Working Paper Series | 19 | 2016

A Highly Efficient Regression Estimator for Skewed and/or Heavy-tailed Distributed Errors

This paper introduces a regression model for extreme events that can be useful for financial market analysis and prediction

Lorenzo Ricci
European Stability Mechanism
Vincenzo Verardi
Université Namur - CRED
Catherine Vermandele
Université libre de Bruxelles

European Stability Mechanism

A Highly Efficient Regression Estimator for Skewed and/or Heavy-tailed Distributed Errors

Lorenzo Ricci[1] European Stability Mechanism
Vincenzo Verardi[2] Université Namur - CRED
Catherine Vermandele[3] Université libre de Bruxelles

Abstract

This paper proposes a simple maximum likelihood regression estimator that outperforms Least Squares in terms of efficiency and mean square error for a large number of skewed and/or heavy tailed error distributions.

Keywords: Skewed and heavy tailed regression; Tukey's *g* and *h* distribution; Maximum approximated likelihood estimator
JEL codes: C13, C16, C51, G17

1 Email: L.Ricci@esm.europa.eu

2 Email: vverardi@ulb.ac.be

3 Email: vermande@ulb.ac.be

ISSN 2443-5503
ISBN 978-92-95085-22-0

doi:10.2852/02035
EU catalog number DW-AB-16-004-EN-N

A highly efficient regression estimator for skewed and/or heavy-tailed distributed errors[1]

LORENZO RICCI [2] VINCENZO VERARDI [3] CATHERINE VERMANDELE [4]

Abstract

In this paper, we propose a simple maximum likelihood regression estimator that outperforms Least Squares in terms of efficiency and mean square error for a large number of skewed and/or heavy tailed error distributions.

Keywords: Skewed and heavy tailed regression; Tukey's g and h distribution; Maximum approximated likelihood estimator

JEL classification: C13, C16, C51, G17

1 Introduction

Ordinary Least Squares (LS) is the simplest and most commonly used estimator for linear regression analysis. Under a set of hypotheses, called Gauss-Markov assumptions, this estimator is the most efficient linear unbiased estimator. With heavy-tailed or asymmetrical distribution of the error term, LS is no longer the most efficient estimator and is outperformed by other maximum likelihood estimators when the error distribution is known (or well approximated). How to improve efficiency when the error distribution is not known beforehand is an old debate (Mandelbrot, 1963; Fama, 1965; Rachev, 2003).[1]

In this paper, we propose a simple maximum likelihood regression estimator that outperforms LS in terms of efficiency for a large number of skewed and/or heavy tailed error distributions.

[1]Corresponding address: Lorenzo Ricci, Université libre de Bruxelles, Av. Roosevelt, 50 - CP139, B-1050 Brussels, Belgium E-mail: `lricci@ulb.ac.be` (L. Ricci), `vverardi@ulb.ac.be` (V. Verardi) and `vermande@ulb.ac.be` (C. Vermandele). Vincenzo Verardi is an associate researcher at FNRS and gratefully acknowledges their financial support.

[2]European Stability Mechanism and Université libre de Bruxelles - ECARES

[3]Université Namur - CRED

[4]Université libre de Bruxelles

[1]In this context, outliers are considered a part of the error distribution.

The underlying idea is to estimate the regression parameters by the maximum likelihood method, assuming that the error distribution belongs to the family of the very flexible Tukey distributions which provide a good adjustment of a large number of commonly used unimodal distributions. We perform a Monte Carlo study to assess the performance of this estimator and find that it behaves better than LS in terms of efficiency (and Mean Squared Error) as soon as the error distribution departs from normality. Consequently, this estimator also leads to more precise predictions under these circumstances.

This paper is structured as follows. Section 2 recalls some characteristics of Tukey's distributions, presents the regression model and our estimation procedure. Section 3 is devoted to the Monte Carlo study. Section 4 is dedicated to the empiric evaluation. In the final section (Section 5) the conclusions are drawn.

