ANNALS OF ECONOMICS AND FINANCE **8-1**, 21–31 (2007)

Empirical Analyses of Industry Stock Index Return Distributions for the Taiwan Stock Exchange

Svetlozar T. Rachev*

School of Economics and Business Engineering, University of Karlsruhe

Emeritus

Department of Statistics and Applied Probability
University of California Santa Barbara

Stoyan V. Stoyanov

Chief Financial Researcher Fin Analytica, Inc.

Chufang Wu

Department of Applied Mathematics, National Donghua University of Taiwan

Frank J. Fabozzi

Finance and Becton Fellow, School of Management, Yale University

We study the daily return distributions for 22 industry stock indexes on the Tai-wan Stock Exchange under the unconditional homoskedastic independent, identically distributed and the conditional heteroskedastic GARCH models. Two distribution hypotheses are tested: the Gaussian and the stable Paretian distributions. The performance of the stable Paretian distribution is better than that of the Gaussian distribution. A back-testing example is provided to give evidence on the superiority of the stable ARMA-GARCH to the normal ARMA-GARCH.

Key Words: Stable distributions; ARMA-GARCH; Heavy tails; Volatility clustering; Value at risk.
JEL Classification Numbers: C13, G10.

* Prof Rachev gratefully acknowledges research support by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California, Santa Barbara and the Deutschen Forschungsgemeinschaft.

1529-7373/2007

1. INTRODUCTION

It is well known that financial returns are non-normal and tend to have fat-tailed distributions. Mandelbrot (1963) strongly rejected normality as a distributional model for asset returns, conjecturing that financial return processes behave like non-Gaussian stable processes (commonly referred to as "stable Paretian" distributions). The autoregressive conditional heteroskedastic (ARCH) models proposed by Engle (1982) and the generalized GARCH proposed by Bollerslev (1986) capture the extra probability mass in the tails. The appealing feature of incorporating conditional volatility is that it allows for a changing distribution over time. However, the distribution of conditional residuals is still not normal (Bollerslev, 1987). The implication for the commonly used risk measure Value-at Risk (VaR) is that risk is still underestimated at high quantiles for fat-tailed results. Moreover, GARCH models also fail to model the asymmetric effect of volatility, where negative return shocks generated by bad news have a larger thrust in increasing future volatility than positive return shocks caused by good news.

One innovation has focused on the power term by which the data are to be transformed. Ding, Granger and Engle (1993) introduced a generalized asymmetric version of the power ARCH (APARCH) model to capture the potentially asymmetric effects of return shocks on future volatility. To further enhance the robustness of the estimation results with respect to non-normality, the errors are considered to follow a t-distribution, called Student-APARCH. Huang and Lin (2004) analyzed the VaR for Taiwan stock data. They assumed the asset returns have fat tails and volatility clustering. At lower VaR confidence levels, the Normal-APARCH model is preferred. However, at high confidence levels, the VaR forecast obtained by the Student-APARCH model is more accurate.

Chiang and Doong (1999) used a generalized M-GARCH(1,1) process and found evidence to reject the hypothesis that the stock excess returns are independent of the real and financial volatilities. The stock excess returns are explained by the predicted volatility of macrofactors and the conditional standard deviation. The volatility of macrofactors consists of the volatilities arising from real (internal) and financial (external) shocks, whereas the time-series volatility is due to previous shocks. The stock excess return is associated with the volatility of macrofactors. The finance industry is more sensitive to a change in economic conditions and has been the leading industry on the Taiwan Stock Exchange (TSE) in the past decade.

In a study of the TSE, Ammermann (1999) found that the stocks trading exhibit nonlinearity and nonstationarity. To capture the full-sample nonlinear serial dependencies found within a number of financial time series, the Normal-GARCH, t-GARCH, and STAR (Student's t autoregressive) models were fitted and compared with the dynamic linear models. The inferences obtained varied from model to model, suggesting the importance of adequately accounting for nonlinear serial dependencies (and of ensuring data stationarity) when studying financial time series.

Rachev and Mittnik (2000) give a very detailed description on the stable Paretian models in finance. The stability property is highly desirable for asset returns. In the context of portfolio analysis and risk management, the linear combinations of different return series follow again a stable distribution. In fact, the Gaussian law shares this feature, but it is only one particular member of a huge class of distributions, which also allows for skewness and heavy tails.

