

ESTIMATING THE DISTRIBUTION OF SHARPE RATIOS*

Nadeem Aftab, Ingo Jungwirth, Tomás Sedliacik, Nadir Virk

June 30, 2008

Abstract

The Sharpe ratio is the most common measure for risk adjusted return of a financial asset. The question of its statistical distribution is of theoretical and practical importance and interest. We will use GMM and ML estimation methods to estimate the distribution of Sharpe ratios of 962 mutual funds and test on it. Sub-sampling gives further insight on the statistical patterns of the data and yields some first results with respect to estimates on the distribution. However, we could not find statistical evidence in favor of the normal distribution, the gamma distribution and the chi-square distribution on the basis of different estimation methods and different tests on distribution.

*This paper was written within the course of 40749 UK Generalized Method of Moments by Prof. Robert M. Kunst at the University of Vienna in the summer term 2008.

Contents

1	Introduction	3
2	Theoretical concepts of the distribution of Sharpe ratios	4
2.1	IID normality	4
2.2	IID generally (no normality of returns)	5
2.3	Stationarity and ergodicity (no IID requirements)	5
3	Literature review	6
4	The data	7
5	Setting up the GMM model	8
6	GMM estimation for normal distribution	10
6.1	Sub-sampling	10
6.2	White covariance matrix	12
7	ML estimation of parameters of distribution	13
8	Tests on statistical distribution	15
8.1	Jarque-Bera test statistic	15
8.2	Cramér-von Mises test statistic	15
8.3	The J-statistic	16
9	Results	16

1 Introduction

The purpose of this paper is to examine the probability distribution of Sharpe ratios using Generalized Method of Moments (GMM) and alternatively Maximum Likelihood (ML) estimation method. Although there have been other methods developed particularly for density estimation, such as Kernel density estimation, also known as the Parzen window method, or vector quantization, just to mention one example of various data clustering techniques used for this purpose, we decided to go an alternative way and use much more general techniques, in particular the GMM. Nevertheless we hope this does not disqualify this method in finding a good estimate for some density function, since it allows to be constructed in different forms in order to meet objectives for the optimal distribution parameters. For what regards the ML estimation, being a quite popular method for finding parameters of a probability distribution, it is more or less applied here for comparison reasons.

The Sharpe ratio is defined as the measure of excess return, usually over the so called risk-free rate, per unit of risk. It is mostly used to evaluate financial assets as some kind of a risk adjusted performance measure, since it combines both the average return of the asset and the corresponding volatility of the return for a particular time period. In contrast to the Jensen's alpha or the Treynor ratio¹ it accounts not only for the systematic but also for the non-systematic risk as well. The following expression gives the exact definition of the Sharpe ratio in mathematical terms

$$s_i = \frac{r_i - r_f}{\sigma_i}, \quad (1)$$

where r_i stands for the average return of an asset i , r_f for the average return of the risk-free asset, and σ_i represents the volatility of the return of asset i , i.e. standard deviation of the return of i .

For financial assets usually log-returns, log-differences of the subsequent market values, instead of simple returns are used, because of their continuous compounding nature. Considering that the simple returns are log-normally distributed, the log-returns are distributed normally and thus the average return, constructed as a sum of n normally distributed independent variables, assuming that the

¹Jensen's Alpha and Treynor Ratio are two alternative risk adjusted measures of an investment performance. To calculate these measures one regresses the excess returns of an asset on the excess returns of some selected benchmark index. Non-systematic risk is ignored that way.

market price evolution of the asset follows a Markov process, must then also be distributed normally. This must then also hold for the average excess return over the risk-free rate. The variance of the log-returns is a chi-square distributed variable. Based on these statements and additionally assuming independence among the average log-returns as well as the corresponding variances of different assets one may try to analytically evaluate the distribution of Sharpe ratios of some cross-sectional data. Unfortunately, the average return of an asset and its variance are not supposed to be two independent random variables, therefore the exact analytical way to find the correct population distribution form for cross-sectional Sharpe ratios becomes a quite difficult task to be handled.

In both, GMM and ML estimation one has to initially specify the form of the distribution function for which parameters are then estimated. In fact the moment conditions are set up for a pre-selected distribution form. Therefore, if we do not know the exact one, we will have to assume some particular distribution form which may seem plausible based on the characteristics of the sample data. Having formulated the corresponding moment conditions we will subsequently estimate the optimal parameters of the assumed distribution function.

