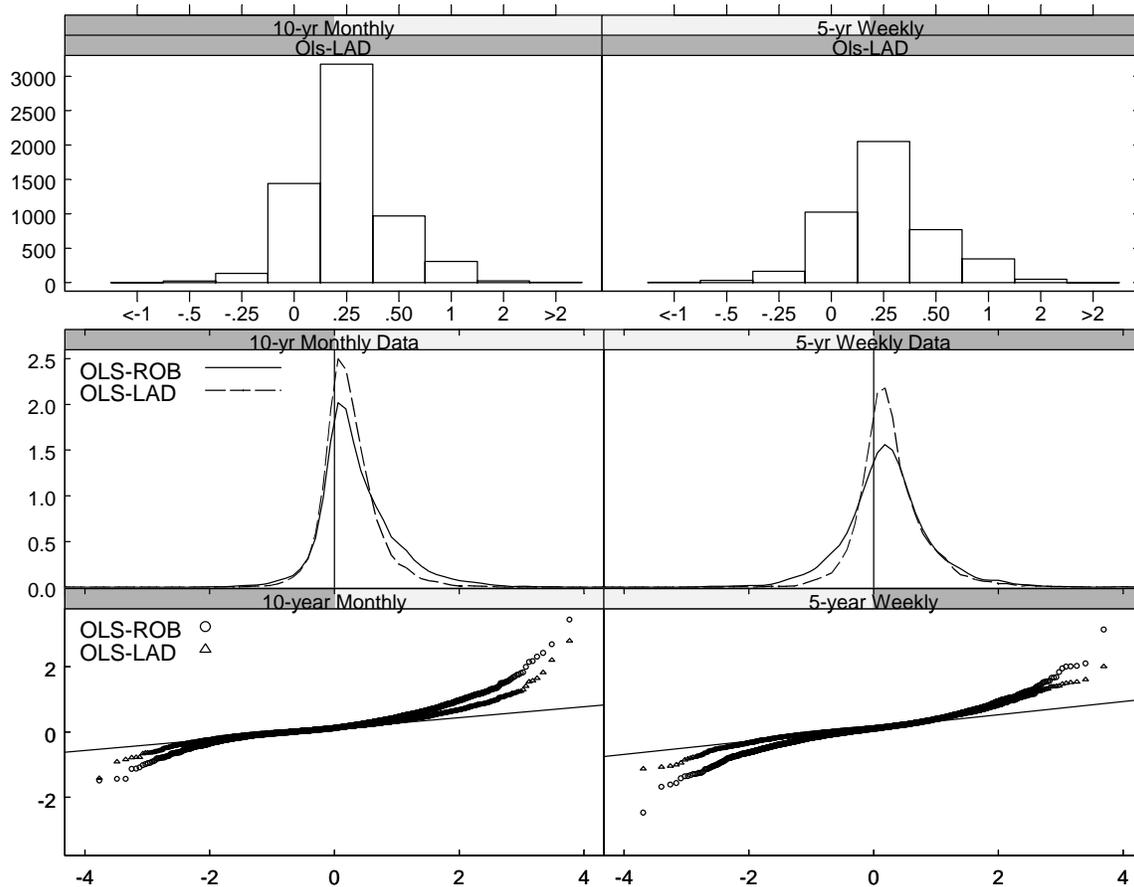


Figure 8. Distributions of OLS – LAD and OLS – ROB. In the upper two panels, the LAD betas for each data set are sorted by their difference from the matching OLS betas. The counts in each bar are the number of betas that fall in a given Δ bin. The center panels are kernel estimates of the probability densities of the difference between OLS and LAD and also for OLS and ROB for the 10-year monthly and 5-year weekly data sets. The bottom panels of the figure contain normal quantile-quantile plots of the difference between the OLS and LAD and also for OLS and robust estimates.

For both the monthly and weekly data sets, 73% of the differences between the OLS and LAD estimates are positive, indicating that on average the OLS estimate is positively biased relative to LAD. However, the greater concentration of the OLS-LAD distribution around zero as compared to the distribution of OLS-ROB is apparent in the plots of Figure 8. This provides evidence that the LAD estimator is less robust toward outliers, i.e., the LAD values tend to be closer to those of OLS than do the values of ROB. At the same time, the normal qq-plots in Figure 8 reveal that the distributions of the differences OLS – LAD and OLS – ROB are both heavy-tailed, with the latter more heavy-tailed than the former. This means that the largest differences in absolute value are larger for

the latter than for the former. This is further evidence that the robust estimate is less biased than the LAD estimate.