In this paper, we study industry stock index return data with respect to: (1) non-Gaussian, heavy-tailed and skewed distributions, (2) volatility clustering (ARCHeffects), (3) temporal dependence of the tail behavior, and (4) short- and long-range dependence. Stable models allow us to generalize Gaussian-based financial theories to build a more general framework for financial modelling. Since asset returns exhibit temporal dependence, the conditional distributions become of interest. We study the daily return distributions for 22 industry stock indexes on the TSE under the unconditional homoskedastic independent, identically distributed (iid) and the conditional heteroskedastic GARCH (varying-conditional-volatility) cases. Two distribution hypotheses were tested: the Gaussian and the stable Paretian distribution. The stable Paretian distribution performed better than that of the Gaussian distribution.

In Section 2, we state the probability models and measures applied in this paper. In Section 3, the numerical analyses results are demonstrated, followed by a back-testing example in Section 4. The conclusion is provided in Section 5.

2. PROBABILITY MODELS

The class of autoregressive moving average (ARMA) models is a natural candidate for conditioning on the past of a return series. These models have the property that the conditional distribution is homoskedastic. Moreover, since financial markets frequently exhibit volatility clustering, the homoskedasticity assumption may be inadequate. On the contrary, the conditional heteroskedastic models, such as ARCH and the GARCH models, combining with an ARMA model, referred to as an ARMA-GARCH model, are common in empirical finance. It turns out that ARCHtype models driven by normally distributed innovations imply unconditional distributions which themselves possess heavier tails. However, many studies have shown that GARCH-filtered residuals are themselves heavy-tailed, so that stable Paretian distributed innovations ("building blocks") would be a reasonable distributional assumption.

A random variable X is said to have a stable distribution if there are parameters: $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\sigma \in [0,\infty)$, $\mu \in R$ such that its characteristic function has the following form:

$$\varphi_X(\theta) = \begin{cases} \exp\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta(sign\,\theta)\tan\left(\frac{\alpha\pi}{2}\right) + i\mu\theta\} & \text{if } \alpha \neq 1 \\ \exp\{-\sigma|\theta|(1 + i\beta\frac{2}{\pi}(sign\,\theta)\ln|\theta|) + i\mu\theta\} & \text{if } \alpha = 1 \end{cases}$$

In the general case, no closed-form expressions are known for the probability density and distribution functions of stable distributions. The parameter α is called the index of stability, which determines the tail weight or densitys kurtosis. The parameters β, σ, and μ are called the skewness parameter, scale parameter, and location parameter, respectively. Stable distributions allow for skewed distributions when $\beta \neq 0$; when β is zero, the distribution is symmetric around μ. Stable Paretian laws have fat tails, meaning that extreme events have high probability relative to the normal distribution, when $\alpha < 2$. The Gaussian distribution is a stable distribution, with $\alpha = 2$. (For more details on the properties of stable distributions see Samorodnitsky, Taqqu (1994).)

The general form of the ARMA(p,q)-GARCH(r,s) model is:

$$\begin{aligned} R_t &= C + \sum_{i=1}^{p} a_i R_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= \sigma_t \delta_t \\ \sigma_t^2 &= K + \sum_{k=1}^{r} \omega_k \varepsilon_{t-k}^2 + \sum_{l=1}^{s} \nu_l \sigma_{t-l}^2 \end{aligned}$$

where a_i, b_j, ω_k, ν_l, C, K are the model parameters, for $i = 1,\dots,p$, $j = 1,\dots,q$, $k = 1,\dots,r$, $l = 1,\dots,s$. δ_t's are called the innovations process and are assumed to be iid random variables which we additionally assume to be either Gaussian or stable Paretian. An attractive property of the ARMA-GARCH process is that it allows a time-varying volatility via the last equation in the above model.

We test the hypotheses in two cases. In the first case, we assume that daily return observations are iid. In the second case, the daily return observations are assumed to follow a GARCH(1,1) model. The first case concerns the unconditional homoskedastic distribution model while the second case belongs to the class of conditional heteroskedastic models.

For both cases, we verify whether the Gaussian hypothesis holds based on the Kolmogorov distance (KD):

$$KD = \sup_{x \in \mathbb{R}} |F_e(x) - F(x)|,$$

where $F_e(x)$ is the empirical sample distribution and $F(x)$ is the cumulative distribution function of the estimated parametric density and emphasizes the deviations around the median of the distribution.

