A Note on Box-Cox Quantile Regression Estimation of the Parameters of the Generalized Pareto Distribution

JM van Zyl

Abstract: Making use of the quantile equation, Box-Cox regression and Laplace distributed disturbances, likelihood estimators are found making use of least absolute deviation methods for the parameters of the generalized Pareto distribution.

Keywords: *Generalized Pareto, quantile, Box-Cox, Laplace distribution*

AMS classification: 62F10,60E05

1 Introduction

The parameters of the generalized Pareto distribution (GPD) are estimated by making use of quantile regression and the Box-Cox form (Zellner ,1996), assuming Laplace (double exponential) distributed errors in the regression. This assumption leads to the least absolute estimation (LAD) form, often used in quantile regression (Kuan (2007), Koenker and Hallock (2001)). DasGupta and Mishra (2003) wrote a review on LAD estimation. Some of the aspects of importance are the asymptotic normality of LAD estimators, iterative estimation procedures and if the disturbances follow a Laplace distribution, likelihood estimation is equivalent to LAD estimation.

The properties and a discussion of estimation for the GPD is given in the book of Johnson, Kotz and Balakrishnan (1994). The two methods most widely to estimate the GPD parameters are the moment estimator (Hosking &Wallis (1987)) and maximum likelihood estimation (Zhang, 2007), (Smith, 1985). There are complications besides the difficult nonlinear aspect of the maximization of the likelihood, and the anamolous behaviour of the likelihood surface is discussed by del Castillo and Daoudi (2009). Teugels and Vanroelen (2004) investigated the use of the Box-Cox transformation for heavy-tailed distributions focusing on the transformation on the regular variation properties of tail quantile functions.

The GPD refers to a two-parameter distribution, but the distribution can be written in a generalized 3-parameter GPD form with a location parameter μ,

$$F(x) = 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, x \geq \mu, \sigma > 0, \xi \geq 0 ,$$

σ a scale parameter and ξ the shape parameter. The shape parameter is often important to estimate when working with extreme values and Pickands (1975) showed that for large n the tail index over a certain threshold can be estimated by using the fact that the distribution of the observations over the threshold follows a generalized Pareto distribution. By rewriting the distribution function, it follows that

$(1-F(x))^{-\xi} - 1 = \dfrac{\xi(x-\mu)}{\sigma}$. Let $F^{(\xi)} = \dfrac{\left(\dfrac{1}{1-F(x)}\right)^{\xi} - 1}{\xi}$,

thus $F^{(\xi)}$ has the linear regression form $F^{(\xi)}(x) = \beta_0 + \beta_1 x, \beta_0 = \mu/\sigma, \beta_1 = 1/\sigma$.
A variation of quantile regression with the observations as the dependent variable was used, but it does not lead to the linear regression form without approximating the distribution (Johnson, Kotz and Balakrishnan , 1994), (Jagger and Elsner , 2008). For most applications μ is zero and the simpler form
$F^{(\xi)}(x) = \beta_1 x, \beta_1 = 1/\sigma$ used.
The likelihood will be derived in section 2 and the results of a simulation study given to compare this estimator with the moment estimator.

2 The likelihood and simulation results

The stochastic model with an observed value $y^{(\xi)} = F^{(\xi)}(x) + u$, where u is an error term assumed to be i.i.d. Laplace distributed with mean zero and variance 2ϕ. It is reasonable to assume that the errors have a symmetric distribution and the Laplace assumption leads to the correct form for quantile regression and more robust estimators. The Laplace density is

$$p_U(u) = \frac{1}{2\phi}\exp(-|u-\theta|/\phi), \quad -\infty < u < \infty, \phi > 0,$$

with variance 2ϕ. The maximum likelihood estimate of ϕ is the mean absolute deviation calculated about the median of the sample It will be assumed that $E(u) = \theta = 0$.

Denote a sample of size n observations from the GPD by $x_1,...,x_n$. The empirical distribution values are $F(x_{(j)}) = \frac{j}{n+1}$ for the ordered sample $x_{(1)},...,x_{(n)}$, Denote the vector of observations by $\mathbf{y}^{(\xi)}$, $\mathbf{u}$ the vector of errors. Let

$$X = \begin{pmatrix} 1 & x_{(1)} \\ \vdots & \vdots \\ 1 & x_{(n)} \end{pmatrix}.$$

The Jacobian of the transformation from the u_j, $j = 1,...,n$ to $y_1,...,y_n$ is

$|J(u \to y)| = \prod_{j=1}^{n} y_j^{\xi-1}$, by noting that $\frac{\partial u_j}{\partial y_i} = (1-y_i)^{\xi+1}$, i=j and 0 otherwise.

