1

Literature review

1.1 Stylized facts of returns

1.2 Volatility

1.2.1 Rolling volatility

When volatility needs to be estimated on a specific trading day, the method used as a descriptive tool would be to use rolling standard deviations. Engle [1] explains the calculation of rolling standard deviations, as the standard deviation over a fixed number of the most recent observations. For example, for the past month it would then be calculated as the equally weighted average of the squared deviations from the mean (i.e. residuals) from the last 22 observations (the average amount of trading or business days in a month). All these deviations are thus given an equal weight. Also, only a fixed number of past recent observations is examined. Engle regards this formulation as the first ARCH model.

1.2.2 ARCH model

Autoregressive Conditional Heteroscedasticity (ARCH) models, proposed by Engle [2], was in the first case not used in financial markets but on inflation. Since then, it has been used as one of the workhorses of volatility modeling. To fully capture the logic behind GARCH models, the building blocks are examined in the first place. There are three building blocks of the ARCH model: returns. the innovation process and the variance process (or volatility function), written out in respectively equation (1.1), (1.2) and (1.3). Returns are written as a constant part (μ) and an unexpected part, called noise or the innovation process. The innovation process is the volatility (σ_t) times z_t, which is an independent identically distributed random variable with a mean of 0 (zero-mean) and a variance of 1 (unit-variance). The independent from iid, notes the fact that the z-values are not correlated, but completely independent of each other. The distribution is not yet assumed. The third component is the variance process or the expression for the volatility. The variance is given by a constant ω, plus the random part which depends on the return shock of the previous period squared (ε_{t-1}^2). In that sense when the uncertainty or surprise in the last period increases, then the variance becomes larger in the next period. The element σ_t^2 is thus known at time $t-1$, while it is a deterministic function of a random variable observed at time $t-1$ (i.e. ε_{t-1}^2).

$$R_t = \mu + \varepsilon_t \tag{1.1}$$

$$\varepsilon_t = \sigma_t * z_t,\ where\ z_t \overset{iid}{\sim} (0,1) \tag{1.2}$$

$$\sigma_t^2 = \omega + \alpha_1 * \varepsilon_{t-1}^2 \tag{1.3}$$

From these components we could look at the conditional moments (or expected returns and variance). We can plug in the component σ_t into the conditional mean innovation ε_t and use the conditional mean innovation to examine the conditional mean return. In equation (1.4) and (1.5) they are derived. Because the random variable z_t is distributed with a zero-mean, the conditional expectation is 0. As a consequence, the conditional mean return in equation (1.5) is equal to the unconditional mean in the most simple case. But variations are possible using ARMA (eg. AR(1)) processes.

$$\mathbb{E}_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\sqrt{\omega + \alpha_1 * \varepsilon_{t-1}^2} * z_t) = \sigma_t \mathbb{E}_{t-1}(z_t) = 0 \tag{1.4}$$

$$\mathbb{E}_{t-1}(R_t) = \mu + \mathbb{E}_{t-1}(\varepsilon_t) = \mu \tag{1.5}$$

For the conditional variance, knowing everything that happened until and including period $t-1$ the condiitional innovation variance is given by equation (1.6). This is equal to σ_t^2, while the variance of z_t is equal to 1. Then it is easy to derive the conditional variance of returns in equation (1.7), that is why equation (1.3) is called the variance equation.

$$var_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\varepsilon_t^2) = \mathbb{E}_{t-1}(\sigma_t^2 * z_t^2) = \sigma_t^2 \mathbb{E}_{t-1}(z_t^2) = \sigma_t^2 \tag{1.6}$$

$$var_{t-1}(R_t) = var_{t-1}(\varepsilon_t) = \sigma_t^2 \tag{1.7}$$

The unconditional variance is also interesting to derive, while this is the long-run variance, which will be derived in (1.11). After deriving this using the law of iterated expectations and assuming stationarity for the variance process, one would get (1.8) for the unconditional variance, equal to the constant c and divided by $1 - \alpha_1$, the slope of the variance equation.

$$\sigma^2 = \frac{\omega}{1 - \alpha_1} \tag{1.8}$$

This leads to the properties of ARCH models.