2 Theoretical concepts of the distribution of Sharpe ratios

The distribution of returns has properties like normality, stationarity, and independently and identically distribution (IID). These properties or a combination of them lead to various derivations of the asymptotic distribution of Sharpe ratio (SR).

2.1 IID normality

This derivation was introduced by Jobson & Korkie in 1981. Their idea is that if the estimators for return (μ) and their variances (σ^2) are asymptotically IID normally distributed, and we are dealing with a finite sample, then we may write

$$\sqrt{T}(\hat{\mu} - \mu) \sim N(0, \sigma^2)$$

$$\sqrt{T}(\hat{\sigma}^2 - \sigma) \sim N(0, 2\sigma^4).$$

Now as we see that when ‘T’ increases without bound the probability distributions of $\sqrt{T}(\hat{\mu} - \mu) \sim N(0, \sigma^2)$ and $\sqrt{T}(\hat{\sigma}^2 - \sigma) \sim N(0, 2\sigma^4)$ approach normal distribution with mean zero and variance σ^2 and $2\sigma^4$, respectively.

This implies that the estimation error can be approximated by $var(\hat{\mu}) \approx \frac{\sigma^2}{T}$, and by $var(\hat{\sigma}^2) \approx \frac{2\sigma^4}{T}$, that is, given above equations as $T \rightarrow \infty$, the estimation error vanishes.

We see that the Sharpe ratio is a function of estimated return and estimated variance. Let $\hat{SR} = g(\hat{\mu}, \hat{\sigma}^2)$ then we may write $\sqrt{T}(\hat{SR} - SR) \sim N(0, V_{iid})$, where $V_{iid} = (\frac{\partial g}{\partial \mu})^2 \sigma^2 + (\frac{\partial g}{\partial \sigma^2})^2 \sigma^4$, where the values within brackets are the weights. Moreover, there would not be any covariance term as the estimated returns and variance are asymptotically independent.

Now we have to find weights in the above equation, and they are as follows:

$\frac{\partial g}{\partial \mu} = \frac{1}{\sigma}$ and $\frac{\partial g}{\partial \sigma^2} = \frac{-(\mu - R_f)}{2\sigma^3}$, which implies that the above expression changes to $V_{iid} = (\frac{1}{\sigma})^2 \sigma^2 + (\frac{-(\mu - R_f)}{2\sigma^3})^2 \sigma^4 = 1 + \frac{1}{2}SR^2$ and the expression for standard error (SE) of the Sharpe Ratio would be given by: $SE(\hat{SR}) = \sqrt{(1 + \frac{1}{2}SR^2)/T}$

2.2 IID generally (no normality of returns)

Merton (2002) purposed derivation of the distribution of Sharpe Ratio when the returns are only IID without being normally distributed. In this case we get following expression:

$\sqrt{T}(\hat{SR} - SR) \sim N(0, V_{iid}^G)$, where variance term is calculated by delta method and it is given by $1 + \frac{1}{2}SR^2 - SR\gamma_3 + SR^2(\frac{\gamma_4 - 3}{4})$, where $\gamma_3 = (\frac{\mu}{\sigma})^3$ and $\gamma_4 = (\frac{\mu}{\sigma})^4$. It is interesting to note that the expression for kurtosis and skewness of the returns just factored in the above model and leave the calculation still very simple.

2.3 Stationarity and ergodicity (no IID requirements)

Christie (2005) has used generalized methods of moments (GMM) – based on single system of moment restrictions that jointly test restrictions – and has derived the asymptotic distribution of SR under the

stationarity and ergodicity. This is a very important derivation as it brings the whole idea testable under the more realistic conditions of time-varying conditional volatilities, serial correlations, and non-iid returns. The most interesting fact is that if we work under more general conditions then Christie's derivation is identical to Merton's derivation of the distribution of SR. Following is the expression for variance in Christie's derivation:

$$var(\sqrt{T}\hat{SR}) = \left\{ \frac{SR^2\mu_4}{4\sigma^4} - \frac{SR[(R_t - R_{ft})(R_t - \mu)^2 - (R_t - R_{ft})\sigma^2]}{\sigma^3} + \frac{2(R_t - \mu)}{\sigma} - \frac{3SR^2}{4} \right\}$$

The whole idea of discussing the above asymptotic properties of the Sharpe Ratio is to provide statistical basis for efficient use of Sharpe Ratio inferentially, under very general and real-world conditions. This statistical discipline is worthwhile in a world of non-normal returns distributions (for example for the hedge funds) on the one hand, and exponential increase in the data recording and retrievability of these huge datasets on the other hand. Therefore, very large-sample properties of the Sharpe Ratio may have greater practical utility in the near future.

