Backtesting Series - Episode 1: Introduction

Introduction

In this article, we present the first episode of a series on backtesting. We want to provide the reader with an immersive analysis of this powerful tool and analyze all the processes that lead to its optimization and implementation. The series will consist of the following parts:

- **Part 1: Introduction**
- Part 2: Cross-Validation techniques
- Part 3: Portfolio Construction & Portfolio Optimization, Factor Models
- Part 4: Backtest Analysis
- Part 5: Introducing TCs, assessing capacity, preparing for production

In this episode, we are going to set the scene for the whole series, giving an overview of how a backtesting pipeline should look like, and give a thorough overview of the Sharpe Ratio, Information Ratio, and Information Coefficient. We are going to cover the non-trivial question of using linear returns vs log-returns and, finally, we are going to explain how we generated the data that we will use to test the robustness of the techniques we will propose in the upcoming episodes.

We will start from the basics by defining alpha. Alpha is a measure of investment performance, mostly used to describe returns that exceed the market or a benchmark index. For active management funds, generating alpha is a primary objective and shows the value added by a portfolio manager or a strategy above what could be gained just by following the market or other factors.

In the development of a trading strategy, the goal is to design a model that delivers indicators, called alpha signals, that are predictive of excess returns and use them to construct a portfolio. Alpha models are automated and predictive. They describe, or decode, some market relation and are designed to have some edge, which allows traders to anticipate the future with enough accuracy that, after allowing for being wrong at least sometimes and for the cost of trading, allow them to still make money. An alpha contains rules for converting input data to positions or trades to be executed in the financial markets.

Data and Look-ahead biases

The starting point for every strategy is data. The process of obtaining high-quality data and checking for potential problems is a tedious one but is also the necessary condition for a robust alpha model. We will not delve into data cleaning in this series, and we will assume that the data has been checked for correctness. What we want to spend some time on, however, are the biases that can occur during the research process and that can undermine the statistical robustness of the backtest.

One of the most common biases that can occur when building an alpha model is look-ahead bias. That happens when you allow future information to influence decisions that are meant to be made only using current or past data. An incorrect time labelling for example can lead to a look-ahead, or a research bug of inadvertently using

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

future data. In contrast, real-time production trading is free of such issues, since it is based only on data available up to the current moment without access to any future data. Researchers using machine learning techniques also can introduce lookahead bias. They may tune some hyper-parameters on the entire data sample and then use those parameters in the backtest. Hyper-parameters should always be tuned using only backward-looking data looking for a correct Train-Test splitting. Similarly, in the area of sentiment analysis, researchers should take note of vendor-supplied sentiment dictionaries that may have been trained on forward-looking data. Just like in backtesting, that can result in an overly optimistic or skewed sentiment classification because the dictionary has been informed by future events that shouldn't have been available at the time.

A more subtle source of lookahead bias can affect large quantitative portfolios that rely on the same forecasts over extended periods. Past trades done for this type of portfolios affect prices by market impact, making the forecast performance look better via a self-fulfilling prophecy. Indeed, if you later backtest a trading strategy on the same period, the results may be skewed. It is a consequence of past trades that have already influenced the prices, making it easier for the model to predict price changes. This creates a self-fulfilling prophecy where the model looks as if it's performing well, but in reality, it's mostly because your trades impacted the prices in your favor.

Moreover, correct lifespan information is important also for avoiding survival bias, a fallacy of running quant research on assets alive today but ignoring those no longer active. Imagine you're testing a trading strategy that invests in S&P 500 stocks. It's important to note that the composition of the S&P 500 today is not the same as it was 10, 20, or 30 years ago. Companies are added or removed from the index over time due to changes in performance, market capitalization, mergers, acquisitions, or bankruptcies. Now, if you backtest your strategy using only the companies that are currently in the S&P 500, you're testing how today's survivors would have performed over the past few decades. This approach ignores companies that were part of the S&P 500 in earlier years, such as those listed in 2000 but have since been removed due to poor performance or bankruptcy. By doing this, you're effectively excluding the "failures," leading to an unrealistic assessment of your strategy's historical performance. Remedies for this issue can be implementing an objective methodology for inclusion that can be implemented at any time or specifying a robust rule in the backtesting process for the event of a delisting.

Another major issue that can occur while working with your dataset is data snooping. It refers to the practice of extensively testing multiple strategies on the same dataset without a proper prior until you find one that appears to perform well. However, the strategy's apparent success is often due to chance rather than a real, generalizable pattern. Data snooping can lead to misleading conclusions and overfitting, where a model or strategy seems to perform well in historical tests but fails to work on new, unseen data. In other words, with data snooping your historical test set is not any more an actual test set and your statistical significance or p-values are not relevant anymore.

