Robust Prediction of Beta

Marc G. Genton[1] and Elvezio Ronchetti[2]

[1] Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland. Marc.Genton@metri.unige.ch
[2] Department of Econometrics, University of Geneva, Bd du Pont-d'Arve 40, CH-1211 Geneva 4, Switzerland. Elvezio.Ronchetti@metri.unige.ch

Summary. The estimation of β plays a basic role in the evaluation of expected return and market risk. Typically this is performed by ordinary least squares (OLS). To cope with the high sensitivity of OLS to outlying observations and to deviations from the normality assumptions, several methods suggest to use robust estimators. It is argued that, from a predictive point of view, the simple use of either OLS or robust estimators is not sufficient but that some shrinking of the robust estimators toward OLS is necessary to reduce the mean squared error. The performance of the proposed shrinkage robust estimator is shown by means of a small simulation study and on a real data set.

Key words: Beta, CAPM, outliers, robustness, shrinkage.

Preamble

In the past 20 years Manfred Gilli has played an important role in the development of optimization techniques and computational methods in a variety of disciplines, including economics and finance, econometrics and statistics. He has played a leading role both in teaching and research in the Department of Econometrics of the University of Geneva, where he has always been a fixed point and a reference in this field and a constant source of advice and intellectual stimulation. It is therefore a great pleasure to contribute to this Festschrift in his honor and to wish him many more years of productive activity.

1 Introduction

In spite of some of its recognized shortcomings, the Capital Asset Pricing Model (CAPM) continues to be an important and widely used model for the estimation of expected return and the evaluation of market risk. Typically, the rate of return r_t of a given security with respect to that of the risk-free asset r_{ft} is regressed on the rate of return of the market index r_{mt} and the following single factor model is fitted by ordinary least squares (OLS)

$$r_t - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \sigma \cdot u_t, \qquad t = 1, \dots, n, \tag{1}$$

where α and β are unknown regression parameters and σ is the standard deviation of the errors. It is well known that from a statistical point of view, standard OLS technology presents several drawbacks. In particular, its high sensitivity in the presence of outliers and its dramatic loss of efficiency in the presence of small deviations from the normality assumption have been pointed out in the statistics literature (see, for instance, the books by Huber (1981), Hampel et al. (1986), Rousseeuw and Leroy (1987), Maronna et al. (2006), in the econometric literature (see, for instance, Koenker and Bassett (1978), Chow (1983), Krasker and Welsch (1985), Peracchi (1990), Krishnakumar and Ronchetti (1997), Ronchetti and Trojani (2001)), and in the finance literature (see, for instance, Sharpe (1971), Chan and Lakonishok (1992), Knez and Ready (1997), Martin and Simin (2003)).

Robust statistics was developed to cope with the problem arising from the approximate nature of standard parametric models; see references above. Indeed robust statistics deals with deviations from the stochastic assumptions on the model and develops statistical procedures which are still reliable and reasonably efficient in a small neighborhood of the model. In particular, several well known robust regression estimators were proposed in the finance literature as alternatives to OLS to estimate β in (1). Sharpe (1971) suggested to use least absolute deviations (or the L_1-estimator) instead of OLS. Chan and Lakonishok (1992) used regression quantiles, linear combinations of regression quantiles, and trimmed regression quantiles. In a study to analyze the risk premia on size and book-to-market, Knez and Ready (1997) show that the risk premium on size estimated by Fama and French (1992) is an artefact of using OLS on data with outliers and it disappears when using robust estimators such as least trimmed squares. Finally, Martin and Simin (2003) proposed to estimate β by means of so-called redescending M-estimators.

These robust estimators produce values of β which are more reliable than those obtained by OLS in that they reflect the majority of the historical data and they are not unduly influenced by abnormal returns. In fact, robust estimators downweight (by construction) outlying observations by means of weights which are not fixed in an arbitrary way by the analyst but which are automatically determined by the data. This is a good practice if one wants to have a measure of risk which reflects the structure of the underlying process as revealed by the bulk of the data. However, a familiar criticism of this approach in finance is that "abnormal returns *are the important observations*," that "the analyst is precisely interested in these data and therefore they *should not be downweighted*." We believe that this criticism does not hold if the main goal of the analysis is to produce an estimate of β which reflects the pattern of historical data, but it has some foundation from the point of view of *prediction*.

Indeed if abnormal returns are not errors but legitimate outlying observations, they will likely appear again in the future and downweighting them by using robust estimators will result in a potentially severe bias in the prediction of β. On the other hand, it is true that OLS will produce in this case unbiased estimators of β but this is achieved by paying a potentially important price of a large variability in the prediction of β. Therefore, we are in a typical situation of a trade-off between bias and variance and we can improve upon a simple use of either OLS or a robust estimator. This calls for some form of shrinkage of the robust estimator toward OLS to achieve the minimization of the mean squared error. Similar ideas have been used in the framework of sample surveys containing outliers to estimate totals and quantiles of a given variable in a finite population; see the original paper by Chambers (1986), Welsh and Ronchetti (1998), and Kuk and Welsh (2001).

The paper is organized as follows. In Section 2 we derive shrinkage robust estimators of β by using a Bayesian argument in the framework of the standard linear model. This leads to a particular form for the shrinkage estimator of β. In Section 3 we provide some empirical evidence to illustrate the behavior of different prediction methods on two data sets. Section 4 presents the results of a small Monte Carlo simulation study and shows the potential gains in prediction power that can be obtained by using shrinkage robust estimators. Finally, some conclusions and a brief outlook are provided in Section 5.

