Stochastic volatility

Yvan Lengwiler
Faculty of Business and Economics
University of Basel

yvan.lengwiler@unibas.ch

File: prep-03.tex

Contents

1 The phenomenon

We have assumed that yields of equities are normally distributed. We have visually seen that this is largely an acceptable assumption empirically if we look at monthly or quarterly or even lower frequency data. At higher frequencies, we have observed that volatility is not constant. Instead, there seems to be temporary outbursts that last for a while and then die down again.

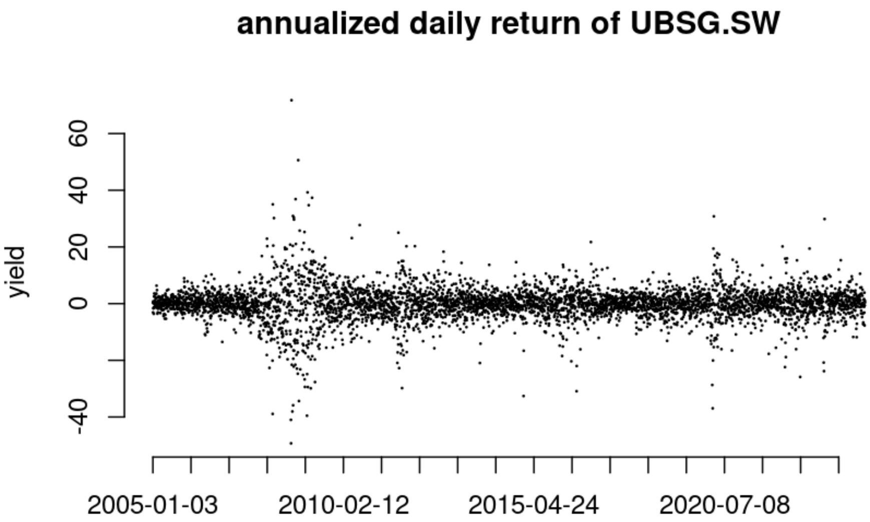

Figure 1: Daily returns of UBS Group equity, SIX market.

A very successful way of modelling this phenomenon is to assume that volatility itself is stochastic, and that the changes of volatility have some persistence.

If volatility of $\tilde{r}$ was constant, we would describe the process as

$$r_t = \mu + \sigma z_t.$$

Here, z_t is a standard normally distributed random variable (so zero mean and unit variance). The idea is to make σ a random variable as well. Moreover, σ_t is influenced by σ_{t-1}. This is what provides persistence of the volatility process.

1.1 *ARCH*

The first such model was proposed by Engle (1982) and goes like this,

$$\begin{aligned}
r_t &= \mu + \varepsilon_t, \\
\varepsilon_t &= \sigma_t z_t, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2.
\end{aligned}$$

The first line is the definition of the residual ϵ, i.e. the deviation of the period yield from the average yield. The second line describes the evolution of variance σ_t^2. This is an auto-regressive process with p lags: variance is modelled as an AR(p) process and the model is called "auto-regressive conditional heteroscadastic", or ARCH. Of course, variance can never become negative. To insure that, we assume that $\alpha_0, \alpha_1, \ldots, \alpha_p > 0$, The larger $\alpha_1, \ldots, \alpha_p$ are, the more persistent the volatility process becomes. However, we also assume that $\alpha_1 + \cdots \alpha_p < 1$, because otherwise the process is explosive and variance could easily diverge to infinity.

1.2 *GARCH*

This looks already promising. In practical applications, the AR-specification proved to be a bit too restrictive. Bollerslev (1986) extended the model by including a moving average component,

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2.$$

Variance is now an ARMA(p,q) process. Again, all α and β must be positive and $\Sigma_i \alpha_i + \Sigma_i \beta_i < 1$. This is the "generalized ARCH", or GARCH model.

The long-run expected variance can be computed by removing all time-indices from the variance equation, recognizing that the expectation of ε^2 is the same as σ^2, and solving for σ^2,

$$\sigma^2 = \omega + \alpha_1 \varepsilon^2 + \cdots + \alpha_p \varepsilon^2 + \beta_1 \sigma^2 + \cdots + \beta_q \sigma^2,$$

and thus

$$\sigma^2 = \frac{\omega}{1 - \Sigma_i \alpha_i - \Sigma_i \beta_i}.$$

2 Estimating GARCH(1,1)

2.1 *The maximum likelihood method*

The simplest GARCH specification is the one that is most often used. The GARCH(1,1) is

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

This model makes most sense with high frequency data, because we have observed this bunching of volatility much more in daily data than in weekly of monthly observations. So we will try to apply this model to our daily return series.

By "applying" I mean estimating the parameters of the GARCH model so that they best describe the observations. A sensible objective function that we want to maximize is the "probability" that the parameters we choose would generate the observed data. The word "probability" is in quotes here because we are dealing with continuous numbers, and the "probability" that a stochastic model generates a particular number is zero. With continuous spaces, we normally use the term "density" instead of "probability". In the present context, we will need "density" to describe a particular point on the density function of the returns at a particular date. We want to maximize the "overall density", meaning the product of the daily densities over each observation date. This product is called the *likelihood*, and we try to find parameters that describe the GARCH process so that the likelihood becomes as large as possible. The method is therefore called *maximum likelihood estimation*, or MLE.

