Jon Danielsson
London School of Economics

Financial Risk Forecasting

Seminar Questions and Solutions

financialriskforecasting.com

2025

Version 4.0

Chapter 1

Financial data

1. Which of the following statements about price-weighted versus value-weighted indices is correct?

   a Price-weighted indices give more weight to companies with higher market capitalization

   b Value-weighted indices are less representative of overall market performance

   c A stock split will decrease a stock's weight in a price-weighted index but not affect its weight in a value-weighted index

   d Price-weighted indices are more commonly used than value-weighted indices

   Solution
   The correct answer is (c). In a price-weighted index, weights are based on stock prices. When a stock splits (e.g., 2-for-1), its price halves, so its weight in the index decreases proportionally. In a value-weighted index, weights are based on market capitalization (price $\times$ shares outstanding). A stock split halves the price but doubles the shares, leaving market cap unchanged, so the weight remains the same.

2. Explain the concept that "risk is latent" and why this creates challenges for risk forecasting models that rely on historical data.

Solution

Risk is latent means that true risk may not be visible or measurable in historical data, especially during calm periods. Markets can appear stable and predictable for extended periods, but underlying vulnerabilities remain hidden until triggered by specific events or conditions.

This creates several challenges for risk forecasting:

Historical data limitations: Past data may not capture all possible risk scenarios. Periods of low volatility can give false confidence about future stability.

Structural breaks: Market conditions can change fundamentally, making historical patterns irrelevant. Models calibrated on past data may fail when underlying relationships shift.

Extreme tail events: Rare but extreme events are by definition underrepresented in historical samples, yet they can have catastrophic impacts.

The implication is that risk models should incorporate scenario analysis, stress testing, and conservative buffers rather than relying solely on historical patterns.

3. A stock's price moves from \$100 to \$120 on day 1, then from \$120 to \$108 on day 2.

   i Calculate the simple returns for both days.

   ii Calculate the continuously compounded returns for both days.

   iii What is the total return over the two days using simple returns?

   iv What is the total return over the two days using continuously compounded returns?

   v Explain why continuously compounded returns demonstrate the "symmetry" property.

Solution

i Simple returns:

$$R_1 = \frac{120 - 100}{100} = 0.20 = 20\%$$
$$R_2 = \frac{108 - 120}{120} = -0.10 = -10\%$$

ii Continuously compounded returns:

$$y_1 = \ln\left(\frac{120}{100}\right) = \ln(1.2) = 0.1823 = 18.23\%$$
$$y_2 = \ln\left(\frac{108}{120}\right) = \ln(0.9) = -0.1054 = -10.54\%$$

iii Total simple return: $\frac{108-100}{100} = 0.08 = 8\%$

iv Total log return: $0.1823 + (-0.1054) = 0.0769 = 7.69\%$

v Log returns are symmetric because $\ln(1.2) = -\ln(1/1.2) = -\ln(0.8333)$. If we had equal magnitude moves (say +20% and -20%), the log returns would be exactly symmetric while simple returns would not result in zero total return.

4. Explain what volatility clustering is, describe how it can be tested statistically, and discuss its implications for risk management.

Solution
Volatility clustering refers to the tendency of large changes in asset prices to follow large changes, and small changes to follow small changes. In financial markets, this manifests as periods of high volatility being followed by periods of high volatility, while calm periods tend to be followed by calm periods. This phenomenon suggests that volatility is not constant over time but exhibits temporal dependence.

Statistical testing for volatility clustering can be performed using several methods. The Ljung-Box test applied to squared returns tests for autocorrelation in the volatility series. Visual inspection of volatility time series plots can also reveal clustering patterns, showing distinct periods of high and low volatility.

The implications for risk management are significant. Since volatility clustering indicates that risk is time-varying rather than constant, traditional risk models that assume constant volatility may be inadequate. Risk managers need models that capture conditional volatility, such as GARCH models, which can adapt to changing market conditions. Risk measures must account for current market conditions rather than relying solely on long-term historical averages. During periods of high volatility, risk limits may need to be adjusted more frequently, and portfolio rebalancing strategies should consider the persistence of volatility regimes.

5. i Consider the autocorrelation function (ACF) plots for daily returns and squared returns of a financial asset, and what to they capture?

   ii How do these ACF patterns relate to market efficiency?

Solution

i The ACF of returns captures linear dependence and predictability in price movements. The ACF of squared returns captures volatility clustering and conditional heteroscedasticity. This reflects the stylised fact that volatility is predictable even when returns are not.

ii Relationship to market efficiency:

- Autocorrelation in returns may violate weak-form efficiency as it suggests predictable price movements. However, weak autocorrelation in returns can be explained by inflation, risk-free rate components, and transaction costs, and so *might not* violate weak-form efficiency.
- Autocorrelation in squared returns (volatility clustering) does not necessarily violate market efficiency because:
  - Volatility forecasting doesn't directly translate to profitable trading
  - High costs of carry for volatility products often exceed predictable components
  - Risk and return are both time-varying, which is consistent with efficient markets

6. Suppose you calculate the first-order autocorrelation of both returns and returns squared. Which of these two would you expect to be more strongly statistically significant?

Solution
The empirical evidence we have seen in the course, obtained by both the Ljung-Box test and the autocorrelation function (ACF) suggest that there is some autocorrelation in returns but it is not very high, while there is significant and long lasting autocorrelation in returns squared. That implies that squared returns are much more strongly statistically significant.

7. Explain the concept of "fat tails" in financial returns.

   i Define fat tails formally.

   ii Provide an intuitive explanation of what this means for extreme events.

   iii How can fat tails be measured or detected?

   iv What distribution is commonly used to model fat tails?

Solution

i Definition: A random variable has fat tails if it exhibits more extreme outcomes than a normally distributed random variable with the same mean and variance.

ii Intuitive meaning: Higher probability of very large positive or negative returns. Extreme events are more common than normal distribution predicts. Market crashes and booms occur more frequently than expected.

iii Detection methods: Kurtosis > 3 (excess kurtosis > 0), Q-Q plots against normal distribution, Jarque-Bera test for normality, or visual inspection of empirical vs. normal density.

iv Modeling: Student-t distribution is commonly used, where the degrees of freedom parameter ν controls tail thickness. Lower ν means fatter tails. As $\nu \to \infty$, the Student-t approaches the normal.

8. You observe the following statistics for daily S&P 500 returns: Mean = 0.05%, Standard deviation = 1.2%, Skewness = -0.3, Kurtosis = 5.2.

   i What do these statistics tell you about the distribution of returns?

   ii How does this compare to a normal distribution?

   iii What test could you use to formally test for normality?

Solution

i The statistics indicate: Small positive mean return (0.05% daily). Negative skewness (-0.3) indicates left tail is heavier (more extreme negative returns). Kurtosis of 5.2 > 3 indicates fat tails (excess kurtosis = 2.2).

ii Compared to normal distribution (skewness = 0, kurtosis = 3): Returns are left-skewed (asymmetric), fat-tailed (higher probability of extreme events), with more extreme outcomes than normal distribution would predict.

iii Tests for normality: Jarque-Bera test (tests skewness and kurtosis jointly), Kolmogorov-Smirnov test, or QQ plots.

9. During the 2008 financial crisis, correlations between major stock indices increased dramatically. For example, correlations between the S&P 500, FTSE 100, and Nikkei 225 rose from around 0.4-0.6 in normal times to over 0.8 during the crisis.

   i What stylized fact does this illustrate?

   ii Why is this problematic for portfolio diversification?

   iii How does this relate to the assumption of joint normality?

Solution

i This illustrates non-linear dependence. The dependence structure between assets changes conditional on market conditions — correlations are higher during stress periods than during normal times.

ii Problematic for diversification because:

- Diversification benefits disappear when you need them most (during crises)
- Risk models based on average correlations underestimate crisis risk
- "Safe" diversified portfolios may become highly correlated during stress

iii Under joint normality, correlations would be constant regardless of market conditions. The empirical evidence of time-varying, state-dependent correlations violates the assumption of elliptical (including normal) distributions and indicates non-linear dependence.

10. What is the difference between a price weighted and a value weighted stock market index? Which type is more useful for financial analysis? Give one example of each.

Solution
A price-weighted index weighs stocks based on their prices, e.g. a stock trading at $100 makes up 10 times more of total than a stock trading at $10.

A value weighted index weighs stocks according to the total market value of their outstanding shares so the impact of change in stock price proportional to overall market value.

A value weighted index is preferable. The S&P 500, FTSE 100 and TOPIX are value weighted while the DJIA and the Nikkei 225 are price weighted.

11. What are simple and continuously compounded returns?

Solution

Simple returns are the percentage changes in prices between two periods,

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Continuously compounded returns are the logarithm of gross return,

$$Y_t = \log(1 + R_t) = \log(P_t) - \log P_{t-1}.$$

12. What are the advantages and disadvantages of using continuously compounded returns?

Solution

The main advantage of continuously compounded returns is that they are symmetric, so a loss and profit of the same magnitude will have the same magnitude return, and play an important role for many financial calculations.

A disadvantage of continuously compounded returns is that portfolio returns cannot be calculated simply as a weighted average of the returns of individual assets.

13. What are the three main stylized facts of financial returns?

Solution
The three main stylized facts of financial returns are:

Volatility clustering: Periods of high volatility tend to be followed by periods of high volatility, and periods of low volatility tend to be followed by periods of low volatility.

Fat tails: The distribution of returns has fatter tails than the normal distribution, meaning extreme events occur more frequently than predicted by normality.

Non-linear dependence: While returns themselves may show little serial correlation, there are non-linear dependencies in the data, particularly in measures of volatility and higher moments.

14. A Japanese investor has JPY 1,000,000 invested in the TOPIX index. The historical annual volatility is 20% and the annual mean return is 2%. The annual Japanese inflation is 0.1%. Based on this information, would you expect the long-run real investment performance to be less than 2%, equal to 2%, or more than 2% annually?

Solution
The long-run real investment performance would be less than 2% annually.

The precise real return calculation uses the Fisher equation:

$$\text{Real return} = \frac{1 + \text{Nominal return}}{1 + \text{Inflation rate}} - 1$$

$$\text{Real return} = \frac{1.02}{1.001} - 1 = 0.018982 \approx 1.90\%$$

For small inflation rates, the approximation Real return $\approx$ Nominal return $-$ Inflation rate $= 2\% - 0.1\% = 1.9\%$ gives essentially the same result.

Therefore, the real return is approximately 1.90%, which is less than the nominal return of 2%. This reflects the erosion of purchasing power due to inflation, even when inflation is relatively low.

Note: This calculation assumes the stated 2% return represents total return including dividends. The volatility of 20% does not affect the expected long-run real return calculation, though it does indicate substantial year-to-year variation around this expected value.

Chapter 2

Volatility forecasting

1. List three major problems with the moving average (MA) volatility model and explain why EWMA is preferred to the MA models.

Solution

Three major problems with MA models:

(a) Window sensitivity: The choice of estimation window W_E strongly affects the forecast, with no clear guidance on optimal window size

(b) Equal weighting: All observations within the window receive equal weight, ignoring the fact that recent observations should be more relevant

(c) Discontinuity: When the window moves, old observations are dropped completely, causing jumps in volatility forecasts

EWMA is preferred because it addresses these issues by using exponentially declining weights, giving more importance to recent observations while maintaining continuity.

2. Consider an EWMA model with $\lambda = 0.94$. If yesterday's return was $y_{t-1} = 0.02$ and yesterday's volatility forecast was $\hat{\sigma}_{t-1} = 0.015$, calculate today's volatility forecast $\hat{\sigma}_t$.

