The Quality of Value at Risk via Univariate GARCH

Patrick Burns
patrick@burns-stat.com
http://www.burns-stat.com

This Draft: 10 October 2002

Abstract: The estimation of value at risk using univariate GARCH models is examined. A long history of the S&P 500 is used to compare these estimators with several other common approaches to value at risk estimation. The test results indicate that GARCH estimates are superior to the other methods in terms of the accuracy and consistency of the probability level. Although all of the GARCH models tested performed relatively well, the quality of the value at risk estimate does depend on which particular GARCH model is used. Weighting recent observations more heavily when fitting the GARCH model seems to be beneficial.

1. Introduction

Value at risk (VaR) has become very popular in risk management because it is an easily understood and obviously relevant concept. It is the amount of money that the portfolio is expected to lose more than over a set time horizon a given fraction of the time. For example if the distribution of tomorrow's expected gains and losses for our portfolio has 5% probability of losses greater than one million dollars, then one million dollars is the 1-day 5% VaR for the portfolio. Jorion (2000) provides an introduction to value at risk as well as discussing its estimation. The www.gloriamundi.org website comprehensively cites the value at risk literature as well as providing other VaR resources.

Even though explaining the concept of VaR is easy, creating the value is non-trivial. In statistical terms, the task is to provide a given quantile for a distribution that—since it continuously changes—is unobservable.

Since the task is difficult, it is sensible to test the quality of the procedures that are proposed. Here there is another problem—the best way to test a VaR estimator is to use a long history, a much longer history than it is typical to have. The present paper employs three test procedures which are increasingly comprehensive. These tests are applied to VaR estimators using almost 70 years of S&P 500 daily data.

The organization of the paper follows. Section 2 covers GARCH estimation of VaR. Section 3 presents the VaR estimators that are compared, and describes the data. Tests of the quality of the VaR estimators on the S&P data are given in sections 4, 5 and 6. Given the results from the pre-selected estimators, two additional estimators were tried—their quality is evaluated in section 7. Section 8 summarizes.

2. Value at Risk with GARCH

GARCH encompasses a broad class of models that estimate and predict volatility (plus correlations in the multivariate case). The original GARCH papers were Engle (1982) and Bollerslev (1986). These along with several other articles are collected into Engle (1995), while Bollerslev, Engle and Nelson (1994) survey the models. Two specific univariate models are fully described in section 3.

To estimate VaR using univariate GARCH, create a history of (daily) returns for the portfolio and then fit the model to these returns. The next step is to do a large number of simulations as many days ahead as the maximum time horizon that is of interest. Once the simulations are done, find the selected percentiles of the distribution of portfolio values within each simulated day. These estimators are similar to the filtered historical simulations of Barone-Adesi, Giannopoulos and Vosper (1999). Pritsker (2001) evaluates the quality of the filtered historical simulations and other estimators.

Exhibit 1 is a small example simulation to make it clear how the simulation operates. This simulation generates only 5 paths and goes 4 days into the future. The task is to find the 20% VaR of a portfolio with initial value 100. First, each simulation is run independently. For example in simulation 1, it gains 4 on the first day, then gains 1, then loses 2, then gains 1. Once all of the simulations are run, each day is examined separately to find the selected percentile of the simulated portfolio values within that day. The selected percentiles are highlighted in the exhibit. Our estimated 20% values at risk are 2 for day 1, 0 for days 2 and 3, and 3 for day 4.

Exhibit 1. Example simulation paths of the portfolio value.

	Day 1	Day 2	Day 3	Day 4
Simulation 1	104	105	103	104
Simulation 2	**98**	92	91	91
Simulation 3	100	104	101	**97**
Simulation 4	99	103	108	102
Simulation 5	97	**100**	**100**	99

Exhibit 2 shows an example using this technique for each day up to ten days ahead and with three probability levels. For this example the starting value of the portfolio is assumed to be 100, that is, we are looking at percent losses. In many cases a graphical presentation of VaR like this will be more useful than a list of one or more numbers.

Exhibit 2. VaR from a univariate GARCH model.

Trading Days after 1 December 2000
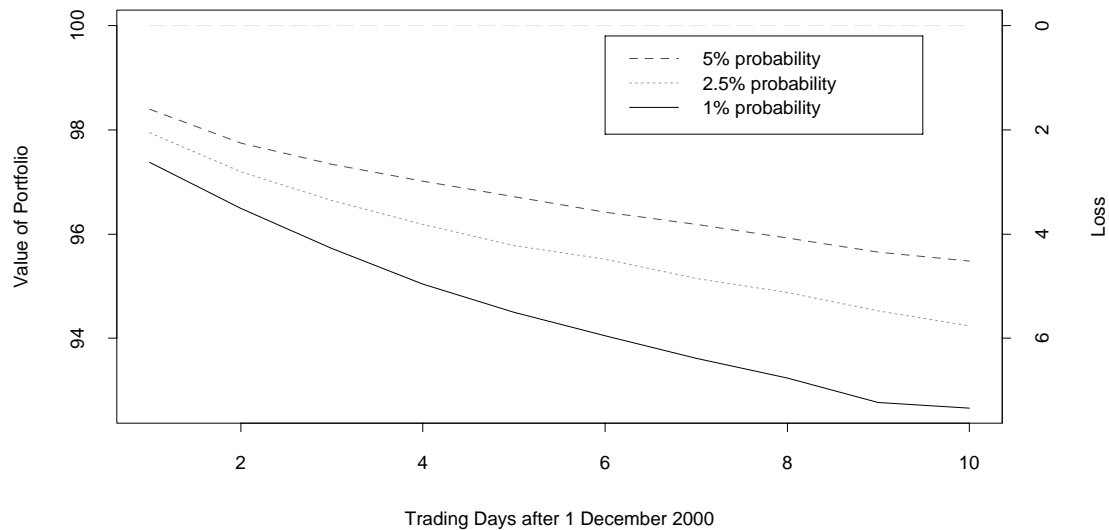
The simulation that produced Exhibit 2 generated 10,000 paths. The number of paths to simulate is determined by a trade-off between speed of computation and accuracy of results. The smaller the probability level, the more simulations needed to achieve a given accuracy. To see this consider generating 1000 simulations—a 5% probability means that there will be 50 points more extreme, but only 10 points more extreme for a 1% probability. Though it is not pursued in this paper, confidence intervals (assuming a correct model) for the value at risk estimate can be formed by using quantiles that surround the quantile of interest—see, for example, Pritsker (1997).

