INTRODUCTION TO PORTFOLIO OPTIMIZATION, COVARIANCE ESTIMATION AND MULTIVARIATE GARCH

Heino Bohn Nielsen and Anders Rahbek
March 30, 2021

1 Introduction

This note introduces the theory for portfolio selection originating from Markowitz (1952). We will mainly focus on the econometric methodology and the empirical implementations of the theoretical models, but this note begins with an overview of the most basic financial theory and the most important insights. We do not dwell on all possible details and do not discuss extensions of the basic framework; for this we refer to standard textbooks in financial theory, *inter alia* Elton, Gruber, Brown, and Goetzmann (2003) or Cvitanic and Zapatero (2004).

We first introduce the notation used throughout and define a *portfolio* of assets. We then turn to the theory in Section 3, where we look at the opportunity set of the investor, i.e. the possible portfolios to invest in. We follow the so-called *mean-variance* approach of Markowitz (1952) and consider the opportunity set in terms of attainable combinations of mean and variance. Next, in Section 4, we introduce the utility function of the investor and discuss the optimal allocation within the opportunity set.

One main advantage of the mean-variance approach is the simplicity, and it is a good way to introduce the economic insights related to portfolio choice, in particular the need to balance expected return (mean) against the associated risk (variance), and the gains from diversification. The mean-variance analysis is a fully sufficient characterization of the portfolio choice possibilities under some particular assumptions on asset return distributions and utility functions, and we state those later. Moreover, it is easy to derive the equilibrium state under the mean-variance assumptions, and it leads directly to the well-known Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and Mossin (1966). The CAPM model has played an important role in the literature of

empirical finance, and it is–together with some basic extensions–introduced in a later lecture note.

Under more general assumptions on utility functions of investing agents or return distributions, the investor may be concerned also with the asymmetry (i.e. skewness) of the distribution or the probability of very large negative shocks (measured e.g. by the fourth moment), and a more elaborate analysis is needed. An alternative and more general approach to the portfolio theory would be to begin with a utility function and choose the portfolio weights to maximize utility directly; we explore this approach briefly in Section 4. Although more general (and possibly more theoretically elegant), it is harder to derive clear economically intuitive results, and it is much harder empirically to implement.

The input to the mean-variance portfolio selection is an estimate of the mean and the covariance matrix of returns. Section 6 gives an introduction to the estimation of covariance matrices in econometrics. The main point is that a $p \times p$ covariance matrix, Ω, is, by definition, a symmetric *positive definite* matrix and the estimation procedure should ensure that. This implies, in particular, that the covariance Ω does not have p^2 free parameters to be estimated. The necessary restrictions have to be imposed explicitly in the estimation procedure or by reparametrizing the matrix. A thorough understanding of the possible reparametrizations of the covariance matrix to impose symmetry and positive definiteness is a good starting point for reading the large, and sometimes confusing, literature on multivariate autoregressive conditional heteroskedasticity (MARCH) models.

Next, Section 7 presents some of the most natural and popular MGARCH models, including the vech model, the BEKK model, the CCC and DCC models as well as a model based on a dynamic specification of the eigenvalues–closely related to the popular OGARCH model. The stochastic properties of the models, in particular conditions for stationarity, ergodicity and existence of moments, are then discussed in Section 8 and Section 9 gives a brief introduction to the asymptotic analysis of MGARCH models. Finally, Section 10 discusses some simple diagnostics for model control.

2 Notation and Preliminary Economic framework

We consider an investor who decides which *securities* or *assets* to invest in. In reality there are many different possibilities, typically categorized as *money market* investments and *capital market* investments, respectively.

Money market instruments are typically characterized as securities with a life shorter than one year and includes, in particular, US Treasury Bills. These securities have different maturities and are issued every month or every week by the US Federal government. They are sold at a given market price below face value, and at maturity they pay face value to the investor. Since TBills are considered to have no risk of default and have an *ex ante* known return they are good approximation to a *risk-free* investment. Other

money market instruments include certificates of deposits (CD) which are bank deposits with a fixed term. Typically the CD is held until maturity, at which time the deposit and the earned interests can be withdrawn.

Capital market instruments have maturity longer than one year, and there are two main types. *Fixed income securities* have a certain specified payment schedule, such as a bond issued by a government of by a real estate agency. *Stocks* or *equity securities* represent an ownership claim on the assets of a firm, and future payments depend on the business performance and policy of the particular firm.

2.1 Portfolios Returns and Diversification

To introduce the notation used in the following, consider a risk-free assets with return R_f and p risky assets with returns R_i per unit invested, $i = 1, 2, ..., p$.[1] We consider the return R_i as a realization of a stochastic variable while R_f is deterministic. Collect the stochastic returns in a vector $R = (R_1, R_2, ..., R_p)'$ and denote by

$$\mu := \mathrm{E}(R) = (\mu_1, \mu_2, ..., \mu_p)' \tag{2.1}$$

the expected value and by

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix} \tag{2.2}$$

the covariance matrix, such that

$$\mathrm{V}[R_i] = \mathrm{E}\left[(R_i - \mu_i)^2\right] = \sigma_{ii} = \sigma_i^2 \tag{2.3}$$

$$\mathrm{cov}[R_i, R_j] = \mathrm{E}\left[(R_i - \mu_i)(R_j - \mu_j)\right] = \sigma_{ij}. \tag{2.4}$$

The diagonal of Ω contains the p variances, $\sigma_{ii} = \sigma_i^2 > 0$, and due to symmetry, the off-diagonal elements contain the $p(p-1)/2$ unique covariances, $\sigma_{ij} = \sigma_{ji}$.

For later use, also define the *excess return* relative to the risk-free asset as $r_i = R_i - R_f$ and $r = (r_1, r_2, ..., r_p)' = (R - \iota R_f)'$, where $\iota = (1, ..., 1)'$ is a $p \times 1$ vector of ones.

Most of this note is formulated in terms of returns. Alternatively we may formulate the problem in terms of asset prices. For asset i we let $P_{i,t}$ and $P_{i,t+1}$ denote the price at time t and $t+1$, respectively. The one period return from t to $t+1$ is given by $R_{i,t+1} = P_{i,t+1}/P_{i,t} - 1$. The theory below is static, however, and in the first part of this note we drop the explicit reference to time.

A *portfolio* of the p risky assets is characterized by the value invested in each asset. Let x_i be the value invested in asset i, such that the total value of the portfolio, or the

[1] In this note 'return' refers to *net return*, such that the pay-off in the next period from investing x in asset i is $x(1 + R_i)$. Be aware that some textbooks refer to *gross return*, $1 + R_i$.

total wealth, is

$$w = x'\iota = \sum_{i=1}^{p} x_i. \tag{2.5}$$

We will in most cases look at the *portfolio weights*, $v_i = x_i/w$, and we will characterize the portfolio by the weights, $v = (v_1, v_2, ..., v_p)'$ with $\sum_{i=1}^{p} v_i = v'\iota = 1$.

To calculate the portfolio return, consider an investor with initial wealth w_t. The number of each asset in the portfolio is given by

$$\delta_i = \frac{x_{i,t}}{P_{i,t}} = \frac{w_t v_i}{P_{i,t}}, \quad i = 1, 2, ..., p, \tag{2.6}$$

i.e. the amount invested in asset i divided with the price. The wealth at time $t+1$ is given by

$$w_{t+1} = \sum_{i=1}^{p} \delta_i P_{i,t+1} = \sum_{i=1}^{p} w_t v_i \frac{P_{i,t+1}}{P_{i,t}}, \tag{2.7}$$

and the rate of return of the portfolio is

$$\bar{R} = \frac{w_{t+1} - w_t}{w_t} = \frac{\sum_{i=1}^{p} w_t v_i \frac{P_{i,t+1}}{P_{i,t}} - w_t}{w_t} = \sum_{i=1}^{p} v_i \left(\frac{P_{i,t+1}}{P_{i,t}} - 1 \right) = \sum_{i=1}^{p} v_i R_i = v'R, \tag{2.8}$$

which is just the weighted average of individual returns. The expected return and the variance of the portfolio is given by, respectively,

$$\bar{\mu} = \mathrm{E}\left[\bar{R}\right] = \mathrm{E}\left[v'R\right] = v'\mu \tag{2.9}$$

$$\bar{\sigma}^2 = \mathrm{V}\left[\bar{R}\right] = \mathrm{V}\left[v'R\right] = v'\Omega v. \tag{2.10}$$

The portfolio may also include the risk-free asset with weight v_f, subject to $v_f + v'\iota = 1$. In this case the return is

$$\bar{R} = v'R + (1 - v'\iota) R_f = R_f + v' \left(R - \iota R_f\right) = R_f + v'r, \tag{2.11}$$

with expectation $R_f + v'(\mu - R_f)$, and variance given by (2.10).

2.2 Diversification and Risk

To illustrate the idea of diversification, consider a portfolio of risky assets with return $\bar{R} = v'R$. The expected return, $\bar{\mu} = v'\mu$, is a linear function of individual expected returns. The variance of the return, on the other hand, is given by:

$$\bar{\sigma}^2 = v'\Omega v = \sum_{i=1}^{p}\sum_{j=1}^{p} v_i v_j \sigma_{ij} = \sum_{i}\sum_{j} v_i v_j \sigma_{ij} = \sum_{i} v_i^2 \sigma_{ii} + \sum_{i}\sum_{j \neq i} v_i v_j \sigma_{ij}, \tag{2.12}$$

which is here decomposed into the p variance terms and the $p(p-1)$ covariance terms.

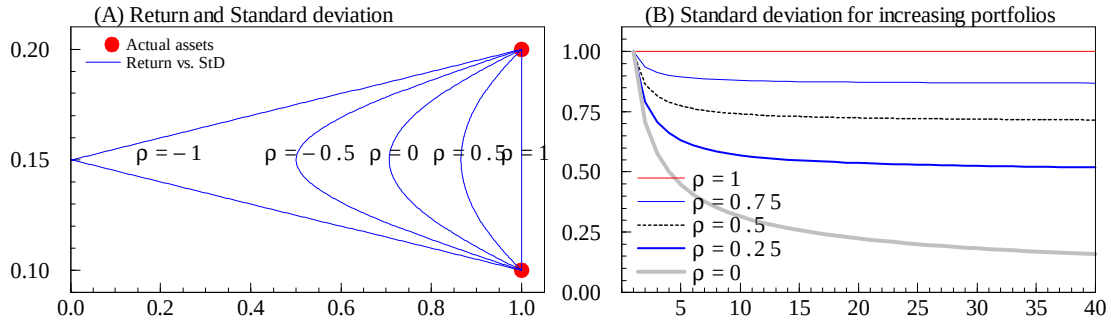

Figure 1: *Effect of diversification. Graph (A) shows the return and standard deviation of portfolios with varying weights $v = (\alpha, 1-\alpha)'$, for different values of asset covariance, ρ. (B) shows the portfolio standard deviation in the case of equal weights, $v_i = p^{-1}$, as a function of p.*

Example 2.1 (Portfolio Variance): To illustrate the portfolio variance in (2.12), consider the case of $p = 3$ assets. It then holds that

$$\bar{\sigma}^2 = v'\Omega v = v_1^2\sigma_{11} + v_2^2\sigma_{22} + v_3^2\sigma_{33} + 2v_1v_2\sigma_{21} + 2v_1v_3\sigma_{31} + 2v_2v_3\sigma_{32}.$$

where symmetry of the covariance implies that $v_1v_2\sigma_{12} = v_2v_1\sigma_{21}$. ♦

To simplify the discussion, consider the case of an equally weighted portfolio, $v_i = 1/p$ for $i = 1, 2, ..., p$. The variance in (2.12) can now be written as

$$\begin{aligned}\bar{\sigma}^2 &= \sum_i \frac{1}{p^2}\sigma_{ii} + \sum_i \sum_{j\neq i} \frac{1}{p^2}\sigma_{ij} \\ &= \frac{1}{p}\left[\sum_i \frac{1}{p}\sigma_{ii}\right] + \frac{p(p-1)}{p^2}\left[\sum_i \sum_{j\neq i} \frac{1}{p(p-1)}\sigma_{ij}\right],\end{aligned} \tag{2.13}$$

where the term in the first square bracket is the average variance and the second square bracket is the average covariance. When p is large, the uncertainty from the individual variances disappears, while the variance of the portfolio return is determined by the covariances alone. We may say that the idiosyncratic risk can be diversified away, while the systematic risk, here measured as the average covariance, cannot.

Example 2.2 (Diversification): To illustrate the idea of diversification and the importance of covariances, consider $p = 2$ assets with expected returns and covariance matrix given by

$$\mu = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where ρ is also the correlation of returns. We construct a portfolio with weights $v = (\alpha, 1-\alpha)'$ and we plot the expected value, $\bar{\mu} = v'\mu$, against the standard deviation, $\bar{\sigma} = \sqrt{v'\Omega v}$. In Figure 1 (A) we show the mean vs. the standard deviation for different

asset covariances, ρ, and varying values of α. Note that assets with perfectly correlated returns, $\rho = 1$, cannot be used for diversification and there is no decrease in portfolio variance. All correlations smaller than one allows for a diversification gain, however, and perfect negative correlation, $\rho = -1$, implies that a risk-free portfolio can be constructed. This is referred to as a *perfect hedge*.

To illustrate the result in (2.13), consider p assets with unit variance and mutual covariances (and hence correlations) of ρ. Graph 1 (B) shows the portfolio standard deviation for an equally weighted portfolio, $v_i = p^{-1}$, for an increasing number of assets, p, and for different values of ρ. Note that for increasing p, the portfolio standard deviation converges to the square root of the average covariance, $\rho^{0.5}$, showing that the idiosyncratic risk can be diversified away, while the systematic risk, measured by the average covariance, cannot. ♦

3 Opportunity Set: Minimum Variance Portfolios

In this section we seek to characterize the opportunity set of the investor. We first consider the case with no risk-free asset so that the investor constructs a portfolio from the p risky assets; this is the Markowitz (1952) mean-variance analysis. Later we introduce a risk-free asset.

3.1 Minimum Variance Portfolios Without a Risk-free Asset

First, we consider the case where the universe of potential investments does not contain a risk-free asset.

Global Minimum Variance Portfolio: Having seen the gains from diversification, we first let the investor choose portfolio weights, $v = (v_1, v_2, ..., v_p)'$, that gives the smallest possible variance, i.e.

$$\min_v v'\Omega v \tag{3.1}$$

subject to the adding-up condition:

$$v'\iota = 1. \tag{3.2}$$

This is a quadratic minimization problem under a linear constraint, which can be easily solved using the Lagrangian method. Introducing a Lagrangian multiplier for the constraint, λ, the Lagrangian function is given by

$$L(v, \lambda) = v'\Omega v + \lambda \left[1 - v'\iota\right]. \tag{3.3}$$

The first order condition with respect to v is given by the p equations

$$\frac{\partial L(v, \lambda)}{\partial v} = 2\Omega v - \lambda\iota = 0, \tag{3.4}$$

and the solution for v is

$$v = \Omega^{-1}\frac{\lambda}{2}\iota, \tag{3.5}$$

depending on the multiplier. We find λ by inserting (3.5) in (3.2):

$$\frac{\lambda}{2} = \frac{1}{\iota'\Omega^{-1}\iota},$$

and inserting back in (3.5) gives the weights

$$v_g = \frac{1}{\iota'\Omega^{-1}\iota}\Omega^{-1}\iota. \tag{3.6}$$

The portfolio defined by v_g is called the *global minimum variance portfolio.* The global minimum variance portfolio has return and variance given by, respectively,

$$\mu_g = v_g'\mu \quad \text{and} \quad \sigma_{gg} = v_g'\Omega v_g = \left(\iota'\Omega^{-1}\iota\right)^{-1}. \tag{3.7}$$

Drawing the Efficient Frontier: Next we want to draw the opportunity set of the investor in the a mean-variance diagram, i.e. all combinations of mean and variance attainable by the investor. To do so we first consider a situation where the investor has a certain required return, $\bar{\mu}^*$, in mind, and choose v to minimize the variance of the portfolio, i.e.

$$\min_v v'\Omega v \tag{3.8}$$

subject to

$$v'\mu = \bar{\mu}^* \tag{3.9}$$

$$v'\iota = 1. \tag{3.10}$$

The global minimum variance portfolio derived above minimizes the variance for $v'\mu \in \mathbb{R}$, and the new restriction in (3.9) ensures the required return $\bar{\mu}^*$. The new Lagrangian function has two restrictions:

$$L(v, \lambda_1, \lambda_2) = v'\Omega v + \lambda_1\left(\bar{\mu}^* - v'\mu\right) + \lambda_2\left(1 - v'\iota\right), \tag{3.11}$$

with multipliers λ_1 and λ_2. The first order condition with respect to v is given by

$$\frac{\partial L(v, \lambda_1, \lambda_2)}{\partial v} = 2\Omega v - \lambda_1\mu - \lambda_2\iota = 0, \tag{3.12}$$

and the solution for v is

$$v = \Omega^{-1}\left(\mu, \iota\right)\begin{pmatrix} \frac{\lambda_1}{2} \\ \frac{\lambda_2}{2} \end{pmatrix} = \Omega^{-1}\left[\frac{\lambda_1}{2}\mu + \frac{\lambda_2}{2}\iota\right]. \tag{3.13}$$

To find the multipliers, λ_1 and λ_2, consider the conditions in (3.9) and (3.10) with (3.13) inserted, i.e.

$$\frac{\lambda_1}{2}\mu'\Omega^{-1}\mu + \frac{\lambda_2}{2}\iota'\Omega^{-1}\mu = \bar{\mu}^* \tag{3.14}$$

$$\frac{\lambda_1}{2}\mu'\Omega^{-1}\iota + \frac{\lambda_2}{2}\iota'\Omega^{-1}\iota = 1, \tag{3.15}$$

or using matrix notation,

$$\begin{pmatrix} \mu'\Omega^{-1}\mu & \iota'\Omega^{-1}\mu \\ \mu'\Omega^{-1}\iota & \iota'\Omega^{-1}\iota \end{pmatrix} \begin{pmatrix} \frac{\lambda_1}{2} \\ \frac{\lambda_2}{2} \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \frac{\lambda_1}{2} \\ \frac{\lambda_2}{2} \end{pmatrix} = \begin{pmatrix} \bar{\mu}^* \\ 1 \end{pmatrix}, \tag{3.16}$$

where we have defined the scalars $a = \mu'\Omega^{-1}\mu$, $b = \iota'\Omega^{-1}\mu$, and $c = \iota'\Omega^{-1}\iota$, and A is the 2×2 matrix. Using the result for the inverse of a 2×2 matrix,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} = \frac{1}{d}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}, \tag{3.17}$$

where $d = ac - b^2$ is the determinant, the solution is given by:

$$\begin{pmatrix} \frac{\lambda_1}{2} \\ \frac{\lambda_2}{2} \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} \begin{pmatrix} \bar{\mu}^* \\ 1 \end{pmatrix} = \frac{1}{d}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix} \begin{pmatrix} \bar{\mu}^* \\ 1 \end{pmatrix} = \frac{1}{d}\begin{pmatrix} -b + c\bar{\mu}^* \\ a - b\bar{\mu}^* \end{pmatrix}. \tag{3.18}$$

Inserting in (3.13) gives the optimal portfolio weights:

$$v = \Omega^{-1}(\mu, \iota) A^{-1} \begin{pmatrix} \bar{\mu}^* \\ 1 \end{pmatrix} = \frac{1}{d}\Omega^{-1}\left((-b + c\bar{\mu}^*)\mu + (a - b\bar{\mu}^*)\iota\right) = \phi_0 - \phi_1\bar{\mu}^*, \tag{3.19}$$

which is a linear function of the required return, $\bar{\mu}^*$, with

$$\phi_0 = \Omega^{-1}(a\iota - b\mu)/d \quad \text{and} \quad \phi_1 = \Omega^{-1}(c\mu - b\iota)/d \tag{3.20}$$

not depending on $\bar{\mu}^*$. The variance of the portfolio is $\bar{\sigma}^2 = v'\Omega v$, and the standard deviation is $\bar{\sigma} = \sqrt{v'\Omega v}$. A plot of $\bar{\mu}^*$ against the corresponding $\bar{\sigma}$ over varying required return gives the so-called *minimum variance frontier* or *efficient frontier*. The linearity in (3.19) implies that if v_a and v_b are two minimum variance portfolios with expected returns $\mu_a \neq \mu_b$ then any minimum variance portfolio can be written as a linear combination of v_a and v_b.

All combinations of mean and variance inside the frontier are attainable by portfolio selection, but if the investor is only interested in the first two moments of returns, he will always choose a portfolio on the part of the frontier above the global minimum variance; this part is called *mean-variance efficient*. The opportunity set reflects the *mean-variance trade-off*: The investor can get a higher return than the return of the global minimum variance portfolio, $v_g'\mu$, but only by taking a higher risk.

Example 3.1 (Minimum Variance Frontier): Consider the case of $p = 3$ risky assets with expected returns and covariance matrix given by

$$\mu = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 0.0625 & & \\ 0.0700 & 0.1225 & \\ 0.1050 & 0.0840 & 0.360 \end{pmatrix}.$$

In this case, $a = 0.865$, $b = 2.398$, $c = 20.093$, and $d = 11.631$. The minimum variance coefficients in (3.19) are given by

$$v = \begin{pmatrix} 2.478 \\ -0.956 \\ -0.522 \end{pmatrix} - \begin{pmatrix} 7.660 \\ -5.321 \\ -2.340 \end{pmatrix} \bar{\mu}^*.$$

For a given required risk-premium of $\bar{\mu}^* = 0.35$, say, the portfolio is given by

$$v = \begin{pmatrix} -0.203 \\ 0.907 \\ 0.297 \end{pmatrix},$$

so for each dollar to be invested, we should sell 0.2 units of asset 1, and buy 0.9 and 0.3 units of asset 2 and 3, respectively. The variance of the portfolio is 0.142, and the standard deviation is 0.376. The portfolio choice is illustrated in Figure 2 (A).

The results emphasize the gains from diversification. To illustrate, graph 2 (B) compares the results above with the case where only asset 1 and 2 are available. With only two assets the portfolio is a linear combination of the assets and they will both be on the frontier. With an additional asset the attainable combinations increase, allowing a smaller variance for each required $\bar{\mu}^*$. ♦

3.2 Including a Risk-Free Asset

Now consider the case where the investor may invest also in a risk-free asset with return R_f. We now have $p+1$ assets with portfolio weights $v = (v_1, ..., v_p)'$ to the risky assets and $v_f = 1 - v'\iota$ for the risk-free asset, or the $p+1$ dimensional weight $w = (v', v_f)'$ with $w'\iota = 1$. The return of the portfolio is

$$\mathrm{E}(\bar{R}) = v'\mu + (1 - v'\iota)\, R_f. \tag{3.21}$$

Global Minimum Variance Portfolio: In this case the global minimum variance portfolio is trivial, because we can get a variance of zero by only investing in the risk free asset, $v_f = 1$.

Drawing the Efficient Frontier: To draw the efficient frontier, minimize the variance for a required return, i.e.

$$\min_v v'\Omega v \tag{3.22}$$

subject to

$$v'\mu + (1 - v'\iota)\, R_f = \bar{\mu}^*. \tag{3.23}$$

The condition in (3.10) is replaced by $w'\iota = 1$, but this restriction is not binding because we minimize over v and determine v_f residually.