2 Shrinkage Robust Estimators of Beta

In this section we derive our shrinkage robust estimator of β in the general setup of the standard linear model

$$y_i = x_i^T \beta + \sigma \cdot u_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $\beta \in \mathcal{R}^p$, $\sigma > 0$, and the u_i's are iid with density $g(\cdot)$.

Our shrinkage robust (SR$_c$) estimator of β is defined by

$$\tilde{\beta}_c = \hat{\beta}_R + \left(\sum_{i=1}^{n} x_i x_i^T \right)^{-1} \sum_{i=1}^{n} \hat{\sigma} \psi_c \left(\frac{y_i - x_i^T \hat{\beta}_R}{\hat{\sigma}} \right) x_i, \tag{3}$$

where $\hat{\beta}_R$ is a robust estimator of β, $\hat{\sigma}$ is a robust estimator of scale, and $\psi_c(\cdot)$ is the Huber function defined by $\psi_c(r) = \min(c, \max(-c, r))$. Typically, $\hat{\beta}_R$ is one of the available robust estimators for regression such as, e.g., an M-estimator (Huber, 1981), a bounded-influence estimator (Hampel et al., 1986), the least trimmed squares estimator (Rousseeuw, 1984) or the MM estimator (Yohai, 1987), and $\hat{\sigma}$ is a robust estimator of scale such as the median absolute deviation (Hampel et al., 1986). The tuning constant c of the Huber function in (3) is typically larger than the standard value 1.345; see below.

The estimator $\tilde{\beta}_c$ defined by (3) can be viewed as a least squares estimator with respect to the "pseudovalues" $\tilde{y}_i = x_i^T\beta + \sigma\psi_c\left(\frac{y_i - x_i^T\beta}{\sigma}\right)$.

The structure of $\tilde{\beta}_c$ is as follows. The parameter β is first estimated by means of a robust estimator $\hat{\beta}_R$ and scaled residuals $r_i = (y_i - x_i^T\hat{\beta}_R)/\hat{\sigma}$ are obtained. Next the Huber function $\psi_c(r_i)$ is applied to each residual. When r_i is larger in absolute value than c, the residual is brought in to the value $c\,\mathrm{sign}(r_i)$, whereas it is not modified otherwise. By using a moderate value of c, i.e., a value larger than the standard value of 1.345 typically used in robust statistics, only very large residuals are trimmed thereby insuring both robustness (low variability) and small bias. The limiting case $c \to \infty$ gives back OLS (unbiasedness but high variability) while a very small value of c leads to $\hat{\beta}_R$ (low variability but large bias).

Further insight can be gained from a predictive Bayesian argument. Assume a prior density $h(\cdot)$ for β and a quadratic loss function. Then, the Bayes estimator of β is given by

$$\begin{aligned}\beta_B(y_1,\ldots,y_n) &= E\left[\beta \mid y_1,\ldots,y_n\right] \\ &= \frac{\int \beta \Pi_{i=1}^n \sigma^{-1} g((y_i - x_i^T\beta)/\sigma)h(\beta)d\beta}{\int \Pi_{i=1}^n \sigma^{-1} g((y_i - x_i^T\beta)/\sigma)h(\beta)d\beta} \\ &= \frac{\int \beta \exp\{-nk(\beta; y_1,\ldots,y_n)\}d\beta}{\int \exp\{-nk(\beta; y_1,\ldots,y_n)\}d\beta}, \end{aligned} \tag{4}$$

where $k(\beta; y_1,\ldots,y_n) = -\frac{1}{n}\left[\sum_{i=1}^n \log\left(\sigma^{-1}g((y_i - x_i^T\beta)/\sigma)\right)\right] - \frac{1}{n}\log h(\beta)$.

Using Laplace's method to approximate the integrals in (4), we can write

$$\beta_B(y_1,\ldots,y_n) = \hat{\beta}\left[1 + O(n^{-1})\right], \tag{5}$$

where $\hat{\beta} = \arg\max_{\beta} k(\beta; y_1,\ldots,y_n)$.

If we choose $g(r) \propto \exp(-\rho(r))$ for some positive, symmetric function ρ, and the prior $h(\beta)$ as $N(\mu, \delta^2 I)$, where I is the identity matrix, then $\hat{\beta}$ satisfies the estimating equation

$$\sum_{i=1}^n \frac{1}{\sigma}\psi\left((y_i - x_i^T\hat{\beta})/\sigma\right)x_i - \frac{\hat{\beta} - \mu}{\delta^2} = 0, \tag{6}$$

where $\psi(r) = \frac{d\rho(r)}{dr}$.