3 Literature review

The Sharpe (1966) portfolio performance ratio, explained as the measure of excess return, usually over the so called risk-free rate, per unit of risk, suffers from a methodological limitation because of the presence of random denominators in its definition, the distribution of Sharpe ratios is difficult to determine (Morey & Vinod (1999)). Thus, the approach may hamper portfolio selection decision making as two portfolios with pretty similar performance or point estimates may come up with varying estimation of risk.

Therefore, fitting a distribution to the cross-sectional data of Sharpe ratios may further require some implications on data in order to correctly specify the asymptotic distributional assumption on the data and for that reason correct specification of distribution is quite important in order to draw inferences from estimation methods. There are studies focusing on the question if the distribution

of Sharpe ratios can be studied in accordance to consumption-based models and if these models are able to explain the cross-sectional dispersion of Sharpe ratios (Paul Söderlind (2003)). But here in this paper we do not emphasize much on these kinds of sensitivities on implications of distributional assumptions and therefore ignore it for simplicity.

The purpose of this paper is to fit a distribution to data with GMM estimation technique and ML estimation for comparability. Generally, there is well reported literature that Sharpe ratios are skewed to the left, fat tailed and very sensitive to small samples (see Goetzmann et al. (2002)). Therefore, we find it reasonable to impose different distributional assumptions, for example, gamma distribution², normal distribution and chi square distribution to study the Sharpe ratio distribution.

4 The data

The cross-section sample data used for our estimation purposes are the Sharpe ratios of 962 mutual funds. The Sharpe ratios have been calculated for the five years period March 2003 - March 2008. For each fund we used its monthly log-returns, their volatility and the average monthly log-return of the 1-month-Euribor as the risk-free rate in order to calculate the corresponding Sharpe ratio for the specified time period. The resulting Sharpe ratios have finally been annualized. It is important to mention that the market values used to calculate the log-returns had to be adjusted in several ways. As market values the net market values were used³. In case of funds traded in a foreign currency the actual market values have been converted to euros according to the actual daily exchange rate of the corresponding date. Market values of funds which distributed capital to their holders during the analyzed time period have been adjusted by the corresponding amounts and recalculated in a way as if each distributed capital amount would have been currently accumulated in the fund and subsequently compounded by its actual return rate.

In order to get some initial impression it seems to be useful to have a look at some descriptive statistics of the data, especially the histogram. Figure 1 presents the most common descriptive statistics

²In order to impose over identification moment conditions for gamma distribution and chi square distribution the data must be non negative and for that reason we excluded the negative Sharpe ratios from the original sample so that we may better explore the distributional convergence of the Sharpe ratio data.

³Market values are used after the subtraction of costs in form of total expense ratio.

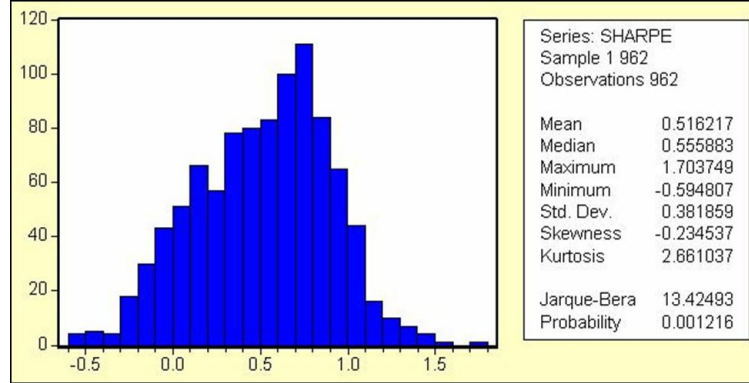


Figure 1:

Histogram and descriptive statistics of the sample of Sharpe ratios

for the underlying sample of Sharpe ratios. Visually it does not seem that the data is too far away from a normal distribution, although the Jarque-Bera test quite unambiguously rejects the null hypothesis of normality. In fact the sample seems to follow some kind of negatively skewed normal distribution which potentially the gamma distribution may be a good candidate for.