By following the same methods as in Zellner (1996), it follows that the log-likelihood simplifies to

$L = const + (\xi+1)\sum_{j=1}^{n}\log(y_j) - n\log(\hat{\tau}(\xi))$, where $\hat{\tau}(\xi) = \frac{1}{n}\sum_{j=1}^{n}|y_j^{(\xi)} - \hat{\beta}_0 - \hat{\beta}_1 x_{(j)}|$

denotes the mean absolute deviation. $\hat{\boldsymbol{\beta}}' = [\hat{\beta}_0 \ \hat{\beta}_1]'$ is such that it is the least absolute estimators (LAD) of β_1, β_2 for the given ξ. The maximum likelihood estimators can be found by finding the values of ξ and $\boldsymbol{\beta}$ for which the log-likelihood is a maximum. In the case of the 3-parameter GPD iteratively re-weighted least squares (IRLS) can be used for a given value of ξ to find the LAD estimators, the log-likelihood calculated and the maximum over all possible $\xi's$ yields the estimator. The method involves the iterative procedure:

$\hat{\boldsymbol{\beta}}^{(j+1)} = (X'W^{(j)}X)^{-1}X'W^{(j)}\mathbf{y}^{(\xi)}$, $W^{(j)}$ a diagonal matrix with diagonal elements:

$$w_i^{(j)} = |y_i^{(\xi)} - X_i\boldsymbol{\beta}^{(j)}|^{-1},\ X_i \ i-th\ row, i = 1,...,n.$$

The weights of zero or very small residuals can be put equal to a large number. This procedure is repeated until convergence. Moment estimators of $\boldsymbol{\beta}$ can be used as a starting value for the iterative procedure. Since the least squares estimates are not very stable for the 3-parameter GPD samples, these estimates was not used as starting values.

The moment estimators are $\hat{\sigma} = \frac{1}{2}\bar{x}(\frac{\bar{x}^2}{s^2}+1)$ and $\hat{\xi} = \frac{1}{2}(\frac{\bar{x}^2}{s^2}-1)$.

The results of a simulation study comparing the moment estimators and the Laplace regression form will be given below. The usual two parameter form will be considered first. It is assumed that $\xi = 0.35, \sigma = 1$. The average estimated parameters for 500 simulations are given in table 1 and 2. Note that $\hat{\beta}_1 = 1/\hat{\sigma}$. The technique will be called the Box-Cox method in the table.

Sample Size $(\sigma = 1)$	**Box-Cox**	**Moment Estimator**
25	1.0674 (0.0448)	0.8534 (0.0581)
50	1.0432 (0.026)	0.8322 (0.0275)
150	1.0100 (0.0009)	0.9031 (0.0126)
250	1.0115 (0.0006)	0.9205 (0.0083)

Table 1. Estimated $\hat{\beta}_1$'s and $\text{var}(\hat{\beta}_1)$ in brackets.

The simulation results for ξ is shown in table 2.

Sample Size $(\xi = 0.35)$	**Box-Cox**	**Moment Estimator**
25	0.3480 (0.0146)	0.1930 (0.0137)
50	0.3488 (0.0043)	0.2212 (0.0117)
150	0.3669 (0.0074)	0.2575 (0.0072)
250	0.3541 (0.0039)	0.2775 (0.0056)

Table 1. Estimated $\hat{\xi}$'s and $\text{var}(\hat{\xi})$ in brackets.

It can be seen that the Box-Cox regression form estimators are stable for the various sample sizes and closer estimates than the moment estimator, especially when estimating ξ. It is a nonlinear procedure, but much easier to perform than the maximization of the GPD likelihood.

Shown below is a histogram of 500 $\hat{\xi}'s$ estimated from samples with 250 observations, the GPD parameters are $\xi = 0.35, \sigma = 1, \mu = 0$. The mean of the $\hat{\xi}'s$ is 0.3578 and covariance 0.0039.