Solution
Using the EWMA equation:

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

Substituting the values:

$$\hat{\sigma}_t^2 = (1 - 0.94) \times 0.02^2 + 0.94 \times 0.015^2$$

$$\hat{\sigma}_t^2 = 0.06 \times 0.0004 + 0.94 \times 0.000225 = 0.000024 + 0.0002115 = 0.0002355$$

Therefore: $\hat{\sigma}_t = \sqrt{0.0002355} = 0.0153$

3. Suppose you estimate a normal ARCH model for a stock, finding the parameter estimates for day t to be

$\hat{\alpha}$	0.5
$\hat{\omega}$	0.00001

Making all necessary assumptions, what is the conditional kurtosis?

Solution
Since

$$y_t = \sigma_t \epsilon_t$$

where

$$\epsilon_t \sim N(0, 1)$$

ϵ_t is the normal, and it follows it is the conditional distribution. Since the kurtosis for the normal is 3, the conditional kurtosis is also 3.

4. Describe the relationship between the EWMA model and GARCH, and explain why EWMA has undefined unconditional volatility.

Solution
Relationship: EWMA is a restricted GARCH(1,1) model where:

- $\omega = 0$
- $\alpha = 1 - \lambda$
- $\beta = \lambda$

Unconditional volatility: For GARCH(1,1):

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

For EWMA with $\omega = 0$:

$$\sigma^2 = \frac{0}{1 - (1 - \lambda) - \lambda} = \frac{0}{0}$$

This is undefined because EWMA has no long-run mean for volatility — it depends entirely on the history of returns. The model can produce any level of volatility depending on the initial conditions and the sequence of returns.

5. A GARCH(1,1) model for daily stock returns has the following parameter estimates: $\hat{\omega} = 0.00001$, $\hat{\alpha} = 0.12$, $\hat{\beta} = 0.85$. Yesterday's return was $y_t = -0.03$ and yesterday's conditional variance was $\hat{\sigma}_t^2 = 0.0009$. Calculate today's conditional variance forecast and determine if this GARCH model is stationary.

Solution

Conditional variance forecast: Using the GARCH(1,1) updating equation:

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

$$\hat{\sigma}_{t+1}^2 = 0.00001 + 0.12 \times (-0.03)^2 + 0.85 \times 0.0009$$

$$\hat{\sigma}_{t+1}^2 = 0.00001 + 0.12 \times 0.0009 + 0.85 \times 0.0009$$

$$\hat{\sigma}_{t+1}^2 = 0.00001 + 0.000108 + 0.000765 = 0.000883$$

Therefore, today's conditional volatility is $\hat{\sigma}_{t+1} = \sqrt{0.000883} = 0.0297$ or 2.97%.
Stationarity check: For GARCH(1,1) stationarity, we require $\hat{\alpha} + \hat{\beta} < 1$.
$\hat{\alpha} + \hat{\beta} = 0.12 + 0.85 = 0.97 < 1$
Since $0.97 < 1$, the model is stationary.

$$\sigma^2 = \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}} = \frac{0.00001}{1 - 0.97} = \frac{0.00001}{0.03} = 0.000333$$

The unconditional volatility is $\sqrt{0.000333} = 0.0182$ or 1.82%.

6. A GARCH(1,1) model has parameter estimates $\hat{\omega} = 0.00001$, $\hat{\alpha} = 0.1$, and $\hat{\beta} = 0.9$. Yesterday's return was $y_t = -0.1$ and yesterday's conditional variance was $\hat{\sigma}_t^2 = 0.0002$. Calculate both the unconditional and conditional volatility, and explain the implications when $\hat{\alpha} + \hat{\beta} = 1$.

Solution

Unconditional volatility: For a GARCH(1,1) model, the unconditional variance is:

$$\sigma^2 = \frac{\hat{\omega}}{1 - \hat{\alpha} - \hat{\beta}}$$

With $\hat{\alpha} = 0.1$, $\hat{\beta} = 0.9$, and $\hat{\omega} = 0.00001$:

$$\sigma^2 = \frac{0.00001}{1 - 0.1 - 0.9} = \frac{0.00001}{0} = \infty$$

The unconditional volatility is infinite (undefined).

Conditional volatility: The conditional volatility for day $t + 1$ is:

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

$$\hat{\sigma}_{t+1} = \sqrt{0.00001 + 0.1 \times (-0.1)^2 + 0.9 \times 0.0002} = \sqrt{0.00019} = 0.01378$$

The conditional volatility is finite: 1.378%.

Reconciliation and implications: The model exhibits the paradox where conditional volatility is finite but unconditional volatility is infinite. This occurs because $\hat{\alpha} + \hat{\beta} = 1$, making the model non-stationary.

Implications: The model is technically integrated GARCH or I-GARCH. Volatility shocks have permanent effects, long-run variance forecasts are meaningless, but short-term conditional forecasts remain valid.

Prevention during estimation: To ensure stationarity, impose the constraint $\hat{\alpha} + \hat{\beta} < 1$ during maximum likelihood estimation. This prevents the optimization from reaching parameter values that would make the unconditional variance infinite.

7. For a GARCH(1,1) model with parameters $\omega = 0.000001$, $\alpha = 0.08$, and $\beta = 0.90$, calculate the unconditional volatility.

Solution
The unconditional volatility for GARCH(1,1) is:

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

Substituting the values:

$$\sigma^2 = \frac{0.000001}{1 - 0.08 - 0.90} = \frac{0.000001}{0.02} = 0.00005$$

Therefore: $\sigma = \sqrt{0.00005} = 0.00707$ or approximately 0.71%

8. Contrast conditional and unconditional volatility and kurtosis in volatility models. Explain how ARCH/GARCH models can produce fat-tailed returns while assuming conditionally normal residuals.

Solution

Volatility: Conditional volatility (σ_t) is time-varying volatility that depends on past information. In GARCH(1,1): $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. Unconditional volatility (σ) is the long-run average volatility. For GARCH(1,1): $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$.

Kurtosis: Conditional kurtosis: If $\epsilon_t \sim N(0,1)$, then conditional kurtosis = 3 (normal). Unconditional kurtosis: For ARCH(1) with normal residuals: $\frac{3(1-\alpha^2)}{1-3\alpha^2} > 3$ (fat-tailed).

Key insight: Returns can be conditionally normal and unconditionally fat-tailed simultaneously. The time-varying conditional volatility creates the fat tails in the unconditional distribution, even though each period's innovation is normally distributed.

Economic interpretation: Periods of high volatility cluster together. During calm periods, returns appear normal with low volatility. During turbulent periods, returns are still normal but with high volatility. This volatility clustering creates the fat tails observed in unconditional return distributions.

9. Explain why the parameter restriction $\alpha + \beta < 1$ is not always imposed in GARCH estimation, despite it being required for stationarity.

Solution

The restriction $\alpha + \beta < 1$ is not always imposed for two main reasons:

(a) It can lead to multiple parameter combinations satisfying the constraint, making volatility forecasts non-unique

(b) The model may be misspecified, and the non-restricted model could provide more accurate forecasts even if theoretically non-stationary

In practice, imposing this constraint can cause optimization problems where the likelihood function becomes flat, leading to unstable parameter estimates during backtesting.

10. In a GARCH(1,1) model, distinguish between the roles of α and β in terms of "news" and "memory".

Solution
In GARCH(1,1), α captures the "news impact" coefficient, measuring how volatility responds to new market shocks through recent squared returns. Typical values range from 0.05 to 0.15.

The parameter β represents "memory" or persistence, determining how much past volatility carries forward. Usually between 0.80 and 0.95, reflecting volatility clustering in financial markets.

Their sum $\alpha + \beta$ measures overall persistence. With $\alpha + \beta = 0.98$, the half-life of a volatility shock is $\ln(0.5)/\ln(0.98) \approx 35$ days. Values near 1 indicate long memory where shocks decay slowly, explaining persistent volatility clustering.

11. Calculate the half-life of volatility shocks for a GARCH(1,1) model with $\alpha = 0.05$ and $\beta = 0.92$.

Solution
The half-life formula is:

$$n^* = 1 + \frac{\log(1/2)}{\log(\alpha + \beta)}$$

Substituting the values:

$$n^* = 1 + \frac{\log(0.5)}{\log(0.05 + 0.92)} = 1 + \frac{-0.693}{log(0.97)} = 1 + \frac{-0.693}{-0.0305} = 1 + 22.7 = 23.7$$

Therefore, it takes approximately 24 days for a volatility shock to decay to half its original impact.

12. Explain the difference between conditional and unconditional fat tails in the context of tGARCH versus normal GARCH models.

Solution

- Normal GARCH: Uses conditionally normal residuals ($\epsilon_t \sim N(0,1)$), so returns are conditionally normal. However, the time-varying volatility makes unconditional returns fat-tailed with kurtosis > 3.
- tGARCH: Uses conditionally fat-tailed residuals ($\epsilon_t \sim t_\nu$), making returns both conditionally and unconditionally fat-tailed.

The key insight is that returns can be conditionally normal and unconditionally fat-tailed simultaneously. The purpose of tGARCH is to make unconditional returns even fatter than they would be with normal GARCH, which can be important for accurate risk measurement.

13. In the APARCH model $\sigma_t^2 = \omega + \alpha(|y_{t-1}| - \zeta y_{t-1})^\delta + \beta\sigma_{t-1}^\delta$, explain how the parameters ζ and δ capture leverage and power effects. Also, explain the economic intuition behind the leverage effect in equity markets.

Solution

- Leverage effect (ζ): When $\zeta > 0$, negative returns ($y_{t-1} < 0$) have a larger impact on volatility than positive returns of the same magnitude. This captures the empirical observation that bad news increases volatility more than good news, reflecting the leverage effect in equity markets.
- Power effect (δ): When $\delta \neq 2$, the model allows for non-quadratic relationships between returns and volatility. This can improve model fit when the standard quadratic form is too restrictive.

If $\zeta = 0$ and $\delta = 2$, the model reduces to standard GARCH(1,1).
Economic intuition: When stock prices fall, the debt-to-equity ratio of companies increases, making them riskier. This increased risk translates to higher volatility. Additionally, bad news tends to spread faster and create more market uncertainty than good news.

APARCH specification:

$$\sigma_t^2 = \omega + \alpha(|y_{t-1}| - \zeta y_{t-1})^\delta + \beta\sigma_{t-1}^\delta$$

When $\zeta > 0$ and $y_{t-1} < 0$ (negative return):

$$|y_{t-1}| - \zeta y_{t-1} = |y_{t-1}| + \zeta|y_{t-1}| = (1 + \zeta)|y_{t-1}|$$

This makes negative returns have a larger impact on volatility than positive returns of the same magnitude.

14. Explain why the likelihood function for GARCH models starts from $t = 2$ rather than $t = 1$.

Solution
The GARCH model requires lagged values:

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

For $t = 1$, we would need y_0 and σ_0, which are unknown since they precede our sample. Therefore, the likelihood function starts from $t = 2$ where we can use y_1 and σ_1. The value of σ_1 is typically set to the sample standard deviation $\hat{\sigma}$ since, with large samples, this choice has minimal impact on parameter estimates.

15. In maximum likelihood estimation, distinguish between local and global maxima problems and suggest how to address them.

Solution

Local maxima: The optimization algorithm finds a peak that is not the highest point of the likelihood function. This can occur when the likelihood has multiple peaks.

Global maxima: The true maximum of the likelihood function that we want to find.

Solutions: Use multiple starting values for the optimization algorithm, try different optimizers (some are more robust than others), check parameter estimates for economic sensibility, and examine the likelihood surface graphically when possible.