Expanding the first term on the left hand side of (6) about β_0 and ignoring terms of order $\|\hat{\beta} - \beta_0\|^2$ or smaller we obtain

$$0 \approx \sum_{i=1}^{n} \frac{1}{\sigma} \psi\left((y_i - x_i^T \beta_0)/\sigma\right) x_i \tag{7}$$

$$- \left(\sum_{i=1}^{n} \frac{1}{\sigma^2} \psi'\left((y_i - x_i^T \beta_0)/\sigma\right) x_i x_i^T \right) (\hat{\beta} - \beta_0) - \frac{\hat{\beta} - \mu}{\delta^2}, \tag{8}$$

and

$$(\mathit{\Omega} + I)\, \hat{\beta} \approx \delta^2 \sum_{i=1}^{n} \frac{1}{\sigma} \psi\left((y_i - x_i^T \beta_0)/\sigma\right) x_i + \mathit{\Omega}\beta_0 + \mu,$$

where $\mathit{\Omega} = \sum_{i=1}^{n} \frac{\delta^2}{\sigma^2} \psi'\left((y_i - x_i^T \beta_0)/\sigma\right) x_i x_i^T$.
Therefore,

$$\begin{aligned} \hat{\beta} &= (\mathit{\Omega} + I)^{-1} \mathit{\Omega} \hat{\beta}_R + (\mathit{\Omega} + I)^{-1} \mu \\ &= (\mathit{\Omega} + I)^{-1} \mathit{\Omega} \left[\hat{\beta}_R + \mathit{\Omega}^{-1} \mu \right], \end{aligned} \tag{9}$$

where $\hat{\beta}_R = \beta_0 + \mathit{\Omega}^{-1} \delta^2 \sum_{i=1}^{n} \frac{1}{\sigma} \psi\left((y_i - x_i^T \beta_0)/\sigma\right) x_i$.

Notice that $\hat{\beta}_R$ is a one-step M-estimator defined by ψ (Hampel et al., 1986, p. 106). Also it is natural to let $\mathit{\Omega}$ be based on $\hat{\beta}_R$. For large δ^2 (which corresponds to a vague prior for β) we have from (9) :

$$\hat{\beta} \approx \hat{\beta}_R + \hat{\mathit{\Omega}}^{-1} \mu, \tag{10}$$

where $\hat{\mathit{\Omega}} = \frac{\delta^2}{\sigma^2} \sum_{i=1}^{n} \psi'\left((y_i - x_i^T \hat{\beta}_R)/\sigma\right) x_i x_i^T$, which corresponds to (3) when we replace $\hat{\mathit{\Omega}}$ by the slightly simpler $\frac{\delta^2}{\sigma^2} \sum_{i=1}^{n} x_i x_i^T$ and we estimate the prior mean μ by $\sum_{i=1}^{n} \sigma \psi_c \left((y_i - x_i^T \hat{\beta}_R)/\sigma\right) x_i$.

Therefore, $\tilde{\beta}_c$ can be viewed as the Bayes estimator in this setup. This argument provides a justification for the use of an additive adjustment in (3). However, it allows for a wide class of such adjustments, interpreted as alternative estimates for the prior mean μ.

The tuning constant c controls the amount of shrinking and is typically larger than standard values used for robust Huber estimators. It can be chosen by cross-validation to minimize the mean squared error or an alternative measure of prediction error; see also Chambers (1986).

3 Empirical Evidence

In this section, we investigate the prediction of beta from two stock returns (Shared Technologies Inc. and Environmental Tectonics Corp.) versus market returns. These data, also analyzed by Martin and Simin (2003), were obtained from the Center for Research on Stock Prices (CRSP) database and are displayed in the top row of Figure 1. Some outlying values can be observed in both plots, especially for the stocks of Environmental Tectonics Corp. In the bottom row of Figure 1, we plot OLS (solid curve) and robust (dotted curve) betas computed using a 104-week moving window. There are dramatic differences between the two curves (solid and dotted) occurring when the outlying value(s) enters the moving window. We also computed a shrinkage robust estimator (dashed curve) of beta obtained from Equation (3) with $c = 1.5$. This estimator shrinks the robust estimator of beta toward the OLS beta in order to account for possible future outliers.

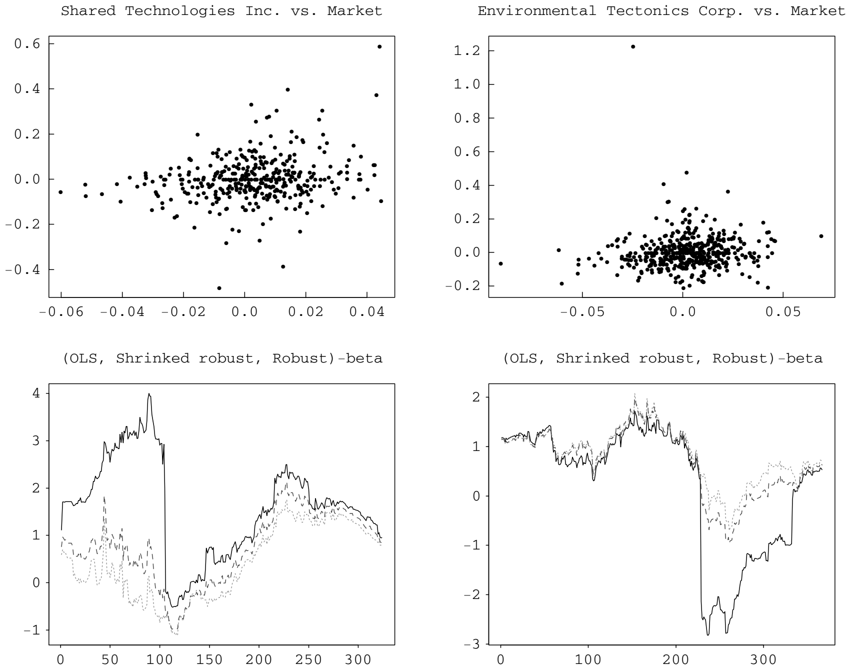


Fig. 1. Top row: plots of stock returns versus market returns. Bottom row: plots of one-week ahead predictions of betas using a 104-week moving window and: OLS-beta (solid curve); shrinkage robust-beta (dashed curve); robust-beta (dotted curve).