The simulation of the GARCH model ideally uses the empirical distribution of the residuals. This allows the data to speak for themselves about the distribution of returns. Using the empirical distribution reduces concern about the shape of the distribution assumed in the model. The concern is not eliminated, however. The distribution that is assumed when the GARCH model is fit does have an effect on the distribution of the residuals. This effect is likely to become more important as the probability level for the VaR gets smaller.

An alternative GARCH estimator for VaR would merely use the predicted volatility given by GARCH and assume a distribution—thus reducing the computational burden. While this technique would likely perform better than many non-GARCH alternatives, the savings in time relative to the GARCH simulation method is unlikely to be worthwhile unless a large number of portfolios are to be examined.

McNeil and Frey (2000) use GARCH in yet another way to get value at risk. They use GARCH to estimate the volatility, and extreme value theory to get tail probabilities.

GARCH simulation can easily provide an estimate of the mean shortfall as well as the value at risk. Mean shortfall, also known as mean excess loss, is the mean of all of the

3

losses in excess of the value at risk. The mean shortfall is quite sensitive to distributional assumptions—the GARCH simulation attempts to circumnavigate such assumptions.

Multivariate GARCH models can also be used to estimate VaR. Univariate models, though, have several advantages. Clearly univariate models are simpler and often have fewer parameters to fit. Multivariate models require a long data history on all of the assets while univariate models can adjust the history when some assets have missing values. However, multivariate GARCH is useful when there are options in the portfolio—the underlying assets for the options can be simulated into the future and the options evaluated.

3. Estimators Studied

Quite a menagerie of methods have been suggested to estimate VaR, they run from assuming that all returns follow the Normal distribution to very elaborate simulation models. Pritsker (1997) studies a number of estimators. A total of 22 estimators are compared here.

Three distributional assumptions are made: the normal distribution, Student's t with 10 degrees of freedom, and Student's t with 5 degrees of freedom. For each of these distributions, four methods of producing the volatility are used. Three of these are exponential smooths, with smoothing parameters of 0.95, 0.97 and 0.99. A rolling estimate is also used which is the standard deviation of the previous 500 observations. For these 12 estimators the 10-day volatility is found by assuming independent observations—that is, the 1-day volatility is multiplied by the square root of 10.

Three simulation methods are performed, using the previous 500, 1000 and 2000 observations. The given quantile is selected from the window of data. This makes no distributional assumption, but implicitly assumes that volatility is constant over the window. (Actually it makes a somewhat stronger assumption.) The 10-day VaR is based on the (overlapping) 10-day returns.

Finally four different GARCH estimators are used. There are two models, and each are estimated using 1000 observations and 2000 observations. The parameters of the models are found using maximum likelihood. The first model is the GARCH(1,1) model assuming the normal distribution. The form of the model is:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$$

where h_t is the variance at time t conditional on past information, and ε_t is the residual at time t. The three parameters in the equation are α, β and ω. These parameters are constrained to be positive in order to ensure that the conditional variance is always positive. There is also a parameter to estimate the mean of the series.

The second GARCH model examined is a components model introduced by Engle and Lee (1999) with a leverage effect as proposed by Glosten, Jagannathan and Runkle

4

(1993), and assuming that the errors follow a Student's t distribution with the degrees of freedom estimated. This model follows the two equations:

$$q_t = \omega + \rho q_{t-1} + \phi(\varepsilon_{t-1}^2 - q_{t-1})$$
$$h_t = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(h_{t-1} - q_{t-1}) + \lambda(\varepsilon_{t-1}^-)^2$$

where the parameters to be estimated are ω, ρ, ϕ, α, β and λ. Again, h_t is the conditional variance which is the quantity of interest. The first equation is the long-term component, the second equation combines the long-term and short-term components plus the leverage term. The notation e^- means $-e$ when e is negative and zero otherwise. There is a parameter for the mean of the series, making 8 parameters in total for the model.

The leverage term is often significant for equities, but tends to be zero for other asset classes. For a long-only equity portfolio, the leverage parameter would be expected to be positive—meaning that downward moves engender more subsequent volatility than upward moves of the same magnitude. Leverage is expected to be negative for a short equity portfolio. When the portfolio is long-short, then the leverage is likely to be smaller and may be of either sign.

The components model usually performs better on portfolios than the GARCH(1,1) model. Fitted components models tend to be more persistent than the GARCH(1,1) fit of the same data, in which case predictions from the components model will converge more slowly to their asymptote. The t distribution is also useful in achieving better fits—the degrees of freedom are often estimated to be in the range of 5 to 8 for daily returns.

All of the GARCH VaR estimates use 5000 simulations per day—the simulations use the empirical distribution of the residuals from the model. The parameters of the GARCH models are re-estimated every 10 trading days. A slight enhancement on the components model will be discussed in section 5, and further enhancements are presented in section 7.

The estimators are tested on the S&P 500 index. The data start on 17 April 1931 and end 29 December 2000. Because up to 2000 observations are used to estimate the models, the first test date is 21 April 1939. All tests are performed out of sample from the estimation—that is, the VaR estimates only use information that would have been available at the time that the estimate would have been made.

The probability levels studied are 1% and 5% with each using a time horizon of 1 day and 10 trading days. The number of test days for the 1-day estimates is 15,510. The number of dates in the 10-day estimate tests is 1550—these are non-overlapping periods.

The 1-day 5% VaR for the test period as estimated by the 2000 observation components GARCH model is plotted in Exhibit 3. The crash of October 1987 is the most significant event during the period, though there are other quite significant events also.

