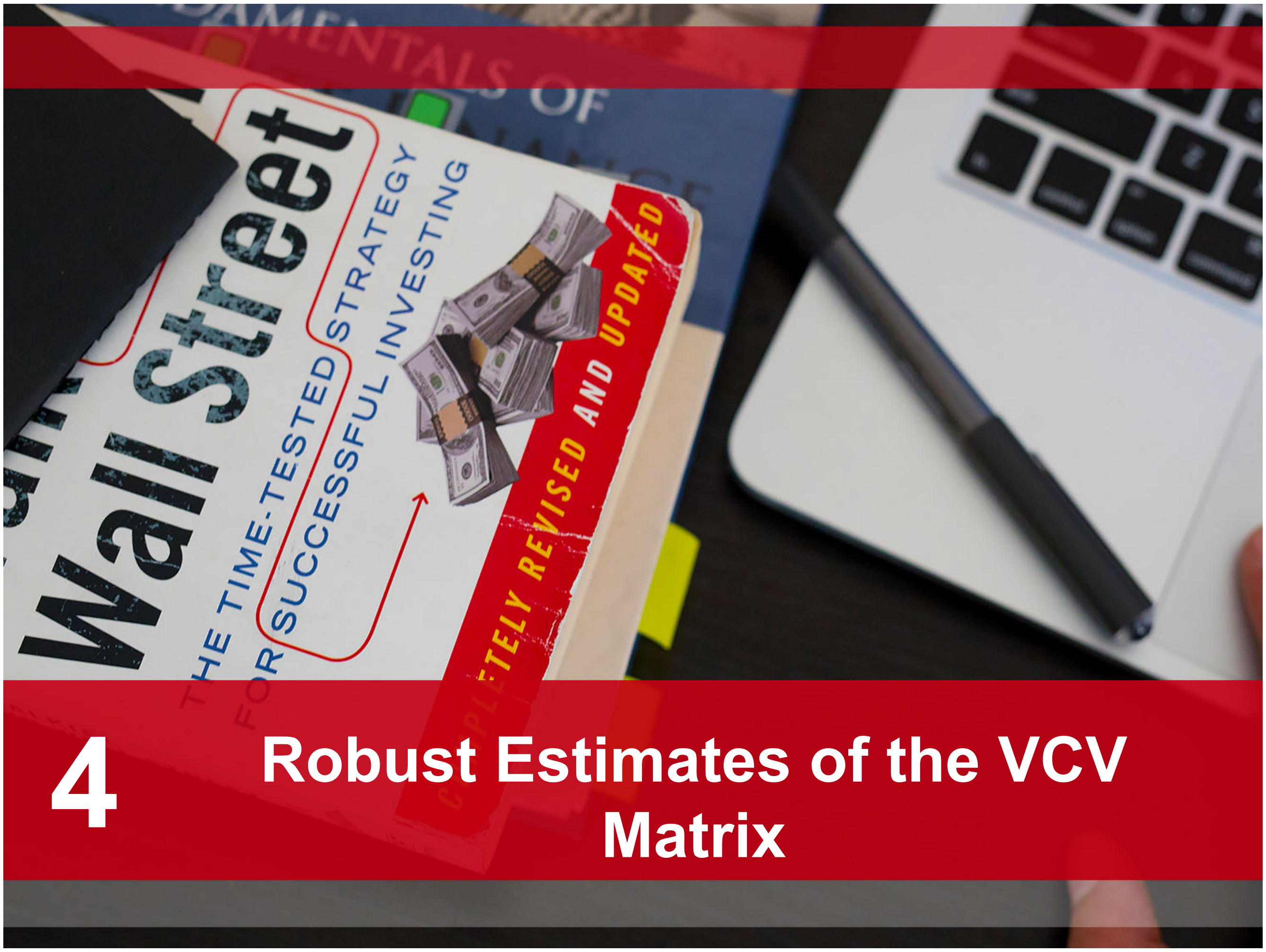
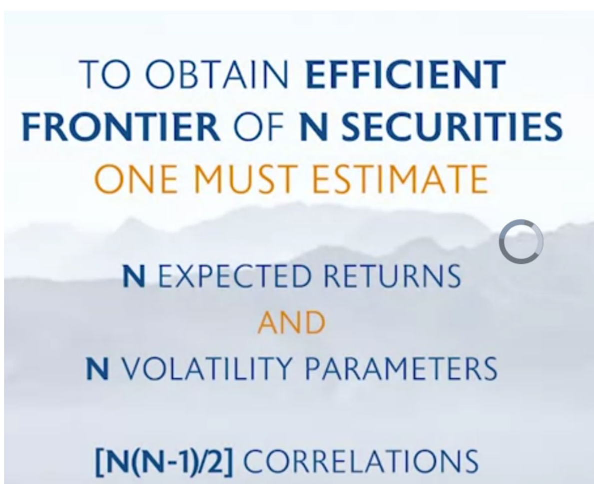