In order to illustrate the effect of outliers on betas, the top panel of Figure 2 depicts the CAPM model fitted by OLS on three weeks of Shared Technologies Inc. vs market data without outlier (solid line) and with one additional outlier (short-dashed line) represented by an open circle. The effect of this single

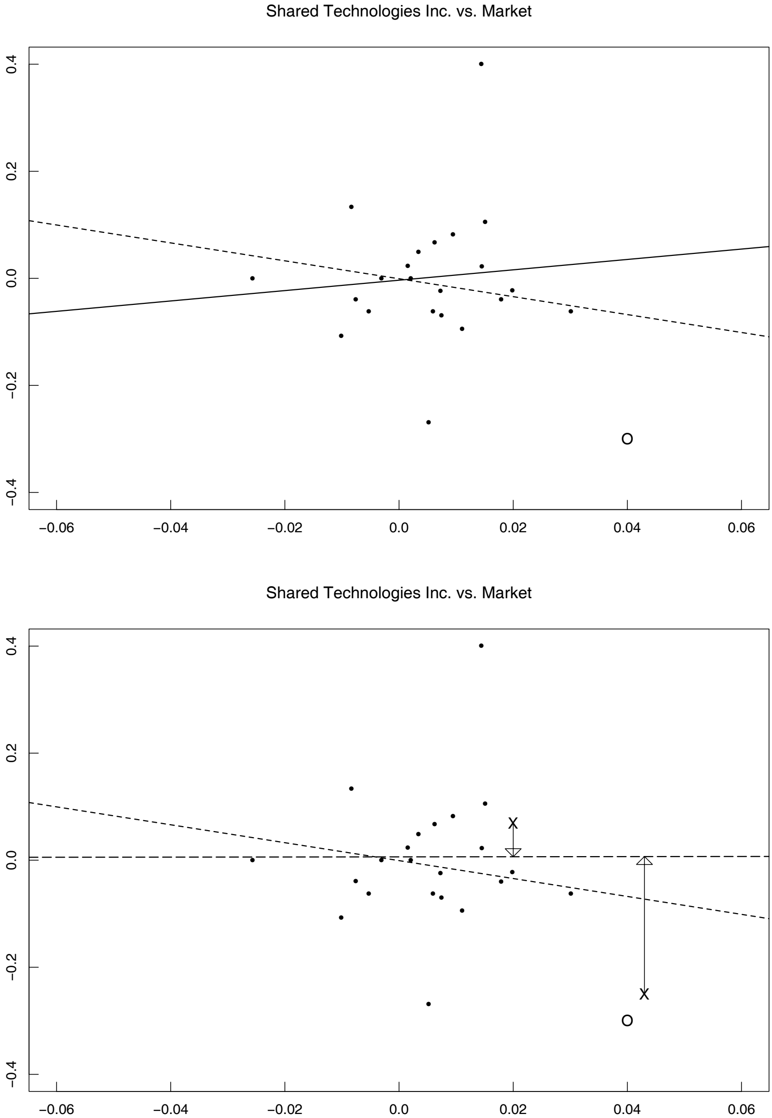


Fig. 2. Top panel: fitted CAPM by OLS on three weeks of Shared Technologies Inc. vs market data: without outlier (solid line), with one additional outlier (short-dashed line). Bottom panel: fitted CAPM by OLS (short-dashed line) and robust procedure (long-dashed line). The prediction of future new points (X) could be inaccurate if outliers occur again.

outlier is a leverage that changes the positive value of beta to a negative one. The bottom panel of Figure 2 depicts the previous contaminated situation with the CAPM model fitted by a robust procedure (long-dashed line). The prediction of future new points represented by an X could be inaccurate if outliers occur again. This points to the need of shrinking the robust estimator toward the OLS estimator.

4 Monte Carlo Simulations

We investigate the robust prediction of beta through Monte Carlo simulations. For the sake of illustration, we consider the model (2) with $p = 1$, that is,

$$y_i = \beta x_i + \sigma \cdot u_i, \qquad i = 1, \dots, n, \tag{11}$$

with $\sigma = 1$, u_i's iid $N(0,1)$, and x_i's iid $N(0,4)$. We choose values of beta that are representative of situations found in real data sets, $\beta = 0.5$ and $\beta = 1.5$, and construct two simulated testing samples from model (11) of size $m = 400$ each. These two testing samples represent future observations that are not available at the modeling stage. In order to reproduce possible outliers found in practice, we include 10% of additive outliers to the x_i's and to the y_i's from a $N(0, \tau^2)$ distribution with $\tau = 5$. Figure 3 depicts these two testing samples where some outliers, similar to the ones found in real data, can be seen.

