CHRISTIAN JOCHUM*

Robust Volatility Estimation

1. Introduction: Why Volatility Forecasting?

For two reasons the exact estimation of the volatility of a financial asset is of crucial importance to the investor. For any application in option pricing, which relies on the formulation proposed by BLACK and SCHOLES (1973), the volatility is the factor ultimately defining the price of the derivative instrument. Secondly, portfolio management and hedging strategies recognize the variance as the variable which describes the risk of the asset.

While, in the standard deviation, statistical theory provides a highly efficient way to estimate the dispersion and consequently the variance of a variable, this measure is depending crucially on the assumption that the underlying distribution is normal in nature. Even slight deviations from this assumption make the standard deviation highly unreliable in the sense that the estimated value will fluctuate considerably around the true value. This of course makes it a hazardous tool for pricing and hedging decisions. That financial markets data do deviate from the ideal model of normality has first been shown by FAMA (1965) and can nowadays be seen as a well established empirical fact. The deviation from normality is characterized by three regularities of return series: (1) Asset returns tend to be leptokurtic (heavy tailed), meaning that one encounters more "far off" observations than the normal distribution would predict. (2) Asset returns show volatility clustering, in the sense that the size of the current return change influences the size of the next change. (3) Finally, in stock price returns evidence for a leverage effect is found, since negative returns tend to be associated with higher levels of volatility.

Among the models proposed to account for and make use of these characteristics, the by far most successful one has been the linear ARCH(q) model (ENGLE, 1982) and a wide variety of extensions. The core idea behind this approach is the insight that even though innovations in market returns are not correlated they are not independent either. This dependence stems from the fact that markets show volatility clustering, which again is closely related to the existence of heavy tails.

If one turns back to the start of the story, namely the problem of volatility estimation in the face of continued deviations from the normal distribution assumption, a second class of models offers some promising insights. Robust estimators of scale

* For discussion during the preparation of the paper the author is much obliged to Wolfgang Eckert, Lars Feld, Cedric Kohler, Gebhard Kirchgaessner and Marcel Savioz. The paper has further benefited from the comments Winfried Stier and Heinz Zimmermann (the referees). Christian Jochum, SIASR-HSG, University of St. Gallen, Dufourstrasse 48, CH-9000 St. Gallen, Switzerland, email: Christian.Jochum@siasr.unisg.ch

(volatility) can be seen as a means to estimate the dispersion of a variable, if the variable's distribution is deviating from normality. Robust estimation describes empirical distributions, which are close to an ideal parameterized one.
LAX (1985) shows that there exists a variety of dispersion estimators, which show considerably better behavior than the standard deviation, when long tailed symmetric distributions are considered. His analysis indicates that robust techniques can minimize the influence of deviations from normality on the performance of volatility estimators. Thus using these estimators should improve pricing and hedging procedures implemented on the basis of volatility estimates.[1]
The structure of this paper is as follows: Section 2 describes the different approaches to volatility estimation and the main characteristics. Section 3 presents an application by modeling the volatility of the Swiss Performance Index (SPI). The forecasting performance is then investigated in section 4 and finally section 5 offers some conclusions. An Appendix presents an extended description of the robust volatility estimators.

2. Volatility Estimation

2.1 ARCH Modeling in Finance

A basic autoregressive conditional heteroscedastic (ARCH) model can be described as follows: The underlying process y is stochastic with a conditional mean μ

$$\mu_t(\theta_0) \equiv E_{t-1}(y_t),$$

where θ_0 is a set of parameters. The expected value for y_t, $E_{t-1}(y_t)$, equals the conditional mean of the process, which allows to define the error term as

$$\varepsilon_t(\theta_0) \equiv y_t - \mu_t(\theta_0).$$

The conditional variance for the error process $\{\varepsilon_t\}$ equals the conditional variance for the $\{y_t\}$ process.
ENGLE (1982) suggests the following parameterization to describe the conditional variance:

$$\sigma_t^2 = \omega + \sum_{i=1}^{n} \alpha_i \varepsilon_{t-i}^2 \qquad \text{(1a)}$$

