Sains Malaysiana 38(5)(2009): 745–749

A Simple Power-Law Tail Estimation of Financial Stock Return

(Penganggaran Hukum-Kuasa Taburan Hujung terhadap Pulangan Saham Kewangan)

CHIN WEN CHEONG*, ABU HASSAN SHAARI MOHD NOR & ZAIDI ISA

ABSTRACT

This study proposes a simple methodology to estimate the power-law tail index of the Malaysian stock exchange by using the maximum likelihood Hill's estimator. Recursive procedures base on empirical distribution tests are use to determine the threshold number of observations in the tail estimation. The threshold extreme values can be selected bases on the desired level of p-value *in the goodness-of-fit tests. Finally, these procedures are apply to three indices in the Malaysian stock exchange.*

Keyword: Goodness-of-fit test; Hill estimator; power-law distribution; stock exchange

ABSTRAK

Kajian ini bertujuan menganggarkan indeks hukum kuasa taburan hujung ke atas bursa saham Malaysia dengan menggunakan penganggar Hill. Prosedur rekursif berdasarkan ujian taburan empirik digunakan untuk menentukan nombor ambang bagi pencerapan di dalam penganggaran hujung. Nilai ambang melampau dipilih berdasarkan kepada aras nilai-p *ujian ketepatan padanan. Akhir sekali, prosedur ini dilaksanakan ke atas tiga indek di bursa saham Malaysia.*

Kata kunci: Bursa saham; penganggar Hill; taburan hukum-kuasa; ujian ketepatan padanan

INTRODUCTION

The power-law distribution (Lux 2001) has successfully described the extreme variations of financial time series (Bauchaud 2001; Lux 1996; Sarah 2000) which includes the stock price changes, volume as well as volatility decay distributions. From economic point of view, the major advantage of these findings provides good understanding on how the extreme asset prices behave under a particular equity market. Especially in risk management (Sarah 2000), the extreme swings in the asset prices have major impacts to the derivatives hedging and portfolio management. Besides the power-law distribution in the extreme tail, the upper and lower tails are also often observe to be asymmetry (Giot, 2004; Lambert & Laurent, 2001) in the financial time series. This phenomenon is very important in risk analysis where the different assets financial positions over a given time period relies seriously on the tail behaviours.

In this study, we use the percentage continuously compounded price changes (return): $r_t = (p_t - p_{t-1}) \times 100\%$, where p_t denotes the natural logarithm of a particular index at time t. The Pareto power-law distribution is a simple and useful model to provide a good fit for the empirical stock return distribution. Given R observations in a return series, the cumulative distribution for Pareto's law, $F(r_t; r_{(\min)}, \alpha) = 1 - (r_{(\min)}/r_t)^{\alpha}$, where $0<r_{(\min)}<r_t$ is the threshold value of the subset of R observations and $\alpha>0$ denotes the shape parameter. In practice, the fit between the empirical and theoretical distributions is often perform by judging the degree of linearity (under the ordinary least squared estimation) in a double logarithmic graphical method by using the ordinary least squared estimates. Others suggested $R^{0.6}$ (Hall 1990) or the ratio n/R ~0.5%- 1.0% (Franke et al. 2004) as the threshold observations that should included in the tail estimation. In general, these approaches encounter drawbacks of less subjectiveness (Goldstein et al. 2004) in selecting the threshold value. If a non-optimal (too large or small) threshold value has been used, the estimations might cause inaccurate, bias and large variance estimators. However, the maximum likelihood estimator in general provids more accurate and robust estimates (Goldstein et al. 2004) than the geometrical method. The optimal threshold selection issue has received great attention from researchers. However, there is no consensus of one particular methodology that out-performs others. For example, Clementi et al. (2006) proposed a subsample semi-parametric bootstrap procedure to minimize the variance estimator to obtain the threshold value, while Coronel-Brizio and Hernandez-Montoya (2005) used the empirical distributed function to identify the best threshold in the Pareto-Levy distribution. In general, recursive procedures are necessary to obtain the threshold in a more objective manner.

In the analysis of Malaysian stock equity market, the proposed method is based on the graphical plots and goodness-of-fit statistics. In order to compromise the trade-off of sample size and mean square error of the estimators, the selection of threshold relies on the *p-value* of the null hypothesis test that give by the goodness-of-fit statistics. The *p-value* approach has the advantage to provide a reference (in term of probability) to observe how well the empirical and theoretical distributions are fit. For example, one must avoid choosing the threshold when the *p-value* is either merely 'rejected' or 'do not reject' conclusion, even though one more observations can be included in the tail estimation to reduce the standard error. A series of steps is introduces to provide the threshold and Pareto distribution for both the upper and lower tails. For the purpose of illustrations, three indices namely, the Composite Index (CI), Finance Index (FIN) and Plantation (PLN) from 1987 to 2007 will be considered in the analyses.

