MODULE 3 UNIT 1
Modelling in financial markets

IN COLLABORATION WITH getsmarter™

Table of contents

Learning outcomes:

LO1: Identify the characteristics of out-of-sample testing and backtesting as methods of statistical verification.

LO2: Review appropriate measures of model performance, including their advantages and disadvantages.

LO3: Discuss how the Sharpe ratio is used to measure a model's validity.

1. Introduction

Understanding the process of how a model is put together is essential for a full understanding of the industry. The purpose of this programme is not to teach you how to build an algorithmic trading model, but rather to instil an understanding that will allow you to engage with the industry and navigate the key issues around a fund's performance.

These notes start with an introduction to mathematical modelling and move into methods for trade strategy validation – how to test that your model actually works. They then move on to covering different measures of performance and when these should be used – having a clear understanding of how a strategy performs compared to others is essential to its evaluation. Finally, you will learn about measures for assessing financial risk, an extremely important consideration for investors.

2. What is modelling?

The idea of using mathematical models to describe the world has been around for thousands for years. Figure 1 shows Pythagoras' theorem, something you may be familiar with. Pythagoras' theorem is a method of Euclidean geometry used to describe the relationship between three sides of a right-angled triangle.

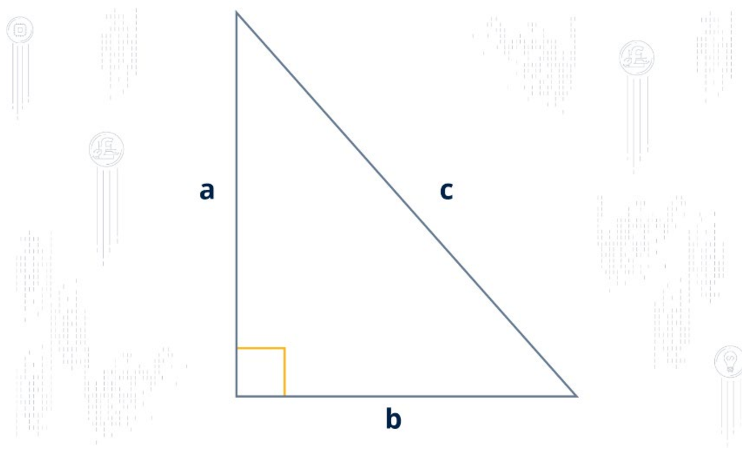

Figure 1: A right-angled triangle.

In Figure 1, *c* is known as the hypotenuse. Its relationship to the other sides is described mathematically as:

$$c^2 = a^2 + b^2$$

or

$$c = \sqrt{a^2 + b^2}$$

In the second formula, the hypotenuse is described as a function of the other two sides. If the lengths of *a* and *b* were known, you could compute *c*. You could even say *c* is the dependent variable and *a* and *b* are the independent variables. This, and the concept of a mathematical model, could be generalised using the following notation:

$H = g(x)$

Here, *H* is the dependent variable, and *x* is the independent variable. The function *g(x)* describes the relationship between the dependent and independent variables (Wolberg, 2000:1).

This is the basic concept of a financial model, but of interest to this programme is how to model and how to use data to model.

2.1 Data modelling

Data modelling is the process used to ascertain what type of modelling techniques should be used to describe the relationship you are interested in. All these techniques fall into two general categories: parametric and non-parametric.

2.1.1 Parametric

When looking at parametric models, the key is to know that they already have a predefined functional form, meaning that *g(x)* is already known or can be determined via the consideration of underlying theories (Wolberg, 2000:2). So, you have a functional form, but perhaps do not have all the parameters defined. A simple example of this is a parabolic equation:

$y = ax^2 + bx + c$

This is the equation's functional form, where *a*, *b*, and *c* are the unknown parameters and *x* is the independent variable. For you to be able to use this to find *y* for a given value of *x*, you would need to use some data to determine values for *a*, *b,* and *c*. This is what data modelling seeks to achieve for more complex parametric models.

2.1.2 Non-parametric

In non-parametric modelling, *g(x)* is unknown, meaning that it makes no assumption about the underlying distribution of the data. This is in contrast to parametric models, which have a predefined form that can be described by a general equation and often require the data to follow a normal distribution and be of a significant size (Yau, n.d.).