This problem is more common in complex models with many parameters.

16. Given the optimization challenges in volatility model estimation, explain why simpler models like GARCH(1,1) are often preferred over more complex alternatives.

Solution

Optimization challenges with complexity: More parameters increase the likelihood of multiple local maxima, flat likelihood surfaces become more common, convergence problems are more frequent, and parameter estimates become less stable.

Advantages of GARCH(1,1): Few parameters reduce optimization problems, the model allows robust estimation with moderate sample sizes, provides good empirical performance despite simplicity, makes it easier to diagnose estimation problems, and produces more stable parameter estimates across different periods.

The "curse of dimensionality" suggests that adding parameters often does not improve out-of-sample performance enough to justify the increased complexity.

17. Compare and contrast the concepts of unconditional and conditional volatility.

Solution
Unconditional volatility is volatility over an entire sample period, while conditional volatility is the volatility in a given time period (like a day), conditional on what happened in the time period before.
Key differences:

- Unconditional volatility: Constant over time, represents long-run average volatility
- Conditional volatility: Time-varying, depends on recent market events and information

Examples:

- Unconditional: Sample standard deviation over entire period
- Conditional: GARCH volatility forecast for tomorrow based on today's information

Conditional volatility is more useful for risk management as it captures volatility clustering and adapts to changing market conditions.

Chapter 3

Multivariate volatility

1. Why must a covariance matrix be positive definite and what happens if this condition is violated?

Solution

A covariance matrix must be positive definite to ensure that portfolio variance $w'\Sigma w \geq 0$ for any portfolio weights w, which guarantees that variance is always non-negative as required by statistical theory. If this condition is violated, portfolio variance could become negative (which is mathematically impossible), risk calculations become meaningless, VaR estimates may be incorrect, and optimization problems may have no solution. Positive definiteness is therefore essential for any meaningful risk analysis or portfolio optimization.

2. How do dimensionality and positive definiteness interact in practice?

Solution

Dimensionality and positive definiteness create fundamental tension in multivariate volatility modelling. The dimensionality constraint requires reducing parameters to make estimation feasible — a 100-asset portfolio has 5,050 unique elements in its covariance matrix.

However, any parameter reduction must preserve positive definiteness to ensure mathematical validity. The practical solution employs structured models like EWMA or DCC that automatically guarantee positive definiteness while dramatically reducing parameters.

This trade-off accepts model restrictions to achieve computational feasibility and statistical reliability, explaining why simple extensions of univariate GARCH to multivariate settings fail for large portfolios.

3. A portfolio manager needs to estimate the covariance matrix for 50 stocks. How many unique parameters need to be estimated and what is this problem called?

Solution
The total number of parameters that need to be estimated is $K + \frac{K(K-1)}{2} = 50 + \frac{50\times 49}{2} = 50 + 1225 = 1275$ parameters. This is called the "curse of dimensionality" and it's challenging because there are too many parameters relative to available data, making estimation unreliable. The computational complexity increases dramatically and there is a high risk of overfitting when the number of parameters approaches or exceeds the sample size.

4. The EWMA model is one of the most widely used approach for multivariate volatility modeling in practice. Write down the EWMA model for forecasting the covariance matrix for two assets.

Solution

The covariance matrix is

$$\Sigma_t = \begin{pmatrix} \sigma_{t,11} & \sigma_{t,12} \\ \sigma_{t,12} & \sigma_{t,22} \end{pmatrix}.$$

The weight λ is assumed to be known, often set at 0.94 for daily returns. The multivariate EWMA form is:

$$\hat{\Sigma}_t = \lambda \hat{\Sigma}_{t-1} + (1-\lambda) y'_{t-1} y_{t-1}$$

with an individual element given by:

$$\sigma_{t,ij} = \lambda \sigma_{t-1,ij} + (1-\lambda) y_{t-1,i} y_{t-1,j}, \quad i,j = 1,\ldots,K.$$

5. Why does EWMA automatically ensure positive definiteness ?

Solution
EWMA automatically ensures positive definiteness through its mathematical structure. The update formula

$$\hat{\Sigma}_t = \lambda\hat{\Sigma}_{t-1} + (1-\lambda)y'_{t-1}y_{t-1}$$

represents a convex combination of two positive definite matrices.

The outer product

$$y'_{t-1}y_{t-1}$$

is always positive semidefinite by construction, and using the same decay parameter λ for all elements preserves the covariance structure. Since convex combinations of positive definite matrices remain positive definite, starting with a valid covariance matrix guarantees all subsequent forecasts maintain positive definiteness without additional constraints or checks.

6. A portfolio has weights $w = [0.6, 0.4]'$ and covariance matrix $\Sigma = \begin{pmatrix} 0.04 & 0.02 \\ 0.02 & 0.09 \end{pmatrix}$.

   i Calculate the portfolio variance.

   ii What is the portfolio volatility?

Solution

i Portfolio variance $= w'\Sigma w = [0.6, 0.4] \begin{pmatrix} 0.04 & 0.02 \\ 0.02 & 0.09 \end{pmatrix} \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$

$= [0.6, 0.4] \begin{pmatrix} 0.032 \\ 0.048 \end{pmatrix} = 0.6(0.032) + 0.4(0.048) = 0.0384$

ii Portfolio volatility $= \sqrt{0.0384} = 0.196$ or 19.6%

7. Identify the main advantage and main limitation of the EWMA model.

Solution
Main advantage: Very simple to implement with no parameter estimation required, guaranteed positive definiteness, and computationally efficient even for large portfolios.

Main limitation: Very restrictive correlation structure that assumes all correlations decay at the same rate, which may not reflect reality where different asset pairs have different correlation dynamics.

8. Compare the CCC and DCC models for multivariate volatility modeling, covering their theoretical foundations, estimation procedures, and practical applications.

   Explain what each acronym stands for and the fundamental difference between the models.

   Solution

   CCC: Constant Conditional Correlation — assumes correlations remain constant over time.

   DCC: Dynamic Conditional Correlation — allows correlations to vary over time following a GARCH-like process.

   Fundamental difference: CCC models assume correlations remain constant over time, making them simpler to estimate but potentially missing important time-varying correlation patterns. In contrast, DCC models allow correlations to vary over time, making them more flexible and realistic but requiring more parameters to estimate.

9. Why is the DCC model generally preferred over CCC and what are the trade-offs?

Solution
DCC advantages:

- Captures time-varying correlations that are empirically observed
- More flexible and realistic for changing market conditions
- Better performance during crisis periods when correlations increase
- Provides good balance between computational difficulties and accuracy

Trade-offs:

- Complexity: More parameters to estimate (ζ, ξ)
- Computation: Higher computational burden
- Overfitting risk: More complex model may overfit with limited data
- When CCC suffices: If correlations are genuinely stable, CCC may be adequate

Conclusion: DCC is generally preferred because it provides a good balance between computational feasibility and accuracy, being rich enough to capture correlation dynamics but restrictive enough to allow estimation even with large numbers of assets.

10. Write down the DCC model equations and explain the two-step estimation procedure.

Solution
DCC Model:

$$R_t = Z_t Q_t Z_t$$

Where matrix Q_t drives the dynamics and matrix Z_t re-scales the dynamics to ensure each element is between -1 and +1.
$\hat{Q}_t$ is given by:

$$\hat{Q}_t = (1 - \zeta - \xi)\overline{Q} + \zeta Y'_{t-1} Y_{t-1} + \xi \hat{Q}_{t-1}$$

where ζ and ξ are parameters to be estimated.
Two-step estimation procedure:

- Step 1: Estimate univariate GARCH models for each asset to obtain standardized residuals: $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$
- Step 2: Estimate correlation dynamics using the standardized residuals from step 1, with Z_t matrix elements:

$$Z_t = \begin{pmatrix} 1/\sqrt{q_{t,11}} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{q_{t,22}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1/\sqrt{q_{t,KK}} \end{pmatrix}$$

11. How does crisis behavior in prices affect the performance of standard multivariate volatility models?

Solution

Crisis correlation behavior creates significant challenges for standard models:

- CCC models fail: Constant correlation assumption is violated during crises
- DCC models struggle: May not capture the speed and magnitude of correlation changes
- EWMA limitations: Single decay parameter cannot capture different crisis dynamics
- Risk underestimation: Models calibrated on normal periods underestimate crisis risk
- Backtesting failures: Models show poor performance during stress periods

The result is that portfolio risk is systematically underestimated during the periods when accurate risk measurement is most critical.

12. What practical approaches can risk managers use to address problems caused by crisis dynamics?

Solution
Risk managers can employ several approaches to address crisis correlation behavior:

- Stress testing: Regularly test portfolios under scenarios where correlations approach unity
- Regime-switching models: Use models that allow for different correlation regimes
- Rolling window estimation: Use shorter estimation windows to capture recent correlation changes
- Tail dependence modeling: Focus on correlations in extreme market conditions
- Copula approaches: Model dependence structure separately from marginal distributions
- Dynamic correlation floors: Assume minimum correlation levels during stress periods

The key is to supplement standard models with approaches that explicitly account for crisis behavior rather than relying solely on historical averages.

Chapter 4

Risk measures

1. Explain why risk is considered a "latent variable" and how this differs from measuring observable quantities like stock prices.

Solution
Risk is a latent variable because it cannot be directly measured or observed. Unlike stock prices, which are unambiguous market-determined numbers, risk must be inferred from the movements of market prices using mathematical models.
Key differences:

- Stock prices: Directly observable, unambiguous, can be measured precisely
- Risk: Must be inferred from price movements, requires a model to translate fluctuations into risk concepts, only accurate to about 2 significant digits

This is why there are infinite possible risk measures and risk measurements — each reflects different modeling assumptions and user preferences. The choice of risk measure depends on the specific application and the user's risk tolerance and investment horizon.

2. Consider three assets A, B, and C, all with identical means ($\mu = 0.08$) and variances ($\sigma^2 = 0.04$) but different distribution shapes. Asset A has normally distributed returns, Asset B has positively skewed returns, and Asset C has negatively skewed returns with fat tails. Explain why different investors might prefer different assets despite the MV model treating them as equivalent.

Solution
The MV model treats all three assets as equivalent while investors have different preferences based on time horizons and risk tolerance.

Asset A (normal) has symmetric distribution with moderate tail risk. Asset B (positive skew) offers frequent small losses with occasional large gains. Asset C (negative skew, fat tails) provides frequent small gains but occasional large losses.

Day traders prefer Asset B because they can cut losses quickly while riding winners, and fat tails are less concerning over short periods.

Fund managers prefer Asset A to avoid extreme outcomes affecting quarterly performance. Asset C creates career risk from large quarterly losses that are difficult to explain to investors.

Pension savers might tolerate Asset C since long horizons allow recovery from occasional large losses, prioritizing long-term wealth accumulation over short-term volatility.

Risk managers view Asset C as most problematic because fat tails increase extreme loss probability and negative skew creates "lottery ticket" mentality that can threaten firm survival.

This demonstrates why the mean-variance model fails to capture all relevant investor preferences and risk characteristics.

3. Does ES satisfies the monotonicity axiom of coherent risk measures? Consider two portfolios: A has returns that are always \$100 lower than B for every possible outcome. Which portfolio should have higher risk according to monotonicity?