METHODOLOGY

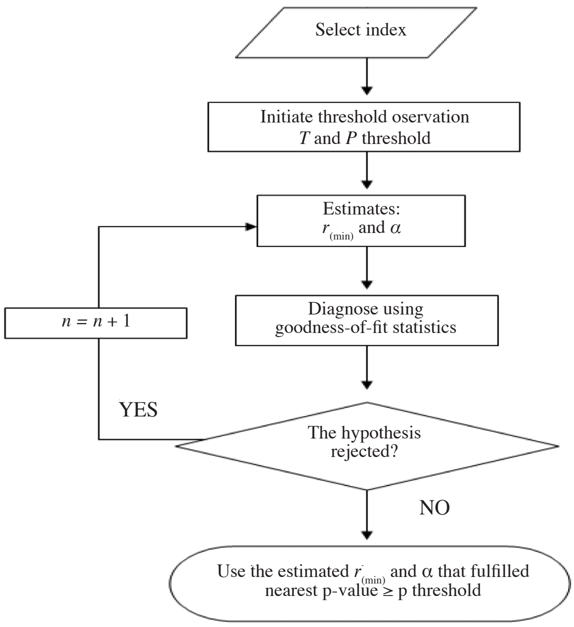


FIGURE 1. Flow chart

The flow chart in Figure 1 summarises the work-flow and the formal procedures are illustrate as follows:

Let $r_{(\min)} = r_{(1)} \le r_{(2)} \le \ldots \le r_{(n)}$ represents a subset of n order statistic of the R observations with the threshold, $r_{(\min)}$:

Step 1:Initiate the model fitting with T observations using the non-negative order statistics The sample size initialisation can be obtained by using the number of observations (outliers) that exceed Q3+3IQR (where Q3 and IQR are the upper quartile and interquartile range respectively).

Step 2: Evaluate the estimates of the Pareto's parameters using the Hill's estimator.

The tail behaviour can be estimated using the Hill's estimator (Hill 1975) with the underlying Pareto type or approximate to Pareto distribution. The log likelihood function can be expressed as $l(\alpha, r_{(1)}) = n\ln\alpha + n\alpha\ln r_{(1)} - (\alpha+1)\sum_{i=1}^{n}\ln r_{(i)}$ where $l(\alpha, r_{(1)})$ is monotonically increasing with $r_{(1)}$. Finally, the estimated α can be obtained by using the analytic partial derivative approach:

$$\hat{\alpha} = \frac{n}{\sum_{i}^{n}\ln r_{(i)} - \ln r_{(\min)}}. \tag{1}$$

Step 3:Use quantile-quantile plot and goodness-of-fit tests to diagnose the fitted distribution.

The goodness-of-fit tests follow the null and alternative hypotheses as follows:

H_0: Both the empirical distribution, $F_1(x)$ and Pareto distribution, $F_0(x)$ are identical;
H_1: H_0 is not true.

Both the quadratic statistics are define as the modified Cramer-von Misses statistic (W^2) and Watson statistic (U^2) as below:

$$W^2 = n\sum_{i=1}^{n}\left[F_1\left(r_{(i)}\right) - F_0\left(r_{(i)}\right)\right]^2 P_0\left(r_{(i)}\right)$$

$$U^2 = n\left\{\sum_{i=1}^{n}\left[F_1\left(r_{(i)}\right) - F_0\left(r_{(i)}\right)\right]^2 P_0(x) - \left(\sum_{i=1}^{n}\left[F_1\left(r_{(i)}\right) - F_0\left(r_{(i)}\right)\right]P_0\left(r_{(i)}\right)\right)^2\right\} \tag{2}$$

where $r_{(n)}$ is the largest observation and $[F_1(r_{(i)}) - F_0(r_{(i)})]^2 P_0(r_{(i)})$ is weighted by a function $\Psi(x) = 1$. If the quadratic deviations between the F_1 and F_0 are large, most likely the two sample cumulative distribution come from different populations.

Step 4: Determine the threshold, n.

If both the W^2 and U^2 statistics fail to reject the null hypotheses at *p-value* = $\alpha_{\text{threshold}}$, stop the computation and the threshold n is obtains. Else go to *Step 1* with (N–1) observations.

These procedures are also apply to the lower tail distribution (negative extreme returns) because the non-negative return can be obtained by a simple sign change.