Defining the investment universe

An important step in developing an alpha model is choosing the target set of assets to be traded, called the investment universe. It is possible to restrict it along one or more dimensions such as asset class, geography, industry or individual instruments. The universe choice is typically driven by the coverage of the input data or the idea that is behind the alpha model, although signals can be designed and tuned specifically for a certain universe even if the data has wider coverage.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

Linear vs Log Returns

We think it is important, first of all, to define returns and their statistical properties. The question of whether returns are Gaussian is an empirical one. We all know assuming cross-sectionally gaussian returns is an approximation (a significant one, even) – we usually do it to make the analysis more tractable, and our returns better-behaved statistically. However, this still does not make them easily tractable in time series analysis. We first define linear returns r_i for a security i as:

$$r_i(t) = \frac{P_i(t+1) - P_i(t) + D_i(t\ +1)}{P_i(t)}$$

Where $P_i(t+1)$ and $P_i(t)$ denote the price of the security at time t and $t+1$ while $D_i(t+1)$ is a factor that adjusts for possible dividends or other sources of income. Also, linear returns are the ones that define the daily portfolio PnL Q_t through the dollar position P_t:

$$Q_t = P_{t-1} * \ r_t$$

On the other hand, we define log returns as:

$$\tilde{r}_i(t) = \log\big(1 + r_i(t)\big)$$

But why would we want to use log-returns, which are just a first-degree approximation? Even if we were to assume linear returns to be normally distributed, the cumulative total return is not normally distributed, and its distribution rapidly diverges from a Gaussian. Instead, log returns are additive over compounding. This implies that the log of the compound return is equal to the sum of the log returns in the same period, and if the log return is normal, so is the log of the compound returns. Moreover, if the returns are independent, the variance of the log of compound log return is equal to the sum of the variances. We define compounded net returns over time with the following common approximation:

$$\prod_t (r_i(t) + 1) - 1 = \exp\left(\sum_t \tilde{r}_i(t)\right) - 1$$

Over short-term horizon, where asset's returns are usually of order 1%, we can approximate the difference between log and linear returns as follows:

$$\tilde{r}_i(t) = r_i(t) + \frac{r_i(t)^2}{2} + O(r_i(t)^3)$$

This delivers a difference between linear and log returns that is of order 10^{-4}. Said that, going back to compounded returns, this allows us to make an even further approximation for short-term horizon:

$$\prod_t (r_i(t) + 1) - 1 = 1 + \sum_t \tilde{r}_i(t) - 1 = \sum_t \tilde{r}_i(t)$$

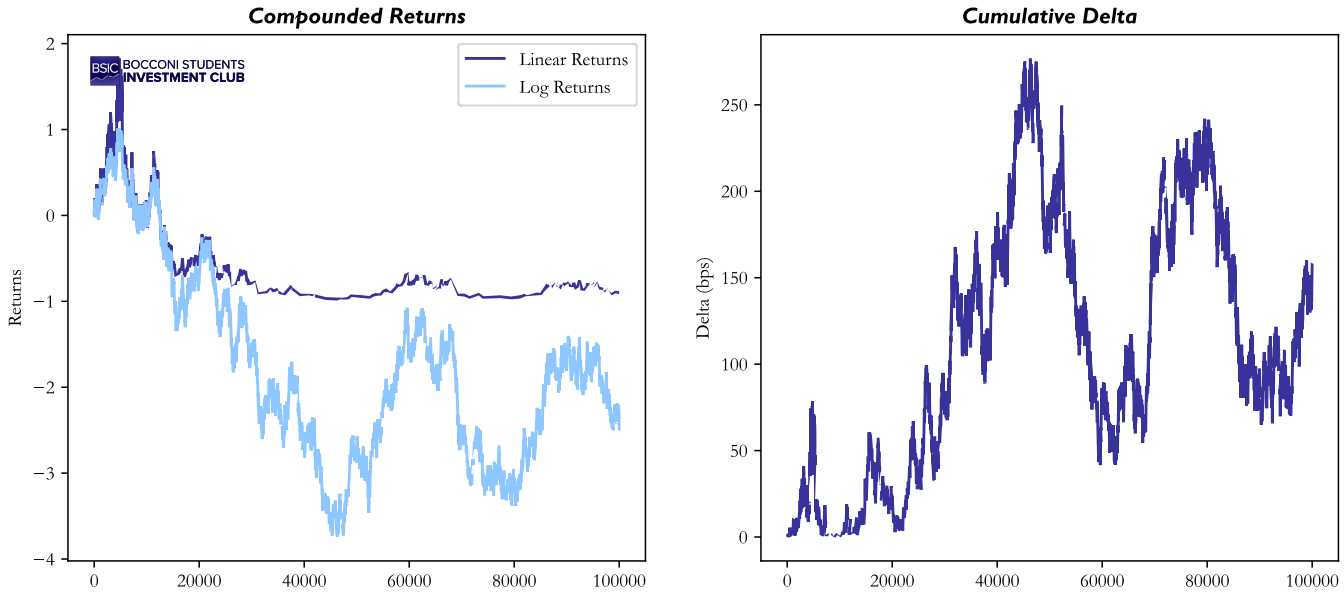
All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

Remark that is always important to verify the accuracy of this approximation and that can vary between different asset classes. For example, when the assets are equities, the approximation is usually considered adequate for daily interval measurements or shorter.

The different statistical properties between linear and log returns affect forecasting especially over longer horizons because the operators of (linear) expectation and (concave) log do not commute, even though the statistical distribution of log returns may have better mathematical properties than those of linear returns. Still, it is the linear return-based PnL that is the target of portfolio optimization. It is then common practice to construct a model on log returns, benefitting of their statistical properties, and then backtest it with linear returns.