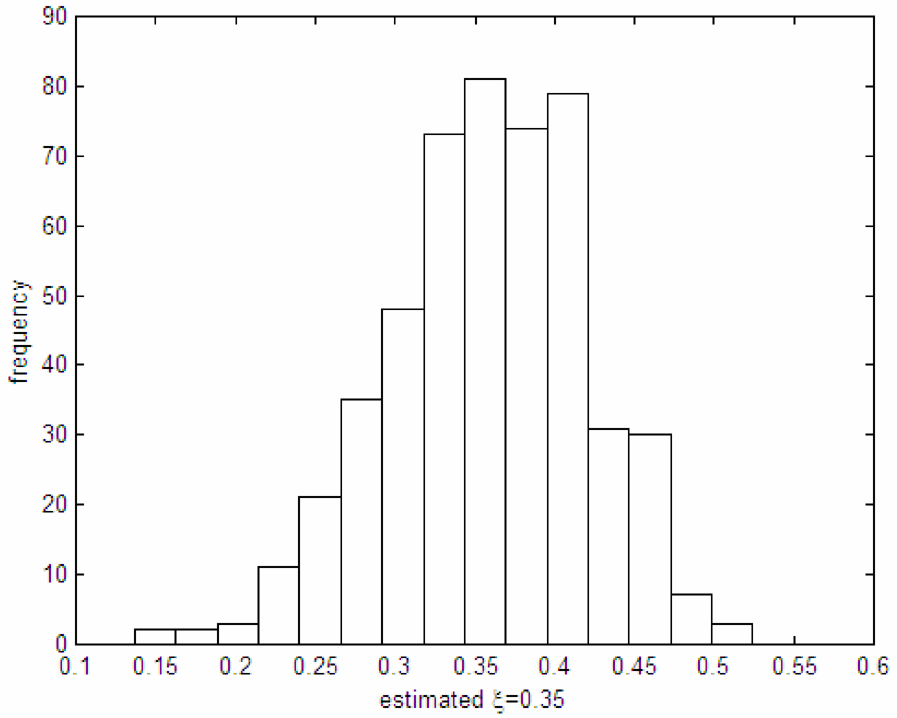

Figure 1. Histogram of 500 estimated $\hat{\xi}'s$, n=250, $\xi = 0.35$.

In the two histograms below, 500 estimated $\beta_0 's, \beta_1 's$ from sample of size 50 for the 3-parameter GPD with parameters $\xi = 0.40, \sigma = 2, \mu = 2$ is shown. Thus $\beta_1 = 0.5$ and $\beta_0 = -1$. The averages of the $\hat{\xi}'s, \hat{\beta}_0 's \ and \ \hat{\beta}_1 's$ are 0.3716, -0.7713 and 0.4315, with covariances 0.0288, 0.1202 and 0.0148 respectively.