For both the iid and the GARCH cases, we compare the goodness-of-fit for the Gaussian and the more general stable Paretian hypotheses. We use two goodnessof- fit measures for this purpose, the KD-statistic and the Anderson-Darling (AD) statistic. The AD-statistic accentuates the discrepancies in the tails and is computed as follows:

$$AD = \sup_{x \in \mathbb{R}} \frac{|F_e(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}}.$$

The data used in this study consist of the daily returns for 22 industry stock indexes (sectors) from the entire TSE. The industries are listed in the first column of Table 1. The sample period of this research spans January 1999 through December 2002. Industry stock returns are defined as the first difference in the log of daily indexes, $R(t) = \log(S(t)/S(t-1))$, where $S(t)$ is the value at t (the returns are adjusted for dividends).

3. MAIN RESULTS

There are several methods that can be employed for estimating the parameters of stable distributions. The most popular methods are Maximum Likelihood (ML), Fourier Transform (FT), and Fast Fourier Transform (FTT), see Rachev and Mittnik (2000), Rachev (2003). Only ML easily allows for estimation of the skewness parameter β; it is also the most accurate method. However it is not the fastest.

3.1. Unconditional homoskedastic iid model

In the simple setting of the iid model, we have estimated the values for the four parameters of the stable Paretian distribution using the ML method. Summary statistics of the various statistical tests and parameter estimates for the entire sample are provided in Table 1.

For the in-sample analyses, we use the standard Kolmogorov-Smirnov test based on the KD. We observe that 59.09% and 27.27% of the industry sectors for which normality is rejected at confidence levels 95% and 99%. On the contrary, 22.73% and 4.55% of the sectors for which the stable Paretian distribution is rejected at confidence levels 95% and 99%. Therefore we have evidence that the stable-Paretian hypothesis is rejected in much fewer cases, hence the stable Paretian distribution fits better than the normal distribution.

For every industry index in our sample, the KD in the stable Paretian case is below that in the Gaussian case. The same is true for the AD. The

TABLE 1.

The MLEs and KS test results for 22 industrial stocks, iid model.

Industry	α	β	σ	μ	KD normal	KD stable	AD normal	AD stable
Cement	1.9906	1.0000	0.0172	−0.0005	0.0516	0.0491	0.1189	0.1088
Foods	1.8690	0.1744	0.0123	−0.0006	0.0457	0.0291	0.1281	0.0682
Plastics	1.8390	0.2939	0.0150	0.0004	0.0392	0.0253	0.1315	0.0956
Textiles	2.0000	0.9777	0.0151	−0.0003	0.0348	0.0353	0.0754	0.0739
Ele. & Machinery	2.0000	0.1657	0.0137	−0.0004	0.0266	0.0283	0.1488	0.0690
Ele . Appliance & Cable	2.0000	0.9547	0.0178	−0.0007	0.0356	0.0357	0.0722	0.0723
Chemicals	2.0000	0.6171	0.0143	−0.0004	0.0277	0.0288	0.1630	0.0736
Glass and Ceramics	1.9821	1.0000	0.0159	−0.0004	0.0440	0.0386	0.1190	0.1296
Paper and Pulp	2.0000	0.9959	0.0170	−0.0004	0.0514	0.0520	0.1087	0.1098
Steel and Iron	1.8060	0.4584	0.0121	0.0004	0.0469	0.0219	0.2016	0.0885
Rubber	2.0000	0.9644	0.0171	−0.0003	0.0304	0.0312	0.0836	0.0844
Automobile	1.6070	0.2436	0.0122	0.0007	0.0644	0.0296	0.2122	0.1272
Electronics	2.0000	0.9996	0.0160	−0.0002	0.0434	0.0443	0.0927	0.0950
Construction	1.9707	1.0000	0.0162	−0.0011	0.0515	0.0433	0.1446	0.1319
Transportation	1.8686	0.4946	0.0151	0.0002	0.0592	0.0397	0.1528	0.0964
Tourism	1.7663	0.4540	0.0084	−0.0005	0.0540	0.0332	0.6864	0.0885
Wholesale and Retail	1.7960	0.1113	0.0108	−0.0004	0.0448	0.0244	0.2566	0.0800
Cement and Ceramics	1.9969	1.0000	0.0155	−0.0005	0.0489	0.0484	0.1080	0.1071
Plastics and Chemical	1.8899	0.2876	0.0139	0.0002	0.0423	0.0296	0.1621	0.0935
Electrical	2.0000	0.9998	0.0157	−0.0002	0.0429	0.0413	0.1000	0.1022
Finance	1.8247	0.5351	0.0135	0.0002	0.0574	0.0262	0.1634	0.0969
Others	1.8990	0.0183	0.0127	−0.0003	0.0374	0.0294	0.1070	0.0790
mean	1.9139	0.6248	0.0144	−0.0002	0.0446	0.0348	0.1608	0.0941
median	1.9764	0.5761	0.0150	−0.0003	0.0444	0.0322	0.1298	0.0942
Q1	1.8464	0.2892	0.0129	−0.0005	0.0378	0.0289	0.1073	0.0792
Q3	2.0000	0.9987	0.0160	0.0001	0.0515	0.0409	0.1628	0.1058
percentage of sectors rejected at 95%					59.09%	22.73%		
percentage of sectors rejected at 99%					27.27%	4.55%		