2 Methodology

2.1 Tukey *g*-and-*h* distribution

In the late 1970s, Tukey (1977) introduced a new family of distributions, called Tukey g-and-h distributions, based on elementary transformations of the standard normal. Let Z be a random variable from the standard normal distribution $N(0,1)$. Define the random variable Y through the transformation

$$Y = \xi + \omega\tau_{g,h}(Z) \tag{1}$$

where $\xi \in \mathbb{R}$, $\omega > 0$, and

$$\tau_{g,h}(z) = \frac{1}{g}\left(e^{gz} - 1\right) e^{hz^2/2} \tag{2}$$

with $g \neq 0$ and $h \in \mathbb{R}$ such that $\tau_{g,h}(z)$ is a one-to-one monotone function of $z \in \mathbb{R}$. Then Y is said to have a Tukey's g-and-h distribution with location parameter (median) ξ and scale parameter ω:

$$Y \sim T_{g,h}(\xi, \omega).$$

Parameter g controls the direction and the degree of skewness, while h controls the tail thickness (or elongation). The family of $T_{g,h}(\xi,\omega)$-distributions is very flexible and approximates well many commonly used distributions (Martinez and Iglewicz, 1984; MacGillivray, 1992; Jiménez and Viswanathan, 2011).

As shown among others by Xu and Genton (2015), the density function of the $T_{g,h}(0,1)$-distributed random variable $T = \tau_{g,h}(Z)$ takes the form:

$$f_{T|g,h}(t) = \frac{\phi\left(\tau_{g,h}^{-1}(t)\right)}{\tau'_{g,h}\left(\tau_{g,h}^{-1}(t)\right)}, \qquad t \in \mathbb{R}, \tag{3}$$

where $\phi(\cdot)$ is the standard normal density function, and $\tau_{g,h}^{-1}(\cdot)$ and $\tau'_{g,h}(\cdot)$ are the inverse and first derivative of the function $\tau_{g,h}(\cdot)$, respectively. Hence, the density function of the $T_{g,h}(\xi,\omega)$-distributed random variable $Y = \xi + \omega T$ can be written as:

$$f_{Y|g,h,\xi,\omega}(y) = f_{T|g,h}\left(\frac{y-\xi}{\omega}\right)\frac{1}{\omega}, \qquad y \in \mathbb{R}. \tag{4}$$

Suppose now that we have a random sample $\{y_1, \ldots, y_n\}$ of n realizations of Y. Then the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ of the parameters vector $\boldsymbol{\theta} = (\xi, \omega, g, h)^{\mathrm{T}}$ is obtained by maximizing the log-likelihood function

$$\begin{aligned} L_n(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log f_{Y|\boldsymbol{\theta}}(y_i) \\ &= \sum_{i=1}^{n} \left[\log \phi \left(\tau_{g,h}^{-1} \left(\frac{y_i - \xi}{\omega} \right) \right) - \log \omega - \log \tau'_{g,h} \left(\tau_{g,h}^{-1} \left(\frac{y_i - \xi}{\omega} \right) \right) \right]. \end{aligned} \tag{5}$$

It is well known that under mild regularity conditions, $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is efficient. Unfortunately, since $\tau_{g,h}^{-1}(\cdot)$ does not have a closed form, numerically evaluating $L_n(\boldsymbol{\theta})$ can be computationally expensive, especially when the sample size is large. For this reason, the existing literature has largely been focused on searching for alternative estimators. One of these alternatives consists of estimating $\boldsymbol{\theta}$ by the values of the parameters that minimize the discrepancy between the empirical and theoretical order statistics of Y, that is, that minimize the following loss function:

$$\sum_{i=1}^{n} \left[y_{(i)} - \{ \xi + \omega \tau_{g,h}(z_{(i)}) \} \right]^2 \tag{6}$$

where $y_{(i)}$ is the i-th order statistic[2] among $y_1, \ldots, y_n$, and $z_{(i)} = \Phi^{-1}\left(\frac{i}{n+1}\right)$ is the quantile of order $\frac{i}{n+1}$ of the standard normal distribution.[3] This estimation technique is a variant of the quantiles least squares method proposed by Xu et al. (2014).

