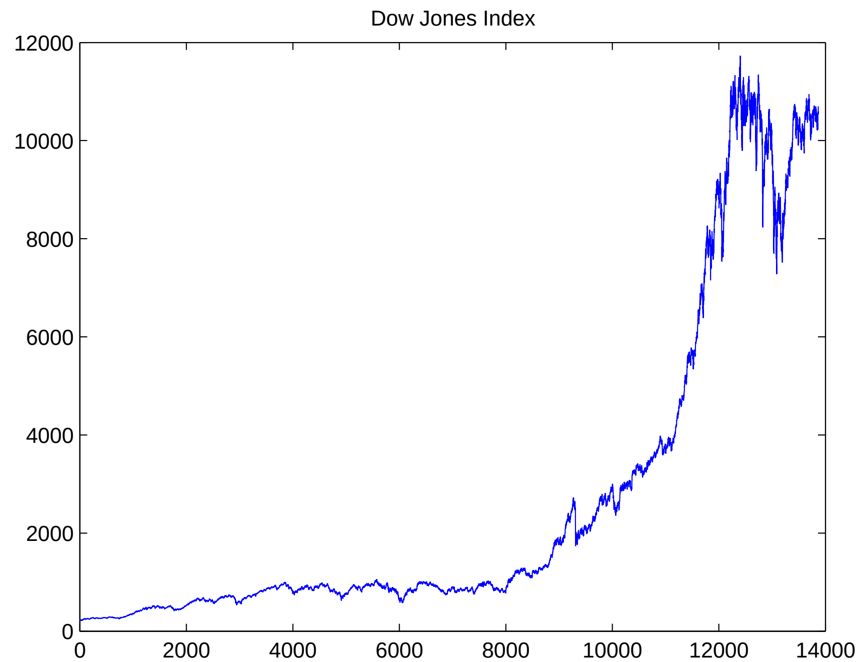
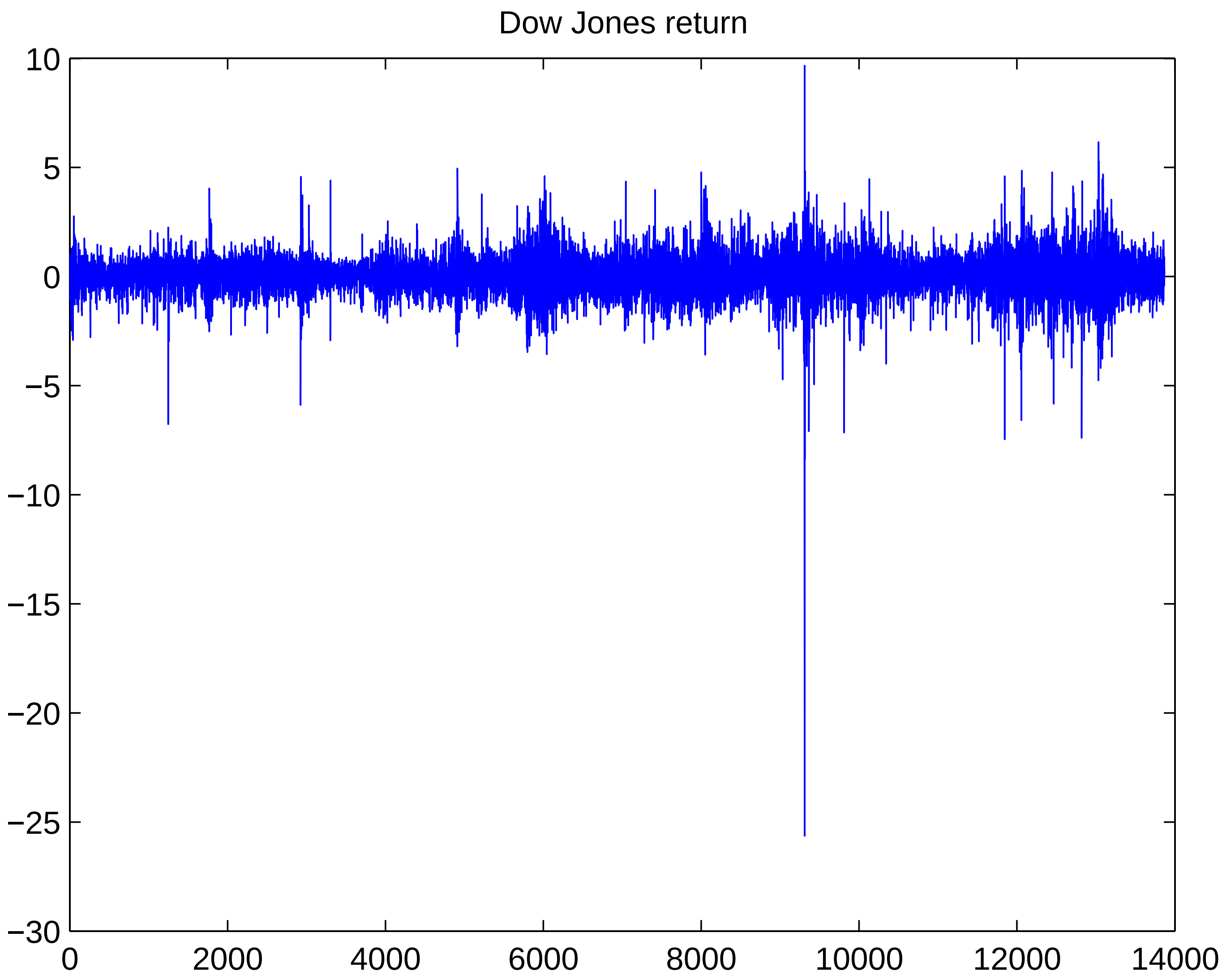
16 Estimating volatilities, correlations and VaR