A sharper comparison of the LAD and ROB estimates is provided by the Trellis display of conditional boxplots of the differences OLS – LAD and OLS - ROB in Figure 9. This

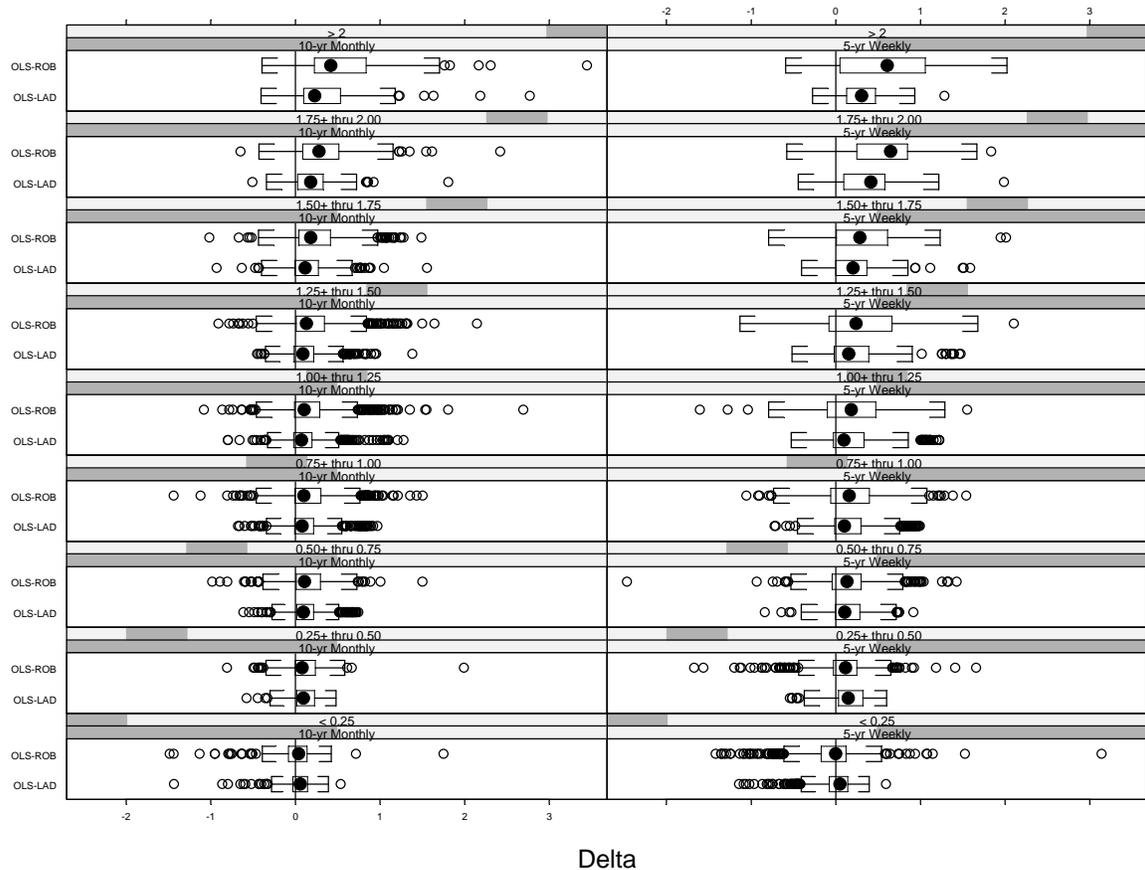


Figure 9. OLS – LAD and OLS – ROB Distributions Conditioned on OLS Beta. The box plots of the difference between the estimates produced by OLS and LAD compared to the difference between the OLS and robust estimates for the 10-year monthly and 5-year weekly data conditional on the OLS beta being in a given bin. The differences between OLS and the two robust betas are sorted based on the size of the OLS beta and the distribution of the differences is then plotted. The solid dot represents the median of the distribution. The box indicates the inter-quartile range (IQR: the middle half of the data), the whiskers cover the range 1.5* IQR and points beyond the whiskers may be considered outliers.

display provides a view of the distributions of the differences OLS – LAD and OLS - ROB conditional on the size of the OLS estimate. The distribution of the difference between OLS and LAD follows a similar pattern to that of OLS and ROB, with an increasing rightward shift and increased spread of the distribution as the conditioning variable OLS increases in. But the shift to the right is more pronounced for the distribution of OLS – ROB, and the largest values of OLS – ROB tend to be larger than the largest values of OLS – LAD, as was noted above.

