

# REGRESSION EQUATION FITTING AS AN APPROACH TO MODELLING FINANCIAL DATA

Carlos N. Bouza Herrera and Sira Allende Alonso, Universidad de La Habana  
Daniel C. Chen, Smith and King College  
Josefina Martinez Barbeito, Universidade La Coruña

## ABSTRACT

Many financial models deal with concepts that are linked to regression. The unpopularity of its use in application is due to the fact that the residuals distribution is not normal. A key example is that of the study of the risk-adjusted return of the portfolio. The general equation is regarded as

$$r - R_f = \alpha + \beta (K_m - R_f) \quad (1)$$

Where  $r$  is the fund's return rate,  $R_f$  is the risk-free return rate, and  $K_m$  is the return of the index. This can be regarded as the usual equation for CAPM excepting the existence of  $\alpha$ .  $\beta$  is the 'beta' derived from the classic Sharpe's representation of equilibrium prices. When fitting a regression we include an error term  $\varepsilon$  and  $\alpha$  represents how much better the fund did than the predicted CAPM. We revise this problem considering that the residuals are distributed according to Stable distributions, not necessarily a normal. Some related financial problems are considered in a similar fashion. Monte Carlo experiments are developed for comparing different methods for estimating the so-called beta-coefficients.

**Key words:** CAPM, robust regression, beta, Monte Carlo experiments, outliers.

MSC: 62P05.

## RESUMEN

Muchos modelos financieros se ocupan de conceptos asociados a la regresión. La impopularidad de su uso en las aplicaciones se debe al hecho de que los residuos tienen una distribución no normal. Un ejemplo clave es el estudio del retorno de un portafolio ajustado al riesgo. La ecuación general es planteada como:

$$r - R_f = \alpha + \beta (K_m - R_f) \quad (1)$$

Donde  $r$  es la tasa de retorno del fondo,  $R_f$  es la tasa de retorno libre de riesgo, y  $K_m$  es el índice de retorno. Esto puede considerarse es la usual ecuación para el CAPM exceptuando la existencia de  $\alpha$ .  $\beta$  es el 'beta' derivado de la representación clásica de Sharpe para los precios de equilibrio. Al ajustar la regresión incluimos un término de error  $\varepsilon$  y  $\alpha$  representa cuan mejor está el fondo que el predicho por el CAPM. Revisamos este problema considerando que los residuos se distribuyen de acuerdo a una distribución estable, no necesariamente una normal. Algunos problemas financieros relacionados son considerados en forma similar. Experimentos de Monte Carlo se desarrollan para comparar diferentes métodos para estimar los llamados beta-coeficientes.

## 1. INTRODUCTION

It is evident that financial data is generally generated by a heavy-tailed distribution or by a mixture of distribution functions. Give a look to the graphs of common series of equity returns data and you will note the no normality of the series. That equity plays a crucial role in finance is well known, see Carleton (1985) for example, and its behaviour is typical for financial data.

Portfolio theory gives models for analysing the risk present in stocks by incorporating it into an assessment of securities. It is based in the following assumptions

1. Investors are risk averse (i.e. less risk is preferred to more)
2. Investors seek to maximise their wealth (i.e. more wealth is preferred to less)
3. The wealth maximisation is represented by an expected rate of return
4. The expected rate of return can be first estimated using past rates of return.

There are two types of risk: unsystematic and systematic. The unsystematic risk can be eliminated by the diversification the portfolio. The systematic risk is undiversificable. For any particular security, the unsystematic component is greater than the systematic one. But, when a security is combined with other securities, the unsystematic risk of each is offset, as long as the securities are not correlated by the other securities in the

portfolio. In fact, in a fully diversified portfolio, no unsystematic risk exists. Therefore it is only appropriate to measure the risk of a security by its systematic risk. That form of risk is indeed the only one that counts because securities would never be held in isolation. The name given to systematic risk is beta, and can be described as the slope of a linear regression equation.

Estimates of market risk premium are given by the historical average return on stocks over and above the government security return. A similar market risk premium is available for other securities than stocks, such as bonds. This allows to generalize the analysis to all markets. The adjustment process occurs through a change in price. The analysis of it and the portfolio theory which leads to it, is generally known as the capital assets pricing model (CAPM).

When we look to the literature on CAPM and the empirical evidence it is clear that the correct approach to estimating the regression parameters is to use a robust point of view. See for example Mandelbrot (1963), Fama (1965), Blume (1975), Bruner, Eades, Harris, Higgins (1996), Sharpe (1971). Cornell and Dietrich (1978) and Martin (1998). The use of L1-norm estimator (LAD) may be considered as a more robust alternative than L2 or Least Squares estimation (LS). This fact may be not valid because LAD performs optimally only when the errors are Laplacian random variables.

In this paper we are going to compare some robust estimators with the usual LS estimation procedure. LS estimates will be compared with LAD and with some robust estimators of beta. The expected result is to establish how some estimators behave, when heavy tailed distributions describe the existent relationships in financial data. We will analyse some series of monthly equity returns data and compare the results through Monte Carlo experiment. They should show how the estimates differ significantly, establishing that the differences are significant not only from the point of view of statistics but also economically. Original data will be used as a basis for generating equivalent series using heavy tailed distributions. The results corroborate the usual statistical recommendation to use a robust estimate as an alternative to LS.