To prove what we have explained above, we drew returns from an i.i.d. normal distribution centered at 0 and with std $\sigma = 0.01$. After computing linear compounded returns and log compounded returns this are the results we obtained:

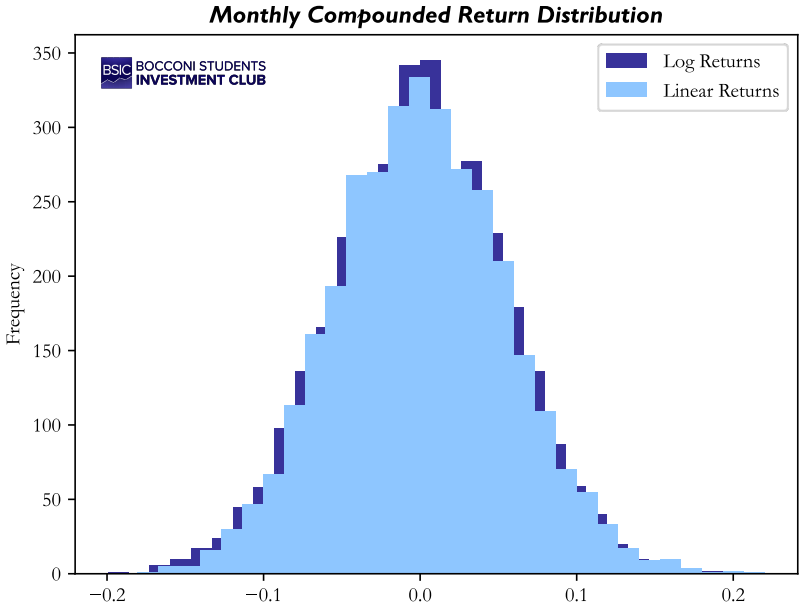


Source: BSIC

From the graph above appears that our approximation for compounded linear and log returns can hold for short period of time but after a while, they begin to diverge. Indeed, at the end of the process, the delta between the two of them reaches 156.6bps after it have reached spikes of 250+bps. But let's have a closer look to the distribution of monthly compounded returns to better understand the statistical properties of log and linear returns.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club



Source: BSIC

Since these plots are not explicable enough at first sight, we extended our analysis to better understand the change in normality between log and linear returns. We applied D'Agostino and Pearson's test, which combines skewness and kurtosis into an omnibus test of normality. The results revealed a p-value of 0.013 for the linear returns distribution, which is below the 0.05 significance threshold. This allows us to reject the null hypothesis of normality for the linear compounded monthly returns distribution. In contrast, the p-value for the log returns distribution increased to 0.441, indicating a significant improvement in normality. This aligns with the statistical properties of log returns we previously discussed.

Portfolio Optimization & Factor Models

Once you have your alpha signals for your universe of assets, you need a way to combine these signals together in an efficient way. This is the step of portfolio construction and optimization, which we are going to elaborate further on in an upcoming episode.

The most popular method is Mean Variance Optimization (MVO), where you maximize the expected return of your portfolio while minimizing the total variance. MVO has its own set of problems, namely that it can lead to sparse solutions (overweight on a handful of assets) and that it can perform very poorly out-of-sample. However, this should not necessarily be traced to a problem with the model itself – often, it is more of a problem with the two inputs: your expected returns and the covariance of the assets. Most importantly, estimation errors in the covariance matrix can be one of the main factors leading to the inferior performance of MVO. We are not going to elaborate on covariance estimation methods here, but one of the main techniques is constructing a factor model, which allows for a more robust estimation of the covariance matrix between assets. Factor models are a great tool, which can be helpful in portfolio construction/optimization, performance attribution, and risk management – we will have a dedicated episode of the series on factor models.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

Backtesting Practice

After constructing your strategy and optimizing your portfolio, it is time to evaluate it and backtest it on available data. We will introduce some of the most common practices that exist to do that, but we will delve deep into this argument in the next episode.

The most common evaluation scheme you can use is Cross-Validation, where the available data is split into a training dataset and a holdout dataset. We will explore two different types of Cross-Validation: K-Fold and Walk-Forward. K-Fold involves splitting the dataset into K equally sized subsets, or "folds". For each fold, the model is trained on the remaining K – 1 folds and validated on the remaining one. This process is repeated K times, with each fold being used exactly once for the validation. In the end, the performance metrics are averaged to provide a more robust estimate of the model's generalization ability. Walk-forward, on the other hand, is the most popular method when working with time series data, and works by splitting based on time. We use historical data up to time t and target returns at time $t + 1$ as our validation set; this process is then repeated expanding the training window forward in time, while the validation window moves step-by-step through the data.

Moreover, in the episode to come, we will confront these common evaluation schemes to a new methodology proposed by Giuseppe Paleologo (2024), called the Rademacher Antiserum. Using Rademacher Complexity, it measures the ability of a set of strategies to avoid overfitting by quantifying how much they covary with random noise. By comparing strategies against random noise, this method helps ensure that the chosen model generalizes well to unseen data, thus providing robust backtesting results. This technique is also able to deliver lower bounds to Information Ratio and Information Coefficient and differently from Cross-Validation, it uses all the data available at the same time, for all the strategies.