In summary, the LAD estimate is substantial improvement over OLS relative to robustness toward bias producing outliers, but the new robust estimate appears to be

consistently better than the LAD estimate in this regard. The new robust estimate can be expected to give better protection against outliers that occur in leverage configurations. In addition it may be noted that the LAD estimate is rather unattractive from the standpoint of constructing confidence intervals and t-statistics. The reason is that the (asymptotic) variance expression involves the value of the error density at the origin, and one is faced with the unappealing task of implementing a reliable nonparametric density estimate in order to construct confidence intervals and t-statistics.

7. PREDICTION PERFORMANCE

A good robust regression estimate of beta produces a fitted line that better represents the bulk of the data than does OLS. Thus one expects that the prediction errors in predicting new equity returns from a model parameterized by the robust technique will be smaller than errors from a model parameterized by OLS. We test this intuition using the market model (1.1) for the weekly time series of Section 5.

For each firm in the sample, the model is parameterized using data in a training period and returns are predicted in a test period. Both one-year and two-year training and test intervals are used as indicated in the first two columns of Table 7. The forecast errors are evaluated by the median of the absolute differences between the predictions from the model estimated during the training period and the actual returns occurring in the test period.

Table VIII

Average Median Prediction Errors

The forecast errors are the difference between the returns predicted by the market model parameterized using the data in the training period and the actual returns in the testing period. The numbers reported are the cross-sectional mean of the median absolute errors of each equity in the weekly data set and the 100 firms with the largest difference between OLS and the robust betas

Training Period	Test Period	Full Data Set		100 Firm Data Set	
		OLS	ROBUST	OLS	ROBUST
1992	1993	0.00722	0.00710	0.00720	0.00717
1993	1994	0.00688	0.00683	0.00686	0.00672
1994	1995	0.00667	0.00662	0.00665	0.00660
1995	1996	0.00702	0.00692	0.00715	0.00692
1992-1993	1994	0.00675	0.00668	0.00667	0.00666
1993-1994	1995	0.00663	0.00656	0.00672	0.00681
1994-1995	1996	0.00681	0.00673	0.00682	0.00661
1992-1993	1994-1995	0.00663	0.00660	0.00666	0.00670
1994-1995	1995-1996	0.00666	0.00659	0.00668	0.00665
1992-1995	1996	0.00678	0.00666	0.00675	0.00654
1992-1994	1994-1996	0.00663	0.00652	0.00666	0.00652

This is a robust measure of prediction-error scale that is not much influenced by outliers in the new returns. This measure seems quite reasonable because one can not expect to predict the outliers with any degree of success (and a method which works well in this regard would be jealously guarded). The analysis was performed on all firms in the weekly data set as well as on a subset consisting of the 100 firms with the largest differences between the OLS and robust beta estimates. The results are reported in Table 7 and in the Trellis dot-chart display of Figure 10.