EMPIRICAL APPLICATION: THE MALAYSIAN STOCK EXCHANGE RETURN DISTRIBUTION

PRELIMINARY ANALYSIS

Table 1 summarises the descriptive statistics for all the return series as well as two simulated normal and student-t distributions. It is evidenced that all the returns series have nearly zero mean and standard deviation around 1.50. Excess kurtosis clearly indicates across the market indices with the maximum of 43.32 (CI) and minimum 5.63 (IND), respectively. The joint tests of skewness (0 for normal distribution) and kurtosis (3 for normal distribution) in Jacque-Bera statistics indicate all the indices are significantly violate from a normal distribution.

TABLE 1. Statistics summary

	KLCI	FIN	IND	NORMAL	STUDENT-t (4)
Mean	0.0346	0.0411	0.1474	0.0333	-0.0086
Median	0.0429	0.0313	0.2195	0.0390	-0.0247
Maximum	20.817	22.627	5.6437	4.8986	13.390
Minimum	-24.153	-24.835	-5.0989	-5.0477	-8.7924
Std. Dev.	1.4829	1.7386	1.4386	1.4728	1.4013
Skewness	0.2872	0.4803	-0.0978	-0.0014	0.4963
Kurtosis	43.3266	36.6967	5.6316	2.9896	10.0374
Jarque-Bera	329041*	229882*	81.1712*	0.0234	10218*

Note: * denotes the 5% significance level.

In addition, the slow decaying tails can be observed through the kernel density estimates. In Figure 2, the kernel density is estimates by $\hat{f}(r_t) = \frac{1}{Rh}\sum_{i=1}^{T}\phi\left(\frac{r_t - \tilde{r}_t}{h}\right)$, where T is the total number of observations, h is the bandwidth and $\phi(\cdot)$ function of standard normal distribution.

EMPIRICAL ANALYSIS OF KLCI

For illustration purposes, only the KLCI index is shown, while the other indices are analyze in a similar manner. There are total of 4855 observations starting from year 6th Jan 1987 to 31st Dec 2007. The empirical analysis explains the step-by-step procedures as follows:

Step 1:Initiate the model fitting with N observations using the non-negative order statistics
The estimated upper quartile and lower quartile are 0.6028 and -0.5367 respectively with approximately 50 observations exceed Q3+3IQR. Therefore, the initial sample size starts with N=50. One may start the calculation bases on the simple rule of thumb method where firstly, the 0.5% of n/R is approximately 242 and secondly, the $R^{0.6}$ is around 162 for the n.

Step 2: Evaluate the estimates of the Pareto's parameters using the Hill's estimator.
Under the maximum likelihood estimation, the following parameters estimates with standard errors quote in the squared brackets

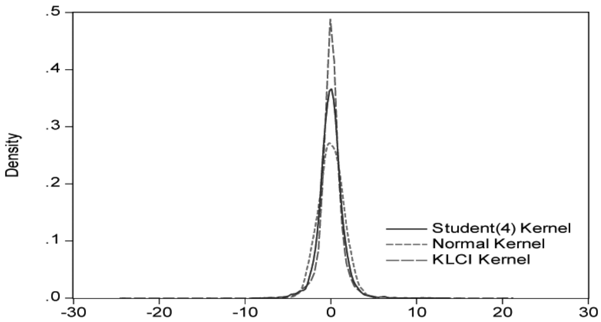
$$\left(\hat{r}_{(\min)},\hat{\alpha}\right) = (3.9299, 2.4702) \quad [(0.0323, 0.3753)]$$

which are both significant at 5% level of significance.

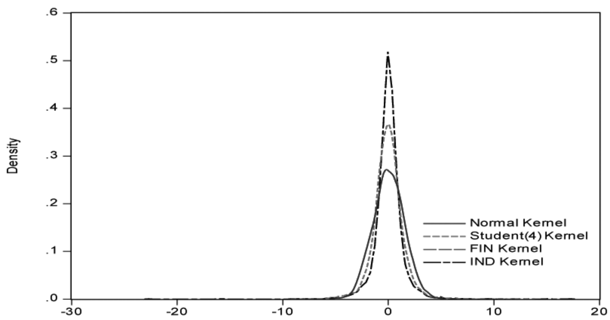
Step 3:Use quantile-quantile plot and goodness-of-fit tests to diagnose the fitted distribution.
Firstly, the Q-Q plot in Figure 3 indicates both the series lie on a straight line. Under the null hypothesis where both the empirical distribution and specified Pareto distribution are identical, the following results are obtained:

Statistically, both the goodness-of-fit statistics fail to reject the null hypothesis.