2.2 Flexible maximum likelihood estimation

Consider linear regression model

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{7}$$

where $\mathbf{x}_i^{\mathrm{T}}$ is the row vector of explanatory variables and $\boldsymbol{\beta}$ is the column vector of regression parameters. Let us assume that the disturbances ε_i are independent and identically distributed according to a $T_{g,h}(0, \omega)$-distribution, with g, h, and ω unknown. In this context, we have to estimate two parameters vectors: $\boldsymbol{\beta}$ and $\boldsymbol{\theta}^* = (0, \omega, g, h)^{\mathrm{T}}$.

The log-likelihood function takes the form:

$$L_n(\boldsymbol{\beta}, \boldsymbol{\theta}^*) = \sum_{i=1}^{n} \left[\log \phi \left(\tau_{g,h}^{-1} \left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}}{\omega} \right) \right) - \log \omega - \log \tau'_{g,h} \left(\tau_{g,h}^{-1} \left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}}{\omega} \right) \right) \right]. \tag{8}$$

However, the joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}^*$ is a quite complex computational problem. We therefore split the problem into two simpler ones that are solved iteratively similarly to what is done in Expectation Maximization algorithms.

The procedure is the following:

1. Take the L_1-estimate $\widehat{\boldsymbol{\beta}}_{\mathrm{L}_1}$ as initial estimate of the regression parameters vector $\boldsymbol{\beta}$: $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\mathrm{L}_1}$.

[2] i is the *rank* of $y_{(i)}$ among the n realizations of Y.

[3] $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

2. Determine the residuals $\widehat{\varepsilon}_i = y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ $(i = 1, \ldots, n)$ and estimate the vector of Tukey's parameters $\boldsymbol{\theta}^*$ by the vector $\widehat{\boldsymbol{\theta}}^* = (0, \widehat{\omega}, \widehat{g}, \widehat{h})^{\mathrm{T}}$ that minimizes the loss function

$$\sum_{i=1}^{n} \left[\widehat{\varepsilon}_{(i)} - \omega\tau_{g,h}(z_{(i)})\right]^2$$

where $\widehat{\varepsilon}_{(i)}$ is the i-th order statistic among $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$, and $z_{(i)} = \Phi^{-1}\left(\frac{i}{n+1}\right)$.

3. Determine the maximum likelihood estimator of $\boldsymbol{\beta}$ assuming that the errors ε_i $(i = 1, \ldots, n)$ in model (7) are distributed according to a $T_{\widehat{g},\widehat{h}}(0, \widehat{\omega})$-distribution: $\widehat{\boldsymbol{\beta}}$ is now the value of the regression parameters vector that minimizes the log-likelihood function (8) in which the unknown vector $\boldsymbol{\theta}^*$ is replaced by its estimate $\widehat{\boldsymbol{\theta}}^*$ obtained in step 2. For this step, since the log-likelihood function does not have an explicit expression, we approximate it by an explicitly computable function proposed by Xu and Genton (2015).

4. Iterate 2 and 3 till the convergence.

3 Simulation Study

In this Section, we describe a simulation study performed to assess the performance of the flexible maximum likelihood (FML) estimator proposed above. The data y_i $(i = 1, \ldots, n)$ are generated according to the following linear model:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

where β_0 and β_1 are set equal to one, and x_1 is normally distributed with zero mean and unit variance. The error term ε is generated from: (i) a $N(0, 1)$-distribution; (ii) a Student distribution with 5 degrees of freedom, $t(5)$; (iii) a Laplace(0,2)-distribution; (iv) a shifted (zero mean) centered Chi-square distribution with 5 degrees of freedom, $\chi^2_c(5)$.

Three different sample sizes — $n = 100$, $n = 500$, and $n = 1000$ — are considered. The number B of replications is equal to 10,000. The FML estimate is computed iteratively as explained in Section 2. The iterations stop when the absolute relative change in the estimate is smaller than 10^{-4} with a maximum number of iterations set to 100.