5 Setting up the GMM model

To formulate our problem in form of a GMM model it is essential to decide about the parameters to be optimized and the moment conditions involved in the overall objective function. Actually the choice of the parameters as well as the formulation of the moment conditions directly depend on which distribution form we assume to be true or at least plausible for the population of the analyzed random variable. The selected distribution form is then well defined by some specified parameters and its moments, centralized or not, can explicitly be expressed as functions of those parameters. Because of the lack of the analytically correct distribution form of Sharpe ratios we simply rely on the descriptive statistics of the sample data as well as the visual inspection of its histogram in order to select a plausible distribution form. In spite of the rejection of the normality hypothesis in the corresponding Jarque-Bera test and the slightly negative skewness measure we decided to choose the

normal and alternatively the gamma distribution⁴, as already mentioned in the data part, as reasonable candidates in this regard.

The normal distribution is fully determined by two parameters, μ as the mean and σ as the standard deviation. The gamma distribution is also completely defined by two parameters k , the shape parameter, and θ the scale parameter. Therefore, in both cases we will have to estimate two parameters in order to fulfill the moment conditions as close as possible. Since for two moment conditions the model would be just identified with optimal values of the population parameters being just the corresponding sample measures, we run the GMM with overidentifying moment conditions for each distribution form. The moment conditions are simply constructed as differences between sample moments and the population moments of a normal distribution with mean μ and standard deviation σ or alternatively of a gamma distribution with shape k and scale θ . By moments we mean the non-centralized moments and we will keep this definition throughout the whole paper. Expression (2) represents the corresponding moment conditions, where $g(\omega)$ is the density function with parameter vector ω to be estimated, n the number of observations and m is the number of moments, equivalently the number of moment conditions, used.

$$E[f_k(s_i, \omega)] = n^{-1} \sum_{i=1}^N s_i^k - \int_{-\infty}^{\infty} x^k g(x | \omega) dx = 0 \quad \forall k \in (1, m) \quad (2)$$

For the calculation of the population moments one does not have to evaluate the integral of the density function for x^k but can use the moment generating function (MGF) instead, see expression (3).

$$MGF = E(e^{tx}) \quad \text{where } x \sim G(\omega) \quad (3)$$

Differentiating the MGF in k -th order by t and setting $t = 0$ produces the k -th moment of the corresponding distribution, expression (4). The exact four moments for both the normal and the gamma distribution can be found in the Appendix.

⁴The use of the gamma distribution has a big advantage since the gamma distribution is quite general and in fact many well known distribution forms such as the chi-square distribution or the exponential distributions are only special cases of the gamma distribution.

$$\int_{-\infty}^{\infty} x^k g(x | \omega) dx = \frac{\partial^k MGF}{\partial^k t} | t = 0 \quad (4)$$

Finally, the optimal vector of parameters ω is the one which minimizes the GMM objective function, see expression (5). f represents the vector of functions f_1 to f_m .

$$\omega^* = \underset{\omega \in \Omega}{argmin} Q_n(\omega) \text{ where } Q_n(\omega) = \left\{ n^{-1} \sum_{i=1}^N f'(s_i, \omega) \right\} W_n \left\{ n^{-1} \sum_{i=1}^N f(s_i, \omega) \right\} \quad (5)$$

6 GMM estimation for normal distribution

6.1 Sub-sampling

We imposed the theoretical population moments to sample moments in order to get the GMM estimates under all visually and theoretically backed distributional assumptions. However, in this section we will only present the normal distributional moment conditions on the sample data. Additionally, we also decided to divide the data into different sub samples and henceforth extracted three more samples of observations 120, 240 and 480 from the original sample of 962 mutual funds through random simulation where for each sample each observation was allowed to enter the sub sample only once. The motive for sub-sampling is that varying sample size might provide some insight to Sharpe ratio's distribution under normality assumption and may be able to achieve better distributional specification under enlarging the sample.

We got GMM estimates for normal distribution for all samples. The null hypothesis, which is that the over identification conditions are satisfied, were rejected for all the samples and for all “over identified” conditions (we estimated all the samples for two and one over identifying moment condition).