The Sharpe Ratio

In the following two paragraphs, we will introduce you to the main evaluation metrics we will use throughout the series, starting from the Sharpe Ratio, the most common measure of the performance of an investment strategy. Informally, the Sharpe ratio can be defined as the risk-adjusted expected excess return of a portfolio with respect to the risk-free rate over the standard deviation of this same excess return:

$$\zeta_{\text{basic}} = \frac{\hat{\mu}}{\hat{\sigma}}$$

Where $\hat{\mu}$ is the sample mean of excess returns and $\hat{\sigma}$ is the sample standard deviation. This basic estimator of the Sharpe Ratio is closely related to the Student's t-statistic for testing whether a random variable, in this case the returns, is significantly different from zero. The t-statistic is the Sharpe Ratio scaled by the square root of the number of trading periods n:

$$t = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} = \sqrt{n}\zeta_{basic}$$

A key problem of the estimator for the SR is that it is biased, meaning that its expected value is not equal to the true Sharpe Ratio ζ. Indeed, the expected value of the basic Sharpe estimator is greater than the true Sharpe Ratio resulting in ζ_{basic} overestimating ζ. Below is the relationship that links unbiased and biased Sharpe Ratio:

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

$$\zeta_{unbiased} = \frac{\zeta_{biased}}{bias\ factor}$$

It is possible to define the Bias Factor in relation to the number of trading periods n as follows:

$$\text{bias factor} = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

Where $\Gamma(x) = (x-1)!, \forall x \in N$ is the Gamma Function. Consult Mattia Rondato (2018) for a more in-depth analysis of the bias factor and the relationship between the expected value of ζ_{basic} and ζ.

It is important to note that as the sample size n increases, the bias factor approaches 1, so bias becomes smaller with more data. Also, a greater sample size reduces the variance of our basic estimator. On the other hand, variance increases as True Sharpe grows larger. Assuming normally distributed i.i.d excess returns, it is possible to derive confidence intervals for ζ_{basic}. We present two ways of doing that. The first one uses asymptotic normality, meaning that for large n, the distribution of the basic Sharpe ratio estimator becomes approximately normal by the Central Limit Theorem. The asymptotic confidence interval is:

$$\left(\hat{\zeta} - z_{\alpha/2} \cdot \hat{se}, \hat{\zeta} + z_{\alpha/2} \cdot \hat{se}\right)$$

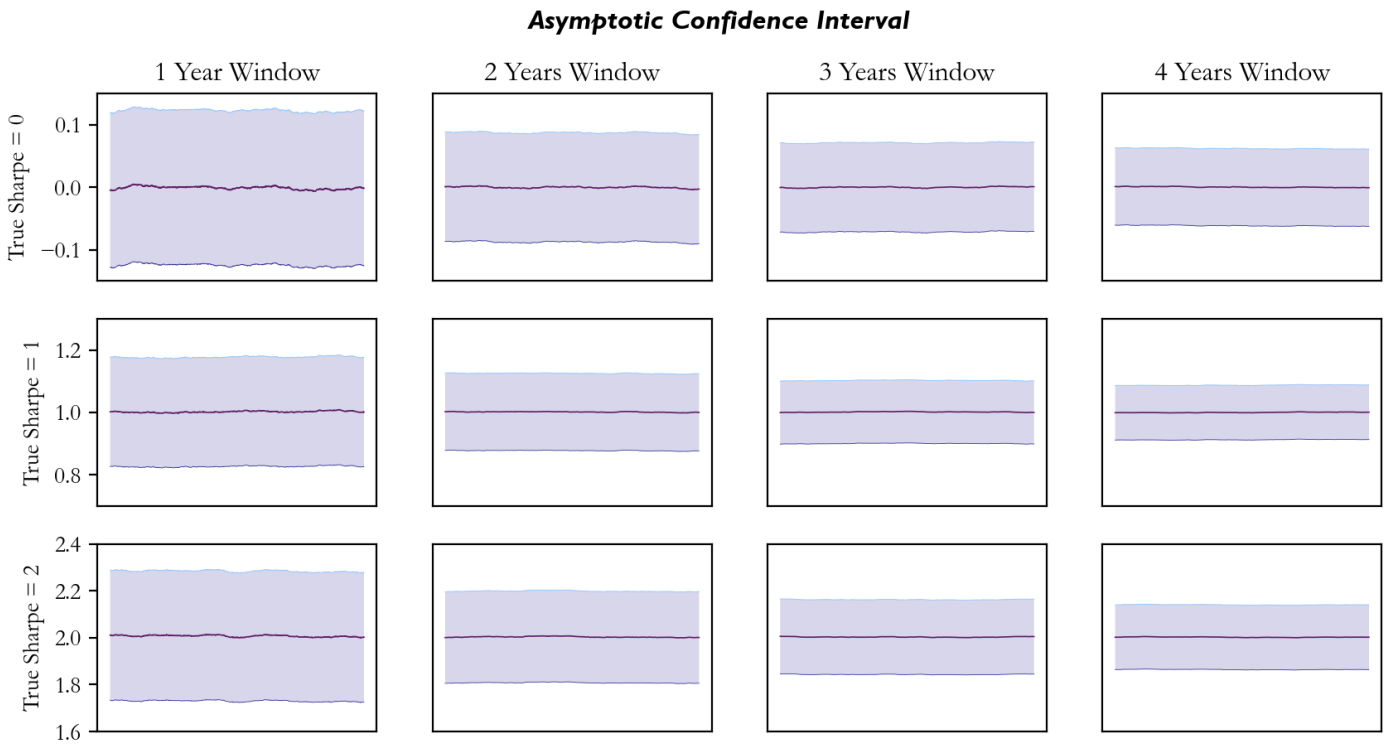
Where $\hat{\zeta}$ is the point of estimate of the true Sharpe, $z_{\alpha/2}$ is the standard normal quantile corresponding to the desired confidence level (e.g., 1.96 for 95% confidence) and $\widehat{se}$ is the standard error computed as $\sqrt{\frac{1+\widehat{\zeta^2}}{n}}$. A more robust method to estimate confidence intervals is using bootstrap, where samples are drawn from the original data to compute a distribution for the Sharpe ratio estimator. This method corrects for potential non-normality in the data, although is much more computationally expensive. It involves defining for each bootstrap sample a "studentized" statistic $T^{(b)} = \frac{\widehat{\zeta^{(b)}} - \hat{\zeta}}{\widehat{se^{(b)}}}$ where $\widehat{\zeta^{(b)}}$ and $\widehat{se^{(b)}}$ are the estimate of the Sharpe and the standard error of the bootstrap. Then after sorting the results in ascending order is possible to form an empirical distribution of the studentized statistic which will be used to define the confidence intervals as follows:

$$C = \left(\hat{\zeta} - T_{\alpha/2}\widehat{se}, \hat{\zeta} - T_{1-\alpha/2}\widehat{se}\right)$$

Here $T_{\alpha/2}$ and $T_{1-\alpha/2}$ represent the critical values from the sorted bootstrap studentized statistics while $\hat{\zeta}$ and $\widehat{se}$ are still respectively the point of estimate of the true Sharpe and the standard error from the whole sample. To give a more comprehensive picture, we decided to provide an empirical example of how these intervals actually behave. We simulated normally distributed excess returns from 1000 different strategies, each time implying different values of True Sharpe. Then we computed the asymptotic confidence intervals on rolling Basic Estimator ζ_{basic} over different window sizes. Due to computational reasons, we did not repeat the same test also for the bootstrap method. Averaging the results from all the 1000 strategies this is what we got:

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

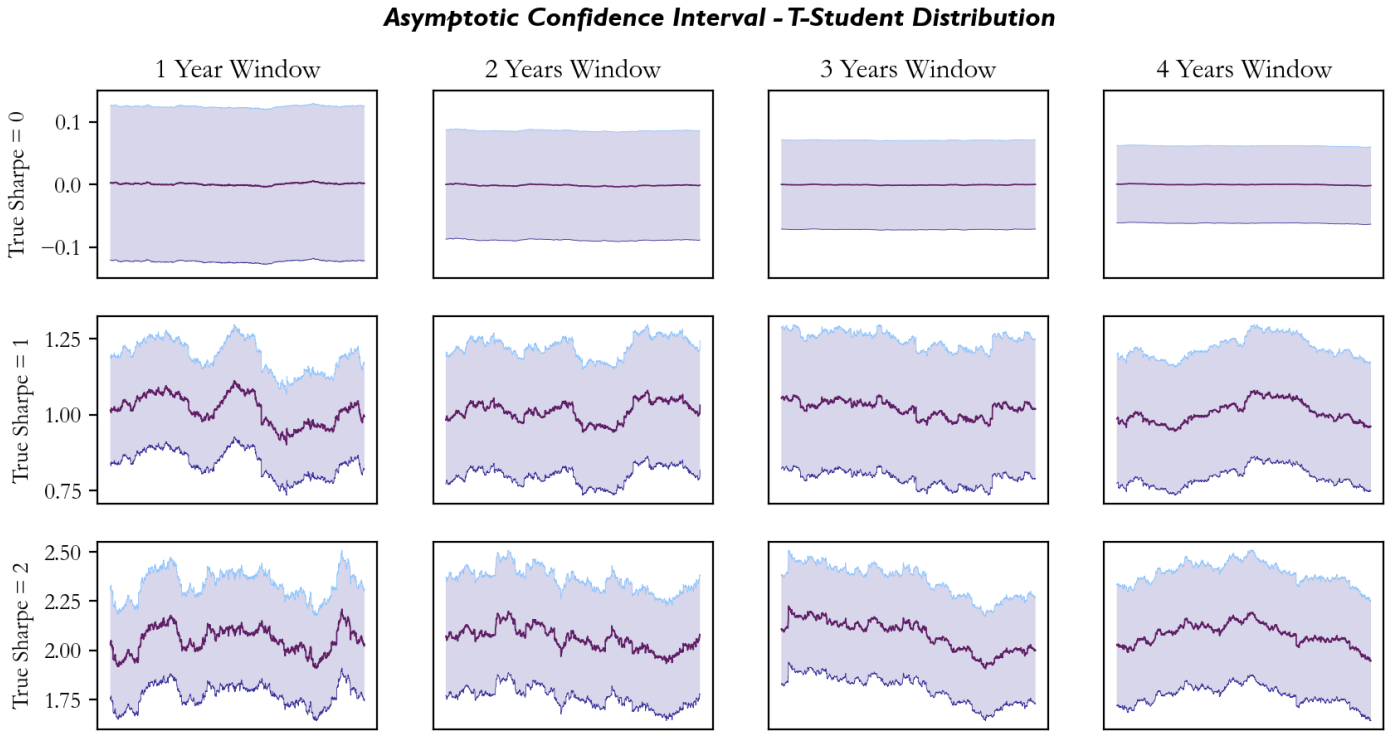


Source: BSIC

The empirical results over the Asymptotic confidence interval confirm our prior assumptions on the bias factor of ζ_{basic}. It appears clear that as the window size gets larger, the confidence interval gets narrower meaning the bias factor is reduced. Also, as True Sharpe grows larger, also the variance of our Basic Estimator increases. We repeated the same process over the same sample of randomly generated data, although this time we drew them from a T-Student distribution with degree of freedom equal to 5, accounting for more heavy-tailed returns. Here are the results:

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club



Source: BSIC

We can see that as soon as the normality assumption of excess returns is relaxed, and the in-sample distribution of Sharpe ratios does not converge to normal and moves from mean 0, the results tend to be more inconsistent. Indeed, the asymptotic confidence interval is based on the assumption that the sampling distribution of the estimator (e.g. the Sharpe ratio) approaches normality as the sample size increases allowing for the use of normal-based confidence interval. Using a more heavy-tailed distribution causes underestimation in the true variability of the estimator, especially in smaller windows and wider confidence intervals for small sample sizes, as the asymptotic normality assumption doesn't hold as well for heavy-tailed distributions. Moreover, it becomes clear how the Basic Estimator that we usually use as a performance metric can either underestimate or overestimate the True Sharpe Ratio.

Our personal take is that if you should take anything out of this, it is that you should be pessimistic about your Sharpes. Furthermore, once you have a decently high Sharpe, the estimation error makes it much more worth it to move your attention to other metrics of your strategy; first of all, controlling drawdowns. Sharpe should be your main priority only when you don't have it.

Information Ratio and Information Coefficient