For this process to be stationary the effect of an innovation has to reduce to zero as t increases.
The general idea behind this very basic ARCH model and all the following models is that, although the innovations ε_t are uncorrelated, it can be shown that large (small) innovations are systematically followed by large (small) innovations. This introduces autocorrelation in ε_t^2 and can be used to describe the clustering effects commonly found in financial markets data.
Among the large number of ARCH type models[2] notable extensions of the original model are the GARCH, the GARCH-T and the EGARCH model. The GARCH-T takes explicitely account of heavy tails by assuming that the underlying distribution is a t-distribution, while the EGARCH models the leverage effect found in financial markets. The relevant model specifications for the GARCH and the GARCH-T model are:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \qquad \text{(1b)}$$

and for the EGARCH model

$$\log(\sigma_t) = \omega + \beta \cdot \log(\sigma_{t-1}) + \gamma \cdot \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}}} - \sqrt{2/\pi} \right], \qquad \text{(1c)}$$

with γ measuring the size of the leverage effect.

2.2 Robust Estimation

What is robust statistics?

Robust statistics is concerned with the problem that many assumptions commonly made in statistics (as normality, linearity, independence) are at most approximations to reality: the theories of classical parametric[3] statistics derive optimal procedures under exact parametric models, but say nothing about their behavior when the models are only approximately valid. At the same time the concept of robust statistics has to be distinguished from the methodology of nonparametric statistics, which explicitely avoids any assumption with regard to the underlying population. Robust statistics is still parametric in the sense that it considers the neighborhood[4] of parametric models. One example for this characteristic is presented in HUBER (1964), who estimates the location parameter ξ of a distribution function F(t) for the independent, identically distributed (iid) random variables $x_1, \ldots, x_n$. This estimation is complicated by the fact that the distribution function F(t) is only approximately known:

Let $$F = (1 - \varepsilon)\Phi + \varepsilon H \tag{2}$$

where $0 \le \varepsilon \le 1$ is a known number,
H is an unknown contaminating distribution function,
Φ is the standard normal distribution function.

The purpose of robust statistics can be seen in the search for estimators which are highly efficient at the parametric model, and whose distribution change little in a small neighborhood of it. Thus, the main aim of this methodology is to describe the structure best fitting the bulk of the data, while taking account of slight deviations from the assumed distribution. Usually the distribution underlying the model is the normal, either for mathematical convenience or because of the central limit theorem[5]. Deviations from this distribution assumption can among other things be caused: (a) by errors in the data set, (b) by the way the data have been filtered (e.g. rounded), (c) or by the fact that the normal distribution was an approximation in the first place (as it is frequently the case for financial markets data). Any of these deviations from the assumed model results in efficiency[6] losses during the parameter estimation. Thus, already slight deviations from the normal distribution cause the variance of the estimated mean and standard deviation to increase significantly.

Scale and Scale estimation

As described in SCHERVISH (1995) a scale (dispersion) estimator should satisfy the following properties: (1) the scale of a sample is nonnegative and is zero only when all the sample observations are identical; (2) the scale is invariant to additive shifts in the location of the sample; (3) multiplying the observation vector X with a constant b increases the scale estimate by the factor b. Thus, S(X) is a scale estimator, if

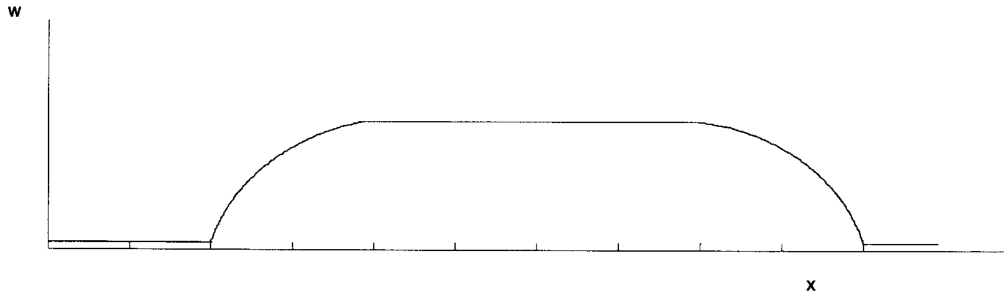
$$S(a + bX) = |b|S(X) \ge 0. \tag{3}$$

The scale estimators used in this paper all satisfy condition (3) and are taken from LAX (1985).