16.1 Volatilities

- Volatility of financial returns is usually not constant.




Dow Jones return
10
5
0
−5
−10
−15
−20
−25
−30
0
2000
4000
6000
8000
10000
12000
14000

- How can we estimate it?
- Consider the case where volatility is constant, $Var[y_t] = \sigma^2$.
- To estimate σ^2, we would use all the data (from $t = 1$ to $t = T$):

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} y_t^2$$

assuming that mean of y_t is zero (or we subtracted the sample mean).

- Next, what if it is σ_t^2 instead of σ^2?

16.1.1 Rolling window

- To capture a variance that changes over time, we can use a rolling window.
- Instead of using all the data we use only the most recent m observations:

$$\hat{\sigma}_t^2 = \frac{1}{m}\sum_{i=1}^{m} y_{t-i}^2$$

- The variance is updated by dropping past returns and adding recent returns.

16.1.2 Weighting schemes

- We want to estimate volatility **today**
- If this volatility is changing over time, does it make sense to put equal weights on all past squared returns as in the rolling window method?
- Why not put bigger weights on recent observations:

$$\hat{\sigma}_t^2 = \sum_{i=1}^{m} \alpha_i y_t^2$$

with $\alpha_i \geq 0$, $\alpha_j < \alpha_i$ if $j > i$ and $\sum_{i=1}^{m} \alpha_i = 1$

16.1.3 Exponential weights

- The exponentially weighted moving average is a particular weighting scheme where we take the weights decreasing exponentially, $\alpha_{i+1} = \lambda \alpha_i$ with λ between 0 and 1.
- Using this scheme, we can write the variance as

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1-\lambda) y_{t-1}^2$$

- To see that the weights decrease exponentially, replace recursively the σ^2 on the right-hand-side. We get this equation:

$$\hat{\sigma}_t^2 = (1-\lambda) \sum_{i=1}^{m} \lambda^{i-1} y_{t-i}^2 + \lambda^m \hat{\sigma}_{t-m}^2$$

- For large m, λ^m is so small that the term $\lambda^m \hat{\sigma}_{t-m}^2$ can be ignored.