In Figure 2, we can draw some first conclusions from the descriptive statistics of sub samples. For the full sample, normality is rejected by the Jarque-Bera test and the smallest sample cannot reject the null in the data and is some what strange with literature. Beside for the smallest sample, the normality null is rejected on the basis on skewness (negative) since the kurtosis is always less than three and also the sub samples mean and volatility are approximately nearing the full sample respective statistics. In

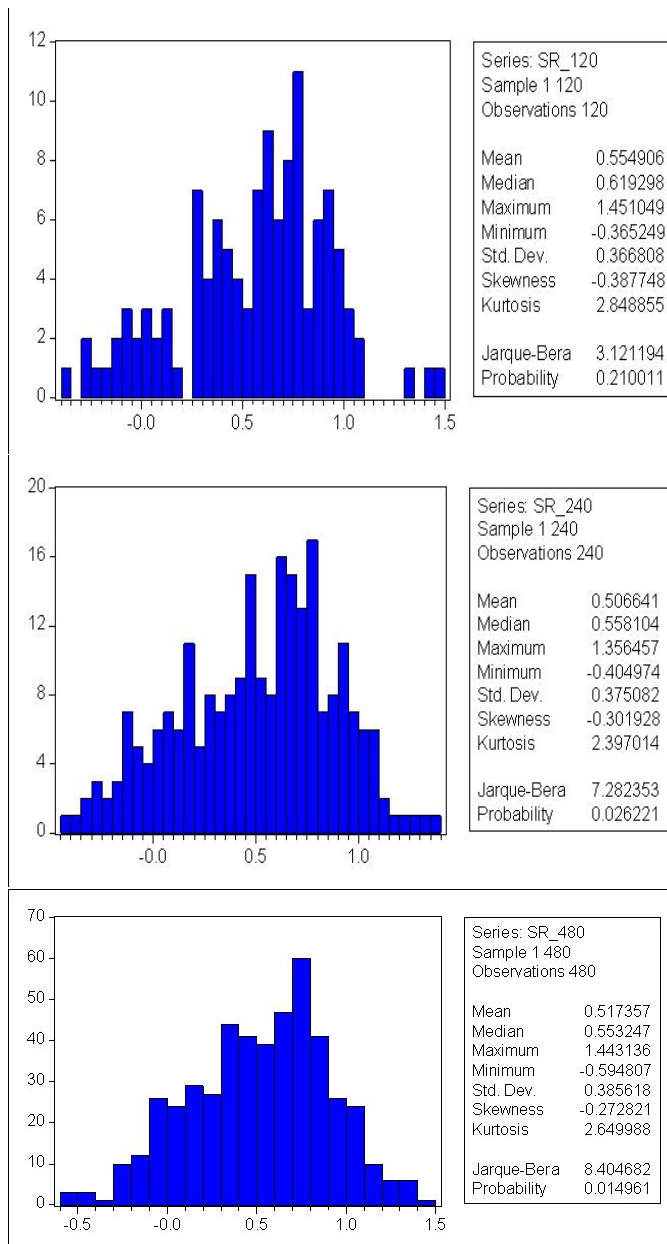


Figure 2: Histograms for the sub-samples

	sr_962		Sr_480		sr_240		sr_120	
	4mc	3mc	4mc	3mc	4mc	3mc	4mc	3mc
C(1)	0.514	0.373	0.508	0.361	0.507	0.336	0.560	0.354
Standard Error	0.012	0.008	0.017	0.011	0.024	0.014	0.033	0.021
C(2)	0.510	0.380	0.512	0.375	0.478	0.374	0.551	0.362
Standard Error	0.012	0.008	0.017	0.011	0.022	0.014	0.033	0.023

Table 1: GMM parameter estimates for normal moment conditions

all the sub samples the distribution appears marginally skewed to the left implying that large negative returns are more likely than the large positive returns.

If we compare the descriptive statistics and the GMM estimates from Table 1 of the entire sub sample groups it is pretty straight forward that they are very similar. This implies that under the normality assumption the GMM parameter estimates correspond to ML estimates, that is sample mean and sample volatility. Also the parameters for all sub samples are statistically consistent. The JB-test and the J-test also rejects the over identification condition null hypothesis for imposed two and one moment conditions. Its comparison with other distribution assumption, ML-estimates and different tests on distribution will be presented in detail in the next section. It is also worth mentioning the fact that we estimated these parameters with GMM cross section (White covariance), which corrects for heteroskedasticity in the covariance matrix. In case of heteroskedasticity in the error term, this may lead to consistent but inefficient parameter estimates.

