Financial Mathematics: Distribution of Stock Returns

Emmanuel Odai Okley
Frank Opoku
Oliver Faulhaber

Seminar in Statistics
Summer Term 2001
University of Kaiserslautern

Abstract

Distributon functions have become an important factor in evaluating the complex areas that are ubiquitous in today's world. Areas to which they are applied include financial markets, engineering and the social sciences. It is apparent that the aforementioned areas are in themselves so complex that very often relevant information concerning them can be obtained only by means of knowing to large accuracy their distribution functions.

Based on a data set consisting of daily prices of 15 DAX shares over a three-year period, we investigate the distribution functions of the stock returns and in particular the normality hypothesis. We introduce a generalized t-student distribution as a model which fits most of the data with high accuracy, but show also its limits and ideas how to overcome them.

Contents

1 Data Treatment **2**

2 Normal distribution **4**

3 T-student distribution **7**

4 Summary **10**

1 Data Treatment

The data consist of daily opening, closing, highest and lowest prices of 15 DAX shares – namely Allianz (ALV), BASF (BAS), Bayer (BAY), BMW, Commerzbank (CBK), Daimler (DAI), Deutsche Bank (DBK), Dresdner Bank (DRB), Hoechst (HFA), Mannesmann (MMW), RWE, Siemens (SIE), Thyssen (THY), Veba (VEB) and Volkswagen (VOW).

The observations are over a three-year period between January 1990 and December 1992, except ALV and RWE which start in March 1990 and May 1991 respectively. We focus on the closing prices and neglect the missing data in BAY.

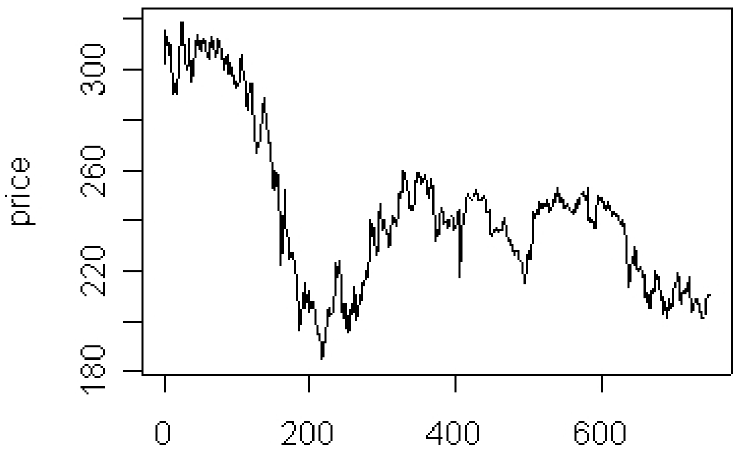

Figure 1: Plot of the raw data for BAS

Looking at the plot of the raw data (FIGURE 1) we realise that the mean depends on time. Hence we plot the sample autocorrelation functions (sacf) to support our claim (FIGURE 2.I). We observe a linear decrease of the sacf, a typical sign for a trend in the data. To remove this trend – and to allow a better comparison between the different stocks – we therefore look at the *returns*, defined by

$$R'_t = (X_t - X_{t-1})/X_{t-1} \tag{1}$$

This transformation eliminates a linear trend and normalizes the data. Instead of working directly with the returns we use the *log returns*, defined by

$$R_t = log(X_t) - log(X_{t-1}) \tag{2}$$

The advantage of the log returns lies in their property that the return over $n-1$ periods $R_t + \ldots + R_{t+n}$ can be rewritten as $log(X_{t+n}) - log(X_{t-1})$ (a necessary condition for a discrete representation of the underlying continuous-time process). This does not hold for the returns.