Exhibit 3. S&P 500 1-day 5% VaR in percent as estimated by a GARCH model.
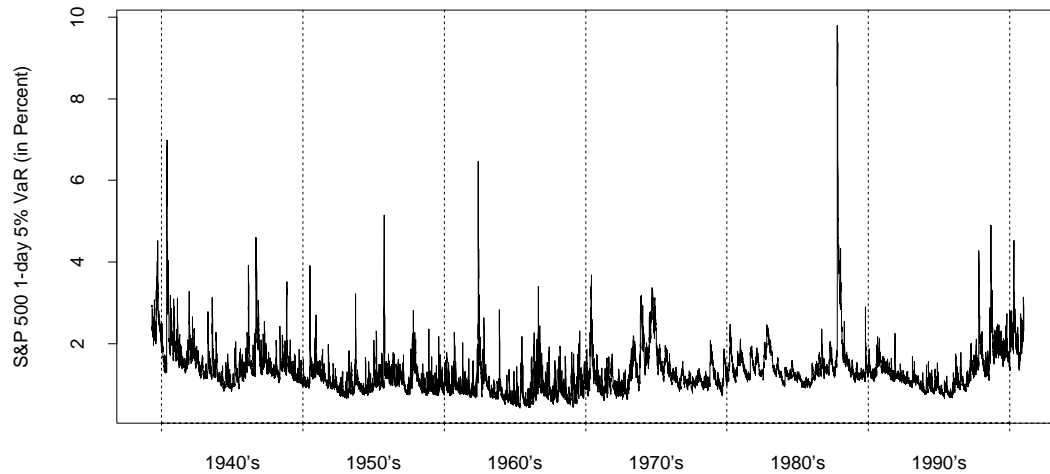
4. Testing the Level

The most basic test of a value at risk procedure is to see if the stated probability level is actually achieved. If for a particular time point the actual loss is greater than the VaR, then we call this a "hit". The hits can be formed into a numeric series which is 1 for time points where there is a hit and 0 when there is not a hit. The mean of this series is the achieved level of the procedure.

If we assume the probability of a hit is constant, then the number of hits follows the binomial distribution. Thus it is possible to form confidence intervals for the level of the VaR estimate. In order for this test to be very sensitive, the expected number of hits can not be small. Even with decades of history, the confidence intervals have appreciable width.

Exhibit 4 shows the achieved level and a 95% confidence interval for each of the 1-day VaR estimates. Bold entries in this and later exhibits indicate that the estimate can not be shown to be faulty at the 5% confidence level. Exhibit 5 presents levels and 95% confidence intervals for the 10-day VaR estimates—the final estimate will be explained in the next section.

6

Exhibit 4. Average probability level and confidence interval for 1-day VaR estimates.

Estimator	1% VaR Estimate	5% VaR Estimate
Normal ExpSmo 95	1.90 (1.68, 2.11)	**5.33 (4.98, 5.69)**
Normal ExpSmo 97	1.86 (1.64, 2.07)	**5.01 (4.67, 5.35)**
Normal ExpSmo 99	1.71 (1.50, 1.91)	4.58 (4.25, 4.91)
Normal Roll 500	1.82 (1.61, 2.04)	**4.71 (4.38, 5.05)**
t10 ExpSmo 95	1.53 (1.34, 1.73)	5.47 (5.11, 5.83)
t10 ExpSmo 97	1.50 (1.31, 1.69)	**5.21 (4.86, 5.56)**
t10 ExpSmo 99	1.39 (1.21, 1.58)	**4.71 (4.37, 5.04)**
t10 Roll 500	1.56 (1.37, 1.76)	**4.89 (4.55, 5.23)**
t5 ExpSmo 95	1.30 (1.12, 1.48)	5.96 (5.59, 6.34)
t5 ExpSmo 97	1.24 (1.06, 1.41)	5.68 (5.32, 6.04)
t5 ExpSmo 99	**1.16 (0.99, 1.33)**	**5.21 (4.86, 5.56)**
t5 Roll 500	1.35 (1.17, 1.54)	**5.29 (4.93, 5.64)**
Simulation 500	1.25 (1.08, 1.43)	**5.22 (4.87, 5.57)**
Simulation 1000	1.27 (1.09, 1.45)	**5.15 (4.80, 5.50)**
Simulation 2000	**1.07 (0.91, 1.23)**	**5.05 (4.71, 5.40)**
GARCH11 Norm 1000	**1.05 (0.89, 1.21)**	**4.94 (4.60, 5.28)**
GARCH11 Norm 2000	**1.01 (0.85, 1.17)**	**4.68 (4.35, 5.01)**
Comp GARCH t 1000	**1.17 (1.00, 1.34)**	**5.11 (4.76, 5.45)**
Comp GARCH t 2000	**1.03 (0.87, 1.18)**	**4.76 (4.42, 5.09)**

Exhibit 5. Average probability level and confidence interval for 10-day VaR estimates.

Estimator	1% VaR Estimate	5% VaR Estimate
Normal ExpSmo 95	3.03 (2.18, 3.89)	7.10 (5.82, 8.38)
Normal ExpSmo 97	2.71 (1.90, 3.52)	6.84 (5.58, 8.10)
Normal ExpSmo 99	2.52 (1.74, 3.30)	**6.06 (4.88, 7.25)**
Normal Roll 500	2.52 (1.74, 3.30)	**5.87 (4.70, 7.04)**
t10 ExpSmo 95	2.39 (1.63, 3.15)	7.48 (6.17, 8.79)
t10 ExpSmo 97	2.13 (1.41, 2.85)	7.23 (5.94, 8.51)
t10 ExpSmo 99	2.32 (1.57, 3.07)	6.26 (5.05, 7.46)
t10 Roll 500	2.06 (1.36, 2.77)	**6.00 (4.82, 7.18)**
t5 ExpSmo 95	2.06 (1.36, 2.77)	8.06 (6.71, 9.42)
t5 ExpSmo 97	1.94 (1.25, 2.62)	7.94 (6.59, 9.28)
t5 ExpSmo 99	1.81 (1.14, 2.47)	6.84 (5.58, 8.10)
t5 Roll 500	1.68 (1.04, 2.32)	6.52 (5.29, 7.74)
Simulation 500	1.81 (1.14, 2.47)	**5.94 (4.76, 7.11)**
Simulation 1000	**1.35 (0.78, 1.93)**	**5.29 (4.18, 6.40)**
Simulation 2000	**0.77 (0.34, 1.21)**	**5.10 (4.00, 6.19)**
GARCH11 Norm 1000	1.94 (1.25, 2.62)	7.10 (5.82, 8.38)
GARCH11 Norm 2000	**1.55 (0.93, 2.16)**	6.71 (5.46, 7.96)
Comp GARCH t 1000	**1.61 (0.99, 2.24)**	7.29 (6.00, 8.58)
Comp GARCH t 2000	**1.48 (0.88, 2.09)**	6.77 (5.52, 8.03)
AR Comp GARCH t 2000	**0.71 (0.29, 1.13)**	**4.97 (3.89, 6.05)**

For the 1-day estimates, most of the estimators have an acceptable level at 5%, but at 1% most—except the GARCH estimators—have too many hits. Though assuming a t distribution in lieu of the normal helps, even the 5 degrees of freedom t gets too many hits. The simulation method using 2000 observations is the only estimator to pass this test in all four combinations of probability and time horizon. Though the GARCH estimators are consistently good at 1-day, they are getting too many hits at 10 days. The reason for this will be discussed in the next section.