We use the least trimmed squares (LTS) regression estimator, $\hat{\beta}_{LTS}$, to obtain robust estimates of beta. It has a bounded influence function and a breakdown-point of nearly 50%. The LTS estimator is defined as the value of β that minimizes the sum of the smallest $\lfloor n/2 \rfloor + 1$ of the squared residuals, where $\lfloor . \rfloor$ denotes rounding down to the next smallest integer. Alternative robust estimators could be used instead of the LTS. In particular, we ran our simulation with the redescending M-estimator used by Martin and Simin (2003) and obtained similar results.

The LTS and OLS estimators of beta on the two testing samples are $\hat{\beta}_{LTS} = 0.379$ and $\hat{\beta}_{OLS} = 0.327$ for the case $\beta = 0.5$, and $\hat{\beta}_{LTS} = 1.492$ and $\hat{\beta}_{OLS} = 0.982$ for the case $\beta = 1.5$. The corresponding regression lines are drawn in Figure 3. They indicate that the effect of outliers on the estimation of beta is not too strong in the first testing sample, whereas it is much stronger in the second one.

The shrinkage robust (SR_c) estimator (3) of β in our context becomes

$$\tilde{\beta}_c = \hat{\beta}_{LTS} + \left(\sum_{i=1}^{n} x_i^2 \right)^{-1} \sum_{i=1}^{n} \hat{\sigma} \psi_c \left(\frac{y_i - \hat{\beta}_{LTS} x_i}{\hat{\sigma}} \right) x_i, \tag{12}$$

where $\hat{\sigma}$ is the median absolute deviation (MAD), a robust estimator of scale. (Again, other robust estimators of scale could be used instead.) In order to analyze the effect of outliers on the estimation of beta, we simulate 1,000

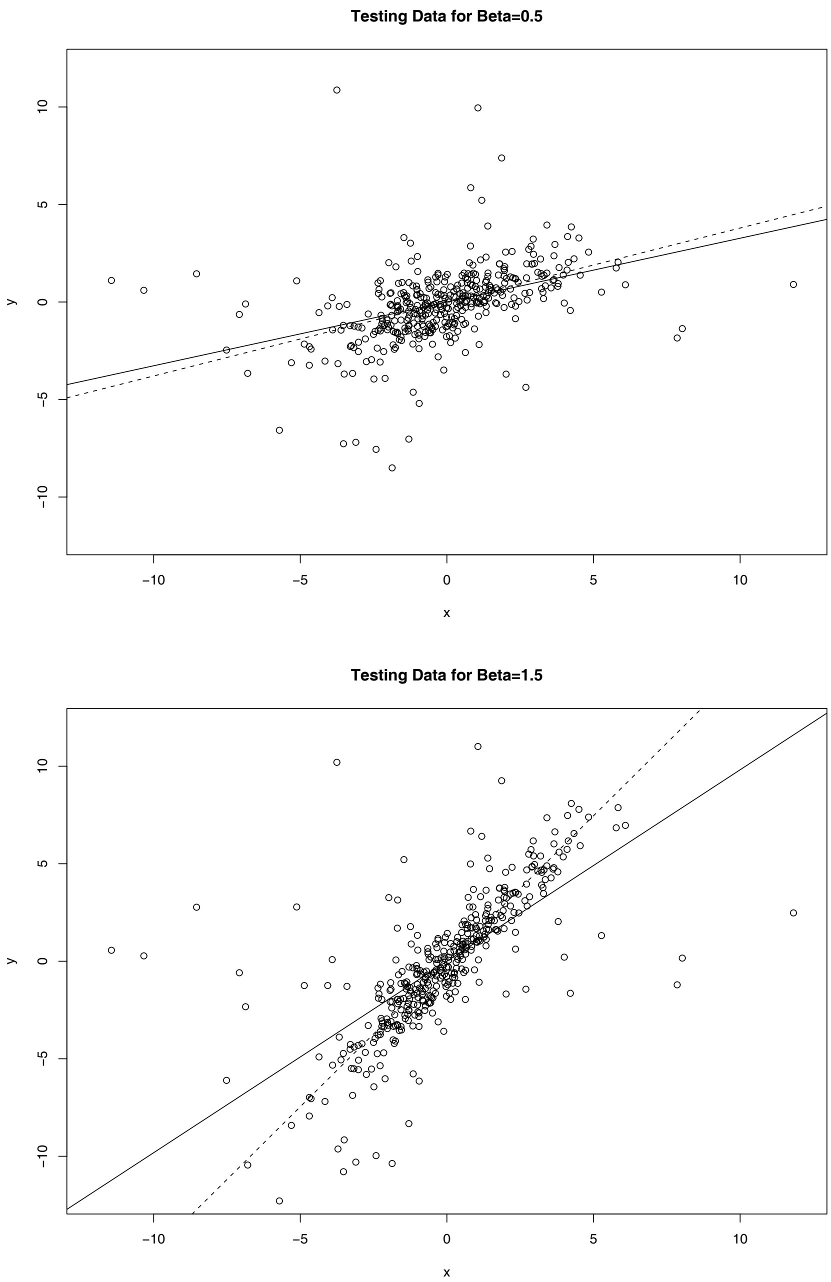


Fig. 3. Scatter plot of $m = 400$ testing data containing 10% outliers with fitted regression line by OLS (solid line) and LTS (dashed line). Top panel: $\beta = 0.5$. Bottom panel: $\beta = 1.5$.

training data sets of size $n = 100$ each, containing outliers, following the same procedure as described above for the testing samples. Hence the structure of the outliers is the same in both the training and the testing samples. For each sample, we estimate β by LTS, OLS, and SR_c with $c = 1, \ldots, 10$. Figure 4 depicts boxplots of these estimates over the 1,000 simulated training data containing outliers for $\beta = 0.5$ (top panel) and $\beta = 1.5$ (bottom panel). The horizontal solid line represents the true value of beta. The robust estimator LTS has a much smaller bias than the OLS estimator, but usually a larger variance. The shrinkage robust estimators behave between the boundary cases SR_0 =LTS and SR_∞ =OLS. Note however that for some values of c, the variance of SR_c is reduced at the cost of a small increase in bias.