4 Robust Estimates of the VCV Matrix

The Curse of Dimensionality

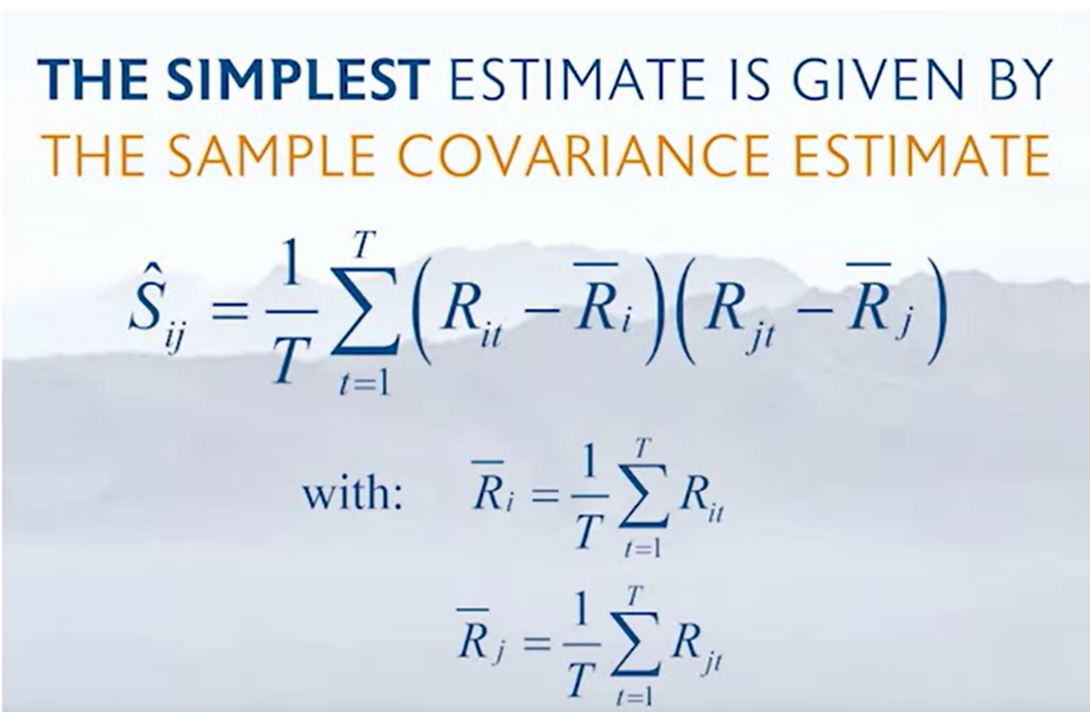


How many parameters do we need to estimate for a portfolio of 100 stocks?

Challenges in Estimating the VCV Matrix

- **Increased Sample Size Requirement**:
  - To reliably estimate the covariance between variables, the number of observations must increase exponentially with the number of variables.
- **Estimation Error**:
  - As dimensionality increases, sample covariance matrices become less stable, leading to larger estimation errors.