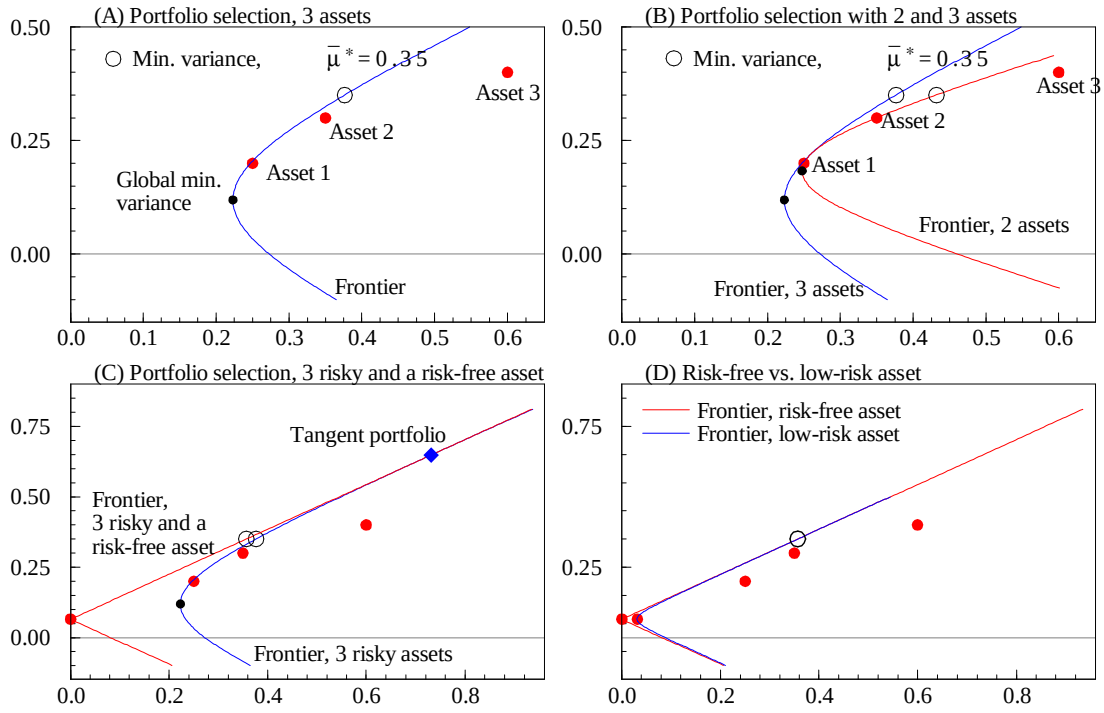

Figure 2: *Mean-variance analysis: Opportunity set of the investor with and without the availability of a risk-free asset. The hollow circles, ◯, indicate the optimal choices for an (arbitrary) return of 35%, cf. Example 3.1.*

The Lagrangian function for the minimization is

$$L(v, \lambda) = v'\Omega v + \lambda \left[\bar{\mu}^* - v'\mu - (1 - v'\iota) R_f\right], \tag{3.24}$$

with first order condition with respect to v given by

$$\frac{\partial L(v, \lambda)}{\partial v} = 2\Omega v - \lambda (\mu - \iota R_f) = 0. \tag{3.25}$$

The solution for v is

$$v = \Omega^{-1} \left[\frac{\lambda}{2} (\mu - \iota R_f)\right], \tag{3.26}$$

depending now on the expected excess returns, $\mu_i - R_f$. We find λ by inserting (3.26) in (3.23):

$$\frac{\lambda}{2} = \frac{\bar{\mu}^* - R_f}{(\mu' - \iota' R_f)\, \Omega^{-1} (\mu - \iota R_f)}, \tag{3.27}$$

and find the minimum variance weights as

$$v = \frac{(\bar{\mu}^* - R_f)}{(\mu' - \iota' R_f)\, \Omega^{-1} (\mu - \iota R_f)} \Omega^{-1} (\mu - \iota R_f) = \tilde{\kappa} \tilde{v}_q, \tag{3.28}$$

with $\tilde{v}_q = \Omega^{-1} (\mu - \iota R_f)$. If we scale $\tilde{v}_q$ to have elements that sum to one, i.e.

$$v_q = \frac{1}{\iota' \Omega^{-1} (\mu - \iota R_f)} \Omega^{-1} (\mu - \iota R_f), \tag{3.29}$$

we get a set of weights defining a portfolio, $v_q'R$, known as the *tangent portfolio*, q, that does not depend on $\bar{\mu}^*$, and the minimum variance weights are linear in the tangent portfolio weights,

$$v = \frac{(\bar{\mu}^*\iota' - R_f\iota')\,\Omega^{-1}\,(\mu - \iota R_f)}{(\mu' - R_f\iota')\,\Omega^{-1}\,(\mu - \iota R_f)} v_q = \kappa v_q, \tag{3.30}$$

with the remaining $v_f = 1 - v'\iota = 1 - \kappa$ for the risk-free asset

Here it is worth noting that the minimum variance weights, v, are proportional to the tangent portfolio weights, v_q, with coefficient κ that depends on the required return, $\bar{\mu}^*$. The mean and standard deviation of the tangent portfolio are given by

$$\bar{\mu}_q = R_f + v'\,(\mu - \iota R_f) = R_f + \kappa v_q'\,(\mu - \iota R_f) \quad \text{and} \quad \bar{\sigma}_q = (v'\Omega v)^{\frac{1}{2}} = \kappa\left(v_q'\Omega v_q\right)^{\frac{1}{2}}, \tag{3.31}$$

which are both linear functions of κ. This implies that the minimum variance frontier in the presence of a risk-free asset traces out a linear function in the $(\bar{\sigma}, \bar{\mu})-$space; the intercept is $(0, R_f)$ and the slope is

$$S(q) = \frac{v_q'\,(\mu - \iota R_f)}{\left(v_q'\Omega v_q\right)^{\frac{1}{2}}} = \frac{\mathrm{E}(r_q)}{\bar{\sigma}_q}. \tag{3.32}$$

Remark 3.1 (**Sharpe Ratio**). *The excess return scaled with the standard deviation, calculated in (3.32) for the tangent portfolio q, is known as the* Sharpe ratio *of a portfolio. We note that the tangent portfolio, and all portfolios on the efficient frontier, has the highest attainable Sharpe ratio. To compare different portfolios, the Sharpe ratio is sometimes used as a summary measure. For later use we state the formula for the linear curve as*

$$\bar{\mu} = R_f + \frac{\mu_q - R_f}{\sigma_q}\bar{\sigma}, \tag{3.33}$$

which is called the capital allocation line *(CAL).*

An additional implication of the solution in (3.30) is that for any required return, the investor will always hold a proportion of the same risky portfolio, q, combined with the risk-free asset. This is sometimes referred to as the *two-fund* of *mutual-fund theorem*: If there existed a mutual fund with a portfolio equal to the tangent portfolio, q, then an investor with any required return, $\bar{\mu}^*$, would always be willing to invest in the mutual fund and in the risk-free asset.

Example 3.2 (Minimum Variance Frontier, Risk-Free Asset): Now reconsider the case in Example 3.1 but include the possibility of a risk-free asset with return 0.065, say. That gives the new linear frontier in graph 2 (C). The tangent portfolio is given by

$$v_q = \begin{pmatrix} -2.497 \\ 2.500 \\ 0.997 \end{pmatrix}$$

with a return of 0.649 and a standard deviation of 0.732. For a required return of $\bar{\mu}^* = 0.35$, we find $\kappa = 0.488$ and the minimum variance weights are given by

$$v = \begin{pmatrix} -1.218 \\ 1.219 \\ 0.486 \end{pmatrix},$$

with the remaining $v_f = 1 - \kappa = 0.512$ to the risk-free asset. The variance of the portfolio is 0.127, and the standard deviation is 0.357, slightly smaller than the portfolio of risky assets alone.

The downward sloping arm in the graph represents a negative weight to (i.e. short position in) the tangent portfolio and a positive weight to the risk-free asset. This is obviously never efficient. ♦

Finally, note that mechanics of the solution in the case of a risk-free asset looks very different from the situation with only risky assets, but the practical difference is only minor and the presence of a low-risk stochastic return will give very similar results. Graph 2 (D) shows the minimum variance frontier in the presence of a risk-free return of 0.065 and the frontier prevailing if an expected return $\mu_4 = 0.065$ has some small variance, here $\mathrm{V}(R_4) = 0.001$.

3.3 Disallowing Short Sales

The optimizations above impose that portfolio weights sum to one, but they may be both positive and negative. A negative portfolio weight corresponds to the investor selling an asset that he does not have, known as *short-selling.* In practice a short sale implies that the asset is physically sold, which is typically made possible because the brokerage firm borrows the asset from another investor and lend it to the short-selling investor. At some future point in time the investor buys back the asset and returns it back to the lender. The original owner is fully compensated during the trades and is normally not informed that someone borrowed the asset. Typically, an interest rate payment is involved in the loan.

Drawing the Efficient Frontier: In some situations, we want to analyze the portfolio decision without the possibility of short sales, and consider first the case of no risk-free asset. In that case the investor minimizes the variance of the return on the portfolio of p risky assets given a required return of $\bar{\mu}^*$, under the additional restriction that weights should be non-negative. In this case we may state the problem as

$$\min_{v} v'\Omega v \tag{3.34}$$

subject to

$$v'\mu = \bar{\mu}^*, \quad v'\iota = 1, \quad \text{and} \quad v_i \geq 0, \quad i = 1, 2, ..., p. \tag{3.35}$$

Due to the inequality constraints in (3.35), the minimization problem is more complicated than with short sales and cannot be solved using the simple Lagrangian method above.

Formally, the problem in (3.34) and (3.35) is a non-linear programming problem, and because the criteria function is quadratic, it is referred to as a *quadratic programming (QP) problem*. Given that the $p+2$ constraints in (3.35) are feasible, which in the portfolio selection case means that we can actually obtain a return of $\bar{\mu}^*$ without short-selling, then the minimization problem has a globally unique solution if the quadratic criteria (3.34) is a strictly convex function of v. A sufficient condition for convexity is that Ω is positive definite. This holds for the covariance matrix unless some assets are perfectly correlated; perfectly correlated returns imply a perfect hedge, however, which is equivalent to the case with a risk-free asset that we discuss below.

When drawing the frontier for the case of no short sales, i.e. the minimum variance for a given return, the only concern is whether a given return is feasible or not without short sales. Intuitively, we cannot obtain an expected portfolio return higher (lower) than the highest (lowest) individual expected return, in which case the portfolio weights are given by a unit vector. All returns between $\min(\mu)$ and $\max(\mu)$ are feasible, but will have a variance which is never smaller than the case allowing short sales.

The *QP* problem can be solved by transforming it to a (much larger) linear programming problem, that can be solved using e.g. a simplex algorithm. The corresponding linear programming problem is composed by the first-order conditions of the Lagrangian wrt. v and the equality constraints. Finally, the inequality constraints are transformed to equality constraints by introducing auxiliary slack variables. In practice the *QP* problem is solved numerically, and many software packages include efficient algorithms for *QP* problems. The Ox notation is briefly given in Example 3.3.

Example 3.3 (Solving Quadratic Programming Problems in Ox): The matrix programming package Ox solves the following, slightly more general, quadratic programming problem:

$$\min_x \left\{ \frac{1}{2} x'Gx + x'g \right\}$$

subject to

$$Ax \geq b, \quad Cx = d, \quad \text{and} \quad x_{lo} \leq x \leq x_{hi}.$$

In this case there is an additional linear term in the criteria, the linear restriction may involve equalities or inequalities, and there are upper and lower bounds for the weights. The general Ox notation is

$$[\texttt{iret}, \texttt{x}, \lambda] = \texttt{SolveQP(G,g,A,b,C,d,x}_{\texttt{lo}}\texttt{,x}_{\texttt{hi}}\texttt{);}$$

Where redundant constraints are indicated by empty vectors, <>. In the case of variance minimization for portfolio choice we would set

$$\texttt{G} = \Omega, \quad \texttt{g} = 0, \quad \texttt{A} = <>, \quad \texttt{b} = <>, \quad \texttt{C} = (\iota, \mu)', \quad \texttt{d} = (1, \bar{\mu}^*)', \quad \texttt{x}_{\texttt{lo}} = 0, \quad \texttt{x}_{\texttt{hi}} = <>,$$

where the zeros are $p \times 1$ vectors

The algorithm returns an array of three components: An indicator for whether the constraints are feasible, `iret` $=$ `0` if a feasible solution is found; the optimal solution, `x`; and the vector of Lagrangian multipliers for the constraints, λ, with zeros for non-binding constraints. ⧫

Example 3.4 (Minimum Variance Frontier, No Short-Sales): Reconsider again the case in Example 3.1 but disallow the possibility of short sales. Graph 3 (A) show the unrestricted frontier also reported in Graph 2 (A) together with the no-short-sale frontier. A section of the new frontier coincides with the unrestricted frontier, but in general the opportunity set is smaller. With unlimited short-selling it is possible to get an infinite expected return (at the expense of an infinite variance). This is not possible without short-selling and we can never get a higher expected return than the highest expected return of the available assets; in this case v is a unit vector. In the present case, the assets are highly positively correlated, and it is not possible to get a portfolio variance smaller than the smallest individual variance. This is not a general result, however, and to illustrate we draw the frontiers for a case with uncorrelated return,

$$\mu = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \end{pmatrix}, \quad \text{and} \quad \Omega = \begin{pmatrix} 0.0625 & & \\ 0 & 0.1225 & \\ 0 & 0 & 0.360 \end{pmatrix}.$$

This case is more favorable in general, and the restriction on short-sale does not change the opportunity set very much for low-variance allocations. ⧫

Allowing A Risk-Free Asset: In the case of a risk-free asset the investors problem is to minimize the variance subject to non-negative weights to all assets, $v_i \geq 0$ and $v_f = 1 - v'\iota \geq 0$, and subject to a certain required return, i.e.

$$\begin{aligned} v'\mu + (1 - v'\iota) R_f &= \bar{\mu}^* \\ v' (\mu - \iota R_f) &= \bar{\mu}^* - R_f. \end{aligned}$$

Formally the problem is formulated as

$$\min_v v'\Omega v \tag{3.36}$$

subject to

$$v'(\mu - \iota R_f) = \bar{\mu}^* - R_f, \quad v'\iota \leq 1, \quad \text{and} \quad v_i \geq 0, \quad i = 1, 2, ..., p. \tag{3.37}$$

This is a quadratic programming problem involving equality and inequality constraints, and it can be solved numerically for v.

Example 3.5 (Minimum Variance Frontier, Risk-free Asset, No Short-Sales): Graph 3 (C) shows the frontiers in the presence of a risk-free asset from Example 3.2.

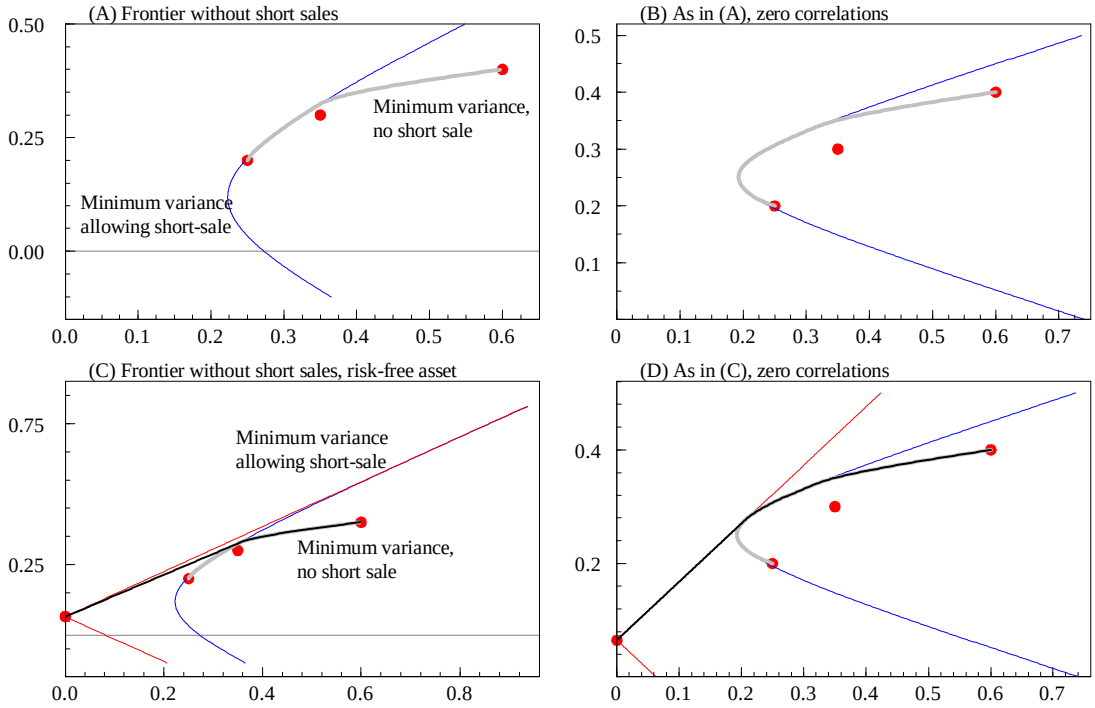

Figure 3: *Minimum variance frontiers without short sales. Compare Figure 2.*

Without short-selling the frontier is still linearly increasing to a tangent portfolio, corresponding to a maximization of the Sharpe ratio. It is not linear after that, however, because the investor cannot short-sell the risk-free asset to gear the investment, and the frontier ends at the maximum individual return. Graph 3 (D) shows the same for the case of uncorrelated return. ♦

4 Utility and the Optimal Portfolio Selection

The above analysis is purely descriptive. It derives the mean-variance frontier, i.e. all attainable combinations of mean and variance. If the investor has preferences formulated in terms of mean and variance only, then the investor will always choose a mean-variance efficient portfolio, i.e. a portfolio on the minimum variance frontier. The exact choice depends on the degree of risk aversion.

To complete the analysis, and pin down a unique optimal portfolio, we can assume a particular utility function and maximize the utility over the attainable combinations of mean and variance. A simple example of a utility function defined over different portfolio returns is the mean-variance expected utility given by

$$\mathrm{E}\left[u(\bar{R})\right] = \mathrm{E}\left[\bar{R}\right] - \frac{\theta}{2}\mathrm{V}[\bar{R}] = \bar{\mu} - \frac{\theta}{2}\bar{\sigma}^2, \tag{4.1}$$

where $\bar{R} = v'R$ depends on portfolio weights, v, and the parameter θ measures the degree of risk aversion. The formulation in (4.1) just reflects the trade-off between return and

risk.

Example 4.1 (Optimal Portfolio Selection): To illustrate, reconsider the case in Example 3.1 and assume the simple mean-variance utility function in (4.1). Graph 4 (A) and (B) show iso-utility curves and the implied optimal choice of portfolio for $\theta = 4$ and $\theta = 2$, respectively. Note that the more risk-averse investor, $\theta = 4$, benefits more from the presence of a risk-free rate. Note again, that in the presence of a risk-free asset, all investors will hold a combination of the tangent portfolio and the risk-free asset, with weights determined by their risk aversion, cf. the two fund separation theorem. ♦

The derivation so far was made in two steps: First finding the opportunity set in terms of the mean-variance efficient frontier and then the optimal portfolio. An alternative approach would be to maximize the expected utility in (4.1) directly. In the case of a risk-free asset this amounts to

$$\max_{v} \left\{ \bar{\mu} - \frac{\theta}{2}\bar{\sigma}^2 \right\} = \max_{v} \left\{ R_f + v'(\mu - \iota R_f) - \frac{\theta}{2} v'\Omega v \right\}. \tag{4.2}$$

The solution is identical to (3.26) with the multiplier $\lambda/2$ replaced by the utility parameter, θ. To characterize the efficient frontier we can trace the solution for varying degree of risk aversion. We may therefore think of the mean-variance analysis in two ways: Either as a simple characterization of the choice set in terms of the first two moments of returns, or, alternatively, as the optimal portfolio selection given the simple mean-variance utility function.

4.1 General Utility Maximization

As an alternative motivation for the mean-variance approach, it turns out that the mean-variance utility function in (4.1) is identical to a more standard choice of utility function if the returns are Gaussian. In particular, consider the negative exponential utility function defined over total wealth, w:

$$u(w) = -\exp(-\theta w), \quad \theta > 0. \tag{4.3}$$

This utility function is increasing and concave,

$$\frac{\partial u(w)}{\partial w} = \theta \exp(-\theta w) > 0 \quad \text{and} \quad \frac{\partial^2 u(w)}{\partial w \partial w} = -\theta^2 \exp(-\theta w) < 0, \tag{4.4}$$

reflecting risk aversion. Moreover, note that the Arrow-Pratt measure of absolute risk-aversion is constant,

$$-\frac{\frac{\partial^2 u(w)}{\partial w \partial w}}{\frac{\partial u(w)}{\partial w}} = \frac{\theta^2 \exp(-\theta w)}{\theta \exp(-\theta w)} = \theta, \tag{4.5}$$

and we refer to (4.6) as having *constant absolute risk aversion* (CARA). Here we think of wealth as being determined by portfolio returns. For a particular portfolio, the wealth would be $w = w_0(1 + \bar{R})$, where w_0 is the initial wealth.

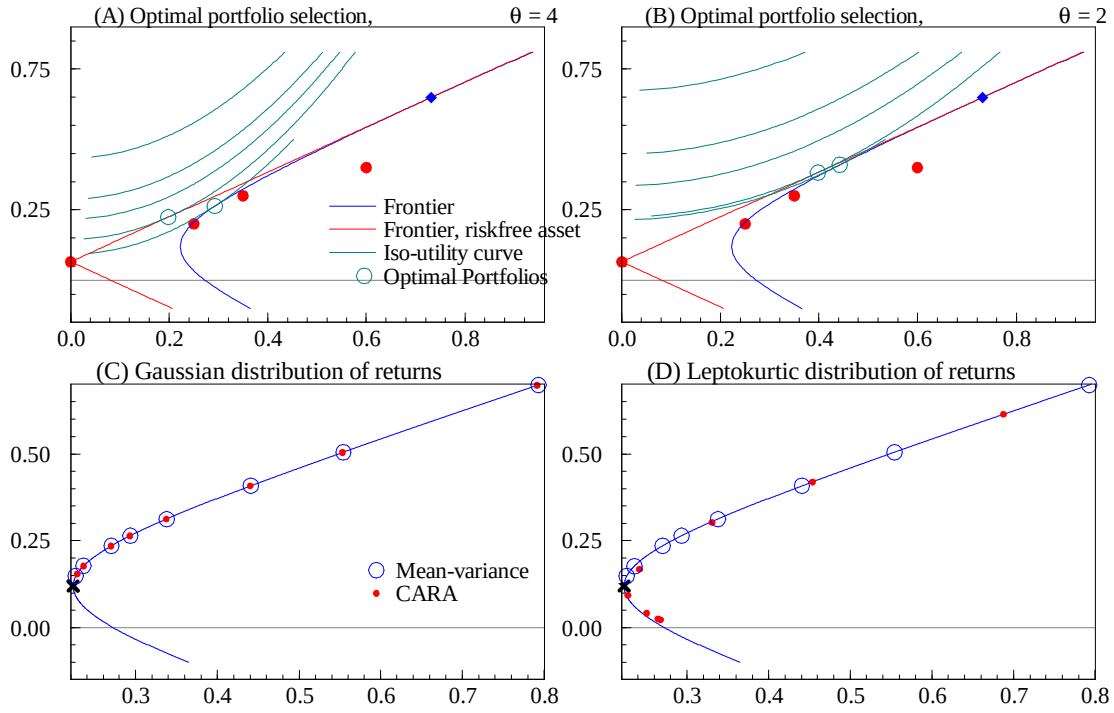

Figure 4: *Optimal portfolio selection. (A) and (B) show the optimal portfolio selection with mean-variance preferences. (C) and (D) compare the mean-variance utility function with the negative exponential CARA utility function for Gaussian and Leptokurtic returns.*

Now assume that returns follow a multivariate Gaussian distribution, i.e. $R \overset{D}{=} \mathrm{N}(\mu, \Omega)$. A given portfolio defines the return $\bar{R} = v'R$, which is also Gaussian, $\bar{R} \overset{D}{=} \mathrm{N}(\bar{\mu}, \bar{\sigma}^2)$, with $\bar{\mu} = v'\mu$ and $\bar{\sigma}^2 = v'\Omega v$. If follows that wealth is distributed as

$$w \overset{D}{=} \mathrm{N}(\bar{\mu}_w, \bar{\sigma}_{ww}), \quad \text{with} \quad \bar{\mu}_w = w_0(1+\bar{\mu}) \quad \text{and} \quad \bar{\sigma}_{ww} = w_0^2\bar{\sigma}^2.$$

We write the density function for the stochastic variable w as

$$f(w) = \frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\frac{(w-\bar{\mu}_w)^2}{2\bar{\sigma}_{ww}}\right\}.$$

The investor is assumed to maximize *expected* utility and has to evaluate the integral

$$\begin{aligned}
\mathrm{E}\left[u(w)\right] &= \int_{-\infty}^{\infty} u(w) f(w) dw \\
&= \int_{-\infty}^{\infty} \left(-\exp(-\theta w)\right) \frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\frac{(w-\bar{\mu}_w)^2}{2\bar{\sigma}_{ww}}\right\} dw \\
&= \int_{-\infty}^{\infty} -\frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\left(\theta w + \frac{(w-\bar{\mu}_w)^2}{2\bar{\sigma}_{ww}}\right)\right\} dw. \qquad (4.6)
\end{aligned}$$

To simplify the expression we want to collect terms that depend on w and those that do not. It turns out to be useful to consider the calculation

$$\frac{\left(w - \bar{\mu}_w + \theta\bar{\sigma}_{ww}\right)^2}{2\bar{\sigma}_{ww}} = \frac{\left(w-\bar{\mu}_w\right)^2 + \theta^2\bar{\sigma}_{ww}^2 + 2\theta\bar{\sigma}_{ww}\left(w-\bar{\mu}_w\right)}{2\sigma_{ww}}$$

$$= \frac{(w-\bar{\mu}_w)^2}{2\bar{\sigma}_{ww}} + \frac{\theta^2\bar{\sigma}_{ww}}{2} + \theta w - \theta\bar{\mu}_w, \tag{4.7}$$

which shows that the exponent in (4.6) can be written as

$$\theta w + \frac{(w-\bar{\mu}_w)^2}{2\bar{\sigma}_{ww}} = \frac{(w-\bar{\mu}_w+\theta\bar{\sigma}_{ww})^2}{2\bar{\sigma}_{ww}} + \theta\left(\bar{\mu}_w - \frac{\theta}{2}\bar{\sigma}_{ww}\right). \tag{4.8}$$

Inserting in (4.6) and taking constants (not depending on the stochastic w) outside the integral, yields

$$\begin{aligned}
\mathrm{E}\left[u(w)\right] &= \int_{-\infty}^{\infty} -\frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\frac{(w-\bar{\mu}_w+\theta\bar{\sigma}_{ww})^2}{2\bar{\sigma}_{ww}} - \theta\left(\bar{\mu}_w - \frac{\theta}{2}\bar{\sigma}_{ww}\right)\right\} dw \\
&= \int_{-\infty}^{\infty} -\frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\frac{(w-\bar{\mu}_w+\theta\bar{\sigma}_{ww})^2}{2\bar{\sigma}_{ww}}\right\} \exp\left\{-\theta\left(\bar{\mu}_w - \frac{\theta}{2}\bar{\sigma}_{ww}\right)\right\} dw \\
&= -\exp\left\{-\theta\left(\bar{\mu}_w - \frac{\theta}{2}\bar{\sigma}_{ww}\right)\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}_{ww}}} \exp\left\{-\frac{(w-\bar{\mu}_w+\theta\bar{\sigma}_{ww})^2}{2\bar{\sigma}_{ww}}\right\} dw.
\end{aligned}$$

Now note that the latter integral is just the cumulated distribution function (cdf) of a Gaussian distribution, $\mathrm{N}(\bar{\mu}_w + \theta\bar{\sigma}_{ww}, \bar{\sigma}_{ww})$, which equals one per definition. Expected utility can be therefore be written as

$$\mathrm{E}\left[u(w)\right] = -\exp\left\{-\theta\left(\bar{\mu}_w - \frac{\theta}{2}\bar{\sigma}_{ww}\right)\right\}. \tag{4.9}$$

To maximize the negative exponential we have to minimize the exponent, which amounts to maximizing the mean-variance expected utility function in (4.1). With Gaussian returns, we may therefore rationalize the mean variance expected utility function as being equivalent to a CARA utility function.