5. Testing Consistency of Level

We want the achieved level of the VaR to be the stated level on average, but we also want to achieve the stated level at all points in time. One approach to testing the consistency of the level is to use the Ljung-Box portmanteau test (Ljung and Box, 1978) on the hit series of zeros and ones. When using Ljung-Box tests, there is a choice of the number of lags in which to look for autocorrelation. If the test uses only a few lags but the autocorrelation occurs over a long time frame, the test will miss some of the autocorrelation. Conversely should a large number of lags be used in the test but the autocorrelation is all in a few lags, then the test won't be as sensitive as if the number of lags in the test matched the autocorrelation.

Exhibits 6, 7, 9 and 10 provide the results of Ljung-Box tests on the VaR estimates. Three different lags have been used for each estimate in order to try to get a good sense of the autocorrelation. A small p-value indicates that the level is not constant throughout time.

Burns (2002) investigates the Ljung-Box test, including in the setting of binary data. From the results of finite sample simulations, the p-values in Exhibit 7 can be taken to be reliable, but the Ljung-Box p-values will not be accurate in the other tables. Even when the p-values are suspect, the test statistic is still relevant.

Exhibit 6. Ljung-Box tests for the 1-day 1% VaR estimates. Given are the Chi-square test statistic and its p-value in parentheses. The p-values are suspect.

Estimator	5 lags	15 lags	50 lags
Normal ExpSmo 95	89.0 (0)	92.7 (3.1e-13)	124 (2.9e-8)
Normal ExpSmo 97	137 (0)	145 (0)	176 (6.7e-16)
Normal ExpSmo 99	232 (0)	301 (0)	352 (0)
Normal Roll 500	478 (0)	777 (0)	965 (0)
t10 ExpSmo 95	64.2 (1.6e-12)	68.9 (7.2e-9)	104 (1.2e-5)
t10 ExpSmo 97	129 (0)	141 (0)	169 (7.1e-15)
t10 ExpSmo 99	222 (0)	287 (0)	337 (0)
t10 Roll 500	461 (0)	717 (0)	894 (0)
t5 ExpSmo 95	59.2 (1.8e-11)	64.6 (4.0e-8)	95.5 (1.1e-4)
t5 ExpSmo 97	110 (0)	118 (0)	145 (3.1e-11)
t5 ExpSmo 99	240 (0)	290 (0)	343 (0)
t5 Roll 500	433 (0)	697 (0)	837 (0)
Simulation 500	294 (0)	507 (0)	672 (0)
Simulation 1000	356 (0)	723 (0)	1116 (0)
Simulation 2000	509 (0)	880 (0)	1334 (0)
GARCH11 Norm 1000	9.2 (0.10)	18.2 (0.25)	71.9 (0.023)
GARCH11 Norm 2000	18.9 (2.0e-3)	27.5 (0.025)	55.3 (0.28)
Comp GARCH t 1000	16.2 (6.4e-3)	24.7 (0.054)	78.8 (5.7e-3)
Comp GARCH t 2000	8.5 (0.13)	15.4 (0.42)	57.4 (0.22)

Exhibit 7. Ljung-Box tests for the 1-day 5% VaR estimates. Given are the Chi-square test statistic and its p-value in parentheses. The p-values are reliable in this case.

Estimator	5 lags	15 lags	50 lags
Normal ExpSmo 95	124 (0)	149 (0)	196 (0)
Normal ExpSmo 97	157 (0)	209 (0)	254 (0)
Normal ExpSmo 99	265 (0)	424 (0)	540 (0)
Normal Roll 500	461 (0)	802 (0)	1296 (0)
t10 ExpSmo 95	128 (0)	155 (0)	209 (0)
t10 ExpSmo 97	151 (0)	199 (0)	248 (0)
t10 ExpSmo 99	259 (0)	416 (0)	536 (0)
t10 Roll 500	455 (0)	791 (0)	1316 (0)
t5 ExpSmo 95	137 (0)	163 (0)	212 (0)
t5 ExpSmo 97	171 (0)	215 (0)	264 (0)
t5 ExpSmo 99	257 (0)	400 (0)	520 (0)
t5 Roll 500	446 (0)	774 (0)	1285 (0)
Simulation 500	342 (0)	646 (0)	1029 (0)
Simulation 1000	433 (0)	927 (0)	1616 (0)
Simulation 2000	444 (0)	934 (0)	1770 (0)
GARCH11 Norm 1000	44.1 (2.2e-8)	52.5 (4.6e-6)	90.9 (3.6e-4)
GARCH11 Norm 2000	53.6 (2.5e-10)	62.7 (8.6e-8)	108 (4.0e-6)
Comp GARCH t 1000	19.1 (1.8e-3)	29.3 (0.015)	74.8 (0.013)
Comp GARCH t 2000	16.3 (6.0e-3)	25.4 (0.045)	73.1 (0.018)

The tests of the 1-day 5% VaR are reliable, so it is possible to assess the quality of these estimates. While the GARCH estimates are not perfect, they perform very much better than the others—their Ljung-Box statistic is often an order of magnitude smaller than those for the other estimators. The components model is definitely better than the GARCH(1,1). As opposed to average hit rate where the simulation method with 2000 observations performed best, it is the worst for the 1-day VaR in terms of autocorrelation.

Exhibit 8 compares the hits in recent years from the components GARCH estimate using 2000 observations with those from the 500 observation simulation method. The cumulative hits from the GARCH method produce almost a straight line (though the results from Exhibit 7 indicate that it is probably not quite straight enough). In contrast the simulation method produces a large number of hits in 1986 and 1987 (even before the crash), and then has almost no hits in 1988 and 1989. This is the characteristic behavior of a volatility estimator that is too "stiff". It doesn't respond fast enough to increases in volatility, nor does it decrease fast enough as volatility subsides.