KD implies that for our sample there is a better fit of the stable Paretian model around the center of the distribution while the AD implies a better fit in the tails. The substantial difference between the AD computed for the stable Paretian model relative to the Gaussian model strongly suggests a much better ability for the stable Paretian model to forecast extreme events and confirms an already noticed phenomenon: the Gaussian distribution fails to describe observed large downward or upward asset price shifts. That is, in reality extreme events have larger probability than predicted by the Gaussian distribution.

3.2. Conditional heteroskedastic GARCH model

In this section we consider the GARCH(1,1) model for the 22 industry indexes daily return time series. The model parameters are estimated using the ML method assuming the normal distribution for the innovations. In this way, we maintain strongly consistent estimators of the model parameters under the stable Paretian hypothesis since the index of stability of the innovations is greater than 1, see Rachev and Mittnik (2000) and references therein. After estimating the GARCH(1,1) parameters, we computed the model residuals and verified which distributional assumption is more appropriate.

A summary of the computed statistics for the residuals of the GARCH(1,1) model is reported in Table 2. Generally, the results imply that the stable Paretian assumption is more adequate as a probabilistic model for the innovations compared to the Gaussian assumption.

We test the hypotheses for the conditional heteroskedastic GARCH(1,1) model. As reported in Table 2, 36.36% and 4.55% of the sectors for which stable distribution is rejected at confidence levels 95% and 99%. In contrast, as can be seen in Table 2, less than 5% of the industry indexes for which the stable distribution is rejected at confidence levels 95% and 99%. Once again, we have evidence that the stable-Paretian hypothesis is rejected in much fewer cases, hence the stable Paretian distribution fits better than the normal distribution.

We observe that the normal distribution is rejected in fewer cases in the GARCH case than in the iid case. A similar situation is observed for the stable distribution.

4. A BACK-TESTING EXAMPLE

In this example, an empirical comparison between the normal ARMA-GARCH (conditional homoskedastic, i.e. constant-conditional-volatility) and the stable ARMAGARCH (conditional heteroskedastic, i.e. varying-conditional-volatility) models is presented. We performed a back-testing analysis for the electrical industry, comparing the performance of the sim-

TABLE 2.

The MLEs and KS test results for 22 industrial stocks, GARCH(1,1) model.