Table 1 shows the gain in efficiency for β_0 (and β_1) of FML with respect to the ordinary least squares (LS), for the different error distributions and sample sizes. The gain in efficiency is defined as $100 - (\text{RMSE}) \times 100$, where the RMSE (Relative Mean Squared Error) is defined as the ratio of the mean squared error of β_0 (and β_1) of FML over that of LS. As can be seen in table 1, the efficiency gain increases substantially when the error term departs significantly from normality. For the Gaussian distribution, the problem of low efficiency for β_1 is reduced when the sample size increases. For symmetric but heavy tailed distributions, the gain is substantial for both small ($n = 100$) and large sample size ($n = 1000$). The gain in efficiency is even larger for skewed distributions such as shifted centered $\chi^2_c(5)$. For the sake of completeness, bias and MSE of the FML estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are presented in Table 2. Since in the simulations the mean (and not the median) of the distribution of the errors is set to zero, the intercept estimated with FML is biased. To have a fair comparison between LS and FML (in terms of MSE) for skewed distributions, we correct the constant of FML (after convergence) by adding mean residual to it.[4]

Finally, Figure 1 shows the gain in efficiency when the error terms are generated from various Tukey g-and-h distributions for a grid of g and h values. A darker grey corresponds to a higher gain in efficiency with respect to LS. The equi-efficiency contours are represented by solid lines while the dashed line is the zero gain equi-efficiency contour. The grid of g spans -0.60 to 0.60 and the grid of h spans -0.05 to 0.60. We locate in the graph the four distributions considered in the previous simulations. The results are shown for a sample of 100 observations. Results are comparable for larger n and available upon request. When g and h increase, the efficiency of FML increases with respect to LS. Overall, FML tends to outperform LS as soon as the distribution of the innovation term becomes notably different from a normal distribution.

Table 1: *Gain in efficiency in %*

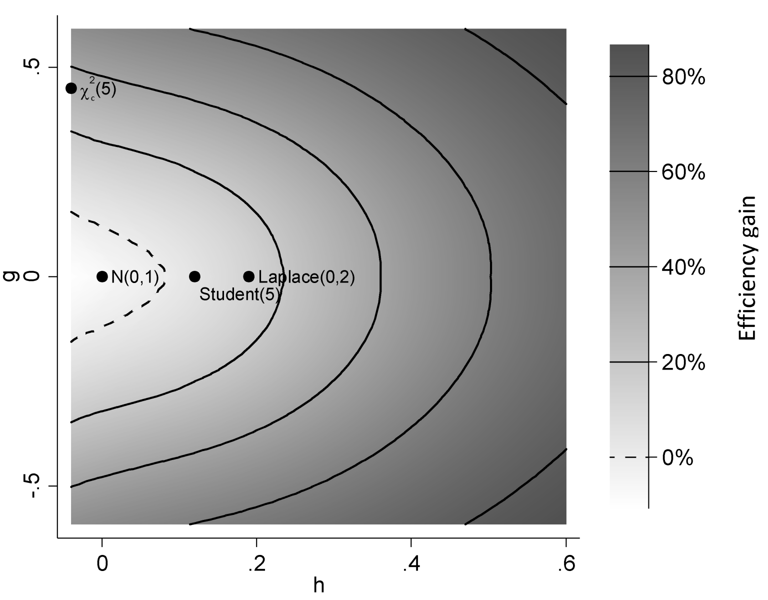
		$n = 100$	$n = 500$	$n = 1000$
$\varepsilon \sim N(0,1)$	β_0	-0.219	0.011	-0.008
	β_1	-18.214	-2.134	-0.680
$\varepsilon \sim t(5)$	β_0	0.106	0.046	0.001
	β_1	11.250	13.178	12.550
$\varepsilon \sim \text{Laplace}(0,2)$	β_0	0.045	0.056	0.029
	β_1	19.570	19.257	18.386
$\varepsilon \sim \chi^2_c(5)$	β_0	0.226	0.129	0.037
	β_1	28.241	38.094	38.915

Table 2: *Bias and MSE of FML*

			$n = 100$	$n = 500$	$n = 1000$
$\varepsilon \sim N(0,1)$	β_0	Bias	-0.002	-0.001	0.000
		MSE	0.010	0.002	0.001
	β_1	Bias	0.003	0.000	0.000
		MSE	0.012	0.002	0.001
$\varepsilon \sim t(5)$	β_0	Bias	0.001	0.000	-0.001
		MSE	0.017	0.003	0.002
	β_1	Bias	0.001	-0.001	0.000
		MSE	0.015	0.003	0.001
$\varepsilon \sim \text{Laplace}(0,2)$	β_0	Bias	0.000	0.000	0.000
		MSE	0.005	0.001	0.001
	β_1	Bias	0.000	0.000	0.000
		MSE	0.004	0.001	0.000
$\varepsilon \sim \chi^2_c(5)$	β_0	Bias	0.002	0.002	-0.002
		MSE	0.100	0.020	0.010
	β_1	Bias	0.002	0.002	-0.001
		MSE	0.072	0.013	0.006