Extreme Example 1:
No model risk, high sample risk



Extreme Example 2:
High model risk, no sample risk

CONSTANT CORRELATION MODEL

$$\begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2N}\sigma_2\sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1}\sigma_N\sigma_1 & \rho_{N2}\sigma_N\sigma_2 & \cdots & \sigma_N^2 \end{pmatrix} \rightarrow \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_N \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_N\sigma_1 & \rho\sigma_N\sigma_2 & \cdots & \sigma_N^2 \end{pmatrix}$$

$$\hat{\sigma}_{ij}^{CC} = \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}$$

Extreme Example 2:
High model risk, no sample risk

CUT THE NUMBER N(N-1)/2 OF CORRELATION PARAMETERS **DOWN TO 1**

THE **OPTIMAL** ESTIMATOR OF THIS CONSTANT CORRELATION IS THE “GLOBAL” AVERAGE

$$\hat{\rho} = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \hat{\rho}_{ij}$$

The Curse of Dimensionality

In the presence of large portfolios: The number of parameters is often larger than the sample size.

Increasing frequency: Increasing frequency is necessary but not always sufficient.

Introducing structure: Introducing structure helps deal with sample risk, but this comes at the cost of model risk.

Factor – Based VCV Estimate

ASSUME **STOCK RETURNS** ARE DRIVEN BY A **LIMITED SET OF FACTORS**

$$R_{it} = \mu_i + \beta_{i1} F_{1t} + \ldots + \beta_{ik} F_{kt} + \ldots + \beta_{iK} F_{Kt} + \varepsilon_{it}$$

where β_{ik} is the sensitivity of asset i with respect to factor k $(k = 1, \ldots, K)$

Factor – Based VCV Estimate

- With two factor model:

VARIANCE

$$\sigma_i^2 = \beta_{i1}^2\sigma_{F1}^2 + \beta_{i2}^2\sigma_{F2}^2 + 2\beta_{i1}\beta_{i2}\mathrm{Cov}(F_1,F_2) + \sigma_{\varepsilon i}^2$$

COVARIANCE

$$\sigma_{ij} = \beta_{i1}\beta_{j1}\sigma_{F1}^2 + \beta_{i2}\beta_{j2}\sigma_{F2}^2 + (\beta_{i1}\beta_{j2} + \beta_{i2}\beta_{j1})\mathrm{Cov}(F_1,F_2) + \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt})$$

INTRODUCE STRUCTURE BY IMPOSING THAT RESIDUALS ARE UNCORRELATED

$$\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$$

Factor – Based VCV Estimate

GENERAL DECOMPOSITION OF RETURNS

$$\mathrm{cov}\left(R_i(t), R_j(t)\right) = \sum_{k=1}^{K} \beta_{ik}\beta_{jk}\sigma_{F_k}^2 + \mathrm{cov}\left(\varepsilon_i(t), \varepsilon_j(t)\right)$$

ASSUME UNCORRELATED **RESIDUALS**

$$\begin{cases} \sigma_{ij} = \mathrm{cov}\left(R_i(t), R_j(t)\right) = \sum_{k=1}^{K} \beta_{ik}\beta_{jk}\sigma_{F_k}^2 \quad \mathit{for}\ i \neq j \\ \sigma_{ii} = \sigma_i^2 = \mathrm{cov}\left(R_i(t), R_i(t)\right) = \sum_{k=1}^{K} \beta_{ik}^2\sigma_{F_k}^2 + \sigma_{\varepsilon_i}^2 \quad \mathit{for}\ i = j \end{cases}$$