The information ratio (IR) and information coefficient (IC) are two other metrics that can be used to evaluate the performance of our alpha signals. In this section, we are going to give an overview of these two measures, connect them to the Sharpe Ratio and our predictive model, and give some useful heuristics regarding what you can expect

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

when running a backtest on real data. We'll move in steps, from a more intuitive definition, to a more rigorous one, and to a proper mathematical definition of these two measures.

First of all, in simple terms, the IR can be seen as a generalization of the Sharpe Ratio; as we will see below, setting our benchmark as the cash portfolio (thus earning the risk-free rate) makes the IR equivalent to the SR. The IC, on the other hand, is a more direct measure of the predicting power of our alpha signal.

Before giving a more rigorous definition of IR and IC, we must briefly introduce the concepts of residual returns/alphas (we will use these two terms interchangeably). Alpha is a rather loose term; the most general definition is that alpha is any excess return over the return of a reference instrument. To give a mathematical "form" to this excess return, we need to decompose the PnL of our portfolio in two parts – the one explained by the benchmark, and the one orthogonal to it. A linear regression is the easiest way to do it – but even then, we can distinguish 3 main types of alphas, which differ in what we consider as our reference instrument.

$$\begin{cases} r_p = \alpha + \beta r_b + \varepsilon \ (Benchmark\ \alpha) \\ r_p = \alpha + \beta r_M + \varepsilon \ (CAPM\ \alpha) \\ r_p = \alpha + \beta_1 f_M + \beta_2 f_2 + \cdots + \beta_k f_k + \varepsilon \ (Multifactor\ \alpha) \end{cases}$$

For the benchmark alpha, we choose the index we are tracking as our reference instrument (e.g. Russell 2000); for the CAPM alpha, we use the market portfolio; for the multifactor alpha, we include all the relevant factors, which will also include the market factor, consistent with what you would do in a factor model.

There are a few conceptual differences in the 3 methods. The benchmark alpha is more relevant to active portfolio managers that track an index; however, the apparent outperformance (alpha) of a PM might just be due to the fact that he chose a bad benchmark. The CAPM alpha solves that, by setting the benchmark as the market portfolio. The multifactor alpha controls for factor exposures, which makes it for a "truer" estimate of PM skill, and for a purer estimate of idiosyncratic returns. That is because if we exclude factors from the regression, it might very well be that the excess returns of the PM were not pure "alpha", but came from harvesting factor risk premia – which is not necessarily a bad thing, but it's something that is arguably easier than being able to predict "true" idiosyncratic returns, orthogonal to known risk factors. A fair critic to this might be that your idiosyncratic returns are just factors you haven't discovered yet – which we don't necessarily disagree with. Regardless of what alpha measure I choose, I would expect my PM to have a positive, statistically significant alpha.