## 2. FINANCIAL DATA AND THE USE OF REGRESSION

The estimation of the slope coefficient is commonly used in the study of different classic market model. See for example, Blume (1975), Bruner, Eades, Harris, Higgins (1996) and Martin (1998). who documented the popularity of beta for measuring market risk and return. The practice establishes that the Capital Asset Price Model (CAPM) leads the task of estimating the cost of equity and in providing information in economical publications. LS is commonly used for calculating beta in

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

Where,  $R_{i,t}$  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.

To use least squares estimator is the common solution suggested in standard econometric books, see Judge **et al.** (1988) for example. Statistical modeling bases the inferences in the normality of the residuals, see Cook-Weisberg (1982) for example, and Gauss Markov Theorem justifies the optimality of LS estimation procedure by proving that it computes optimal estimates of the linear regression model coefficients. Usually, an analysis of large databases on equity returns establishes that the normality is not acceptable, because the empirical frequency distribution is markedly leptokurtotic and skewed. This fact establishes the existence of a considerable number of outliers. The presence of outliers influences negatively in the proclaimed properties of LS estimates. Mandelbrot (1963) developed a large study of market data and his empirical investigations evidenced that equity returns are not generated by a Gaussian. Furthermore was argued that it was more credible to assume a heavy-tailed,  $t$ -distribution or a Gaussian mixture distributions as the distribution of the errors. Fama (1965) provided more empirical evidence. He also suggested modeling the distribution of the errors using a stable distribution. Other authors developed empirical studies of large series of financial data and suggested that, a normal mixture distribution permitted to model the existence of outliers in financial returns. See Kon (1984), Roll (1988), Connolly (1989), and Smith (1994).

As an alternative to LS in economics and finance studies of regression the model the more ancient alternative is the *least absolute deviations* (LAD). It is perhaps the oldest and most widely known alternative to least squares. The existence of a large number of outliers suggests that the normality is not longer valid. This evidence may motivate to consider that LAD procedure should be used to produce estimates of the beta. In the study of stocks it has been used for calculating the Dow Jones Industrial Average, see the reported

experiences of Sharpe (1971). Cornell and Dietrich (1978) used LAD in the study of 100 companies selected by a random sampling design from the S&P 500 from 1962 to 1975. Unfortunately Sharpe (1971) and Cornell and Dietrich (1978) did not obtain that to use LAD improves the performance of LS in a conclusive manner. Theoretically this fact is easily explained because LAD is also affected when the underlying distribution is not the assumed as optimal by the theory. For LAD the unacceptability that a Laplace (Double Exponential) as the distribution of the residuals has in the estimation of beta a similar effect as non Gaussianity (normality) in LS. From a practical point of view the explanation is in the lack of influential outliers in the returns of the relatively large sized firms and mutual funds considered.

Koenker and Basset (1978) proposed to use a class of estimators based on regression-quantiles. They were connected with the theory on L- estimates and the robustness, Hubert (1981). The proposal was the so-called regression-quantile L-estimates. Its theoretical properties are supported by the general theory on L-estimation. Connolly (1989) provided elements on how robustness worked when L or M-estimates with respect to LS. The practical result for finance data was to establish that the weekend effect was not as large as supposed when LS is used. Chan and Lakonishok (1992) continued this line of analysis in his study of beta and derived that the L-regression quantile method was considerably more efficient than LS. Mackinlay and Richardson (1991) considered that the returns were generated by a leptokurtotic multivariate T-student distribution. The unconditional CAPM beta was considered and a generalized method of moments was used for computing the bias. Small deviations increased the amplitude of the calculated multivariate normality rejection regions. Richardson and Smith (1994) reproduced the experience for stock prices evaluated the effect of modeling them by a mixture of normal distributions model.

Equity returns on firm size were analyzed by Knez and Ready (1997) using least trimmed squares (LTS) regression method, due to Rousseeuw's (1984). They provided elements for considering that LS estimators computed negative predictions of the risk premium on size, as reported by Fama and French (1992), are caused by the existence of not more than a 1% of outliers in the data. They generally corresponded to exceptionally large returns for small sized firms. Eliminating these outliers the correlation was positive when the returns and firm size were considered.

[Eldletrabajo] analyzed these problem and conjectured that the joint distribution of equity returns and index returns, should be modeled by a mixture of distributions. The dominant should be a normal central component and the outlier-generating component another one. Allende-Bouza (1995) proposed as a solution for these problems a parametric programming approach.

The empirical evidence, accumulated by several economic studies, suggests that LS is not a good estimation procedure for fitting beta the study of CAPM due to the intrinsic non-normality of the financial data.

### 3. REGRESSION EQUATION FITTING AS AN APPROACH TO MODELING CAPM

A line can represent the security market. It models the relationship between return and risk measured by beta for all stocks. The return is on the vertical axis and the risk on the horizontal. The risk free rate is at point on the return axis. If a certain security is off the line, market forces in the form of trading decisions by investors, would push it back on to the line. A security located above the line means that its return is higher than the risk adjusted return of securities with the same amount of risk. Hence the investor must be interested in acquiring this stock. This interest generates a rise of its price and a diminish of its return, thus placing it back on the line. If the security is below the line, the same reasoning will produce the opposite results in the investor and the stock: a lower risk adjusted return will generate the selling the stock, as a consequence cause its return rises and put it back on the line.

One of the uses of CAPM is to analyze the performance of mutual funds and other portfolios. It looks for comparing historical risk-adjusted returns of the fund with those of an appropriate index. See for example Blume (1971) for its use in risk assessment. The usual approach is to use ordinary least squares to determine a straight line.