2.2 *The likelihood function*

Consider some day t we have in our sample. We first compute μ as the mean of our observed yields r_t and then compute the residual as $\varepsilon_t = r_t - \mu$. The parameters (α, β) together with an initial variance σ_1^2 and the GARCH variance equation implies a sequence of variances, $\sigma_2^2, \sigma_2^2, \ldots, \sigma_t^2, \ldots$.

With these components, and assuming that the residuals are normally distributed on each day (with shifting variance though), we can compute the density of the observation at period t, namely r_t, or likewise, the residual at this period, $\varepsilon_t = r_t - \mu$. The density at the point ε_t in period t (when volatility of σ_t), is given by the density function of the normal distribution,

$$f(\varepsilon_t) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\varepsilon_t}{\sigma_t}\right)^2}.$$

Note that this is the normal density assuming a zero mean. This is justified here because the residuals ε_t should have zero mean by definition in an estimated model. In R, this function (the density of the normal distribution) is conveniently available as `'dnorm(...)'`, so we do not even have to code this ourselves.

We search for parameters $(\omega, \alpha, \beta, \sigma_1^2)$ so that the "overall density", or *likelihood* over all days, is as large as possible. The likelihood is the product of the individual daily densities,

$$L = \prod_t f(\varepsilon_t).$$

But maximizing a product is numerically expensive and not very stable. It is better to maximize the log of L.

$$\log L = \sum_t \log f(\varepsilon_t).$$

Again, coveniently the logarithm of the normal density is also available in R as `'dnorm(..., log=TRUE)'` :-)

1 Question. Why are we allowed to maximize $\log L$ instead of just plain L? Does this not skew our result? After all, the logarithm is not a linear transformation!

Your answer here:

Note: This is a non-linear optimization over four parameters, $(\omega, \alpha, \beta, \sigma_1^2)$. This should not be a substantial challenge for a modern computer. You can nevertheless simplify the problem by not optimizing over the initial variance σ_1^2. The effect of this parameter on the overall likelihood is small as it affects only the first few dates (directly in period 1 and indirectly through influence on $\sigma_2^2, \sigma_3^2, \ldots$), but this influence should die down relatively fast. Software packages therefore often mechanically set σ_1^2 equal to the unconditional variance $\omega/(1-\alpha-\beta)$ instead of optimizing it. You can do that, too, if you prefer. You should get only a slightly worse fit.

2.3 *Putting it into code*

2 Task. Use the file `GARCH-unfinished.r` and complete the script so that it uses the MLE method to fit a GARCH(1,1) model to the daily yields.

In addition, try to think of useful charts that would show what is going on. For instance, the simplest chart would simply plot σ_t. Maybe you can combine that with ε_t somehow?

The script downloads prices of UBS shares from the Swiss market and begins in 2005. This allows us to see the global financial crisis in a company that was deeply hit by it. Once your program runs, play around, investigate other companies, try to understand how different turmoils affect various companies differently.

Your answer here:

---

3 Simulation and Value at Risk

We are now combining the technologies we have learned this week (GARCH) and last week (Monte Carlo) into one realistic application that is used in practice.

3.1 *Simulating the GARCH process*

Just as we can Monte Carlo simulate a process with normally distributed innovations, we can Monte Carlo simulate a GARCH process. Just draw a random sequence of standard normally distributed z_t, start from an initial value σ_1^2, and generate ε_t and σ_t^2 for arbitrary t. The process is not fundamentally different than the Monte Carlo we have done last week, but implementing it correctly in code requires some diligence.

3 Task. Start from the file `GARCH-MC-unfinished.r` to generate a Monte Carlo simulation of a fitted GARCH(1,1) model for UBSG.SW for the duration of one month, starting at 2008-09-19 (in the middle of the global financial crisis), 2008-12-19 (three months later), and at 2024-04-02 (recently). Make charts of the simulated volatility and the simulated price of UBSG.SW.

PS: "One month" refers to one calendar month, which is about 22 working days. Since we work with working day data, you should simulate 22 days.

Your answer here:

⋆ ⋆ ⋆

I expect to get to this point in the first lecture. For next week, go though the rest of this handout, as well as through next week's handout as preparation for the lecture. That lecture will be online via Zoom.

Have this ready for the Zoom meeting. Please prepare notes, charts, programs etc so that you can share the in Zoom. If you failed to complete a task, that is not a problem. It happens to all of us. But do prepare an explanation that makes clear where exactly you struggled.

⋆ ⋆ ⋆

3.2 *An application: Value at Risk*

Value at Risk, or *VaR*, is a very common measure of financial risk. It is defined by two parameters: the probability α and the horizon t. $\text{VaR}_\alpha(t)$ is the minimum amount a portfolio or company or project looses over the next t days in the α percent worst cases. Mathematically, this is the α quantile of the value in t days.

VaR is incredibly popular because it focuses on the worst possible outcomes and captures that risk in just one number. VaR is normally given in units of money, not percent. Using our GARCH Monte-Carlo, we can easily compute the VaR for UBSG.SW shares.

4 Task. Suppose your company has a very non-diversified portfolio: It just holds CHF 1 million in UBS shares (UBSG.SW), nothing else. Compute the $\alpha = 1\%$ and $\alpha = 5\%$ VaR over one month (22 days) at the three dates used in the previous task. Comment on what these numbers and their changes over time mean.

Note: In reality, this exercise would be performed on a diversified portfolio. In that case, you just estimate the GARCH using the past returns of the portfolio and go from there.

Your answer here:

References

Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics*, 31, 307–327.

Engle, Robert F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, 50(4), 987–1007. `https://www.jstor.org/stable/1912773`.