Outliers and the weighting of the observations

The basic idea behind robust statistics is to treat observations, which are "too far" away from the center, differently from observations which lie close to the center of the distribution. This is achieved by defining influence functions, which describe the approximate effect of an additional observation x on a particular statistic, given a large sample with a certain distribution.
The basic concept to achieve this was proposed by HUBER (1964), who suggested solving the following equation for a scale estimate S:

Figure 1: Weight of observation x



$$\frac{1}{n-1}\sum_{i=1}^{n}\psi^2(u_i) = E\left[\psi^2(z)|z \approx N(0,1)\right] \qquad (4)$$

with $u_i = (x_i - T)/S$,

T an estimate of the center,

and the weighting function:

$$\psi_H(u_i) \begin{cases} = -b, & u_i < -b, \\ = u_i, & b \geq u_i \geq -b, \\ = b, & u_i > b. \end{cases}$$

The deviations of observations that are far from the center T are limited in their influence, because their weight is set to b^2 rather than u_i^2. The concept has been further developed to include redescending ψ functions (as used in the sine A or biweight A estimator), which are characterized by:

$\psi(u_i) \rightarrow 0$ as $u_i \rightarrow \infty$.

Thus observations which are located too far from the center receive only a minimal influence on the estimate, while observations in the intermediate area are partly downweighted and those close to the center carry full weight, as figure 1 shows for observations x and weights w.

The following scale estimators are chosen for estimation: the standard deviation and the robust estimators: trimmed standard deviation, modified sine A, biweight A.
While the standard deviation is taken as a benchmark, the other estimators are chosen because of their performance as reported in table 1 of LAX (1985). The trimmed standard deviation has been included, because BAHRA (1994) reports that this estimator is performing particularly well under the assumption of fat-tailed distributions[7] .
The aim of the results below is to establish the ability of various volatility measures to provide exact volatility estimates and volatility forecasts in the presence of significant deviations from the normal distribution.

3. Empirical Analysis

3.1 The Data

Since ARCH models work particularly well for stock return data, the Swiss Performance Index (SPI) is chosen as the basis for the empirical investigation. The return data [$r_t = ((S_t - S_{t-1})/S_{t-1})*100$] are quoted daily and span the time from January 02, 1989 to December 31, 1995.

The estimates reported in Table 1 describe some of the characteristics mentioned when ARCH models and robust estimators were introduced. Compared to the mean and the standard deviation both the minimum and the maximum values are far too large to fit a normal distribution. Combining this with the significant measure for excess kurtosis shows that the return distributions are clearly fat-tailed. The null-hypothesis that the data are normal distributed is rejected.
The test for the presence of ARCH effects is modeled as proposed by ENGLE (1982) and the null-hypothesis of constant conditional variance is rejected at the 1% level for the SPI returns.

3.2 Volatility Estimation

Since it has been shown that ARCH effects are present and that the SPI returns show the characteristics, which justify the use of robust estimators, the procedures are first implemented to estimate the different volatility measures over the whole sample of 1825 observations. To calculate the robust estimates and the standard deviation a simple moving average over a 20 day window[8] is used. From the variety of available ARCH type models a GARCH, an EGARCH and a GARCH-T[9] representation is chosen to calculate volatilities under the assumption of autoregressive behavior. Figure 2 shows that the standard deviation of the SPI returns has varied considerably over the sample period.
The statistics presented in Table 2 demonstrate the way robust estimators work by giving less weight to observations which are far away from the mean of the distribution: the average level of volatility as measured by the standard deviation is higher than the average of the robust estimates. This is because the standard deviation equally weights all observation when calculating the dispersion, while the robust estimators use weighting functions (e.g. the mod. sine A estimator) or leave off certain observations altogether (the trimmed standard deviation). This also explains why among the robust estimators the trimmed standard devia-

Table 1a: Descriptive Statistics for daily return data (1825 observations)

	mean	Std.Dev	Min.	Max.	Skewness	Ex. Kurt	Normality
SPI	0.044	0.817	–9.852	4.924	–1.690	20.171	1865.5

The values for skewness and excess kurtosis of a standard normal distribution are both 0.
The normality test (DOORNIK, HANSEN (1994)) has a chi-squared distribution with 2 dgf.

