Empirical Finance: Week 1 Study Guide

Handout 1: Introduction & Basic Concepts (Slides 1–97)

1 Introduction & Why Empirical Finance (Pages 1–8)

Professor's Key Teaching Points

- Finance is NOT an experimental science – you cannot run controlled experiments
- Finance is both art and science; empirical work is essential
- Historical examples matter: LTCM collapse (1998), Tulip Bubble (1636)
- Start with the research QUESTION first, then find the simplest tool (Occam's Razor)
- Never shop around for applications to fit a tool you learned – that's the wrong approach

Exam Learning Objectives

- Understand why empirical methods are critical in finance
- Know the correct research approach: Question → Simplest Tool

Key Formulas

None for this section – conceptual introduction only.

2 Simple (Discrete) Returns (Pages 9–41)

Professor's Key Teaching Points

- Simple returns are what you actually observe when you buy/sell assets
- The subscript convention matters: R_{t+1} means you observe the return when you SELL at $t+1$
- Interest rates from Bloomberg/Refinitiv are ALWAYS quoted as percentage per annum – you must scale them
- Day count conventions differ by country:
  - US, Germany, Italy: Actual/360
  - UK, Japan, Australia: Actual/365
  - Brazil: Business days/252
- "The devil is always in the details" – always check day count conventions on trading floors
- For this course: assume 30 days/month, 360 days/year (shortcut)
- Simple returns are CROSS-SECTIONALLY ADDITIVE (portfolio returns)
- The approximation $R^r_{t+1} \approx R_{t+1} - \Pi_{t+1}$ only holds for low inflation
- Taylor Rule: If inflation is 1% above target, raise interest rates by 1.5%

Exam Learning Objectives

- Calculate one-period and multi-period simple returns

- Understand why geometric average (not arithmetic) is used for average returns
- Convert nominal returns to real returns
- Understand riskless return notation (R_t^f not R_{t+1}^f) and why
- Calculate excess returns (equity premium)
- Adjust returns for dividends (cash vs reinvested)
- Calculate portfolio returns as weighted average of individual returns

Key Formulas to Remember

Gross Simple Return (One Period):

$$R_{t+1} = \frac{P_{t+1}}{P_t}$$

Net Simple Return:

$$R_{t+1} - 1 = \frac{P_{t+1} - P_t}{P_t}$$

Multi-Period Return:

$$R_{t,t+k} = \prod_{j=1}^{k} R_{t+j} = R_{t+1} \times R_{t+2} \times \cdots \times R_{t+k}$$

Average Return (Geometric Mean):

$$\bar{R} = \left[\prod_{j=1}^{k} R_{t+j}\right]^{1/k}$$

Real Return:

$$R_{t+1}^r = \frac{R_{t+1}}{\Pi_{t+1}} \approx R_{t+1} - \Pi_{t+1} \quad \text{(for low inflation)}$$

Riskless Return:

$$R_t^f = \frac{B_{t+1}}{B_t} \quad \text{(subscript } t \text{ because known at time } t\text{)}$$

Excess Return (Equity Premium):

$$R_{t+1}^e = R_{t+1} - R_t^f$$

Return with Dividends:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Portfolio Return:

$$R_{t+1}^p = \sum_{i=1}^{N} w_{i,t} \cdot R_{i,t+1} \quad \text{where } \sum_i w_{i,t} = 1$$

Day Count Conversion (US Example):

$$\text{Return} = \text{Quoted Rate} \times \frac{\text{Actual Days}}{360}$$

3 Log (Compounded) Returns (Pages 42–58)

Professor's Key Teaching Points

- Log returns are TIME ADDITIVE by construction – no approximation needed
- Simple returns are cross-sectionally additive; log returns are time additive
- With log returns, gains/losses are charged continuously (every millisecond)
- With simple returns, gains/losses are charged discretely (end of period)
- Capital R = gross discrete return; lowercase r = log return (always net)
- The VIX pricing builds exactly on the spread between discrete and log returns
- Volatility drag explains why average log return < average simple return

Exam Learning Objectives

- Convert between simple and log returns
- Understand time additivity of log returns
- Calculate multi-period log returns by summation
- Understand the volatility drag relationship
- Know when to use log vs simple returns

Key Formulas to Remember

Log Return (One Period):

$$r_{t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right) = \ln(R_{t+1})$$

Time Additivity (Two Periods):

$$r_{t,t+2} = \ln\left(\frac{P_{t+2}}{P_t}\right) = \ln\left(\frac{P_{t+1}}{P_t}\right) + \ln\left(\frac{P_{t+2}}{P_{t+1}}\right) = r_{t+1} + r_{t+2}$$

Multi-Period Log Return:

$$r_{t,t+k} = \sum_{j=1}^{k} r_{t+j}$$

Average Log Return:

$$\bar{r} = \frac{1}{k}\sum_{j=1}^{k} r_{t+j}$$

Log Return with Dividends:

$$r_{t+1} = \ln(P_{t+1} + D_{t+1}) - \ln(P_t)$$

Real Log Return:

$$r^r_{t+1} = r_{t+1} - \pi_{t+1} \quad \text{(exact, no approximation)}$$ ln

Volatility Drag (Taylor Expansion):

$$r_{t+1} \approx (R_{t+1} - 1) - \frac{1}{2}(R_{t+1} - 1)^2$$

Expected Log vs Simple Return:

$$E[r_{t+1}] \approx E[R_{t+1} - 1] - \frac{1}{2}\text{Var}(R_{t+1} - 1)$$

4 Stylized Facts of Financial Returns (Pages 59–68)

Professor's Key Teaching Points

- These are empirical regularities observed across many asset classes
- Fat tails: extreme events occur more often than normal distribution predicts
- Negative skewness: large negative returns more likely than large positive
- At aggregate level (monthly, quarterly), returns look more normal
- No serial correlation in returns, but volatility clusters
- Correlations are time-varying and increase during crises

