ESTIMATING MARKET MODEL BETAS USING LEAST ABSOLUTE VALUE ESTIMATION: DOES IT MAKE A DIFFERENCE?

Jan G. De Gooijer*, Stuart Bretschneider† and Peter Koveos†

ABSTRACT: The beta coefficient is the traditional measure of systematic risk in investment and portfolio analysis. Typically beta estimates are generated by the ordinary least squares (*OLS*) method. However, the validity of this method is based on certain assumptions which have been challenged in a number of empirical studies. The purpose of this note is to apply a robust estimation procedure to estimating market model betas and compare its performance with results obtained by *OLS*. The results indicate that statistically significant differences in the beta estimates can be found between the two estimation procedures. The nature and significance of the differences varies with the number of observations available for each firm and may be construed as an "interval" or as a "firm maturity" effect.

*Corresponding author: Department of Economic Statistics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Tel. +31 - 20 - 525 4244; e-mail: jandeg@butler.fee.uva.nl.

†The Maxwell School, Syracuse University, Syracuse, NY 13244, U.S.A.

1 INTRODUCTION

The market model has been a focal point of research in finance. It has been therefore subjected to extensive testing. Beta estimates have been generated under alternative sample characteristics, including interval length and market proxies. For example, beta stability has been examined in terms of both its performance of mean and mean absolute deviations with varying conclusions about its performance; see, *e.g.*, Alexander and Chervany (1980). Typically beta estimates are generated through some application of Ordinary Least Squares (OLS) estimation. Applying OLS techniques, however, requires that certain assumptions about the distribution of the error term be true. Unfortunately, these assumptions may not be justified. As a result there has been an increassed interest in robust estimation methods; see, *e.g.*, Huber (1977). One such robust method of estimation is the Least Absolute Value (LAV)[1] estimation; see, *e.g.*, Bloomfield and Steiger (1983). The LAV regression method has become popular in recent years, largely due to the relative insensitivity of LAV estimators to outliers, and the development of fast algorithms for computing the estimators. In this paper, we apply the LAV estimation procedure to the estimation of market model betas. The principle question is: does the use of (LAV) approach, as an alternative to OLS estimation, make a systematic difference in the quantification of risk?

According to the market model, the relationship between the return of an individual security and the market index is depicted by

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_t \tag{1}$$

where R_{it} is the return on asset i $(i = 1, \ldots, n)$ at time t $(t = 1, \ldots, N)$ and R_{mt} is the return on a market index. Assuming that the error term ϵ_t, has zero mean and constant variance, then: $E(R_{it}) = \alpha_i + \beta_i E(R_{mt})$. The total variance in the return of the asset is

$$Var(R_{it}) = \beta_i^2 Var(R_{mt}) + Var(\epsilon_t). \tag{2}$$

This last relation is fundamental in finance. The first component on the right-hand side represents the systematic or common risk. The second component is the non-systematic

[1] LAV is not the only acronym used. Others include MAD (minimum absolute deviation), LAD (least absolute deviation), LAE (least absolute error), and $L1$ norm.

risk, attributable to the peculiarities of the individual return. If the distribution of ϵ_t is symmetric with finite variance, then the non-systematic component can be reduced to a value not significantly different from zero, by diversifying the number of securities included in a portfolio; see, *e.g.*, Evans and Archer (1968) and Fama (1965a).