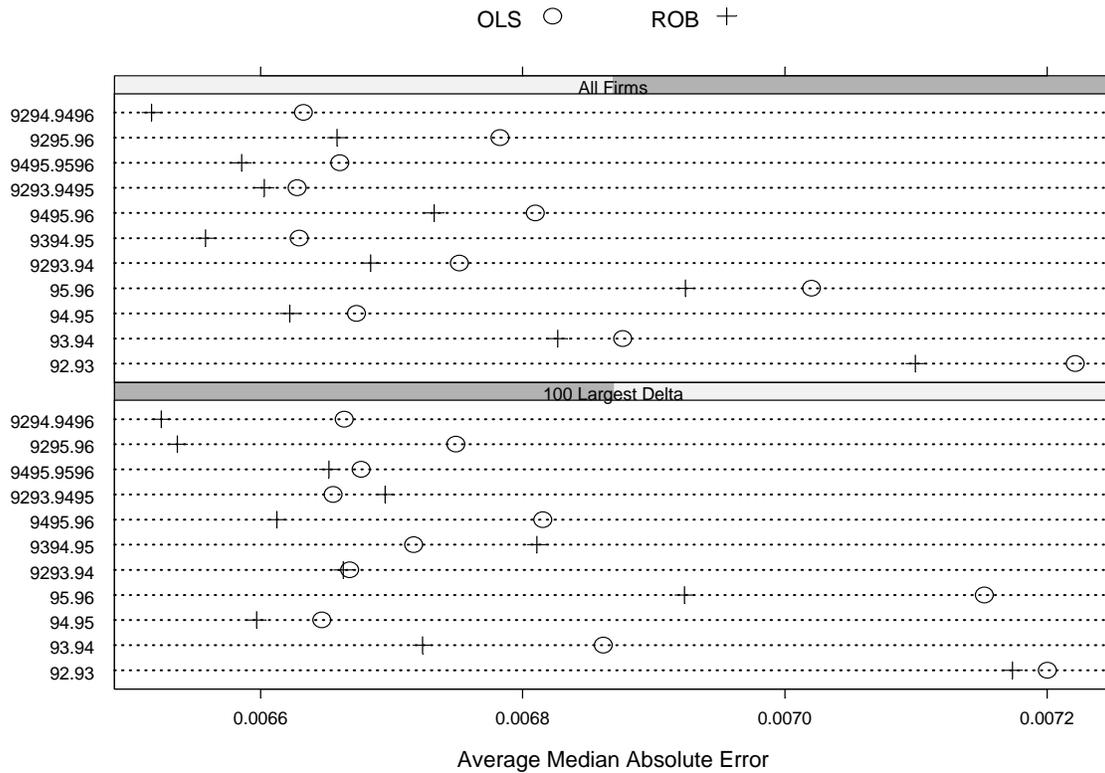


Figure 10. Median Absolute Error of Returns Predictions. The prediction errors are the difference between the returns predicted by the market model parameterized using the data in the training period (years on the left of the decimal point) and the actual returns in the (years on the right of the decimal point) testing period. The numbers reported are the cross-sectional mean of the median absolute errors of each equity in the weekly data set (upper panel) and the 100 firms with the largest difference between OLS and the robust betas (lower panel).

For all the different subsets defined by the training and testing periods, the robust estimator exhibits smaller median absolute errors on average than does OLS when the entire cross-section of firms is considered

When the cross section is restricted to those firms having the largest difference between the OLS and robust betas, there are only two instances when OLS clearly appears to predict returns better on average than the robust technique. Overall, the predictions using the robustly parameterized model are on average better than those produced by OLS.

8. CONCLUDING DISCUSSION

Because a few outliers can have such a strong influence on the OLS beta, classical OLS may provide a distorted view of the true risk versus expected return characteristics of a particular equity. The intuition is straightforward. A very small number of exceptionally large outlier returns giving rise to a large value of beta, say in excess of 2, do not justify an expectation of future return movements twice those of the market. Furthermore, if the outliers have the same positive sign, it does not seem appropriate to say that the equity is twice as risky as the market (at least outside of some empirical Bayesian model).

On the other hand, the robust beta accurately reflects the risk versus expected return characteristics of the *bulk* of the returns for a given equity, without regard to the behavior of outliers. If one believes that the bulk of the data is the only predictable part, then the robust beta is reasonable to use with regard to the future (if one had a useful predictor of outliers, one could reap large profits indeed!). The prediction results in Section 7 lend credibility to the robust beta being a better predictor than the OLS beta for the bulk of future returns.

The result of comparing the behavior of both the classical OLS estimate and a new robust estimate of beta for historical equity returns show emphatically that outliers are a real cause for concern. In particular, the effects of a very small fraction of outliers, even a single large outlier, can substantially bias the OLS estimate of beta. The proposed robust regression method yields a beta estimate that is resistant to outliers. Consequently, large differences between the OLS and robust beta estimates occur with regard to both investor subjective metrics and a formal robust statistical test. Our results show that the large differences between OLS and robust beta's are associated with small market capitalization firms. Table VI and Figure 7 reveal the particular form of the relationship. There is a clear negative relationship between the location and scale of the distribution of firm market capitalization and the size of the differences between OLS and robust beta's.

It is fair to say that when influential outlier returns exist, neither the OLS beta nor the robust beta provides an adequate summary of an equities risk versus expected return characteristics. However, the two estimates may be used together with considerable effectiveness. When the difference between the two estimates is less than a specified Δ , based on either economic significance or statistical significance, report the classical result. Otherwise, report the value of both estimates, along with supporting information. A minimal report should include the times of occurrence and sizes of outliers that are given zero weights by the robust regression estimate of beta. A more extensive report would include event information, idiosyncratic and/or systematic, associated with the times of occurrences of outliers.