Each data point is the risk-adjusted return of the portfolio and the index over one time period in the past. The general equation is regarded in common economic models as

$$R - R_f = \alpha + \beta (K_m - R_f) \tag{1}$$

Where  $r$  is the fund's return rate,  $R_f$  is the risk-free return rate, and  $K_m$  is the return of the index. This can be regarded as the usual equation for CAPM excepting the existence of  $\alpha$ . Then  $\beta$  is the " derived from the

classic Sharpe's representation of equilibrium prices. When fitting a regression we include an error term  $\varepsilon$  and  $\alpha$  represents how much better the fund did than the predicted CAPM. The quality of the fit may be analyzed using different statistical techniques.

CAPM uses a single factor, beta, to compare a portfolio with the market as a whole. Another approach was proposed by Fama et. al. (1996) who considered that the observation of two classes of stocks performs better than the observation of the whole market. They added two more factors to CAPM to reflect the portfolio's exposure to these two classes: The model is

$$r - R_f = \alpha + B_3 (K_m - R_f) + B_s \text{ SMB} + B_v \text{ HML}$$

Here  $r$  is the portfolio's return rate,  $R_f$  is the risk-free return rate, and  $K_m$  is the return of the whole stock market. SMB and HML stand for "small [cap] minus big" and "high [book/price] minus low". They measure the historic excess returns of small caps and "value" stocks over the market as a whole. This model is called the "three factor" beta and is similar to the classical beta. Once SMB and HML are defined, the corresponding coefficients  $B_s$  and  $B_v$  take values on a scale in the interval  $[0,1]$ : Note that  $B_s = 1$  in the case of a small cap portfolio and  $B_s = 0$  for large cap. Similarly we can establish that  $B_v = 1$  for a portfolio with a high book/price ratio.

One of the variations of CAPM allows several different types of systematic risk to be affecting a security market line, such as interest rate risk, purchasing power risk and economic risk, instead of just market risk used above. This version of CAPM is known as arbitrage pricing theory (APT) and is a more general formulation of the same risk adjustment process but involves a series of equations instead of just one.

The validity of CAPM is essential for testing if prices adjust as well as it should be theoretically. The tests are conducted using the hypothesis that we are working within the framework of the efficient market (prices reflect all the relevant information). The tests show how the prices react. The efficient market hypothesis has always been rejected by securities industry practitioners and recent findings bring in even some doubts on its validity. The verification of CAPM would be an actual lining up of risk - return values along the security market line. Clearly in practice the relationship is not a line. It is an ideal representation. Then we are not able to estimate the true expected return using the model.

The investors should rely on prior data. They want to use the models of CAPM for investing successfully. Hence they want to make predictions and look for the behaviour of their decisions in the future. Then they look at the data and wants to estimate (predict) the parameters involved in the description of CAPM. Using regression theory we have that:

$$R_p - R_f = \alpha + B (R_m - R_f) + \varepsilon$$

That is

$$Y = \alpha + BX + \varepsilon$$

Where

$Y$  = *portfolio return* - Market or benchmark return

$\alpha$  = risk adjusted excess return

$B$  = market risk

$X$  = Risk-free proxy return- Market or benchmark return

$\varepsilon$  = Residual risk or regression error

When the CAPM reflects a portfolio's exposure to the two classes the regression model is:

$$r - R_f = \alpha + B_3 (K_m - R_f) + B_s \text{ SMB} + B_v \text{ HML} + \varepsilon$$

Using the standard regression notation it is

$$Y = \alpha + B_3 X_3 + \beta_2 X_2 + B_3 X_3 + \varepsilon$$

The efficient estimation problem of beta is a significant limitation of the theory. Though ordinary least squares (OLS) regression is widely used to create software prediction models, and it seems to perform just as well or better than most other, non-regression, prediction models. That is the reason of a questioning LS as a good measure, see Eubank-Zumwalt, (1979) Elton (1978) and Klemkosky-Martin, (1975) for a discussion of this theme. Software data sets may however exhibit certain characteristics that do not always comply with the requirements of OLS. Generally the distribution of residuals exhibit heavy tails, high kurtosis, and/or skewness. Under these circumstances, OLS regression may be less efficient than certain, see Hubert (1973) methods.

Andrews **et al.** (1972) provides a long discussion on the possibilities of robust methods.

The relaxing of the hypothesis with respect to the disturbances leads to symmetric  $\alpha$ -stable distribution.

Financial data are related with high volatile and low volatile phases. These facts, known as volatile clustering, generates unconditional high-peaked distributions around zero.

There is a need to use  $\alpha$ -stable distributions to model effectively CAPM.

## 2. STABLE DISTRIBUTIONS

Stable laws were introduced by Lévy, P. (1925) as a result of the investigation of the probabilistic behaviour of the sums of independent random variables. These distributions are usually called  $\alpha$ -stable. The sums of two variables with index  $\alpha$  is also  $\alpha$ -stable distributed with the same index. If the index varies the invariance is not longer valid.

Financial asset returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time. The effect is that the data of financial asset returns usually have heavy tails. As the assumption of normality is not acceptable the use of stable distributions usually provides the best approach because:

1. The Generalized Central Limit Theorem states that stable laws are the only possible limit distributions for properly normalized and centered sums of independent, identically distributed random variables, see Laha and Rohatgi. (1979).
2. Stable distributions are leptokurtic and they can model the existence of heavy tails and asymmetry.