Exhibit 8. Cumulative hits of the 1-day 5% VaR estimates from the 2000 components GARCH and the 500 observation simulation methods.
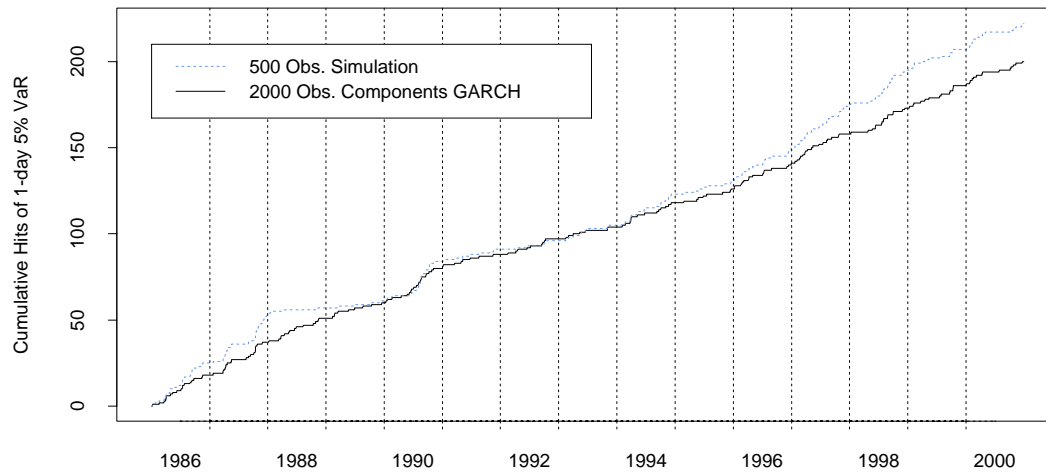

Exhibit 9. Ljung-Box tests for the 10-day 1% VaR estimates. Given are the Chi-square test statistic and its p-value in parentheses. The p-values are suspect.

Estimator	5 lags	15 lags	50 lags
Normal ExpSmo 95	5.1 (0.41)	15.8 (0.40)	63.2 (0.10)
Normal ExpSmo 97	3.7 (0.60)	10.1 (0.81)	46.0 (0.63)
Normal ExpSmo 99	17.3 (3.9e-3)	24.9 (0.052)	50.1 (0.47)
Normal Roll 500	30.4 (1.2e-5)	42.5 (1.9e-4)	64.9 (0.077)
t10 ExpSmo 95	2.8 (0.73)	8.7 (0.89)	50.1 (0.47)
t10 ExpSmo 97	2.5 (0.78)	11.6 (0.71)	57.7 (0.21)
t10 ExpSmo 99	6.0 (0.31)	15.6 (0.41)	42.5 (0.76)
t10 Roll 500	23.7 (2.5e-4)	35.5 (2.1e-3)	54.5 (0.31)
t5 ExpSmo 95	2.4 (0.79)	12.2 (0.67)	52.9 (0.36)
t5 ExpSmo 97	2.5 (0.78)	14.0 (0.52)	54.7 (0.30)
t5 ExpSmo 99	6.6 (0.25)	15.9 (0.39)	38.5 (0.88)
t5 Roll 500	8.5 (0.13)	24.7 (0.054)	42.6 (0.76)
Simulation 500	10.7 (0.057)	20.0 (0.17)	46.3 (0.62)
Simulation 1000	16.5 (5.6e-3)	25.7 (0.041)	45.5 (0.65)
Simulation 2000	9.4 (0.094)	19.3 (0.20)	59.0 (0.18)
GARCH11 Norm 1000	2.2 (0.83)	16.1 (0.37)	51.2 (0.43)
GARCH11 Norm 2000	2.6 (0.76)	16.4 (0.36)	52.5 (0.38)
Comp GARCH t 1000	3.1 (0.69)	9.8 (0.83)	28.7 (0.99)
Comp GARCH t 2000	3.7 (0.60)	9.2 (0.87)	29.6 (0.99)
AR Comp GARCH t 2000	0.4 (1.00)	12.2 (0.66)	60.1 (0.16)

Exhibit 10. Ljung-Box tests for the 10-day 5% VaR estimates. Given are the Chi-square test statistic and its p-value in parentheses. The p-values are suspect.

Estimator	5 lags	15 lags	50 lags
Normal ExpSmo 95	5.0 (0.42)	13.4 (0.57)	70.3 (0.030)
Normal ExpSmo 97	5.9 (0.31)	16.1 (0.38)	74.4 (0.014)
Normal ExpSmo 99	13.9 (0.016)	27.6 (0.024)	70.2 (0.031)
Normal Roll 500	22.1 (5.0e-4)	37.0 (1.3e-3)	79.9 (4.6e-3)
t10 ExpSmo 95	2.9 (0.72)	11.6 (0.71)	76.7 (9.0e-3)
t10 ExpSmo 97	6.2 (0.28)	15.3 (0.43)	72.7 (0.020)
t10 ExpSmo 99	18.9 (2.0e-3)	31.0 (8.8e-3)	81.1 (3.5e-3)
t10 Roll 500	21.2 (7.6e-4)	36.9 (1.3e-3)	80.5 (4.1e-3)
t5 ExpSmo 95	3.6 (0.61)	12.7 (0.63)	71.7 (0.024)
t5 ExpSmo 97	7.6 (0.18)	16.6 (0.34)	63.4 (0.097)
t5 ExpSmo 99	14.2 (0.014)	23.3 (0.078)	67.7 (0.048)
t5 Roll 500	28.2 (3.4e-5)	42.5 (1.9e-4)	83.4 (2.1e-3)
Simulation 500	19.9 (1.3e-3)	28.9 (0.017)	64.1 (0.086)
Simulation 1000	37.1 (5.7e-7)	48.7 (2.0e-5)	81.9 (2.9e-3)
Simulation 2000	18.8 (2.1e-3)	28.8 (0.017)	74.9 (0.013)
GARCH11 Norm 1000	3.4 (0.64)	15.8 (0.39)	78.4 (6.3e-3)
GARCH11 Norm 2000	1.8 (0.87)	12.4 (0.65)	63.7 (0.092)
Comp GARCH t 1000	1.8 (0.87)	11.8 (0.69)	70.0 (0.032)
Comp GARCH t 2000	2.0 (0.84)	9.9 (0.83)	82.2 (2.8e-3)
AR Comp GARCH t 2000	0.6 (0.99)	17.8 (0.27)	58.5 (0.19)

The imperfect performance of the GARCH methods—especially the failure of the 10-day estimates to have good average level—was surprising at first glance. Some investigation led to Exhibit 11 which shows the rolling mean level of hits for the 2000 observation components GARCH estimate. There is a definite pattern that corresponds almost precisely to the amount of autocorrelation in the data as shown in Exhibit 12.