The distribution of ϵ_t, however, is not necessarily symmetric with constant variance; see Kon (1984). Instead, a stable Paretian distribution with characteristic exponent δ in the range $0 < \delta < 2$ has been found by some researchers for ϵ_t (*e.g.* for $\delta = 1$ and an index of skewness equal to zero, the Pareto distribution becomes a symmetric Cauchy distribution); see, *e.g.*, Fama and Roll (1968, 1971). If $\delta < 2$, the variance of ϵ_t does not exist. Also evidence has been reported that the distribution of security returns can be characterized by a Student t-distribution with few degrees of freedom. For example, Akgiray and Booth (1988) reject the stable law model for stock returns, finding that empirical tail shapes are thicker than those of a normal distribution but thinner than stable tails. These findings are corroborated by Nieuwland (1993), using extreme value theory. Hence, OLS may not be the appropriate estimation procedure. Indeed, Fama (1965b) suggests that a possible alternative procedure for estimating the relationship between returns on individual securities and the market is provided by the absolute value regression. The regression, according to Fama (1965b), estimates a set of coefficients so as to minimize the sum of absolute deviations between actual and predicted values of the dependent variable. This is an application of LAV.

Following Fama's suggestion, estimates of beta from the market model are generated applying both OLS and LAV techniques. The next section of this study describes the current state of the theory that underlies LAV estimation. It is followed by a discussion of the data and methodology employed in our study. The results are explained and the appropriate conclusions are drawn in the last sections.

2 LEAST ABSOLUTE VALUE ESTIMATION

Traditional approaches to model estimation are predicated on certain assumptions about the nature of the underlying error distribution, which are rarely tested empirically. Some work in finance suggests that these conditions do not uniformly exist; see, *e.g.*, Fama

(1965b). Alternative approaches to estimating relationships exist that do not rely on minimizing the sum of squared errors. For quite some time analysts have known that by minimizing the sum of absolute deviations between a hypothesized model, such as the linear regression model, and data results in robust parameter estimates. Nevertheless, several important problems have prevented LAV approaches from being utilized by researchers.

The initial problem was computational. Unlike OLS estimation, there is no closed form formula based solution to the problem of minimizing the sum of absolute deviations. This changed when Charnes, Cooper and Ferguson (1955) demonstrated that LAV estimation could be accomplished directly through an application of linear programming. Consider the multiple regression model

$$Y_t = \sum_k \beta_k X_{t,k} + \epsilon_t, \;\; (t = 1, \ldots, N), \tag{3}$$

where Y_t is the tth observation of the dependent variable, $X_{t,k}$ is the tth observation of the kth independent variable, β_k is the parameter associated with the kth independent variable, and N is the number of observations. The LAV estimates for β_k, $\hat{\beta}_k$, are those which minimize

$$\sum_t Y_t - \sum_k \beta_k X_{t,k} \tag{4}$$

and the linear programming problem that generates a solution is

$$Minimize \; Z = \sum_t (p_t + q_t) \tag{5}$$

subject to

$$\begin{aligned} Y_t - \sum_k \hat{\beta}_k X_{t,k} + p_t - q_t &= 0, \;\; (t = 1, \ldots, N) \quad (6) \\ p_t, q_t &> 0, \;\; (t = 1, \ldots, N) \quad (7) \end{aligned}$$

where $\hat{\beta}_k$, $(k = 1, \ldots, r)$, are unrestricted in sign, and where p_t and q_t are the positive and negative deviations associated with the tth observation. The problem stated in equations (5)-(7) can be solved directly with the revised simplex algorithm though several researchers have developed specialized codes for dealing with this problem; see, *e.g.* Barrodale and Roberts (1978). Accompanying the above theoretical developments, computer codes for implementation have be written, and diffused. Since 1979, *SAS* has made available *PROCLAV* in its user supplemental library for *LAV* multiple regression analysis. Though there are still problems associated with estimation[2], these developments essentially make it possible to calculate *LAV* estimates for multiple regression analysis quite easily.

A second problem in using *LAV* estimates is that until recently little was known about the sample distribution of the estimates. Various simulation studies have demonstrated that *LAV* estimates are more efficient than traditional *OLS* estimates for small samples (N=10 and 50) where the residuals have Cauchy, Laplace, stable Paretian distributions with the characteristic exponent less than or equal to 1.5, and certain contaminated normal distributions; see, *e.g.*, Rosenberg and Carlson (1977) and Bloomfield and Steiger (1983). These results are significant in that they suggest that situations where error distributions have "fat-tails" (*e.g.* Cauchy, Laplace distributions) or "long-tails", *LAV* estimates will be more efficient than *OLS* estimates, hence generating smaller confidence intervals.