Factor – Based VCV Estimate

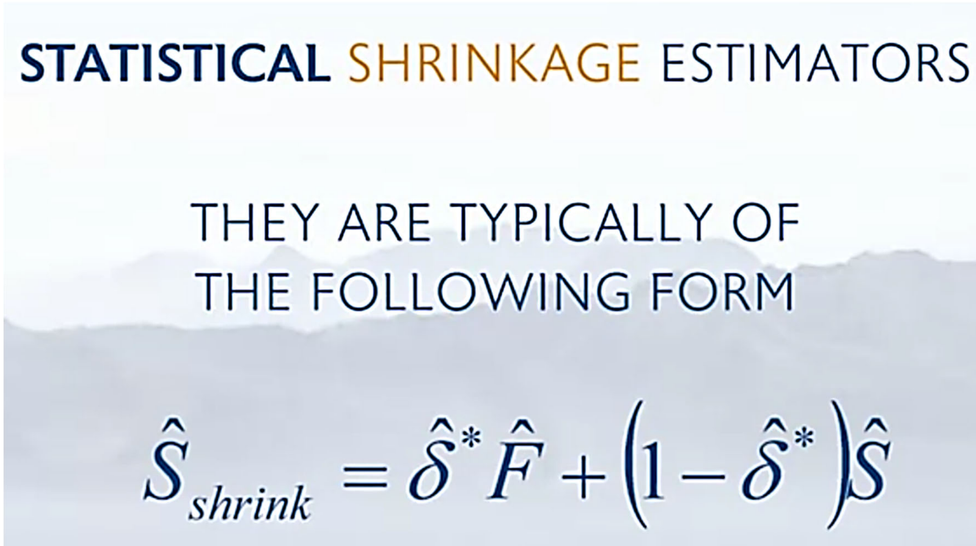
- **Using a factor model is a convenient way** to reduce the number of risk parameters to estimate while introducing a reasonable amount of model risk.
- **An implicit factor model is often preferred** since it lets the data tell us what the relevant factors are, thus alleviating model risk.
  - PCA analysis

Honey I Shrunk the Covariance Matrix! Shrinkage Approach

- Trade-off between Sample risk and Model risk

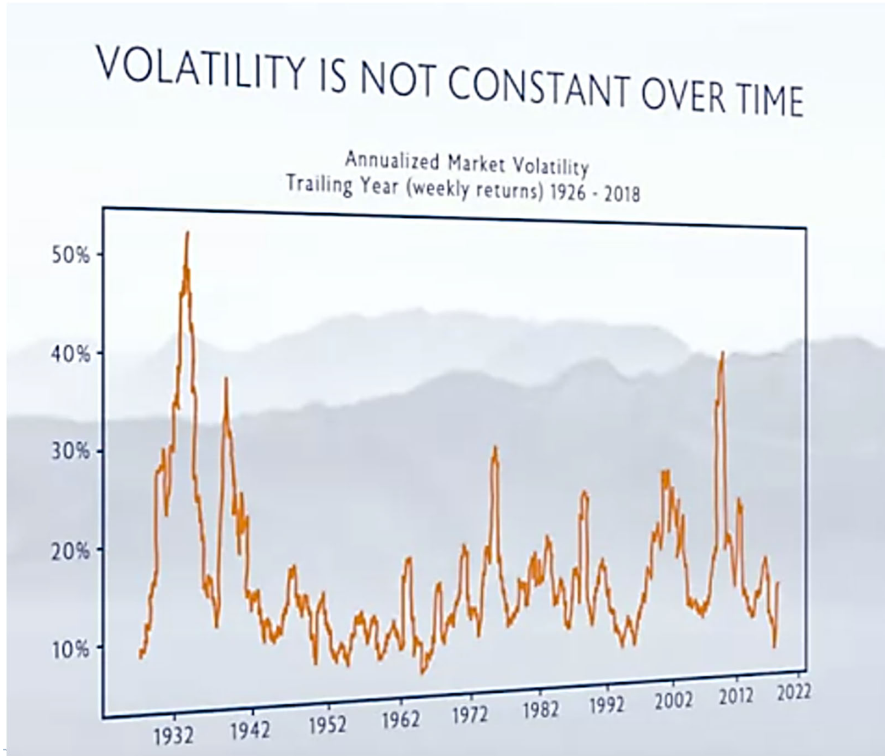
- **Shrinkage**: Do not choose between sample risk and model risk but instead mixing sample risk and model risk.