Solution
Monotonicity requires: If $A \leq B$ (meaning A always has worse outcomes than B), then $\varphi(A) \geq \varphi(B)$ (A should have higher risk than B).
Given that Portfolio A has returns that are always \$100 lower than Portfolio B:

$$A_i = B_i - 100 \text{ for all outcomes } i$$

This means $A \leq B$ (A dominates B in the sense of having consistently worse outcomes).
According to monotonicity, we should have:

$$ES(A) \geq ES(B)$$

ES satisfies monotonicity because:

- If A always performs \$100 worse than B, then A's worst outcomes are also \$100 worse
- The expected loss conditional on being in the worst tail will be \$100 higher for A
- Therefore: $ES(A) = ES(B) + 100$, which satisfies $ES(A) \geq ES(B)$

This makes economic sense — if one portfolio consistently performs worse than another across all scenarios, it should be considered riskier.

4. Test whether volatility satisfies the translation invariance axiom of coherent risk measures. If an asset has volatility $\sigma_A = 0.03$ and we add a constant $c = 0.001$ to all returns, what happens to the volatility?

Solution
Translation invariance requires: $\varphi(A + c) = \varphi(A) - c$
For volatility:

$$\begin{aligned}\sigma(A + c) &= \sqrt{E[(A + c - E[A + c])^2]} \\ &= \sqrt{E[(A + c - \mu_A - c)^2]} \\ &= \sqrt{E[(A - \mu_A)^2]} \\ &= \sigma(A)\end{aligned}$$

Therefore: $\sigma(A + c) = \sigma(A) = 0.03$

Since $\sigma(A+c) = \sigma(A) \neq \sigma(A)-c$, volatility does not satisfy translation invariance.

This makes economic sense — adding a constant to all returns shifts the mean but doesn't change the variability around that mean, so volatility remains unchanged.

5. Demonstrate positive homogeneity for VaR. If a portfolio worth $5 million has a 1% daily VaR of $100,000, what is the 1% VaR if the portfolio size doubles to $10 million?

Solution
Positive homogeneity requires: $\varphi(cA) = c\varphi(A)$ for $c > 0$
Given:

- Original portfolio value: $5 million
- Original VaR: $100,000
- New portfolio value: $10 million (so $c = 2$)

Under positive homogeneity:

$$VaR(2 \times \text{portfolio}) = 2 \times VaR(\text{portfolio}) = 2 \times \$100,000 = \$200,000$$

VaR satisfies positive homogeneity when portfolio composition remains unchanged and we simply scale the position sizes. The risk doubles when the portfolio value doubles, which is economically reasonable for most market conditions.

Note: This can be violated in practice due to liquidity effects — larger positions may be harder to liquidate quickly.

6. Show how VaR can violate subadditivity using the example from Chapter 4. Two independent assets A and B each have a 4.9% probability of losing 100 and 95.1% probability of no loss. Calculate the 5% VaR for each asset individually and for an equally weighted portfolio.

Solution
For each individual asset:

- 4.9% probability of -100 loss
- 95.1% probability of 0 loss

Since the 5th percentile falls at 4.9

$$VaR^{5\%}(A) = VaR^{5\%}(B) = 0$$

For the equally weighted portfolio $\frac{1}{2}A + \frac{1}{2}B$:

- Probability of (-100, -100): $0.049 \times 0.049 = 0.24\%$
- Probability of (-100, 0) or (0, -100): $2 \times 0.049 \times 0.951 = 9.3\%$
- Probability of (0, 0): $0.951 \times 0.951 = 90.4\%$

The cumulative probability up to -50 is $0.24\% + 9.3\% = 9.54\%$, which exceeds 5%.

Therefore: $VaR^{5\%}(\text{portfolio}) = 50$

Subadditivity violation:

$$VaR^{5\%}(\text{portfolio}) = 50 > 0.5 \times 0 + 0.5 \times 0 = 0$$

The portfolio appears riskier than the individual assets, violating the diversification principle.

7. Using the square root of time scaling rule, calculate the 10-day VaR from a 1-day VaR of $50,000. Explain when this scaling rule is accurate and when it may fail.

Solution
Using the square root of time rule:

$$VaR_{10\text{ days}} = VaR_{1\text{ day}} \times \sqrt{10} = \$50,000 \times 3.162 = \$158,100$$

When the rule is accurate: Returns must be independently and identically distributed (IID). For volatility scaling, it works regardless of distribution (if variance is defined). For VaR scaling, it's only accurate if returns are IID and normally distributed.

When the rule may fail: The rule fails when returns exhibit volatility clustering (GARCH effects), are not normally distributed (fat tails), show serial correlation, or experience structural breaks or regime changes over the scaling period.

The rule is a rough approximation that becomes less reliable as the scaling period increases and when returns exhibit the stylized facts observed in financial markets.

8. Explain how a trader might manipulate VaR using derivatives.

Solution
A trader can artificially lower VaR by using a combination of put options.

Strategy: Buy a put option with strike price just above the desired VaR level (VaR_1) and sell a put option with strike price below the original VaR level (VaR_0).

Effect on risk profile: The bought put provides protection around the new VaR level, reducing measured risk. The sold put creates exposure to large losses below the original VaR level. This creates a "gap" in the loss distribution, artificially lowering VaR.

Consequences: Lower expected profit (premium paid > premium received), increased tail risk (potential for much larger losses beyond the original VaR), and misleading risk measure (VaR appears lower but actual economic risk may be higher).

This manipulation demonstrates why VaR alone may be insufficient for risk management and why measures like ES, which consider the entire tail distribution, can be more robust.

Chapter 5

Implementing risk forecasts

1. A portfolio manager uses HS to calculate 1% VaR with 500 days of return data. The sorted returns (from smallest to largest) show the 5th smallest return is -0.032 and the 6th smallest is -0.029. For a portfolio worth \$25 million, calculate the 1% VaR and explain the key assumption underlying this method.

   Solution
   Calculation: With 500 observations and 1% probability: $W_E \times \rho = 500 \times 0.01 = 5$
   The 5th smallest return is -0.032, so:

   $$\text{VaR}^{1\%} = \text{Portfolio Value} \times |\text{5th smallest return}| = \$25,000,000 \times 0.032 = \$800,000$$

   Key assumption: HS assumes that history repeats itself. Specifically, it assumes that one of the historical observations will occur tomorrow with equal probability, making the past distribution a perfect predictor of future risk.

   Advantages: HS is distribution-free (nonparametric), simple to implement, and captures actual tail behavior from historical data.

   Limitations: The method assumes the future will resemble the past, cannot capture events not in the historical sample, and requires large sample sizes for extreme quantiles.

2. A portfolio manager has collected the following 20 daily returns (in %) for a stock over the past month:

-2.1	1.3	-0.8	2.4	-1.5	0.9
-3.2	0.7	1.8	-2.8	0.4	-1.1
2.2	-0.6	1.4	-4.1	0.3	1.7
-1.9	-0.5				

For a portfolio value of $50,000, calculate the 1-day 5% VaR and ES using HS, and comment on the reliability of these estimates given the sample size.

Solution
Step 1: Sort returns from smallest to largest -4.1, -3.2, -2.8, -2.1, -1.9, -1.5, -1.1, -0.8, -0.6, -0.5, 0.3, 0.4, 0.7, 0.9, 1.3, 1.4, 1.7, 1.8, 2.2, 2.4

Step 2: Calculate 5% VaR With $W_E = 20$ observations and $\rho = 0.05$: $(W_E \times \rho)^{\text{th}} = (20 \times 0.05)^{\text{th}} = 1^{\text{st}}$ smallest return
The 1st smallest return is -4.1%

$$VaR = 4.1\% \times \$50,000 = \$2,050$$

Step 3: Calculate 5% ES ES is the mean of all observations less than or equal to the VaR threshold. Since VaR corresponds to the 1st smallest observation (-4.1%), and there is only one observation in the tail:

$$ES = 4.1\% \times \$50,000 = \$2,050$$

Note: With only one observation in the tail, ES equals VaR in this case.

Step 4: Sample size reliability assessment The minimum recommended sample size for reliable VaR estimation is $\frac{3}{\rho} = \frac{3}{0.05} = 60$ observations.

Reliability concerns: The current sample of 20 observations is below the minimum threshold. The VaR estimate is based on a single extreme observation, which increases estimation uncertainty. With such a small sample size, results may not be representative of true tail risk. Recommendation: Collect more historical data before using these estimates for risk management.

3. Outline one advantage and two disadvantages of using HS for forecasting risk.

Solution
Advantage: HS requires no distributional assumptions. It uses the empirical distribution directly, avoiding model specification errors and capturing actual return patterns including skewness and fat tails.

Disadvantage 1: Equal weighting makes HS slow to adapt. All observations receive equal weight regardless of recency, so recent volatility changes are ignored until older data rolls out of the window.

Disadvantage 2: Ghost effects create artificial jumps. When extreme observations drop from the estimation window, risk measures can change dramatically overnight without any actual market movement, creating spurious volatility in risk estimates.

4. A stock is currently priced at \$100. Calculate the 1-day 5% VaR using both simple returns and continuously compounded returns, assuming daily volatility is 2% and returns are normally distributed. Comment on the difference.

Solution

Simple returns approach: For simple returns, VaR is calculated as:

$$\text{VaR} = -\sigma\Phi^{-1}(\rho) \times P_{t-1}$$

With $\sigma = 0.02$, $\Phi^{-1}(0.05) = -1.645$, and $P_{t-1} = \$100$:

$$\text{VaR}_{\text{simple}} = -0.02 \times (-1.645) \times \$100 = \$3.29$$

Continuously compounded returns approach: For log returns, VaR is:

$$\text{VaR} = -P_{t-1}(e^{\sigma F_y^{-1}(\rho)} - 1)$$

$$\text{VaR}_{\text{log}} = -\$100(e^{0.02\times(-1.645)} - 1) = -\$100(e^{-0.0329} - 1)$$

$$\text{VaR}_{\text{log}} = -\$100(0.9676 - 1) = -\$100(-0.0324) = \$3.24$$

Comparison:

- Simple returns VaR: \$3.29
- Log returns VaR: \$3.24
- Difference: \$0.05 (1.5%)

Key insights: The difference is small for daily returns with typical volatilities. Log returns produce slightly lower VaR because the exponential transformation creates asymmetry - large positive returns are capped while large negative returns approach -100%. For practical daily risk management, the choice between simple and log returns has minimal impact on VaR calculations.

5. Calculate the 1-day 5% VaR for a $50,000 portfolio with normally distributed returns, zero mean, and daily volatility of 1.8%.

Solution
For normally distributed returns with zero mean, the VaR formula is:

$$VaR = -\sigma\Phi^{-1}(\rho) \times \text{Portfolio Value}$$

Given:

- $\sigma = 0.018$ (daily volatility)
- $\rho = 0.05$ (5% probability level)
- Portfolio Value = $50,000

For the standard normal distribution: $\Phi^{-1}(0.05) = -1.645$

$$VaR = -0.018 \times (-1.645) \times \$50,000 = \$1,481$$

There is a 5% probability that the portfolio will lose $1,481 or more in one day.

6. Suppose you estimate a t-GARCH model for a stock with the following parameter estimates for day t:

$\hat{\alpha}$	0.1
$\hat{\omega}$	0.00001
$\hat{\beta}$	0.85
$\hat{\nu}$	∞
y_t	-0.1
$\hat{\sigma}_t^2$	0.0002

What is the 1-day 1% VaR forecast for day $t+1$? State all necessary assumptions.

Solution

We need to make three assumptions: probability, portfolio value and the conditional distribution.

