

1

Literature review

1.1 Stylized facts of returns

When analyzing returns as a time-series, we look at log returns. The log returns are similar to simple returns so the stylized facts of returns apply to both. One assumption that is made often in financial applications is that returns are iid, or independently and identically distributed, another is that they are normally distribution. Are these valid assumptions? Below the stylized facts¹ following Annaert [1] for returns are given.

- Returns are *small and volatile* (with the standard deviation being larger than the mean on average).
- Returns have very little serial correlation as mentioned by for example Bollerslev [2].
- Returns exhibit conditional heteroskedasticity, or *volatility clustering*. This effect goes back to Mandelbrot [3]. There is no constant variance (homoskedasticity), but it is time-varying. Bollerslev [2] describes it as “rates of return data are characterized by volatile and tranquil periods”. Alexander [4] says this will have implications for risk models: following a large shock to the

¹Stylized facts are the statistical properties that appear to be present in many empirical asset returns (across time and markets)

market, the volatility changes and the probability of another large shock is increased significantly

- Returns also exhibit *asymmetric volatility*, in that sense volatility increases more after a negative return shock than after a large positive return shock. This is also called the *leverage effect*. Alexander [4] mentions that this leverage effect is pronounced in equity markets: usually there is a strong negative correlation between equity returns and the change in volatility.
- Returns are *not normally distributed* which is also one of the conclusions by Fama [5]. Returns have tails fatter than a normal distribution (leptokurtosis) and thus are riskier than under the normal distribution. Log returns **can** be assumed to be normally distributed. However, this will be examined in our empirical analysis if this is appropriate. This makes that simple returns follow a log-normal distribution, which is a skewed density distribution. A good summary is given by Alexander [4] as: “In general, we need to know more about the distribution of returns than its expected return and its volatility. Volatility tells us the *scale* and the mean tells us the *location*, but the dispersion also depends on the *shape* of the distribution. The best dispersion metric would be based on the entire distribution function of returns.”

Firms holding a portfolio have a lot of things to consider: expected return of a portfolio, the probability to get a return lower than some threshold, the probability that an asset in the portfolio drops in value when the market crashes. All the previous requires information about the return distribution or the density function. What we know from the stylized facts of returns that the normal distribution is not appropriate for returns. In appendix we summarize some alternative distributions (T, GED, ST, SGED, SGT) that could be a better approximation of returns than the normal one.

1.2 Volatility modeling

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 [6] 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 [7], 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 for an ARCH(1) 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).

$$y_t = \mu + \varepsilon_t \quad (1.1)$$

$$\varepsilon_t = \sigma_t \times z_t, \text{ where } z_t \stackrel{iid}{\sim} (0, 1) \quad (1.2)$$

$$\sigma_t^2 = \omega + \alpha_1 \times \varepsilon_{t-1}^2 \quad (1.3)$$

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.

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

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

For the conditional variance, knowing everything that happened until and including period $t - 1$ the conditional 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.

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

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

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.

$$\sigma^2 = \frac{\omega}{1 - \alpha_1} \quad (1.8)$$

This leads to the properties of ARCH models: Stationarity² condition for variance: $\omega > 0$ and $0 \leq \alpha_1 < 1$. But also, zero-mean innovations and 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 stylized fact of returns).

$$3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \quad (1.9)$$

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 $var_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}^2) = \sigma_{t+1}^2 = c + \alpha_1 \times \varepsilon_t^2$. 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.

²Stationarity implies that the series on which the ARCH model is used does not have any trend and has a constant expected mean. Only the conditional variance is changing.

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

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 saw in part ???. 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 $\sigma_{T+1}^2 = \omega + \alpha_1 \times \varepsilon_T^2$ from equation (1.3).

$$\begin{aligned} \mathbb{E}_T(\varepsilon_{T+k}^2) &= \omega \times (1 + \alpha_1 + \dots + \alpha^{k-2}) + \alpha^{k-1} \times \sigma_{T+1}^2 \\ &= \omega \times (1 + \alpha_1 + \dots + \alpha^{k-1}) + \alpha^k \times \sigma_T^2 \end{aligned} \quad (1.11)$$

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_T^2 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.

$$\mathbb{E}_T(\varepsilon_{T+k}^2) - \sigma^2 = \alpha_1^k \times (\varepsilon_T^2 - \sigma^2) \quad (1.12)$$

1.2.3 Univariate GARCH models

An improvement on the ARCH model is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH)³. This model and its variants come in to play because

³*Generalized* as it is a generalization by Bollerslev [8] of the ARCH model of Engle [7]. *Autoregressive*, as it is a time series model with an autoregressive form (regression on itself).

of the fact that calculating standard deviations through rolling periods, gives an equal weight to distant and nearby periods, by such not taking into account empirical evidence of volatility clustering, which can be identified as positive autocorrelation in the absolute returns. GARCH models are an extension to ARCH models, as they incorporate both a novel moving average term (not included in ARCH) and the autoregressive component. Furthermore, a second extension is changing the assumption of the underlying distribution. As already explained, the normal distribution is an unrealistic assumption, so other distributions which are described in part ?? will be used. As Alexander [4] explains, this does not change the formulae of computing the volatility forecasts but it changes the functional form of the likelihood function⁴. An overview (of a selection) of investigated GARCH models is given in the following table.