The assumption that equity returns are generated by the mixture model (2.2) suggests estimating the probability distribution model from the data via maximum likelihood. One might even be able to use a parametric normal distribution for h . For monthly and weekly data, estimation of the mixture distribution parameters is difficult given the sample sizes and the typically small fraction of equity returns observations generated by

the density h . Moreover, both the fraction of outliers and the form of h may change with time (and more rapidly than that of f_o) reducing the reliability of estimates of the mixture model distribution. In fact, the situation faced when estimating beta is exactly that addressed by robustness. To paraphrase Huber (1981), “Robustness is doing well *near* a parametric model”. Here the parametric model is f_o in (2.3), and nearness is measured by the mixing parameter γ which appears to be typically quite small for equity returns. When there is little hope of an adequate parametric or non-parametric approach, robust techniques may provide a reasonable solution.

We have stressed robustness in the sense of combining the following two properties: (a) control of bias due to outliers generated by the mixture model (2.2) with arbitrary density h , and (b) achieving a specified high-efficiency in the event that the data is normally distributed. The first property is relevant given that the empirical behavior of the data suggests that such a mixture model is reasonable. The second property is desirable since bias can be the dominating component of mean-squared-error in sample sizes such as those encountered when estimating beta. The advantages of this estimator are not limited to the case of data from a bias producing mixture distribution.

The robust estimator is not constrained to use with data from a mixture distribution. The choice of ψ , based on efficiency-constrained bias optimality is fortuitous even if the outliers are generated by a very heavy-tailed symmetric density f_o , rather than by the mixture density h . In this case bias will not be an issue but controlling variance becomes paramount. It has been known for some time that use of a non-monotonic ψ -function similar to the one we use is needed to obtain high-efficiency at very heavy-tailed symmetric distributions as well as at the normal distribution (the Huber type ψ -function is not adequate in this regard). See for example Andrews et. al. (1972). Thus it does not matter whether the outliers are generated by a heavy-tailed symmetric error distribution f_o or by a bias producing mixture model density h . The ψ we use will do well in either case.

It is worth noting several open issues. The first is that of using relative differences versus log-ratio definitions of returns (discrete versus continuous compounding of interest). In section 2.3 we pointed out the asymmetry effect of outliers induced by use of the relative difference definitions and reported that computation on a few examples revealed using the log-ratio definition did little to reduce the size of the difference between the OLS and robust estimates. Determination of whether the existence of outliers leads to preferring the log-ratio definition to the relative difference definition of returns would be efficacious. Another issue is that the existence of heteroscedasticity is plausible even from visual inspection of many plots such as those in Figure 1. It would be useful to have a heteroscedastic consistent estimate of standard errors for our robust regression estimate, analogous to that of White (1980). In this regard it is to be noted that under strong heteroscedasticity, a robust estimate may down-weight residuals in a very non-optimal manner. Finally, the frequent non-normal character of the bivariate distribution of equity returns versus market returns motivates more in-depth study of modeling this distribution.

BIBLIOGRAPHY

- Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J. W. Tukey, "Robust Estimates of Location: Survey and Advances," Princeton University Press, Princeton N.J, 1972.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter, "Statistics for Experimenters," *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1978.
- Becker, R. A. and W. S. Cleveland, "The Design and Control of Trellis Display," *Journal of Computational and Statistical Graphics*, 5, (1996) 123-155.
- Blume, M. E., "On the Assessment of Risk," *Journal of Finance*, 26:1 (March 1971) 1-10.
_____, "Betas and Their Regression Tendencies," *Journal of Finance*, 30:3 (June 1975) 785-795.
- Bruner, R. F., K. M. Eades, R. S. Harris, R. C. Higgins, "'Best Practices' in Estimating the Cost of Capital: Survey and Synthesis," *University of Washington Working Paper*, (March 1996).
- Carleton, W. T. and J. Lakonishok, "Risk and Return on Equity: The Use and Misuse of Historical Estimates," *Financial Analysts Journal*, (January 1985) 38-47.
- Chan, L. K. C. and J. Lakonishok, "Robust Measurement of Beta Risk," *Journal of Financial and Quantitative Analysis*, 27:2 (June 1992) 265-282.
- Connolly, R. A., "An Examination of the Robustness of the Weekend Effect," *Journal of Financial and Quantitative Analysis*, 24:2 (June 1989) 133-169.
- Cook, R. D. and S. Weisberg, "Residuals and Influence in Regression," Chapman and Hall, New York, (1982).
- Cornell, B. and J. K. Dietrich, "Mean-Absolute-Deviation Versus Least-Squares Regression Estimation of Beta Coefficients," *Journal of Financial and Quantitative Analysis*, 13 (March 1978) 123-131.
- Elton, E. J., M. J. Gruber, and T. J. Urich, "Are Betas Best?" *Journal of Finance*, 33:5: (December 1978) 1375-1384.
- Eubank, A. A. Jr. and J. K. Zumwalt, "An Analysis of the Forecast Error Impact of Alternative Beta Adjustment Techniques and Risk Classes," *Journal of Finance*, 34:3 (June 1979) 761-776.
- Fama, E. "The Behavior of Stock Prices," *Journal of Business*, 38 (Jan. 1965) 34-105.