Honey I Shrunk the Covariance Matrix! Shrinkage Approach



Honey I Shrunk the Covariance Matrix! Shrinkage Approach

Portfolio Construction with Time-Varying Risk Parameters



Rolling Windows

- The 'historical' covariance matrix is calculated on a T-day window that is rolled through time, each day adding the new return and taking off the oldest return:

$$\mathbf{H}_t = \frac{1}{T}\sum_{i=1}^{T} \mathbf{r}_{t-i}\mathbf{r}_{t-i}^{'}.$$

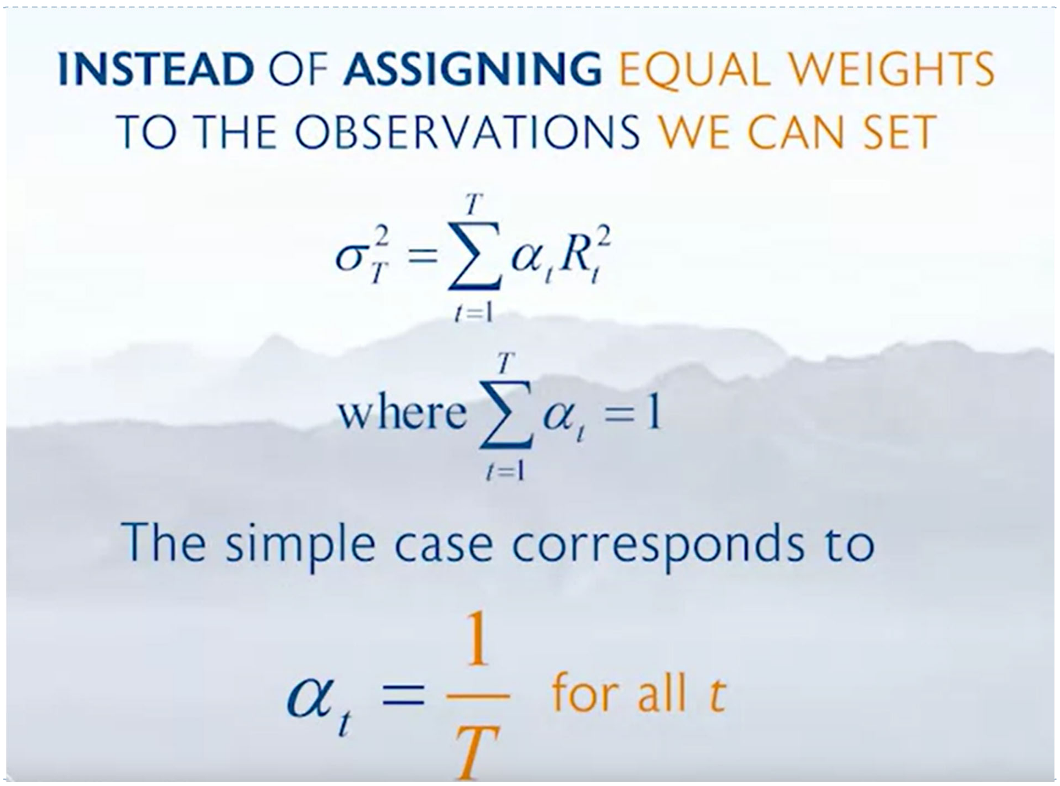
- The sophistication of this model lies in the choice of the window length T. If the length is short, the estimate may be noisy. The longer the window, the less noisy the estimate, but the more biased it is when far more distant observations, which may not be relevant today, are included in the calculation. Hence, the length of the window T directly determines the trade-off between the sampling error and the unbiasedness of the estimate.