Assumptions:

(a) Probability level: 1% ($\Phi^{-1}(0.01) = -2.33$)

(b) Portfolio value: \$1,000

(c) Conditional distribution: Normal (since $\hat{\nu} = \infty$, the t-GARCH reduces to normal GARCH)

Calculation: First, forecast the conditional variance for day $t+1$:

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

$$\hat{\sigma}_{t+1}^2 = 0.00001 + 0.1 \times (-0.1)^2 + 0.85 \times 0.0002 = 0.000181$$

Therefore: $\hat{\sigma}_{t+1} = \sqrt{0.000181} = 0.01345$

VaR calculation:

$$VaR_{t+1} = -\hat{\sigma}_{t+1} \times \Phi^{-1}(0.01) \times \text{Portfolio Value}$$

$$VaR_{t+1} = -0.01345 \times (-2.33) \times \$1,000 = \$31.34$$

7. Using the t-GARCH parameter estimates from Question 6, calculate the 2-day 1% VaR forecast using the square-root-of-time rule.

Solution

The square-root-of-time rule states that for independent returns, VaR scales with the square root of the time horizon:

$$VaR_T = VaR_1 \times \sqrt{T}$$

where T is the number of time periods.

From Question 6 1-day VaR = \$31.34

2-day VaR calculation:

$$VaR_{2-day} = VaR_{1-day} \times \sqrt{2}$$

$$VaR_{2-day} = \$31.34 \times \sqrt{2} = \$31.34 \times 1.414 = \$44.33$$

Key assumption: The square-root-of-time rule assumes that returns are independent and identically distributed (IID), which may not hold perfectly with GARCH effects where volatility clustering occurs.

8. A portfolio worth JPY 1,000,000 has annual volatility of 20% and annual mean return of 2%. Calculate the 1-month and 1,000-year VaR using the square-root-of-time rule, and explain what fundamental issues the long-term VaR result reveals about the limitations of VaR as a risk measure. Use 1% probability level and assume IID normally distributed returns.

Solution

One-month VaR calculation:
Time scaling parameters:

- Time horizon: $T = \frac{1}{12}$ years (one month)
- Monthly volatility: $\sigma_m = 0.20 \times \sqrt{\frac{1}{12}} = 0.20 \times 0.2887 = 0.0577$
- Monthly mean: $\mu_m = 0.02 \times \frac{1}{12} = 0.00167$

For 1% VaR with $\Phi^{-1}(0.01) = -2.33$:

$$VaR_{\text{1-month}} = -0.0577 \times (-2.33) - 0.00167 = 0.1344 - 0.00167 = 0.1327$$

$$VaR_{\text{1-month}} = 0.1327 \times JPY 1,000,000 = JPY 132,700$$

1,000-year VaR calculation:
Time scaling for 1,000 years:

- Scaled volatility: $\sigma_{1000} = 0.20 \times \sqrt{1,000} = 6.324$
- Scaled mean: $\mu_{1000} = 0.02 \times 1,000 = 20$

$$VaR_{\text{1000-year}} = -6.324 \times (-2.33) - 20 = 14.735 - 20 = -5.265$$

$$VaR_{\text{1000-year}} = -5.265 \times JPY 1,000,000 = -JPY 5,265,000$$

Fundamental issues revealed by negative long-term VaR:

(a) Contradiction of VaR definition: A negative VaR suggests gains rather than losses, contradicting VaR's purpose as a loss measure.

(b) Mean-volatility scaling problem: Over long horizons, mean scales linearly (μT) while volatility scales with square root ($\sigma\sqrt{T}$). Eventually mean dominates, creating meaningless results.

Conclusion: This demonstrates that VaR is appropriate only for short-term risk measurement (typically 1-10 days). Long-term risk assessment requires different methodologies that account for parameter uncertainty and structural changes.

9. Explain what "model risk" means in the context of VaR estimation and provide two specific examples of how it can manifest in practice.

Solution
Model risk refers to the uncertainty and potential errors that arise from using mathematical models to estimate risk measures like VaR. Since risk is a latent variable that cannot be directly observed, different models can produce substantially different risk estimates for the same portfolio.

Two specific examples:
1. Distributional assumptions: Consider a portfolio with daily volatility of 2%. The 1% VaR estimates would be:

- Normal distribution: VaR = 2% × 2.33 = 4.66%
- t-distribution ($\nu = 4$): VaR = 2% × 3.75 = 7.50%

The choice of distribution creates a 61% difference in VaR estimates, with potential capital allocation and risk management implications.

2. Estimation methodology: Using the same return data, different VaR approaches can yield:

- HS: Uses empirical quantiles from past data
- GARCH: Uses time-varying volatility with parametric distributions
- EWMA: Uses exponentially weighted variance with constant distribution

These methods often produce VaR estimates that can differ by 50% or more, especially during volatile periods.

Implications: Model risk means that risk managers must acknowledge the inherent uncertainty in their estimates, validate models through backtesting, and often use multiple models or conservative approaches to account for this uncertainty.

10. Using the t-GARCH parameter estimates from Question 6, suppose instead that $\hat{\nu} = 3$ degrees of freedom. Calculate the 1-day 1% VaR forecast for day $t+1$.

Solution
Key difference: With $\hat{\nu} = 3$, we now have a true t-GARCH model with fat-tailed conditional distribution instead of the normal distribution.

From the calculation in Question 6 $\hat{\sigma}_{t+1} = 0.01345$ (conditional volatility remains the same)

Critical value for t-distribution: For a t-distribution with $\nu = 3$ degrees of freedom at the 1% level: $t^{-1}_{0.01,3} = -4.54$

VaR calculation:

$$VaR_{t+1} = -\hat{\sigma}_{t+1} \times t^{-1}_{0.01,3} \times \text{Portfolio Value}$$

$$VaR_{t+1} = -0.01345 \times (-4.54) \times \$1,000 = \$61.06$$

Comparison:

- Normal distribution VaR: \$31.34
- t-distribution ($\nu = 3$) VaR: \$61.06

The t-distribution produces a VaR that is 95% higher due to the fatter tails, reflecting the higher probability of extreme events with lower degrees of freedom.

Note: This calculation uses the unstandardized t-distribution. If using a standardized t-GARCH where the variance is scaled to unity, an additional correction factor $\sqrt{\frac{\nu-2}{\nu}}$ would be needed.

11. For a portfolio with normally distributed returns having mean $\mu = 0$ and standard deviation $\sigma = 0.03$, calculate both the 5% VaR and 5% ES. Explain why ES is always greater than VaR and what this difference represents.

Solution
5% VaR calculation: For normally distributed returns:

$$VaR^{5\%} = -\sigma \times \Phi^{-1}(0.05) = -0.03 \times (-1.645) = 0.04935$$

5% ES calculation: For normally distributed returns:

$$ES^{5\%} = -\sigma \times \frac{\phi(\Phi^{-1}(0.05))}{0.05}$$

Where $\phi(\Phi^{-1}(0.05)) = \phi(-1.645) = 0.1031$

$$ES^{5\%} = -0.03 \times \frac{0.1031}{0.05} = -0.03 \times 2.062 = 0.06186$$

Results:

- 5% VaR = 4.935%
- 5% ES = 6.186%

Why ES > VaR: VaR measures the threshold loss that occurs with 5% probability, while ES measures the expected loss conditional on being in the worst 5% of outcomes. ES captures the severity of losses beyond the VaR threshold, providing a more complete picture of tail risk.

Economic interpretation: The difference (6.186% - 4.935% = 1.251%) represents the additional expected loss severity in the tail beyond the VaR threshold. This "tail risk premium" shows that when losses exceed VaR, they average 1.251 percentage points worse than the VaR threshold itself.

This is why ES is considered superior for risk management - it provides information about the magnitude of extreme losses, not just their probability threshold.

12. Why is the mean often ignored in daily VaR calculations?

Solution
The mean is often ignored in daily VaR calculations because:

- It's much smaller than the volatility component at daily frequency
- It's difficult to estimate accurately compared to volatility
- The error from ignoring it is typically very small
- For daily VaR, $\sigma \gg \mu$, so the volatility term dominates

For example, with a daily volatility of 1.5% and daily mean of 0.02%, the volatility component ($\sigma\Phi^{-1}(\rho)$) is approximately 75 times larger than the mean component for 1% VaR. Including the mean typically changes VaR estimates by less than 1%, making the approximation of zero mean practically acceptable for daily risk management.

13. A risk manager compares VaR estimates from three different models for the same portfolio:

Model	1-day 5% VaR
HS	\$450,000
Normal GARCH	\$420,000
t-GARCH ($\nu = 4$)	\$510,000

Explain why these estimates differ and which model would be most appropriate for stress testing purposes.

Solution

VaR estimates differ due to different distributional assumptions and methodologies:

HS (\$450,000): Uses empirical distribution of historical returns with no parametric assumptions, capturing actual tail behavior from past data.

Normal GARCH (\$420,000): Assumes conditional normality with thin tails, giving the lowest VaR but may underestimate tail risk.

t-GARCH (\$510,000): Uses Student's t-distribution with fat tails ($\nu = 4$ indicates significant excess kurtosis), reflecting heavier tail risk.

For stress testing, t-GARCH is most appropriate because:

- Stress testing requires conservative risk estimates
- Fat tails better capture extreme market conditions
- Higher VaR provides buffer against model risk

The t-GARCH model's \$90,000 higher estimate provides valuable capital buffer during stressed conditions when tail events are more likely to occur.

14. Explain what “burn-in time” means in the context of EWMA volatility models and why it is necessary for reliable risk forecasting.

Solution

Burn-in time is the initial period required for an EWMA model to converge to reliable volatility estimates after being initialized with an arbitrary starting value.

The EWMA model:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) y_{t-1}^2$$

Why burn-in time is necessary: The model requires an initial variance σ_1^2, but this choice is arbitrary. The influence of this initial condition decays exponentially at rate λ^t. With $\lambda = 0.94$, after 30 days only 16% influence remains. Using estimates before convergence leads to unreliable VaR calculations and poor risk forecasting.

Recommended practice: Use a burn-in period of approximately 30 days for $\lambda = 0.94$, waiting until $\lambda^t < 0.1$ (less than 10

15. Explain why financial risk measurements are only accurate to about 2 significant digits. If a risk model outputs a VaR of 7.8432%, what should be reported and why?

Solution

Financial risk measurements are only accurate to about 2 significant digits because risk is a latent variable that cannot be directly observed. Risk must be inferred from price movements using models, and the stochastic process of asset prices cannot be measured with high accuracy. Model assumptions introduce uncertainty, limited sample sizes for extreme events reduce precision, and structural breaks and regime changes affect stability.

For a VaR of 7.8432%, only 7.8% should be reported. The digits beyond the second significant digit (4, 3, 2) are not informative and create a false sense of precision. Reporting excessive precision can mislead users about the reliability of the risk measurement.

This limitation applies to all financial risk measures, not just VaR, because they all depend on inferring unobservable risk from observable price data through mathematical models.

Chapter 6

Analytical methods

1. Briefly explain the problem of asymmetry in bond risk analysis?

Solution
Bond prices exhibit positive convexity: symmetric yield changes produce asymmetric price changes favouring the bondholder.

When yields rise (say 3% to 5%), bond prices fall less than duration predicts. When yields fall equally (3% to 1%), bond prices rise more than duration predicts. This asymmetry occurs because the price-yield relationship is convex, not linear.

Mathematically: $\Delta P \approx -D \times P \times \Delta y + \frac{1}{2} \times C \times P \times (\Delta y)^2$, where convexity $C > 0$ always adds value regardless of yield direction.

2. What does the modified duration of a bond measure?

Solution
The modified duration of a bond is defined as the negative first derivative of the bond-pricing function divided by its price. It measures the price sensitivity of a bond to interest rate movements.

3. What factors affect the accuracy of duration-normal VaR for bonds?

Solution

Duration-normal VaR assumes bond prices change linearly with yields: $\Delta P \approx -D \times P \times \Delta y$. This breaks down for large yield changes where convexity matters.