The justification for focusing on the log returns instead of the returns is the

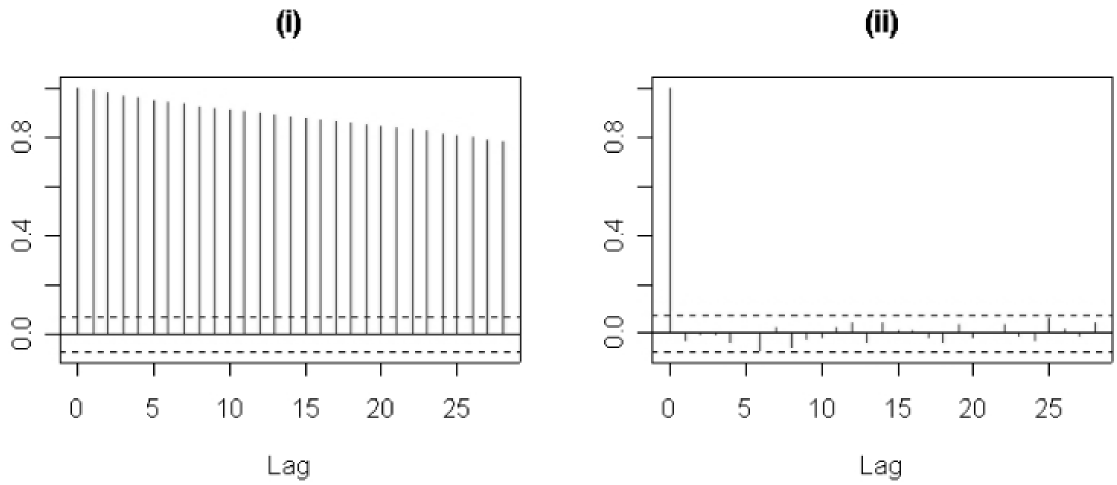

Figure 2: Sample autocorrelation function for (i) the raw data and (ii) the log returns of BAS

following: The returns can be rewritten as

$$R'_t = \frac{X_t}{X_{t-1}} - 1 \tag{3}$$

and the log returns as

$$R_t = log(\frac{X_t}{X_{t-1}}) \tag{4}$$

Now we replace $\frac{X_t}{X_{t-1}}$ by x. As the values for x are close to 1 (a share typically doesn't gain or lose more than 10% of its value within one day), the two functions $log(x)$ and $x-1$ are nearly identical – due to the Taylor Expansion of $log(x)$ at $x = 1$, the error introduced is $O(h^2)$). Thus we can from now on concentrate on the log returns.

A look at the sacf of the log returns (FIGURE 2.II) shows, that we successfully removed the trend and can now start with the analysis of their distribution.

Table 1: Kurtosis, skewness and result of the GOF tests for normal distribution

	Kurtosis	Skewness	χ^2_N
ALV	-1.018	12.995	43.21
BAS	-0.743	11.617	94.38
BAY	-0.264	8.595	91.56
BMW	-0.319	9.338	79.47
CBK	-0.870	12.542	71.00
DAI	-0.093	7.645	75.82
DBK	-0.435	10.126	92.74
DRB	-0.736	12.090	94.03
HFA	-0.485	10.338	95.80
MMW	-0.856	13.569	61.96
RWE	-0.059	10.415	92.63
SIE	-0.193	9.212	41.21
THY	-0.068	8.372	56.31
VEB	-0.460	11.411	82.39
VOW	-0.170	10.044	43.95

2 Normal distribution

In general it is assumed that stock returns are normally distributed and that has been the motivation to start with this distribution. We are going to check whether in reality this is true.

Histograms and the empirical densities of the raw data are plotted (FIGURE 4) and then fitted with the normal distribution function with sample mean and sample variance as parameters. The graphs show wide deviation from normality. To confirm this a few heuristical tests for normality are carried out. These are quantile-quantile (qq) plots, skewness and kurtosis values. The *qq plots* (FIGURE 3) show a obvious deviation from a straight line and thus from normality. It indicates that there is considerably more mass around the origin and in the tails than the standard normal distribution provides us. The estimated values for the *skewness* defined as

$$\frac{\sum_{k=1}^{N}(x_k - \overline{\mu})^3}{N\widehat{\sigma}^3} \tag{5}$$

and the *kurtosis* defined as

$$\frac{\sum_{k=1}^{N}(x_k - \overline{\mu})^4}{N\widehat{\sigma}^4} \tag{6}$$

confirm the leptokurtic characteristic of the distribution (TABLE 1). The kurtosis is far above the expected value of 3 and the slightly negative skewness might indicate even the presence of an asymmetry.