The Curse of Non-Stationarity

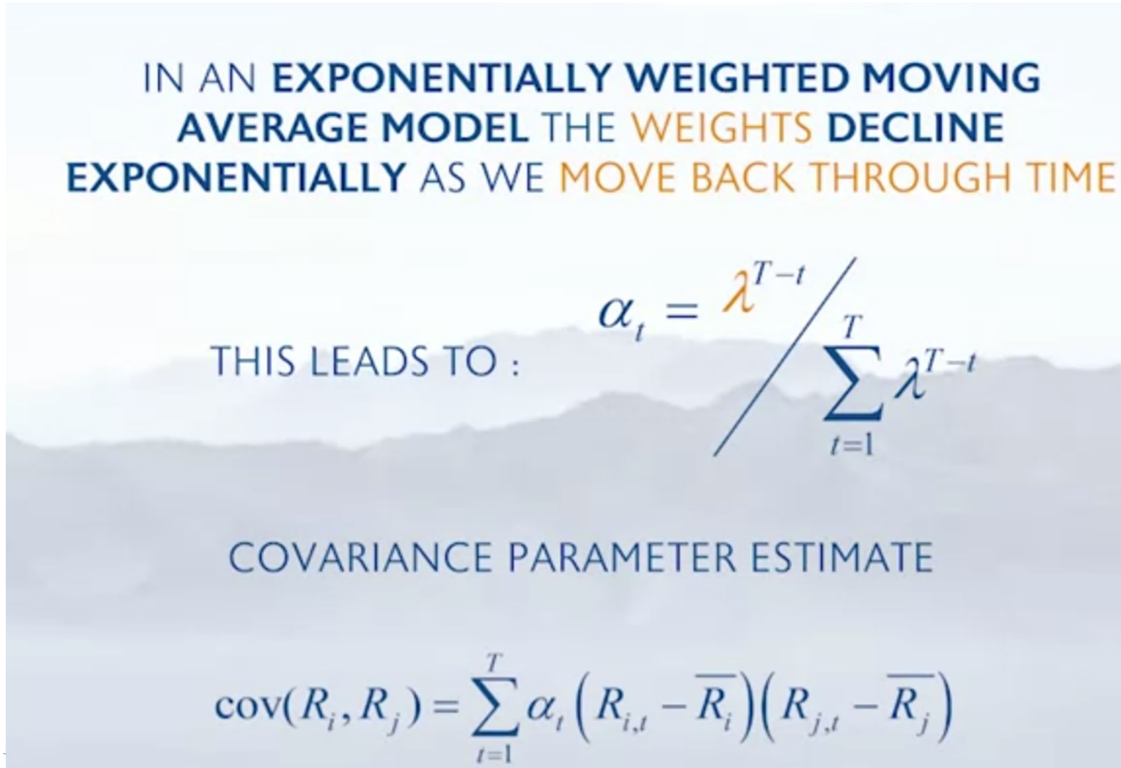
Increasing frequency is better
than increasing sample period in case of non-stationary return distributions.

In this context, using **rolling windows** is better than **expanding windows.**

The EWMA Model



The EWMA Model



The EWMA model

- The EWMA covariance matrix has the following specification:

$$\mathbf{H}_t = (1-\lambda)\mathbf{r}_{t-1}\mathbf{r}_{t-1}' + \lambda\mathbf{H}_{t-1}, \tag{3.3}$$

where λ is the decay factor $(0 < \lambda < 1)$.

The first term of the right hand side of (3.3), $(1-\lambda)\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}'$, denotes the response of volatility to one-period news, while the second term, $\lambda\mathbf{H}_{t-1}$, determines the persistence in volatility. The higher the value of λ, the more persistent the process and the slower the response to new shocks. However, in the EWMA model, the reaction and persistence parameters are not independent because they sum to one.

By backward substitution, the covariance matrix can be written as:

$$\mathbf{H}_t = (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}\mathbf{r}_{t-i}\mathbf{r}_{t-i}^{'}. \tag{3.4}$$

The model derives its name from this formulation, in which the elements of the covariance matrix are the exponentially weighted moving averages of past squares and cross products of returns. In practice, the process is often estimated with a cut-off time T, scaling the infinite sum in (3.4) by

$$\mathbf{H}_t = \frac{(1-\lambda)}{1-\lambda^T}\sum_{i=1}^{T}\lambda^{i}\mathbf{r}_{t-i}\mathbf{r}_{t-i}^{'}. \tag{3.5}$$

Under the RiskMetrics (1994), λ takes the values of 0.94 and 0.97 for daily and weekly forecasts, respectively. The EWMA process is hence easily estimated in a spreadsheet for any dimensional system.