Key accuracy factors: magnitude of yield changes (larger changes increase convexity error), bond duration (longer bonds have greater convexity), yield curve shape (parallel shift assumption rarely holds), and volatility level.

The method works best for short-maturity bonds with small yield movements. For a 10-year bond, a 100bp yield change can cause 5% pricing error from ignoring convexity.

4. What computational issue arises when using the Delta-Gamma method for option VaR?

   Solution
   We end up with a non-central chi-square distribution, which is computationally challenging to work with and requires specialized numerical methods to evaluate accurately.

5. Why are the methods in this chapter not recommended for most applications?

Solution
They are only based on approximations and Monte Carlo is usually much faster and easier to implement.

6. Explain why bond prices are more sensitive to interest rate changes for longer maturities. Include the concept of duration in your explanation.

Solution
Duration and Interest Rate Sensitivity:
Mathematical explanation: Bond price sensitivity to interest rate changes is measured by duration. For a bond with cash flows c_n at time n:

$$\text{Duration} = \frac{\sum_{n=1}^{N} n \cdot \frac{c_n}{(1+I)^n}}{P}$$

Why longer maturities are more sensitive:

(a) Time factor: Each cash flow is weighted by its time n in the duration formula. Longer maturity bonds have more cash flows at distant times, increasing duration.

(b) Present value effect: Distant cash flows are more sensitive to interest rate changes because they're discounted by $(1 + I)^n$ where n is large.

(c) Convexity: The relationship between bond prices and interest rates is convex. For longer maturities, this convexity is more pronounced.

Chapter 7

Simulation methods

1. Explain the trade-off between doing too few and too many simulations in Monte Carlo analysis. How should a risk manager determine the optimal number of simulations?

Solution
Too few simulations produce unreliable results with high sampling error and unstable risk estimates. Too many waste computational resources with diminishing accuracy gains — error decreases at rate $\frac{1}{\sqrt{S}}$, so quadrupling simulations only halves the error.

Optimal simulation count depends on convergence testing: start with 1,000 simulations, double repeatedly, and stop when VaR estimates stabilize within acceptable tolerance (typically 1-2% change).

Practical guidelines: use 5,000-10,000 for daily risk management, 10,000-50,000 for regulatory reporting, and 100,000+ for model validation. The baseline of 10,000 simulations works well for most applications, balancing accuracy against computational cost.

2. What is the benefit of simulations over analytical methods in quantitative financial analysis?

Solution
Simulations excel where analytical methods fail or become intractable. Path-dependent options like Asian or lookback options have no closed-form solutions, but Monte Carlo simulation handles them easily.

Simulations offer flexibility to model complex portfolios, incorporate jumps or regime switches, and test multiple scenarios simultaneously. Changing assumptions requires modifying parameters rather than rederiving mathematics.

While analytical methods provide exact solutions when available, simulations trade precision for versatility. The same framework prices exotic derivatives, measures portfolio risk, and performs stress testing without mathematical reformulation.

3. A risk manager wants to use Monte Carlo simulation to estimate the 1% VaR for a portfolio. Explain the key components of a pseudo-random number generator, discuss the concept of period in RNGs, and calculate how many simulations are theoretically possible using the Mersenne Twister RNG before the sequence repeats.

Solution

A pseudo-random number generator (RNG) has several key components. The deterministic algorithm is a function that generates the next number based on the previous one: $u_{i+1} = h(u_i)$ where u_i is the i-th random number and $h(\cdot)$ is the RNG function. The seed is the initial value that starts the sequence, with different seeds producing different sequences. The period is the number of unique numbers the RNG can produce before repeating. RNGs typically produce numbers uniformly distributed on $[0, 1]$.

Random number generators do repeat themselves - they do not create infinitely long sequences. This is because all practical RNGs are deterministic algorithms that eventually return to a previous state. Once an RNG returns to a previous internal state, it will produce exactly the same sequence from that point forward.

An example of a simple RNG is the Linear Congruential Generator (LCG): $u_{i+1} = (a \times u_i + b) \bmod m$ where a is the multiplier, b is the increment, and m is the modulus. Simple linear congruential generators have periods as low as $2^{31} - 1$ (about 2 billion), while poor quality RNGs may have periods as small as thousands or millions.

The Mersenne Twister has a period of $2^{19{,}937}-1$. To put this in perspective, $2^{10} = 1{,}024 \approx 10^3$, so $2^{19{,}937} \approx 10^{6{,}000}$. Comparing this to physical quantities, the age of the universe is approximately 1.4×10^{10} years and the number of atoms in the observable universe is approximately 10^{80}. This period is so astronomically large that repetition is never a practical concern in financial applications.

4. A portfolio consists of 2,000 shares of stock (current price \$35, daily volatility 2.5%) and 100 call options on the same stock (strike \$40, 2 months to expiration, Black-Scholes price \$1.80). Calculate the 1% VaR using Monte Carlo simulation, explaining how options modify the standard procedure.

Solution

Monte Carlo simulation for portfolios containing options requires modifications to the standard 6-step procedure to handle non-linear payoffs and complex option dynamics. The given parameters include 2,000 shares at \$35 with daily volatility of 2.5%, 100 call options with strike \$40, time to expiration $T = 2/12 = 0.167$ years, current Black-Scholes price \$1.80, and an assumed risk-free rate of 5% annually.

Step 1 is modified to calculate initial portfolio value as $\vartheta_t = x^b P_t + x^o g(P_t, X, T, \sqrt{250}\sigma, r)$ where $g()$ is the Black-Scholes pricing function. This gives $\vartheta_t = 2,000 \times \$35 + 100 \times \$1.80 = \$70,000 + \$180 = \$70,180$.

Steps 2 and 3 proceed as normal: simulate one-day returns $R_{t+1,s} \sim N(0, \sigma^2) = N(0, 0.025^2) = N(0, 0.000625)$ and calculate simulated stock prices $P_{t+1,s} = P_t \times (1 + R_{t+1,s}) = 35 \times (1 + R_{t+1,s})$.

Step 4 is critically modified to calculate simulated portfolio value as $\vartheta_{t+1,s} = x^b P_{t+1,s} + x^o g(P_{t+1,s}, X, T - \frac{1}{365}, \sqrt{250}\sigma, r)$. The key change is that time to expiration decreases by 1 day: $T - \frac{1}{365} = 0.167 - 0.00274 = 0.164$ years. For each simulation, the Black-Scholes formula must be re-evaluated using the new stock price and reduced time.

Steps 5 and 6 calculate profit/loss as $q_{t+1,s} = \vartheta_{t+1,s} - \vartheta_t$ and find VaR by sorting all $\{q_{t+1,s}\}_{s=1}^{S}$ values to identify the 1% quantile.

An example simulation illustrates the non-linear effects. Suppose $R_{t+1,s} = -0.06$ (6% stock decline). The simulated stock price becomes $P_{t+1,s} = 35 \times (1 - 0.06) = 35 \times 0.94 = \$32.90$. Using Black-Scholes with $P_{t+1,s} = \$32.90$, $X = \$40$, and $T = 0.164$, the option is deep out-of-the-money and the option price might fall to approximately \$0.20. The simulated portfolio value becomes $\vartheta_{t+1,s} = 2,000 \times \$32.90 + 100 \times \$0.20 = \$65,800 + \$20 = \$65,820$, creating a profit/loss of $q_{t+1,s} = \$65,820 - \$70,180 = -\$4,360$.

With 10,000 simulations, if the 100th smallest loss (1% of 10,000) equals -\$8,500, then the 1% VaR equals \$8,500.

5. Consider using Monte Carlo methods for forecasting risk. Outline how we forecast the risk of a portfolio consisting of two stocks and an option on one of them.

Solution
We use current prices to get the current value of the portfolio. We then use the covariance matrix of the stocks to simulate a bivariate normal random shock and use that to get simulated prices one day into the future, and hence calculate the simulated portfolio. We then subtract the initial portfolio value to obtain simulated profits or losses. We repeat that number of times (simulations) to obtain a vector of simulated profits or losses. Call the number of simulations S. If we indicate the VaR probability by p, the simulated VaR is the $p \times S$ value of sorted simulated profits and losses. This last step is the same procedure as in HS.

Chapter 8

Backtesting

1. Why is it important to evaluate the quality of a risk forecast model, and why is it difficult to use operational criteria for this evaluation?

Solution
Evaluating risk forecast models is crucial because they drive critical investment decisions and capital allocation. Poor models can result in excessive losses threatening institutional solvency. Regulators require comprehensive backtesting evidence for model approval under Basel regulations, and evaluation enables systematic comparison of competing approaches.

Operational criteria evaluation faces significant difficulties. Reliable assessment requires many years of data to observe sufficient extreme events, creating substantial delays. Market regimes change over time, and extreme events occur infrequently, providing insufficient data points for robust statistical testing. Risk models continuously evolve through updates, creating mismatches between evaluated and current versions. Backtesting addresses these challenges by using historical data to provide timely feedback without waiting years for operational outcomes.

2. Define the following key backtesting terms: estimation window, testing window, VaR violation, and violation ratio.

Solution
Estimation Window (W_E): The number of historical observations used to estimate risk model parameters, also called the training window. Typically ranging from 250 to 1,000 daily observations, it calibrates volatility models or builds historical simulation samples.

Testing Window (W_T): The out-of-sample period over which VaR forecasts are evaluated, containing days with both forecasts and actual outcomes. Typically ranges from 250 to 2,500 daily observations following the estimation period.

VaR Violation: An event where actual portfolio loss exceeds the VaR forecast, mathematically expressed as $q_t < -\mathrm{VaR}_t(\rho)$. For 1
Violation Ratio: The ratio of observed to expected violations, calculated as $VR = \frac{\text{Number of observed violations}}{\rho \times W_T}$. A ratio of 1.0 indicates perfect calibration, while ratios < 1 suggest conservative models and ratios > 1 indicate aggressive models. Statistical tests determine if deviations from 1.0 are significant.

3. Consider the coverage test used in backtesting VaR models. Explain how it works, identify its main advantage and main disadvantage.

Solution
The coverage test evaluates whether observed VaR violation rates match expected rates using a likelihood ratio test. The null hypothesis states that the true violation probability p equals the VaR confidence level ρ, while the alternative hypothesis allows them to differ. The test statistic is $LR_{CC} = -2\ln\left(\frac{L(\rho)}{L(\hat{p})}\right)$, where $\hat{p} = \frac{N}{W_T}$ is the observed violation rate over the testing window. Under the null hypothesis, $LR_{CC} \sim \chi^2(1)$ asymptotically, with rejection occurring when $LR_{CC} > \chi^2_{1,\alpha}$.

The main advantage is statistical rigor, providing a formal framework that removes subjectivity from model evaluation through clear acceptance/rejection criteria. This standardized approach quantifies whether deviations are statistically meaningful and is widely accepted in regulatory frameworks.

The main disadvantage is asymptotic approximation problems since the χ^2 distribution assumption requires large samples, but typical backtesting applications have small samples. For 1% VaR over 1,000 days, only $\approx$ 10 violations are expected, making the asymptotic approximation poor and reducing test power to detect model failures.

4. Explain why backtesting ES is significantly more challenging than backtesting VaR. Provide specific reasons and discuss potential approaches to address these challenges.

Solution
Unlike VaR, which can be directly compared to actual outcomes, ES cannot be directly observed or verified on any single day because it represents the average loss in the worst $\rho\%$ of cases. ES is not elicitable on its own (cannot be scored with a simple loss function) and depends on the full shape of the tail distribution, not just a single threshold breach.