6.2 White covariance matrix

The parameter covariance matrix estimator applies the assumption of independent and not identical distribution on residuals and therefore takes care of heteroskedasticity in the fixed regressors along with the case in which observations are obtained not from a controlled experiment (as the fixed regressor assumption requires) but rather from a stratified cross-section. White (1980) covariance matrix estimator applies the following assumptions besides (INID),

2.a) There exist positive finite constants δ and Δ such that for all i ,

	sr_963		sr_480		sr_240		sr_120	
	4mc	3mc	4mc	3mc	4mc	3mc	4mc	3mc
J-Stat	23.43749	12.32159	24.27216	9.144	23.91624	8.328	5.112	5.166
p- value	0.0001	0.0004	0.0001	0.0025	0.0001	0.0039	0.08	0.08

Table 2: J-statistic for normal moment conditions

$$E(|\epsilon_i^2|)^{1+\delta} < \Delta \text{ and } E(|X_{ij}X_{ik}|)^{1+\delta} < \Delta \quad (j, k = 1, \dots, K).$$

2.b) $\bar{M} = n^{-1}\Sigma X_i'X_i$ is nonsingular for (all) n sufficiently large, such that $\det(\bar{M}) > \delta > 0$.

3.a) There exist positive finite constants δ and Δ such that for all i ,

$$E(|X_{ij}X_{ik}|)^{1+\delta} < \Delta \quad (j, k = 1, \dots, K);$$

3.b) The average covariance matrix $\bar{V} = n^{-1}\sum_{i=1}^N E(|\epsilon^2 X_i'X_i|)$ is nonsingular for n sufficiently large, such that $\det(\bar{V}) > \delta > 0$.

For a detailed review of covariance matrix estimator see White (1980).

The reported J-statistics rejects normality for the over identification case in all samples. Only for the smallest sample the normality is rejected at a 90% confidence interval, otherwise it is rejected at 99% for all samples. We also estimated the GMM (HAC) for normal distribution which were pretty close to sample mean and sample volatility. The selection of different kernel methods and bandwidths came up with similar results as GMM cross-section (White covariance). There is not any impression for non-rejection of normality and the results for all samples and all moment specification are provided in the Appendix. Because of the rejection of normal distributional assumptions under GMM estimation we try some other distributional assumptions which will be done in the next section.

7 ML estimation of parameters of distribution

The ML estimation procedure for the parameters of a normal distribution is straight forward. It is the value of μ and σ^2 that would generate the observation sample Y (which is assumed to be normal distributed, i.e. $Y \sim N(\mu, \sigma^2)$) most likely. Thus, the ML estimation requires an assumption on the

distribution of the population. The probability density function (pdf) of the normal distribution is $f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$, $y \in \mathbb{R}$. Using the density function of the normal distribution with (μ, σ^2) , the log-likelihood function is defined as

$$\ell(\mu, \sigma^2) = -\frac{n}{2}(\ln 2\pi - \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu)^2.$$

The log-likelihood function is maximized by setting the first derivatives with respect to μ and σ^2 to zero:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \mu) := 0 \quad \frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (y_i - \mu)^2 := 0.$$

The ML estimators \bar{x} and s^2 for μ and σ^2 respectively are thus defined as

$$\bar{x} = \sum_{i=1}^N y_i \text{ and } s^2 = n^{-1} \sum_{i=1}^N (y_i - \bar{x})^2.$$

Assuming a gamma distribution of the sample, i.e. $Y \sim \Gamma(k, \theta)$, with the pdf $f(y; k, \theta) = \frac{y^{k-1} e^{-y/\theta}}{\theta^k \Gamma(k)}$ for $y > 0$ and $k, \theta > 0$ the log-likelihood function is defined as

$$\ell(k, \theta) = (k-1) \sum_{i=1}^N \ln(y_i) - \theta^{-1} \sum_{i=1}^N y_i - Nk \ln(\theta) - N \ln \Gamma(k).$$

Because of the use of the natural logarithm in the formula, negative y_i have to be excluded from the sample when applying the ML estimation. As seen in Choi and Wette (1969) there is no closed form solution of the parameters. Therefore, a numerical solution has to be found using an iterative algorithm. The algorithms provided by software packages include the Berndt-Hall-Hausman algorithm presented in Berndt et al. (1974) and the Levenberg-Marquardt algorithm (see Marquardt (1963) and Levenberg (1944)).

Assuming $Y \sim \chi^2(k)$ with $f(y; k) = \frac{0.5^{0.5k}}{2^{0.5k} \Gamma(0.5k)} y^{0.5k-1} e^{-0.5y}$ being the pdf of the Chi-square distribution the log-likelihood function becomes

$$\ell(k) = (0.5 \cdot k - 1) \sum_{i=1}^N \ln(y_i) - N \cdot 0.5k \ln(2) - N \ln \Gamma(0.5k) - 0.5 \sum_{i=1}^N y_i.$$

The parameter k can again be obtained by an iterative algorithm

8 Tests on statistical distribution

8.1 Jarque-Bera test statistic

The Jarque-Bera Test statistic allows to test on normal distribution for a sample of N observations. It measures the difference of the skewness and kurtosis of the sample with those from the normal distribution, i.e.

$$Jarque - Bera = \frac{N - k}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right)$$

where K represents the sample kurtosis, S represents the sample skewness and k stands for the number of estimated coefficients. Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is χ^2 distributed with 2 degrees of freedom.