Exam Learning Objectives

- Identify and explain the six stylized facts
- Understand implications for risk management
- Know why normal distribution is inadequate for daily returns

Key Formulas

This section is conceptual. Know the six stylized facts:

1. Fat tails (leptokurtosis)
2. Asymmetry (negative skewness)
3. Aggregate normality (CLT effect at lower frequencies)
4. Absence of serial correlation in returns
5. Volatility clustering (ARCH effects)
6. Time-varying correlations

5 Moments: Mean, Variance, Skewness, Kurtosis (Pages 69–85)

Professor's Key Teaching Points

- Always PLOT your data first before running any algorithm
- A single outlier can completely distort sample skewness and kurtosis
- Outliers are raised to power 3 (skewness) or 4 (kurtosis) – massive impact
- Outliers can be genuine (COVID crash) or fat finger errors
- Never use Python packages as a black box – know what you're doing
- Kim & White (2004) robust measures: SK2, KU2 are NOT distorted by outliers
- SK3, KU3 account for outliers but remove distortion
- SK4, KU4 if you want to keep outlier influence (genuine outliers)
- KU3 may have finite sample problems with < 500 observations
- "Unless I tell you to remember the formula, you don't need to memorise it"

Exam Learning Objectives

- Calculate sample mean, variance, skewness, kurtosis
- Understand why traditional measures are sensitive to outliers
- Know which robust measure to use in different situations
- Be able to plug numbers into formulas if given

Key Formulas to Remember

Sample Mean:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

Sample Variance:

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \hat{\mu})^2$$

Sample Skewness (Traditional):

$$\widehat{Sk}_1 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - \hat{\mu}}{\hat{\sigma}}\right)^3$$

Sample Kurtosis (Traditional):

$$\widehat{Ku}_1 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - \hat{\mu}}{\hat{\sigma}}\right)^4$$

Robust Skewness – Bowley (SK2):

$$Sk_2 = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Robust Skewness – Groeneveld-Meeden (SK3):

$$Sk_3 = \frac{\mu - Q_2}{E|x - Q_2|}$$

Robust Skewness – Pearson (SK4):

$$Sk_4 = \frac{\mu - Q_2}{\sigma}$$

Robust Kurtosis – Moors (KU2):

$$Ku_2 = \frac{(E_7 - E_5) + (E_3 - E_1)}{E_6 - E_2} - 1.23$$

where E_i are octiles

Robust Kurtosis – Hogg (KU3):

$$Ku_3 = \frac{U_{0.05} - L_{0.05}}{U_{0.5} - L_{0.5}} - 2.59$$

where U_α *= mean of top* $\alpha\%$, L_α *= mean of bottom* $\alpha\%$

Robust Kurtosis – Crow-Siddiqui (KU4):

$$Ku_4 = \frac{F^{-1}(0.975) - F^{-1}(0.025)}{F^{-1}(0.75) - F^{-1}(0.25)} - 2.91$$

6 Robust Estimation: OLS vs LAD (Pages 86–97)

Professor's Key Teaching Points

- OLS minimises sum of SQUARED errors – outliers have massive impact
- LAD minimises sum of ABSOLUTE errors – much more robust to outliers
- If you suspect outliers: run OLS, then check with LAD
- LAD = Least Absolute Deviations = Quantile Regression at median
- If OLS and LAD give very different results, investigate the outliers
- "Before you get to conclusions, there's lots of data cleaning underneath"
- Simulated examples show LAD line barely moves with outliers, OLS line shifts dramatically

Exam Learning Objectives

- Understand the difference between OLS and LAD objectives
- Know when to use LAD as a robustness check
- Interpret differences between OLS and LAD estimates

Key Formulas to Remember

OLS Estimator:

$$\hat{\beta}_{OLS} = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

LAD Estimator:

$$\hat{\beta}_{LAD} = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} |y_i - \alpha - \beta x_i|$$

Summary: What to Remember for the Exam

1. **Notation**: R_{t+1} for risky returns (observed at $t+1$), R_t^f for riskless (known at t)
2. **Simple returns** are cross-sectionally additive (portfolios)
3. **Log returns** are time additive (multi-period)
4. **Day count conventions**: Actual/360 (US), Actual/365 (UK)
5. **Interest rates** are always quoted per annum – must scale
6. **Volatility drag**: $E[r] \approx E[R-1] - \frac{1}{2}\text{Var}(R-1)$
7. **Taylor Rule**: Inflation 1% above target $\Rightarrow$ raise rates 1.5%
8. **Fisher Equation**: Real rate = Nominal rate − Expected inflation

9. **Robust measures**: SK2, KU2 for outlier-resistant estimation
10. **LAD** as robustness check against OLS when outliers suspected
11. **Six stylized facts** of financial returns
12. Always **plot data first**, check summary statistics before running algorithms