Industry	α	β	σ	μ	KD normal	KD stable	AD normal	AD stable
Cement	1.9606	0.5767	0.6935	0.0072	0.0445	0.0375	0.1383	0.0821
Foods	1.9056	0.0966	0.6739	−0.0005	0.0366	0.0247	0.2550	0.0814
Plastics	1.9075	0.2629	0.6779	0.0018	0.0342	0.0253	0.1106	0.0861
Textiles	1.9998	−0.9314	0.7068	−0.0148	0.0293	0.0293	0.0796	0.0800
Elec. & Machinery	1.9855	−1.0000	0.6987	−0.0181	0.0217	0.0236	0.3045	0.0636
Elec Appliance & Cable	2.0000	0.9962	0.7068	−0.0111	0.0272	0.0264	0.0551	0.0536
Chemicals	1.9821	−1.0000	0.6987	−0.0236	0.0193	0.0209	0.2392	0.0560
Glass and Ceramics	1.9494	1.0000	0.6904	−0.0029	0.0383	0.0259	0.1232	0.0604
Paper and Pulp	2.0000	0.0527	0.7069	−0.0150	0.0459	0.0460	0.0960	0.0961
Steel and Iron	1.8406	0.4880	0.6473	0.0223	0.0409	0.0196	0.3650	0.0771
Rubber	2.0000	0.9952	0.7068	−0.0028	0.0251	0.0255	0.0836	0.0826
Automobile	1.7695	0.2465	0.6267	0.0267	0.0511	0.0304	0.3947	0.0972
Electronics	1.9934	−1.0000	0.7041	−0.0197	0.0366	0.0375	0.0973	0.1013
Construction	1.9407	0.8549	0.6888	0.0151	0.0446	0.0291	0.1202	0.0863
Transportation	1.9359	0.2575	0.6855	0.0070	0.0423	0.0368	0.1418	0.0855
Tourism	1.8682	0.6649	0.6605	0.0056	0.0504	0.0255	0.2295	0.0652
Wholesale and Retail	1.9285	−0.2314	0.6814	−0.0269	0.0338	0.0259	0.1806	0.0789
Cement and Ceramics	1.9551	0.2501	0.6904	−0.0041	0.0434	0.0386	0.1585	0.0805
Plastics and Chemical	1.9328	0.2064	0.6847	−0.0085	0.0339	0.0265	0.1376	0.0855
Electrical	1.9909	−1.0000	0.7029	−0.0221	0.0358	0.0383	0.1079	0.0892
Finance	1.9205	0.5158	0.6820	0.0129	0.0491	0.0345	0.1495	0.0896
Others	1.9591	−0.0153	0.6908	−0.0269	0.0318	0.0304	0.2550	0.0759
mean	1.9421	0.1039	0.6866	−0.0044	0.0371	0.0299	0.1710	0.0797
median	1.9494	0.2465	0.6904	−0.0041	0.0366	0.0291	0.1418	0.0814
Q1	1.9225	−0.1774	0.6816	−0.0173	0.0323	0.0255	0.1086	0.0762
Q3	1.9895	0.5614	0.7019	0.0066	0.0443	0.0362	0.2250	0.0862
percentage of sectors rejected at 95%					36.36%	4.55%		
percentage of sectors rejected at 99%					4.55%	0%		

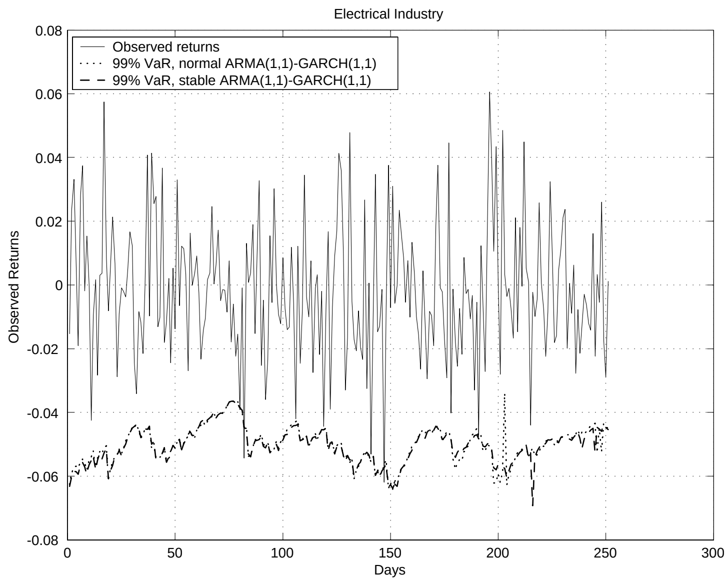

FIG. 1. Electrical industry out-of-sample comparisons between normal and stable ARMA(1,1)-GARCH(1,1) models.

pler ARMA(1,1)-GARCH(1,1) model from the ARMAGARCH family with stable and normal innovations using the VaR risk measure at 99% confidence level. The choice of $p = q = r = s = 1$ proved appropriate because the serial correlation in the residuals disappeared. The performance is compared in terms of the number of exceedances for the VaR measure; that is, how many times the forecast of the VaR is above the realized asset return. We verify if the number of exceedances is in the 95% confidence interval for the corresponding back-testing period.[1]

The VaR exceeding model comparison between normal and stable distributions is performed in Figure 1. In the last 250 days of the study period, the number of exceedances is 3 for the stable, 4 for the normal, and the 95% confidence bound is $[0, 5]$. The normal and stable ARMA-GARCH models both demonstrate good performance.