Next, we investigate the effect of outliers and robust estimates of β on the prediction of future observations represented by the two testing data plotted in Figure 3. Specifically, for each of the 1,000 estimates $\hat{\beta}$, we compute the predicted values $\hat{y}_i = \hat{\beta} x_i$, $i = 1, \ldots, m$ with $m = 400$, on the two testing samples.

A measure of quality of prediction should depend on the given problem and there is a large amount of literature in econometrics and economics on the choice of the loss function for prediction purposes. Some authors (see, e.g., Leitch and Tanner (1991) and West et al. (1993)) argue that the loss function should reflect an economically meaningful criterion, while others consider statistical measures of accuracy which include, e.g., mean squared error, absolute error loss, linex loss among others. While we agree that these discussions are important, it is not our purpose in this paper to compare the different proposals. Instead we argue that the robustness issue (and therefore the possible deviations from the stochastic assumptions of the model) should also be taken into account when choosing the loss function.

In the absence of specific information, a typical choice is the root mean squared error. Note, however, that this measure is related directly or indirectly to the Gaussian case. In non-Gaussian cases there are no standard choices and a good recommendation is to compute several such measures. To illustrate this point, we focus here on a particular class of measures defined by

$$Q(p) = \left(\frac{1}{m} \sum_{i=1}^{m} \mid y_i - \hat{y}_i \mid^p \right)^{1/p} . \tag{13}$$

Three choices of p give the following quantities: $p = 2$ yields the root mean squared error (RMSE); $p = 1$ the mean absolute error (MAE); and $p = 1/2$ the square root absolute error (STAE). We use these measures to evaluate the quality of the prediction. For each of them, Table 1 reports the percentage of selection of a minimum measure of prediction for a range of values of $c = 0, 1, 2, \ldots, 10, \infty$ over the 1,000 simulation replicates for the scenarios with $\beta = 0.5$ and $\beta = 1.5$. The highest percentage for each measure is identified in bold fonts. As can be seen, the optimal c which minimizes a certain measure of quality of prediction is not exclusively concentrated at the boundary esti-