By recursive substitution, the h-step forecast is equal to the one-step ahead forecast:

$$\mathbf{H}_{t+h} = \mathbf{H}_{t+h-1} = \ldots = \mathbf{H}_{t+2} = \mathbf{H}_{t+1}. \tag{3.7}$$

Assuming returns are serially uncorrelated, the expected covariance matrix over k cumulative steps is given by $\mathbf{H}_{t+1:t+k} = k \times \mathbf{H}_{t+1}$. The multiple-period forecast is a simple product of the one-day forecast with the forecast horizon, k. This is also known as the 'square root of time' rule for volatility forecasts. The EWMA model can thus be thought of as a random walk model, where a shock will have a permanent effect on the expectation of future variance and covariance. The volatility process in the EWMA model is not mean-reverting, which is quite counterfactual since financial return volatility tends to eventually converge to its long-run average.

The ARCH and GARCH model

- Observing that squared residuals are often autocorrelated even though residuals themselves are not, Engle (1982) sets the stage for the new class of time-varying conditional volatility models with the Autoregressive Conditional Heteroskedasticity (ARCH) model. The new model has inspired a huge amount of related research on its development, generalisation and application, and deserved Engle a Nobel Prize in Economics in 2003

The ARCH model

IN AN **ARCH(T) MODEL** WE ALSO ASSIGN SOME WEIGHT TO THE LONG-RUN VARIANCE V_L

$$\sigma_T^2 = \gamma V_L + \sum_{t=1}^{T} \alpha_t R_t^2$$

$$\text{where } \gamma + \sum_{t=1}^{T} \alpha_t = 1$$

ARCH(1) MODEL

$$\sigma_T^2 = \gamma V_L + \alpha R_T^2$$

$$\text{where } \gamma + \alpha = 1$$

α_i

The ARCH model

- The ARCH model is thus able to capture the volatility clustering observed in asset returns. One advantage of the ARCH model is that the weight α can be estimated from historical data, based on, e.g., the Maximum Likelihood procedure, even though the 'true' volatility is never observed.
- The ARCH model captures the two most common features of real high frequency financial asset returns, i.e., volatility clustering and heavy-tailed unconditional distributions.

The GARCH model

- In the ARCH*(p)* model, past shocks of more than *p* periods ago have no effect on the current volatility, hence the order *p* determines how long a shock is persistent to volatility. For financial time series, it typically requires a very high order *p* to capture the dependence. Bollerslev (1986) proposes a parsimonious way to handle with this problem, introducing the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) model.

The GARCH model

IN **GARCH (1,1)** WE ADDITIONALLY ASSIGN SOME WEIGHT TO THE PREVIOUS VARIANCE ESTIMATE TO **CAPTURE VOLATILITY CLUSTERING**

$$\sigma_T^2 = \gamma V_L + \alpha R_T^2 + \beta \sigma_{T-1}^2$$

with $\gamma + \alpha + \beta = 1$

The GARCH model

GARCH(p,q)

$$\sigma_T^2 = \omega + \sum_{i=1}^{p} \alpha_i R_{T-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{T-j}^2$$

$$\omega = \gamma V_L$$

λ

The GARCH model

- It obviously follows that the GARCH model is also an exponentially weighted moving average process. However, there are two major differences between the GARCH and the EWMA models. First, while the parameter λ of the EWMA process is often set *ad hoc*, the parameters of the GARCH process have to be estimated by rigorously statistical methods, normally using the Maximum Likelihood procedure. Second, the GARCH model allows the volatility process to eventually revert to its long-run level.

The Multivariate GARCH model

- Orthogonal GARCH
- The DCC model

LAB Section:

- Use the file:

  lab_22_ShrinkageVCV.ipynb