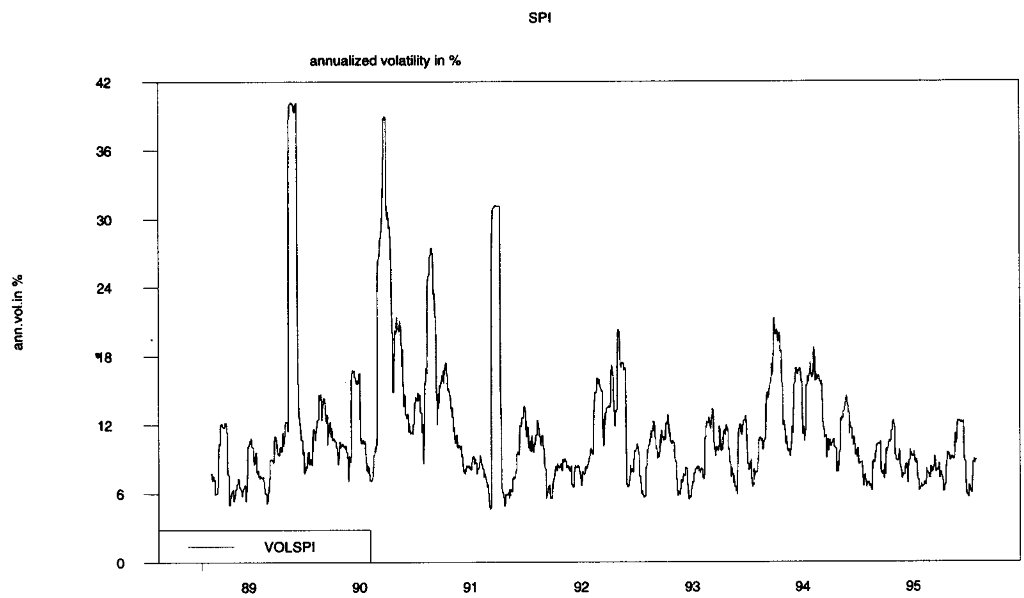
Table 1b: Testing for Autocorrelation in the Returns and the Presence of ARCH Effects

	autocorrelation				ARCH effects up to	
	lag1	lag2	lag3	lag4	lag5	lag20
SPI	0.023	–0.006	–0.008	0.043	present	present

Testing for the presence of ARCH involves running the regression:

$\varepsilon_t^2 = \alpha + \beta_1 \varepsilon_{t-1}^2 + \cdots + \beta_k \varepsilon_{t-k}^2$ and testing $H_0 := \beta_1 = \cdots = \beta_k = 0$

Figure 2: Market volatility



tion has by far the lowest mean. Including the biweight estimator made some additional calculations necessary, because this estimator has proved very sensitive to the choice of the constant factor c, which is used to standardize the deviation from the mean. While LAX (1985) reports values for c around 9, a grid search[10] for c results in values which are higher.

Finally, the volatility estimates from the GARCH, the EGARCH and the GARCH-T model are on average higher, which is due to the high amount of persistence found in all ARCH models, but show significantly less variation than the standard deviation.

Table 2: Statistics on the standard deviation, robust volatility estimators, and ARCH volatility for the SPI

	mean	std.dev.	Minimum	Maximum
std. dev.	11.73	5.89	4.64	40.21
mod. sine A	10.61	4.23	3.20	39.53
trim. std. dev.	7.46	2.50	3.53	15.94
biweight	11.67	5.56	4.63	60.96
GARCH	12.73	3.26	10.38	50.99
GARCH-T	12.05	4.35	8.50	60.40
EGARCH	12.50	3.19	7.61	45.89

The statistics refer to 1696 observations from July 1 1989 to December 30 1995. All rates are annualized.

4. Forecasting

4.1 Generating the Forecasts

The purpose of this section of the paper will be to test the reliability of robust estimators by using them as the basis for volatility forecasts and compare their performance with forecasts based on ARCH estimates.