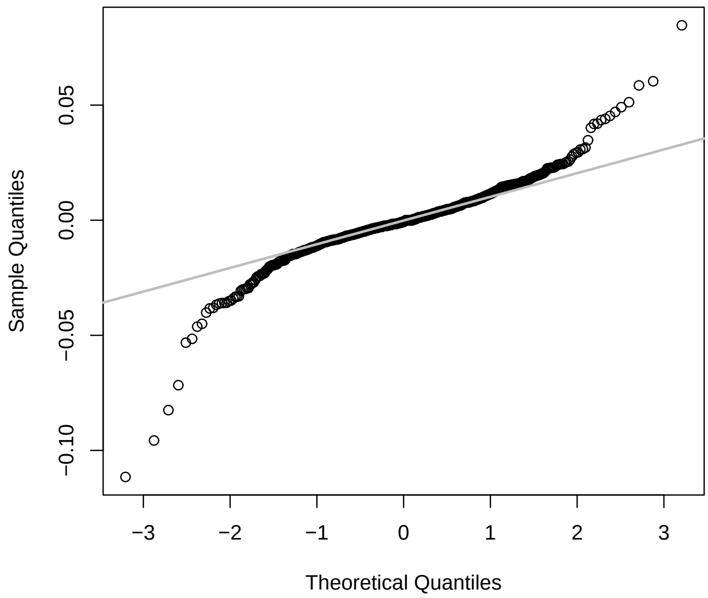

Figure 3: Quantile-quantile plot for the log returns of BAS

To formalize these heuristical results we conduct a χ^2 test, a quantitative yet powerful method for testing the goodness of fit. We test the hypothesis

$$H_0 : \text{the data has distribution F}$$

(where D is fixed up to p parameters, which have to be estimated), against the alternative

$$H_1 : \text{the data has not distribution F}$$

We then partition the data domain into d intervals and compare the number of expectations E_k in each interval (under H_0) with the observations O_k. The statistic

$$\chi^2 = \sum_{k=1}^{d} \frac{(O_k - E_k)^2}{E_k} \tag{7}$$

is chi-square distributed with $d-1-p$ degrees of freedom, hence we accept the hypothesis H_0 if

$$\chi^2 < c_{d-1-p,1-\alpha} \tag{8}$$

where $c_{d-1-p,1-\alpha}$ is the $(1-\alpha)$-quantile of a χ^2_{d-1-p}-distribution.

In applying this to our case, we test the hypothesis "normal distributed" and choose α to be 0.05. The unknown parameters μ and σ are estimated by means of sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}$ respectively. Ten intervals are chosen such that the first interval ranges from $-\infty$ to the 10% quantile of the

$N(\hat{\mu}, \hat{\sigma})$ distribution, the second one from the 10% to the 20% quantile and so on. Hence we expect 10% of all observations in each interval.

The χ^2 values for the stock returns of the 15 companies (as shown in TABLE 1) exceed the critical value $c_{7,0.95} = 14,08$ by far and thus we can reject the hypothesis for all stocks.

3 T-student distribution

As shown in the last chapter the stock returns are not normally distributed. The t-distribution is considered, as it provides heavier tails and thus can model the leptokurtic character of the data better.

Definition 1 *Let $X_0, X_1, \ldots, X_N$ be standard normal distributed, then Z is defined to be t-student distributed with N degrees of freedom if*

$$Z = \frac{\sqrt{N} \times X_0}{\sqrt{\sum_{i=1}^{N} X_i^2}} \tag{9}$$

Its density function is then given by

$$f(x) = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi \cdot N} \cdot \Gamma(\frac{N}{2})} \cdot (1 + \frac{x^2}{N})^{-\frac{N+1}{2}} \tag{10}$$

To adapt the distribution to the data, three generalizations have to be considered: A scaling parameter σ and a location parameter μ are introduced and the integer condition on N is relaxed. The resulting density function is given by

$$f(x) = \frac{\Gamma(\frac{N+1}{2})}{\sqrt{\pi \cdot N} \cdot \Gamma(\frac{N}{2}) \cdot \sigma} \cdot (1 + \frac{(x-\mu)^2}{N \cdot \sigma^2})^{-\frac{N+1}{2}} \tag{11}$$

As shown in FIGURE 4 this generalized t-distribution can provide a much better fit to the data, if the parameters are well chosen. To get good estimates for these parameters we numerically approximate the maximum likelihood estimates as given in TABLE 2. We then conduct another χ^2 test, now with the hypothesis "t-distributed". To get better results we increased the number of intervals to 25, resulting in a critical value $c_{21,0.95} = 32,67$. As shown in TABLE 2 the hypothesis can be rejected for only four stocks, namely ALV,CBK, DBK and DRB.