However, ES is jointly elicitable with VaR (Fissler and Ziegel, 2016), enabling meaningful backtesting. The Acerbi-Székely (2019) test uses:

$$Z_{\text{ES}}(\text{ES}_t, \text{VaR}_t, q_t) = \text{ES}_t - \text{VaR}_t - \frac{1}{\rho}(q_t + \text{VaR}_t)_+$$

where $(x)_+ = \max(x, 0)$. This compares forecast ES with VaR plus the average excess loss beyond VaR.

The test has a bias term that is zero when VaR predictions are perfect and remains small when VaR accuracy is within $\pm 15\%$. Importantly, this bias is prudential — it makes the test more conservative, not more lenient. The test statistic follows $\mathcal{N}(0, 1)$ under correct forecasts, enabling standard hypothesis testing.

ES backtesting requires many more observations than VaR backtesting and critically depends on VaR prediction quality. When ES provides the same signal as VaR (when VaR is subadditive), VaR backtesting may be more reliable, explaining why regulatory frameworks often retain VaR despite ES's theoretical advantages.

5. What is the independence test in backtesting and what is its main disadvantage?

Solution
The independence test examines whether VaR violations cluster together over time, testing the null hypothesis that violations are independently distributed across time periods.

Main disadvantage:

The test makes a strong assumption about the pattern of clustering, specifically that violations only occur on consecutive days. However, during financial crises, markets often experience a different pattern: a large drop, followed by a correction, and then another large drop. This means violations might occur on days one and three rather than days one and two, which the standard independence test would fail to detect.

This limitation makes the test less effective at capturing the true clustering patterns that occur during periods of market stress.

6. Identify one reason why backtesting ES is more challenging than backtesting VaR.

Solution
When backtesting VaR, we can directly compare whether an actual return exceeds the VaR threshold — this provides a clear binary outcome (violation or no violation).

However, ES represents the expected value of losses in the tail beyond the VaR threshold. Since we typically observe only one actual return per day, we cannot directly measure whether ES is violated because there is no single observable outcome to compare it against.

Challenges with ES backtesting:

- ES is an expectation, not a threshold, so direct comparison is impossible
- We can use indirect methods such as normalized shortfall, but these require much more data to be statistically reliable
- The mathematical properties of ES backtests are more complex to derive and interpret
- Statistical tests for ES accuracy have lower power than equivalent VaR tests

This fundamental difference makes ES backtesting significantly more complex and data-intensive than VaR backtesting.

7. How would we apply backtesting to evaluate the quality of volatility forecast models?

Solution
Traditional backtesting cannot be applied to volatility forecast models because volatility is a latent variable that cannot be directly observed. We only observe returns, not the underlying volatility process that generated them.

8. Consider four GARCH(1,1) parameter estimates for daily stock returns:

Set	ω	α	β	$\alpha + \beta$
A	0.000010	0.05	0.90	0.95
B	0.000015	0.15	0.84	0.99
C	0.000030	0.08	0.90	0.98
D	0.000010	0.08	0.90	0.98

Which parameter sets would be most appropriate for (a) the 2008 financial crisis and (b) the calm period of 2004-2006? Explain your reasoning for each choice.

Solution
(a) Most appropriate for 2008 crisis: Parameter set B

Parameter set B's high α (0.15) captures markets' extreme sensitivity to news during crises, as 2008 featured dramatic daily reactions to bankruptcy announcements and policy changes. The high $\alpha + \beta$ (0.99) reflects near-unit root behavior common in financial crises, where volatility shocks persist for months creating long-lasting market memory effects.

(b) Most appropriate for 2004-2006 calm period: Parameter set A

Parameter set A's low α (0.05) reflects markets' muted reactions to news during the "great moderation" period, while moderate persistence ($\alpha + \beta = 0.95$) allows volatility shocks to decay at reasonable rates without crisis-level persistence. The low ω provides the lowest baseline volatility, consistent with the stable economic conditions of the mid-2000s low-volatility regime.

9. Consider two GARCH(1,1) parameter sets: Set A ($\omega = 0.000010$, $\alpha = 0.05$, $\beta = 0.90$) and Set B ($\omega = 0.000015$, $\alpha = 0.15$, $\beta = 0.84$). Calculate the unconditional variance and half-life of volatility shocks for each set. What do these calculations reveal about different market regimes, and what are the implications for backtesting?

Solution
Comparison and regime interpretation:

Measure	Set A (Normal)	Set B (Stressed)
Unconditional volatility	1.41%	3.87%
Half-life	13.5 days	68.7 days

Market regime implications:

The stressed regime has 2.7× higher long-run volatility than the normal regime, with 3.87% daily volatility translating to 61% annual volatility (crisis level) compared to 1.41% daily volatility translating to 22% annual volatility (normal level). Persistence differences are even more dramatic, as stressed regime shocks persist 5× longer than normal regime shocks, with crisis volatility shocks taking 69 days to halve versus 14 days in normal periods.

Risk management implications include that stressed regimes require longer risk horizon considerations, portfolio rebalancing frequency should differ between regimes, and VaR models need regime-dependent parameters. The economic interpretation reveals that normal periods feature efficient information processing and quick mean reversion, while stressed periods exhibit market fragility where shocks have lasting impact on volatility dynamics.

10. A bank's VaR model produces the following 1% daily VaR forecasts and actual P&L outcomes over 10 consecutive trading days (in millions):

Day	VaR Forecast	Actual P&L	Violation?
1	-2.5	1.2	No
2	-2.8	-3.1	Yes
3	-2.3	-0.8	No
4	-2.7	-2.9	Yes
5	-2.4	0.5	No
6	-2.6	-1.8	No
7	-2.5	-4.2	Yes
8	-2.9	-0.3	No
9	-2.2	-2.8	Yes
10	-2.8	1.8	No

Calculate the violation rate, perform the coverage test at 5% significance level, and interpret the results.

Solution
Violation identification: A violation occurs when actual loss exceeds the VaR forecast (actual P&L < -VaR). From the table, violations occur on days 2, 4, 7, and 9, giving 4 violations out of 10 observations.

Violation rate calculation: The observed violation rate is $\hat{p} = \frac{4}{10} = 0.40 = 40\%$. For a 1% VaR model, the expected violation rate is $\rho = 0.01 = 1\%$.

Coverage test: The likelihood ratio test statistic is $LR_{CC} = -2\ln\left[\frac{\rho^N(1-\rho)^{W_T-N}}{\hat{p}^N(1-\hat{p})^{W_T-N}}\right]$ where $N = 4$ violations and $W_T = 10$ days. Substituting: $LR_{CC} = -2\ln\left[\frac{0.01^4(0.99)^6}{0.40^4(0.60)^6}\right] = -2\ln\left[\frac{1\times10^{-8}\times0.941}{0.0256\times0.047}\right] = -2\ln(7.88\times10^{-6}) = 23.5$.

Critical value: At 5
Interpretation: The model fails the coverage test, indicating it significantly underestimates risk. With 40% violations versus the expected 1%, the model is clearly inadequate and requires recalibration or replacement.

Chapter 9

Extreme Value Theory

1. State and briefly describe the three types of tail distributions.

Solution

Weibull Thin tails where the distribution has a finite endpoint (e.g. the distribution of mortality and insurance/re-insurance claims)

Gumbel Tails decline exponentially (e.g. the normal and log-normal distributions)

Fréchet Tails decline by a power law; such tails are known as "fat tails" (e.g. the Student-t and Pareto distributions)

2. What is the formal definition of fat tails?

Solution

Regular variation A random variable, X, with distribution $F(.)$ has fat tails if it varies regularly at infinity; that is there exists a positive constant ι such that:

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\iota}$$

where ι is the tail index

3. What are the two common approaches to POT (peaks over thresholds)? Which approach would you prefer?

Solution

There are two common approaches to POT:

(a) The Generalized Pareto distribution or GPD (Fully parametric model)

(b) The Hill estimator (Semi-parametric model)

The choice between approaches depends on the specific application:

- GPD: More robust and theoretically grounded, but requires strong distributional assumptions
- Hill estimator: More flexible and requires fewer assumptions, but can be sensitive to threshold selection

In practice, the Hill estimator is often preferred for its simplicity and flexibility, though both approaches should be considered depending on the data characteristics and modeling objectives.

Chapter 10

Endogenous risk

1. Define the terms endogenous and exogenous risk, and provide examples of each.

   Solution
   Exogenous risk is where market participants do not influence market outcomes. An example would be an asteroid hitting the City of London or playing the roulette in a casino. Endogenous risk is where market participants do influence outcomes. An example is playing poker in the casino.

2. The former general manager of the BIS, Andrew Crockett stated in 2000:

> "The received wisdom is that risk increases in recessions and falls in booms. In contrast, it may be more helpful to think of risk as increasing during upswings, as financial imbalances build up, and materialising in recessions."

How does his view on risk relate to endogenous risk, and what is the implication for governments' policies on financial stability?

Solution
He is describing a form of endogenous risk. During economic upswings, market participants collectively engage in increasingly risky behavior, believing that good times will continue. This collective behavior creates systemic imbalances that are not visible in traditional risk measures. The risk builds up endogenously through the actions of market participants themselves, even though individual risk measures may appear low.

The implication for government policy is that governments should look at low measured risk as a potential predictor of future crises, rather than a sign of stability. This suggests the need for counter-cyclical policies - tightening regulation and supervision during boom periods when risk appears low, rather than waiting for crises to materialize. It also implies the importance of macroprudential regulation that looks at system-wide risks rather than individual institution risks.

3. It has been said that financial risk models are least reliable when we need them the most. Explain this statement.

Solution
Risk models perform well during calm markets when they're least needed — historical patterns hold, correlations remain stable, and volatility stays within normal bounds.

During crises when accurate risk measurement becomes critical, models fail systematically. Correlations converge to one as diversification disappears, extreme events far exceed historical calibrations, and endogenous feedback loops emerge as everyone reacts to the same signals. Liquidity evaporates, invalidating transaction cost assumptions.

This paradox — models work when unnecessary but fail when essential — arises because models are calibrated on normal market behaviour while crises represent structural breaks. The solution requires stress testing, scenario analysis, and maintaining scepticism about model outputs during market stress.

4. You run a Japan-oriented hedge fund with a target leverage of five. You own stock in only one company, where the price of one stock is ¥20,000. The number of stocks in your portfolio is 10. Your hedge fund is very small and you have no price impact when you trade. Suppose the price of the stock falls by ¥1,000. Making all necessary assumptions, how many stocks would be left in your portfolio after the necessary rebalancing?

Solution

Initially, with 10 stocks at ¥20,000 each, assets are ¥200,000. With target leverage of 5, equity equals ¥40,000 and debt equals ¥160,000.

After the stock price falls to ¥19,000, assets become ¥190,000 while debt remains ¥160,000, reducing equity to ¥30,000. To maintain leverage of 5, the fund must sell stocks and reduce debt.

Let Q_1 be the final number of stocks. The leverage constraint requires:

$$5 = \frac{19,000Q_1}{19,000Q_1 - [160,000 - 19,000(10 - Q_1)]}$$

Simplifying: $5 = \frac{19{,}000Q_1}{30{,}000}$, which gives $Q_1 = \frac{150{,}000}{19{,}000} = 7.89$ stocks.

Answer: 7.89 stocks would remain in the portfolio.

5. Starting with the previous question, suppose instead that your fund is large and exerts significant pricing power. In particular, for every ¥1,000,000 you trade, the price of the stock moves by ¥2,000 in the direction of the trade (up if you buy, down if you sell). Suppose the price of the stock falls for exogenous reasons by the same amount as in the previous question. Making all necessary assumptions, how many stocks would be left in your portfolio after the necessary rebalancing?