8.2 Cramér-von Mises test statistic

The Cramér-von Mises test allows to test on statistical distributions of a sample. In a first step the distribution parameters are estimated by ML procedure derived from the underlying pdf of the distribution to be tested. In the second step the test statistic is calculated as the quadratic difference between the estimated empirical cumulative distribution function (cdf) F_n and the true theoretical cdf F . The null hypothesis is that the assumed distribution was reasonable (i.e. that the squared difference of both cdf is sufficiently small).

$$w^2 = \int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 dF.$$

Anderson (1962) provides the probability distribution and a corresponding table for levels of sig-

Assumption	ML-estimation*		# Mom. cond.	GMM-estimation**		p-value of test		
	Param(1)	Param(2)		Param(1)	Param(2)	Jarque-Bera	Cramer-von Mises	J-test
Normal distribution	0.516	0.382	4	0.513	0.373	0.001	0.00	0.00
			3	0.510	0.379	-		
Chi-square (sample+const.)	1.423	-	4	1.402	-	-	0.00	0.00
Gamma (sample+const.)	0.003	201.34	4	-3.143	2.105	-	0.00	0.00
* Manquardt								
** Cross section (white coef.)								

Table 3:

nificance. The Cramér-von Mises test can not only be taken out on basis of ML estimation methods. The estimated empirical cdf F_n could also be obtained by GMM.

8.3 The J-statistic

The J-statistic is derived from the GMM-estimators and the corresponding weighting matrix as follows:

$$J_n = nQ_n(\hat{\theta}_n) = n^{-1}u(\hat{\theta}_n)'Z\hat{S}^{-1}nZ'u(\hat{\theta}_n).$$

A simple application of the J-statistic is to test the validity of overidentifying restrictions. Under the null hypothesis that the overidentifying restrictions are satisfied, the J-statistic is asymptotically χ^2 – distributed with degrees of freedom equal to the number of overidentifying restrictions.

9 Results

Table 3 shows that ML- and GMM-estimation yield fairly similar coefficients for the normal distribution. Also differences between the GMM-estimators with four and three moment conditions are minimal. This implies that for our sample over identification does not yield any advantage. However, the test on normal distribution is rejected on the basis of both estimation methods. Results for the Chi-square distribution are broadly in line for both estimation methods. Interestingly, estimators for the gamma distribution differ manifestly for GMM and ML. However, on basis of the tests in use no statistical evidence can be found in favor of these distributions.

Among other factors, problems in finding an adequate distribution might be caused by a reporting bias and by the survival bias. Actually many weak performing funds are canceled and fund managers

tend to be proud on their successful products while they are discreet about their bad ones. The question of how Sharpe ratios are distributed remains an interesting and vivid field of research.

References

- [1] Anderson, T. W. (1962): "On the Distribution of the Two-Sample Cramer-von Mises Criterion", The Annals of Mathematical Statistics 33, #3, p1148-1159.
- [2] Berndt, E., Hall, B., Hall, R. and Hausman, J., (1974): "Estimation and Inference in Nonlinear Structural Models", Annals of Social Measurement, Vol. 3, 653-665.
- [3] Choi, S.C. and Wette, R. (1969): "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias", Technometrics, 11(4) 683-690.
- [4] Christie, S. (2005): "Is the Sharpe Ratio Useful in Asset Allocations?", MAFC Research Paper Nr.1, Applied Finance Center, Macquarie University.
- [5] Goetzmann, W.N., Ingersoll, J.E. Jr., Spiegel, M.I. and Welch, I. (2002): "Sharpening Sharpe Ratios", Yale School of Management Working Papers , ysm 273, Yale School of Management.
- [6] Jobson, J. and Korkie, B. (1981): "Performance Hypothesis testing with Sharpe Ratio and Treynor Measures", The Journal of Finance, Vol 36(4), 889-908.
- [7] Levenberg, K. (1944): "A Method for the Solution of Certain Non-Linear Problems in Least Squares", The Quarterly of Applied Mathematics 2, 164-168.
- [8] Marquardt, D. (1963): "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", SIAM Journal on Applied Mathematics 11, 431-441.
- [9] Morey, M.R. and Vinod, H.D. (2001): "A Double Sharpe Ratio", Advances in Investment Analysis and Portfolio Management, Vol. 8, JAI-Elsevier Science pp. 57-65.
- [10] Opdyk, J.D. (2007): "Comparing Sharpe Ratios: So Where are the p-values?", Journal of Asset Management, 8(5).