Alternatively, we can rationalize the mean-variance analysis as being a second order approximation of utility maximization. To illustrate, let $u(w)$ be a utility function with $w = w_0(1+\bar{R})$. Now consider a Taylor series expansion of $u(\cdot)$ around the expected terminal wealth

$$u(w) = u(\mathrm{E}[w]) + u'(\mathrm{E}[w])\,(w-\mathrm{E}[w]) + \frac{1}{2}u''(\mathrm{E}[w])\,(w-\mathrm{E}[w])^2 + Z_3, \tag{4.10}$$

where $u'(\cdot)$, $u''(\cdot)$, and $u^{(i)}(\cdot)$ are derivatives, and $Z_3 = \sum_{i=3}^{\infty} \frac{1}{i!}u^{(i)}(\mathrm{E}[w])\,(w-\mathrm{E}[w])^i$ is the remainder term. If $\mathrm{E}[Z_3]$ is small we can approximate expected utility by

$$\mathrm{E}[u(w)] \approx u(\mathrm{E}[w]) + \frac{1}{2}u''(\mathrm{E}[w])\,\mathrm{V}(w). \tag{4.11}$$

Since $u(\cdot)$ is increasing and concave there is a preference for higher expected return and a variance aversion.

Pulley (1983) and Kroll, Levy, and Markowitz (1984) compare utility maximization and the mean-variance approach for some empirically relevant cases and they both conclude that the mean-variance analysis is a good approximation to more realistic or conventional utility approaches.

Example 4.2 (Mean-Variance Analysis vs. Utility Maximization): To illustrate the importance of the return distribution and utility function graph 4 (C) and (D) show the mean-variance efficient frontier in the case from Example 3.1. They also depict the mean and standard deviation of the optimal portfolio choices given the mean-variance utility function in (4.1) and the CARA utility function in (4.6) for $\theta \in (1, 1.5, 2, 3, 4, 5, 10, 20)$.

In graph (C) the returns are Gaussian. In this case the two utility functions coincide and for increasing risk aversion the optimal portfolio converges to the global minimum variance. In (D) the returns are Leptokurtic, i.e. having fatter tails than the normal distribution, here taken from a student$-t$ distribution with 6 degrees of freedom. In the leptokurtic case the optimal portfolios for the two utility functions differ. For low values of θ the CARA optimal portfolios are on the efficient frontier, but for increasing θ that is no longer the case. Intuitively the risk of the portfolios are no-longer well approximated by the variance. For $\theta = 4$, say, the distribution of the optimal portfolio returns are characterized by the following moments:

	Mean-variance	Negative Exponential
Mean	0.264	0.093
Variance	0.293	0.228
Skewness	0.000	−0.008
Kurtosis	5.318	4.454

According to the CARA utility function it is preferable to trade-off a huge decrease in the mean return for a reduction in variance and kurtosis. For larger θ, an increasing variance is even tolerated in order to lower the fourth and higher-order moments.

In all cases $w = 1+\bar{R}$, and the expected utility defined by the integral $\int u(w)f(w)dw$ is approximated by the average of $u(w)$ over 10^6 random numbers drawn from the specified $f(w)$. ♦

Example 4.3 (Portfolio Selection for Industry Indices): The examples above are all artificial to illustrate the mathematics. We continue by considering an empirical example based on Standard & Poor's 500 composite index, SP500, and 10 industry-sorted indices, that is: Consumer Discretionary (ConsD), Consumer Staples (ConsS), Energy (Energy), Financials (Financial), Health Care (Health), Industrials (Industrials), Information Technology (IT), Materials (Materials), Telecommunication Services (Tele), and Utilities (Utils). We use daily data, $t = 1, 2, ..., 5046$, covering September 11, 1989 to November 11, 2009. First recall that the considered portfolio theory is static and based on our expected period mean, μ, and variance, Ω. For the empirical implementation we use the historical data to make expectations of the coming period. The most simple implementation is to use the full sample averages of the daily percentage returns,

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t, \quad \text{and} \quad \hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'. \tag{4.12}$$

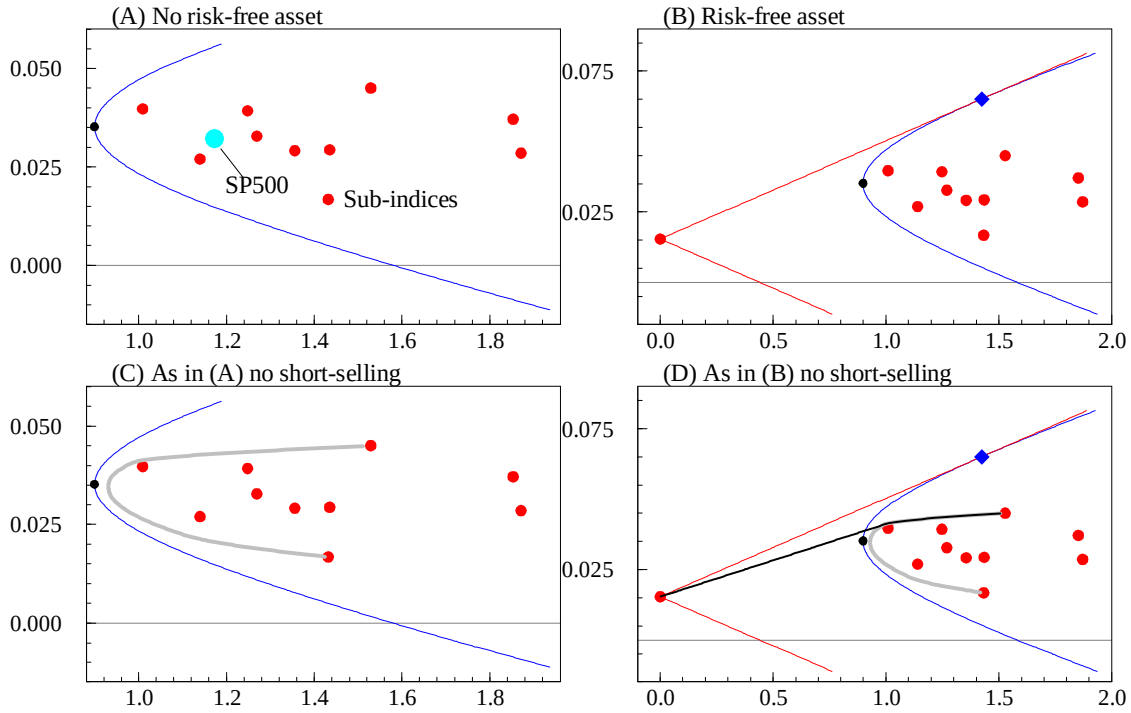

Figure 5: *Full sample frontiers for 10 industry-sorted sub-indices of SP500.*

As will be shown in Lemma 6.1 in Section 6.1, $\hat{\Omega}$ in (4.12) is the maximum likelihood estimator of Ω under the assumption that R_t is i.i.d. $\mathrm{N}(\mu, \Omega)$.

Based on $\hat{\mu}$ and $\hat{\Omega}$ we get the opportunity sets reported in Figure 5. Graph (A) shows the case with no risk-free asset. The average return on the sub-indices have been largely comparable, but the variances differ greatly. The total SP500 has a low variance compared to most sub-indices, but the full diversification gain is not utilized, and a lower variance could have been obtained by a difference weighting of the sub-indices. Graph (C) repeats the analysis in the case where short-selling is not allowed. Graph (B) and (D) extend the results to the case of a risk-free asset. Here we use the return on the 3 month Tbill as a risk-free investment. ♦

Remark 4.1 (Problems and Refinements). *One problem with the above approach is that the estimation error may be large relative to the actual mean, and the mean-variance analysis may be quite sensitive to that. If the returns are i.i.d., then the variance of the estimators are given by*

$$\mathrm{V}(\hat{\mu}_i) = T^{-1}\sigma_{ii} \quad \textit{and} \quad \mathrm{V}(\hat{\sigma}_{ii}) = 2T^{-1}\sigma_i^4,$$

where σ_i is the standard deviation of R_i. For monthly data, reasonable values could be $\mu = 1\%$ and $\sigma = 4\%$, and with 5 years of monthly data, i.e. $T = 60$, the estimation uncertainty would be $\mathsf{se}(\hat{\mu}) = 4/\sqrt{60} = 0.52\%$, which is large compared to the mean return of 1%, and $\mathsf{se}(\hat{\sigma}_i^2) = \sqrt{2}4^2/\sqrt{60} = 2.92\%$. In such a situation the estimates may not be very helpful. Chopra and Ziemba (1993) consider the relative importance of errors in the

mean and covariance, and conclude that estimating the mean is particularly important, and they suggest to only use the estimated covariance, and set all means equal.

If the number of assets is very large, for example all stocks listed on the stock exchange, then it is particularly hard to obtain reliable estimates of the $p \times p$ *covariance matrix. In particular* $\hat{\Omega}$ *may be close to singular and* $\hat{\Omega}^{-1}$*, that enters the formulas, will be extremely sensitive to uncertainty in* $\hat{\Omega}$*. One possible solution is to impose more structure on the matrix before estimation, i.e. to limit the number of free parameters to estimate.*

If p *is reasonably small, a natural way to estimate the expected mean and variance, is to estimate a multivariate autoregressive conditional heteroskedastic (MARCH) model for a set of historical data, forecast the time varying conditional mean and variance and plug these into the above framework. There is a large empirical literature on this.*

5 Introduction to Multivariate ARCH Models

The (vast) class of multivariate (G)ARCH (MGARCH) models are typically – as for univariate GARCH models – specified by *(i):* a conditional distribution of the p-dimensional vector of observations X_t, and *(ii):* a specification of the time-varying conditional covariance Ω_t. Consider here the case of X_t being normal distributed conditionally on the past realizations $\mathcal{F}_{t-1} = (X_{t-1}, X_{t-2}, ...)$ with mean zero and *conditional* covariance matrix Ω_t. That is,

$$X_t | \mathcal{F}_{t-1} \overset{D}{=} N_p \left(0_p, \Omega_t\right).$$

Here $0_p = \underbrace{(0, 0, ..., 0)}_{p \text{ elements}}'$ while Ω_t is a $(p \times p)$ covariance matrix with entries $(\sigma_{ijt})_{i,j=1,2,...,p}$,

$$\Omega_t = \begin{pmatrix} \sigma_{11t} & \sigma_{12t} & \cdots & \sigma_{1pt} \\ \sigma_{21t} & \sigma_{22t} & & \\ \vdots & & & \vdots \\ \sigma_{p1t} & & \cdots & \sigma_{ppt} \end{pmatrix}. \tag{5.1}$$

What is substantially different from the univariate GARCH models is the constraints needed for the dynamic specification of Ω_t. That is, the specification of a multivariate GARCH (MGARCH) model differs from the univariate case ($p = 1$), where X_t conditional on $\mathcal{F}_{t-1}$ is $N(0, \sigma_{11t})$. In the univariate case, with $\sigma_{11t} = \sigma_t^2$, the only constraint for the dynamics of the scalar σ_t^2 sequence is that it should be positive for all t, $\sigma_t^2 > 0$. As is well-known, this allows a large number of specifications such as the classic GARCH(1,1) – the workhorse of ARCH models –

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where the parameter constraints $\omega > 0$ and $\alpha, \beta \geq 0$ indeed ensure $\sigma_t^2 > 0$.

The challenge in the MGARCH case is to formulate a dynamic specification of Ω_t which ensures that Ω_t is positive definite at each point in time. Obviously, at the same time

the dynamic specification should be empirically "relevant", that is, the model should capture the dynamics of the (portfolio) data well, and making it possible to interpret the obtained results in the context of risk analyses for example. At the same time the MGARCH model should also be possible to estimate, in the sense that there are not "too many parameters" to estimate, and not "too many" (and difficult to implement) constraints on the (matrix) parameters leading to highly non-linear likelihood functions to be optimized.

To address these issues, and thereafter present existing MGARCH models, we consider first different reparametrizations – as well as likelihood-based estimation – of a constant covariance matrix, $\Omega_t = \Omega$.

A firm understanding of the structure of the covariance matrix and the possible reparametrizations suitable for imposing positive definiteness, is a good starting point for understanding the parametrizations proposed for multivariate ARCH models.

6 Covariance Parameterization and Estimation

This section provides an introduction to estimation and reparametrization of Ω, where Ω is a $p \times p$ covariance matrix. As Ω is a symmetric positive definite matrix, a crucial issue is the discussion of whether different reparametrizations imply this. That is, discuss if positive definiteness can be imposed directly – and in different ways – in the estimation procedure, which in turn will be useful when turning to the formulation of MGARCH models. Stated differently, when turning to more sophisticated structures of a time varying (conditional) covariance matrix, such as in the conditional variance modelling in multivariate ARCH models and factor analysis etc. it is useful to initially discuss different ways of computing the maximum likelihood estimator (MLE) of Ω. We do so by considering well-known different ways of writing Ω in terms of freely varying parameters θ.

Before addressing reparametrizations, we consider initially the classic multivariate approach to likelihood-based way of estimating the covariance matrix in an i.i.d. setting.

6.1 Sample-Covariance MLE

Consider a sample $\{X_t\}_{t=1}^T$ where $X_t = (x_{1t}, ..., x_{pt})' \in \mathbb{R}^p$ is assumed to be $p-$dimensional and i.i.d. $\mathrm{N}_p(0, \Omega)$,

$$X_t = \varepsilon_t, \quad \text{for } t = 1, 2, ..., T, \tag{6.1}$$

with ε_t i.i.d. $\mathrm{N}_p(0, \Omega)$, where Ω is positive definite, $\Omega > 0$. Alternatively, using the later introduced definition of a square root of a matrix, $\Omega^{1/2}$, we may write the model as

$$X_t = \Omega^{\frac{1}{2}} z_t, \quad \text{with} \quad z_t \text{ i.i.d. } \mathrm{N}_p(0, I_p). \tag{6.2}$$

By definition positive definite means that Ω is symmetric, and $v'\Omega v > 0$ for any p-dimensional vector v, where $v'v \neq 0$. Equivalently, all linear combinations of X_t, (the portfolios) $v'X_t$, have a strictly positive variance.

Similar to (5.1), one may alternatively write Ω in terms of its individual entries parameters as $\Omega = (\sigma_{ij})_{i,j=1,\ldots,p}$, that is

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \\ \vdots & & & \vdots \\ \sigma_{p1} & & \cdots & \sigma_{pp} \end{pmatrix}. \tag{6.3}$$

In terms of parameters, Ω has $n = p(p+1)/2$ unique entries (due to symmetry). More specifically, Ω consists of p variances σ_{ii} (often, as for the univariate GARCH models, written as σ_i^2 to emphasize that $\sigma_i^2 = \sigma_{ii} > 0$) on the diagonal and $p(p-1)/2$ off-diagonal covariance elements σ_{ij}, $i \neq j$. And we may also write,

$$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & & \\ \vdots & & & \vdots \\ \sigma_{p1} & & \cdots & \sigma_p^2 \end{pmatrix}.$$

In terms of estimation, the log-likelihood function for the model given by (6.1) is given by,

$$\begin{aligned} L(\Omega) &= -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T} X_t'\Omega^{-1}X_t = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}\operatorname{tr}\left(\Omega^{-1}X_tX_t'\right) \\ &= -\frac{T}{2}\left(\log|\Omega| + \operatorname{tr}\left(\Omega^{-1}S_{xx}\right)\right) \end{aligned} \tag{6.4}$$

where $S_{xx} = T^{-1}\sum_{t=1}^{T} X_tX_t'$, and $|\cdot|$ denotes the determinant. Moreover, properties of the trace operator, such as $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for any matrices A and B have been used.

Maximization of $L(\Omega)$, subject to the matrix constraint that Ω is positive definite, $\Omega > 0$, has an explicit solution $\hat{\Omega}$ in terms of the well-known sample-covariance matrix of X_t. This is stated in the next lemma, which is proved in the Appendix.

Lemma 6.1. *With $L(\Omega)$ given by (6.4),*

$$\hat{\Omega} = \arg\max_{\Omega>0} L(\Omega) = S_{xx} \tag{6.5}$$

where $S_{xx} = T^{-1}\sum_{t=1}^{T} X_tX_t'$.

Remark 6.2. *Note that $S_{xx} \to_p \Omega > 0$.*

Remark 6.3 (Non-Zero Level). *If $\mathrm{E}(X_t) = \mu \neq 0$, it follows that the model can be written as*

$$X_t = \mu + \Omega^{\frac{1}{2}} z_t, \quad \text{with} \quad z_t \text{ i.i.d. } N_p(0, I_p), \tag{6.6}$$

and the covariance estimator is given by,

$$\hat{\Omega} = S_{xx\cdot c}, \quad \text{with} \quad S_{xx\cdot c} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \hat{\mu})(X_t - \hat{\mu})',$$

and $\hat{\mu} = T^{-1}\sum_{t=1}^{T} X_t$.

Remark 6.4 (BEKK). *In terms of MGARCH models one may view the direct parametrization here with $\Omega > 0$ (as a matrix restriction) as the origin of the explicit dynamic specification of BEKK-type MGARCH models, see the discussion of BEKK-MGARCH.*

Remark 6.5 (Covariance Targeting). *The sample covariance estimator is often applied in the context of so-called covariance targeting in multivariate dynamic MGARCH models. The idea of covariance targeting can be exemplified by considering the univariate ARCH model as given by*

$$X_t = \sigma_t z_t,$$

with z_t i.i.d.$N(0,1)$ and $\sigma_t^2 = \omega + \alpha x_{t-1}^2$. It follows that if $0 \leq \alpha < 1$, $\mathrm{E}\, X_t^2 < \infty$ and the unconditional variance of X_t is given by

$$\mathrm{V}(X_t) = \sigma^2 = \omega/(1-\alpha).$$

This can be used to re-write the ARCH model as,

$$X_t = \sigma_t(\theta) z_t, \;\; \sigma_t^2(\theta) = \sigma^2(1-\alpha) + \alpha x_{t-1}^2, \;\; \text{and } \theta = \left(\sigma^2, \alpha\right)'.$$

A two-step estimator of the ARCH model is then given by the following two steps. First set $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} x_t^2$, that is the sample variance. Next, optimize the log-likelihood function $L_{ARCH}(\theta)$ which is given by,

$$L_{ARCH}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\sigma_t^2(\alpha) - x_t^2/\sigma_t^2(\alpha)\right), \qquad \sigma_t^2(\alpha) = \sigma^2(1-\alpha) + \alpha x_{t-1}^2,$$

with $\sigma^2 = \hat{\sigma}^2$ fixed. That is, variance targeting estimator of α is given by

$$\hat{\alpha}_V = \arg\max_{0\leq\alpha<1} L_{ARCH}\left(\hat{\sigma}^2, \alpha\right).$$

Summarizing, the two-step estimator is based on: (i) *an initial estimation of the unconditional variance (by the sample variance) – the "variance has been targeted" – and,* (ii) *a second step where the nonlinear function $L_{ARCH}(\theta)$ is maximized with σ^2 fixed at $\hat{\sigma}_T^2$ and subject to $\alpha \geq 0$ (and $\alpha < 1$). By definition this differs from the MLE of (σ^2, α) which is based on maximizing $L_{ARCH}(\theta)$ for both ω (or σ^2) and α, with $\omega > 0$ and $\alpha \geq 0$.*

In the univariate case, the variance targeting approach of first obtaining the unconditional variance, and next maximizing the likelihood function with the unconditional variance fixed, "saves" one parameter for the second nonlinear optimization. While this may not seem of much importance in the univariate case, this is different for higher dimensional MGARCH models. There, even for moderate p, saving $n = p(p+1)/2$ parameters in an initial step by targeting the unconditional variance may be useful, see Pedersen and Rahbek (2014).

Next, we *reparametrize* the symmetric and positive matrix Ω in different ways. Each different parametrization may be viewed as representing the origin of one (or more) member(s) of the huge class of various advanced multivariate autoregressive conditional heteroskedastic models, the MGARCH models. Consider first the so-called vech reparametrization.

6.2 vech(Ω) | Parametrization

The first parametrization of Ω to be considered is based on using the so-called vech $(\cdot)$ operator. As shown vech (Ω) leads to a reparametrization in terms of θ, $\theta \in \mathbb{R}^n$ with $n = p(p+1)/2$ parameters.

The background for this are the two key operations on matrices: vec$(\cdot)$ and vech$(\cdot)$. Here vec$(\cdot)$ is the vector obtained by stacking the columns of a matrix. That is, with $\Omega = (\sigma_{ij})_{i,j=1,...p}$, vec (Ω) is the p^2 dimensional vector,

$$\text{vec}(\Omega) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{p1} & \cdots & \sigma_{1p} & \cdots & \sigma_{pp} \end{pmatrix}'.$$

As noted Ω is symmetric, and instead the vech $(\cdot)$ ("vec-half") operator stacks all non-identical elements of Ω by stacking all columns from the diagonal and below, that is, the vec of the lower triangular part of Ω. Specifically, vech (Ω) is the $n = p(p+1)/2$ dimensional vector given by,

$$\text{vech}(\Omega) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{p1} & \sigma_{22} & \cdots & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}'. \tag{6.7}$$

Various rules for calculus in terms of these exist and we shall state these as we need them in the following, see also Henderson and Searle (1979) and Magnus and Neudecker (1988). For the vech (Ω) one can define the parameters in θ, $\theta = (\theta_1, ..., \theta_n)'$, by simply setting θ_i equal to the ith element of vech (Ω),

$$(\theta_1, ..., \theta_n) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{p1} & \sigma_{22} & \cdots & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

Given Ω, then $\theta = \text{vech}(\Omega)$, while given θ, $\Omega = \Omega(\theta)$ is found by inverting the vech$(\cdot)$ operation (denoted by "unvech$(\cdot)$") and constructing $\Omega(\theta)$ as,

$$\Omega(\theta) = \text{unvech}(\theta) := \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \\ \theta_2 & \theta_{p+1} & & \\ \vdots & & & \vdots \\ \theta_p & & \cdots & \theta_n \end{pmatrix}.$$

Note that $\Omega(\theta)$ is symmetric for any choice of θ. However, it may not be positive definite unless the parameters θ are further restricted. And hence it is *not* a reparametrization unless the parameters θ_i are subject to the further constraint that $v'\Omega(\theta)v > 0$ for any $v \in \mathbb{R}^p, v \neq 0$, which is difficult to implement.

Therefore the vech parametrization does not seem to be a very useful parametrization for developing MGARCH model specifications.

As an illustration consider the bivariate case of the vech parametrization:

Example 6.1 (vech-parametrization): With $p = 2$, write Ω as

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \quad \sigma_{12} = \sigma_{21}. \tag{6.8}$$

The parameters to be estimated are given by

$$\theta = (\theta_1, \theta_2, \theta_3)' = \text{vech}(\Omega) = (\sigma_{11}, \sigma_{21}, \sigma_{22})'.$$

Now consider maximizing $L(\Omega)$ in terms of θ, $L_{\text{vech}}(\theta) = L(\Omega(\theta))$, using numerical optimization with $\theta \in \mathbb{R}^3$, that is, with no restrictions on the parameters in θ. Insertion gives,

$$\begin{aligned} -2L_{\text{vech}}(\theta)/T &= \log|\Omega(\theta)| + \text{tr}\left(\Omega(\theta)^{-1} S_{xx}\right) \\ &= \log\left(\theta_1\theta_3 - \theta_2^2\right) + \text{tr}\left(\frac{1}{\theta_1\theta_3 - \theta_2^2}\begin{pmatrix} \theta_3 & -\theta_2 \\ -\theta_2 & \theta_1 \end{pmatrix} S_{xx}\right). \end{aligned}$$

By maximizing $L_{\text{vech}}(\theta)$ (or minimizing $-2L_{\text{vech}}(\theta)/T$) over θ, with θ freely varying, the resulting estimator $\hat{\Omega}_{\text{vech}}$, say, while symmetric by definition, may not be positive definite as argued. This is because one ignores the constraint implied by $\Omega > 0$. ♦

Example 6.2 (Restrictions to Ensure Positive Definiteness): From the previous considerations it follows that $\hat{\Omega}_{\text{vech}} \neq \hat{\Omega}$, as the $\hat{\Omega}$ is the MLE which satisfies the positive definite constraints. In the simple bivariate case considered one may actually impose directly the further constraints for the vech parametrization to ensure positiveness. The constraints for $\Omega > 0$ for the bivariate case are simply,

$$\theta_1 > 0, \quad \theta_3 > 0, \quad \text{and} \quad \theta_1\theta_3 > \theta_2^2. \tag{6.9}$$

Optimization which takes these restrictions into account would indeed lead to $\hat{\Omega}_{\text{vech}} > 0$ and $\hat{\Omega}_{\text{vech}} = \hat{\Omega}$. In particular, the two different numerical computations *(i)* the sample-covariance, $\hat{\Omega}$ and, *(ii)* the maximization approach of $L_{\text{vech}}(\theta)$ subject to the further constraints $\Omega(\theta) > 0$, leads to the same estimator. Hence, with the further restrictions imposed it is indeed a reparametrization in the bivariate case. ♦

Example 6.3 (Positivity of a Scalar Parameter): The first two constraints in (6.9) are simple to implement (positivity of a parameter), while the last is not so easy (and becomes more complicated as the dimension p increases). Most likely, this is one possible explanation as to why the vech (re-)parametrization is not much used in practice.