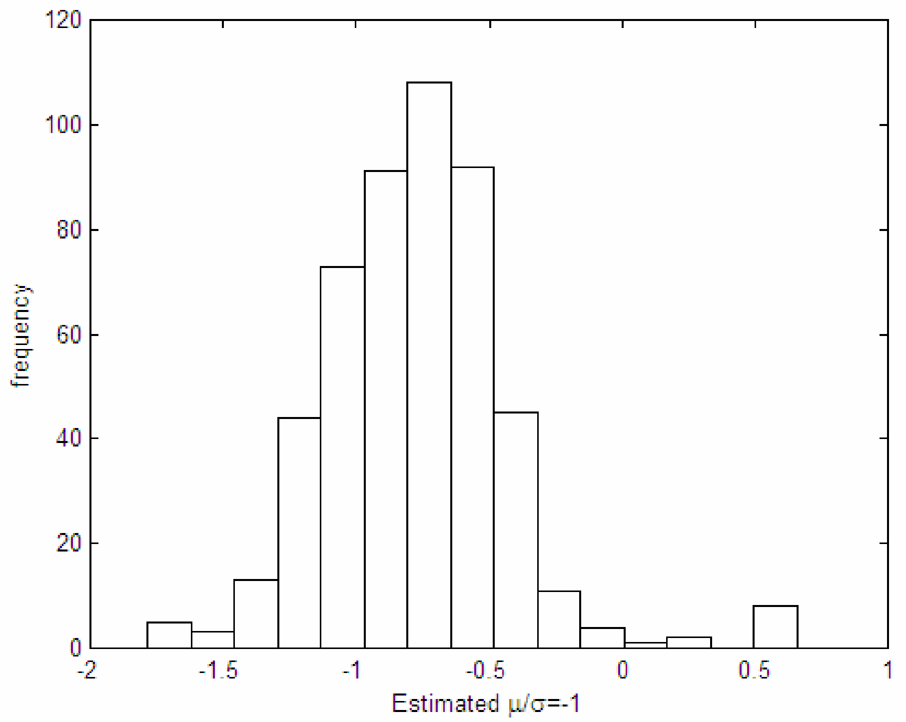

Figure 2. Histogram of 500 estimated $\beta_0 = -\mu/\sigma = -1$, n=50, $\xi = 0.40$.

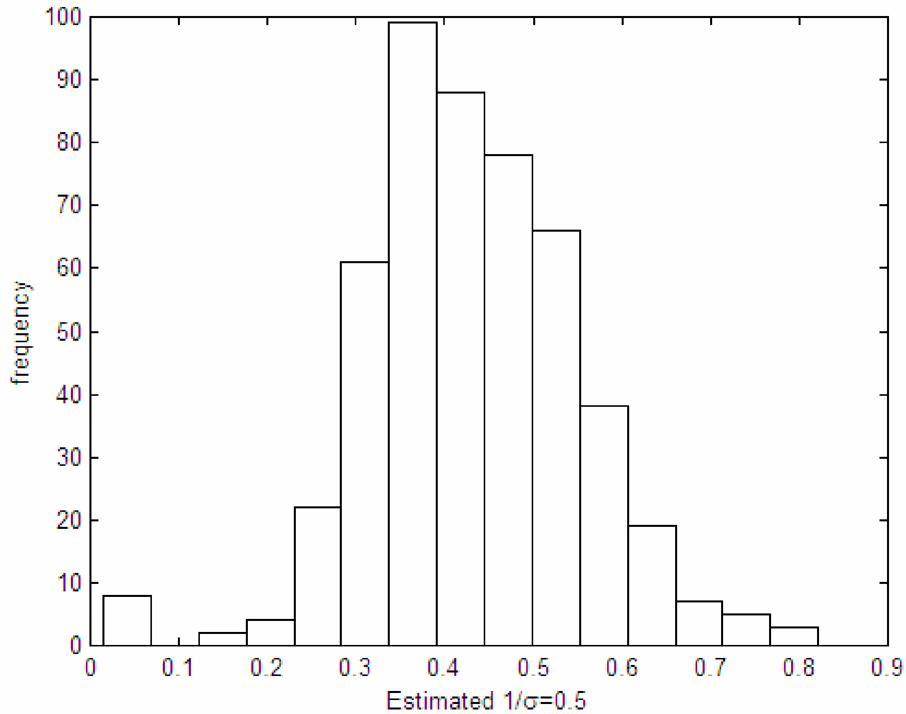

Figure 3. Histogram of 500 estimated $\beta_1 = 1/\sigma = 0.5$, n=50, $\xi = 0.40$.

3. Application to the absolute NYSE Composite log returns

This technique was applied to the 150 largest absolute log returns of the NYSE Composite index, daily closing values index, for the period beginning 2000 till the end of 2009. There were 2514 observations. The estimated parameters are: $\hat{\xi} = 0.1920, \hat{\sigma} = 0.0126, \hat{\mu} = 0.0248.$ The index, the absolute log returns and the fitted cdf of the GPD versus the empirical cdf are shown in figures 4 - 6.