To support our results we conduct a Kolmogorov-Smirnov test as a second Goodness-of-Fit test. Again we test the hypothesis

$$H_0 : \text{the data has distribution F}$$

against the alternative

$$H_1 : \text{the data has not distribution F}$$

For this test the empirical cumulated distribution function $E(x)$ is compared with the distribution function $F(x)$ given in H_0. The statistic D is then defined as

$$D = max_x \mid E(x) - F(x) \mid \tag{12}$$

We accept H_0, if D is smaller than some $d_{n,\alpha}$, and reject it otherwise.

Although the KS test is said to be distribution-free and values for $d_{n,\alpha}$ can be easily looked up in tables (or, for $n > 30$, approximated by $1.36/\sqrt{n}$), this

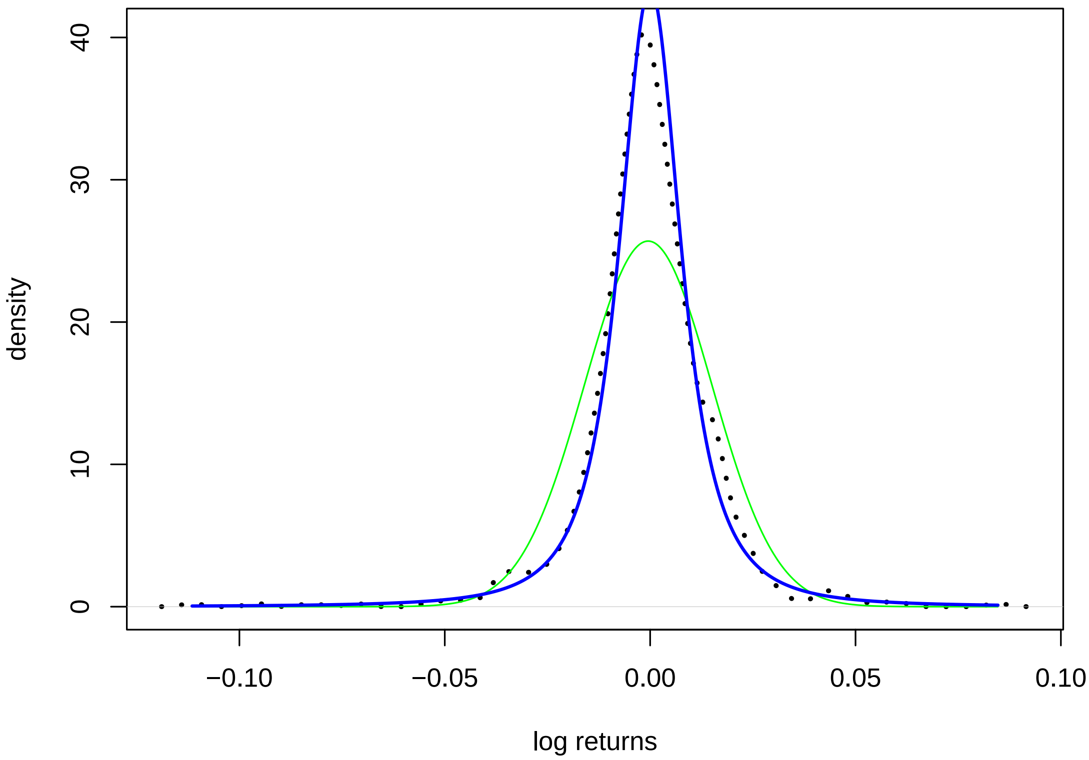

Figure 4: Fitted t-student distribution for BAS (black solid line) compared to the normal fitting (gray solid), the dotted line is the smoothed empirical density

Table 2: Result of the GOF tests for t-student distribution including the MLE for the three estimated parameters. The last coloumns represent the KS test statistics, the critical values for simple hypotheses and the estimated quantiles.