Specifically, a simple way to handle the constraint of positivity is to write,

$$\theta_i = \exp(\gamma_i).$$

In this case, for all $\gamma_i \in \mathbb{R}$, $\theta_i > 0$, and the mapping from θ_i to γ_i is one-to-one ($\gamma_i = \log\theta_i$ and $\theta_i = \exp(\gamma_i)$). This is an example of a (further) reparametrization which removes a constraint since maximizing over $\theta_i > 0$ is equivalent to maximizing over $\gamma_i \in \mathbb{R}$. If $\theta_i \geq 0$, likewise one may parametrize as $\theta_i = \gamma_i^2$, $\gamma_i \in \mathbb{R}$, leading to $\theta_i \geq 0$ by definition. ♦

Remark 6.6 (Probability Theory and vech). *In terms of MGARCH models, (a recursive dynamic version of the)* vech *formulation is typically not applied empirically,*

but instead applied when considering probability theoretical arguments. That is, most – if not all – of the MGARCH specifications can be re-cast by using the vech *operator and some general theory for the stochastic behavior can be derived using this, see Section 8.*

6.3 Correlation Reparametrization

The idea upon which the so-called constant conditional correlation MGARCH, or the CCC-MGARCH model is build on is given by factorizing the covariance matrix Ω in terms of variances $(\sigma_i^2)_{i=1}^p$ and correlations $(\rho_{ij})_{i\neq j}$, where

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

More precisely, with $D = \operatorname{diag}\left((\sigma_i)_{i=1}^p\right)$ the $(p \times p)$-dimensional diagonal matrix with standard deviations on the diagonal, and Γ the correlation matrix,

$$\Gamma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & & \vdots \\ \vdots & & \ddots & \rho_{(p-1)p} \\ \rho_{1p} & \cdots & \rho_{(p-1)p} & 1 \end{pmatrix}$$

it follows that Ω can be factorized as,

$$\Omega = D\Gamma D. \tag{6.10}$$

Thus in terms of the parameters in

$$\theta = \left(\left(\sigma_i^2\right)_{i=1}^p, \left(\rho_{ij}\right)_{i=1,\ldots,p,j>i}\right).$$

the correlation likelihood function $L_{\text{Corr}}(\theta)$ is given by $L_{\text{Corr}}(\theta) = L(\Omega(\theta))$ with $L(\Omega)$ defined in (6.4) and

$$\Omega(\theta) = D(\theta)\,\Gamma(\theta)\,D(\theta).$$

Direct insertion gives,

$$\begin{aligned} -\frac{2}{T}L_{\text{Corr}}(\theta) &= \log|D(\theta)\,\Gamma(\theta)\,D(\theta)| + \operatorname{tr}\left(\left(D(\theta)\,\Gamma(\theta)\,D(\theta)\right)^{-1} S_{xx}\right) \\ &= \log|\Gamma(\theta)| + \sum_{i=1}^{p}\log\theta_i^2 + \operatorname{tr}\left(\Gamma(\theta)^{-1} D(\theta)^{-1} S_{xx} D(\theta)^{-1}\right) \end{aligned} \tag{6.11}$$

By definition, the MLE $\hat{\theta}_{\text{Corr}} = \operatorname{argmax} L_{\text{Corr}}(\theta)$, where maximization is subject to the constraints that $\theta_i = \sigma_i^2 > 0$ and that $\Gamma(\theta)$ is a correlation matrix (in terms of ρ_{ij}). Also by definition, $\hat{\Omega}_{\text{Corr}} = \Omega(\hat{\theta}_{\text{Corr}})$ is identical to $\hat{\Omega}$.

The complication is the maximization of $L_{\text{Corr}}(\theta)$ over θ, imposing $\Gamma(\theta)$ to be a correlation matrix. An immediate necessary restriction is that $|\rho_{ij}| < 1$, which can be implemented by using for example a further reparametrization, $\theta_{ij} = \log\left(\frac{1+\rho_{ij}}{1-\rho_{ij}}\right)$, such

that θ_{ij} takes values in all of $\mathbb{R}$, $\theta_{ij} \in \mathbb{R}$. And ρ_{ij} can be found by the inverse mapping given by,

$$\rho_{ij} = \frac{\exp(\theta_{ij}) - 1}{\exp(\theta_{ij}) + 1}.$$

Importantly, however, while this is sufficient and necessary for $\Gamma(\theta)$ to be a correlation matrix in the bivariate case of $p = 2$, and hence $\rho_{12} = \rho$ to be *the* correlation, it is *not sufficient* for the case of $p > 2$. That is, for the case of $p > 2$ maximization of $L_{\text{Corr}}(\theta)$ by numerical optimization is non-trivial as by definition the correlation matrix $\Gamma(\theta)$ is a positive definite matrix with unit entries on the diagonal (1/2 vech reparametrization).

A general and explicit solution is to factor the correlation matrix $\Gamma(\theta)$ as

$$\Gamma(\theta) = G(\theta) G(\theta)',$$

where $G(\theta)$ is defined in terms of hyper-spherical parameters, that is in terms of trigonometric functions $\cos(\cdot)$ and $\sin(\cdot)$. For example, for $p = 2$ and with $\rho_{12} = \cos(\theta_{12})$, $G(\theta)$ can be defined as,

$$G(\theta) = \begin{pmatrix} 1 & 0 \\ \cos(\theta_{12}) & \sin(\theta_{12}) \end{pmatrix}. \tag{6.12}$$

It follows directly that in this case,

$$G(\theta) G(\theta)' = \begin{pmatrix} 1 & 0 \\ \cos(\theta_{12}) & \sin(\theta_{12}) \end{pmatrix} \begin{pmatrix} 1 & \cos(\theta_{12}) \\ 0 & \sin(\theta_{12}) \end{pmatrix} = \begin{pmatrix} 1 & \cos(\theta_{12}) \\ \cos(\theta_{12}) & 1 \end{pmatrix}.$$

and hence as $-1 < \cos(\theta_{12}) < 1$ by definition for any θ_{12}, $G(\theta) G(\theta)'$ is indeed a correlation matrix. In particular, observe that for the determinant of $\Gamma(\theta)$ it holds that,

$$|\Gamma(\theta)| = |G(\theta) G(\theta)'| = 1 - \cos^2(\theta_{12}) = \sin^2(\theta_{12}) > 0,$$

if (for example) $\theta_{12} \in (0, \pi/2)$.

For dimensions $p > 2$, $\Gamma(\theta) = G(\theta) G(\theta)'$ can likewise be parametrized in terms (of products) of cos and sin functions, see Creal, Koopman, and Lucas (2011, eq.(21)) for an explicit expression. See also Example 6.6 and Appendix A.2 for more details. Citation

Remark 6.7 (Two-Step Estimation). *An alternative to numerical optimization of $L_{\text{Corr}}(\theta)$ is a two-step estimation method, as sometimes applied in the context of CCC and DCC MGARCH estimation. The two-step estimation approach can be presented as follows: For each variable X_{it} in X_t, define $\hat{\theta}_{TS} = \left((\hat{\sigma}_i^2)_{i=1}^p, (\hat{\rho}_{ij})_{i=1,\ldots,p, j>i} \right)$ as follows. Set*

$$\hat{\sigma}_i^2 = S_{x_i x_i} = T^{-1} \sum_{t=1}^{T} X_{it}^2, \ \ i = 1, \ldots, p$$

That is, $\hat{\sigma}_i^2$ is set equal to the sample variance of the variable X_{it}. With $\hat{D}_{TS} = \text{diag}(\hat{\sigma}_i)$, define next

$$\hat{X}_t = \hat{D}_{TS}^{-1} X_t,$$

such that $\hat{X}_{it} = X_{it}/\hat{\theta}_i$. *Next, consider maximization of* $L_{\text{Corr}}(\theta)$ *which, with* $\sigma_i^2 = \hat{\sigma}_i^2$ *fixed, reduces to minimization of "the concentrated" likelihood* $L_{TS}(\rho_{ij})$ *given by,*

$$L_{TS}(\rho_{ij}) = \log|\Gamma(\rho_{ij})| + \text{tr}\left(\Gamma(\rho_{ij})^{-1} S_{\hat{x}\hat{x}}\right).$$

Here $S_{\hat{x}\hat{x}} = T^{-1}\sum_{t=1}^{T} \hat{X}_t \hat{X}_t'$ *and since the correlation matrix is positive definite, then as in Lemma 6.1,*

$$\hat{\Gamma}_{TS} = \Gamma(\hat{\rho}_{ij}) = S_{\hat{x}\hat{x}} = \hat{D}^{-1} S_{xx} \hat{D}^{-1}.$$

Collecting terms,

$$\hat{\Omega}_{TS} = \hat{D}_{TS} \hat{\Gamma}_{TS} \hat{D}_{TS}.$$

Note that for general $p \geq 2$ *the two-step (TS) method is equivalent to:* (i) *First estimate the covariance matrix* Ω *by the sample covariance matrix,* $\hat{\Omega} = S_{xx}$, *and next* (ii), *use the definition of correlations to compute each sample correlation entry in* $\hat{\Gamma}$. *In other words,* $\hat{\Omega}_{TS}$ *is identical to* $\hat{\Omega}$, *the MLE.*

Example 6.4 (Two-Step Estimation, p=2): Consider the bivariate case where $p = 2$. It follows directly that with $S_{x_1x_2} = T^{-1}\sum_{t=1}^{T} X_{1t}X_{2t}$, the TS estimator is given by,

$$\begin{aligned}\hat{\Gamma}_{\text{TS}} = \hat{D}_{\text{TS}}^{-1} S_{xx} \hat{D}_{\text{TS}}^{-1} &= \begin{pmatrix} 1/\hat{\sigma}_1 & 0 \\ 0 & 1/\hat{\sigma}_2 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_1^2 & S_{x_1x_2} \\ S_{x_1x_2} & \hat{\sigma}_2^2 \end{pmatrix} \begin{pmatrix} 1/\hat{\sigma}_1 & 0 \\ 0 & 1/\hat{\sigma}_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & S_{x_1x_2}/\hat{\sigma}_1\hat{\sigma}_2 \\ S_{x_1x_2}/\hat{\sigma}_1\hat{\sigma}_2 & 1 \end{pmatrix}.\end{aligned}$$

Hence indeed this leads to the sample correlation matrix as $\hat{\Gamma}_{12} = \hat{\rho}_{12}$ is given by

$$S_{x_1x_2}/\hat{\sigma}_1\hat{\sigma}_2 = \sum_{t=1}^{T} X_{1t}X_{2t} \Big/ \left(\sum_{t=1}^{T} X_{1t}^2 \sum_{t=1}^{T} X_{2t}^2\right)^{1/2},$$

and

$$\hat{\Omega}_{\text{TS}} = \hat{D}_{\text{TS}} \hat{\Gamma}_{\text{TS}} \hat{D}_{\text{TS}} = S_{xx} = \hat{\Omega}.$$

♦

6.4 Eigenvalue Reparametrizations

MGARCH specifications may also be based on the idea of a decomposition of Ω in terms of its eigenvalues, the so-called *spectral decomposition.* Examples include the λ-MGARCH, or eigenvalue-MGARCH in Hetland, Pedersen, and Rahbek (2019), as well as the GO-GARCH in Boswijk and van der Weide (2011) and variants thereof.

Recall that the eigenvalue problem for Ω, $\Omega > 0$, is given by solving,

$$|\lambda I_p - \Omega| = 0,$$

for eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p > 0$. The eigenvectors $(v_i)_{i=1}^p$ corresponding to (each of) the eigenvalues $(\lambda_i)_{i=1}^p$, satisfy by definition $v_i \in \mathbb{R}^p$, $v_i' v_j = \mathbb{I}(i = j)$ and

$$\Omega v_i = \lambda_i v_i.$$

Collecting the orthogonal eigenvectors in $V = (v_1, v_2, ..., v_p)$, and similarly the eigenvalues in

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_p) = \text{diag}\left((\lambda_i)_{i=1}^p\right),$$

it follows that as $v_i' v_j = 0$ for $i \neq j$,

$$\Omega = V \Lambda V', \quad VV' = V'V = I_p. \tag{6.13}$$

Using this, the eigenvectors in V can be used to diagonalize Ω,

$$V' \Omega V = \Lambda. \tag{6.14}$$

The choice of eigenvectors is unique as they satisfy $V'V = VV' = I$; that is, the eigenvectors are chosen as orthogonal vectors with length one (referred to as orthonormal).

6.5 Reparametrization | Eigenvectors and Eigenvalues

Using the eigenvalue problem one may reparametrize Ω in terms of eigenvalues and eigenvectors. Consider initially the case of $p = 2$:

Example 6.5 (Eigenvectors and Eigenvalues, p=2): Consider as an example the case of $p = 2$. With $\theta = (\theta_1, \theta_2, \theta_3)'$ one can set

$$\Lambda(\theta) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}, \text{ with } \theta_1 = \lambda_1 > 0 \text{ and } \theta_2 = \lambda_2 > 0.$$

Moreover, one can explicitly parametrize the eigenvector matrix, $V(\theta)$, as a rotation matrix in terms of $\cos(\cdot)$ and $\sin(\cdot)$ as follows:

$$V(\theta) = (v_1(\theta), v_2(\theta)) = \begin{pmatrix} \cos(\theta_3) & -\sin(\theta_3) \\ \sin(\theta_3) & \cos(\theta_3) \end{pmatrix}, \theta_3 \in \mathbb{R}.$$

To see this, note that by definition $V(\theta)' V(\theta) = V(\theta) V(\theta)' = I_2$ for any $\theta_3 \in \mathbb{R}$.

With $L(\cdot)$ given by (6.4), the likelihood function L_{EE} to be optimized over θ is given by,

$$\begin{aligned} L_{\text{EE}}(\theta) &= L\left(V(\theta) \Lambda(\theta) V(\theta)'\right) \\ &= -\frac{T}{2} \log\left|V(\theta) \Lambda(\theta) V(\theta)'\right| - \frac{1}{2} \sum_{t=1}^{T} \text{tr}\left(\Lambda(\theta)^{-1} V(\theta)' X_t X_t' V(\theta)\right) \\ &= -\frac{T}{2} \left(\log|\Lambda(\theta)| + \text{tr}\left(\Lambda(\theta)^{-1} V(\theta)' S_{xx} V(\theta)\right)\right) \end{aligned}$$

$$= -\frac{T}{2}\sum_{i=1}^{2}\left(\log|\theta_i| + S_{\tilde{x}_i\tilde{x}_i}/\theta_i\right)$$

where $S_{\tilde{x}_i\tilde{x}_i} = T^{-1}\sum_{t=1}^{T}\tilde{X}_{it}^2$, $\tilde{X}_{1t} = \cos(\theta_3)X_{1t} + \sin(\theta_3)X_{2t}$ and $\tilde{X}_{2t} = \cos(\theta_3)X_{1t} + \sin(\theta_3)X_{2t}$. In other words the multiplication of X_t by $V(\theta)$,

$$V(\theta)'X_t = \begin{pmatrix}\tilde{X}_{1t}\\ \tilde{X}_{2t}\end{pmatrix},$$

rotates the X_t such that the rotated $V(\theta)'X_t = (v_1'(\theta)X_t, v_2'(\theta)X_t)'$ consists of two independent Gaussian variables, each with variance θ_i. ♦

For dimensions $p > 2$ one may likewise parametrize in terms of *(i)* the eigenvalues $(\theta_i)_{i=1}^p$, and *(ii)* the (rotation) matrices $V(\theta)$, or the eigenvectors $v_i(\theta)$. This is 1/2 correlation matrix parametrization and explicit formulas for $V(\theta)$ for general $p > 2$ are discussed in Hetland, Pedersen, and Rahbek (2019), and given in Appendix A.2.

Using the identity

$$|\Omega(\theta)| = \left|V(\theta)\Lambda(\theta)V(\theta)'\right| = |\Lambda(\theta)| = \prod_{i=1}^{p}\theta_i,$$

for general p, the MLE $\hat{\theta}_{\text{EE}}$ of θ is by definition given by,

$$\hat{\theta}_{\text{EE}} = \arg\max L_{\text{EE}}(\theta),\ L_{\text{EE}}(\theta) = -\frac{T}{2}\sum_{i=1}^{p}\left(\log|\theta_i| + S_{\tilde{x}_i\tilde{x}_i}/\theta_1\right),$$

where $\tilde{X}_t = V(\theta)'X_t$ and $S_{\tilde{x}_i\tilde{x}_i} = T^{-1}\sum_{t=1}^{T}\tilde{X}_{it}^2$. The constraints imposed are $\theta_i > 0$ for $i = 1, ..., p$, while for the remaining parameters $(\theta_j)_{j=p+1,...,p(p+1)/2}$ – if they as in the bivariate case are parameters, or arguments, in the trigonometric functions $\cos(\cdot)$ and $\sin(\cdot)$ then – a simple requirement is to let $\theta_j \in (0, \pi/2)$, see Hetland, Pedersen, and Rahbek (2019) for further details.

Note also that as the $\Omega(\theta)$ is by definition positive definite,

$$\hat{\Omega}_{\text{EE}} = \Omega(\hat{\theta}_{\text{EE}}) = \hat{\Omega}.$$

6.6 Matrix Functions

A further feature of the eigenvalue problem decomposition of Ω in (6.14) is that it makes it possible to define various functions of matrices which are generalizations of well-known functions of scalars. This is because any *scalar* function can be applied to the *scalar* diagonal elements of Λ.

First, by definition, the matrix *inverse* of Ω can be defined by,

$$\Omega^{-1} = V\Lambda^{-1}V', \quad \text{with} \quad \Lambda^{-1} = \text{diag}\left(\left(\lambda_i^{-1}\right)_{i=1}^{p}\right).$$

Likewise, the *square root* of a matrix, $\Omega^{1/2}$, can be defined by,

$$\Omega^{1/2} = V\Lambda^{1/2}V', \quad \text{with} \quad \Lambda^{1/2} := \operatorname{diag}\left(\left(\lambda_i^{1/2}\right)_{i=1}^{p}\right). \tag{6.15}$$

Note that it is straightforward to check that indeed,

$$\Omega^{1/2}\Omega^{1/2} = V\Lambda^{1/2}V'V\Lambda^{1/2}V' = V\Lambda^{1/2}\Lambda^{1/2}V' = V\Lambda V' = \Omega.$$

In the same manner, we may define the inverse matrix square-root

$$\Omega^{-1/2} = V\Lambda^{-1/2}V', \quad \text{with} \quad \Lambda^{-1/2} := \operatorname{diag}\left(\left(\lambda_i^{-1/2}\right)_{i=1}^{p}\right). \tag{6.16}$$

Two other key examples of matrix functions are the (matrix) $\log(\cdot)$ and $\exp(\cdot)$ functions of the matrix Ω, which similarly are defined by

$$\log(\Omega) = V\log(\Lambda)V', \quad \text{with} \quad \log(\Lambda) = \operatorname{diag}\left(\log(\lambda_i)_{i=1}^{p}\right),$$

and

$$\exp(\Omega) = V\exp(\Lambda)V', \quad \text{with} \quad \exp(\Lambda) = \operatorname{diag}\left(\exp(\lambda_i)_{i=1}^{p}\right).$$

Note that as $\lambda_i > 0$, $\log(\lambda_i) \in \mathbb{R}$, with $\exp(\log(\lambda_i)) = \lambda_i > 0$. That is, by definition

$$\log(\Omega) = V\log(\Lambda)V'$$

is a symmetric matrix, which is not necessarily positive definite (and need not be $1/2$ $\log(x)$ function in the univariate case). Likewise,

$$\Omega = \exp(\log(\Omega)) \tag{6.17}$$

is positive definite as $\exp(\cdot) > 0$. The identity in (6.17), follows by

$$\exp(\log(\Omega)) = \exp(V\log(\Lambda)V') = V\exp(\log(\Lambda))V' = V\Lambda V' = \Omega,$$

which is a generalized version of the classic scalar identity $\exp(\log\sigma^2) = \sigma^2$.

6.7 Reparametrization | $\log(\Omega)$

Using the $\log(\Omega)$ function, one may reparametrize as follows,

$$\log(\Omega(\theta)) = V(\theta)\log(\Lambda(\theta))V(\theta)',$$

where $V(\theta)$ are as discussed above for the eigenvalues and eigenvectors reparametrization, while

$$\log(\Lambda(\theta)) = \operatorname{diag}\left((\theta_i)_{i=1}^{p}\right),$$

with $\theta_i \in \mathbb{R}$. That is, as the $\log(\cdot)$ is not necessarily positive (definite), the θ_i can take any values.

The likelihood function $L_{\log}(\theta)$ is given by $L(\Omega(\theta))$ with $L(\cdot)$ defined in (6.4), and

$$\log \Omega(\theta) = V(\theta) \log(\Lambda(\theta)) V(\theta)'.$$

Insertion gives immediately,

$$L_{\log}(\theta) = L(\Omega(\theta)) = -\frac{T}{2}\left(\log|\Omega(\theta)| + \operatorname{tr}\left(\Omega(\theta)^{-1} S_{xx}\right)\right).$$

Next, by $V(\theta)' V(\theta) = I_p$ and the definition of matrix $\log(\cdot)$ and $\exp(\cdot)$, observe that

$$\begin{aligned}\Omega(\theta) &= \exp\left(V(\theta) \log(\Lambda(\theta)) V(\theta)'\right) \\ &= V(\theta) \exp(\log(\Lambda(\theta))) V(\theta)' \\ &= V(\theta) \operatorname{diag}\left(\exp(\theta_i)_{i=1}^p\right) V(\theta)'.\end{aligned}$$

Thus the first term in $L_{\log}(\theta)$ simplifies using this identity (as well as $V(\theta)' V(\theta) = I_p$),

$$\begin{aligned}\log|\Omega(\theta)| &= \log|V(\theta) \operatorname{diag}\left((\exp(\theta_i))_{i=1}^p\right) V(\theta)'| \\ &= \log|\operatorname{diag}\left((\exp(\theta_i))_{i=1}^p\right)| \\ &= \log \prod_{i=1}^p \exp(\theta_i) = \sum_{i=1}^p \theta_i.\end{aligned}$$

Similarly, using the definition of the matrix inverse, and setting $\tilde{X}_t = V(\theta)' X_t$,

$$\begin{aligned}\operatorname{tr}\left(\Omega(\theta)^{-1} S_{xx}\right) &= \operatorname{tr}\left(\operatorname{diag}\left((\exp(\theta_i))_{i=1}^p\right)^{-1} V(\theta)' S_{xx} V(\theta)\right) \\ &= \sum_{i=1}^p S_{\tilde{x}_i \tilde{x}_i} / \exp(\theta_i).\end{aligned}$$

Collecting terms one gets,

$$L_{\log}(\theta) = -\frac{T}{2} \sum_{i=1}^p \left(\theta_i + S_{\tilde{x}_i \tilde{x}_i} / \exp(\theta_i)\right).$$

By definition,

$$\hat{\theta}_{\log} = \arg\max L_{\log}(\theta),$$

with $\theta_1, ..., \theta_p > 0$ and for $(\theta_j)_{j=p+1}^n$, one may use $\theta_j \in (0, \pi/2)$. Also we have,

$$\hat{\Omega}_{\log} = \Omega(\hat{\theta}_{\log}) = \hat{\Omega}.$$

Remark 6.8 (**Rotation**). *As for the eigenvector and eigenvalue reparametrization, a key role here is the rotation* $\tilde{X}_t = V(\theta)' X_t$ *as the likelihood function simplifies into* p *terms, each corresponding to a classic univariate i.i.d. model for* $\tilde{X}_{it}$ *where* $\log \sigma_i^2 = \theta_i$, *or* $\sigma_i^2 = \exp(\theta_i)$.

Remark 6.9 (**Matrix Functions in Programming Languages**). *In most programming languages the* $\exp(\cdot)$ *and* $\log(\cdot)$ *functions of a matrix are not defined as above, but rather as an element-wise operation. To circumvent this the matrix functions defined here have to be implemented using an eigenvalue-problem algorithm (which are in all programming languages), and define matrix functions as above.*

6.8 Comparing Covariance Parametrizations

To emphasize in which sense the (re-)parametrizations so far produces identical results for the MLE, $\hat{\Omega}$ of Ω, we consider a small example for each parametrization.