- Stationarity condition for variance: $\omega > 0$ and $0 \leq \alpha_1 < 1$.
- Zero-mean innovations
- Uncorrelated innovations

Thus a weak white noise process ε_t

The unconditional 4th moment, kurtosis $\mathbb{E}(\varepsilon_t^4)/\sigma^4$ of an ARCH model is given by equation (1.9). This term is larger than 3, which implicates that the fat-tails (a stylised fact of returns).

$$3\frac{1-\alpha_1^2}{1-3\alpha_1^2} \tag{1.9}$$

Another property of ARCH models is that it takes into account volatility clustering. Because we know that $var(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma^2 = \omega/(1-\alpha_1)$, we can plug in ω for the conditional variance $var_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}^2) = \sigma_{t+1}^2 = c + \alpha_1 * \varepsilon_t^2$. Thus it follows that equation (1.10) displays volatility clustering. If we examine the RHS, as $\alpha_1 > 0$ (condition for stationarity), when shock ε_t^2 is larger than what you expect it to be on average σ^2 the LHS will also be positive. Then the conditional variance will be larger than the unconditional variance. Briefly, large shocks will be followed by more large shocks.

$$\sigma_{t+1}^2 - \sigma^2 = \alpha_1 * (\varepsilon_t^2 - \sigma^2) \tag{1.10}$$

Excess kurtosis can be modeled, even when the conditional distribution is assumed to be normally distributed. The third moment, skewness, can be introduced using a skewed conditional distribution as we will see in part 1.2.4. The serial correlation for squared innovations is positive if fourth moment exists (equation (1.9), this is volatility clustering once again.

The estimation of ARCH model and in a next step GARCH models will be explained in the methodology. However how will then the variance be forecasted?

Well, the conditional variance for the k-periods ahead , denoted as period $T+k$, is given by equation (1.11). This can already be simplified, while we know that $\sigma^2_{T+1} = \omega + \alpha_1 * \varepsilon^2_T$ from equation (1.3).

$$\mathbb{E}_T(\varepsilon^2_{T+k}) = \omega * (1 + \alpha_1 + ... + \alpha^{k-2}) + \alpha^{k-1} * \sigma^2_{T+1} \mathbb{E}_T(\varepsilon^2_{T+k}) \quad = \omega * (1 + \alpha_1 + ... + \alpha^{k-1}) + \alpha^k * \sigma^2_T \tag{1.11}$$

It can be shown that then the conditional variance in period $T+k$ is equal to equation (1.12). The LHS is the predicted conditional variance k-periods ahead above its unconditional variance, σ^2. The RHS is the difference current last-observed return residual ε^2_T above the unconditional average multiplied by α_1^k, a decreasing function of k (given that $0 \leq \alpha_1 < 1$). The further ahead predicting the variance, the closer α_1^k comes to zero, the closer to the unconditional variance, i.e. the long-run variance.

$$\mathbb{E}_T(\varepsilon^2_{T+k}) - \sigma^2 = \alpha_1^k * (\varepsilon^2_T - \sigma^2) \tag{1.12}$$

1.2.3 Univariate GARCH models

Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models come in because of the fact that rolling period standard deviations give an equal weight to the deviations, by such not taking into account volatility clustering, which can be identified as positive autocorrelation in the absolute returns.

sGARCH model
eGARCH model
iGARCH model
gjrGARCH model
nGARCH model
tGARCH model
tsGARCH model

1.2.4 Conditional distributions

1.3 Value at Risk

1.4 Conditional Value at risk

References

[1] Robert Engle. “GARCH 101: The use of ARCH/GARCH models in applied econometrics”. In: *Journal of Economic Perspectives* (2001).

[2] R. F. Engle. “Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation,journal of Econometrica, Volume 50, Issue 4 (Jul., 1982),987-1008.” In: *Econometrica* 50.4 (1982), pp. 987–1008.