Table 1.1: GARCH models, the founders

Author(s)/user(s)	Model
Engle [7]	ARCH model
Bollerslev [8]	GARCH model
Bollerslev [8]	IGARCH model
Nelson [9]	EGARCH model
Glosten, Jagannathan, and Runkle [10]	GJRGARCH model
Engle and Ng [11]	NAGARCH model
Zakoian [12]	TGARCH model
Taylor [13] and Schwert [14]	TSGARCH (or AVGARCH) model
Morgan Guaranty Trust Company [15]	EWMA or RiskMetrics model

1.3 ACD models

An extension to GARCH models was proposed by Hansen [16], the autoregressive conditional density estimation model (referred to as ACD models, sometimes ARCD). It focuses on time variation in higher moments (skewness and kurtosis), because the degree and frequency of extreme events seem to be not expected by

Conditional heteroscedasticity, while time variation in conditional variance is built into the model [4].

⁴which makes the maximum likelihood estimation explained in part ?? complex with more parameters that have to be estimated.

traditional models. Some GARCH models are already able to capture the dynamics by relying on a different unconditional distribution than the normal distribution (for example skewed distributions like the SGED, SGT), or a model that allows to model these higher moments. However, Ghalanos [17] mentions that these models also assume the shape and skewness parameters to be constant (not time varying). As Ghalanos mentions: “the research on time varying higher moments has mostly explored different parameterizations in terms of dynamics and distributions with little attention to the performance of the models out-of-sample and ability to outperform a GARCH model with respect to VaR.” Also one could question the marginal benefits of the ACD, while the estimation procedure is not simple (nonlinear bounding specification of higher moment distribution parameters and interaction). So, are skewness (skewness parameter) and kurtosis (shape parameters) time varying? The literature investigating higher moments has arguments for and against this statement. In part ?? the specification is given.

1.4 Value at Risk

Value-at-Risk (VaR) is a risk metric developed simultaneously by Markowitz [18] and Roy [19] to calculate how much money an investment, portfolio, department or institution such as a bank could lose in a market downturn, though in this period it remained mostly a theoretical discussion due to lacking processing power and industry demand for risk management measures. Another important document in literature is the *1996 RiskMetrics Technical Document*, composed by RiskMetrics⁵, Morgan Guaranty Trust Company [15] (part of JP Morgan), gives a good overview of the computation, but also made use of the name “value-at-risk” over equivalents like “dollars-at-risk” (DaR), “capital-at-risk” (CaR), “income-at-risk” (IaR) and “earnings-at-risk” (EaR). According to Holton [19] VaR gained traction in the last decade of the 20th century when financial institutions started using it to determine their regulatory capital requirements. A VaR_{99} finds the amount that would be

⁵RiskMetrics Group was the market leader in market and credit risk data and modeling for banks, corporate asset managers and financial intermediaries [4].

the greatest possible loss in 99% of cases. It can be defined as the threshold value θ_t . Put differently, in 1% of cases the loss would be greater than this amount. It is specified as in (1.13). Christofferson2001 puts forth a general framework for specifying VaR models and comparing between two alternatives models.

$$Pr(y_t \leq \theta_t | \Omega_{t-1}) \equiv \phi \quad (1.13)$$

With y_t expected returns in period t , Ω_{t-1} the information set available in the previous period and ϕ the chosen confidence level.

1.5 Conditional Value at Risk

One major shortcoming of the VaR is that it does not provide information on the probability distribution of losses beyond the threshold amount. As VaR lacks subadditivity of different percentile outcomes, [Artzner1998] reject it as a coherent measure of risk. This is problematic, as losses beyond this amount would be more problematic if there is a large probability distribution of extreme losses, than if losses follow say a normal distribution. To solve this issue, they provide a conceptual idea of a Conditional VaR (CVaR) which quantifies the average loss one would expect if the threshold is breached, thereby taking the distribution of the tail into account. Mathematically, a $cVaR_{99}$ is the average of all the VaR with a confidence level equal to or higher than 99. It is commonly referred to as expected shortfall (ES) sometimes and was written out in the form it is used by today by [20]. It is specified as in (1.14).

To calculate θ_t , VaR and CVaR require information on the expected distribution mean, variance and other parameters, to be calculated using the previously discussed GARCH models and distributions.

$$Pr(y_t \leq \theta_t | \Omega_{t-1}) \equiv \int_{-\infty}^{\theta_t} f(y_t | \Omega_{t-1}) dy_t = \phi \quad (1.14)$$

With the same notations as before, and f the (conditional) probability density function of y_t .