A considerable problem lies in the fact that assuming a certain ARCH type model determines the forecasting procedure[11], whereas using a robust estimator does not point to a certain underlying process, which has to be used for forecasting purposes. This introduces the problem that whenever the forecasting ability of a robust estimator is tested, the validity of the implicitly assumed process connecting the robust estimates is under scrutiny as well. This might bias the forecasting performance of the robust estimators downwards.

The procedures proposed by HEYNEN and KAT (1994) for the ARCH models are used to calculate out-of-sample-forecasts as a function of the deviation of the last estimated variance from the unconditional variance, the forecasting horizon and the coefficients of the volatility model. One way to address the problem posed by the robust estimators is the Box Jenkins forecasting method, as discussed in BOX and JENKINS (1976), which involves the process of constructing an AR(I)MA(p,d,q) model and the generation of forecasts from this model. The model found to fit the data best is assumed to describe the process[12] connecting the robust estimates.

4.2 Testing the predictive power

JORION (1995) shows that the predictive power of a volatility forecast can be estimated by regressing the realized volatility on the predicted volatility:

$$\sigma_t = \alpha + \beta_1 \hat{\sigma}_t^F + \varepsilon_t \qquad (10)$$

To estimate the forecasts a moving window procedure is employed which first estimates the respective model making use of four years of data and then calculates a k (5,10,20,40) step ahead forecast. By moving the window one step forward and repeating the process, a day to day series of out-of-sample volatility forecasts is created. The same moving window methodology is also used to generate a series of k steps ahead forecasts based on the robust volatility estimates. While this approach makes use of all the data available and returns a large number of forecast values it also introduces overlaps in the error terms of above regression of order k – 1. Traditionally, this problem is avoided by defining the sampling interval to be equal to the forecast interval. In this case this would imply to move the window not by 1 but by k steps per cycle. Another way around this problem is HANSEN's (1982) Generalized Method of Moments (GMM), which allows to account for unknown orders of serial correlation and heteroscedasticity while explicitely holding the instrument vector orthogonal to the error terms of the regression. Past values of the regression's explanatory variable (which is the volatility forecast) are used as instruments. If the forecast is efficient in the sense that all past information (including past forecasts) is used to generate the current forecast then the covariance between the instruments and the error term of the regression should be zero. GMM estimates are not efficient compared to a well-specified GLS model, but by using GMM the danger of inconsistent parameter estimates can be avoided.

4.3 Results

A first characteristic of the results is a significant decline in the goodness of fit, as measured by the R^2, for longer forecast horizons both for the robust and the ARCH type estimators. While this is true for all estimators, some indication can be found that the relative performance of the ARCH models is improving as the horizon is increased. As shortly mentioned above the use of the robust estimators necessitates the modeling of an underlying process as well. This implicitly makes any test on the forecast ability of the robust estimators a joint hypothesis test: that the volatility is correctly estimated and that the correct underlying process has been chosen.

Table 3: Results from the forecasting regression (5, 10, 20, 40 days)

$$\sigma_t = \alpha + \beta_1 \hat{\sigma}_t^F + \varepsilon_t$$

where σ_t is the standard deviation and $\hat{\sigma}_t^F$ the period t forecast

Interval	const.	stoch.	trim	biweight	mod.sine	GARCH-T	EGARCH	GARCH	R^2	GMM
5	0.299	0.973**							0.74	0.04
	4.71**		0.772**						0.30	0.94
	–0.38			1.020**					0.81	0.02
	1.55**				0.853**				0.64	0.09
	–4.50**					1.242**			0.55	0.10
	–2.24*						1.006**		0.35	0.27
	–1.39							0.906**	0.31	0.72
10	2.31**	0.755**							0.48	0.23
	5.70**		0.613**						0.18	0.98
	1.73**			0.797**					0.50	0.20
	3.01**				0.695**				0.42	0.67
	–5.57**					1.312**			0.49	0.57
	–1.42						0.939**		0.30	0.92
	0.34**							0.752**	0.23	0.93
20	5.87**	0.386**							0.13	0.70
	7.17**		0.381**						0.07	0.96
	5.38**			0.424**					0.14	0.77
	6.04**				0.382**				0.13	0.85
	–0.57					0.870**			0.17	0.96
	–0.71						0.866**		0.14	0.95
	2.47							0.582*	0.11	0.99
40	6.38**	0.340**							0.06	0.29
	7.62**		0.333*						0.04	0.35
	5.48**			0.418**					0.06	0.29
	6.49**				0.343**				0.05	0.36
	–0.171					0.818**			0.08	0.52
	0.77						0.748**		0.07	0.31
	2.18							0.592**	0.08	0.38