A permutation test was performed on the hit series. It showed that if the hits were randomly ordered, then there would be about a one in a thousand chance that the rolling mean would hit or exceed 14.5%, as it actually does.

Exhibit 11. 200-observation rolling mean of hits from 10-day 5% VaR estimate from components GARCH 2000-observation estimate.
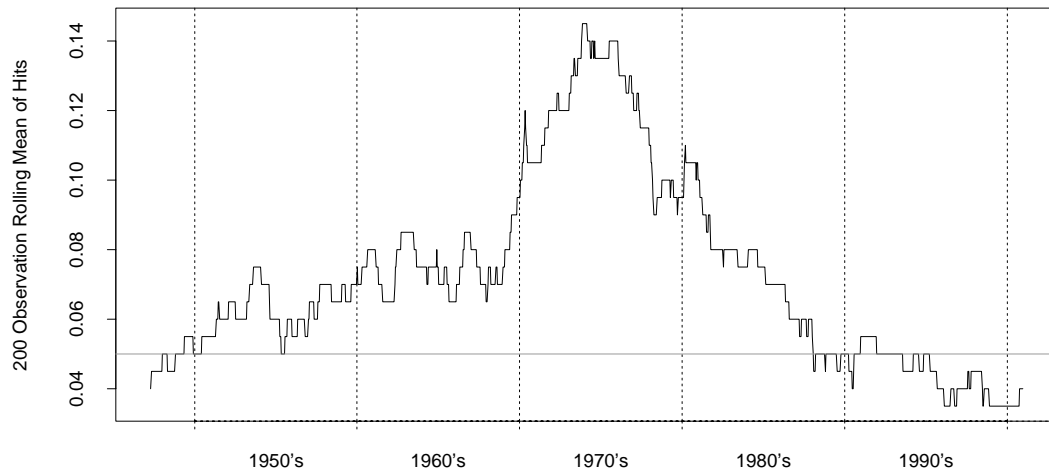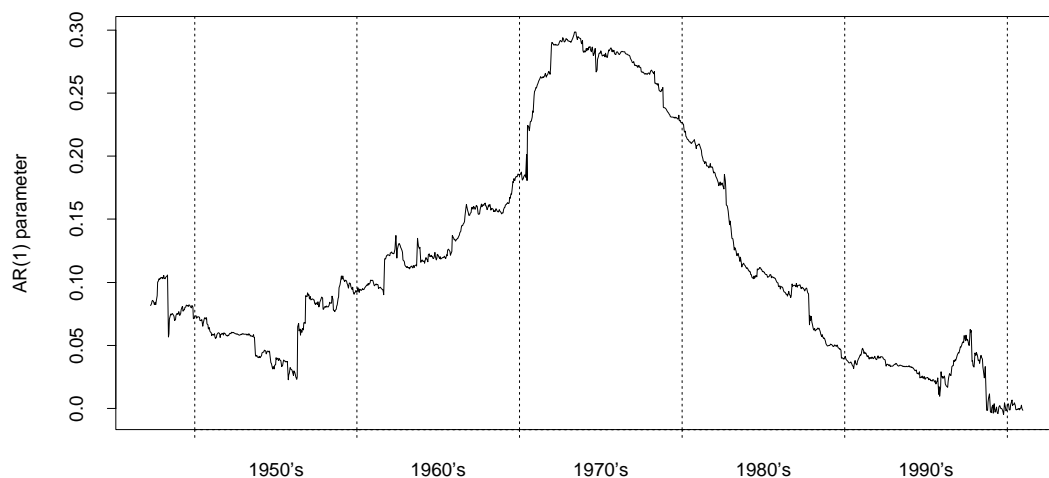

Exhibit 12. Estimate of an AR(1) parameter for the S&P data using a 2000 observation window.

An additional estimate for the 10-day VaR was computed using the 2000-observation components GARCH model but with the addition of an AR(1) parameter for the returns. The AR parameter was constrained to be non-negative, in the belief that a negative estimate would likely be noise rather than an indication of truly negative autocorrelation. The results for this estimator given in Exhibits 5, 9 and 10 show an improvement. Exhibit 13 shows the rolling mean of hits for this new estimator. This is a definite improvement over the estimate that did not include the AR parameter. The rise up to

13

10% in the 70's may still be too large though—a permutation test shows a maximum of at least 10% to occur about 5% of the time.

Exhibit 13. 200-observation rolling mean of hits from 10-day 5% VaR estimate from components GARCH with AR estimate.
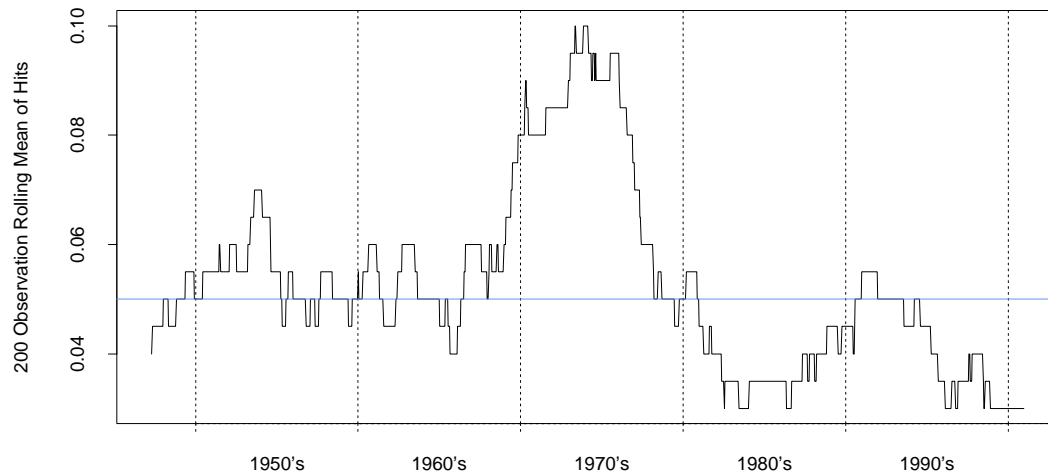
6. Testing Predictability

There is another property that VaR procedures should possess in addition to constantly having the stated probability level—the hits should be unpredictable with all information available at the time that the estimate is made.