Step 4: Determine the threshold, n
Says, we has selected *p-value* = $p_{\text{threshold}} \sim 0.100$. Since the p-values are above the $p_{\text{threshold}}$, the computation is proceeds to *Step 1* with 50+1 data in the following estimation.



(a)



(b)

FIGURE 2. Kernel density plots for KLCI, FIN and IND return series

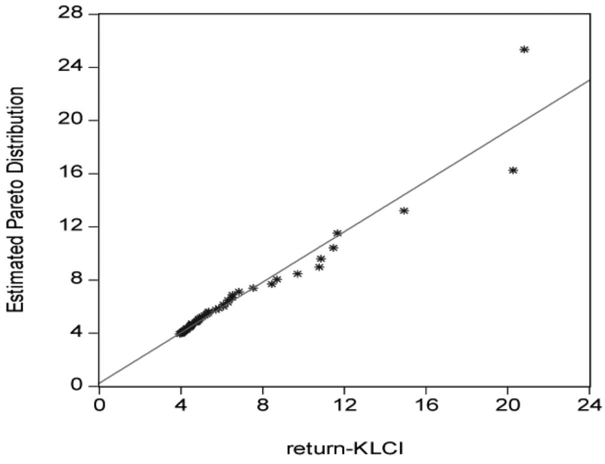


FIGURE 3. QQ plot between KLCI return series against Pareto distribution

TABLE 2. Goodness-of-fit test

Method	Test statistic	p-value
W^2	0.0499	0.7288
U^2	0.0396	0.7688

Note: the dotted line indicates one of the test statistic (W^2 and U^2) exceeds $p_{threshold}$ = 0.100

A summary of the threshold, *n* selection in the Hill's estimator is shown in Figure 4. The plot depicts the outcomes of both the W^2 and U^2 statistics for the KLCI index associate with the Hill's estimations. The measure of discrepancy indicated consistent growth in both the test statistics after the threshold, *n* exceeded 270. However, the choice of the threshold is bases on the $p_{\text{threshold}}$, which falls at the 273th descending order statistics. As show in Table 3, the threshold, *n* is selects as 273 since both the test statistics indicate nearest to $p_{\text{threshold}}$.

Base on the extreme outliers observations, Table 4 shows that the estimated shape parameters (αs) which are all exceed 2 for both the upper and lower tails and indicate the presence of finite means and variances whereas the moments of order higher than 2 are unbounded. According to Loretan and Phillips (1994), not necessary all the moment higher than 2, such as kurtosis is finite. The positive estimated α implies that the tails on both tails of the innovation distributions are heavy.

For thickness comparison of upper and lower tails, three indices (KLCI, FIN and IND) indicate slightly heavier tails at the upper tails where the smaller the shape parameter (α), the heavier the density mass of the tail. The asymmetric distributed tails provide useful information to the market participants who involve in portfolio investment or risk management analysis. For example in market risk

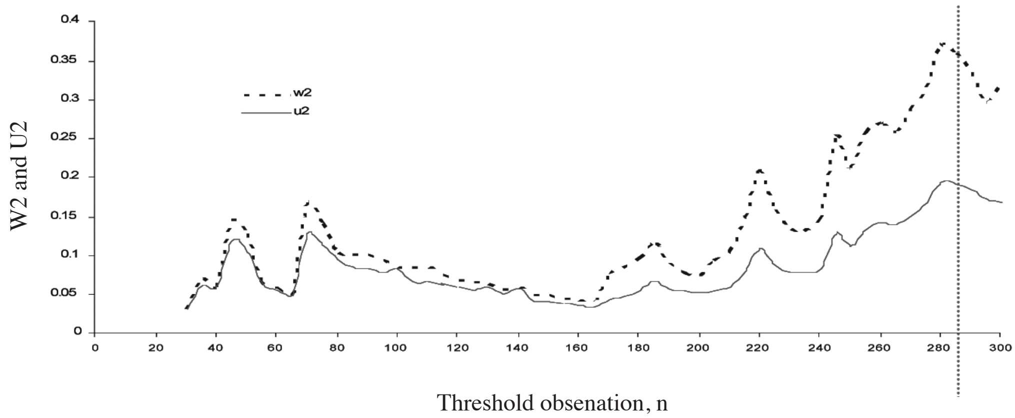


FIGURE 4. goodness-of-fit statistics (W2 and U2) versus value of threshold (*n*) for KLCI

TABLE 3. Threshold selection for KLCI

Threshold, n	Method	Test statistic	p-value	$r_{(min)}$	α
273	W^2	0.1496	0.1463	1.8483 (0.0031)	2.2103 (0.1355)
	U^2	0.0955	0.2186		
274	W^2	0.1838	*0.0872*	1.837556 (0.0031)	2.1902(0.1340)
	U^2	0.1138	0.1424		

Note: The value in the parenthesis denotes the standard error.