An  $\alpha$ -stable distribution depends of four parameters:

- The index of stability (tail index, tail exponent or characteristic exponent)  $\alpha \in ]0,2]$ , which determines the rate at which the tails of the distribution taper off.
- The skewness parameter  $\beta \in [-1,1]$ . If  $\beta$  is positive, the distribution is skewed to the right say that the right tail is thicker. When  $\beta$  is negative, it is skewed to the left and  $\beta = 0$  determines that the distribution is symmetric about  $\mu$ .
- The scale parameter  $\sigma > 0$  which determines the width.
- The location parameter  $\mu \in \Re$  that represents the shift of the mode (the peak) of the distribution.

$\alpha$  is the main parameter of a stable distribution because analyzing it are fixed the following properties of the distribution?

- It is a Gaussian if  $\alpha = 2$ .
- For  $\alpha < 2$  the variance is infinite. As a result the tails are asymptotically equivalent to a Pareto law.
- $E(X^p) = \mu_p$  exists only if  $p < \alpha$ . Then for  $\alpha > 1$  is granted the existence of  $E(X) = \mu$ .
- As  $\alpha$  approaches to 2 the distribution tends to a Gaussian for any  $\beta$ .

The stable distributions have not an explicit representation excepting some particular cases: is that, with the exception of three special cases, its probability density function (PDF) and cumulative distribution function (CDF) do not have closed form expressions. These exceptions include the well known

- The Gaussian ( $\alpha = 2$ )
- The Cauchy ( $\alpha = 1$  and  $\beta = 0$ )
- The Lévy ( $\alpha = 0,5$  and  $\beta = 1$ )

They are generally described by its characteristic function  $\phi(t)$ . Different representations are given but the most popular is obtained by parametrizing the characteristic function of  $X$  with distributing according to  $S_\alpha(\sigma, \beta, \mu)$ , see for a detailed discussion Samorodnitsky and Taqqu (1994) and Weron (1996). The use of the representation  $S_\alpha(\sigma, \beta, \mu)$  yields as characteristic function:

$$\log \phi(t) = \begin{cases} -\sigma^\alpha |t| \left[ 1 - i\beta \text{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right) \right] + i(\mu t) & \text{if } \alpha \neq 1 \\ -\sigma^\alpha |t| \left[ 1 + i\frac{2\beta}{\pi} \log |t| \text{sign}(t) \right] + i(\mu t) & \text{if } \alpha = 1 \end{cases}$$

$\sigma$  is not the standard deviation as in the case of the Gaussian distribution but a scale parameter that equals to  $\sigma_{\text{stdev}} = \sigma\sqrt{2}$

A popularly used formula is , see Nolan, J.P. (1997).

$$\log \phi_0(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \left[ 1 + i\beta \text{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right) \left[ (\sigma|t|)^{1-\alpha} - 1 \right] \right] + i(\mu_0 t) & \text{if } \alpha \neq 1 \\ -\sigma |t| \left[ 1 + i\frac{2\beta}{\pi} \log(\sigma|t|) \text{sign}(t) \right] + i(\mu_0 t) & \text{if } \alpha = 1 \end{cases}$$

li is derived from the representation  $S_\alpha^0(\sigma, \beta, \mu)$ .

The quantiles and convergence to the power-law tail vary in a continuous way as function of  $\alpha$  and  $\beta$ . The location parameters of the two representations are related by

$$\mu = \begin{cases} \mu_0 - \beta\sigma \tan(\pi\alpha/2) & \text{if } \alpha \neq 1 \\ \mu_0 - \beta\sigma \tan(2/\pi) & \text{if } \alpha = 1 \end{cases}$$

Zolotarev, V.M. (1986) proposed a parametrization taking  $\zeta = -\beta \tan(\pi\alpha/2)$  then we have that  $S^0 = S_\alpha^0(1, \beta, 0)$ .

In this case the density function can be expressed by:

- If  $x > \zeta$  and  $\alpha \neq 1$ .

$$f(x | \alpha, \beta) = \frac{\alpha(x - \zeta)^{\frac{1}{\alpha-1}}}{\pi|\alpha - 1|} \int_{-\theta_0}^{\frac{\pi}{2}} V(\theta | \alpha, \beta) \exp\left\{ -(x - \zeta)^{\frac{\alpha}{\alpha-1}} V(\theta | \alpha, \beta) \right\} d\theta$$

- If  $x = \zeta$  and  $\alpha \neq 1$ .

$$f_x(\alpha, \beta) = \frac{\Gamma\left(\frac{\alpha+1}{\alpha}\right)}{\pi(1 + \zeta^2)^{\frac{1}{2\alpha}}} \cos(\xi)$$

- If  $x > \zeta$  and  $\alpha \neq 1$ .

- $f(x|\alpha, \beta) = f(-x|\alpha, -\beta)$

If  $\alpha = 1$ .

$$f(x|1, \beta) = \begin{cases} \frac{\exp(\pi\xi/2\beta)}{2|\beta|} \int_{-\pi/2}^{\pi/2} V(\theta|1, \beta) \exp[\exp(\pi\xi/2\beta)V(\theta|1, \beta)d\theta] & \text{if } \beta \neq 0 \\ \frac{1}{\pi(1+x^2)} & \text{if } \beta = 0 \end{cases}$$