Suppose that our 1% VaR procedure has predictable hits. If the probability of a hit is greater than 1%, we should increase the VaR estimate. Likewise, we should decrease the VaR on days when the probability of a hit is less than 1%.

There is a trivial way to get a VaR that is clearly not what we want, yet exactly satisfies the testing criteria of the previous two sections. The VaR is declared to be some very large number except on days that are randomly chosen (with the given probability level) when the VaR is declared to be some very large negative number. Thus we are generating hits only on the randomly chosen days. To eliminate the behaviour that this pathological estimate takes to the extreme, the hits should not be predictable by the VaR estimate.

These non-predictability requirements are tested by the dynamic quantile test of Engle and Manganelli (1999). This test is easy to perform as it is a simple function of the results of a least squares regression, and any variables that might be thought to influence the hits can be included. Should the test imply that there is predictability, the results of the regression can be examined to identify which variables have predictability.

The dynamic quantile test was performed for the S&P VaR estimates using an intercept, the VaR, 3 lags of the hits, the mean of the previous 50 hits, and the decades (the 1930's were put in with the 1940's, the 2000's in with the 1990's) for a total of 11 parameters. Thus the statistic for these tests follows a Chi-square with 11 degrees of freedom under the null hypothesis that there is no predictability. Exhibit 14 shows the test results for the 1-day VaR estimates, and Exhibit 15 contains the results for the 10-day estimates.

Exhibit 14. Dynamic quantile test statistic (p-value) for 1-day VaR estimates.

Estimator	1% VaR Estimate	5% VaR Estimate
Normal ExpSmo 95	276 (0)	131 (0)
Normal ExpSmo 97	316 (0)	139 (0)
Normal ExpSmo 99	349 (0)	243 (0)
Normal Roll 500	750 (0)	442 (0)
t10 ExpSmo 95	145 (0)	142 (0)
t10 ExpSmo 97	201 (0)	145 (0)
t10 ExpSmo 99	231 (0)	245 (0)
t10 Roll 500	563 (0)	462 (0)
t5 ExpSmo 95	97.5 (5.6e-16)	185 (0)
t5 ExpSmo 97	102 (0)	188 (0)
t5 ExpSmo 99	180 (0)	271 (0)
t5 Roll 500	407 (0)	486 (0)
Simulation 500	259 (0)	390 (0)
Simulation 1000	450 (0)	522 (0)
Simulation 2000	456 (0)	548 (0)
GARCH11 Norm 1000	33.0 (5.2e-4)	67.0 (4.5e-10)
GARCH11 Norm 2000	39.5 (4.4e-5)	70.6 (9.3e-11)
Comp GARCH t 1000	43.1 (1.0e-5)	43.5 (9.0e-6)
Comp GARCH t 2000	**14.8 (0.19)**	23.1 (0.017)

Exhibit 15. Dynamic quantile test statistic (p-value) for 10-day VaR estimates.

Estimator	1% VaR Estimate	5% VaR Estimate
Normal ExpSmo 95	126 (0)	53.2 (1.6e-7)
Normal ExpSmo 97	91.5 (8.4e-15)	50.7 (4.6e-7)
Normal ExpSmo 99	94,9 (1.8e-15)	44.6 (5.6e-6)
Normal Roll 500	125 (0)	48.9 (9.7e-7)
t10 ExpSmo 95	63.5 (2.0e-9)	63.9 (1.8e-9)
t10 ExpSmo 97	53.1 (1.7e-7)	54.5 (9.6e-8)
t10 ExpSmo 99	67.7 (3.3e-10)	51.8 (2.9e-7)
t10 Roll 500	91.3 (9.2e-15)	49.7 (7.1e-7)
t5 ExpSmo 95	50.3 (5.4e-7)	79.8 (1.6e-12)
t5 ExpSmo 97	47.1 (2.0e-6)	80.4 (1.2e-12)
t5 ExpSmo 99	44.4 (6.2e-6)	59.2 (1.3e-8)
t5 Roll 500	45.4 (4.1e-6)	67.7 (3.4e-10)
Simulation 500	34.8 (2.7e-4)	56.0 (5.2e-8)
Simulation 1000	45.9 (3.3e-6)	49.4 (8.1e-7)
Simulation 2000	**8.81 (0.64)**	32.3 (6.9e-4)
GARCH11 Norm 1000	51.2 (3.8e-7)	47.3 (1.9e-6)
GARCH11 Norm 2000	28.1 (3.2e-3)	38.4 (6.7e-5)
Comp GARCH t 1000	**19.6 (0.052)**	52.1 (2.6e-7)
Comp GARCH t 2000	24.9 (9.6e-3)	39.9 (3.7e-5)
AR Comp GARCH t 2000	**6.24 (0.86)**	**11.9 (0.37)**

For the 1-day VaR the GARCH estimates are again very much better than the others, yet they still show some predictability. In the 10-day case, the AR GARCH estimator appears to be superior to the rest.

7. Weighted GARCH Estimates

The GARCH estimators that were selected at the start of the simulation—while in general better than the other estimators—could be improved upon. After these results were seen, it was decided to investigate if weighted GARCH estimates would perform better. The hypothesis is that a GARCH model slowly evolving over time is a better model than a static GARCH. For example from Exhibit 12 it appears that the autoregression can change quite dramatically in a short period of time. Time-weighting the observations should keep the parameterization more up-to-date.