- This approach is designed to track volatility. It puts a lot of weights on recent observations. For $\lambda = 0.95$,

$$\begin{aligned}\sum_{i=1}^{10}(1-\lambda)\lambda^{i-1} &= 0.4013\\ \sum_{i=1}^{25}(1-\lambda)\lambda^{i-1} &= 0.7226\\ \sum_{i=1}^{50}(1-\lambda)\lambda^{i-1} &= 0.9231\end{aligned}$$

- The RiskMetrics database (J.P. Morgan) uses $\lambda = 0.94$.

16.1.4 GARCH(1,1)

- What is (probably) the most popular volatility model is the GARCH(1,1) model
- The model is

$$\hat{\sigma}_t^2 = \omega + \alpha y_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2$$

with $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$.

- See the difference with the exponential scheme?
- The $(1, 1)$ is because there is one lagged y_t^2 and one lagged $\hat{\sigma}_t$ on the RHS.

16.1.5 Estimation by maximum likelihood

- There are unknown parameters in the exponential weights scheme (λ) and in the GARCH model (ω, α, β).
- The most common way to estimate them is by maximum likelihood.
- A density function is $f(y|\theta)$: probabilities for different values of y for a given value of the parameter θ.
- A likelihood is a function $f(\theta|y)$: probabilities for different values of the parameters θ for a given value of y.
- Maximum likelihood: given observations $y_1, y_2, \ldots, y_T$, which value of θ is the most likely (probable)? We denote this value of θ by $\hat{\theta}$.
- Numerical search (optimization)

- For example, suppose $y_t \sim N(0, \sigma_t^2)$, then the likelihood for T observations is

$$f(\theta|y_1, y_2, \ldots, y_T) \;=\; \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2\sigma_t^2} y_t^2\right)$$

  - If it's exponential weights, $\sigma_t^2 = \lambda\sigma_t^2 + (1-\lambda)y_{t-1}^2$ and $\theta = \lambda$.
  - If it's GARCH, $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$ and $\theta = \{\omega, \alpha, \beta\}$.

16.2 Correlations

- Just like variances, covariances are also time-varying
- To estimate them, there are schemes similar to the ones for volatilities
- An additional constraint is that the covariance matrix has to be Positive Semi-Definite (PSD)
- For the exponential weights for example, we would have to use the same λ for pair of assets:

$$\hat{\sigma}_{ij,t} = \lambda\hat{\sigma}_{ij,t-1} + (1-\lambda)y_{i,t-1}y_{j,t-1}$$

- Another possibility is to note that we can always write a covariance as the product of a correlation and two standard deviations.
- A standard deviation is the square root of a variance. We know how to estimate variances.
- Idea: estimate variance, create standardized returns $u_t = y_t/\sqrt{\sigma_t^2}$, estimate correlations using u_t.
- For example, if we assume constant correlations:

$$\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T} u_{i,t}u_{j,t}$$

16.3 Value-at-Risk

- We saw different measure of risk for assets:
  - For stocks, we have the variance, covariances (+the underlying density: Normal, Student-t, ...)
  - For options/derivatives, we have Δ, Γ, vega
- To report the riskiness of a portfolio (stocks + options), you would need all of the above (if not more). This might be too much information (too much is like not enough).
- The Value-at-Risk (VaR) is an attempt to provide a single number summarizing all the risk in a portfolio.

- Definition: VaR is the amount such that over the next N days there is a probability Y that I will lose more than this amount.
- For example, if for a 1 day horizon and a 1% probability, the VaR is $10M, then it means that the probability that my portfolio will lose more than $10M tomorrow is at most 1%.
- It's a measure of how bad things can get
- Widely used. For example, banks have to hold reserves 3-5 times their "10-day 1%" VaR.

- Graphically,

- The challenge is to come up with the distribution of the future gains/losses for a portfolio over the next N days.
- For stocks, you combine GARCH models for the variances with a model for the correlations and you pick a distribution (Normal, Student-t)
- For derivatives (value is a non-linear function of the underlying asset), you can use Taylor expansions to build a linear or quadratic approximation (think of Δ and Γ of an option). Can be messy.
- If you have bonds, then you need to write a model for the evolution of the term structure. You combine it for the duration to get a distribution over the change in the value of the bond. Even more messy.