The small sample results generated through simulation studies have been supported by the development of large sample asymptotic properties for *LAV* estimates; see Bassett and Koenker (1978). The essence of this result is that for large samples, the sampling distribution of the *LAV* regression estimates are asymptotically normal with a mean given by the true parameter value β and a variance-covariance matrix $\sigma^2(X'X)^{-1}$ where $\sigma/\sqrt{N}$ is the asymptotic standard deviation of the sample median for random samples of size N taken from the residual distribution, and X is the matrix of all values of $X_{t,k}$, the independent variables. Estimation of σ is problematical though in a summary paper Dielman and Pfaffenberger (1982) demonstrate one approach that results in a consistent estimator.

[2]One problem that remains is that the linear programming solution may not be unique. Most codes can determine the uniqueness of the solution by examination of the dual variables at the optimum. In the case of *SAS*'s *PROCLAV* the output indicates if the solution is unique. Other approaches to estimation of *LAV* exist, such as iteratively weighted *LS*, but these only result in approximate solutions and give no indication as to the potential existence of alternative optima.

The value of LAV estimation then is best understood as an alternative to OLS in situations where the error distribution is thought to be non-normal. In particular, when empirical tests for normality of residual distributions generated by OLS regression result in rejection, this suggests that LAV estimates could be more efficient. This would result in tighter confidence intervals around each parameter estimate, and smaller standard errors. LAV estimates are unbiased estimates of the conditional median but biased estimates of the conditional mean. Consequently, in the presence of skewed error distribution, or error distributions with fat tails it would mean that median regression is a superior approach to summarizing and modelling than traditional conditional mean regression. Another value of LAV estimation is that LAV parameter estimates will not be as strongly affected by outliers in a data set since these observations will not be as heavily weighted as in OLS estimation. This feature is particularly useful when estimating the market model (1) since it is known that estimates of beta, using only a small number of historical observations which can be obscured by outliers, tend to lack stability. In fact, LAV is a member of two important classes of robust regression estimators, M and R; see, *e.g.*, Huber (1977).

3 DATA, METHODOLOGY AND RESULTS

This study utilized monthly returns for 90 firms, selected randomly from the *COMPUSTAT* tape. The tape includes monthly return data from 1962 through 1986. The exact dates of coverage for each firm selected, varied, with 57 companies containing over 200 observations, 14 companies containing between 100 and 199 observations, and the remaining 19 firms with fewer than 100 observations. One firm had only 2 observations and was not used, leaving a total of 89 firms. The highest number of returns generated in the sample was 287, while the lowest was 11. As a market portfolio, the Standard and Poor's 500 Composite Average was used.

For each firm, relation (1) was estimated using both the OLS and the LAV methods of estimation. In terms of testing the applicability of the market model: the OLS procedure generated estimates of beta that were statistically significant at the 1% level in 47 cases, at the 5% level 14 more times and at the 10% level for an additional 7 firms. Overall, 21 firms did not have statistically significant betas when applying OLS procedures. Similar

Table 1: Description of slope differences (OLS slope - LAV slope).

Average	0.0350
$T-$ratio for average	0.8477
Standard deviation	0.3890
Coefficient of variation	1112.8900
Number of times $OLS > LAV$	57.0000
Number of times $LAV > OLS$	32.0000
Skewness	-3.6684
Kurtosis	26.4605
Maximum	1.3057
3rd quartile	0.1690
Median	0.0699
1st quartile	-0.0226
Minimum	-2.6301
Sign rank test statistic	824.5000
Non-parametric test H_0: Mean=0	
$P-$value	0.0007

results were obtained for the LAV procedure. Using the large sample theory developed above, 48 firms had statistically significant LAV beta estimates at the 1% level, 12 at the 5% level and 6 at the 10% level. There were 22 firms where the LAV beta estimate was not significant. Generally the decision to accept or reject the hypothesis that the beta is zero was the same firm by firm, regardless of the estimation procedures.

