Lesson 5.3: The Limits of Prediction – Quiz

Digital Finance v4

Question 1

A time series is stationary if which of the following holds?

A. Its price always increases
B. Its mean, variance, and autocorrelation structure do not change over time
C. It has no missing values
D. It can be modeled with linear regression

Question 1

A time series is stationary if which of the following holds?

A. Its price always increases

B. Its mean, variance, and autocorrelation structure do not change over time

C. It has no missing values

D. It can be modeled with linear regression

Answer: B

Stationarity requires constant statistical properties over time. Financial prices are non-stationary (trending, volatility clustering), which is why returns are used instead.

Question 2

The autocorrelation of daily stock returns is typically:

A. Strongly positive (returns predict returns)
B. Near zero (returns are approximately unpredictable)
C. Strongly negative (reversals every day)
D. Exactly 1.0 (perfect predictability)

Question 2

The autocorrelation of daily stock returns is typically:

A. Strongly positive (returns predict returns)
B. Near zero (returns are approximately unpredictable)
C. Strongly negative (reversals every day)
D. Exactly 1.0 (perfect predictability)

Answer: B
Daily return autocorrelation is near zero, consistent with weak-form market efficiency. However, the autocorrelation of absolute returns is significantly positive (volatility clustering).

Question 3

A model trained on 2015–2019 data fails in March 2020. The most likely cause is:

A. The model had too few features
B. A regime change caused by the COVID-19 pandemic
C. The model was underfitting
D. The training data had too many rows

Question 3

A model trained on 2015–2019 data fails in March 2020. The most likely cause is:

A. The model had too few features

B. A regime change caused by the COVID-19 pandemic

C. The model was underfitting

D. The training data had too many rows

Answer: B

The training period was a low-volatility bull market. COVID created an extreme regime shift (VIX spiked from 15 to 82), producing data far outside the training distribution.

Question 4

You test 200 trading rules on the same dataset. Five show $p < 0.05$. What is this an example of?

A. Survivorship bias
B. Data snooping / multiple testing problem
C. Look-ahead bias
D. Successful alpha discovery

Question 4

You test 200 trading rules on the same dataset. Five show $p < 0.05$. What is this an example of?

A. Survivorship bias

B. Data snooping / multiple testing problem

C. Look-ahead bias

D. Successful alpha discovery

Answer: B

At $p = 0.05$, we expect $200 \times 0.05 = 10$ false positives by chance alone. Finding 5 significant results is consistent with random noise, not genuine signals.

Question 5

A backtest uses restated earnings data that was not available at the original decision time. This is:

A. Survivorship bias
B. Data snooping
C. Look-ahead bias
D. Selection bias

Question 5

A backtest uses restated earnings data that was not available at the original decision time. This is:

A. Survivorship bias
B. Data snooping
C. Look-ahead bias
D. Selection bias

Answer: C
Look-ahead bias occurs when future information (restated earnings) that was unavailable at the time is used in historical decisions, inflating backtest performance.

Question 6

Which form of EMH is most strongly supported by empirical evidence?

A. Strong form
B. Semi-strong form
C. Weak form
D. All three forms are equally supported

Question 6

Which form of EMH is most strongly supported by empirical evidence?

A. Strong form
B. Semi-strong form
C. Weak form
D. All three forms are equally supported

Answer: C
Weak-form efficiency (past prices cannot predict future returns) is well-supported. Semi-strong is debated, and strong form is rejected because insiders earn abnormal returns.

Question 7

Why does backtesting only on current S&P 500 members introduce bias?

A. The S&P 500 is too small
B. Survivorship bias – failed companies that were once members are excluded
C. The data is too recent
D. Index membership changes are random

Question 7

Why does backtesting only on current S&P 500 members introduce bias?

A. The S&P 500 is too small

B. Survivorship bias – failed companies that were once members are excluded

C. The data is too recent

D. Index membership changes are random

Answer: B

Companies that went bankrupt or were delisted were once in the index but are excluded from current-member backtests, overstating returns by $\sim$1–2% per year.