- _____, and Kenneth R. French, "Cross-section of Expected Stock Returns," *Journal of Finance*, 47 (1992) 427-466.
- Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, "Robust Statistics. The Approach Based on Influence Functions," *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1986.
- Huber, Peter J., "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, 35 (1964) 73--101.
- _____, "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1 (1973) 799--821.
- _____, "Robust Statistics," *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1981. (Reviewer: Rudolph J. Beran)
- Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T. C. Lee, "Introduction to the Theory and Practice of Econometrics," Second edition. John Wiley & Sons, Inc., New York, 1988. (Reviewer: Kajal Lahiri)
- Klemkosky, R. C. and J. D. Martin, "The Adjustment of Beta Forecasts," *Journal of Finance*, 30:4 (September 1975) 1123-1129.
- Knez, P. J. and M. J. Ready (1997), "On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions," *Journal of Finance*; 52(4), September 1997, pages 1355-82.
- Koenker, R., and G. Bassett, "Regression Quantiles," *Econometrica*, 46 (Jan. 1978) 33-50.
- Kon, S. "Models of Stock Returns – A Comparison," *Journal of Finance*, 39 (March 1984) 147-165.
- Mackinlay, A. C. and M. P. Richardson, "Using Generalized Method of Moments to Test Mean-Variance Efficiency" *Journal of Finance*, 2 (June 1991).
- Mandelbrot, B. "The Variation of Certain Speculative Prices," *Journal of Business*, 36 (Oct. 1963) 394-419.
- Martin, R. D. and R. H. Zamar, "Asymptotically Min-Max Bias Robust M-estimates of Scale for Positive Random Variables," *Journal of American Statistical Assoc.*, 84 (1989) 494-501.

- _____, V. J. Yohai, and R. H. Zamar "Min-Max Bias Robust Regression," *The Annals of Statistics*, 17:4 (1989) 1608-1630.
- _____, and R. H. Zamar, "Bias Robust Estimation of Scale," *The Annals of Statistics*, 21:2 (1993) 991-1017.
- _____, "Trellis Displays for Deep Understanding of Financial Data," *Financial Engineering News*, 2:1 (1998).
- Richardson, Matthew and Tom Smith, "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information," *Journal of Financial and Quantitative Analysis*, (1994) 101-116.
- Roll, R., "R²" *Journal of Finance*, 43 (July 1988) 541-566.
- Rousseeuw, P. J. and V. J. Yohai, "Robust Regression by Means of S-estimators," *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics*, 26, Springer, New York, (1984) 265-272.
- _____, "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, (December 1984) 871-880.
- _____, and Annick M. Leroy, "Robust Regression and Outlier Detection," *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons, Inc., New York, 1987. (Reviewer: V. J. Yohai)
- Sharpe, W. F., "Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios," *Management Science*, 18:2 (Oct. 1971) B1-B13.
- Vasicek, O., "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas," *Journal of Finance*, 28 (December 1973) 1233-1239.
- White, H., "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity," *Econometrica*, 48, (1980), pp. 817-838.
- Wiggins, J. B., "Betas in Up and Down Markets," *The Financial Review*, 27:1 (Feb. 1992) 107-123.
- Yohai, V. J., "High Breakdown-point and High Efficiency Estimates for Regression," *The Annals of Statistics*, 15 (1988) 642-665.
- _____, W. A. Stahel, and R. H. Zamar, "A Procedure for Robust Estimation and Inference in Linear Regression," in Stahel, W. A. and S. W. Weisberg, Eds., *Directions in Robust Statistics and Diagnostics, Part II*, Springer-Verlag, New York (1991).