Since the normal distribution is a special case of general stable distribution, one may think that the stable model should produce better VaR estimates than the normal distribution (no matter what confidence level

[1]With 250 observations, the "exact" 95% confidence interval for the number of exceedances of the 99% VaR is $[0, 5.6]$, but since we need an integer for the upper bound, we round it to 5. The "exact" interval will be $[0, 5]$ if the confidence level is about 88.6%. Furthermore, the "exact" confidence intervals would be $[0, 4]$ and $[0, 3]$ if the confidence level is 65.6% and 24.4%, respectively.

is) all the time. But this is not true. The stable distribution contains four parameters, the normal distribution only two. Therefore in the out-of sample analysis, the forecasting properties of the two-parameter-model can beat that of the four-parameter-model in some cases.

5. CONCLUSIONS

We investigated the empirical daily return distribution properties of the 22 industry stock indexes on the TSE. Our in-sample analyses show that we can reject the Gaussian and the stable Paretian hypotheses at both the 95% and 99% confidence levels using the iid model. But the stable Paretian hypothesis is not rejected at both the 95% and 99% confidence levels in the GARCH(1,1) model, whereas the Gaussian is still rejected at the 95% and 99% confidence levels. Therefore the empirical evidence is overwhelming that stable laws are superior to the Gaussian distribution for the market; we do observe heavy tails from the empirical data. This finding is consistent with equity markets in other countries and in financial markets for other asset classes.

For the out-of-sample performance the empirical evidence suggests that the normal ARMA(1,1)-GARCH(1,1) model is reasonable, based on the number of exceedances observed. There is no contradiction in this finding because the normal ARMA(1,1)-GARCH(1,1) is already a heavy-tailed model. Nevertheless, we can say that the stable ARMA(1,1)-GARCH(1,1) is still better because the number of exceedances is 3, which is closer to the average value of about 2.5 exceedances. The fact that the normal ARMA(1,1)-GARCH(1,1) is doing practically as good as the stable ARMA(1,1)-GARCH(1,1) should be alarming, because the normal ARMA(1,1)-GARCH(1,1) model is unconditionally heavy-tailed. So the stable ARMA-GARCH has the flexibility to add a bit of heavy-tailedness, but in some cases that might not lead to a very significant improvement of the Gaussian ARMA-GARCH model.

REFERENCES

Ammermann, Peter A., 1999, Nonlinearity and overseas capital market : Evidence from the Taiwan stock exchange. Doctoral dissertation, Virginia Polytechnic Institute and State University.

Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

Bollerslev, Tim, 1987, A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* **69**, 542-547.

Chiang, Thomas C. and Shuh-Chyi Doong, 1999, Empirical analysis of real and financial volatilities on stock excess returns: Evidence from Taiwan industrial data. *Global Finance Journal* **10**, 187-200.

Chiang, Thomas C. and Shuh-Chyi Doong, 1999, Empirical analysis of real and financial volatilities on stock excess returns: evidence from Taiwan industrial data. *Global Finance Journal* **10**, 187-200.

Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 276-287.

Huang, Yu Chuan and Bor-Jing Lin, 2004, Value-at-risk analysis for Taiwan stock index futures: fat tails and conditional asymmetries in return innovations. *Review of Quantitative Finance and Accounting* **22**, 79-95.

Mandelbrot, Benoit, 1963, The variation of certain speculative prices. *Journal of Business* **26**, 394-419.

Rachev, Svetlozar and Stefan Mittnik, 2000, *Stable Paretian Models in Finance.* Chichester: John Wiley.

Rachev, Svetlozar, (Ed.), 2003, *Handbook of Heavy Tailed Distributions in Finance.* Amsterdam: Elsevier/North-Holland.

Samorodnitsky, Gennady and Murad T. Taqqu, 1994, *Stable Non-Gaussian Random Processes, Stochastic models with Infinite Variance.* New York: Chapman and Hall/CRC.