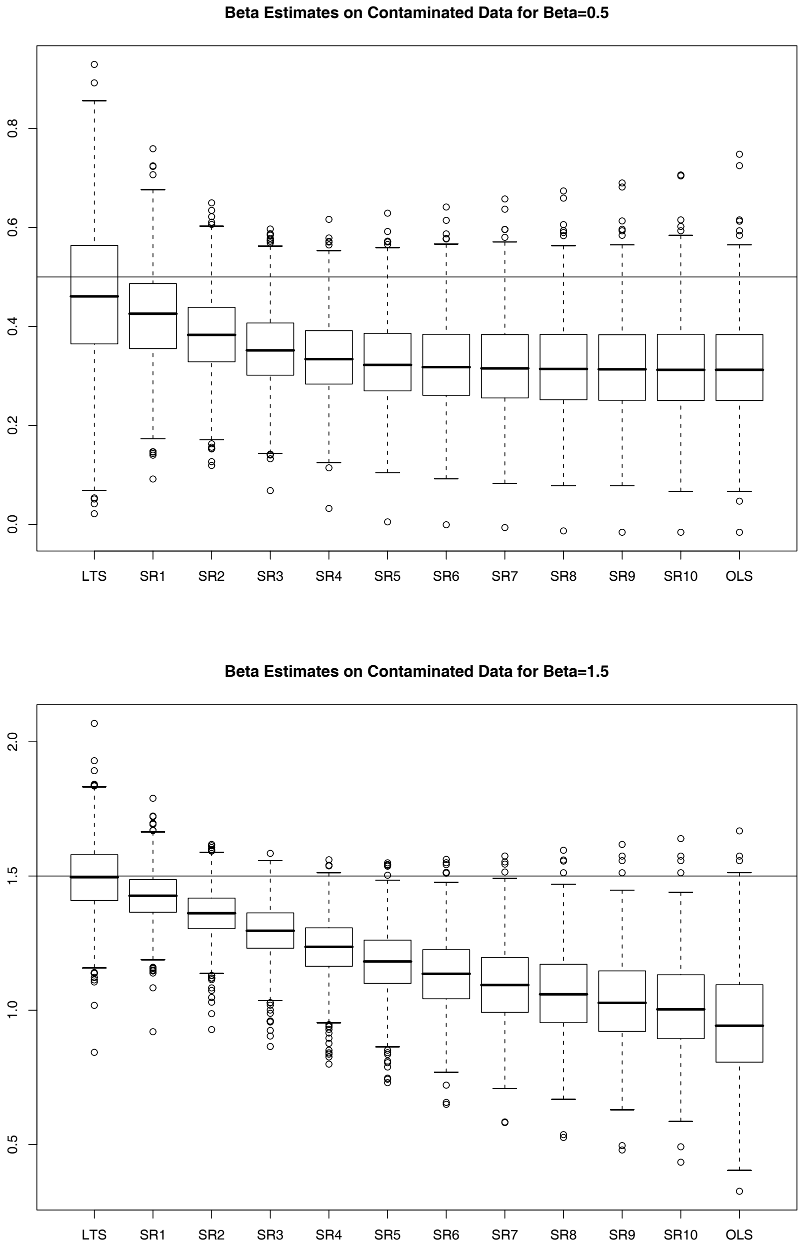


Fig. 4. Boxplots of estimates of beta on 1,000 simulated training data containing 10% outliers. The horizontal solid line represents the true value of beta. Top panel: $\beta = 0.5$. Bottom panel: $\beta = 1.5$.

mators LTS and OLS. Although a more rigorous comparison would involve formal tests of predictive accuracy such as the standard Diebold and Mariano (1995) test or its more general extension by Giacomini and White (2006), the effects seem to be clear: it is fair to say that better predictions of future beta can be obtained by making use of shrinkage robust estimators, that is, by shrinking the robust estimator toward OLS to minimize the trade-off between bias and variance.

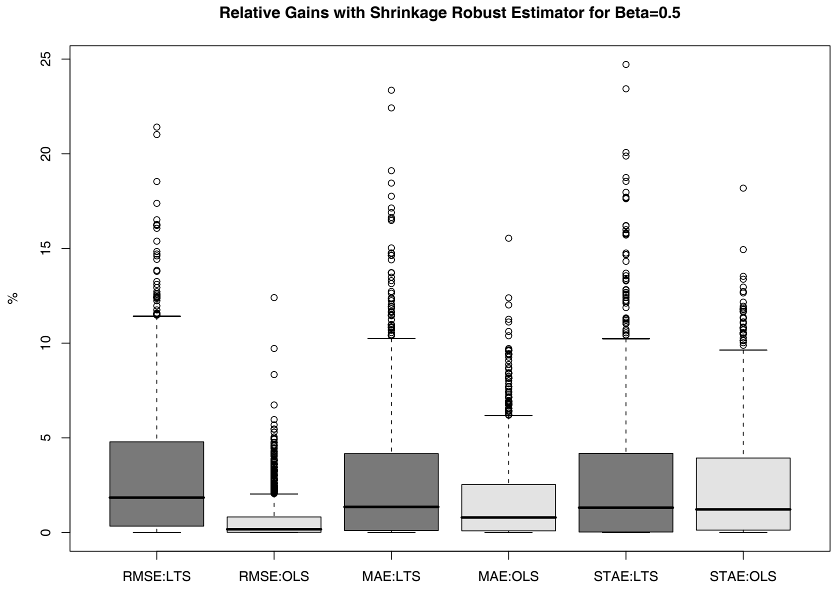
Table 1. Percentage of selection of a minimum measure of prediction (RMSE, MAE, STAE) for a range of values of c $(0, 1, 2, \ldots, 10, \infty)$ over 1,000 simulation replicates for the scenarios with $\beta = 0.5$ and $\beta = 1.5$. The highest percentage for each measure is identified in bold fonts.

	LTS	SR_1	SR_2	SR_3	SR_4	SR_5	SR_6	SR_7	SR_8	SR_9	SR_{10}	OLS
$\beta = 0.5$												
RMSE	10.2	10.6	14.2	16.5	11.6	9.3	5.3	2.6	1.2	0.9	0.6	**17.0**
MAE	18.0	18.2	**22.5**	13.1	6.5	4.6	3.3	1.2	1.3	0.4	0.5	10.4
STAE	**22.8**	21.5	21.2	11.1	5.2	3.9	2.6	2.1	1.2	0.7	0.7	7.0
$\beta = 1.5$												
RMSE	1.3	0.0	0.3	0.9	2.4	6.5	10.2	8.1	9.6	7.5	11.4	**41.8**
MAE	15.9	25.2	**31.2**	14.0	6.5	2.8	1.8	0.8	0.6	0.2	0.2	0.8
STAE	32.5	**35.0**	20.6	7.6	2.9	0.7	0.4	0.1	0.1	0.1	0.0	0.0

The relative gains in reduction of the measure of quality of prediction obtained with shrinkage robust estimators compared to LTS and OLS are investigated next. Specifically, we define the relative gain by

$$RG_{c,d}(p) = \frac{Q_d(p) - Q_c(p)}{Q_d(p)}, \tag{14}$$

where $Q_c(p)$ denotes a measure of quality of prediction obtained with a shrinkage robust estimator SR_c. Denote by c^* the optimal value of c, that is, the value of c minimizing $Q_c(p)$ for a fixed p. Figure 5 depicts boxplots over the 1,000 simulation replicates of $RG_{c^*,0}(p)$ and $RG_{c^*,\infty}(p)$ for $p = 2, 1, 1/2$. In terms of RMSE, the gains of the SR_{c^*} compared to the OLS are fairly small, although they can reach 10-20%. In terms of MAE and STAE, the gains are more important and can reach up to 40% in the case $\beta = 1.5$. The gains compared to LTS go in the other direction.