Example 6.6 (Comparing Covariance Parametrizations): Consider a time series $\{X_t\}_{t=1}^T$ with $T = 500$ observations generated from a known DGP with X_t i.i.d. $\mathrm{N}_p(\mu, \Omega)$ with parameters given by

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 1 & 0.8 & 0.9 \\ 0.8 & 2 & 1.2 \\ 0.9 & 1.2 & 3 \end{pmatrix}. \tag{6.18}$$

Also consider the statistical model with a constant term, as in (6.6),

$$X_t = \mu + \Omega^{1/2} z_t, \quad \text{with} \quad z_t \text{ i.i.d. } \mathrm{N}_p(0, \Omega),\ \Omega > 0, \tag{6.19}$$

with likelihood function given by equation (6.4), i.e.

$$L(\Omega, \mu) = -\frac{T}{2} \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} (X_t - \mu)\Omega^{-1}(X_t - \mu). \tag{6.20}$$

With $\Omega(\theta)$ and $\mu(\theta)$ specified as functions of the $n = 3(3+1)/2 + 3 = 9$ parameters in $\theta = \{\theta_1, ..., \theta_9\}$, consider next the different parametrizations.

Sample Covariance: The closed form estimators are given by

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} X_t \quad \text{and} \quad \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} (X_t - \hat{\mu})(X_t - \hat{\mu})',$$

see Lemma 6.1, produce the results

$$\hat{\mu} = \begin{pmatrix} -0.029379 \\ -0.072781 \\ -0.15085 \end{pmatrix} \quad \text{and} \quad \hat{\Omega} = \begin{pmatrix} 0.98337 & 0.84912 & 0.79463 \\ 0.84912 & 2.0755 & 1.0780 \\ 0.79463 & 1.0780 & 3.0904 \end{pmatrix}.$$

These are quite close to the true values in (6.18). All parametrizations below give the same estimate for the mean, $\hat{\mu} = (\hat{\theta}_7, \hat{\theta}_8, \hat{\theta}_9)'$, and to save space, we focus on the variance decompositions for the remaining parametrizations of Ω as a function of $\theta_\Omega = (\theta_1, ..., \theta_6)'$.

VECH Parametrization: For the direct vech-parametrization, we maximize the likelihood in (6.20) numerically and obtain $\hat{\theta}_{\text{vech}} = (\hat{\theta}_1, ..., \hat{\theta}_6, ..., \hat{\theta}_9)$ such that

$$\hat{\Omega}_{\text{vech}} = \text{unvech}(\hat{\theta}_{\text{vech}}) = \begin{pmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \hat{\theta}_3 \\ \hat{\theta}_2 & \hat{\theta}_4 & \hat{\theta}_5 \\ \hat{\theta}_3 & \hat{\theta}_5 & \hat{\theta}_6 \end{pmatrix} = \begin{pmatrix} 0.98337 & 0.84912 & 0.79463 \\ 0.84912 & 2.0755 & 1.0780 \\ 0.79463 & 1.0780 & 3.0904 \end{pmatrix}.$$

The vech-parametrization gives a symmetric matrix by construction, and in this case, with $X_t \in \mathbb{R}^3$, the unrestricted estimates result in a positive definite matrix. Recall, however, that for larger dimensions, the unrestricted numerical optimization may fail to produce a positive definite estimate.

Cholesky Parametrization: For the Cholesky parametrization in Appendix A.1, $\Omega = LL'$, with L a lower triangular matrix. Numerical optimization of the likelihood function in (6.20) or (A.2) produces the lower triangular matrix

$$\hat{L} = L(\hat{\theta}_{\text{chol}}) = \begin{pmatrix} \hat{\theta}_1 & 0 & 0 \\ \hat{\theta}_2 & \hat{\theta}_4 & 0 \\ \hat{\theta}_3 & \hat{\theta}_5 & \hat{\theta}_6 \end{pmatrix} = \begin{pmatrix} 0.99165 & 0 & 0 \\ 0.85627 & 1.1586 & 0 \\ 0.80132 & 0.33819 & 1.5277 \end{pmatrix},$$

and it follows that

$$\hat{\Omega}_{\text{chol}} = \Omega(\hat{\theta}_{\text{chol}}) = \hat{L}\hat{L}' = \begin{pmatrix} 0.98337 & 0.84912 & 0.79463 \\ 0.84912 & 2.0755 & 1.0780 \\ 0.79463 & 1.0780 & 3.0904 \end{pmatrix},$$

which is positive definite by construction. Observe that there is no direct interpretation of entries in L in terms of the covariances in Ω, apart from the first entry,

$$\hat{L}_{11} = 0.99165 = 0.98337^{1/2} = \hat{\Omega}_{11}^{1/2},$$

which is the standard deviation of X_{1t}.

Correlation Parametrization: For the parametrization in terms of standard deviations and correlations,

$$\Omega = D\Gamma D,$$

numerical optimization of the likelihood function in (6.20) and (6.11) yields estimates of the standard deviations given by

$$\hat{D} = \begin{pmatrix} \hat{\sigma}_1 & 0 & 0 \\ 0 & \hat{\sigma}_2 & 0 \\ 0 & 0 & \hat{\sigma}_3 \end{pmatrix} = \begin{pmatrix} e^{\hat{\theta}_1} & 0 & 0 \\ 0 & e^{\hat{\theta}_2} & 0 \\ 0 & 0 & e^{\hat{\theta}_3} \end{pmatrix} = \begin{pmatrix} 0.99165 & 0 & 0 \\ 0 & 1.4407 & 0 \\ 0 & 0 & 1.7580 \end{pmatrix},$$

where the exponential function is used to impose positivity of $\hat{\sigma}_i$, $i = 1, 2, 3$. For the correlations,

$$\Gamma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix},$$

we use the parametrization in terms of hyper-spherical coordinates in (6.12), which generalizes to $p = 3$ as

$$G(\theta_{\text{Corr}}) = \begin{pmatrix} 1 & 0 & 0 \\ \cos\theta_4 & \sin\theta_4 & 0 \\ \cos\theta_5 & \cos\theta_6 \sin\theta_5 & \sin\theta_6 \sin\theta_5 \end{pmatrix},$$

such that, with $\Gamma(\theta_{\text{Corr}}) = G(\theta_{\text{Corr}})G(\theta_{\text{Corr}})'$,

$$\Gamma(\theta_{\text{Corr}}) = \begin{pmatrix} 1 & \cos\theta_4 & \cos\theta_5 \\ \cos\theta_4 & 1 & \cos\theta_4\cos\theta_5 + \cos\theta_6\sin\theta_4\sin\theta_5 \\ \cos\theta_5 & \cos\theta_4\cos\theta_5 + \cos\theta_6\sin\theta_4\sin\theta_5 & 1 \end{pmatrix}. \tag{6.21}$$

The ML estimates are given by $(\hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6)' = (0.93432, 1.0975, 1.3529)'$, such that

$$\hat{\Gamma} = \Gamma(\hat{\theta}_{\text{Corr}}) = \begin{pmatrix} 1 & \hat{\rho}_{12} & \hat{\rho}_{13} \\ \hat{\rho}_{12} & 1 & \hat{\rho}_{23} \\ \hat{\rho}_{13} & \hat{\rho}_{23} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.59436 & 0.45582 \\ 0.59436 & 1 & 0.42563 \\ 0.45582 & 0.42563 & 1 \end{pmatrix}.$$

Finally,

$$\hat{\Omega}_{\text{Corr}} = \Omega(\hat{\theta}_{\text{Corr}}) = \hat{D}\hat{\Gamma}\hat{D} = \begin{pmatrix} 0.98337 & 0.84912 & 0.79463 \\ 0.84912 & 2.0755 & 1.0780 \\ 0.79463 & 1.0780 & 3.0904 \end{pmatrix}.$$

Eigenvalue and Eigenvector Decomposition: For the spectral decomposition using eigenvalues and eigenvectors, $\Omega = V\Lambda V'$. As above, we use the exponential function to restrict eigenvalues to be positive, and with $\hat{\Lambda} = \Lambda(\hat{\theta}_{\text{EE}})$, we find

$$\hat{\Lambda} = \begin{pmatrix} e^{\hat{\theta}_1} & 0 & 0 \\ 0 & e^{\hat{\theta}_2} & 0 \\ 0 & 0 & e^{\hat{\theta}_3} \end{pmatrix} = \begin{pmatrix} e^{1.429} & 0 & 0 \\ 0 & e^{0.3873} & 0 \\ 0 & 0 & e^{-0.6914} \end{pmatrix} = \begin{pmatrix} 4.175 & 0 & 0 \\ 0 & 1.473 & 0 \\ 0 & 0 & 0.5009 \end{pmatrix}.$$

The orthonormal eigenvectors are parametrized in terms of rotation angles, which in the case $p = 3$ is given by,

$$V = \begin{pmatrix} \cos\theta_4 & -\sin\theta_4 & 0 \\ \sin\theta_4 & \cos\theta_4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta_5 & 0 & -\sin\theta_5 \\ 0 & 1 & 0 \\ \sin\theta_5 & 0 & \cos\theta_5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_6 & -\sin\theta_6 \\ 0 & \sin\theta_6 & \cos\theta_6 \end{pmatrix}.$$

With $(\hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6)' = (1.0102, 0.88855, 1.7366)'$ we find

$$\hat{V} = V(\hat{\theta}_{\text{EE}}) = \begin{pmatrix} 0.33528 & -0.26725 & 0.90342 \\ 0.53401 & -0.73609 & -0.41593 \\ 0.77616 & 0.62189 & -0.10408 \end{pmatrix},$$

such that $\hat{V}\hat{V}' = I_3$, and

$$\hat{\Omega}_{\text{EE}} = \Omega(\hat{\theta}_{\text{EE}}) = \hat{V}\hat{\Lambda}\hat{V}' = \begin{pmatrix} 0.98337 & 0.84912 & 0.79463 \\ 0.84912 & 2.0755 & 1.0780 \\ 0.79463 & 1.0780 & 3.0904 \end{pmatrix}.$$

Observe that the rotation matrix $V(\theta)$ is a product of three rotation matrices, each depending on one parameter, which as mentioned in Section 6.5 generalizes the

bivariate case. The parameter $\hat{\theta}_4$ controls the rotation angle between eigenvectors one and two, $\hat{\theta}_5$ controls the rotation angle between eigenvectors one and three, and $\hat{\theta}_6$ controls the rotation angle between eigenvectors two and three. In the general case, $X_t \in \mathbb{R}^p$, the rotation matrix, V, is a product of $p(p-1)/2$ individual rotations. ♦

7 Multivariate GARCH Models | MGARCH

In the previous sections different parametrizations of Ω were discussed, *(i)* vech (Ω); *(ii)* Correlation decomposition, $\Omega = D\Gamma D$; *(ii)* Eigenvector and eigenvalue decomposition, $\Omega = V\Lambda V'$, and *(vi)* Matrix function decompositions, $f(\Omega) = Vf(\Lambda)V'$, in particular with $f(\cdot)$ the exp and log matrix functions.

As mentioned the vast class of existing MGARCH models can be presented based on these decompositions, or reparametrizations. The MGARCH models discussed here all have the form

$$X_t = \Omega_t^{1/2} z_t, \quad t = 1, ..., T$$

where the $z_t's$ are i.i.d. $N_p(0, I_p)$ and Ω_t is the conditional variance of X_t given the past $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = (X_t, X_{t-1,...})$.

Dynamic specifications of Ω_t are given in terms of parameters in θ and lagged (functions of) X_t, Ω_t and possibly explanatory variables. Similar to univariate GARCH models, the notation $\Omega_t = \Omega_t(\theta)$ emphasizes the dependence on θ.

This section present the MGARCH models and likelihood-based estimation of these, while stochastic properties of the MGARCH processes are discussed in Section 8, and asymptotic properties of the MLE in Section 9. Specifically, the following specifications are discussed: First the *(i)* vech-MGARCH and *(ii)* BEKK-MGARCH models are discussed. Next, the *(iii)* Constant Conditional Correlation (CCC) MGARCH model (and Dynamic CC, or DCC-GARCH); and, finally, *(iv)* the Dynamic Conditional Eigenvalue GARCH, λ-MGARCH model (as well as GO-GARCH). For an alternative survey of MGARCH models, see for example Bauwens, Laurent, and Rombouts (2006).

7.1 Vech-MGARCH

With vech (Ω) defined in (6.7) the vech-MGARCH is a simple vector generalization of the univariate GARCH model(s).

Specifically, consider initially the simplest vech analogy of the univariate ARCH(1), the vech-MARCH(1) model. With vech (Ω_t) by definition an n-dimensional vector, where $n = p(p+1)/2$, the vech-MARCH(1) model is given by a linear vector autoregressive scheme as given by,

$$\text{vech}(\Omega_t(\theta)) = C + A\,\text{vech}\left(X_{t-1}X_{t-1}'\right), \tag{7.1}$$

where A is $(n \times n)$ dimensional, while C is a n-dimensional vector. An immediate drawback of the vech-MARCH(1) model (and in general vech-MGARCH models) is that

$$\Omega_t(\theta) = \text{unvech}\left(C + A\,\text{vech}\left(X_{t-1}X_{t-1}'\right)\right)$$

may not be positive definite, unless further restrictions are made on C and A.

The vech-MGARCH(1,1) – generalizing the GARCH(1,1) model – is given by,

$$\text{vech}\left(\Omega_t(\theta)\right) = C + A\,\text{vech}\left(X_{t-1}X_{t-1}'\right) + B\,\text{vech}\left(\Omega_{t-1}(\theta)\right). \tag{7.2}$$

Corresponding to the discussion of the vech-reparametrization, some immediate drawbacks with the vech-GARCH model are that: *(i)* As for the vech-MGARCH(1) discussed above, $\Omega_t(\theta)$ may not be positive definite, and *(ii)* the number of parameters in θ, that is $\dim_\theta = n(2n+1)$, where $n = (p+1)\,p/2$, is a large number already for small p. For example $\dim_\theta = 210$ for $p = 4$, and $\dim_\theta = 78$ for $p = 3$. The latter has – as for many of the other MGARCH models – lead to propositions of restrictive versions of the vech-GARCH model, where for example A and B are assumed diagonal (which is denoted diagonal vech-GARCH, or, DVEC-GARCH).

In terms of estimation, the parameter θ is given by $\theta = (C, A, B)$, where C is n-dimensional and A, B are $(n \times n)$ dimensional matrices. The log-likelihood function $L_{\text{vech}}(\theta)$ is given by $L(\cdot)$ in (6.4) evaluated in $\Omega_t(\theta)$,

$$-\frac{2}{T}L_{\text{vech}}(\theta) = -\frac{2}{T}L\left(\Omega_t(\theta)\right) = T^{-1}\sum_{t=1}^{T}\left(\log|\Omega_t(\theta)| - \text{tr}\left(\Omega_t(\theta)^{-1}X_tX_t'\right)\right).$$

The MLE of θ is defined by,

$$\hat{\theta}_{\text{vech}} = \arg\max L_{\text{vech}}(\theta),$$

where maximization is over $\theta = (C, A, B)$ with A, B and C freely varying. And finally, $\hat{\Omega}_{\text{vech},t}$ is given recursively by

$$\hat{\Omega}_{\text{vech},t} = \Omega_t(\hat{\theta}_{\text{vech}}) = \text{unvech}\left(\hat{C} + \hat{A}\,\text{vech}\left(X_{t-1}X_{t-1}'\right) + \hat{B}\,\text{vech}(\Omega_{t-1}(\hat{\theta}_{\text{vech}}))\right).$$

Remark 7.1 (Curse of Dimensionality). *The vech-GARCH model is so general (or, has so many parameters in general) that combined with the positive definiteness issue, this has meant that it is not applied much in practice. However, in terms of probabilistic properties the vech-recursive scheme is sometimes used. Specifically, for $\theta = \theta_0$ (the true value) with $\Omega_t = \Omega_t(\theta_0)$ it follows that,*

$$\text{vech}\,\Omega_t = C_0 + A_0\,\text{vech}\left(\Omega_{t-1}^{1/2}z_{t-1}z_{t-1}\Omega_{t-1}^{1/2}\right) + B_0\,\text{vech}\,\Omega_{t-1}.$$

That is, Ω_t is a (nonlinear) function of Ω_{t-1} and the i.i.d. sequence z_t, $\Omega_t = f(\Omega_{t-1}, z_{t-1})$. Therefore various multivariate techniques (for stochastic recurrence equations, or SREs) similar to univariate applications of the drift criteria from Markov chain theory can be

applied to discuss ergodicity, stationarity and existence of moments for Ω_t – and hence of X_t, see Section 8. The properties of the vech-scheme are used for some of the alternative MGARCH processes which can be re-stated as a vech*-GARCH process (with various constraints, see Section 8 for details).*

Observe that the recursive definition in (7.2) is given in terms of cross-product terms in $X_{t-1}X_{t-1}'$, where

$$X_{t-1}X_{t-1}' = \begin{pmatrix} X_{1t-1}^2 & X_{1t-1}X_{2t-1} & \cdots & & X_{1t-1}X_{pt-1} \\ & \ddots & & & \\ & & & & \\ & & & X_{p-1,t-1}^2 & X_{p-1,t-1}X_{pt-1} \\ X_{pt-1}X_{1t-1} & X_{pt-1}X_{2t-1} & \cdots & & X_{pt-1}^2 \end{pmatrix}. \tag{7.3}$$

As for univariate GARCH models, other variants can be formulated. For example Ω_t may depend on some (non-linear) function of $X_{t-1}X_{t-1}'$ similar to asymmetric univariate GARCH models, by for example adding additional terms of the form,

$$A_+ X_{t-1}X_{t-1}'A_+' \mathbb{I}(X_{t-1} > 0).$$

Here the parameter A_+ is $(p \times p)$-dimensional, and $X_{t-1} > 0$ means $X_{it-1} > 0$ for all $i = 1, 2, ..., p$ such that $\mathbb{I}(X_{t-1} > 0) = 1$ provided all $X_{it-1} > 0$ and zero otherwise.

7.2 BEKK-MGARCH

The BEKK-MGARCH in its simplest form is given by $X_t = \Omega_t(\theta)^{1/2} z_t$, with

$$\Omega_t(\theta) = \Omega + AX_{t-1}X_{t-1}'A', \tag{7.4}$$

with parameter $\theta = (\Omega, A)$, where $\Omega > 0$ and A is a $(p \times p)$-dimensional matrix. With $X_{t-1}X_{t-1}'$ given in (7.3), it follows with $A = (a_{ij})_{i,j=1,...,p}$ that the term $AX_{t-1}X_{t-1}'A'$ is given by the $(p \times p)$ dimensional matrix,

$$\begin{pmatrix} \left(\sum_{j=1}^p a_{1j}X_{jt-1}\right)^2 & \cdots & \left(\sum_{j=1}^p a_{1j}X_{jt-1}\right)\left(\sum_{j=1}^p a_{pj}X_{jt-1}\right) \\ \vdots & \ddots & \vdots \\ \left(\sum_{j=1}^p a_{1j}X_{jt-1}\right)\left(\sum_{j=1}^p a_{pj}X_{jt-1}\right) & \cdots & \left(\sum_{j=1}^p a_{pj}X_{jt-1}\right)^2 \end{pmatrix}.$$

Thus the term $AX_{t-1}X_{t-1}'A'$ in $\Omega_t(\theta)$ loads all (cross-)products of

$$\sum_{j=1}^p a_{ij}X_{jt-1},\ i = 1, ..., p,$$

and in this sense the conditional variance is allowed to depend on (squared and cross product) linear combinations of X_{t-1}.

In its is simple form the BEKK is analogous to the univariate ARCH(1) model where $X_t = \left(\omega + \alpha X_{t-1}^2\right) z_t$. That is, A "corresponds" to $\sqrt{\alpha}$, and the model may therefore be denoted as BEKK-ARCH(1). The BEKK-ARCH(1), as well as the general BEKK-MGARCH discussed below, has the important property that $\Omega_t(\theta)$ is positive definite for any θ. This is stated in the next lemma.

Lemma 7.2. *With $\Omega_t(\theta)$ given by (7.4), $\Omega_t(\theta) > 0$ for any $\Omega > 0$ and A any $(p \times p)$ matrix.*

To see this recall that $\Omega_t(\theta) > 0$ is equivalent to $v'\Omega_t(\theta) v > 0$ for any $v \in \mathbb{R}^p$, $v \neq 0$. As $v'AX_{t-1}$ is a scalar, which may or may not equal zero, it follows that $v'\Omega_t(\theta) v > 0$ by the following inequalities,

$$v'\Omega_t(\theta) v = v'\Omega v + \left(v'AX_{t-1}\right)^2 \geq v'\Omega v > 0.$$

Here it has been used that by definition (the matrix) $\Omega > 0$, such that $v'\Omega v > 0$, and the scalar $v'AX_{t-1}$, satisfies $\left(v'AX_{t-1}\right)^2 \geq 0$.

The same argument can be applied to the general BEKK-MGARCH(1,1), or BEKK-GARCH(1,1,k) (which generalizes the univariate GARCH(1,1)) as given by

$$\Omega_t(\theta) = \Omega + \sum_{i=1}^{k} \left(A_i X_{t-1} X_{t-1}' A_i' + B_i \Omega_{t-1} B_i'\right). \tag{7.5}$$

One should note the $2k$ loading matrices $(A_i, B_i)_{i=1}^k$ each of dimension $(p \times p)$, which all load only the first lag of $X_t X_t'$ and Ω_t. That several A_i (and B_i) load the same lag is a "truly" multivariate feature of the BEKK-MGARCH model; in contrast, for the univariate case one cannot have

$$\sigma_t^2 = \sum_{i=1}^{k} \left(\alpha_i X_{t-1}^2 + \beta_i \sigma_{t-1}^2\right),$$

as for $k > 1$ the parameters $(\alpha_i, \beta_i)_{i=2}^k$ are not identified. Correspondingly, for the multivariate case, the choice of k is not trivial as it on the dimension p in the sense that $k \leq p$ has to chosen such that all parameters (in the) A_i and B_i (matrices) are identified. A detailed discussion of this can be found in Engle and Kroner (1995).

By definition, the BEKK-MGARCH likelihood function $L_{\text{BEKK}}(\theta)$ is given by $L_{\text{BEKK}}(\theta) = L\left(\Omega_t(\theta)\right)$ with $\Omega_t(\theta)$ given by (7.5) (or a restricted version, where A_i and B_i are restricted) while $L(\cdot)$ is defined in (6.4). It follows by insertion that,

$$-\frac{2}{T} L_{\text{BEKK}}(\theta) = T^{-1} \sum_{t=1}^{T} \left(\log\left|\Omega_t(\theta)\right| + \operatorname{tr}\left(\Omega_t^{-1}(\theta) X_t X_t'\right)\right), \tag{7.6}$$

and hence the MLE $\hat{\theta}_{\text{BEKK}}$ of θ is defined by,

$$\hat{\theta}_{\text{BEKK}} = \arg\max L_{\text{BEKK}}(\theta),$$

where maximization is over $\theta = \left(\Omega, (A_i, B_i)_{i=1}^k\right)$ with A_i, B_i and Ω freely varying, with $\Omega > 0$. And, using the notation introduced in the context of vech-MGARCH,

$$\hat{\Omega}_{\text{BEKK},t} = \Omega_t(\hat{\theta}_{\text{BEKK}}).$$

As for the vech-MGARCH, the BEKK-MGARCH in general has the drawback that too many parameters enter the model for large, or even moderate p. Most likely, therefore an applied – and much studied – model in the literature is the case with $k = 1$ such that,

$$\Omega_t(\theta) = \Omega + AX_{t-1}X_{t-1}'A' + B\Omega_{t-1}B',$$

where the number of parameters is $n + 2p^2 = p(p+1)/2 + 2p^2$. Thus provided A and B are not restricted further, this choice reduces the number of parameters to 42 for $p = 4$, and 24 for $p = 3$. And, correspondingly typical applications of the unrestricted BEKK-GARCH(1,1) model are for small to moderate p, $p < 10$ say.

For large p, again various restrictive versions are considered in order to reduce the number of parameters. A classic restrictive version is the scalar-BEKK-MGARCH where, with S some *a priori* known $(p \times p)$ selection matrix (for example $S = I_p$) and α, β scalars,

$$\Omega_t(\theta) = \Omega + \alpha\left(SX_{t-1}X_{t-1}'S'\right) + \beta\left(S\Omega_{t-1}S'\right).$$

Here $\theta = (\Omega, \alpha, \beta)$, with $\Omega, \alpha, \beta > 0$ and the number of parameters is reduced to $n + 2$; however, note that such simplifications, while appealing to "classic univariate GARCH reasoning", they may be of less empirical relevance due to the restriction of matrices to scalars.