Solution
With the same initial conditions (10 stocks at ¥20,000, leverage 5, debt ¥160,000) and exogenous price drop to ¥19,000, the fund must solve an iterative problem since selling stocks causes additional price declines through market impact of -0.002 times value sold.

Starting with the previous answer of 7.89 stocks, selling 2.11 stocks worth ¥40,090 creates additional price impact of -¥80.18, reducing price to ¥18,919.82. Recalculating with this new price gives $Q_1 = \frac{150{,}000}{18{,}919.82} = 7.93$ stocks.

The second iteration confirms convergence at 7.93 stocks with final price ¥18,921.67. This demonstrates endogenous amplification where price impact created additional losses beyond the original shock, creating a deleveraging spiral where forced selling caused further price declines. The additional price decline of ¥78.33 (0.41%) shows how large players can destabilize markets through feedback loops.

Answer: 7.93 stocks would remain in the portfolio.

6. Consider the price of a typical stock and the VIX index. Are they mean reverting? Explain your reasoning and discuss the implications for risk management.

Solution
Stock prices are not mean reverting as they approximately follow a random walk with no tendency to revert to any particular level. Stock prices can theoretically reach any positive value or approach zero with no fundamental equilibrium price to return to. Unit root tests typically fail to reject non-stationarity, and autocorrelation functions show persistence rather than mean reversion.

The VIX index is mean reverting because it measures implied volatility, which has natural economic bounds. Volatility cannot remain extremely high or low indefinitely, with the VIX rarely falling below 10 or exceeding 80, typically oscillating around 15-20. The VIX shows strong negative autocorrelation at longer lags, with high levels typically followed by declines and a half-life of shocks around 2-6 months.

These properties have important risk management implications. For stock prices, losses may persist indefinitely with no guarantee of recovery, making diversification critical and stop-loss strategies potentially necessary. For the VIX, high levels may signal buying opportunities as volatility typically normalizes over time, and extreme VIX levels are usually temporary, affecting options strategies and crisis management approaches.

The different mean reversion properties require fundamentally different trading approaches: momentum strategies may work for stock prices while contrarian strategies suit the VIX. Understanding these properties is crucial for effective portfolio management and risk control.

7. Recall the case of the Long-Term Capital Management (LTCM) hedge fund and its investment strategies related to volatility. In the months before LTCM's default in 1998, the VIX was steadily rising, causing increasing distress for LTCM. Describe in detail why the rising VIX caused the default of LTCM and analyze this as an example of endogenous risk.

Solution
LTCM was betting on mean reversion in market volatility, taking positions that would profit from VIX declining back to normal levels after being considerably above its historical mean in summer 1998. The fund was essentially short volatility across multiple markets, expecting profits from volatility convergence trades based on models suggesting VIX levels were unsustainably high.

The rising VIX caused LTCM's distress through multiple mechanisms. As VIX continued rising, short volatility positions lost value through daily mark-to-market losses that accumulated rapidly and were amplified by leverage. Rising VIX triggered margin calls requiring additional collateral, creating increasing liquidity pressure. Unable to meet margin calls, LTCM was forced to liquidate positions at the worst possible time, realizing losses and contributing to further market volatility. The fund was right about mean reversion but wrong about timing, running out of liquidity before volatility normalized.

This exemplifies endogenous risk through self-reinforcing feedback loops where LTCM's distress contributed to market volatility, which worsened their position and created a vicious cycle. LTCM was not alone in volatility convergence trades, and collective distress led to widespread forced selling as individual rational strategies became collectively destabilizing. The crisis broke model assumptions as correlations increased dramatically and "25-sigma events" occurred multiple days in a row.

LTCM's failure demonstrates how market participants' collective actions create the conditions that cause their individual strategies to fail. When many participants adopt similar strategies in crowded trades, diversification benefits disappear and systemic risk increases. Risk management systems forced selling at the worst times through procyclical behavior, while interconnectedness meant LTCM's problems spread beyond the fund's actual size, requiring Federal Reserve intervention to prevent broader crisis.

Chapter 11

Regulation

1. Explain the key changes introduced by Basel III for market risk measurement compared to Basel II, focusing on the shift from Value-at-Risk to ES.

Solution
Basel III replaced Value-at-Risk with Expected Shortfall as the primary risk measure, shifting from 99% to 97.5% confidence levels. The framework introduced different liquidity horizons for different asset classes and separate capital requirements for non-modellable risk factors that cannot be reliably modeled.

ES was chosen because it is a coherent risk measure satisfying subadditivity, unlike VaR, and captures tail risk severity beyond the VaR threshold while properly accounting for portfolio diversification effects. This reduces incentives for risk measure manipulation and provides better regulatory stability.

Implementation challenges include increased model complexity requiring sophisticated techniques, more complex backtesting procedures, higher data requirements for historical stress periods, and greater computational burden. The shift generally resulted in higher capital requirements, particularly for portfolios with significant tail risk, requiring banks to upgrade risk management systems.

2. The Basel Accords are overseen by the Basel Committee on Banking Supervision (BCBS) and implemented by G20 countries. Explain the institutional framework and describe the key components of the Basel III capital requirements.

Solution
The institutional framework consists of the G20 (20 largest economies) coordinating global financial regulations, with all members committed to implementing Basel recommendations. The Basel Committee on Banking Supervision, hosted at the Bank for International Settlements in Basel, Switzerland, develops common standards but lacks formal supervisory powers, leaving implementation to member countries.

Basel III capital requirements include core components of Tier 1 capital (high-quality capital, primarily Common Equity Tier 1) and Tier 2 capital (supplementary instruments). Variable buffers include the capital conservation buffer and countercyclical buffer adjusted based on economic cycles. Additional requirements apply to systemically important banks through G-SIB capital and Total Loss-Absorbing Capacity (TLAC).

Capital instruments are designed for loss absorption, permanency, payment flexibility during stress, and freedom of operational action. Common Equity Tier 1 represents the highest quality (shareholders' equity), Additional Tier 1 includes convertible and perpetual bonds, while Tier 2 comprises subordinated debt and long-term bonds. These requirements evolved from Basel I's basic framework through Basel II's risk-sensitive approach to Basel III's post-2008 crisis enhancements, serving as buffers against unexpected losses and limits to excessive leverage.

3. A bank's trading book has the following VaR statistics over the past 250 trading days: current day 1% VaR = \$8 million, 60-day average VaR = \$10 million, and 6 violations occurred in the testing period. Calculate the market risk capital requirement under Basel II regulations and explain the traffic light system.

Solution
Calculation:

With 6 violations in 250 days, the multiplication factor is M = 3.0 + 0.2(6-4) = 3.4 (amber zone). Using max(\$8M current VaR, \$10M average VaR) = \$10M, the market risk capital = 3.4 $\times$ \$10M = \$34M.

The traffic light system penalizes VaR model violations through multiplication factors. Green zone ($\leq$ 4 violations) uses factor 3.0 with no regulatory concern since expected violations for 1% VaR over 250 days is 2.5. Amber zone (5-9 violations) increases the factor by 0.2 per excess violation, indicating potential model underestimation requiring supervisory attention. Red zone ($\geq$ 10 violations) applies maximum penalty factor 4.0, requiring model recalibration.

The system creates direct capital costs for model inadequacy. With 6 violations versus 4, the bank pays \$4M additional capital (13.3% increase), encouraging conservative model calibration. This graduated response provides regulatory flexibility while ensuring backtesting results directly impact capital costs, though it only considers violation frequency rather than magnitude.

4. The Swiss bank UBS failed in 2008 due to $19 billion in losses on collateralized debt obligations (CDOs) composed of U.S. sub-prime mortgages. Despite these massive losses, UBS's VaR models showed zero risk for these positions. Analyze this case study and explain how this illustrates the limitations of VaR as a risk measure.

Solution

UBS suffered $19 billion in losses on CDO positions despite VaR models showing zero risk. This occurred because VaR measures daily losses with 1% probability (99% confidence), but CDOs had high default probability over longer time horizons. Since daily loss probability was below 1%, VaR registered zero risk even though structural credit risk was enormous.

This case reveals fundamental VaR limitations. VaR typically uses 1-10 day horizons but credit risks materialize over months or years. The 1% probability threshold misses rare but devastating events beyond VaR scope. VaR was designed for market risk, not the embedded credit risk in CDOs, and models based on historical data failed to capture unprecedented CDO risk characteristics.

The failure was institutional as well as technical. Risk managers chose inappropriate tools, senior management failed to question zero risk readings, and external oversight collapsed as auditors and regulators accepted VaR measurements at face value. The case demonstrates how quantitative models created false security and risk management became a box-ticking exercise.

The UBS case catalyzed major reforms including Basel III's replacement of VaR with Expected Shortfall, stressed risk measures, and longer time horizons. Key lessons include that VaR should be one tool among many, not the sole measure, and that qualitative assessment must supplement quantitative models. The case emphasizes the need for healthy skepticism about model outputs and comprehensive risk frameworks beyond single metrics.

5. Basel III introduced significant reforms to the trading book regulations, replacing the 10-day 99% VaR with 97.5% ES calculated over various holding periods. Explain the key changes from Basel II to Basel III in market risk measurement and analyze the advantages and challenges of these reforms.

Solution
Basel III replaced Basel II's 10-day 99% VaR with 97.5% ES over multiple holding periods. ES captures tail risk beyond the VaR threshold, measuring expected losses given losses exceed the 97.5% quantile. The framework calculates ES directly for asset-specific holding periods based on liquidity, eliminating square-root scaling.

Key advantages of ES:

- Coherent risk measure satisfying subadditivity
- Captures magnitude of tail losses beyond threshold
- Stressed ES requirement using max(current ES, stressed ES)

Implementation challenges:

- Computational complexity requiring conditional expectation estimation
- Difficult backtesting as ES violations cannot be directly observed
- Higher capital requirements and significant systems upgrade costs

6. Explain why financial institutions are more heavily regulated than most other private firms. Discuss the specific externalities and information asymmetries that justify this regulatory treatment.

Solution
Financial institutions require heavy regulation due to systemic importance and market failures. Bank failures create contagion effects throughout the economy, unlike other industries where failures remain isolated.

Information asymmetries prevent depositors from assessing bank risk due to opaque balance sheets and complex derivatives positions. Bank failures generate severe negative externalities through credit crunches and payment system disruption.

Moral hazard arises from deposit insurance and too-big-to-fail policies, creating excessive risk-taking incentives. Additionally, maturity transformation — funding long-term assets with short-term liabilities — creates inherent instability requiring regulatory oversight to prevent bank runs and maintain financial stability.

7. Capital serves as a buffer against unexpected losses and as a limit to leverage. A commercial bank has the following simplified balance sheet (in billions):

Assets		Liabilities & Equity	
Cash	$50	Deposits	$800
Loans	$900	Bonds	$200
Securities	$100	Equity	$50
Total	$1,050	Total	$1,050

The bank expects loan losses of $20 billion over the next year, but actual losses could range from $10 billion to $40 billion. Analyze how capital serves as a buffer against unexpected losses and calculate the bank's leverage ratio.

Solution
Capital Buffer Analysis: Maximum unexpected loss = $40B - $20B = $20B. In the worst case, this reduces capital from $50B to $30B, allowing survival but with reduced cushion. Expected losses are absorbed by provisions; unexpected losses require capital.

Leverage Calculation: Current leverage ratio = $1,050B/$50B = 21. After $40B losses: assets = $1,010B, equity = $30B, leverage = 33.7. Basel III maximum leverage is 33.3, so worst-case losses would exceed regulatory limits.