_____, and R. H. Zamar, "Optimal Locally robust M-estimates of Regression,"
Working Paper, (1997).

APPENDIX

A.1 The Analytic Expressions for the ρ and ψ -functions

These are the functions displayed in Figure 2.

$$\rho(r;c) = \begin{cases} 3.25 \cdot c^2 & |r/c| > 3 \\ c^2 \cdot \left[1.792 - 0.972 \cdot \left(\frac{r}{c}\right)^2 + 0.423 \cdot \left(\frac{r}{c}\right)^4 - 0.052 \cdot \left(\frac{r}{c}\right)^6 - 0.002 \cdot \left(\frac{r}{c}\right)^8 \right] & 2 < |r/c| \leq 3 \\ 0.5 \cdot r^2 & |r/c| \leq 2 \end{cases} \quad (\text{A.1.1})$$

and

$$\psi(r;c) = \begin{cases} 0 & |r/c| > 3 \\ c \cdot \left[-1.944 \cdot \left(\frac{r}{c}\right) + 1.728 \cdot \left(\frac{r}{c}\right)^3 - 0.312 \cdot \left(\frac{r}{c}\right)^5 + 0.016 \cdot \left(\frac{r}{c}\right)^7 \right] & 2 < |r/c| \leq 3 \\ r & |r/c| \leq 2 \end{cases} \quad (\text{A.1.2})$$

A.2 The Computational Algorithm

The key to obtaining a good local minimum to the M-estimation problem (3.1) when using a bounded, non-convex loss function such as ρ is to compute a good, highly robust initial estimate $\hat{\beta}^0$. We do so by using the regression S-estimate approach of Rousseeuw and Yohai (1984).

The S-estimate approach has as its foundation an M-estimate \hat{s} of an unknown scale parameter s for observations y_1, y_2, \dots, y_n , assumed to be robustly centered (i.e., by subtracting a robust location estimate). The M-estimate \hat{s} is obtained by solving the equation

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{y_i}{\hat{s}}\right) = .5 \quad (\text{A.2.1})$$

where ρ is a symmetric, bounded function. It is known that such a scale estimate has a breakdown point of one-half (Huber, 1981), and that one can find min-max bias robust M-estimates of scale (Martin and Zamar, 1989, 1993).

Rousseeuw and Yohai (1984) introduced the following regression S-estimate method. Consider the linear regression model modification of (A.2.1):

$$\frac{1}{n-p} \sum_{i=1}^n \rho\left(\frac{y_i - x_i^T \beta}{\hat{s}(\beta)}\right) = .5. \quad (\text{A.2.2})$$

For each value of β we have a corresponding robust scale estimate $\hat{s}(\beta)$. A regression S-estimate is the value $\hat{\beta}^0$ that minimizes $\hat{s}(\beta)$:

$$\hat{\beta}^0 = \arg \min_{\beta} \hat{s}(\beta). \quad (\text{A.2.3})$$

Once the initial S-estimate $\hat{\beta}^0$ is computed, take the minimum value of $\hat{s}(\beta)$ attained in (A.2.3) as an initial robust scale estimate \hat{s}^o .

The computation (A.2.3) is a non-linear optimization problem, one for which a solution is usually sought by using a random resampling algorithm to find an preliminary estimate, $\tilde{\beta}^0$ followed by a local search refinement to find $\hat{\beta}^0$, as described in Yohai, Stahel and Zamar (1991). For our study of robust estimation of beta where there are just two coefficients, the slope and intercept, and the sample sizes are not too large ($n \leq 260$), we have replaced the random resampling algorithm with an exhaustive search over all pairs of observations.

The final M-estimate $\hat{\beta}$ is then obtained as the nearest local minimum of (2.1) to $\hat{\beta}^0$, using the high-efficiency bias-robust function ρ and with \hat{s}^o as the robust scale estimate in (2.1). A final robust scale estimate \hat{s} is then obtained by using $\hat{\beta}$ in (A.2.2) and solving for $\hat{s} = s(\hat{\beta})$.

A.3 Approximate Standard Error Estimation

Standard errors are obtained by computing the estimate $\hat{C}_{\hat{\beta}}$ in (3.5). The natural estimate of \hat{v} in (3.5) is

$$\hat{v} = \frac{(\hat{s}^o)^2 \sum_{i=1}^n \psi^2(\tilde{r}_i)}{\left(\sum_{i=1}^n \psi'(\tilde{r}_i) \right)^2} \quad (\text{A.3.1})$$

where

$$\tilde{r}_i = \frac{y_i - x_i^T \cdot \hat{\beta}^o}{\hat{s}^o}. \quad (\text{A.3.2})$$