Remark 7.3 (**Covariance Targeting**). *Also note the previously discussed (co-)variance targeting reduces the numerical problems with higher dimensions. In covariance targeting, see Pedersen and Rahbek (2014) for a complete statistical analysis, the covariance is first estimated by an initial (simple) estimator (such as for example the sample covariance matrix, S_{xx}), which is kept fixed in the next step. In the second step the likelihood function is maximized over the remaining parameters A_i and B_i. Specifically, for the MARCH(1,1), the unconditional variance $\mathrm{V}(X_t) = \Sigma$ is (implicitly) given as the solution to*

$$\Sigma = \mathrm{V}(X_t) = \Omega + A\Sigma A'. \tag{7.7}$$

Using (7.7), rewrite $\Omega_t(\theta)$ as follows,

$$\Omega_t(\theta) = \Omega + AX_{t-1}X_{t-1}'A' = \Sigma - A\Sigma A' + AX_{t-1}X_{t-1}'A'. \tag{7.8}$$

With this formulation covariance target estimation consists of the following two estimation steps: (i) Set $\hat{\Sigma} = S_{xx}$, the sample-covariance (or set $\hat{\Sigma}$ equal to some other estimator); (ii) Maximize the L_{BEKK} in (7.6) over A for fixed Σ, using instead of $\Omega_t(\theta) = \Omega_t(\Omega, \Sigma)$ the alternative expression,

$$\Omega_t(A) = \Omega_t(\hat{\Sigma}, A) = \hat{\Sigma} - A\hat{\Sigma}A' + AX_{t-1}X_{t-1}'A'.$$

Remark 7.4 (BEKK Unconditional Variance). *Note that the equation in (7.7) can be solved explicitly by using properties of the* vec *operator. It follows that* $\text{vec}(\Sigma)$ *solves the equation,*

$$\text{vec}(\Sigma) = \text{vec}(\Omega) + (A \otimes A)\,\text{vec}(\Sigma),$$

which has solution,

$$\text{vec}(\Sigma) = \left(I_{p^2} - (A \otimes A)\right)^{-1} \text{vec}(\Omega),$$

provided $\rho(A \otimes A) < 1$. *Here* $\rho(A \otimes A)$ *denotes the largest eigenvalue of the matrix* $(A \otimes A)$ *in absolute value and* $(A \otimes A)$ *is the* $(p^2 \times p^2)$ *matrix given by,*

$$(A \otimes A) = \begin{pmatrix} a_{11}A & & a_{1p}A \\ & \ddots & \\ a_{p1}A & & a_{pp}A \end{pmatrix}.$$

The condition on $(A \otimes A)$ *is equivalent to the condition that* X_t *is stationary and ergodic with finite second order moments,* $\mathrm{E}\|X_t\|^2 < \infty$, *see Section 8. In particular, it implies that* X_t *is covariance-stationary with* $\mathrm{V}(X_t) = \Sigma$, *where* Σ *is finite and positive definite.*

7.3 CCC-MGARCH

The constant conditional correlation (CCC) MGARCH model applies the correlation representation in (6.10) of Ω, where $\Omega = D\Gamma D$ with the diagonal (square root of) variance matrix $D = \text{diag}(\sigma_i)_{i=1}^p$, while $\Gamma = (\rho_{ij})_{i,j=1,\ldots,p}$ is the correlation matrix.

Similar to vech-MARCH(1) and BEKK-MARCH(1), the CCC-MARCH(1) model generalizes the univariate ARCH(1) by defining the variances σ_i^2 as time-varying, σ_{it}^2, with an ARCH(1) formulation for each. Specifically, the conditional covariance MARCH(1) model is given by $X_t = \Omega_t^{1/2}(\theta)\, z_t$ with,

$$\Omega_t(\theta) = D_t(\theta)\,\Gamma(\theta)\,D_t(\theta), \tag{7.9}$$

and z_t i.i.d. $\mathrm{N}(0, I_p)$. Moreover,

$$D_t(\theta) = \text{diag}\left((\sigma_{it}(\theta))_{i=1}^p\right), \text{ with } \sigma_{it}^2(\theta) = \omega_i + \alpha_i X_{it-1}^2, \tag{7.10}$$

while $\Gamma(\theta)$ is (the constant conditional) correlation matrix,

$$\Gamma(\theta) = (\rho_{ij})_{i,j=1,\ldots,p} \text{ with } \rho_{ii} = 1.$$

Thus, each X_{it} is assumed to marginally follow an ARCH(1) specification with the constant conditional (CC) correlation matrix allowing for interactions through the CC coefficients. For example in the bivariate case, $\Omega_t(\theta)$ with $\rho_{12} = \rho$ is given by,

$$\Omega_t(\theta) = \begin{pmatrix} \sigma_{1t}^2(\theta) & \sigma_{1t}(\theta)\,\sigma_{2t}(\theta)\,\rho \\ \sigma_{1t}(\theta)\,\sigma_{2t}(\theta)\,\rho & \sigma_{2t}^2(\theta) \end{pmatrix}.$$

More generally, the specification of the individual time-varying conditional variances σ_{it}^2 can be chosen as any of the well-known (G)ARCH specifications. An example is an asymmetric GARCH(1,1) specification with some exogenous explanatory (news) variable N_t entering,

$$\sigma_{it}^2 = \omega_i + \alpha_i X_{it-1}^2 + \beta_i \sigma_{it-1}^2 + \gamma N_{t-1} \mathbb{I}\left(N_{t-1} > 0\right),$$

or the following, which is sometimes referred to as extended CCC (ECCC) GARCH,

$$\sigma_{it}^2 = \omega_i + \alpha_i X_{it-1}^2 + \beta_i \sigma_{it-1}^2 + \alpha_j X_{jt-1}^2, \text{ for } j \neq i\,.$$

As for estimation, the CCC-MGARCH likelihood function $L_{\text{CCC}}(\theta)$, is given by

$$\begin{aligned} -\frac{2}{T} L_{\text{CCC}}(\theta) &= T^{-1} \sum_{t=1}^{T} \left(\log \left| \Omega_t(\theta) \right| + \operatorname{tr}\left(\Omega_t^{-1}(\theta) X_t X_t' \right) \right) \\ &= T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{p} \log \sigma_{it}^2(\theta) + \left(\log \left| \Gamma(\theta) \right| + \operatorname{tr}\left(\Gamma(\theta)^{-1} X_{D,t} X_{D,t}' \right) \right) \right) \end{aligned} \tag{7.11}$$

where $X_{D,t} = D_t(\theta)^{-1} X_t$. The MLE is defined by

$$\hat{\theta}_{\text{CCC}} = \arg\max L_{\text{CCC}}(\theta)\,,$$

where maximization is over θ with θ containing the parameters parametrizing the univariate GARCH specifications, together with the parameters parametrizing the correlation matrix $\Gamma(\theta)$.

Remark 7.5 (**Two-Step Estimation**). *Similar to the discussion in Section 6.3, an alternative (often applied) to the ML estimation is two-step (TS) estimation. The idea is to first estimate the p univariate ARCH models, and define $\hat{X}_t = \hat{D}_t^{-1} X_t$ where $\hat{D}_t = \operatorname{diag}\left((\hat{\sigma}_{it,TS})_{i=1}^p \right)$ with $\hat{\sigma}_{it,TS}^2$ the estimated time-varying ARCH conditional variances. With $\hat{\sigma}_{it,TS}^2$ fixed, and hence $\hat{X}_t$ known, the likelihood in (7.11) reduces to the concentrated version,*

$$L_{TS,\ CCC}(\theta) = \log \left| \Gamma(\theta) \right| + \operatorname{tr}\left(\Gamma(\theta)^{-1} S_{\hat{x}\hat{x}} \right), \quad \text{where } S_{\hat{x}\hat{x}} = T^{-1} \sum_{t=1}^{T} \hat{D}_t^{-1} X_t X_t' \hat{D}_t.$$

Applying Lemma 6.1, it follows that

$$\hat{\Gamma}_{TS} = S_{\hat{x}\hat{x}} = T^{-1} \sum_{t=1}^{T} \hat{D}_t^{-1} X_t X_t' \hat{D}_t.$$

Collecting terms, $\hat{\theta}_{TS,CCC} = (\hat{\theta}_{GARCH}, \hat{\Gamma}_{TS})$, with θ_{GARCH} the parameters parametrizing the univariate GARCH models, e.g. $\theta_{GARCH} = (\omega_i, \alpha_i)_{i=1}^p$ if the CCC-MARCH(1) model is considered.

7.3.1 Dynamic Conditional Correlation (DCC) MGARCH

A natural extension of the CCC-MGARCH model is to make the CC matrix Γ time-varying as well, which leads to the (class of) dynamic conditional correlation (DCC) MGARCH models as originally proposed by Engle (2002).

With $D_t(\theta)$ as for the CCC-MGARCH models, the challenge is to formulate a time varying sequence Γ_t of correlation matrices. One immediate option is to consider using a time-varying equivalent of the decomposition in Section 6.3 of Γ as $\Gamma(\theta) = G(\theta) G(\theta)'$. This way,

$$\Gamma_t = G_t(\theta) G_t(\theta)',$$

where $G_t(\theta)$ is time varying and defined in terms of trigonometric functions $\cos(\cdot)$ and $\sin(\cdot)$. For $p = 2$,

$$G_t(\theta) = \begin{pmatrix} 1 & 0 \\ \cos(\gamma_t(\theta)) & \sin(\gamma_t(\theta)) \end{pmatrix},$$

with $\gamma_t(\theta)$ a time-varying angle, $\gamma_t(\theta) \in (0, \pi/2)$, suitably parameterized in terms of θ. However, this may be difficult to interpret and an alternative is the approach in Engle (2002), which can be motivated by applying again the correlation decomposition.

Thus with Q any positive definite (covariance) matrix, then $Q = \operatorname{diag}(Q)^{1/2} \Gamma \operatorname{diag}(Q)^{1/2}$. That is, for any $Q > 0$,

$$\Gamma = \operatorname{diag}(Q)^{-1/2} Q \operatorname{diag}(Q)^{-1/2}$$

is by definition a correlation matrix. In particular with $Q_t(\theta)$ a time-varying covariance matrix parametrized by θ, then

$$\Gamma_t(\theta) = \operatorname{diag}(Q_t(\theta))^{-1/2} Q_t(\theta) \operatorname{diag}(Q_t(\theta))^{-1/2},$$

is a time-varying, or dynamic, conditional correlation matrix. Adapting the notation from the CCC model and setting $X_{D,t} = D_t(\theta)^{-1} X_t$, a BEKK-ARCH specification of $Q_t(\theta)$ is for example given by,

$$Q_t(\theta) = Q - AQA' + AX_{D,t-1}X_{D,t-1}'A'.$$

Note that $\Sigma = Q - AQA'$ is the unconditional variance of $X_{D,t}$ (see the discussion of covariance targeting in the BEKK model in Remark 6.5). This is a generalized version of the scalar BEKK-ARCH, as given by

$$Q_t(\theta) = (1 - \alpha) Q + \alpha X_{D,t-1}X_{D,t-1}',$$

see Engle (2002). Most applications of the DCC model use the scalar BEKK-GARCH formulation

$$Q_t(\theta) = (1 - \alpha - \beta) Q + \alpha X_{D,t-1}X_{D,t-1}' + \beta Q_{t-1}(\theta).$$

7.4 Dynamic λ-MGARCH

The dynamic λ-MGARCH model in Hetland, Pedersen, and Rahbek (2019) applies the decomposition of Ω in terms of eigenvalues (λ) and eigenvectors in Section 6.5. As the eigenvalues are key-determinants of the dynamics of a system, see also the discussion of principal components in Appendix B, these are allowed to be time-varying in the λ-MGARCH model.

Recall initially that the eigenvalue decomposition in Section 6.5 is given by $\Omega = V\Lambda V'$, where

$$\Lambda = \text{diag}\left((\lambda_i)_{i=1}^p\right)$$

and $V = (v_1, ..., v_p)$ with $V'V = VV' = I_p$.

The λ-MGARCH(1,1) model is given by $X_t = \Omega_t^{1/2}(\theta)\, z_t$, where Λ is time-varying, such that

$$\Omega_t(\theta) = V(\theta)\,\Lambda_t(\theta)\,V(\theta), \quad \Lambda_t(\theta) = \text{diag}\left((\lambda_{it}(\theta))_{i=1}^p\right), \tag{7.12}$$

where $V(\theta)'V(\theta) = V(\theta)V(\theta)' = I_p$.

In terms of parametrization of the constant conditional eigenvectors $V(\theta) \in \mathbb{R}^{p\times p}$ a useful parametrization for $V(\theta)$ is as discussed in Section 6.5 in terms of trigonometric functions. For the specification of the time-varying conditional eigenvalues $\lambda_{it}(\theta)$ in $\Lambda_t(\theta)$ a GARCH(1,1)-type scheme can be applied for the vector

$$\lambda_t(\theta) = (\lambda_{1t}(\theta), \ldots, \lambda_{pt}(\theta))'$$

of strictly positive eigenvalues $\lambda_{it}(\theta) > 0$. Specifically, consider here,

$$\lambda_t(\theta) = C + A(V(\theta)'X_{t-1})^{\odot 2} + B\lambda_{t-1}, \tag{7.13}$$

where for any vector $X = (X_1, ..., X_p)'$, $X^{\odot 2} = \left(X_1^2, ..., X_p^2\right)'$.

In addition to the parameters parametrizing the conditional eigenvectors (or, rotation matrices) $V(\theta)$, the parameters of the model are therefore given by C, A and B. Here C is a p-dimensional vector $C = (c_1, ..., c_p)'$ with strictly positive entries, $c_i > 0$, while the $(p \times p)$ dimensional matrices $A = (a_{ij})_{i,j=1,..,p}$ and $B = (b_{ij})_{i,j=1,..,p}$ are assumed to have non-negative entries, $a_{ij}, b_{ij} \geq 0$.

The λ-MGARCH MLE $\hat{\theta}_\lambda$ is by definition given by,

$$\hat{\theta}_\lambda = \arg\max L_\lambda(\theta),$$

where the log-likelihood function $L_\lambda(\theta)$ is given by,

$$-\frac{2}{T}L_\lambda(\theta) = T^{-1}\sum_{t=1}^{T}\left(\log|\Omega_t(\theta)| + \text{tr}\left(\Omega_t^{-1}(\theta)X_tX_t'\right)\right).$$

An important interpretation of the λ-MGARCH model can be given by observing that $L_\lambda(\theta)$ simplifies. That is, using the identity,

$$\log|\Omega_t(\theta)| = \sum_{i=1}^{p}\log\lambda_{it}(\theta),$$

and setting $\tilde{X}_t = V(\theta)' X_t$ such that

$$X_t' \Omega_t^{-1}(\theta) X_t = \tilde{X}_t' \Lambda_t(\theta)^{-1} \tilde{X}_t,$$

the likelihood function can be re-stated as

$$-\frac{2}{T} L_\lambda(\theta) = \frac{1}{T} \sum_{i=1}^{p} \left(\sum_{t=1}^{T} \left(\log \lambda_{it}(\theta) + \tilde{X}_{it}^2 / \lambda_{it}(\theta) \right) \right).$$

That is, the model may be viewed as p univariate (GARCH) models for the rotated (by $V(\theta)$) $\tilde{X}_{it}$ in $\tilde{X}_t = V(\theta)' X_t$. This is analogous to the GO-GARCH and Rotated-GARCH models studied in Boswijk and van der Weide (2011) and Noureldin, Shephard, and Sheppard (2014) and references therein, where similar models for rotations, or linear combinations, of the original returns X_t are considered.

8 Stochastic properties of MGARCH processes

In this section, stationarity and ergodicity as well as existence of (unconditional) variance X_t are discussed for the vech-MGARCH, BEKK-GARCH, CCC-MGARCH and λ-MGARCH processes.

The results presented rely on general Markov chain theory, and the application of the well-known drift criterion from there, see e.g. Nielsen and Rahbek (2014). This is similar to the approach when verifying stochastic properties of univariate GARCH models. However, as the processes under consideration here are multivariate, the choice of drift function is more delicate than for univariate processes. Moreover, while in the univariate case the conditional variance σ_t^2 is a positive scalar, the multivariate conditional variance Ω_t is a positive definite matrix. In other words, the state space for (the univariate Markov chain) σ_t^2 is the positive real numbers, which has a simple structure. On the other hand, the state space of positive definite matrices for (the multivariate Markov chain) Ω_t has a highly complicated structure, requiring sophisticated theory, see e.g. the discussion in Boussama, Fuchs, and Stelzer (2011).

Throughout this section it is assumed that the p-dimensional process X_t is given by,

$$X_t = \Omega_t^{1/2} z_t,$$

with z_t i.i.d. $\mathrm{N}_p(0, I_p)$ and Ω_t is given by one of the discussed MGARCH specifications.

Note that the dependence on θ in Ω_t is suppressed as all statements are probability theory statements and hence the parameter θ is fixed, and not varying as for the statistical model. Similarly for other quantities in the expressions the θ is suppressed.

8.1 The BEKK-MARCH(1)

To illustrate some of the considerations consider first the BEKK-MARCH(1) process (BEKK-ARCH henceforth), where we apply Markov chain theory as in the proof of Theorem 1 in Nielsen and Rahbek (2014).

The BEKK-ARCH is given by $X_t = \Omega_t^{1/2} z_t$, with

$$\Omega_t = \Omega + A X_{t-1} X_{t-1} A'. \tag{8.1}$$

By definition X_t is a Markov chain on $\mathbb{R}^p$ with a Gaussian transition, or conditional, density f given by,

$$f\left(X_t | X_{t-1}\right) = (2\pi)^{-p/2} \left|\Omega_t\right|^{-1/2} \exp\left(-\frac{1}{2} X_t' \Omega_t^{-1} X_t\right),$$

with $f(\cdot) > 0$, and continuous in X_t and X_{t-1}. Hence the drift criterion can be used to establish geometric ergodicity and stationarity of X_t.

As to the choice of drift function we rewrite X_t as a random coefficient vector autoregression (RCA) as a rich theory applying drift criteria to RCA processes exist. It follows that,

$$X_t = \Phi_t X_{t-1} + \epsilon_t, \tag{8.2}$$

where ϵ_t and Φ_t are random mutually independent i.i.d. sequences. Here ϵ_t is p-dimensional and $N_p(0, \Omega)$ distributed, while Φ_t is a $(p \times p)$-dimensional matrix with a Gaussian (matrix) distribution. Specifically, $\mathrm{E}(\Phi_t) = 0_{p \times p}$, and a has a covariance structure defined as,

$$\mathrm{E}\left(\Phi_t \otimes \Phi_t\right) = \left(A \otimes A\right).$$

To see that indeed $X_t = \Omega_t^{1/2} z_t$ with Ω_t given by (8.1) and X_t given by (8.2) are identical in distribution we consider the conditional mean and variance of each representation. This is sufficient as the conditional distribution is Gaussian, and hence equality of mean and variance implies equality. Clearly, for both representations $\mathrm{E}\left(X_t | X_{t-1}\right) = 0$, as

$$\begin{aligned}
\mathrm{E}\left(X_t | X_{t-1}\right) &= \Omega_t^{1/2} \mathrm{E}\left(z_t\right) = 0 \text{ and} \\
\mathrm{E}\left(X_t | X_{t-1}\right) &= \mathrm{E}\left(\Phi_t\right) X_{t-1} + \mathrm{E}\left(\epsilon_t\right) = 0.
\end{aligned}$$

Next, consider $\mathrm{V}\left(X_t | X_{t-1}\right) = \mathrm{E}\left(X_t X_t' | X_{t-1}\right)$. For X_t given by (8.1), obviously, $\mathrm{V}\left(X_t | X_{t-1}\right) = \Omega_t$ which can be stated in terms of the vec as[2],

$$\begin{aligned}
\mathrm{vec}\left(\mathrm{V}\left(X_t | X_{t-1}\right)\right) &= \mathrm{vec}\left(\mathrm{E}\left(X_t X_t' | X_{t-1}\right)\right) = \mathrm{vec}\left(\Omega_t\right) \\
&= \mathrm{vec}\left(\Omega\right) + \left(A \otimes A\right) \mathrm{vec}\left(X_{t-1} X_{t-1}'\right).
\end{aligned}$$

Next, with X_t given by (8.2),

$$\begin{aligned}
\mathrm{vec}\left(\mathrm{V}\left(X_t | X_{t-1}\right)\right) &= \mathrm{vec}\left(\mathrm{E}\left(X_t X_t' | X_{t-1}\right)\right) = \mathrm{E}\left(\mathrm{vec}\left(X_t X_t'\right) | X_{t-1}\right) \\
&= \mathrm{E}\left(\left(\Phi_t \otimes \Phi_t\right) \mathrm{vec}\left(X_{t-1} X_{t-1}'\right)\right) + \mathrm{vec}\,\mathrm{E}\left(\epsilon_t \epsilon_t'\right) \\
&= \left(A \otimes A\right) \mathrm{vec}\left(X_{t-1} X_{t-1}'\right) + \mathrm{vec}\left(\Omega\right).
\end{aligned}$$

[2] Using the identity

$$\mathrm{vec}\left(ABC\right) = \left(C' \otimes A\right) \mathrm{vec}\left(B\right)$$

for any matrices A, B and C (such that ABC is well-defined),

The next lemma is based on Nielsen and Rahbek (2014, Theorem 1) where a drift function of the form,

$$\delta(X) = 1 + \|X\|^{\varepsilon}$$

is applied, with $\|X\|^2 = X'X$. Likewise, with A a matrix, the matrix norm is given by $\|A\|^2 = \operatorname{tr}(A'A)$ (other matrix norm can be applied) in terms of which we can state the following lemma. The proof is given in Appendix 8.1.

Lemma 8.1. *Consider* $X_t = \Omega_t^{1/2} z_t$, z_t *i.i.d.* $N(0, I_p)$ *and* $\Omega_t = \Omega + AX_{t-1}X_{t-1}'A'$. *Define the Lyapunov coefficient* ϱ_L *by,*

$$\varrho_L = \lim_{m\to\infty} (\tfrac{1}{m} \operatorname{E} \log ||\prod_{t=1}^{m} \Phi_t||),$$

where Φ_t *is an i.i.d. Gaussian* $(p \times p)$ *sequence, with* $\operatorname{E}(\Phi_t) = 0$ *and* $\operatorname{E}(\Phi_t \otimes \Phi_t) = (A \otimes A)$.

If $\varrho_L < 0$, *then* X_t *is geometrically ergodic and has a stationary solution with* $\operatorname{E}\|X_t\|^{\delta} < \infty$ *for some small* $\delta > 0$. *Moreover,* X_t *is geometrically ergodic and stationary with finite variance or,* $\operatorname{E}\|X_t\|^2 < \infty$, *provided the stronger condition,*

$$\varrho_2 = \rho(A \otimes A) < 1,$$

holds. Here $\rho(A \otimes A)$ *is the spectral radius of* $(A \otimes A)$, *i.e. the largest eigenvalue of* $(A \otimes A)$ *in absolute value.*

Recall that for the ARCH(1) process, the condition of second order moments is that $\alpha < 1$ while for stationarity, the condition is that $\log \alpha < -E \log(z^2)$

$$\operatorname{E} \log(\alpha z^2) = \varrho_L < 0, \text{ or } \alpha < \exp(-\operatorname{E} \log(z^2)) \simeq 3.56.$$

Thus as $\varrho_2 < 1$ implies $\varrho_L < 0$ also for the multivariate case, then – analogous to the univariate GARCH processes – the condition on (strict) stationarity, although implicit, is milder than the (strong) requirement of stationarity and finite variance (covariance stationarity).

Some more remarks are in order:

Remark 8.2 (Analogy to ARCH(1))**.** *In line with this for this simple BEKK-ARCH(1) – as argued in ?– it follows that* $\Phi_t \overset{D}{=} A\eta_t$, *where* η_t *is a i.i.d.* $N(0,1)$ *sequence. Hence using the identity* $||A^m||^{1/m} \to \rho(A)$,

$$\varrho_L = \lim_{m\to\infty} (\tfrac{1}{m} \operatorname{E} \log ||\prod_{t=1}^{m} \Phi_t||) = \operatorname{E} \log |\eta_t| + \lim_{m\to\infty} \log ||A^m||^{1/m} = \tfrac{1}{2} E \log(\eta_t^2) + \log \rho(A),$$

such that the condition $\varrho_L < 0$ *becomes,* $\rho(A) < \exp(-\frac{1}{2} E \log(\eta_t^2))$. *Equivalently,*

$$\rho(A \otimes A) < \exp(-E \log(\eta_t^2)) \simeq 3.56.$$

Remark 8.3 (Lyapunov Coefficient). *The coefficient ϱ_L is the so-called Lyapunov coefficient. In general no explicit representation can be derived and hence the condition in practice is verified by numerical simulations.*

Note that if $\Phi_t = \Phi$, that is Φ_t is non-stochastic and constant, then

$$\frac{1}{m} \log || \prod_{i=1}^{m} \Phi_i | = \log ||\Phi^m||^{1/m}.$$

As by definition the spectral radius $\rho(\Phi)$ satisfies, $\rho(\Phi) = \lim_{m\to\infty} ||\Phi^m||^{1/m}$, the condition thus becomes $\varrho_L = \log \rho(\Phi) < 0$, which is equivalent to the well-known condition,

$$\rho(\Phi) < 1.$$

That is, if $X_t = \Phi X_{t-1} + \epsilon_t$, it reduces to the well-known condition for stationarity and existence of all moments.

8.2 vech and BEKK

Consider first the vech-MGARCH process as discussed in Section 7.1.