Taking

$$\xi = \begin{cases} \frac{\arctan(-\zeta)}{\alpha} & \text{if } \alpha \neq 1 \\ \frac{\pi}{2} & \text{if } \alpha = 1 \end{cases}$$

and

$$V(\theta|\alpha, \beta) = \begin{cases} \left( \frac{\cos(\alpha\xi)}{\cos(\theta)} \right)^{\frac{1}{\alpha+1}} \left( \frac{\cos(\theta)}{\sin(\alpha[\xi + \theta])} \right) \frac{\cos[\alpha\xi + (\alpha-1)\theta]}{\cos(\theta)} & \text{if } \alpha \neq 1 \quad \frac{\alpha}{\alpha+1} \\ \frac{2}{\pi} \left( \frac{\frac{\pi}{2} + \beta\theta}{\cos(\theta)} \right) \exp \left[ \frac{1}{\beta} \left( \frac{\pi}{2} + \beta\theta \right) \tan(\theta) \right] & \text{if } \alpha = 1 \text{ and } \beta \neq 0 \end{cases}$$

As stable densities and distribution functions do not have necessarily a closed form formulas we need to use either numerical approximation or direct numerical integration. Then for dealing with stable distributions we face the need of using computational devices, with an increase of computational costs, for obtaining approximate solutions, with an increase in inaccuracy. The importance of this problematic is the leit motiv of numerous papers looking for good and uncostly approximations. We can mention the Tables developed by Holt, D.R. and Crow, E.L. (1973), the formulas given by Borak, Sz., Härdle, W. and Weron, R. (2004), the algorithm proposed by Mitnik, S., Doganoglu, T. and Chenyao, D. (1999) and Nolan, J.P. (1997, 1999).

The commercial softwares systems do not provide adequate package for dealing with stable distributions. The needed computations can be performed through the use of libraries provided by specialists, see for example Würtz's collection of software packages for S-plus/R, see Würtz (2004). For dealing with commercial risk management, solutions for derivatives pricing and portfolio optimization based on the assumption of stably distributed returns etc., is provided in <http://www.finanalytica.com>.

#### 4. ESTIMATION OF THE REGRESSION PARAMETERS

The estimation of stable law parameters of stable distributions is complicated for the vast majority of them because they have not a closed-form density. Then numerical approximation or direct numerical integration is needed. The first method that seems to be usable when the statistician's toolbox is consulted is a maximum likelihood (ML) estimation algorithm. It will rely on these approximations but to implementing it is time consuming for financial applications.

Take a sample  $x = (x_1, \dots, x_n)$  of independent and identically distributed (i.i.d.) selected from  $S_\alpha(\sigma, \beta, \mu)$  and the estimates  $a, s, b, y, m$  of the four parameters of this law parameters.

The quantiles estimators provide adequate estimates if the law is symmetric, that is if  $\beta = \mu = 0$  when  $\alpha > 1$ . If  $x_p$  is the estimate of the  $p$ -th quantile an adequate estimate of  $\sigma$  is

$$s = \frac{X_{0.72} - X_{0.28}}{1.654}$$

Fama and Roll (1971) based their estimations of  $\sigma$  in that estimator

Continuing with the suggestions of Fama-Roll (1971) the characteristic exponent  $\alpha$  can be estimated from the tail behaviour of the distribution using  $p = 0.95, 0.96$  or  $0.97$  as the best suited values

$$S_a\left(\frac{X_p - X_{1-p}}{2s}\right) = p$$

For  $\alpha \in ]1, 2[$  a stable distribution has finite mean  $\mu$  and the sample mean is a consistent estimate of it. The truncated sample mean for  $p=0.50$  is more robust and must be used when the range of  $\alpha$  is unknown.

When the OLS method is used for estimating the regression coefficient is given by the computation of the Maximum Likelihood estimator

$$B_{ols} = [X^T X]^{-1} X Y$$

Where  $Y$  is the  $n$ -dimensional vector of response variables and  $X$   $n \times k$  matrix of the control variables.

The broadest approach to the solution of the problematic associated to the non optimality of OLS is to use robust estimators but the most popular is to use LAD. OLS estimation is based on the use of L2 norm which leads to the Quadratic Program

$$\min \sum_{i=1}^n \left( Y_i - \sum_{j=1}^K B_j X_{ij} \right)^2 = \sum_{i=1}^n \varepsilon_i^2$$

Its solution is the  $B_{ols}$ .

The use of Least Absolute Deviation (L1 norm) is based on the Manhattan (City Block) distance and is the solution of the Linear Programming model, see

$$B_{LAD} = \text{Argmin} \sum_{i=1}^n \left| Y_i - \sum_{j=1}^K B_j X_{ij} \right| = \sum_{i=1}^n |\varepsilon_i|$$

subject to

$$- \varepsilon_i \leq Y_i - \sum_{j=1}^K B_j X_{ij} \leq \varepsilon_i, \quad i = 1, \dots, n$$

The roots of robust regression are in the paper of Hubert (1973). The main idea is that a function of the residuals should be minimized. The use of robust procedures for estimating beta is documented in Martin. Simin (1999): The large residuals must have a smaller weight in the calculation of the estimates. The basic model is to define an adequate symmetric function  $\rho(\bullet)$  which settles the basic optimization problem

$$\min \sum_{i=1}^n \rho \left( \frac{Y_i - \sum_{j=1}^K B_j X_{ij}}{\sigma} \right) = \sum_{i=1}^n \rho(\varepsilon_i)$$

Where  $\sigma$  is a scale parameter  $\rho(e)$  has a minimum at 0 and generally  $\sigma$  is unknown. A solution is to use as estimation the absolute deviation with respect to the median of the residuals  $M(\varepsilon)$  given by

$$\sigma_M = 1.4825 M[|\varepsilon_i - M(\varepsilon_i)|]$$

Note that if we use OLS  $\rho(\varepsilon_i) = \varepsilon_i^2$  and if LAD  $\rho(\varepsilon_i) = |\varepsilon_i|$ .