	μ	σ	N	χ^2_t	D	d	quant.
ALV	-0.000303	0.008744	3.81282	35.27	0.0411	0.0645	0.99
BAS	-0.000341	0.009088	2.74397	26.65	0.0292	0.0498	0.97
BAY	-0.000103	0.009140	2.78866	32.08	0.0268	0.0498	0.94
BMW	-0.000083	0.010507	3.09120	25.65	0.0333	0.0498	0.99
CBK	-0.000163	0.010369	3.36758	40.26	0.0371	0.0498	0.99
DAI	-0.000522	0.010376	2.93690	17.87	0.0239	0.0498	0.83
DBK	-0.000223	0.008264	2.85336	49.31	0.0236	0.0498	0.82
DRB	-0.000144	0.008370	2.90883	35.50	0.0272	0.0498	0.95
HFA	-0.000422	0.009647	2.77630	21.29	0.0254	0.0498	0.91
MMW	-0.000487	0.012574	3.67121	18.81	0.0278	0.0498	0.95
RWE	-0.000021	0.008091	2.52124	17.93	0.0227	0.0529	0.65
SIE	-0.000376	0.009166	3.98599	16.40	0.0142	0.0498	0.04
THY	-0.000355	0.012293	3.56287	30.34	0.0281	0.0498	0.96
VEB	0.000158	0.009161	2.98931	27.86	0.0220	0.0498	0.74
VOW	-0.001051	0.013841	4.12965	17.27	0.0265	0.0498	0.92

holds only for simple hypotheses. As we had to estimate the parameters μ, σ and N the $d_{n,\alpha}$ can no longer be looked up. (Using them would cause the test to be very conservative and would lead us to accept the hypothesis for all stocks.)

Hence we simulate the data set 1.000 times by means of bootstrapping, conduct a KS test in each case and compare D with the simulated statistics (cf. [2] for more details about this). The estimated quantiles obtained are shown in TABLE 2 . As we test on a 95% level, the hypothesis is rejected for those quantiles bigger than 0.95. This is the case for ALV, BASF, BMW, CBK and THY.

4 Summary

As we saw in the last section the t-student distribution provides a good fit in half of the cases, the remaining 7 data sets are only marginally rejected. One possibility for further research would be the slightly simplified hyperbolic distribution as suggested in [3] or even the generalized hyperbolic distribution introduced by Eberlein and Keller in [1]. An example for the former one with suitable parameters is given in FIGURE 5. Although there is a slight improvement compared to the previously fitted t-distribution which could even be big enough to make the hypothesis pass the χ^2 test, we leave further research about that open as it would go beyond the scope of this project.