Question 8

A model achieves Sharpe ratio 4.5 in backtest. Your immediate reaction should be:

A. Invest immediately
B. Suspect overfitting, look-ahead bias, or missing transaction costs
C. The model is excellent – no further checks needed
D. Sharpe ratio does not matter

Question 8

A model achieves Sharpe ratio 4.5 in backtest. Your immediate reaction should be:

A. Invest immediately

B. Suspect overfitting, look-ahead bias, or missing transaction costs

C. The model is excellent – no further checks needed

D. Sharpe ratio does not matter

Answer: B

Sharpe ratios above 3.0 in backtests almost always indicate a methodological error. The best quant hedge funds sustain Sharpe ratios of 1.5–2.5 after costs.

Question 9

In financial sentiment analysis, what does NER stand for?

A. Neural Earnings Regression
B. Named Entity Recognition
C. Negative Equity Ratio
D. Natural Environment Risk

Question 9

In financial sentiment analysis, what does NER stand for?

A. Neural Earnings Regression
B. Named Entity Recognition
C. Negative Equity Ratio
D. Natural Environment Risk

Answer: B
Named Entity Recognition identifies and classifies entities (companies, people, amounts, dates) in text, enabling sentiment to be mapped to specific financial entities.

Question 10

Alpha decay occurs because:

A. Models run out of battery
B. Competitors discover and trade the same signal, compressing returns
C. Interest rates always rise
D. Alpha is a constant of nature

Question 10

Alpha decay occurs because:

A. Models run out of battery
B. Competitors discover and trade the same signal, compressing returns
C. Interest rates always rise
D. Alpha is a constant of nature

Answer: B
When multiple market participants exploit the same signal, their collective trading arbitrages away the edge. Publication of academic factors accelerates this decay.

Question 11

Why is random train/test splitting inappropriate for financial time series?

A. It takes too long to compute
B. It introduces look-ahead bias by allowing future data to leak into training
C. Financial data cannot be split
D. Random splits always produce better results

Question 11

Why is random train/test splitting inappropriate for financial time series?

A. It takes too long to compute

B. It introduces look-ahead bias by allowing future data to leak into training

C. Financial data cannot be split

D. Random splits always produce better results

Answer: B

Random splits break temporal ordering, allowing the model to train on data from 2023 and test on 2021. Walk-forward validation preserves time order and prevents leakage.

Question 12

What is the "sim-to-real gap" in reinforcement learning for trading?

A. The gap between simulation speed and real-time speed
B. The difference between simulated market conditions and actual live market behavior
C. The distance between a SIM card and a router
D. The time between market open and close

Question 12

What is the "sim-to-real gap" in reinforcement learning for trading?

A. The gap between simulation speed and real-time speed

B. The difference between simulated market conditions and actual live market behavior

C. The distance between a SIM card and a router

D. The time between market open and close

Answer: B

RL agents trained in simulation experience perfect fills, no slippage, and no market impact. Real markets have partial fills, latency, and order book dynamics the simulator ignores.

Question 13

Earnings surprise is defined as:

A. Earnings announced on an unexpected date
B. The difference between actual earnings and analyst consensus estimate
C. Earnings that are negative
D. Earnings reported in a foreign currency

Question 13

Earnings surprise is defined as:

A. Earnings announced on an unexpected date
B. The difference between actual earnings and analyst consensus estimate
C. Earnings that are negative
D. Earnings reported in a foreign currency

Answer: B
Earnings surprise measures how much actual results deviate from expectations. Positive surprises typically trigger bullish reactions, with post-earnings announcement drift (PEAD) lasting weeks.

Question 14

The Adaptive Market Hypothesis (Lo, 2004) differs from EMH by arguing that:

A. Markets are always inefficient
B. Markets are adaptively efficient – inefficiencies exist but are competed away over time
C. Only insiders can profit
D. Prices are completely random at all times

Question 14

The Adaptive Market Hypothesis (Lo, 2004) differs from EMH by arguing that:

A. Markets are always inefficient

B. Markets are adaptively efficient – inefficiencies exist but are competed away over time

C. Only insiders can profit

D. Prices are completely random at all times

Answer: B

Lo's framework uses evolutionary concepts: market participants adapt, creating and eliminating inefficiencies. Alpha exists temporarily but decays as competitors adapt.