Table 1 describes the distribution of the difference between the two slope estimates. The mean difference was 0.0350. Statistically, this difference was not significantly different from zero at the 5% level but the measures of skewness and kurtosis suggest that the distribution of differences is not symmetric. Since the number of observations available per security varied considerably and the distribution of differences is skewed, further examination of these differences was warranted.

Table 2 examines this distribution cumulatively by iteratively slicing away cases based on sample size. This table includes both the average and median differences between OLS and LAV estimates, plus for each grouping a test for normality was conducted. Once the three firms with less than 20 observations are removed from the sample the mean difference become significant and positive. The differences change as the various classes of firms are considered. The average and median difference increases initially, but starts to decline after all firms with less than 140 observations (12 years) have been removed. These results suggest a relationship between the extent to which OLS and LAV estimates differ and the sample size used in the estimation. Despite the theoretical evidence that for small sample sizes LAV estimation leads to smaller confidence ellipsiods (see, *e.g.*, Bloomfield and Steiger, 1983, p. 51) than OLS estimation, this was only noticed for four series having N <100.

It is well known that the presence of heteroskedasticity in the disturbances of an otherwise properly specified linear model leads to consistent but inefficient parameter estimates (whatever the estimation method) and inconsistent covariance matrix estimates. As a result faulty inferences may be drawn when testing statistical hypotheses in the presence of heteroskedasticity. To test for homoskedasticity of errors the Goldfeld-Quandt test is used. It requires that the observations are ordered according to increasing error variance, assuming that heteroskedasticity does exist. Next two regressions are run, one using the first $(N-c)/2$ observations and the other using the last $(N-c)/2$ observations. Here c denotes the number of middle observations. Then, assuming that the errors are distributed normally, the two residual sum of squares are compared using an F test. Taking $c = N/3$, the results of the Goldfeld-Quandt test exhibit that only six of the 89 stocks rejected the hypothesis of no difference in the subsample pure residual variances at the 10% level. The second test of heteroskedasticity is the Spearman's rank correlation coefficient. The results of this test indicate that only seven of the 89 stocks tested showed significant correlation at the 10% level. Following the above test results it is reasonable to claim that only 7% of the stocks under study illustrate evidence of heteroskedasticity in the pure residual error. Therefore, no effort has been made to correct the stock prices for the presence of heteroskedasticity. Moreover, this would lead to the estimation of (G)ARCH processes (see, *e.g.*, Bollerslev, Chou and Kroner, 1992) which lies beyond the scope of this paper.

Table 3 presents the results of a regression relating the difference between the two estimation approaches (OLS slope - LAV slope) and the number of observations per security. This is done for all firms (n=89) as well as for firms having more than 200 observations (n=57). According to the estimated parameters, increases in the number of observations per firm are associated with increases in the difference between the OLS

Table 2: Slope differential and number of observations per firm. $N >$ indicates that the firms included in the calculations had a number of observations greater than 20,...; * indicates all mean and median differences are statistically different at a 5% level.

$N >$	No. of firms (n)	Average* difference	Median* difference	$P-$value H_0: Normality
20	86	0.0589	0.0720	<0.01
40	83	0.0624	0.0740	<0.01
60	77	0.0833	0.0782	<0.01
80	74	0.0947	0.0776	<0.01
100	72	0.1029	0.0806	<0.01
120	71	0.1044	0.0831	<0.01
140	71	0.1044	0.0831	<0.01
160	70	0.1040	0.0807	>0.15
180	61	0.0739	0.0740	>0.15
200	57	0.0647	0.0612	0.193
220	48	0.0771	0.0720	0.899
240	27	0.0592	0.0699	0.952
260	23	0.0704	0.0740	0.846
280	19	0.0801	0.0740	0.360

Table 3: Slope differences and number of observations: Regression results. Differences(i) = $\beta_0 + \beta_1$(Observations(i)) + ϵ(i); * indicates statistical significance at a 5% level.