An example of how non-parametric modelling can help is laid out in *Expert Trading Systems: Modeling Financial Markets with Kernel Regression* by John R. Wolberg. Wolberg's students are tasked with producing a program that can estimate height as a function of position for a given area. So, if height is *H* and position is (*x*, *y*), then they want to produce a model in the form *H(x, y)* and have a defined functional form with parameters that can be estimated. The problem is that the area is extremely mountainous, and so there is no reason to assume that two points next to each other have similar heights. Compounding this problem is the fact that the placement of hills or elevated regions is random. How can this be solved? To do this in a parametric manner would require intensive labour and multiple equations for each small section of the area.

It is extremely hard to estimate the value of a dependent variable that is a function of multiple independent variables when a complex relationship between these independent variables exists. Attempting to come up with a general equation that explains the dependent variable (height, in this example) as a function of independent variables (*x* and *y)* in cases like this is pointless. A non-parametric method for height estimation could be used instead.

When modelling financial markets, it would be ideal if there were an overall functional form of an equation that could predict market movements. However, it is unlikely that such a model exists, and certainly no one has discovered it yet. Hence, most modelling of financial markets is of the non-parametric form. Typically, these are known as "data-driven models" (Wolberg, 2000:3).

3. Modelling in financial markets

What is required here is a mechanism for predicting market movements – fundamentally, some truth that will continue to last and is right most of the time. This belief must exist, as it is your premise and essentially what a model is. It can be very hard to see this truth when looking at only one year or one market, as market movements are not just affected by the underlying truth being sought (the model), but also by noise.

Noise in this context can be thought of as follows: You can see that market movements happen because of an underlying truth or economic principle, but the movements are also a result of some other unknown interaction (or some unknown variable). This variable is the noise. It is for this reason that you need a lot of data, or a lot of markets to work in, to be able to see beyond the noise and find the underlying truth on which you will base a model. Anthony Ledford's presentation in Module 1 covers this topic nicely, and it is recommended that you refer back to it if necessary.

Typically, what the model would focus on is future volatility (recall the use of predicted volatility in Ledford's presentation) or future price movements (Wolberg, 2000:7). This is the base that you want to build on with the idea of moving towards creating an algorithmic trading system that will allow you to trade in multiple markets at various frequencies. As mentioned in Ledford's presentation, algorithmic traders do not focus on only one market. Instead, their method is to trade in multiple markets (known as diversifying), which is a common financial principle.

There are specific challenges for modelling in financial markets that are common for non-parametric data:

- The amount of data that you require to make a functional model is unknown. In other words, there is no measure for how many independent variables you need. Therefore, it is unknown whether it is possible to produce a working model with the available data.

- Given a defined combination of independent variables, there is not necessarily only one value for the dependent variable. A very simplified representation of this concept is shown in Figure 2. As seen on the graph, Point *A* is the *x* value and Points *B* and *C* are the *y* values. So, when *x* equals *A*, *y* can be either *B* or *C*. In financial market modelling, the result is that, for each combination of independent variables, there is a range of possible values for the dependent variable.

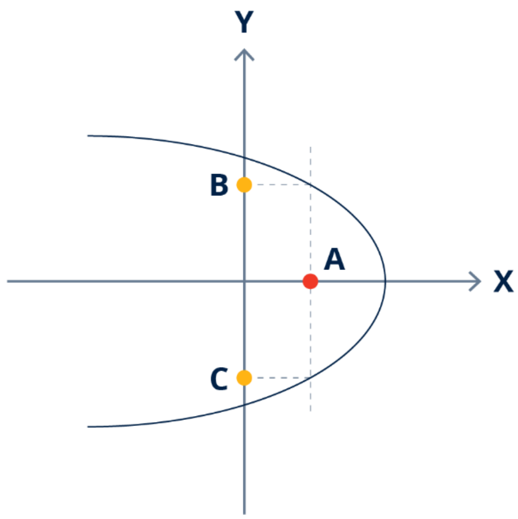

Figure 2: A graph with two *y* values for one *x* value.

- No matter how much data you have, your model will never be error-free. Think back again to Anthony Ledford's presentation: Before certain transformations, the data looked like noise; the signal he did manage to obtain was, as he said, extremely weak. What you want is "a signal with enough predictive power to make the model useful".

(Wolberg, 2000:9)

Keeping these challenges in mind, it is time to think about some practical steps that would need to be taken to begin modelling in financial markets. The next section talks briefly about candidate predictors and their relevance in developing a model.

3.1 Candidate predictors

As mentioned earlier, there is no universal equation for defining financial markets. Consider a theoretical market "M". It is thought that conditions X, Y, and Z affect market M. So, you know that X, Y, and Z affect M, but that is not the same as knowing the exact relationship between X and M, Y and M, and Z and M. A simple example is that it can be reasonably assumed that negative earnings expectations for a firm would lower its stock value, but by how much and for how long?