The above approach to estimating v was proposed in Yohai, Stahel and Zamar (1991), who also proposed the following important weighting of X to obtain $\hat{C}_{\hat{\beta}}$. Let

$$w_i = \frac{\psi(\tilde{r}_i)}{\tilde{r}_i}, \quad i = 1, \dots, n \quad (\text{A.3.3})$$

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i \quad (\text{A.3.4})$$

$$\tilde{X} = \frac{1}{\sqrt{\bar{w}}} \text{diag}(\sqrt{w_i}) \cdot X. \quad (\text{A.3.5})$$

Then use \tilde{X} to compute

$$C \hat{\beta} = \frac{1}{n} (\tilde{X}^T \tilde{X})^{-1} \cdot \hat{v} \quad (\text{A.3.6})$$

with \hat{v} as described above.

The motivation for this last modification is that sufficiently large residuals are rejected in the estimation procedure, and therefore the associated increased variability of the coefficient estimates needs to be reflected by replacing the corresponding rows of X with vectors shrinking to zero for sufficiently large outliers.

Endnotes

¹ Estimates of beta from various commercial suppliers such as Bloomberg, Value Line, and Dow Jones Markets differ in use of market index, data frequency, and/or adjustments for the reversion of beta towards 1 over time. For instance, Value Line regresses the logarithm of the price ratio (P_{t+1}/P_t) on the logarithm of the price ratio for the NYSE composite without an intercept using the latest five years of weekly data and then adjusting beta by $0.67\hat{\beta} + .35$. Bloomberg provides a variety of betas with the default beta being adjusted as above and using 2 years of weekly data with the S&P 500 as the market index. Dow Jones provides betas for up and down markets by including a dummy variable, equal to 1 if the market risk premium is positive and 0 if negative, in the standard regression. For a review of different adjustment techniques and factors affecting commercially estimated betas see Vasicek (1971), Blume (1971, 1975), Klemkosky and Martin (1975), Elton, Gruber, and Urich (1978), Eubank and Zumwalt (1979), Carleton and Lakonishok (1985), and Wiggins (1992).

² See for example Judge et. al. (1988), Huber (1981), Rousseeuw and Leroy (1987), and Hampel et. al. (1986).

³ The MADM estimate is scaled to be a consistent estimate of the standard deviation when the data is normally distributed.

⁴ It may be noted that in the regression context the influence of an outlier can be particularly potent when it is a *leverage* point, namely when it is associated with an outlying (normed) value of the predictor vector. In the context of (1.1) this occurs when a particular market return is an outlier relative to the bulk of the market returns. Further discussion of outliers with leverage may be found for example in Judge (1988) and Cook and Weisberg (1982).

⁵ Mackinlay and Richardson (1991) discuss the sensitivity of inference when asset returns are not normally distributed.

⁶ Since the variance of an M-estimate goes to zero like n^{-1} under regularity, squared bias is the dominant component of mean-squared error in large samples. Furthermore, the dominance of the bias component can occur at surprisingly small sample sizes, as is shown by Martin and Zamar, 1989 in the context of robust estimation of scale.

⁷ An alternative data-oriented consideration is that mixture distributions of the type (2.2) – (2.3) can give rise to outliers with *leverage*, which have considerable influence on regression M-estimates of the Huber and LAD type. One way to avoid such leverage effects is to use a regression M-estimate with a bounded loss function and corresponding redescending psi-function.

⁸ Martin, Yohai and Zamar (1989) showed that a certain *bounded- ρ* M-estimates highly robust by virtue of minimizing the maximum asymptotic bias, and that these M-estimates are very well-approximated by the least-median-of-squares (LMS) estimate introduced by Rousseeuw (1984). Unfortunately, like the LMS approximating estimate, the min-max bias robust estimates have a slow rate of convergence and zero asymptotic efficiency when the regression model errors are normal.

⁹ Martin and Zamar (1993) obtained an efficiency constrained bias-robust estimate of location, but over the class of monotonic ψ functions.

¹⁰ It is interesting to note that there does not exist a maximum-likelihood with a score function ψ that has value zero outside a finite interval.

¹¹ The discontinuity in the hard-rejection psi-function causes a loss of efficiency at the Gaussian model for a given level of maximum bias control. The discontinuity also makes estimation of the variance expression difficult, the reason being that the variance expression then involves the probability density of the errors.

¹² It may be noted that the ratio of the OLS standard errors to robust standard errors, i.e., the robust estimator's efficiency in terms of standard errors, at the Gaussian model is $0.084/.114 \cong .84$. This is rather lower than expected in comparison with the asymptotic standard error efficiency of .92.