Boussama, Fuchs, and Stelzer (2011, Theorem 2.4) establishes conditions for the vech (Ω_t) parametrization which is given by

$$\text{vech}(\Omega_t) = \text{vech}(C_{\text{vech}}) + A_{\text{vech}} \text{vech}\left(X_{t-1}X_{t-1}'\right) + B_{\text{vech}} \text{vech}(\Omega_{t-1}), \tag{8.3}$$

where A_{vech}, B_{vech} and C_{vech} are $(p \times p)$ dimensional matrices with $C_{\text{vech}} > 0$. In terms of the definition of the vech-MGARCH in (7.2), $C = \text{vech}(C_{\text{vech}})$, while $A_{\text{vech}} = A$ and $B_{\text{vech}} = B$. In other words, here $\text{unvech}(C) = C_{\text{vech}}$ is positive definite by assumption, while this was not assumed for C in the original $\text{vech}(\Omega_t)$ parametrization.

With $X_t = \Omega_t^{1/2} z_t$, Ω_t given by (8.3) and with z_t i.i.d. $\mathrm{N}(0, I_p)$, Theorem 2.4 in Boussama, Fuchs, and Stelzer (2011) establishes that the vector $(p+n)$-dimensional vector consisting of X_t and $\text{vech}(\Omega_t)'$ is: *(i)* a Markov chain which is geometrically ergodic, and has a stationary representation, and *(ii)* $\mathrm{E}\|X_t\|^2 < \infty$, provided the following condition holds,

$$\varrho_2 = \rho(A_{\text{vech}} + B_{\text{vech}}) < 1. \tag{8.4}$$

Similar to the arguments applied to the BEKK-ARCH process above, the proof in Boussama, Fuchs, and Stelzer (2011) is based on rewriting the vector $\left(X_t', \text{vec}(\Omega_t)'\right)'$ as a Markov chain and apply the drift criterion. However, a crucial difference is that here the state space of Ω_t (or $\text{vec}\,\Omega_t$) is considered, which is why the condition given is the *strong* condition in terms of ergodicity and finite variance. In contrast, for the BEKK-ARCH the state space of X_t is simple, and the condition for ergodicity and stationarity (and fractional moments), $\varrho_L < 0$, was stated in addition to the stronger condition, $\varrho_2 < 1$.

Some further remarks are:

Remark 8.4 (**Positive Definiteness**). *In terms of the* vech*-MGARCH process in Section 7.1, the condition becomes,* $\rho(A+B)<1$ *and the requirement that* C *is such that* C *is given by the "*unvech*" of some positive definite matrix.*

Remark 8.5 (**GARCH(1,1) Condition**). *The condition in (8.4) is analogous to the univariate condition for a GARCH(1,1) process,* $\alpha+\beta<1$.

Next turn to the BEKK-MGARCH process in Section 7.2, where Ω_t is given by,

$$\Omega_t = C + AX_{t-1}X_{t-1}'A' + B\Omega_{t-1}B'.$$

It follows by using properties of the vec $(\cdot)$ operator,

$$\operatorname{vec}(\Omega_t) = \operatorname{vec}(C) + (A\otimes A)\operatorname{vec}\left(X_{t-1}X_{t-1}'\right) + (B\otimes B)\operatorname{vec}(\Omega_{t-1}).$$

Using further properties of the vec (and vech) one can write the BEKK of the vech form, and it follows that the condition $\rho(A_{\text{vech}}+B_{\text{vech}})<1$ from Boussama, Fuchs, and Stelzer (2011) for the BEKK-MGARCH process reduces to

$$\varrho_2 = \rho((A\otimes A)+(B\otimes B))<1, \tag{8.5}$$

see also Pedersen and Rahbek (2014).

8.2.1 λ-GARCH

Consider next the λ-MGARCH process in Section 7.4, with Ω_t and λ_t satisfying (7.12) and (7.13),

$$\begin{aligned}\Omega_t &= V\Lambda_t V', \text{ with } \Lambda_t = \operatorname{diag}\left((\lambda_{it})_{i=1}^p\right) = \operatorname{diag}(\lambda_t),\\ \lambda_t &= C + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1}.\end{aligned} \tag{8.6}$$

Here $C=(c_1,...,c_p)'$, $c_i>0$ while A and B are $(p\times p)$ matrices with non-negative (larger than or equal to zero) entries.

By multiplication with V, it follows by (8.6) that $V'X_t = V'\Omega_t^{1/2}z_t$ satisfies,

$$V'X_t = \Lambda_t^{1/2}z_t, \text{ with } \lambda_t = C + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1}. \tag{8.7}$$

Next, observe that one can rewrite the dynamics of λ_t in (8.7) in terms of lagged λ_t and z_t. That is, one can rewrite as a stochastic recurrence equation given by,

$$\lambda_t = C + A\operatorname{diag}\left((z_{i,t}^2)_{i=1}^p\right)\lambda_{t-1} + B\lambda_{t-1} = C + \Phi_t\lambda_{t-1}, \tag{8.8}$$

where $z_t=(z_1,...,z_{pt})'$. The matrix coefficient Φ_t is stochastic and is given by,

$$\Phi_t = A\operatorname{diag}\left((z_{i,t}^2)_{i=1}^p\right) + B, \tag{8.9}$$

such that the expression for λ_t in (8.8) is similar to a RCA process.

Theorem 3.1 in Hetland, Pedersen, and Rahbek (2019) states that for $X_t = \Omega_t^{1/2} z_t$ with Ω_t given by (8.6), then X_t is stationary and geometrically ergodic if, and only if, the Lyapunov coefficient ϱ_L is negative, that is

$$\varrho_L := \lim_{m\to\infty} \frac{1}{m} \mathrm{E}\left[\log || \prod_{t=1}^{m} \Phi_t ||\right] < 0,$$

Moreover, X_t is stationary with $\mathrm{E}\,||X_t||^2 < \infty$ if, and only if, the classic GARCH-type restriction applies,

$$\varrho_2 := \rho\,(A + B) < 1.$$

9 Asymptotics for the MLE in MGARCH models

In general, as for univariate GARCH models, each MGARCH model requires a detailed analysis to get a full understanding of which regularity conditions are needed for consistency and asymptotic normality of the MLE $\hat{\theta}$ of θ. Existing papers in the literature provide asymptotic theory for some of the models already discussed, see also below. A common feature of the different regularity conditions is that it is typically assumed that process X_t is stationary and ergodic, often with some high-order finite moment condition such as $\mathrm{E}\,||X_t||^k < \infty$ for some $k \geq 6$. Such moment conditions, while sufficient, may not be necessary. Their origin is from various applications of matrix inequalities in the proofs and may possibly reflect that the theory is not complete yet. At least the assumptions are often challenged in simulations, which even for simulated processes with only finite $2+\delta$ or $4+\delta$ moments, where δ is some small positive number, indicate that asymptotic normality of $\hat{\theta}$ still holds.

This is similar to the theory for univariate GARCH(1,1) model, where for example the classic requirement of a finite second order moment, or equivalently, $\alpha + \beta < 1$ is not needed for asymptotic normality of the MLE, see for example Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b).

Asymptotic theory is given in several papers, which include (see also the references therein): The BEKK-MGARCH is discussed in Comte and Lieberman (2003) and Pedersen and Rahbek (2014). Asymptotic theory for the λ-MGARCH is given in Hetland, Pedersen, and Rahbek (2019), and for the CCC-MGARCH model in Pedersen (2017) and Francq and Zakoïan (2012).

9.1 The scalar BEKK-ARCH

In this section the scalar version of the BEKK-ARCH(1) model is briefly considered in order to introduce the underlying reasoning for the asymptotic theory.

The scalar BEKK-ARCH considered is given by $X_t = \Omega_t^{1/2} z_t$ with z_t i.i.d. $\mathrm{N}(0, I_p)$ and

$$\Omega_t\,(\theta) = \Omega + \alpha X_{t-1} X_{t-1}',$$

where the scalar α is non-negative, $\alpha \geq 0$, while here $\Omega > 0$ is kept fixed. As Ω is fixed, $\theta = \alpha$ is a scalar which simplifies the arguments somewhat.

The likelihood function is given by (7.6), and with $\ell_T(\alpha) = -\frac{2}{T} L_{\text{BEKK}}(\alpha)$, it holds that

$$\ell_T(\alpha) = T^{-1} \sum_{t=1}^{T} \left(\log |\Omega_t(\alpha)| + \operatorname{tr}\left(\Omega_t(\alpha)^{-1} X_t X_t' \right) \right).$$

Asymptotic theory for MGARCH models, and non-linear models in general, relies on establishing properties as given in Jensen and Rahbek (2004a, Lemma 1) for derivatives of $\ell_T(\alpha)$. These ensure the validity of applying Taylor expansion(s) of (the derivative of) $\ell_T(\alpha)$. The conditions to be validated are: *(i)* asymptotic normality of the normalized score, $s_T(\alpha) = T^{1/2} \partial \ell_T(\alpha) / \partial \alpha$ at $\alpha = \alpha_0$; and, *(ii)* convergence in probability of the information, $i_T(\alpha) = -\partial^2 \ell_T(\alpha) / \partial \alpha^2$ at $\alpha = \alpha_0$. Moreover, it is required that, *(iii)* that $\sup_\alpha |\partial^3 \ell_T(\alpha) / \partial \alpha^3| \leq c_T \xrightarrow{p} c > 0$, where the sup is over all α for which $0 < \alpha_L \leq \alpha \leq \alpha_U$, with $\alpha_0 \in (\alpha_L, \alpha_U)$.

By Lemma 1 in Jensen and Rahbek (2004a), provided *(i)-(iii)* hold, $\hat{\alpha} \xrightarrow{p} \alpha_0$ (in a neighborhood of α_0) and

$$T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{D} N\left(0, \sigma_\alpha^2\right),$$

where $\sigma_\alpha^{-2} = \operatorname{plim} i_T(\alpha_0)$. Pedersen and Rahbek (2014) verify that *(i)-(iii)* hold if X_t is geometrically ergodic and stationary, with $\mathrm{E}\|X_t\|^6 < \infty$. This we illustrate by considering here the score where it is sufficient to require X_t to be geometrically ergodic and stationary, with $\mathrm{E}\|X_t\|^2 < \infty$, or equivalently, $\alpha < 1$, see Section 7.2.

In terms of the score, observe initially that (by standard matrix calculus),

$$\begin{aligned} \partial \log |\Omega_t(\alpha)| / \partial \alpha &= \operatorname{tr}\left(\Omega_t^{-1}(\alpha) X_{t-1} X_{t-1}' \right), \text{ and} \\ \partial \Omega_t(\alpha)^{-1} / \partial \alpha &= -\Omega_t^{-1}(\alpha) X_{t-1} X_{t-1}' \Omega_t^{-1}(\alpha). \end{aligned}$$

Hence, with $\Omega_t(\alpha_0) = \Omega_t = \Omega + \alpha_0 X_{t-1} X_{t-1}'$, and using $z_t = \Omega_t^{-1/2} X_t$,

$$\begin{aligned} s_T(\alpha_0) &= T^{1/2} \partial \ell_T(\alpha_0) / \partial \alpha \\ &= T^{-1/2} \sum_{t=1}^{T} \left(\operatorname{tr}\left(\Omega_t^{-1} X_{t-1} X_{t-1}' \right) - \operatorname{tr}\left(\Omega_t^{-1} X_{t-1} X_{t-1}' \Omega_t^{-1} X_t X_t' \right) \right) \\ &= T^{-1/2} \sum_{t=1}^{T} \left(\operatorname{tr}\left(\Omega_t^{-1/2} X_{t-1} X_{t-1}' \Omega_t^{-1/2} \right) - \operatorname{tr}\left(\Omega_t^{-1/2} X_{t-1} X_{t-1}' \Omega_t^{-1/2} z_t z_t' \right) \right) \\ &= T^{-1/2} \sum_{t=1}^{T} X_{t-1}' \Omega_t^{-1/2} \left(I - z_t z_t' \right) \Omega_t^{-1/2} X_{t-1}. \end{aligned}$$

By the identity, $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec}(B)$, it follows that $s_T(\alpha_0) = T^{-1/2} \sum_{t=1}^{T} s_t$, with the scalar sequence s_t defined by

$$s_t = v_t' \operatorname{vec}\left(I_p - z_t z_t' \right), \text{ where } v_t = \left(\Omega_t^{-1/2} X_{t-1} \otimes \Omega_t^{-1/2} X_{t-1} \right) \in \mathbb{R}^{p^2}.$$

Here the p^2-dimensional vector $\text{vec}\,(I_p - z_t z_t')$ is an i.i.d. sequence with

$$\text{E}\,\text{vec}\,(I_p - z_t z_t') = 0_{p^2},$$

since z_t are i.i.d. $\text{N}(0, I_p)$. Moreover, the vector v_t depends on X_{t-1}, and with $\mathcal{F}_t = (X_t, X_{t-1}, ...)$ it therefore holds that

$$\text{E}\,(s_t|\mathcal{F}_{t-1}) = v_t'\,\text{E}\,\text{vec}\,(I_p - z_t z_t') = 0.$$

Thus the scalar $s_T(\alpha_0)$ is a sum of martingale differences s_t wrt. $\mathcal{F}_t$ (strictly speaking, provided $\mu_\alpha = \text{E}\,|s_t| < \infty$), and standard application of the central limit theorem for martingale differences, gives

$$s_T(\alpha_0) \xrightarrow{D} N\left(0, \sigma_\alpha^2\right) \text{ as } T \to \infty,$$

with $\sigma_\alpha^2 = \text{E}\,s_t^2$.

As to the existence of finite moments of X_t, note that σ_α^2 (and hence also μ_α) is finite, is implied by $\text{E}\,\|X_t\|^2 < \infty$. The condition of $\text{E}\,\|X_t\|^4 < \infty$ is needed for the information to have finite expectation, while the strong condition $\text{E}\,||X_t||^6 < \infty$ is used to validate *(iii)*.

A detailed discussion of these issues can be found in Pedersen and Rahbek (2014) and references therein.

Remark 9.1 (**Necessity Versus Sufficiency**). *Note that the condition of* $\text{E}\,\|X_t\|^6 < \infty$ *is conjectured in Pedersen and Rahbek (2014) to be possible to relax to* $\text{E}\,\|X_t\|^4 < \infty$.

10 Standardized Residuals and Model Control

Consider the MGARCH process,

$$X_t = \mu_t + \epsilon_t, \quad \epsilon_t = \Omega_t^{1/2} z_t, \quad \text{with} \quad z_t \text{ i.i.d. } \text{N}_p(0, I_p), \tag{10.1}$$

where

$$\mu_t = \text{E}(X_t \mid \mathcal{F}_{t-1}) \tag{10.2}$$
$$\Omega_t = \text{V}(X_t \mid \mathcal{F}_{t-1}) > 0, \tag{10.3}$$

denote the first two moments of X_t conditional in the information set $\mathcal{F}_{t-1}$. Here μ_t could be a constant, $\mu_t = \mu$, or could be a function of $\mathcal{F}_{t-1}$, e.g. as a vector autoregression, while Ω_t is one of the introduced MGARCH models (with X_t replaced by $X_t - \mu_t = \epsilon_t$). This formulation imposes assumptions on:

1. The specification of the conditional mean, μ_t. The assumption implies that $\epsilon_t = X_t - \mu_t$ is serially uncorrelated (and therefore independent in the Gaussian case).

2. The specification of the conditional variance, Ω_t, such that z_t is conditionally homoskedastic.

3. The shape of the conditional distribution, here the Gaussian.

In addition, each specification of μ_t and Ω_t imposes a number of regularity conditions, necessary for application of laws of large numbers and central limit theorems in the asymptotic analysis. As discussed above, these regularity conditions typically involve the existence of unconditional moments, e.g. $\mathrm{E}\left\|X_t\right\|^k < \infty$ for some $k \geq 4$.

In this section we introduce some diagnostics and misspecification tests that can be used to check the model assumptions. First, we define the standardized residuals:

Standardized Residuals: The model assumptions are formulated in terms of the unobserved error terms, z_t. To check the assumptions upon estimation, we define the *standardized residuals* by inserting estimated parameters for the true values, i.e.

$$\hat{z}_t = \hat{\Omega}_t^{-1/2}(X_t - \hat{\mu}_t). \tag{10.4}$$

To implement this, we need to calculate the inverse matrix square-root, $\hat{\Omega}_t^{-1/2}$, at each point in time, which we can find using (6.16), i.e. based on an eigenvalue decomposition of Ω_t at each point in time $t = 1, 2, ..., T$. Note that as

$$(X_t - \mu_t)|\mathcal{F}_{t-1} \overset{D}{=} N(0, \Omega_t),$$

such that

$$\hat{z}_t|\mathcal{F}_{t-1} = \hat{\Omega}_t^{-1/2}(X_t - \hat{\mu}_t)|\mathcal{F}_{t-1} \overset{D}{=} N(\hat{\Omega}_t^{-1/2}(\mu_t - \hat{\mu}_t), \hat{\Omega}_t^{-1/2}\Omega_t\hat{\Omega}_t^{-1/2}) \approx N(0, I_p),$$

where the last approximation follows from consistency, i.e.

$$\hat{\Omega}_t^{-1/2}(\mu_t - \hat{\mu}_t) \xrightarrow{p} 0 \quad \text{and} \quad \hat{\Omega}_t^{-1/2}\Omega_t\hat{\Omega}_t^{-1/2} \xrightarrow{p} I_p,$$

as $T \to \infty$.[3]

Diagnostic Checking: The standard tools for diagnostic checking are based the sample autocorrelation function,

$$C_{Z_t}(k) = \frac{1}{T-k}\sum_{t=k+1}^{T}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})', \tag{10.5}$$

where $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$. By definition, $C_{Z_t}(k)$ is a $(p \times p)$ matrix with diagonal elements given by estimated autocorrelations, $\rho_{ii}(k) = \widehat{\mathrm{Corr}}(Z_{it}, Z_{it-k})$, and off-diagonal elements given by cross auto-correlations, $\rho_{ij}(k) = \widehat{\mathrm{Corr}}(Z_{it}, Z_{jt-k})$ for $i \neq j$. As a diagnostic

[3] Observe, that we could also standardize using the lower triangular Cholesky decomposition, $\hat{L}_t$, because $\hat{L}_t\Omega_t\hat{L}_t' \xrightarrow{p} I_p$. In this case, however, the standardization depends on the ordering of the variables.

for autocorrelation and ARCH-effects in the original data, we may let Z_t be X_t and X_t "squared", respectively, and as a diagnostic check of an estimated model, we may let Z_t be $\hat{z}_t$ and $\hat{z}_t$ "squared".

As a first simple diagnostic, we can look at simple graphs of the autocorrelations, $\rho_{ii}(k)$ for $k = 1, 2, ..., m$, and cross auto-correlations, $\rho_{ij}(k)$ for $i \neq j$ and $k = 1, 2, ..., m$. For the original data and under the assumption of X_t i.i.d., the variance of $\rho_{ii}(k)$ is T^{-1} and we can use $\pm 2T^{-1/2}$ as confidence bounds as a first measure of significance. For the standardized residuals, the variance is unknown, and the diagnostics should be interpreted with care.

A more formal diagnostic is the portmanteau statistic given by

$$P(m) = T^2 \sum_{k=1}^{m} \frac{1}{T-k} \operatorname{tr}\left\{C_{Z_t}(0)^{-1} C_{Z_t}(k) C_{Z_t}(0)^{-1} C_{Z_t}^{-1}(k)\right\}, \tag{10.6}$$

see Hayashi (2000) for discussion of properties and possible alternative definitions. To test for no-autocorrelation in the data or residuals, we may choose Z_t as X_t or $\hat{z}_t$, respectively. In the case of observed data the statistic is $\chi^2_{p^2 m}$ under the null hypothesis of no autocorrelation. This reference distribution is typically also used in the case of $\hat{z}_t$, although this neglects the fact that $\hat{z}_t$ is estimated.

To test for no ARCH-effects, we may choose Z_t as $\operatorname{vech}(X_t X_t')$ or $\operatorname{vech}(\hat{z}_t \hat{z}_t')$, respectively. In the case of observed data the statistic is $\chi^2_{(p(p+1)/2)^2 m}$ under the null hypothesis of no ARCH.

An alternative test is based on a linear regression, with an auxiliary model given by, e.g.

$$\operatorname{vech}(\hat{z}_t \hat{z}_t') = A + B_1 \operatorname{vech}(\hat{z}_{t-1} \hat{z}_{t-1}') + ... + B_m \operatorname{vech}(\hat{z}_{t-m} \hat{z}_{t-m}') + \eta_t,$$

with hypothesis of no ARCH effects given by $B_1 = ... = B_m = 0$.

To test the distributional shape, standard normality tests can be applied, such as the well-known *Jarque-Bera* (JB) test. Under the assumption of normality, the estimated skewness, (S) and kurtosis (K) are asymptotically normal and

$$\xi_S = \frac{T}{6} \cdot \mathsf{S}^2 \xrightarrow{D} \chi_1^2 \quad \text{and} \quad \xi_K = \frac{T}{24} \cdot (\mathsf{K} - 3)^2 \xrightarrow{D} \chi_1^2.$$

Because ξ_S and ξ_K are asymptotically independent, the joint test for normality, $\mathsf{S} = (\mathsf{K} - 3) = 0$ is given by

$$\text{JB} = \xi_S + \xi_K \xrightarrow{D} \chi_2^2.$$

Under the model assumptions, the $\hat{z}_t$ components are independent, and the multivariate statistic is just the sum of the individual statistics, and it is distributed as a χ^2_{2p} under the null hypothesis.

Remark 10.1 (**Alternative Error Distributions**). *In this note, we have considered only the Gaussian QMLE, but the likelihood analysis of the MGARCH model can be extended to other assumptions of the distributional shape. Examples include the Student's t-distribution or the Skew-Student's t-distribution that allow for more probability mass in the tails and potentially also skewness.*

11 Illustration | MGARCH and Portfolio Choice

To illustrate the use of MGARCH to estimate the *conditional* mean and variance as inputs for the portfolio choice, consider daily observations from 2012 to 2019 ($T = 1995$) for $p = 4$ large-cap Danish stocks: Carlberg B, Coloplast B, Demant, and Rockwool B. The daily returns are reported in Figure 6.

As a model for the conditional mean and variance of returns, $X_t \in \mathbb{R}^4$, consider a BEKK model with $k = 1$ loading matrix,

$$\begin{aligned} X_t &= \mu + \epsilon_t \\ \epsilon_t &= \Omega_t^{1/2} z_t \\ \Omega_t &= LL' + A\epsilon_{t-1}\epsilon_{t-1}'A' + B\Omega_{t-1}B', \end{aligned}$$

for $t = 1, 2, ..., T$ and with z_t i.i.d. $\mathrm{N}_4(0, I_4)$. For the observed sample, the average returns have been positive, and we allow for a non-zero expected return by including a constant term, μ, in the equation for the condition mean. Also note, that we have used a Cholesky parametrization to ensure positive definiteness of the constant matrix, $\Omega = LL'$, see also Appendix A.1. The MLEs for the full sample are given by

$$\hat{\mu} = \begin{pmatrix} 0.044473 \\ 0.087868 \\ 0.032557 \\ 0.056367 \end{pmatrix}$$

$$\hat{L} = \begin{pmatrix} 0.10687 & 0 & 0 & 0 \\ -0.49015 & 1.0660 & 0 & 0 \\ -0.13111 & 0.16656 & 2.9530 \cdot 10^{-6} & 0 \\ 0.082351 & -0.014231 & -3.8814 \cdot 10^{-6} & 5.7442 \cdot 10^{-6} \end{pmatrix}$$

$$\hat{A} = \begin{pmatrix} 0.133490 & -0.010297 & 0.023581 & 0.013220 \\ -0.061279 & 0.343640 & -0.054730 & 0.064413 \\ -0.035822 & -0.014544 & 0.321760 & 0.091719 \\ -0.028214 & 0.015913 & 0.086982 & 0.296850 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 0.975460 & 0.115810 & -0.064833 & -0.025603 \\ 0.255290 & -0.065932 & 0.287610 & 0.007857 \\ -0.025237 & 0.601590 & 0.628750 & -0.216170 \\ 0.026451 & -0.245380 & 0.430890 & 0.817720 \end{pmatrix}.$$

Individually the parameters are hard to interpret, but we observe that L is close to being only positive semi-definite.

For each point in time, $t = 1, 2, ..., T$ we may use the estimated mean, $\hat{\mu}$, and the covariance, $\hat{\Omega}_t$, to select a portfolio. Both are measurable at time $t - 1$ and can be considered *ex-ante* at time t.