We now have the necessary ingredients to introduce the Information Ratio: using the results from the previous paragraph, the IR is defined as the expected annual residual return over the volatility of the residual return:

$$\text{IR} = \frac{E[\alpha + \varepsilon]}{\sqrt{Var[\alpha + \varepsilon]}} = \frac{\alpha}{stdev(\varepsilon)}$$

From this, we can easily derive that if we assume our benchmark to be the cash portfolio, thus earning the risk-free rate, our Information Ratio will be

$$\text{IR} = \frac{E[r - r_f]}{\sqrt{Var[r - r_f]}} \approx SR$$

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

It is important to make a distinction between the ex-ante IR and ex-post IR. When running a backtest, we are in the dream world of the ex-ante IR. We can see the ex-ante IR as the maximum IR we can get ex-post – that is, if we squeeze all the information out of our signal when we implement our portfolio. We can repeat the same reasoning for the SR. If our portfolio construction/optimization process is sub-optimal, our realized IR will be lower than the in-sample one. The delta between the two has a specific name: Information Loss (IL).

Moving on to the IC – the information coefficient is the correlation between our alpha signals and the returns of our asset. There is a very useful rule of thumb relating the IC to the IR, which is called the fundamental law of active management, which says

$$IR = IC \cdot \sqrt{Breadth}$$

Where breadth is the number of independent predictions we make. This should be taken more as a diagnostic measure rather than being used for portfolio construction, and basically says that we can improve our IR either by improving our model (IC) or by expanding our asset universe or how often we make predictions (the breadth). If we can improve one of the two without sacrificing the other, we should always do so. An interesting empirical question would be which one of the two is easier to improve: our intuitive take is that it is easier to find a high conviction signal that works for just a few assets (high IC, low breadth), rather than a "silver bullet signal" having lower conviction on many more assets (low IC, high breadth).

The IC sounds so similar to the R2 of a regression – a natural question is whether there is a mathematical link between the two. The answer is obviously yes: it can be proved that, in expectation, by running a cross-sectional regression of the asset returns on our alphas,

$$E[R^2] = (IC)^2$$

You might want to reflect on the equation above – it is very intuitively linked to both measures being "correlations". You might also find peace in the fact that there is a mathematical relationship between the IR and the R2 of our regression: in fact, it can be proven that

$$IR = \sqrt{R^2 n}$$

Where n is the number of assets in our cross-sectinal regression, while the annualized IR is a function of the per-period cross-sectional R2:

$$IR = \sqrt{R^2 nT} = IC\sqrt{nT}$$

Consult Paleologo, 2024 for a more thorough discussion of this relationship.

Now for a more practical question – what IC/IR can you expect when backtesting a strategy on actual data? Well, the answer is obviously it depends on the asset, but there are some heuristics.

First of all, remember we are in the financial world, where the signal to noise ratio is very low (I would make you a market around .1) – you reasonably can't expect to get an IC higher than the signal-to-noise, and if you do you are just fitting to noise. Secondly, if we agree that markets are close to efficient (if you don't, I have a 2/20 active fund to sell you), you wouldn't expect to find many signals that surmount transaction costs on their own. This

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

means that it's easier to find higher ICs for assets for which (i) liquidity is low, (ii) TCs/vol is high (think of swap spreads vs S&P Futures).

Taking a step aside, does this mean that you should scrap signals that do not surpass TCs on their own? Absolutely not! They can be useful when combined with other signals and can definitely improve the marginal SR of your portfolio, both when used in an active way (creating a composite signal) and in a passive way (conditioning rebalancing on the unprofitable signals).

Generating Fictional Data

We need a robust way to compare the different methods we are going to propose over the series. We choose to simulate data, for the following reasons: (i) we do not have enough independent (and statistically significant) signals to test on, and creating them is outside the scope of the series, (ii) generating signals allows us to have a wider sample, and to know the true statistics that we are trying to estimate (SR, IC, IR…), (iii) generating signals allows us to impose different levels of correlation between the target and the alpha signals, and between alpha signals themselves. We take the same approach as in Paleologo, 2024, which we explain below.

When generating alpha signals, we draw returns and signals from a multivariate normal distribution. We will tweak the vol of the asset, while keeping, without loss of generality, unit-variance for the signals (thus representing a z-score). We can set different correlations between the signals and the asset, and between the signals themselves. We can also generate strategy returns directly. In that case, we relax the assumption of normal returns and draw from a multivariate t-distribution; we will tweak the expected returns and volatilities of the strategies, along with the correlations between them.