The components model with leverage and one autoregressive parameter was fit on a window of 2000 observations. The most recent observation was given three times the weight of the most ancient observation, and the weights were linear in between. Once the GARCH model was estimated, two forms of the simulation were performed. The first simulates in the usual way—each observation is selected with equal probability—while the second selects observations with the same weights as was used to fit the model. The

method that weighted the simulation so that more recent standardized residuals are more likely to be selected is referred to as doubly weighted.

Exhibit 16 shows the confidence intervals achieved with these methods. The Ljung-Box tests are given in Exhibit 17, and the dynamic quantile tests in Exhibit 18. The only test that the weighted GARCH estimates do not pass is the dynamic quantile test for the 1-day 1% VaR. There is also a hint that the probability level for the 10-day 1% VaR may be too small. The doubly weighted estimator seems to have a slight advantage over the estimator that uses the weights in the GARCH estimation but not in the simulation.

Exhibit 16. Average probability level and confidence interval for weighted GARCH estimators.

Estimator	1% VaR Estimate	5% VaR Estimate
Weighted, 1-day	**1.05 (0.89, 1.21)**	**4.86 (4.52, 5.20)**
Doubly weighted, 1-day	**1.04 (0.88, 1.20)**	**4.89 (4.55, 5.23)**
Weighted, 10-day	**0.65 (0.25, 1.04)**	**4.90 (3.83, 5.98)**
Doubly weighted, 10-day	**0.77 (0.34, 1.21)**	**4.90 (3.83, 5.98)**

Exhibit 17. Ljung-Box tests for the weighted GARCH estimators. Given are the chi-square statistic (p-value) for the 15 lag test.

Estimator	1% VaR Estimate	5% VaR Estimate
Weighted, 1-day	**20.0 (0.17)**	**13.1 (0.59)**
Doubly weighted, 1-day	**14.2 (0.51)**	**12.6 (0.63)**
Weighted, 10-day	**1.00 (1.00)**	**15.3 (0.43)**
Doubly weighted, 10-day	**10.4 (0.79)**	**14.9 (0.46)**

Exhibit 18. Dynamic quantile test statistic (p-value) for the weighted GARCH estimators.

Estimator	1% VaR Estimate	5% VaR Estimate
Weighted, 1-day	23.4 (0.015)	**11.8 (0.38)**
Doubly weighted, 1-day	21.1 (0.032)	**11.6 (0.40)**
Weighted, 10-day	**6.93 (0.80)**	**9.38 (0.59)**
Doubly weighted, 10-day	**6.43 (0.84)**	**13.6 (0.25)**

The regressions for the dynamic quantile test on the 1-day 1% VaR estimates were examined. It appears that there is predictability from VaR, the lag 1 hit, and possibly the decades.

8. Discussion

Though we don't want to make too sweeping of statements from the analysis of one dataset, there are at least tentative conclusions that can be drawn. When using traditional estimates of VaR, it appears to be a daunting task to find a combination of distributional assumption and volatility estimate to arrive at the stated probability level. It also appears that the desire to have the level constant throughout time can be at odds with getting the average level correct.

The simulation method has the nice property of not assuming a distribution, however, it does inherently assume constant volatility. This inherent assumption is onerous.

The univariate GARCH method gives quite good results—it keeps the good qualities of the simulation method while fixing its bad properties. Even so, the quality of the VaR estimate is sensitive to the quality of the GARCH model for the returns. Using time weights for the fitting of the GARCH model and the simulation appears to be valuable.

The CAViaR models of Engle and Manganelli (1999) are also likely to produce good quality estimates. These estimate the desired quantile directly rather than pulling the quantile out of an estimated distribution.

The S&P 500 data used to test the estimators is, in some ways, an easy case. There is not asynchronous data and it doesn't contain options, to name two common complications.

If the portfolio is global or multi-regional, then daily data will be asynchronous. For example if important news occurs in the afternoon New York time, then prices in New York will react to the news but Asia and Europe will be closed so they can only react when they next open. The significance of asynchrony is that the correlations between the assets appear to be smaller than they actually are. For portfolios that are all long (or all short) this will tend to have the effect of making VaR estimates too small. Burns, Engle and Mezrich (1998) provide a multivariate solution to asynchrony. It is possible that asynchrony could be suitably accounted for in univariate GARCH VaR estimates by adding a moving average term to the GARCH model, but this is a topic that needs to be researched.

The existence of options in the portfolio is probably a harder problem for univariate GARCH. A number of ad hoc approaches come to mind of how to adapt the approach to the existence of options. This is another avenue for further research.

Acknowledgements

References

Barone-Adesi, G., K. Giannopoulos and L. Vosper (1999). VaR without correlations for nonlinear portfolios. *Journal of Futures Markets* **19** (April) 583-602.

Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31** p307-327.

Bollerslev, Tim, Robert Engle and Daniel Nelson (1994). ARCH Models, in *Handbook of Econometrics, Vol IV*, ed. R. F. Engle and D. McFadden, North Holland, Amsterdam, p2959-3038.

Burns, Patrick (2002). Robustness of the Ljung-Box test and its rank equivalent. Burns Statistics working paper. http://www.burns-stat.com

Burns, Patrick, Robert Engle and Joseph Mezrich (1998). Correlations and Volatilities of Asynchronous Data, *The Journal of Derivatives* **5** p7-18.

Engle, Robert F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* **50**(4) p987-1007.

Engle, Robert F. (1995). *ARCH Selected Readings*. Oxford University Press.

Engle, R. and G. J. Lee (1999). A permanent and transitory component model of stock return volatility, in ed. R. Engle and H. White, *Cointegration , Causality and Forecasting: A Festschrift in Honor of Clive W. J. Granger*. Oxford University Press. pp 475-497.

Engle, Robert F. and Simone Manganelli (1999). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, University of California, San Diego, Department of Economics Working Paper 99-20.

Glosten, L., R. Jagannathan, and D. Runkle (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48** 1779-1801.

Jorion, Philippe (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*. (2nd Edition). McGraw-Hill, New York.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika* **65** p297-303.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7** 271-300.

Pritsker, Matthew. (1997). Evaluating value at risk methodologies: accuracy versus computational time, *Journal of Financial Services Research* **12** October/December, p201-242.

Pritsker, Matthew. (2001). The Hidden Dangers of Historical Simulation. FEDS Working Paper 2001-27, The Federal Reserve Board.