Quantitative Investing

Interview preparation summary · HF track

IN THIS SUMMARY

- Signals, data and statistics
- Backtests, portfolios and live trading

INT.HU.1.a · Signals, data and statistics

Explain what systematic quantitative investing is, and define, score and combine signals (value, momentum, quality, low risk, carry) into one cross-sectional score.

Quantitative investing uses rules and data rather than a manager's judgement on each company, so that a model ranks a broad universe and a portfolio is built from the ranking. Equity quant funds trade hundreds or thousands of stocks with small bets, while multi-asset quant funds apply the same approach to futures, currencies and bonds. A signal is a number for each asset that predicts its return relative to others. The standard families are value (cheap on earnings, book or sales), momentum (strong recent returns, usually over 12 months excluding the latest month), quality (high profitability and stable earnings), low risk (low volatility or beta) and carry (the return from holding the asset if prices do not move). To score a signal, define it precisely, take it for each asset on a date, remove extreme values, standardise it across the universe into a z-score (value minus the average, divided by the standard deviation) and often neutralise it within sectors. Scores are then combined, with equal or chosen weights, into one composite. Combining helps most when signals are not highly correlated. The principle is diversification of ideas, as in Risk and Portfolio Construction.

WORKED EXAMPLE: SCORING TWO SIGNALS FOR SIX STOCKS

A quant researcher ranks six invented stocks on earnings yield (value) and 12-1 month momentum. Earnings yields are 8.0%, 5.0%, 6.5%, 3.0%, 4.5%, 9.5% and momentum figures are -5%, 25%, 10%, 30%, -10%, 5%, for Ardent, Brisa, Calder, Dovey, Elmar, Fenn. Build a composite z-score with equal weights and say which stocks the model would buy and sell.

Answer: The composite ranks Fenn, Brisa, Calder, Dovey, Ardent, Elmar, so the model buys Fenn (+0.64) and Brisa (+0.30) and sells Elmar (-1.02). Each signal is first put on the same scale so that neither dominates.

> **Interview tip:** Say that a z-score puts different signals on one scale, and that a composite is a weighted average. Explain why you combine signals that are not highly correlated.

INT.HU.1.b · Signals, data and statistics

Describe the quant research process and its data: point-in-time data, cleaning, outliers and corporate actions, and the promise and limits of alternative data.

Research follows a loop: an idea from theory or data, a precise definition, a test on past data, a check of whether the effect survives costs and out-of-sample periods, and then a small live trial. Data decides the quality of the result. Prices must be adjusted for splits and dividends, or a split looks like a crash. Fundamental data must be point-in-time, meaning that only what was known on each date is used, with the delay between the end of a period and its publication, and without later restatements. Identifiers must follow a company through name changes and mergers. Outliers, such as a decimal-point error, can distort a standardised signal, so values are checked, capped or winsorised at fixed percentiles. Missing values are handled with a rule rather than a fill that uses the future. Alternative data, such as card spending, web traffic, satellite images or job postings, may contain information before it appears in accounts, but it often covers part of the market, has a short history, changes its coverage and costs money, so it must pass the same tests. Data engineering and governance are a large part of the work. See AI in Hedge Funds and Asset Management for the wider issues.

WORKED EXAMPLE: A DECIMAL SLIP IN A VALUE SIGNAL

A researcher standardises the book-to-price ratio of ten invented stocks: 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3 and 13.0. The last value should be 1.3 and was entered with a misplaced decimal point. What does the error do to the scores, and what is the right control?

Answer: The error pushes the stock at 1.2 from +0.95 to -0.25 and the stock at 0.5 from -1.61 to -0.44, because the standard deviation rises from 0.27 to 3.64. A winsorisation rule and range checks would have caught it.

Interview tip: Describe both checks: that the numbers are right, and that the data was available on the date used. A good answer names one example of each failure.

INT.HU.1.c · Signals, data and statistics

Judge whether a signal is real with the information coefficient, t-statistics, standard errors, the Sharpe ratio and a correction for multiple testing.

The information coefficient (IC) is the correlation between a signal and the returns that follow, measured each period across the assets. An average IC of 0.03 to 0.05 is typical for a useful signal, and it looks small because markets are noisy. To judge it, take the average IC across periods, the standard deviation of the IC, and the standard error, which is that standard deviation divided by the square root of the number of periods. The t-statistic is the average divided by the standard error; values of about 2 or more are the usual starting point, but not enough on their own. The Sharpe ratio is return above the risk-free rate divided by its volatility, and a rough rule links it to skill: the information ratio is about the IC times the square root of the number of independent bets a year (the fundamental law of active management). The big risk is multiple testing. If you test 100 ideas, a few will look good by chance, so the t-statistic must be raised, for example with the Bonferroni rule, which divides the significance level by the number of tests, or the result must be confirmed out of sample. Use the probability tools from Probability and Market-Making Maths where needed.

WORKED EXAMPLE: IS A 3.1 T-STATISTIC ENOUGH

A signal has an average monthly IC of 0.04 with a standard deviation of 0.10 over 60 months. The researcher tried 100 variants and reports the best. What is the t-statistic, what is the chance that the best of 100 useless signals would look this good, and what threshold would a 5% Bonferroni test use? If the signal could be applied to 400 independent bets a year, what information ratio would the IC suggest?

Answer: The t-statistic is 3.10, but with 100 trials the chance of such a result from luck is 18% and the Bonferroni threshold is about 3.48, so the signal is not yet proved. It needs out-of-sample confirmation.

Interview tip: Always say how many ideas were tried. A t-statistic without the number of trials is not evidence.