According to the BIS framework, banks need to calculate both VaR_{99} and $VaR_{97.5}$ daily to determine capital requirements for equity, using a minimum of one year of daily observations [21]. Whenever a daily loss is recorded, this has to be registered as an exception. Banks can use an internal model to calculate their VaRs, but if they have more than 12 exceptions for their VaR_{99} or 30 exceptions for their $VaR_{97.5}$ they have to follow a standardized approach. Similarly, banks must calculate $CVaR_{97.5}$.

1.6 Past literature on the consequences of higher moments for VaR determination

Here comes the discussion about studies that have looked at higher moments and VaR determination. Also a summary of studies that discusses time-varying higher moments, but not a big part, while it is also only a small part of the empirical findings (couple of GARCH-ACD models).

Table 1.2: Higher moments and VaR

Author	Higher moments
Hansen [16]	
\@harvey1999	

Brooks et al. [22]

References

- [1] Jan Annaert. *Quantitative Methods in Finance*. Version 0.2.1. Antwerp Management School, Jan. 2021.
- [2] Tim Bollerslev. “A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return”. In: *The Review of Economics and Statistics* 69.3 (1987). Publisher: The MIT Press, pp. 542–547. URL: <https://www.jstor.org/stable/1925546>.
- [3] Benoit Mandelbrot. “The Variation of Certain Speculative Prices”. In: *The Journal of Business*. University of Chicago Press 36 (1963), p. 394.
- [4] Carol. Alexander. *Market risk analysis. Volume 2, Practical financial econometrics*. 2nd ed. The Wiley Finance Series. Chichester, England ; Wiley, 2008.
- [5] Eugene F. Fama. “The Behavior of Stock-Market Prices”. In: *The Journal of Business* 38.1 (1965), pp. 34–105. URL: <http://www.jstor.org/stable/2350752>.
- [6] Robert Engle. “GARCH 101: The use of ARCH/GARCH models in applied econometrics”. In: *Journal of Economic Perspectives* (2001).
- [7] 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.
- [8] Tim Bollerslev. “Generalized Autoregressive Conditional Heteroskedasticity”. In: *Journal of Econometrics* 31 (1986), pp. 307–327.
- [9] Daniel B. Nelson. “Conditional Heteroskedasticity in Asset Returns: A New Approach”. In: *Econometrica* 59.2 (Mar. 1991). Publisher: JSTOR, pp. 347–347.
- [10] Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”. In: *The Journal of Finance* 48.5 (Dec. 1993). Publisher: John Wiley and Sons, Ltd, pp. 1779–1801. URL: <http://doi.wiley.com/10.1111/j.1540-6261.1993.tb05128.x>.
- [11] Robert F. Engle and Victor K. Ng. “Measuring and Testing the Impact of News on Volatility”. In: *The Journal of Finance* 48.5 (Dec. 1993). Publisher: John Wiley and Sons, Ltd, pp. 1749–1778.
- [12] Jean Michel Zakoian. “Threshold heteroskedastic models”. In: *Journal of Economic Dynamics and Control* 18.5 (1994), pp. 931–955.
- [13] Stephen J. Taylor. *Modelling financial time series*. Chichester: John Wiley and Sons, Ltd., 1986.
- [14] G. William Schwert. “Why Does Stock Market Volatility Change Over Time?” In: *The Journal of Finance* 44.5 (1989), pp. 1115–1153.

- [15] Morgan Guaranty Trust Company. *RiskMetricsTM—Technical Document*. Tech. rep. 1996.
- [16] Bruce E. Hansen. “Autoregressive Conditional Density Estimation”. In: *International Economic Review* 35.3 (1994), pp. 705–730.
- [17] Alexios Ghalanos. *racd: Autoregressive Conditional Density Models*. <http://www.unstarched.net>, <https://bitbucket.org/alexiosg/>. 2016.
- [18] Harry Markowitz. “Portfolio Selection”. In: *Journal of Finance* 7.1 (1952), pp. 77–91.
- [19] Glyn A Holton. “History of Value-at-Risk: 1922-1998”. In: (2002). Contingency Analysis Working Paper. URL: <http://www.contingencyanalysis.com>.
- [20] Dimitris Bertsimas, Geoorey J Lauprete, and Alexander Samarov. “Shortfall as a risk measure: properties, optimization and applications”. In: *Journal of Economic Dynamics and Control* 28 (2004), pp. 1353–1381.
- [21] Basel Committee on Banking Supervision. *Minimum capital requirements for market risk*. Tech. rep. Issue: January Publication Title: Bank for International Settlements. 2016, pp. 92–92. URL: https://www.bis.org/basel_framework/.
- [22] Chris Brooks et al. “Autoregressive conditional kurtosis”. In: *Journal of Financial Econometrics* 3.3 (2005), pp. 399–421.