The proposal of Huber (1973) was to use



$$\rho(e) = \begin{cases} \frac{e^2}{2} & \text{if } |e| \leq c \\ c|e| - \frac{c^2}{2} & \text{if } |e| > c \end{cases}$$

The parameter  $c$  is defined has the role of tuning the efficiency level of the residuals. A common selection is to use  $c = 1.345$  for achieving an efficiency of 95%. This is called Hubert's M estimator. It is very popular in robust regression applications , see for example Martin and Zamar, (1989). We will denote it by BM.

Deriving we have a system of  $K$  non linear equations

$$\sum_{i=1}^n \frac{\partial \left( \rho \left( \frac{Y_i - \sum_{j=1}^K B_j X_{ij}}{\sigma} \right) \right)}{\partial B_j} = \sum_{i=1}^n \psi \left( \frac{Y_i - \sum_{j=1}^K B_j X_{ij}}{\sigma} \right) (X_{ij}) = 0 \quad j = 1, \dots, K$$

Different proposals have been developed in a series of papers. We will consider some that have exhibited a good behaviour in dealing with regression fitting with heavy tailed distributed residuals.

One of them is the proposal of Andrews **et al.** (1972)

$$\psi(e) = \begin{cases} c \sin\left(\frac{e}{c}\right) & \text{if } |e| < c\pi \\ 0 & \text{otherwise} \end{cases}$$

Generally is used  $c = 1.339$ . We will denote it by  $B_A$ .

The Biweight uses as default parameter  $c = 4.685$  and is defined as

$$\psi(e) = e(1 - [e/c^2]^2) \text{ if } |e| \leq c$$

Cauchy weight function is

$$\psi(e) = e/(1 - [e/c^2]) \text{ for any } e$$

The usually the value of the tuning parameter is  $c=2.385$ . A similar proposal is the Fair function where  $c = 1.4$  is used broadly and

$$\psi(e) = e/(1 - |e|/c) \text{ for any } e$$

The last weight function taken into account is the Logistic where  $c = 1.205$  is the default value and

$$\psi(e) = c[\tanh(e/c)] \text{ for any } e.$$

The calculation of the estimators based on this weight functions area available in the software provided by S-Plus and by R. We will denote them by BI, where I = Biweight, Cauchy, Fair and Logistic.

The behaviour of robust methods is very important in issues such as outlier detection see Rousseeuw **et al.** (1987). It is expected that they reduce the percent of outliers.

## 5. COMPARISON OF DIFFERENT ESTIMATORS

We will consider some of the series of returns used by Martin-Simin (1999). The data were collected during the period 1962-1996 from the NYSE, AMEX, NASDAQ of the firms A W Computer Systems Inc, Oil City Petroleum Inc. and Biopharmaceutics Inc. They are the excess of the one-month nominal rate for the Treasury bill closest to 30 days to maturity and were reported in the CRSP's files. We used their means and standard

deviations for generating new series using a Gaussian, a Cauchy and a Lévy. They were considered as the location and scale parameters and are given in Table 1.

**Table 1.** Statistics Summary for Equity Returns for the Firms in the Experiment.

	A W Computer Systems Inc.	Oil City Petroleum Inc.	Biopharmaceutics Inc.
Mean	0.068	0.017	0.058
Median	- 0.037	- 0.002	- 0.055
Std. Dev.	0.644	0.509	0.645

We generated a series of 5 000 entries and computed the different estimator of beta. They were compared with the estimate computed using LS and

$$\Delta_{ols,l} = \text{Mean of } |B_{OLS} - B_l|$$

was calculated for 10 and 5 years periods. The results are given in Tables 1 and 3.

**Table 2.** Means of the Absolute Differences for the Experiment for a 10 Years Period.

$\Delta_{ols,l} = \text{Means }  B_{OLS} - B_l $									
$B_l$	A W Comp. Systems Inc.			Oil City Petrol. Inc.			Biopharm. Inc.		
	Normal	Cauchy	Lévy	Normal	Cauchy	Lévy	Normal	Cauchy	Lévy
$B_{Lad}$	0.06	0.15	0.11	0.18	0.27	0.25	0.14	0.22	0.19
$B_M$	0.05	0.21	0.34	0.07	0.33	0.21	0.04	0.18	0.24
$B_A$	0.08	0.16	0.27	0.02	0.29	0.33	0.09	0.25	0.25
$B_B$	0.14	0.11	0.33	0.18	0.09	0.08	0.18	0.19	0.27
$B_C$	0.19	0.22	0.24	0.04	0.38	0.41	0.14	0.21	0.29
$B_F$	0.23	0.15	0.28	0.08	0.17	0.30	0.11	0.10	0.17
$B_L$	0.19	0.28	0.26	0.09	0.11	0.19	0.08	0.19	0.35

The results in Table 2 suggest that BM has a good behavior compared with the optimum estimator represented by the OLS in the normal case. LAD behaves similarly to OLS only in the first firm. BA is also a good alternative in the normal case. The differences between the other estimators and OLS should be complemented with the analysis of Table 3 with the percent of outliers generated by the use of each estimator. OLS has a small percent of outliers only when the Gaussian distribution generates the variables. LAD is an adequate substitute of it under the normality assumption and behaves better than OLS when the other distributions generate the errors. BM has similar percent in all the cases varying within the interval [6.21 9.23]. Hence we can consider it as very stable and with rather small percent of outliers. The rest of the robust estimators behave better than OLS and LAD but in some cases the percent of outliers is rather large. It is remarkable that BL is very good when the errors are generated by a Levy distribution.