** significance at the 1% level, * significance at the 5% level
The column "GMM" reports the significance value of the specification test for the GMM regression: values smaller than 5% (0.05) reject the nullhypothesis that the regression is correctly specified.

The reported loss of relative strength of the robust estimators is an indication that this problem is of relevance. Nevertheless it has to be seen that particularly over horizons of 5 and 10 days the forecasts explain a considerable part of the variation in the market volatility. Considering that in today's financial markets 10 days are not exactly a short period of time the R^2 reported here are promising and justify the methodology used.

Still, achieving better forecasts than the simple benchmark of the stochastic volatility assumption (reported as STOCH.) remains a considerable problem, although the simplest process (AR(1)) and the standard deviation is used to generate these forecasts. Nevertheless, Table 3 shows that a simple stochastic forecast can be outperformed, namely by forecasts based on the BIWEIGHT estimator for shorter and the GARCH-T model for longer forecast horizons. This result contrasts to HEYNEN and KAT's (1994) work, which shows the stochastic volatility model outperforming the ARCH models used as comparison. Along with the reasonable R^2 this shows the volatility prediction beyond purely stochastic models is rewarding.

By taking a closer look at the individual results it becomes evident that there is a difference between smoothing (robust) dispersion estimators by weighting down outliers and the complete removal of the most distant observations. This effect becomes particularly pronounced if one compares the trimmed standard deviation estimator (TRIM) and the BIWEIGHT estimator. The significantly worse performance of the TRIM estimator for all horizons shows that simply cutting off distant observations generates an undesirable loss of information. The BIWEIGHT estimator on the other hand, which uses all observations by assigning them different weights, shows that there is indeed improvement possible by acknowledging the presence of outliers.

A further traditional measure for the quality of a forecast is the joint test on the forecasting regression (as in Table 3) that the coefficient α on the constant is zero and the coefficient β on the forecast equals one. A F-Test run for the regressions presented consistently rejects this null-hypothesis. Inspection of the (significant) coefficients shows that this is largely due to the coefficient α which diverges considerably from zero: the coefficients on the constant are increasing in absolute value while the values for β are decreasing. This reflects the mean reverting nature of the procedures used to generate the out-of-sample-forecasts.

Finally, the question remains open which of the two groups (ARCH type or robust estimators) are on average better able to predict volatility. To test the assumption that the robust estimators are better in forecasting future volatilities a Mann-Whitney test statistic is used. This test can be employed to compare the central locations of two populations with unknown distribution functions. This characteristic is useful since the information which describes the forecasting performance is given by R^2, whose distribution is not obvious. The nullhypothesis tested is, that the two distributions of the R^2 given by the regressions on the ARCH models (GARCH, GARCH-T, EGARCH) and the robust forecasts (TRIM, BIWEIGHT, MOD. SINE) have the same central location; or more generally that there is no difference in the forecasting power. The calculated rank sum for the robust forecasts in Table 3 is 145.5 versus 154.5 for the (E)GARCH ones. This gives a value for the test statistic of 0.015 as compared to a 5% critical value of –1.65: using the realized standard deviation as measure, forecasts based on robust estimators cannot be said to offer consistently better results than the ones calculated from ARCH models. This is mainly due to the fact that the ARCH models are outperforming for longer horizons. Thus the Mann-Whitney statistic corroborates our above result that the relative performance changes with time and supports the conclusion that choosing an underlying process for the robust estimator from the variety of ARMA models might not be the ultimate step towards achieving optimal forecasts.