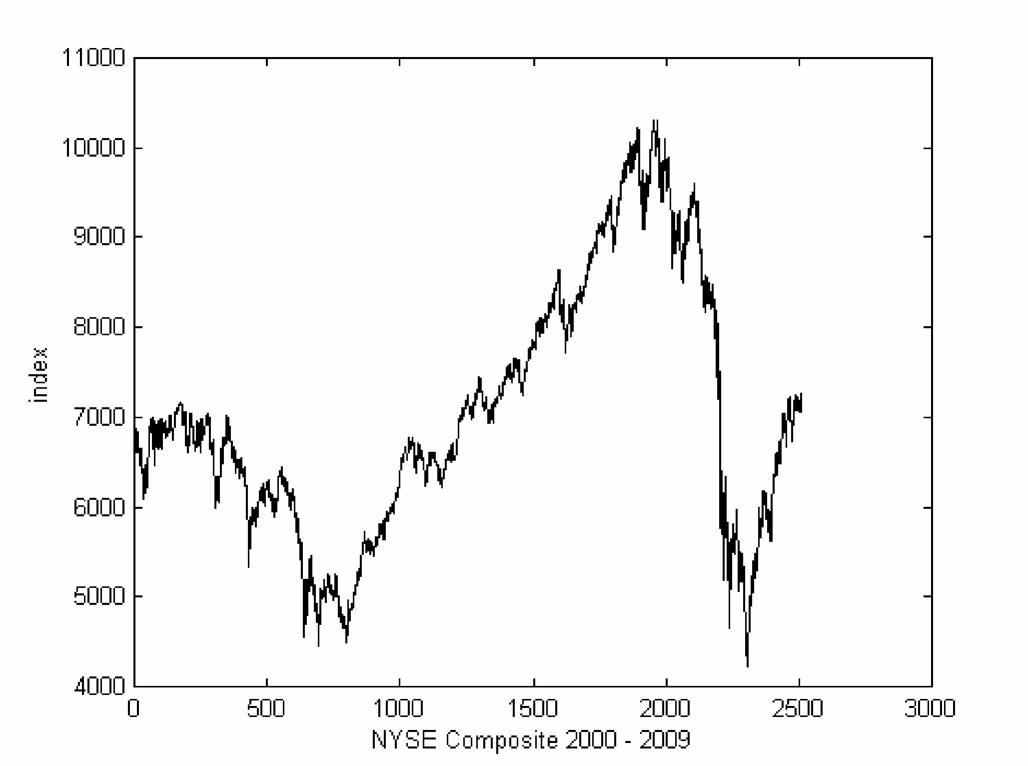

figure 4. Daily closing values of the NYSE Composite Index, 2000 - 2009.

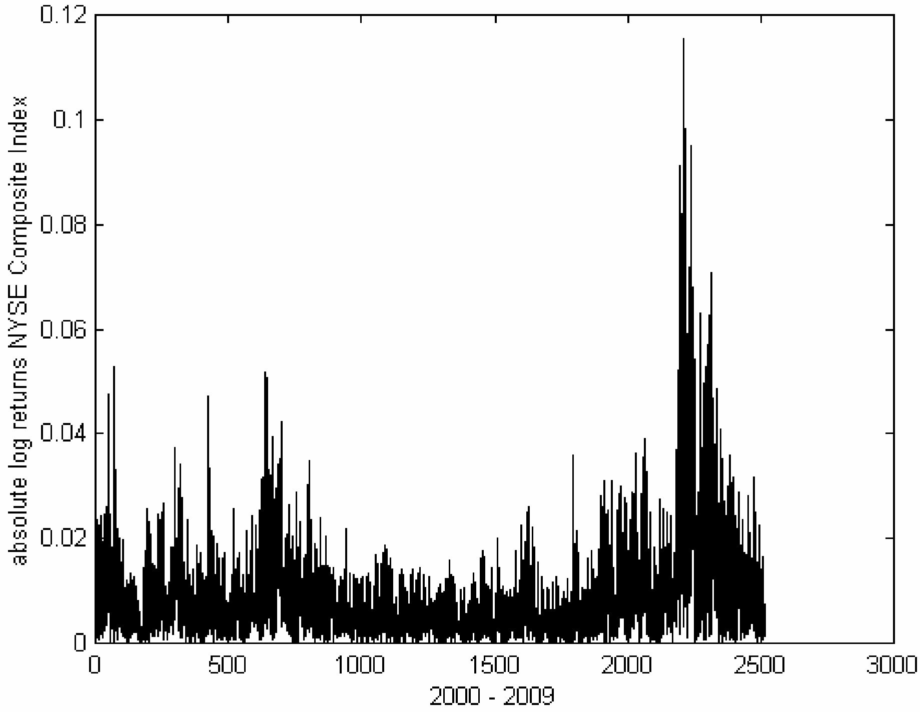

figure 5. Absolute log returns, NYSE Composite Index.