**Table 3.** Percent of Outliers in the Experiment for a 10 Years Period.

Percent of residuals out of the interval [-3 3]									
$B_l$	A W Comp. Systems Inc.			Oil City Petrol. Inc.			Biopharm. Inc.		
	Normal	Cauchy	Lévy	Normal	Cauchy	Lévy	Normal	Cauchy	Lévy
$B_{OLS}$	3.58	25.64	21.84	2.11	27.88	27.66	4.23	23.64	20.61
$B_{Lad}$	7.56	10.54	16.90	12.66	9.94	10.57	6.58	7.69	9.97

B <sub>M</sub>	6.21	7.95	9.23	5.40	7.51	8.59	7.91	7.09	7.62
B <sub>A</sub>	7.94	10.57	11.59	6.96	9.96	10.44	6.59	9.77	10.62
B <sub>B</sub>	12.0	15.54	8.94	13.66	15.54	8.94	13.51	13.57	8.31
B <sub>C</sub>	10.51	10.82	12.47	13.87	9.97	10.85	12.41	10.02	10.71
B <sub>F</sub>	10.66	9.83	10.20	11.45	9.89	10.35	1.21	9.70	10.67
B <sub>L</sub>	23.54	12.94	4.58	21.51	12.52	4.33	22.44	11.94	5.28

The analysis of the result for the 5 years period yields a similar valorization of the estimators, see Table 4. The percent of outliers is considerably smaller for the Hubert M-estimator than for the 10 years period. BL continues being the best alternative for estimating beta for Levy's distribution.

**Table 4.** Mean of the Absolute Differences for a 5 Years Period.

$\Delta_{ols,i} = \text{Mean} B_{OLS} - B_i $									
B <sub>i</sub>	A W Comp. Systems Inc.			Oil City Petrol. Inc.			Biopharm. Inc.		
	Normal	Cauchy	Lèvy	Normal	Cauchy	Lèvy	Normal	Cauchy	Lèvy
B <sub>Lad</sub>	0.04	0.12	0.19	0.09	0.32	0.21	0.08	0.21	0.14
B <sub>M</sub>	0.03	0.25	0.29	0.09	0.31	0.19	0.06	0.16	0.21
B <sub>A</sub>	0.07	0.38	0.31	0.08	0.27	0.31	0.05	0.25	0.23
B <sub>B</sub>	0.12	0.14	0.34	0.16	0.11	0.07	0.12	0.16	0.20
B <sub>C</sub>	0.15	0.19	0.30	0.09	0.25	0.35	0.16	0.23	0.24
B <sub>F</sub>	0.12	0.11	0.27	0.10	0.14	0.31	0.12	0.12	0.11
B <sub>L</sub>	0.17	0.18	0.31	0.18	0.12	0.21	0.10	0.17	0.29

Then we can recommend the use of BM as the best alternative when we do not know if a Gaussian, a Cauchy or a Levy is generating the errors. The evidence that it is a Levy will make recommendable to use BL for estimating the beta in CAPM.

**Table 5.** Percent of Outliers in the Experiment for a 5 Years Period.

Percent of residuals out of the interval [-3 3]									
B <sub>i</sub>	A W Comp. Systems Inc.			Oil City Petrol. Inc.			Biopharm. Inc.		
	Normal	Cauchy	Lèvy	Normal	Cauchy	Lèvy	Normal	Cauchy	Lèvy
B <sub>OLS</sub>	2.25	22.67	25.11	2.11	27.11	27.66	4.22	21.84	20.54
B <sub>Lad</sub>	6.55	8.23	15.88	12.74	9.47	10.66	1.51	8.33	10.07
B <sub>M</sub>	3.54	4.21	5.48	1.57	5.14	6.41	3.69	2.87	3.45
B <sub>A</sub>	6.51	11.24	10.27	7.95	10.54	10.81	1.95	10.34	10.92
B <sub>B</sub>	12.5	12.27	6.71	13.20	12.90	8.72	12.84	12.46	7.42
B <sub>C</sub>	11.51	10.12	12.47	11.51	9.15	14.08	12.87	11.02	11.44
B <sub>F</sub>	11.64	9.42	11.88	11.51	9.57	10.81	8.50	9.66	10.99
B <sub>L</sub>	25.49	15.94	1.31	20.16	12.71	2.64	22.06	10.39	2.91