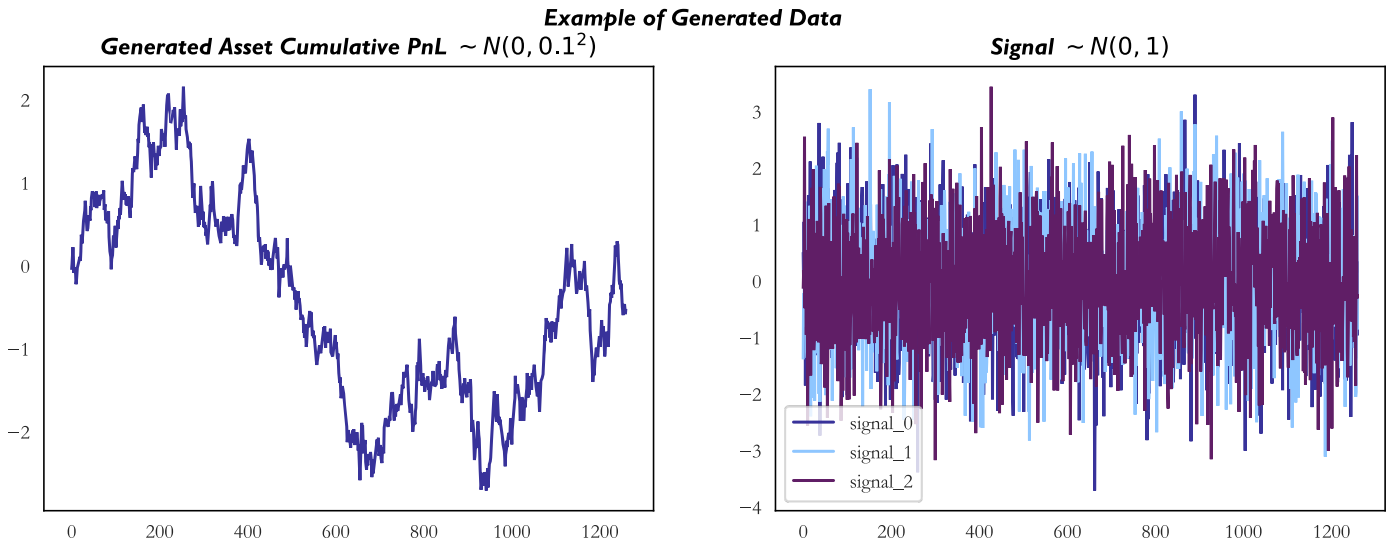
All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club

```python
def generate_covariance_matrix(
        n_signals: int,
        return_signal_correlations: Union[np.ndarray, list],
        returns_stdev: float = 1.,
        signals_stdev: float = 1.,
        signals_correlation: float = 0.
):
    # if the correlations are in a list convert to a numpy array
    if isinstance(return_signal_correlations, list):
        return_signal_correlations = np.array(return_signal_correlations)

    # initialize the covariance matrix
    covariance_mat = np.zeros((n_signals + 1, n_signals + 1), dtype=float)

    # initialize the covariance matrix between the signals
    signals_covariance = np.eye(n_signals, dtype=float) * signals_stdev ** 2 * (1-signals_correlation) +
np.ones((n_signals, n_signals), dtype=float) * signals_correlation * signals_stdev ** 2

    # insert in the final covariance matrix
    covariance_mat[1:, 1:] = signals_covariance

    # compute the asset-signal covariances
    returns_signal_covariance = return_signal_correlations * returns_stdev * signals_stdev

    # add the covariance between the returns and the signal
    covariance_mat[0, 0] = returns_stdev ** 2
    covariance_mat[0, 1:] = returns_signal_covariance
    covariance_mat[1:, 0] = returns_signal_covariance

    return covariance_mat

def generate_asset_signal(n_periods: int, means: np.ndarray, covariance: np.ndarray, return_df: bool = False):
  # check the means and covariance have consistent shape
  assert means.shape[0] == covariance.shape[0], "Mean and covariance must have the same shape"

  generated_data = np.random.multivariate_normal(means, covariance, n_periods)

  # if returning a df, assign the column names and return the dataframe
  if return_df:
      df = pd.DataFrame(generated_data, columns=['asset'] + [f'signal_{i}' for i in range(means.shape[0] - 1)])
      return df

  # return the generated numpy array
  return generated_data


def generate_strategy_returns(n_periods: int, mean: np.ndarray, covariance: np.ndarray, df: int = None):
    if df is not None:
        distr = multivariate_t(mean, covariance, df)
        returns = distr.rvs(size=n_periods)
        return returns

    returns = np.random.multivariate_normal(mean, covariance, n_periods)
    return returns
```

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club



Source: BSIC

Conclusions

In this first episode of our Backtesting series, we've laid the groundwork by introducing key concepts. Understanding these fundamentals is crucial for building a robust backtesting pipeline. In the next episode, we will dive deeper into cross-validation techniques, exploring how to optimize the training and testing of your models to ensure more reliable results. Stay tuned as we discuss K-Fold and Walk-Forward validation methods, along with a novel approach using Rademacher Complexity, all of which are essential for refining your strategies and avoiding common pitfalls in backtesting.

References

[1] Grinold, Kahn, "Active Portfolio Management", 1999

[2] Paleologo, Giuseppe, "The Elements of Quantitative Investing", 2024

[3] Paleologo, Giuseppe, "Advanced Portfolio Management: A Quant's Guide for Fundamental Investors", 2021

[4] Isichenko, Michael, "Quantitative Portfolio Management: The Art and Science of Statistical Arbitrage", 2021

[5] Chincarini, Kim, "Quantitative Equity Portfolio Management", 2006

[6] Pav, Steven, "The Sharpe Ratio: Statistics and Applications", 2022

[7] Andrew W. Lo "The Statistics of Sharpe Ratios", 2003

TAGS: sharpe, quant, backtesting

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Copyright © 2024 BSIC | Bocconi Students Investment Club