Parameter	OLS^+ Estimate	$T-$ratio (H_0: β_j=0)	Trimmed^{++} OLS Estimate	$T-$ratio (H_0: β_j=0)
Intercept β_0	-0.2271	-2.198*	-0.0469	-0.280
Slope β_1	0.0013	2.748*	0.0011	1.982*

Note: + these estimates are based on n=89, $F-$value=7.553, R^2=0.0693;
++ trimmed 4 outliers, n=57 (N >200), R^2=0.0489.

Table 4: Slope differences, number of observations, and firm assets; Differences(i) = β_0 + β_1(Observations(i)) + β_2(Firm Assets (i)) + ϵ(i).

Parameter	Estimate	$T-$ratio (H_0: β_j=0)
Intercept β_0	-0.1771	0.0935
Observations β_1	0.0011	2.2450*
Firm Assets β_2	0.0000005	0.0900

Note: n=87, $F-$value=2.655, R^2=0.0595.

Table 5: Calibration of LAV and OLS slope; LAV Slope(i) = $\beta_0 + \beta_1$ OLS Slope(i) + β_2(Observations(i)) + ϵ(i); * indicates statistical significant at a 5% level.

Parameter	OLS^+ Estimate	T−ratio (H_0: β_j=0)	LAV Estimate	Trimmed^{++} OLS Estimates
Intercept β_0	0.43221	3.387*	0.07153	0.06188
OLS Slope β_1	0.81975	11.782*	0.89832*	0.89292*
Cases β_2	-0.00131	-2.829*	-0.00014	-0.00007
Sample size	89		89	85

Note: + these estimates are based on n=89, R^2=0.6217;
++ trimmed 4 outliers, n=85, R^2=0.9085.

and LAV-estimated market model slopes (betas). There are several possible explanations for this result. One is related to asymptotic sampling behavior of the estimates. The larger the samples the more likely differences in distributional factor can be detected, hence differences between conditional mean and median estimates become larger. An alternative explanation is that the sample size is a surrogate for other important factors related to risk, age or maturity of the company for example.

Table 3 also indicates that the dependent variable used in the regression, number of observations, accounts for a very small portion of the variation in the differences between the slopes. The use of the particular dependent variable reflected only indirectly on specific firm characteristics. The availability of a long time series may be viewed as evidence for firm maturity. There is justification, then, for inclusion of explanatory variables explicitly reflecting specific firm characteristics. One such variable is the size of the firm. Empirical studies have indeed attempted to explain the higher average returns accruing to small listed firms. For instance, according to Roll (1981), this difference in average returns may be due to improper measurement of these returns. The hypothesis of this study is that differences in the estimated returns is explained by the use of the OLS approach. To test this hypothesis, a size variable is included in the regression. The inclusion of firm total assets results in the estimations presented in Table 4. According to the table, however, the size variable employed does not contribute to the explanation of the slope differential.

The final analysis of data attempts to answer the question: is there a difference between OLS and LAV estimates of beta coefficients-market risk? Table 5 reports the results of a model that relates the *LAV* estimate to both the *OLS* estimate and the sample size. Straight *OLS* regression analysis of this relationship does not take into account the outliers present in the conditional error distributions, consequently both a "trimmed" *OLS* and a *LAV* estimation were also done for this proposed relationship. The results indicate that there is no fixed bias or effect of sample size. Essentially the results indicate that the *LAV* estimate is 89% of the *OLS* estimate. Testing the hypothesis that the *OLS* and *LAV* are equal (*i.e.* $\beta_1=1$) results in rejection at the 1% level. *LAV* estimates systematically reflect less market risk than do *OLS* estimates.