5. Conclusion

This work set out to show that, given certain empirical deviations from the normal distribution in financial data, along with the approach based on the ARCH methodology there is a second class of volatility estimators which also takes account of phenomena like clustering or the presence of leverage effects. Whatever the name, one realizes, that the whole discussion burns down to the question on how to treat outliers in the data. While the ARCH models recognize them as relevant realizations of an underlying process, the theory of robust estimation treats them as data points which should be given less weight in the calculation of the mean or the dispersion of a variable. When used to calculate measures of volatility for the SPI returns the robust estimates are less volatile and on average lower than the standard deviation, which is in line with their theoretical set-up. When robust and ARCH based volatilities are used to forecast future volatility no significant difference in their performance can be found. Nevertheless the statistics presented here show that along with the ARCH models the robust variance estimators provide useful instruments in the presence of deviations from normality.

While the nature of the problem that we report, namely deviations from the normality assumption in various financial time series, appears like a rather technical issue, the practical implications cannot be neglected. This is particularly true for the pricing of derivatives, e.g. options, where the volatility measure to a large degree determines the price of the financial instrument. The standard deviation, as the traditional volatility measure, rapidly looses accuracy as soon as the underlying data series exhibits leptokurtosis or skewness, which automatically invalidates the prices inferred from the Black-Scholes formula. Consequently, the mispricing of options and the problem of exact volatility estimation cannot be seen as separate issues.

Appendix

The Estimators:

(1) Standard Deviation:
Given a vector of sample observations $X = \{x_1, x_2, \ldots, x_n\}$ with an average $\bar{x}$, $d_i = x_i - \bar{x}$ are the deviations from the mean. The standard deviation is:

$$S = \left(\sum_{i=1}^{n} d_i^2 / (n-1) \right)^{0.5}. \qquad (5)$$

(2) Trimmed Standard Deviation:
A two-sided p% trimmed mean, $M_{2,p}$, can be calculated by sorting the observations, eliminating the (pn/2) smallest and the (pn/2) largest observations and then computing the mean of the remaining observations. The deviation from this trimmed mean $M_{2,p}$ is called d_i.

$$d_i = \left| x_i - M_{2,p} \right|$$

Analogously a one sided trimmed mean $M_{1,r}$ can be obtained by cutting off the r% highest observations. The trimmed standard deviation then is defined as

$$S_{trim} = \left(M_{1,r}(d_1^2, d_2^2, \ldots, d_n^2) \right)^{0.5}. \qquad (6)$$

(3) Biweight A:
Let T be the median of a vector $\{x_1, x_2, \ldots, x_n\}$ of observations, S_0 its MAD[13] , c a positive constant, and compute

$$u_i = (x_i - T) / cS_0. \qquad (7)$$

An increase in the constant c increases the relative importance of large deviations from T. The weighting function for the observations is given by

$$w_i = (1 - u_i^2)^2 \quad \text{for } |u_i| \le 1$$
$$= 0 \quad \text{for } |u_i| > 1.$$

Given these weights for the respective "standardized" observations the biweight A estimator is:

$$S_i^b = \frac{n}{(n-1)^{0.5}} * \frac{\left\{ \sum_{|u_i \le 1|} (x_i - T)^2 (1 - u_i^2)^4 \right\}^{0.5}}{\left| \sum_{|u_i \le 1|} (1 - u_i)^2 (1 - 5u_i^2) \right|}. \qquad (8)$$

(4) modified Sine A:
Let T be the median of a vector $\{x_1, x_2, \ldots, x_n\}$ of observations, S_0 its MAD, c a positive constant, and compute

$$u_i = (x_i - T) / cS_0.$$

The weighting function for the observations is calculated as

$$b(u_i) = \sin(u_i) \quad \text{if } |u_i| \le \pi$$
$$= 0 \quad \text{else.}$$

The modified sine A estimator is then calculated as

$$S_i^m = \frac{ncS_0}{(n-1)^{0.5}} * \tan^{-1} \left\{ \left(\sum_{|u_i| \le \pi} \sin^2(u_i) \right)^{0.5} * \left(\sum_{|u_i| \le \pi} \cos(u_i) \right)^{-1} \right\}. \qquad (9)$$

Footnotes

[1] GESKE, TOROUS (1990) show that "robust estimation of common stock return volatility eliminates the model's mispricing with respect to sample skewness and sample kurtosis".

[2] For an excellent survey of the ARCH literature see BOLLERSLEV et al. (1993).