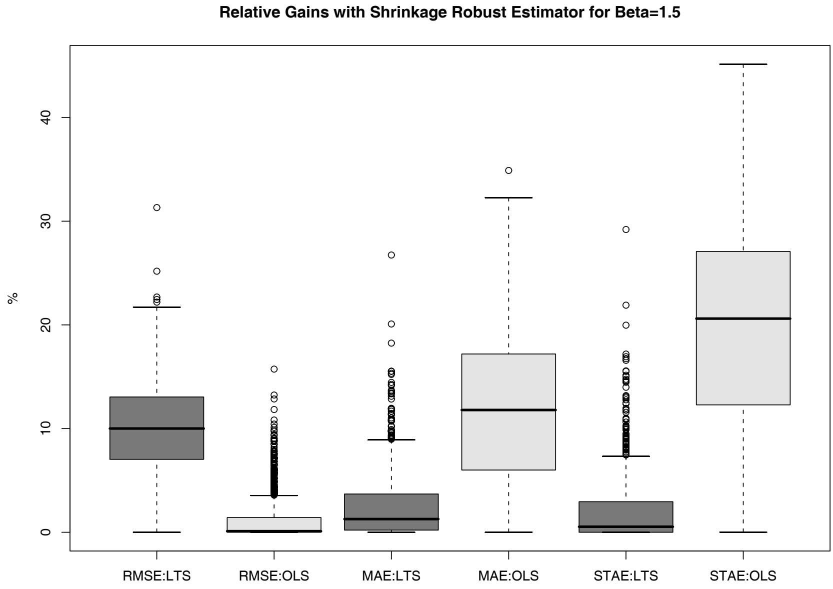


Fig. 5. Relative gains obtained with shrinkage robust estimators compared to LTS and OLS on various measures of prediction (RMSE, MAE, STAE). Top panel: $\beta = 0.5$. Bottom panel: $\beta = 1.5$.

5 Conclusion

In this paper we presented a shrinkage robust estimator which improves upon OLS and standard robust estimators of β from a predictive point of view in the presence of deviations from the normality assumption. The ideas developed here can be used beyond the simple framework of the CAPM and the standard linear model and are basic building blocks to introduce robustness in any prediction problem. This is particularly useful in financial applications but also in many other fields in economics and natural sciences.

Acknowledgements

The authors would like to thank the referees for helpful comments and relevant references.

References

Chambers, R.L.: 1986, Outlier robust finite population estimation, *Journal of the American Statistical Association* **81**, 1063–1069.

Chan, L.K.C. and Lakonishok, J.: 1992, Robust measurement of beta risk, *Journal of Financial and Quantitative Analysis* **27**, 265–282.

Chow, G.: 1983, *Econometrics*, McGraw-Hill, New York.

Diebold, F.X. and Mariano, R.S.: 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**, 253–263.

Fama, E.F. and French, K.R.: 1992, The cross-section of expected stock returns, *The Journal of Finance* **47**, 427–466.

Giacomini, R. and White, H.: 2006, Tests of conditional predictive ability, *Econometrica* **74**, 1545–1578.

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

Huber, P.J.: 1981, *Robust Statistics*, Wiley, New York.

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

Koenker, R. and Bassett, G.: 1978, Regression quantiles, *Econometrica* **46**, 33–50.

Krasker, W.S. and Welsch, R.E.: 1985, Resistant estimation for simultaneous equations models using weighted instrumental variables, *Econometrica* **53**, 1475–1488.

Krishnakumar, J. and Ronchetti, E.: 1997, Robust estimators for simultaneous equations models, *Journal of Econometrics* **78**, 295–314.

Kuk, A.Y.C. and Welsh, A.H.: 2001, Robust estimation for finite populations based on a working model, *Journal of the Royal Statistical Society Series B* **63**, 277–292.

Leitch, G. and Tanner, J.E.: 1991, Economic forecast evaluation: Profits versus the conventional error measures, *American Economic Review* **81**, 580–590.

Maronna, R.A., Martin, R.D. and Yohai, V.J.: 2006, *Robust Statistics: Theory and Methods*, Wiley, New York.

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

Peracchi, F.: 1990, Robust M-estimators, *Econometric Reviews* **9**, 1–30.

Ronchetti, E. and Trojani, F.: 2001, Robust inference with GMM estimators, *Journal of Econometrics* **101**, 37–69.

Rousseeuw, P.J.: 1984, Least median of squares regression, *Journal of the American Statistical Association* **79**, 871–880.

Rousseeuw, P.J. and Leroy, A.M.: 1987, *Robust Regression and Outlier Detection*, Wiley, New York.

Sharpe, W.F.: 1971, Mean-absolute-deviation characteristic lines for securities and portfolios, *Management Science* **18**, B1–B13.

Welsh, A.H. and Ronchetti, E.: 1998, Bias-calibrated estimation from sample surveys containing outliers, *Journal of the Royal Statistical Society Series B* **60**, 413–428.

West, K.D., Edison, H.J. and Cho, D.: 1993, A utility-based comparison of some models of exchange rate volatility, *Journal of International Economics* **35**, 23–45.

Yohai, V.: 1987, High breakdown-point and high efficiency robust estimates for regression, *The Annals of Statistics* **15**, 642–656.