Remark 11.1 (Timing of Trades). *Here we assume that we can form the portfolio weights, v_{t-1}, at the end of day $t-1$, and look at the return the next day $v_{t-1}'X_t$. This is*

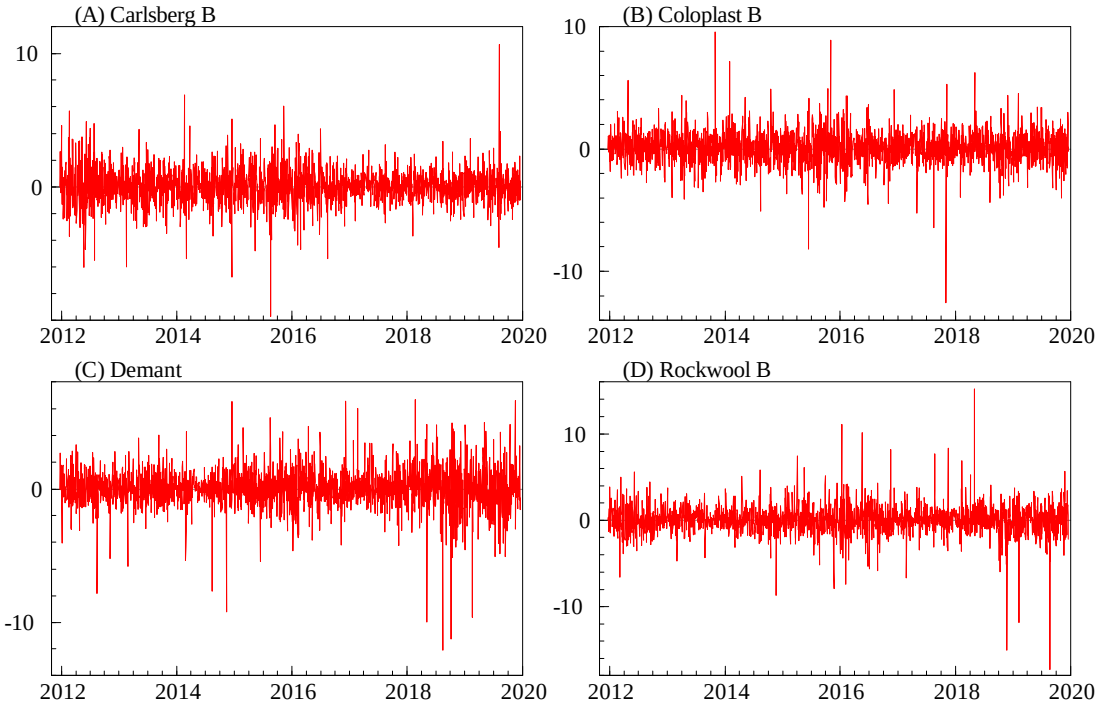

Figure 6: *Daily returns.*

slightly overoptimistic because it is not, in general, possible to buy at the closing price. Alternatively we could have looked at $v'_{t-1}X_{t+1}$. *These implementation issues are extremely important in practice, but outside the scope of this note.*

To illustrate the ideas, we consider the *global minimum variance (GMV) portfolio* with weights given in (3.6) and a *utility optimization (UO)* based on the mean-variance utility in (4.1) with risk aversion given by the parameter $\theta = 0.06$. Figure 7, panels (A)-(C), illustrate the portfolio choice at three different points in time. In July 2013 (T_1) the conditional variances are quite close and the implied GMV portfolio, calculated as a function of $\hat{\mu}$ and $\hat{\Omega}_t = \Omega_t(\hat{\theta})$ for each point in time, is given by the weights

$$\hat{v}_{T_1} = (0.25306, 0.21281, 0.29155, 0.24258)',$$

which is quite close to being equally weighted. In August 2019 (T_2), however, the conditional variance-covariances matrix, $\hat{\Omega}_{T_2}$, is very different and the portfolio weights are given by

$$\hat{v}_{T_2} = (0.47583, 0.37474, -0.051092, 0.20052)'$$

which implies that Demant is sold short.

The cumulated return of the GMV and the UO portfolios are presented in Figure 7 (D). We see that the UO portfolio delivers the highest cumulated return and both portfolios obtain higher cumulated return than a simple equally weighted portfolio.

To compare the performance in more details, the following table report some summary

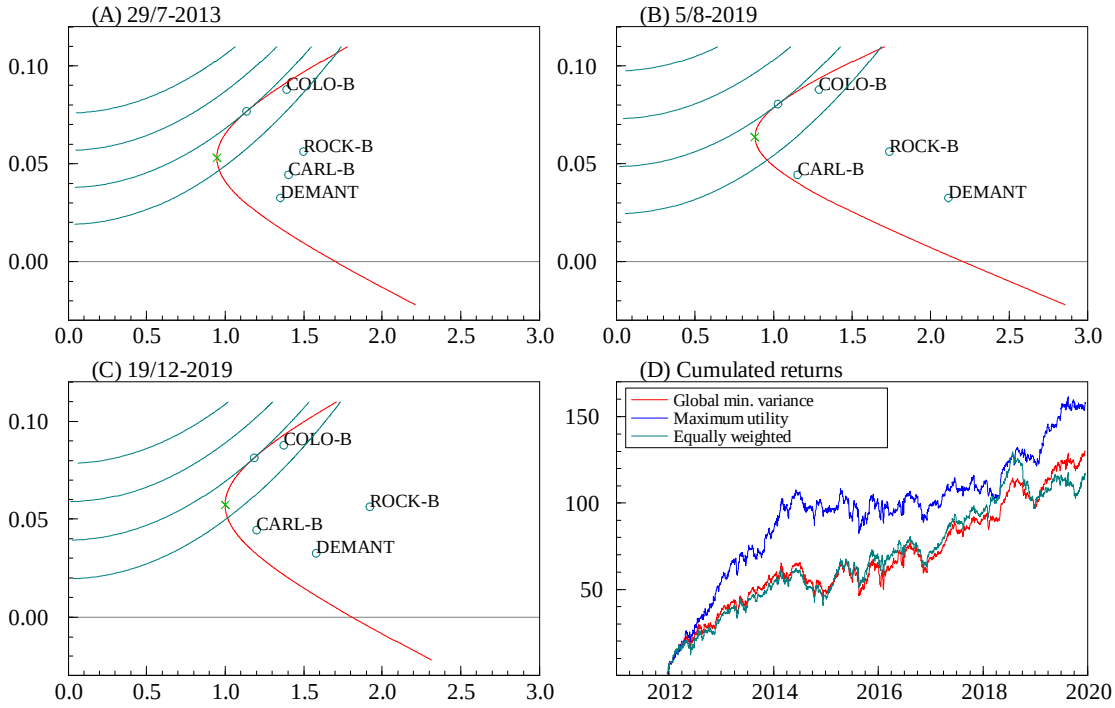

Figure 7: *Opportunity set and cumulated returns.*

statistics

	Global min. variance	Utility optimization	Equally weighted
Average return	0.065237	0.079244	0.058272
Standard deviation	0.97864	1.1717	1.0320
Skewness	−0.19563	−0.27737	−0.29419
Kurtosis	4.9364	6.8955	4.8677
Sharpe ratio	0.066661	0.067632	0.056465

As intended, the GMV portfolio has the smallest variance (and a small fourth moment), but also an average return which is only marginally higher than the simple equally weighted portfolio. The utility is optimized for an agent which is less risk averse, and this portfolio construction produces a higher average return, but also a higher variance (and higher fourth moment).

Measured in terms of the sharpe ratio, which is a simple trade-off between return and risk, differences are small, but the UO is slightly preferred over the GMV portfolio, and both are preferred over the equally weighted portfolio.

Remark 11.2 (**Trading Costs**). *In practice, trading costs are very important, and the implemented portfolios rebalance every day. Other strategies could have been considered, e.g. weekly trading etc., but that is beyond the scope of this illustration.*

Remark 11.3 (**Real-Time Analysis**). *In principle, the analysis should have been implemented in* real time*–such that only information up to time* $t-1$ *is included in the*

decision in time $t-1$. This requires, however, that the model is reestimated at each point in time, $t = T_0, T_0 + 1, ..., T$, where T_0 is the smallest possible sample length required to precisely estimate all parameters. This is time consuming, however, and has not be considered here.

Appendices:

A Other Reparametrizations

A.1 Cholesky Decomposition | Reparametrization

The so-called Cholesky decomposition of positive definite matrices is a much used way to parametrize covariance matrices. It is also the starting point of some existing multivariate ARCH model formulations.

Consider initially the bivariate case where $p = 2$:

Example A.1: With $p = 2$ and Ω given in (6.8), rewrite Ω as follows,

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \theta_{11} & 0 \\ \theta_{12} & \theta_{22} \end{pmatrix} \begin{pmatrix} \theta_{11} & \theta_{12} \\ 0 & \theta_{22} \end{pmatrix} = U'U, \tag{A.1}$$

with U "upper-triangular". For the identity $\Omega = U'U$ to hold we see that,

$$\sigma_{11} = \theta_{11}^2, \quad \sigma_{12} = \theta_{11}\theta_{12}, \quad \text{and} \quad \sigma_{22} = \theta_{22}^2 + \theta_{12}^2,$$

by direct calculation. ♦

In general, a Cholesky decomposition for Ω can be written as,

$$\Omega = U(\theta)' U(\theta),$$

or simply, $\Omega = U'U$, where U is a $(p \times p)$ *upper* triangular matrix of the form,

$$U = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1p} \\ 0 & \theta_{22} & & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \theta_{pp} \end{pmatrix},$$

with again $p(p+1)/2$ elements $\theta_{ij}, \theta_{ij} \in \mathbb{R}$. We may use the introduced vech $(\cdot)$ to write θ as a vector in a simple way,

$$\theta = \text{vech}(U'), \quad \theta \in \mathbb{R}^n \text{ and } n = p(p+1)/2.$$

Note that by setting $L = U'$, $\Omega = LL'$, that is a Cholesky decomposition in terms of *lower*-triangular L. Both the U and the L versions are used in the literature, and they are of course identical.

Some fundamental issues with the Cholesky decomposition of a covariance matrix are worth emphasizing here: First of all, while the entries in Ω have direct interpretations (as variances and covariances respectively) there is no such interpretation of the entries

of the Cholesky factor U (apart from θ_{11}, see the bivariate example). The log-likelihood function in terms of the Cholesky decomposition and hence θ, is given by

$$L_{\text{chol}}(\theta) = L(\Omega(\theta))$$

where $L(\Omega)$ is given by (6.4). Thus it follows by simple insertion that

$$\begin{aligned} L_{\text{chol}}(\theta) = L(\Omega(\theta)) &= -\frac{T}{2}\log\left|U(\theta)'U(\theta)\right| - \frac{1}{2}\sum_{t=1}^{T}\text{tr}\left(\left(U(\theta)'U(\theta)\right)^{-1}X_tX_t'\right) \quad \text{(A.2)} \\ &= -\frac{T}{2}\left(\log\left|U(\theta)'U(\theta)\right| + \text{tr}\left(\left(U(\theta)'U(\theta)\right)^{-1}S_{xx}\right)\right) \end{aligned}$$

Maximization of $L_{\text{chol}}(\theta)$ over $\theta \in \mathbb{R}^n$ is usually working well, but as mentioned the estimates in $\hat{U} = U(\hat{\theta})$ do not have a direct interpretation. However, one can use the identity $\Omega = U'U$, to see that

$$\hat{\Omega} = \Omega(\hat{\theta}) = \hat{U}'\hat{U}.$$

In practice, what remains in terms of estimation is the computation of standard errors for $\hat{\Omega}$ from (the simple to compute) standard errors of $\hat{U}$. As Ω is parametrized in terms of θ (or U), the standard errors can be derived by applying a Taylor expansion of $\Omega(\theta)$ in terms of θ, sometimes referred to as the "delta-method".

A second issue is that the Cholesky decomposition is not unique. Specifically, in terms of the example, observe that for example the signs of the first row of U are not unique. This can be seen directly in the bivariate example, where

$$\Omega = U'U = U_-'U_- \text{ where } U = \begin{pmatrix} \theta_{11} & \theta_{12} \\ 0 & \theta_{22} \end{pmatrix} \text{ and } U_- = \begin{pmatrix} -\theta_{11} & -\theta_{12} \\ 0 & \theta_{22} \end{pmatrix}.$$

In short, if U is an upper Cholesky factor of Ω, then so is any other obtained by multiplying any of the rows in U by (-1). This on the other hand means that we have 2^p different parametrizations in terms of θ_{ij} which represent the same Ω, and hence the non-uniqueness leads to lack of identification, which is often solved by assuming that all diagonal elements of U are positive. To summarize, a unique Cholesky decomposition is given by,

$$\Omega(\theta) = U(\theta)'U(\theta), \quad U(\theta) = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1p} \\ 0 & \theta_{22} & & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \theta_{pp} \end{pmatrix} \quad \text{if} \quad \theta_{ii} > 0, \qquad i = 1, ..., p. \tag{A.3}$$

In terms of implementation, Cholesky decompositions of a positive definite matrix can be computed in most software. With this choice one can find $\hat{U} = U(\hat{\theta})$, by maximizing the likelihood function in (A.2) as a function of θ,

$$\hat{\theta} = \text{argmax}_{\theta=(\theta_{ij})_{i,j=1,...,p},\theta_{ii}>0} L_{\text{chol}}(\theta),$$

with $U = U(\theta)$, or $\theta = \text{vech}(U')$, given by (A.3). Note that, similar to the previous discussions, the constraint that $\theta_{ii} > 0$ may be implemented explicitly by reparametrizing θ_{ii} in terms of $\gamma_{ii} \in \mathbb{R}$:

$$\theta_{ii} = \exp(\gamma_{ii}) \quad \text{or} \quad \theta_{ii} = \gamma_{ii}^2, \quad \gamma_{ii} \in \mathbb{R}.$$

This way the optimization can be performed without any constraints; we have indeed as desired found a reparametrization of Ω in terms of θ, or $U(\theta)$, which is implementable in practice although it may still have numerical instabilities depending on the data at hand.

Remark A.1. *Existing literature on Cholesky-decomposition based MGARCH models include Dellaportas and Pourahmadi (2012).*

A.2 General parametrization of $V(\theta)$

For the general case the $(p \times p)$ dimensional orthonormal matrix V can defined in terms rotation matrices $R(i,j) = (R(i,j)_{kl})_{k,l=1,\ldots,p}$, see Boswijk and van der Weide (2011). That is, $V = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} R(i,j)$, where

$$R(i,j)_{kk} = 1 \text{ if } k \neq i,j, \qquad R(i,j)_{kl} = 0 \text{ if } k \neq l \text{ and } k \neq i,j,$$

$$R(i,j)_{ii} = R(i,j)_{jj} = \cos(\phi_{ij}), \text{ and } R(i,j)_{ij} = -R(i,j)_{ji} = \sin(\phi_{ij}).$$

B Principal Components

For later use we finally state the relationship between the discussed eigenvalue problem and so-called *principal components*. First note that by solving the eigenvalue problem from above,

$$|\lambda I_p - \Omega| = 0,$$

one has $V'\Omega V = \Lambda$ and V orthonormal. This can be used for decomposing both Ω and the variables X_t. To see this, observe that as $V'\Omega V = \Lambda$,

$$v_i'\Omega v_j = 0 \text{ for } i \neq j,$$

which is used to define the principal components of X_t. With X_t i.i.d. $\mathrm{N}_p(0,\Omega)$, define as before (the rotated) $\tilde{X}_t = V'X_t$, or $\tilde{X}_{it} = v_i'X_t$. It follows that

$$\mathrm{V}(\tilde{X}_{it}) = v_i'\Omega v_i = \lambda_i, \quad \text{while} \quad \text{cov}(\tilde{X}_{it}, \tilde{X}_{jt}) = v_i'\Omega v_j = 0.$$

Thus the *principal components* $\tilde{X}_{it}$ are independent (uncorrelated if normality is violated), and by ordering the eigenvalues of Ω,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0,$$

then the first component, $\tilde{X}_{1t} = v_1'X_t$, will be the linear combination of X_t which has highest variance and is independent (uncorrelated) with the others, $\tilde{X}_{jt}$, $j \neq 1$. Likewise,

$\tilde{X}_{2t}$ is the linear combination of X_t which has second highest variance and is independent (uncorrelated) with the other components. And so forth for $\tilde{X}_{kt}$, $k > 2$.

One may rewrite X_t in terms of the principal components $\tilde{X}_{it}$ as follows,

$$X_t = VV'X_t = \sum_{i=1}^{p} v_i \tilde{X}_{it}.$$

And similarly, the covariance can be decomposed as follows,

$$\Omega = V\Lambda V' = \lambda_1 v_1 v_1' + ... + \lambda_p v_p v_p',$$

that is, rewrite Ω in terms of the eigenvalues λ_i and the corresponding *factor loadings* v_i.

Remark B.1. *In practice the following considerations for principal components are often used. Suppose that the last $(p-k)$ of the eigenvalues λ_i are "so small" that,*

$$\Omega = \sum_{i=1}^{p} \lambda_i v_i v_i' \simeq \sum_{i=1}^{k} \lambda_i v_i v_i', \quad k < p.$$

That is, the covariance of X_t is essentially described by the first k principal components, $\tilde{X}_{it} = v_i' X_t$.

Remark B.2. *One can show that $(\mathrm{tr}\,(\Omega - Q))^2$, with Q a $p \times p$ dimensional symmetric matrix of rank $k \leq p$ (and positive semi-definite), is minimized by $Q = \sum_{i=1}^{k} \lambda_i v_i v_i'$, see Magnus and Neudecker (1988, Theorem 3, p.355).*

As to estimation of the v_i and λ_i, one may proceed as above and parametrize as

$$\Omega(\theta) = V(\theta)\,\Lambda(\theta)\,V(\theta)',$$

and then set $\hat{v}_i$ equal to the ith vector in $V(\hat{\theta})$ corresponding to the ordered eigenvalues $\hat{\lambda}_i$. However, a much used alternative in the principal components literature is based on using the sample covariance as an estimator of Ω, $\hat{\Omega} = S_{xx}$, and then back out $\hat{v}_i$ and $\hat{\lambda}_i$ (which indeed gives the MLE provided the eigenvalues of Ω are distinct).

C Proofs of Lemmas

C.1 Proof of Lemma 6.1:

With $L(\Omega)$ given by (6.4), $f(\Omega) = -2L(\Omega)/T$ can be written as

$$\begin{aligned} f(\Omega) &= \log|\Omega| + \mathrm{tr}\left(\Omega^{-1}S_{xx}\right) = \log|S_{xx}| + \mathrm{tr}\left(\Omega^{-1}S_{xx}\right) - \log\left|\Omega^{-1}S_{xx}\right| \\ &= \log|S_{xx}| + \mathrm{tr}\left(\Omega^{-1/2}S_{xx}\Omega^{-1/2}\right) - \log\left|\Omega^{-1/2}S_{xx}\Omega^{-1/2}\right| \end{aligned}$$

With $A = \Omega^{-1/2}S_{xx}\Omega^{-1/2}$ consider the eigenvalue problem,

$$|\lambda I - A| = 0,$$

with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p > 0$ and corresponding eigenvectors v_i. With $V = (v_1, ..., v_p)$ and $\Lambda = \text{diag}(\lambda_1, ..., \lambda_p)$, by definition

$$V'V = VV' = I_p \text{ and } V'AV = \Lambda.$$

Using this rewrite the terms in $f(\Omega)$ as follows:

$$\log\left|\Omega^{-1}S_{xx}\right| = \log|A| = \log|V\Lambda V'| = \log|\Lambda| = \sum_{i=1}^{p}\log\lambda_i.$$

Moreover,

$$\text{tr}\left(\Omega^{-1}S_{xx}\right) = \text{tr}(A) = \text{tr}(V\Lambda V') = \text{tr}(\Lambda) = \sum_{i=1}^{p}\lambda_i.$$

Collecting terms, minimizing $f(\Omega)$ over $\Omega > 0$, is equivalent to minimizing,

$$\sum_{i=1}^{p}(\lambda_i - \log\lambda_i)$$

over $\lambda_i > 0$. This is minimized by $\lambda_i = 1$ and hence by $A = V\Lambda V' = I_p$. That is, $A = \Omega^{-1/2}S_{xx}\Omega^{-1/2} = I_p$, or

$$\hat{\Omega} = S_{xx}.$$

□

C.2 Proof of Lemma 8.1:

By Tjøstheim (1990) a drift function $\delta(\cdot)$ can be applied to the $m-$step process, X_{mt}, rather than X_t itself, where X_{mt} is obtained by simple recursion from (8.2) as,

$$X_{mt} = \Phi_{(mt,m)}X_{m(t-1)} + \sum_{j=0}^{m-1}\Phi_{(mt,j)}\epsilon_{tm-j},$$

using the notation $\Phi_{(m,j)} := \Phi_m\Phi_{m-1}\cdots\Phi_{m-(j-1)}$ for $j \geq 1$, while $\Phi_{(m,0)} := I_p$. With drift function, $\delta(X) = 1 + \|X\|^{\varepsilon}$, where $\varepsilon > 0$ is chosen appropriately below, it follows one can choose $\varepsilon \in (0,1)$ and m such that

$$\text{E}\left\|\Phi_{(mt,m)}\right\|^{\delta} < 1.$$

Hence, with $c > 0$ a constant,

$$\text{E}\left(\delta(X_{mt}) \mid X_{m(t-1)} = X\right) \leq \text{E}\left\|\Phi_{(mt,m)}\right\|^{\kappa}\|X\|^{\kappa} + c.$$

It follows, with M some (large) constant, and for $\delta(X) > M$, we have

$$\text{E}\left\|\Phi_{(mt,m)}\right\|^{\kappa}\|X\|^{\kappa} + c = \left[\frac{\text{E}\left\|\Phi_{(mt,m)}\right\|^{\kappa}\|X\|^{\kappa}+c}{1+\|X\|^{\kappa}}\right]\delta(X) < \beta\delta(X),$$

where β is some constant with $|\beta| < 1$. Thus X_t is stationary and geometrically ergodic, provided

$$\varrho := \lim_{m \to \infty} \left[\tfrac{1}{m} \operatorname{E} \log \left\| \prod_{t=1}^{m} \Phi_t \right\| \right] < 0.$$

Next, the result for $\operatorname{E} ||X_t||^2 < \infty$, follows by considering the drift function, $\delta_2 (X) = 1 + ||X||^2$. □

References

Bauwens, L., S. Laurent, and J. V. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21(1), 79–109.

Boswijk, H., and R. van der Weide (2011): "Method of moments estimation of GO-GARCH models," *Journal of Econometrics*, 163, 118–126.

Boussama, F., F. Fuchs, and R. Stelzer (2011): "Stationarity and geometric ergodicity of BEKK multivariate GARCH models," *Stochastic Processes and their Applications*, 121, 2331–2360.

Chopra, V. K., and W. T. Ziemba (1993): "The effect of errors in means, variances, and covariances on optimal portfolio Choice," *Journal of Portfolio Management*, 19(2), 1–6.

Comte, F., and O. Lieberman (2003): "Asymptotic theory for multivariate GARCH processes," *Journal of Multivariate Analysis*, 84(1), 61–84.

Creal, D., S. J. Koopman, and A. Lucas (2011): "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," *Journal of Business & Economic Statistics*, 29(4), 552–563.

Cvitanic, J., and F. Zapatero (2004): *Introduction to the Economics and Mathematics of Financial Markets*. The MIT Press.

Dellaportas, P., and M. Pourahmadi (2012): "Cholesky-GARCH models with applications to finance," *Statistics and Computing*, 22, 849–855.

Elton, E. J., M. J. Gruber, S. J. Brown, and W. Goetzmann (2003): *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, sixth edn.

Engle, R. (2002): "Dynamic Conditional Correlation," *Journal of Business and Economic Statistics*, 20(3), 339–350.

Engle, R. F., and K. F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11(01), 122–150.

Francq, C., and J.-M. Zakoïan (2012): "QML Estimation of a Class of Multivariate Asymmetric GARCH Models," *Econometric Theory*, 28, 179–206.

Hayashi, F. (2000): *Econometrics*. Princeton University Press, Princeton.

Henderson, H. V., and S. R. Searle (1979): "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics," *The Canadian Journal of Statistics*, 7(1), 65–81.

Hetland, S., R. S. Pedersen, and A. Rahbek (2019): "Dynamic Conditional Eigenvalue GARCH," *Working paper*.

Jensen, S. T., and A. Rahbek (2004a): "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20(6), 1203–1226.

——— (2004b): "Asymptotic normality the QMLE estimator of arch in the nonstationary case," *Econometrica*, 72(2), 641–646.

Kroll, Y., H. Levy, and H. M. Markowitz (1984): "Mean-Variance versus Direct Utility Maximization," *Journal of Finance*, 39(1), 47–61.

Lintner, J. (1965): "Security Prices, Risk, and Maximal Gains From Diversification," *The Journal of Finance*, 20(4), 587–615.

Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Markowitz, H. M. (1952): "Portfolio Selection," *The Journal of Finance*, 7(1), 77–91.

Mossin, J. (1966): "Equilibrium in a Capital Asset Market," *Econometrica*, 34(4), 768–783.

Nielsen, H. B., and A. Rahbek (2014): "Unit Root Vector Autoregression with Volatility Induced Stationarity," *Journal of Empirical Finance*, 29, 144–167.

Noureldin, D., N. Shephard, and K. Sheppard (2014): "Multivariate rotated ARCH models," *Journal of Econometrics*, 179(1), 16–30.

Pedersen, R. (2017): "Inference and testing on the boundary in extended constant conditional correlation GARCH models," *Journal of Econometrics*, 196, 23–36.

Pedersen, R., and A. Rahbek (2014): "Multivariate variance targeting in the BEKK-GARCH model," *The Econometrics Journal*, 17, 24–55.

Pulley, L. B. (1983): "Mean-Variance Approximations to Expected Logarithmic Utility," *Operations Research*, 31(4), 685–696.

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *The Journal of Finance*, 19(3), 425–442.

Tjøstheim, D. (1990): "Non-Linear Time Series and Markov Chains," *Advances in Applied Probability*, 22, 587–611.