[3] The term parameter is generally used to indicate a characteristic of the population. For example the class of all normal distibutions $\{N(\mu,\sigma^2)\}$ is parametric with the population parameters μ and σ^2.

[4] Neighborhood in this context can be understood as a "true" distribution lying near a parametric model.

[5] Stating that, whatever the distribution of x_i (as long as the variance is finite), for large n the distribution of $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ converges to the standard normal.

[6] Efficiency in this context is defined as follows (e.g. scale): Let V_{min} be the smallest known variance of a scale (measure of dispersion) estimator in repeated samples from a given distribution. Then an estimator with variance V has the relative efficiency E = 100 (V_{min} /V). Thus the relative efficiency of an estimator is lower the higher its variance.

[7] An exact description of the robust scale estimators used in the following estimates can be found in the appendix of this paper.

[8] see JORION (1995) and BRAILSFORD, FAFF (1996). A 20 day window covers one trading month, yielding 'monthly' volatility estimates. It can be shown that volatility shocks decay at a rate of $(\alpha+\beta)$ for the GARCH(1,1) model. The value of $(\alpha+\beta) = 0.84$ for the estimation below. A shock in period t return will thus have only very limited effect on the volatility estimate in period t+20. Consequently, a 20 day window is sufficient to estimate the volatility during this period.

[9] For the GARCH-T the normal distribution underlying the GARCH structure is replaced with a t-distribution in order to account for leptokurtosis. For equations (1b) and (1c) it was sufficient to set p = q = 1.

[10] Low values of c generate highly volatile results for the biweight estimator. Thus, the value of c, which is associated with the lowest amount of variance in the series of biweight estimates, is used in equation (7). The forecasting results support this approach.

[11] The procedures are derived and first used in HEYNEN, KEMNA, VORST (1994) and HEYNEN, KAT (1994). This paper makes use of their procedures.

[12] Usually an AR(1) process is sufficient.

[13] MAD is defined as the median absolute deviation from the median of the observations $\{x_1, x_2, \ldots, x_n\}$.

References

BAHRA B. (1994): „Robust Estimators of Volatility and their Forecasting Properties", MSc Thesis, Warwick Business School (unpublished).

BLACK F. and M. SCHOLES (1973): „The Pricing of Option Contracts and Corporate Liabilities", Journal of Political Economy 81, pp. 637-659.

BOLLERSLEV T., R. ENGLE and D. NELSON (1993): „ARCH models", Discussion Paper 93–49, University of California, San Diego.

BOX G. and G. Jenkins (1976): Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

BRAILSFORD T. and R. FAFF (1996): „An Evaluation of Volatility Forecasting Techniques", Journal of Banking & Finance 20, pp. 419–438.

DOORNIK J. and H. HANSEN (1994): „A Practical Test of Multivariate Normality", Nuffield College, (unpublished).

ENGLE R. (1982): „Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica 50, pp. 987–1008.

FAMA E. (1965): „The Behavior of Stock Market Prices", Journal of Business 38, pp. 34–105.

GESKE R. and W. TOROUS (1990): „Black-Scholes Option Pricing and Robust Variance Estimation", in: S. Hodges (ed.): Options – Recent Advances in Theory and Practice, Manchester: Manchester University Press.

HANSEN L. (1982): „Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, pp. 1029–1054.

HEYNEN R. and H. KAT (1994): „Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH(1,1) and EGARCH(1,1) Models", The Journal of Derivatives, pp. 50–65.

HEYNEN R., A. KEMNA and T. VORST (1994): „Analysis of the Term Structure of Implied Volatilities", Journal of Financial and Quantitative Analysis 29, pp. 31–56.

HUBER P. (1964): „Robust Estimation of Location Parameter", The Annals of Mathematical Statistics 35, pp. 73–101.

JORION P. (1995): „Predicting Volatility in the Foreign Exchange Market", The Journal of Finance Vol. L, pp. 507–528.

LAX D. (1985): „Robust Estimators of Scale: Finite Sample Performance in Long Tailed Symmetric Distributions", Journal of the American Statistical Association 80, pp. 736–741.

SCHERVISH M. (1995): Theory of Statistics, New York: Springer Verlag.