Question 15

A fund discovers a profitable signal trading $10M. When scaled to $1B, returns collapse. Why?

A. The fund manager got less skilled
B. Market impact – large orders move prices, eroding the edge
C. Regulators intervened
D. The signal improved but returns didn't

Question 15

A fund discovers a profitable signal trading $10M. When scaled to $1B, returns collapse. Why?

A. The fund manager got less skilled

B. Market impact – large orders move prices, eroding the edge

C. Regulators intervened

D. The signal improved but returns didn't

Answer: B

Market impact scales approximately with $\sqrt{\text{trade size}}$. Small signals that work at $10M capacity become unprofitable at $1B due to price impact and reduced available liquidity.

Question 16

What is "reward shaping" in the context of RL for trading?

A. Giving bonuses to human traders
B. Designing the reward function the RL agent optimizes – wrong rewards lead to wrong behavior
C. Reducing trading costs
D. Shaping the stock price chart

Question 16

What is "reward shaping" in the context of RL for trading?

A. Giving bonuses to human traders

B. Designing the reward function the RL agent optimizes – wrong rewards lead to wrong behavior

C. Reducing trading costs

D. Shaping the stock price chart

Answer: B

If the reward is naive daily P&L, the agent may take excessive leverage. A well-shaped reward includes risk penalties, drawdown limits, and transaction cost awareness.

Question 17

The Loughran-McDonald dictionary is used in finance for:

A. Translating financial reports to other languages
B. Dictionary-based sentiment analysis calibrated for financial text
C. Spell-checking SEC filings
D. Defining accounting terms

Question 17

The Loughran-McDonald dictionary is used in finance for:

A. Translating financial reports to other languages
B. Dictionary-based sentiment analysis calibrated for financial text
C. Spell-checking SEC filings
D. Defining accounting terms

Answer: B
Loughran-McDonald (2011) created a word list specifically for financial sentiment. General dictionaries misclassify finance-specific terms (e.g., "liability" is negative in general but neutral in finance).

Question 18

LTCM's 1998 collapse primarily demonstrated the failure of:

A. Internet technology
B. Models that assumed normal distributions and ignored tail risk under extreme leverage
C. Index fund investing
D. Government regulation

Question 18

LTCM's 1998 collapse primarily demonstrated the failure of:

A. Internet technology

B. Models that assumed normal distributions and ignored tail risk under extreme leverage

C. Index fund investing

D. Government regulation

Answer: B

LTCM used 25:1 leverage with models assuming normally distributed returns. The Russian default created a tail event the models never anticipated, causing $4.6B in losses.

Question 19

Volatility clustering means:

A. Volatility is constant over time
B. Large price moves tend to be followed by large moves, and small by small
C. Volatility only occurs in clusters of stocks
D. Volatility is unpredictable

Question 19

Volatility clustering means:

A. Volatility is constant over time

B. Large price moves tend to be followed by large moves, and small by small

C. Volatility only occurs in clusters of stocks

D. Volatility is unpredictable

Answer: B

Volatility clustering is captured by GARCH models and shows up as significant positive autocorrelation in absolute or squared returns. It is one of the few predictable features of financial markets.

Question 20

A colleague shows you an ML model with 95% accuracy predicting stock direction. The most important next question is:

A. "What GPU did you use?"
B. "What is the out-of-sample performance using walk-forward validation with realistic costs?"
C. "How many features does it have?"
D. "Can I invest my savings?"

Question 20

A colleague shows you an ML model with 95% accuracy predicting stock direction. The most important next question is:

A. "What GPU did you use?"
B. "What is the out-of-sample performance using walk-forward validation with realistic costs?"
C. "How many features does it have?"
D. "Can I invest my savings?"

Answer: B
In-sample accuracy is meaningless without out-of-sample validation. Walk-forward testing with realistic transaction costs, slippage, and market impact separates genuine signals from overfitting artifacts.