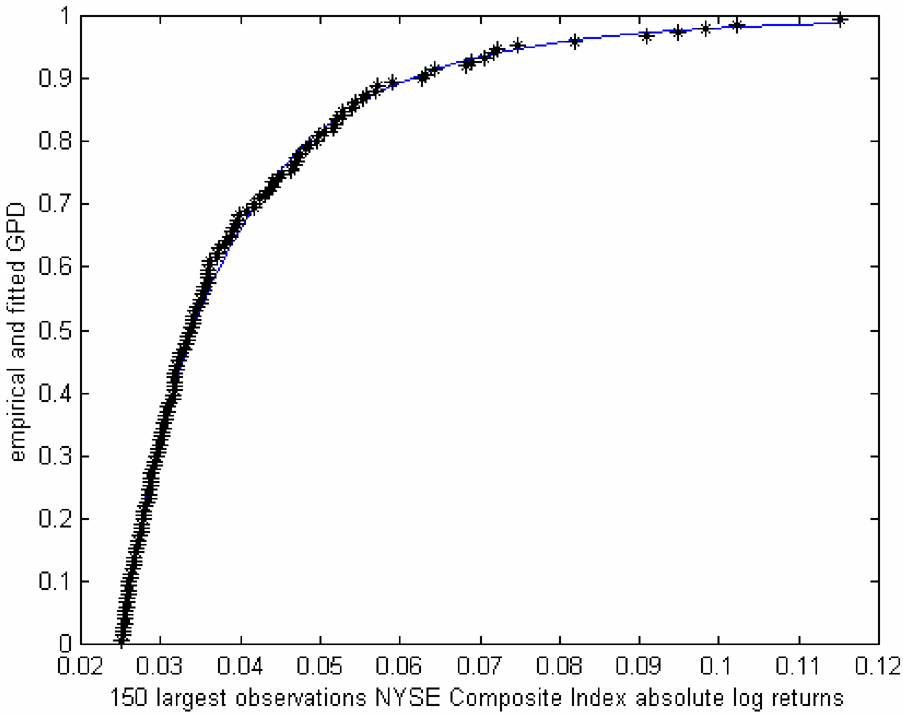

Figure 6. The fitted GPD and empirical cumulative distribution function.

The extreme values is clearly well fitted by the GPD showing that the largest absolute log returns are regularly varying, but the index is greater than 4, which is an important condition needed to fit GARCH models.

4 Conclusion

The quantile regression form assuming Laplace errors, outperforms the method of moments. Although involving nonlinear LAD estimation, it is much easier than maximizing the likelihood. The estimators have good asymptotic properties since the estimators are maximum likelihood estimators. Possible refinements can be developed to ensure fast convergence of the log-likelihood maximization.

References

DasGupta, M. and Mishra, R.K.,2003. Least Absolute Deviation Estimation of Multi-Equation Linear Econometric Models: A Study Based on Monte Carlo Experiments. North-Eastern Hill University (NEHU) Economics Working Paper No. skm/02.

Del Castillo, J. and Daoudi, J., 2009. Estimation of the generalized Pareto distribution. Statistics and Probability Letters, 79, 684 – 688.

Hosking, J.R.M. and Wallace, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics, 29(3), 339 – 349.

Jagger, T.H. and Elsner, J.B., 2008. Modelling tropical cyclone intensity with quantile regression. International Journal of Climatology, 29,10, 1351 – 1361.

Johnson, N.L., Kotz, S. and Balakrishnan, N., 1994. Continuous Univariate Distributions. Volume 1. Wiley, New York.

Koenker, R. and Hallock, K., 2001. Quantile regression. Journal of Economic Perspectives, 15, 143 – 156.

Kuan, C-M, 2007. An introduction to quantile regression. Institute of Economics, Academia Sinica.

Pickands, J.,1975. Statistical inference using extreme order statistics, Ann. Statistics, 3, 119 – 131.

Smith, R.L., 1987. Estimating tails of probability distributions. Ann. Statist.,15, 1174 -1207.

Teugels, J.F. and Vanroelen, G., 2004. Box-Cox transformations and heavy-tailed distributions. Stochastic Methods and Their Applications, 41, 213 – 227.

Zellner, A., 1996. An Introduction to Bayesian Inference in Econometrics. New York: Wiley.

Zhang, J., 2007. Likelihood moment estimation for the generalized Pareto distribution. Austr. N.Z. J. Stat. 49(1), 69 -77.