TABLE 4. Hill estimator for lower and upper tail

Index	Lower tail					Upper tail				
	$r_{(min)}$	α	W^2	U^2	n	$r_{(min)}$	α	W^2	U^2	n
KLCI	2.2696 (0.0045)	2.4933 (0.1794)	0.1562 (0.1319)	0.1006 (0.1941)	200	1.8483 (0.0031)	2.2103 (0.1355)	0.1496 (0.1463)	0.0955 (0.2186)	273
FIN	2.7526 (0.0065)	2.4366 (0.1896)	0.1723 (0.1035)	0.1162 (0.1347)	172	2.3393 (0.0041)	2.2537 (0.1448)	0.1370 (0.1783)	0.1009 (0.1925)	249
IND	2.0308 (0.0040)	2.4323 (0.1728)	0.1554 (0.1336)	0.0948 (0.2221)	205	1.9300 (0.0034)	2.4276 (0.1621)	0.1205 (0.2321)	0.0884 (0.2583)	231

Notes:
Value in the parenthesis for goodness-of-fit statistics denotes the p-value;
Value in the parenthesis for Pareto parameter estimates denotes the standard error;

determination bases on Value-at-Risk (Jorion 2002), the FIN market suggests that short trading (upper tail) might encounter higher risk as compare to long trading (lower tail) investments for all the indices.

CONCLUSION

This study proposes an objective method for fitting the power-law distribution to extreme variations in Malaysian stock indices. This methodology empirically shows that the goodness-of-fit statistics (W^2 and U^2) can be used to determine the optimal threshold parameter more subjectively than the simple rule of thumb method. From economic viewpoint, the estimated Pareto's tail index of return distribution is expects to help the investors in market risk determination where large amount of money can be lost due to failure of underestimating the market risks.

ACKNOWLEDGEMENT

We would like to thank the anonymous referee for helpful comments. We also would like to gratefully acknowledge the financial support from the Fundamental Research Grant Scheme and Research University Grant Scheme.

REFERENCES

Bouchaud, J.P. 2001. Power laws in economics and finance: Some ideas from physics. *Quantitative Finance* 1(1): 105-112.

Clementi, F., Matteo, T.D. & Gallegati, M. 2006. The power-law tail exponent of income distributions. *Physica A* 370: 49-53.

Coronel-Brizio, H.F. & Hernandez-Montoya, A.R. 2005. On fitting the Pareto-Levy distribution to stock market index data: selecting a suitable cutoff value. *Physica A* 354: 437-449.

Franke, J., Hardle W.K. & Hafner, C.M. 2004. *Introduction to Statistics of Financial Markets*. Germany: Springer-Verlag.

Giot, P. 2004. Modelling daily value at risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11(3): 379-398.

Goldstein, M.L., Morris, S.A. & Yen, G.G. 2004. Problems with fitting to the power-law distribution. *The European Physical Journal B* 41(2): 255-258.

Hall, P. 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis* 32: 177-203.

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3: 1163-1173.

Jorion, P. 2002. *Value-at-Risk: The new benchmark for controlling market risk*. Chicago: McGraw-Hill.

Lambert, P. & Laurent, S. 2001. *Modelling financial time series using GARCH–type models and a skewed Student density*. Mimeo. Université de Liège.

Loretan, M. & Phillips, P.C.B. 1994. Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1: 211-248.

Lux, T. 1996. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Applied Financial Economics* 6(6): 463-475.

Lux, T. 2001. The limiting extremal behaviour of speculative returns, an analysis of intradaily data from the Frankfurt Stock Exchange. *Applied Financial Economics* 11(3): 299-315.

Sarah, L. 2000. Estimation of Value at risk by extreme value methods. *Extremes* 3(2): 107-144.

Chin Wen Cheong*
Faculty of Information Technology
Multimedia University
63100 Cyberjaya Selangor, D.E.
Malaysia

Abu Hassan Shaari Mohd Nor
Faculty of Economic and Business
University Kebangsaan Malaysia
43600 UKM Bangi Selangor, D.E.
Malaysia

Zaidi Isa
Faculty of Science Technology
University Kebangsaan Malaysia
43600 UKM Bangi Selangor, D.E.
Malaysia

*Corresponding author; email: wcchin@mmu.edu.my

Received: 22 September 2008
Accepted: 11 November 2008