## REFERENCES

- ALLENDE, S.M. and C.N. BOUZA (1995): "Multiple linear regression curve fitting: a quadratic Programming solution". In **Approximation & Optimization** (J. Guddat et al. Eds) P. Lang Verlag, Frankfurt. 1-3.
- ANDREWS, D. F. et al. (1972): **Robust Estimates of Location: Survey and Advances**, Princeton University Press, Princeton. N. Jersey.
- BLUME, M. E. (1971): "On the Assessment of Risk," **Journal of Finance**, 26:1 1-10.
- \_\_\_\_\_ (1975): "Betas and Their Regression Tendencies," **Journal of Finance**, 30, 785-795.
- BORAK, S.Z.; W. HÄRDLE and R. WERON (2004): Stable Distributions, in P. Cizek, W. Härdle, R. Weron (eds.) **Statistical Tools for Finance and Insurance**, Springer.
- BRUNER, R.F.; K.M. EADES; R.S. HARRIS and R.C. HIGGINS (1996): "'Best Practices' in Estimating the Cost of Capital: Survey and Synthesis", University of Washington Working Paper.
- CARLETON, W.T. and J. LAKONISHOK (1985): "Risk and Return on: The Use and Misuse of Historical Estimates, **Financial Analysts Journal**, 19, 38-47.
- CHAN, L.K.C. and J. LAKONISHOK (1992): "Robust Measurement of Beta Risk", **Journal of Financial and Quantitative Analysis**, 27:265-282.
- CONNOLLY, R.A. (1989): "An Examination of the Robustness of the Weekend Effect", **Journal of Financial and Quantitative Analysis**, 24:133-169.
- COOK, R.D. and S. WEISBERG (1982): "Residuals and Influence in Regression", Chapman and Hall, New York.
- CORNELL, B. and J.K. DIETRICH (1978): "Mean-Absolute-Deviation Versus Least-Squares Regression Estimation of Beta Coefficients", **Journal of Financial and Quantitative Analysis**, 13: 123-131.
- ELTON, E.J.; M.J. GRUBER and T.J. URICH (1978): "Are Betas Best?" **Journal of Finance**, 33:1375-1384.
- EUBANK, A.A. Jr. and J.K. ZUMWALT (1979): "An Analysis of the Forecast Error Impact of Alternative Beta Adjustment Techniques and Risk Classes," **Journal of Finance**, 34, 761-776.
- FAMA, E. (1965): "The Behavior of Stock Prices", **Journal of Business**, 38:34-105.
- FAMA, E. and KENNETH R. FRENCH (1992): "Cross-section of Expected Stock Returns", **Journal of Finance**, 47:427-466.
- HOLT, D.R. and E.L. CROW (1973): Tables and graphs of the stable probability density functions, **Journal of Research of the National Bureau of Standards B** 77B:143-198.
- HUBER, P.J. (1973) "Robust Regression: Asymptotics, Conjectures and Monte Carlo," **Annals of Statistics**, 1:799-821.
- HUBER, P.J. (1981): "Robust Statistics", Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- JUDGE, G.G. et al. (1988): "Introduction to the Theory and Practice of Econometrics," Second edition. John Wiley & Sons, Inc., New York.

- KLEMKOSKY, R.C. and J.D. MARTIN (1975): "The Adjustment of Beta Forecasts," **Journal of Finance**, 30, 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, 1355-82.
- KOENKER, R. and G. BASSETT (1978): "Regression Quantiles", **Econometrica**, 46:33-50.
- KON, S. (1984): "Models of Stock Returns – A Comparison", **Journal of Finance**, 39 147-165.
- LAHA, R.G. and V.K. ROHATGI (1979): **Probability Theory**, Wiley, New York.
- LEVY, P. (1925): **Calcul des Probabilites**, Gauthier Villars. Paris
- MACKINLAY, A.C. and M.P. RICHARDSON (1991): "Using Generalized Method of Moments to Test Mean-Variance Efficiency", **Journal of Finance**, 2 , 234-247.
- MANDELBROT, B. (1963): "The Variation of Certain Speculative Prices," **Journal of Business**, 36:394-419.
- MARTIN, R.D. (1998): "Trellis Displays for Deep Understanding of Financial Data", **Financial Engineering News**, 2:123-137.
- MARTIN, R.D. and R.H. ZAMAR (1989): "Asymptotically Min-Max Bias Robust M-estimates of Scale for Positive Random Variables", **Journal of American Statistical Assoc.**, 84:494-501.
- MARTIN, R.D. and T. SIMIN (1999): **Robust estimation of beta**. Technical Report no. 350 Department Of Statistics, University of Washington.
- MITTNIK, S.; T. DOGANOGLU and D. CHENYAO (1999): Computing the Probability Density Function of the Stable Paretian Distribution, **Mathematical and Computer Modelling** 29: 235-240
- NOLAN, J.P. (1997): Numerical Calculation of Stable Densities and Distribution Functions, *Communications in Statistics-Stochastic Models* 13: 759-774.
- \_\_\_\_\_ (1999): An Algorithm for Evaluating Stable Densities in Zolotarev's (M) Parametrization, **Mathematical and Computer Modelling** 29:229-233.
- RICHARDSON, M. and T. SMITH (1994): "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information", **Journal of Financial and Quantitative Analysis**, 101-116.
- ROUSSEEUW, P.J.; V.J. YOHAI and ANNICK M. LEROY (1987): "Robust Regression and Outlier Detection", **Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics**. John Wiley & Sons, Inc., New York.
- SAMORODNITSKY, G. and M.S. TAQQU (1994): **Stable Non-Gaussian Random Processes**, Chapman & Hall. New York.
- SHARPE, W.F. (1971): "Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios", **Management Science**, 18:B1-B13.
- VASICEK, O. (1973): "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas," **Journal of Finance**, 28:1233-1239.
- WERON, R. (1996): "Correction to: On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables", **Research Report HSC/96/1**, [www.im.pwr.wroc.pl](http://www.im.pwr.wroc.pl)).
- WIGGINS, J. B. (1992): "Betas in Up and Down Markets", **The Financial Review**, 27:107-123.

WÜRTZ, S. (2004): Nolan's STABLE programs, see <http://www.robustanalysis.com>, collection of software packages for S-plus/R. [see <http://www.itp.phys.ethz.ch>].

ZOLOTAREV, V.M.(1986): One-Dimensional Stable Distributions, American Mathematical Society.  
New York