- [11] Sharpe, W.F. (1966): "Mutual Fund Performance", The Journal of Business, Vol. 39, No. 1, 119-138.
- [12] Söderland, P. (2004): "C-CAPM and the Cross-Section of Sharpe Ratios", Working Paper Series In Finance, WP 6.
-

Appendix

First four population moments of a normal distributed variable with mean μ and standard deviation σ :

$$m_1 = \mu$$

$$m_2 = \mu^2 + \sigma^2$$

$$m_3 = 3\mu\sigma^2 + \mu^3$$

$$m_4 = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4$$

First four population moments of a gamma distributed variable with scale parameter θ and shape parameter k :

$$m_1 = \theta k$$

$$m_2 = \theta^2 k(k+1)$$

$$m_3 = \theta^3 k(k+1)(k+2)$$

$$m_4 = \theta^4 k(k+1)(k+2)(k+3)$$

$$m_j = \theta^j \prod_{i=0}^{j-1} (k+i)$$

Since the chi-square distribution is a special case of the gamma distribution with $\theta = 2$ and $k = \nu/2$ where ν represents the number of degrees of freedom of the chi-square distribution, substituting this to the moments of the gamma distribution automatically produces the following first four moments of a chi-square distributed variable with ν degrees of freedom:

$$m_1 = \nu$$

$$m_2 = \nu(\nu + 2)$$

$$m_3 = \nu(\nu + 2)(\nu + 4)$$

$$m_4 = \nu(\nu + 2)(\nu + 4)(\nu + 6)$$

$$m_j = \prod_{i=0}^{j-1} (\nu + 2i)$$

	HAC Time Series						HAC Time Series (Prewhitening by VAR (1))					
	Bartlett			Quadratic			Bartlett			Quadratic		
	Fixed	Variable	Andrews	Fixed	Variable	Andrews	Fixed	Variable	Andrews	Fixed	Variable	Andrews
J-Stat*963	12.92	7.21	10.64	11.57	8.26	11.00	11.35	7.06	12.76	11.25	8.25	14.16
p-value	0.00	0.03	0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.00	0.02	0.00
J-Stat*480	19.08	18.16	24.18	17.57	18.69	24.39	19.26	18.29	25.93	17.60	18.75	25.92
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
J-Stat*240	20.62	17.09	22.59	20.29	20.29	22.38	20.04	16.82	20.40	20.06	20.29	20.41
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
J-Stat*120	3.90	4.90	5.12	3.59	3.28	5.14	3.81	4.67	4.60	3.57	3.28	4.60
p-value	0.14	0.09	0.08	0.17	0.19	0.08	0.15	0.10	0.10	0.17	0.19	0.10

Table 4:
GMM estimation with two over identifying conditions

	HAC Time Series						HAC Time Series (Prewhitening by VAR (1))					
	Bartlett			Quadratic			Bartlett			Quadratic		
	Fixed	Variable	Andrews	Fixed	Variable	Andrews	Fixed	Variable	Andrews	Fixed	Variable	Andrews
J-Stat*963	6.39	3.17	4.58	6.39	3.68	4.88	5.69	3.12	6.71	5.39	3.68	7.67
p-value	0.04	0.21	0.10	0.04	0.16	0.09	0.06	0.21	0.03	0.07	0.16	0.02
J-Stat*480	8.44	8.05	9.71	8.16	9.55	9.82	8.55	8.12	10.00	8.18	9.64	10.00
p-value	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01
J-Stat*240	7.95	7.79	8.00	8.28	7.93	7.77	7.70	7.61	7.19	8.20	7.79	7.19
p-value	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01
J-Stat*120	3.84	4.84	4.67	3.55	4.51	3.37	3.74	4.32	4.31	3.53	3.27	4.30
p-value	0.05	0.03	0.03	0.06	0.03	0.07	0.05	0.04	0.04	0.06	0.07	0.04

Table 5:
GMM estimation with one over identifying condition