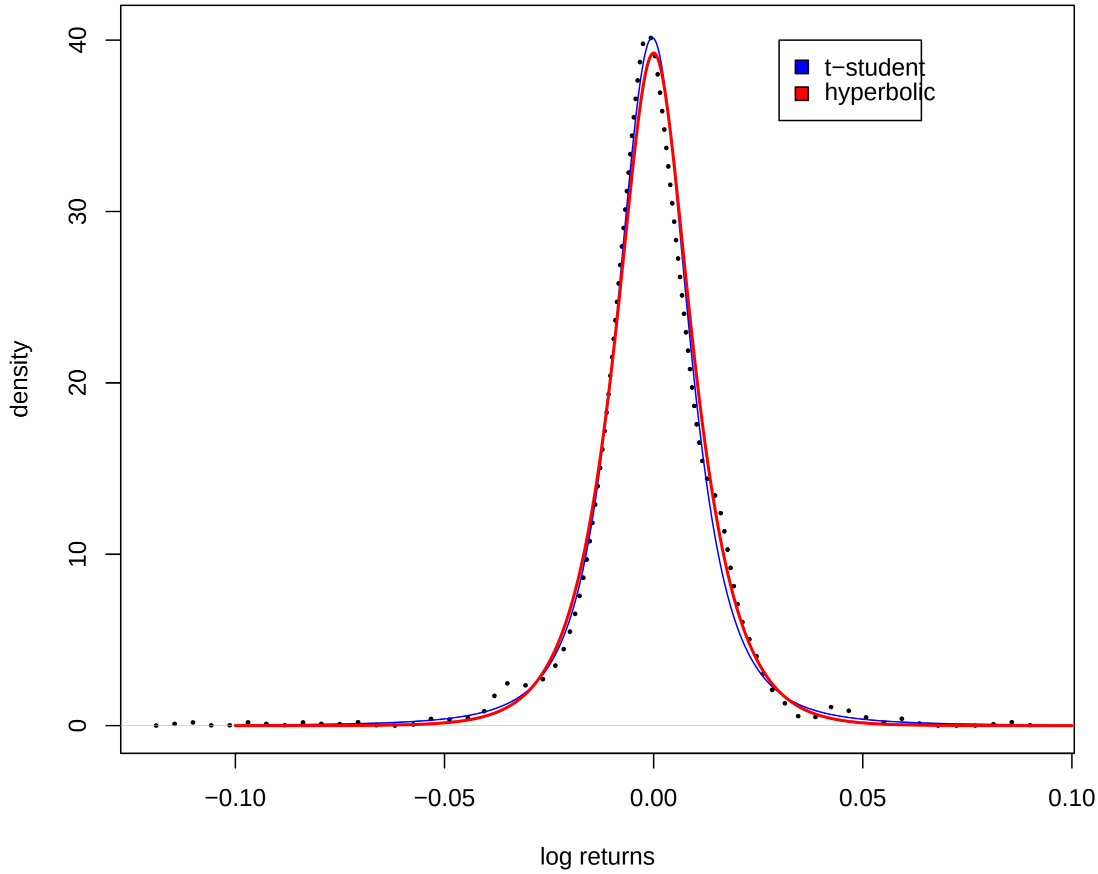

Figure 5: Fitted hyperbolic distribution for BAS

The last task we work on is finding criteria to group the 15 stocks by means of their distributions. As a first natural criterion we divided the stock returns into one group which passed the χ^2 test and a second group which didn't pass.

Surprisingly the elements of the latter group belong all to the financial sector (four banks and one insurance company).

As μ is for all data sets not significantly different from 0, we can restrict ourselves to comparing the values of σ and N. They range from around 0.008 to 0.012 and 2.5 to 4 respectively. If we plot these two parameters against each other, we see that the values of DRB and DBK are extremly close. The same similarity can be observed for two other "clusters": the two automotive companies (BMW and DAI) and the three companies of the chemical sector (BAS, BAY and HFA). All three groups are illustrated in Figure 6.

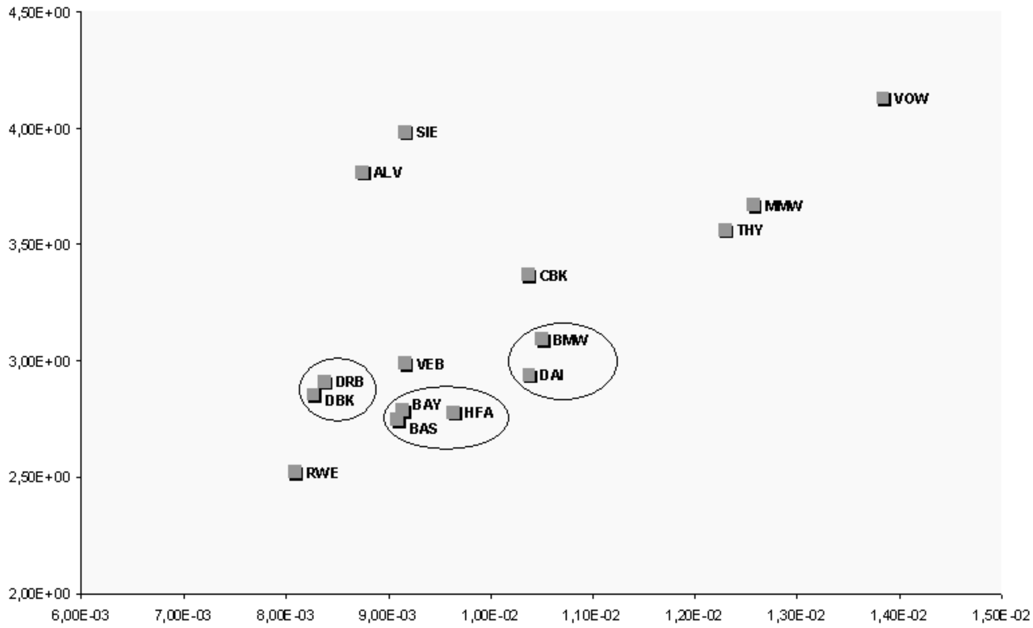

Figure 6: Plot of N vs. σ

References

[1] Ernst Eberlein and Ulrich Keller (1995), "Hyperbolic Distributions in Finance", Bernoulli 1/1995, 281-299.

[2] J. P. Romano (1988), "A Bootstrap Revival of Some Nonparametric Distance Tests", Journal of the American Statistical Association, vol. 83, 698-708.

[3] Frithjof Weber (2000), "Modellrisiko bei value-at-risk-Schaetzungen: eine empirische Untersuchung fr den schweizerischen Aktien- und Optionenmarkt", Schweizerische Zeitschrift fuer Volkswirtschaft und Statistik, 136(1), 99-121.