INT.HU.2.a · Backtests, portfolios and live trading

Identify the pitfalls of a backtest (look-ahead bias, survivorship bias, overfitting, transaction costs) and design out-of-sample tests that expose them.

A backtest simulates how a rule would have performed. Its results are only as good as its honesty. Look-ahead bias means using information that was not available on the date, such as accounts before they were published or index membership known only later. Survivorship bias means using a universe of companies that exist today, so failures are missing and returns look better. Overfitting, or data mining, means tuning the rule until it fits the past, so that part of the result is noise; the more parameters and variants tested, the bigger the risk. Transaction costs, including spreads, commissions, market impact and the cost of borrowing shares to sell short, can turn a good gross result into a poor net result, especially when turnover is high. Good practice is to define the rule before looking at results, record the number of variants tried, charge realistic costs, and hold out data that is never used for choosing the rule. A walk-forward test chooses parameters on past data and trades the next period, and a final test on a later period is used only once. A rule with few parameters and a clear reason is more likely to survive. See Probability and Market-Making Maths for the sampling ideas.

WORKED EXAMPLE: FROM A GROSS BACKTEST TO A BELIEVABLE NUMBER

A backtest shows a gross return of 6% a year with volatility of 5%. The strategy trades a total of 4 times its capital a year at a cost of 15 basis points per unit traded. In a later period, kept aside, the net Sharpe ratio was 0.60. How much of the result survives?

Answer: Costs lower the Sharpe ratio from 1.20 to 1.08, and the held-out period gives 0.60, a haircut of 44%. A researcher should quote the out-of-sample number and the number of variants tested.

> **Interview tip:** Walk through the list of biases and say which one you would check first for the result in front of you. Quote net, out-of-sample figures.

INT.HU.2.b · Backtests, portfolios and live trading

Explain how a signal becomes a portfolio: risk models, optimisation, neutralisation, turnover and cost control.

A portfolio constructor turns scores into weights while controlling risk and cost. A risk model describes how assets move together, normally through factors such as market, sector, size, value and momentum plus asset-specific risk, and it estimates the risk of any portfolio. Optimisation chooses weights that maximise expected return for the risk taken, subject to limits such as maximum position size, gross and net exposure, and liquidity. Neutralisation removes bets that the model did not intend: a sector-neutral portfolio has no net long or short in any sector, and a market-neutral portfolio has a net beta near zero, so that returns come from the signal rather than from the market or a sector. Turnover is the amount traded to move from old to new weights, and costs grow with it, so constructors add a penalty or a no-trade band to avoid small trades that cost more than they earn. A signal that flips every month needs a large alpha to pay for itself. Estimated risk models are themselves uncertain, so the portfolio should be tested for sensitivity. This builds on Risk and Portfolio Construction.

WORKED EXAMPLE: NEUTRALISING SECTOR BETS AND COSTING A REBALANCE

A model scores six invented stocks: S1 1.2, S2 0.8, S3 0.1 (all Tech, beta 1.2) and S4 minus 0.2, S5 minus 0.9, S6 minus 1.0 (all Retail, beta 0.8). Weights are proportional to the score with gross exposure of 100%. Then scores are demeaned within each sector. The current weights are S1 12%, S2 10%, S3 minus 5%, S4 0%, S5 minus 9% and S6 minus 18%, and trading costs 10 basis points per unit traded. What does neutralisation change, and what does the rebalance cost?

Answer: Neutralisation takes the sector net from +50% and -50% to zero and the beta from 0.20 to 0.00. The rebalance trades 65.6% of capital and costs about 6.6 basis points.

> **Interview tip:** Say what exposures you are neutralising and why, and that trades must earn more than they cost.

INT.HU.2.c · Backtests, portfolios and live trading

Assess alpha decay, capacity, crowding and execution costs, the promise and limits of machine learning, and explain a result and its risks clearly.

A signal's edge fades. Alpha decay is the loss of predictive power as time passes after the signal is formed or as more money copies it, so speed of trading matters: a signal with a short half-life loses value if trading is delayed. Capacity is the amount of money a strategy can run before its own trading costs remove the return. Market impact grows with the size of the trade relative to the volume, and a common estimate is that impact rises with the square root of the participation rate, so doubling the size raises the cost per trade by about 40%. Crowding occurs when many funds hold similar positions, which can reduce returns and create sudden losses when several sell together, as seen in quant funds in some past episodes. Execution uses algorithms to trade over time and reduce cost. Machine learning can find non-linear patterns in large data sets, but it has a high risk of overfitting, needs much data and is harder to explain, so it must pass strict out-of-sample tests. When explaining a result, state the evidence, the number of trials, the costs, the capacity and what would make the signal stop working. See AI in Hedge Funds and Asset Management for the wider picture.

WORKED EXAMPLE: HOW MUCH MONEY CAN THE STRATEGY RUN

A strategy earns 6% a year gross and trades 5 times its capital a year. At 500m of capital each trade is 2% of average daily volume, daily volatility is 2%, and market impact per unit traded is 0.8 times volatility times the square root of the participation rate. How much alpha is left at 500m and at 2,000m, and at what size does it fall to 3%?

Answer: Net alpha is 4.87% at 500m and 3.74% at 2,000m, and it falls to 3% at about 3,516m. A one-month delay costs about 21% of the signal if its half-life is 3 months.

> **Interview tip:** Say that impact rises with the square root of size, so alpha shrinks as assets grow, and name what you would do about it.

Summary of the Charta interview-preparation topic. Companies and figures in the examples are invented. Draft content, not yet expert reviewed.