4 CONCLUSIONS

The theoretical and practical importance of beta coefficients necessitate the deployment of the most appropriate estimation techniques. To estimate betas, the use of *OLS* has prevailed by default. The results of this study revealed that there are small, but noteworthy, differences between the estimates generated by *OLS* and those associated with *LAV*. One qualification of the study may arise from the use of individual security, rather than portfolio, beta estimates. It should be emphasized, however, that the interest of this study is in the differences between *OLS* and *LAV* estimates, and not in the characteristics of each estimate. According to the approach followed in this study, statistically significant differences in the slope estimates generated by *OLS* and *LAV* can be found. The nature and significance of these differences varies with the number of observations available for each firm. This could be construed as an "interval" effect or even as a "firm maturity" effect. The association between firm size and slope differential, however, was not verified. An interesting follow-up of this paper, would be to investigate the difference between *LAV* and *OLS* beta estimates, given a certain portfolio optimization model. For the Tokyo stock market, such an investigation has recently been carried out by Konno and Yamazaki (1991) employing the classical Markowitz's model.

REFERENCES

Akgiray, V., and Booth, G.G. (1988). The stable-law model of stock returns. *Journal of Business and Economic Statistics* 6: 51-57.

Alexander, G., and Chervany, N.L. (1980). On the Estimation and Stability of Beta. *Journal of Financial and Quantitative Analysis* 15: 123-38.

Barrodale, I., and Roberts, F. (1978). An Efficient Algorithm for Discrete L1 Linear Approximation with Constant Constraints. *SIAM Journal of Numerical Analysis* 15: 603-611.

Bassett, G. Jr., and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Association* 73: 618-622.

Bloomfield, P., and Steiger, W.L. (1983). *Least Absolute Deviations Theory, Applications.* Birkhäuser, Boston, MA.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52: 5-59.

Charnes, A., Cooper, W.W., and Ferguson, R.O. (1955). Optimal Estimation of Executive Compensation by Linear Programming. *Management Science* 1: 138-151.

Dielman, T., and Pfaffenberger, R. (1982). *LAV* (Least Absolute Value) Estimation in Linear Regression: A Review. In *TIMS Studies in the Management Sciences* 19: 31-52.

Evans, J., and Archer, S. (1968). Diversification and the Reduction of Dispersion: An Empirical Analysis. *Journal of Finance* 23: 761-767.

Fama, E.F. (1965a). Portfolio Analysis in a Stable Paretian Market. *Management Science* 11: 404-419.

Fama, E.F. (1965b). The Behavior of Stock Market Prices. *Journal of Business* 38: 34-105.

Fama, E.F., and Roll, R. (1968). Some Properties of Systematic Stable Distributions. *Journal of the American Statistical Association* 66: 817-36.

Fama, E.F., and Roll, R. (1971). Parameter Estimates for Symmetric Stable Distributions. *Journal of the American Statistical Association* 68: 331-38.

Huber, P.J. (1977). *Robust Statistical Procedures.* SIAM, Philadelphia.

Kon, S. (1984). Model of Stock Returns - A Comparison. *Journal of Finance* 34: 147-65.

Konno, H., and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science* 37: 519-531.

Nieuwland, F.G.M.C. (1993). *Speculative Market Dynamics.* Ph.D. University of Limburg, The Netherlands.

Roll, R. (1981). A Possible Explanation of the Small Firm Effect. *Journal of Finance* 36: 879-888.

Rosenberg, B., and Carlson, D. (1977). A Simple Approximation of the Sampling Distribution of Least Absolute Residuals in Regression Estimates. *Communications in Statistics* B6: 421-38.

Ontvangen: 2-8-1993
Geaccepteerd: 31-3-1994