[4] In practice, we never know the true distribution of the error term. Hence, we do not advise making any correction to empirical studies. In any case, the issue of the location parameter of the error distribution would only affect the constant.

Figure 1: *Gain in efficiency for β_1 varying g and h*



The figure reports the gain in efficiency for β_1 for various g and h (see text). Results are shown for a sample size of 100 observations.

4 Empirical application

In this section, we use the proposed methodology to study how AAA bond rates react to changes in 10-year bond rates. As Nolan and Ojeda-Revah (2013), we select from the Federal Reserve Board the rates for 10-year U.S. constant maturity bonds and AAA corporate bonds for the time period 2002-2014.[5] Week-to-week differences are computed and the difference in AAA bond rates are regressed on the difference in 10-year bond rates. Table 3 shows the summary statistics for residuals. We test normality with the Jarque-Bera test. The null hypothesis (that data are normally distributed) is rejected at less than 1%. The distribution of the residuals is heavy tailed and slightly skewed, which makes FML preferable to LS. The FML-estimates of the intercept and of the slope are respectively: -0.002 and 0.756. Table 4 reports estimated parameters ($\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}^*$) and the bootstrapped standard errors (in parentheses) using FML.

Table 3: *Summary statistics for residuals*

	# Obs.	Mean	Std. Dev.	Skew.	Kurt.	Jarque-Bera test (p-value)
residuals	678	0.001	0.046	0.499	6.707	0.000

Table 4: *Regression analysis using FML*

$\widehat{\beta}_0$	-0.002 (0.002)
$\widehat{\beta}_1$	0.756*** (0.016)
$\widehat{\omega}$	0.008*** (0.002)
$\widehat{g}$	0.067 (0.049)
$\widehat{h}$	0.170*** (0.028)

Bootstrapped standard errors (in parentheses). *** significant at 1%.

5 Conclusion

This paper introduces a flexible maximum likelihood regression estimator with Tukey g-and-h distributed errors for linear regression. Simulations show that FML outperforms LS in terms of efficiency as soon as the error distribution departs from normality. This methodology can be applied across a broad range of finance and economic topics.

[5] Available online at www.federalreserve.gov/releases/h15/data.htm.

References

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

Jiménez, J. A. and A. Viswanathan (2011). Using Tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk. *The Journal of Risk 13*(4), 95–116.

MacGillivray, H. L. (1992). Shape properties of the g-and-h and Johnson families. *Communications in Statistics - Theory and Methods 21*, 1233–1250.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business 36*(4), 394–419.

Martinez, J. and B. Iglewicz (1984). Some properties of the tukey g and h family of distributions. *Communications in Statistics-Theory and Methods 13*(3), 353–369.

Nolan, J. P. and D. Ojeda-Revah (2013). Linear and nonlinear regression with stable errors. *Journal of Econometrics 172*, 186–194.

Rachev, S. T. (2003). *Handbook of Heavy Tailed Distributions in Finance*, Volume 1. Amsterdam: North-Holland.

Tukey, J. W. (1977). Modern techniques in data analysis, nsf sponsored regional research conference. *Southern Massachusetts University, North Dartmouth, Massachusetts*.

Xu, G. and M. G. Genton (2015). Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution. *Computational Statistics & Data Analysis 91*, 78–91.

Xu, Y., B. Iglewicz, and I. Chervoneva (2014). Robust estimation of the parameters of g-and-h distributions, with applications to outlier detection. *Computational statistics & data analysis 75*, 66–80.

European Stability Mechanism

6a Circuit de la Foire Internationale
L-1347 Luxembourg

Tel: +352 260 292 0

www.esm.europa.eu

info@esm.europa.eu