Practically speaking, you are trying to predict future movements of a time series, specifically the changes in a security's value over time. So, the conditions mentioned earlier (X, Y, and Z) would also be some form of time series. Their exact relationship with market M is unknown, but candidate predictors can be suggested based off X, Y, and Z. These predictors will form the basis of your modelling process (Wolberg, 2000:34).

Suppose you are trying to predict price movements (which you can call Y) one day into the future. So, if the price is £100 on Day 1 and £105 on Day 2, then Y at Day 1 is £5. This is a forward-looking measure. A basic example of a candidate predictor for this would be to first look at the daily price changes over the last *n* days, which is a backward-looking measure. You would need to select which values of *n* you are interested in. Perhaps you want to look at *n* = 1, 2, and 3.

Looking at Table 1, X1 is the price difference over one day (*n* = 1). This is calculated by subtracting the price at Day 1 from the price at Day 2. Similarly, X2 is the price change over two days (*n* = 2). This is calculated by subtracting the price at Day 1 from the price at Day 3. Lastly, X3 is the price difference over three days (*n* = 3). The X1, X2, and X3 columns are the candidate predictors for Y. You can calculate the values of the different X columns for all the days, as shown in Table 1. Thereafter, if it is possible to come up with a model based on the X values, then it would be possible to use this model to predict Y.

Table 1: Candidate predictors.

Day	Price	Y	X1	X2	X3
1	100	5			
2	105	-1	5		
3	104	-5	-1	4	
4	99	-19	-5	-6	-1
5	80	26	-19	-24	-25
6	106	14	26	7	2

7	120	30	14	40	21
8	150	-27	30	44	70
9	123	-13	-27	3	17
10	110	1	-13	-40	-10

The suggested predictors in this table are a basic illustration of how you would begin to develop the rules that your model will follow.

4. What is a valid model?

After you have used your data to estimate a model, you need to test that it works. Professor Vulkan highlighted that effective model validation is an essential component in the model development process. This section examines two methods of validation: out-of-sample testing and backtesting.

Before moving on to these topics, it is important to understand the concept of overfitting. This is a problem faced by any model that tries to predict future human behaviour based on historical performance, and can be consistently observed in the world. Overfitting occurs when additional parameters are added to your model to better fit historical data. This often results in a highly complex model that appears to fully explain the past but really just fits itself to the noise in the data.

What happens is that the model gives equal weighting to the noise as to any underlying economic principle that you are attempting to base your model on. Imagine you are interested in betting on horse racing. Predicting the winner of the next race would be of great value to you. You have a room of six people; one person is an expert on the topic, and the other five know nothing of value or have false information. You have asked them for advice, and they all give you their opinions on who is likely to win the next race. You give each of their opinions an equal weighting and choose the horse that the group, on average, thought would win.

Clearly, more weight should be placed on the opinion of the expert. Ideally, only the expert should be in the room, but unfortunately, you cannot control this; there will always be non-experts. In this analogy, the five people who are not experts represent the noise, and the expert represents the underlying economic principle you hope to target. You need to be aware of the noise in the data, or the other people in the room. If you are not aware of the underlying principle, or expert, at all times, then your model will see the noise as a signal.

So, when you give your model too much freedom to fit data, it will easily achieve this – almost perfectly in the case of complex models. But think back to the discussion about noise. Some movements are simply random. So, an overfitted model is focused on explaining all the past movements rather than adhering to some underlying economic principle. This restricts it from seeing beyond the past.

Further reading:

Overfitting is a tricky consideration to navigate when building a model. Learning how to detect and prevent overfitting is crucial to build a valid model.

In addition to overfitting, the structural concerns with trading in a real-life environment must be considered, including transaction costs and slippage:

- **Transaction costs:** The cost of actioning a trade can influence the potential profit of a model. This is of special importance when considering high-frequency trading, as more transactions mean more transaction costs.

- **Slippage:** Slippage is the difference between what an investor thinks they will pay for a security and what they actually pay. If the model expects to buy a security for £40 but buys it for £41, and this happens over hundreds of trades, it could impact the profitability of the model.

4.1 Out-of-sample testing

One way to test the model's assumptions and compare its predictive performance against other models is to use a meaningful out-of-sample (OOS). The OOS comprises data that is withheld from the model training process and is used to test model accuracy. This testing is achieved by allowing the model to make predictions using the withheld data (also known as hold-out data) and comparing whether performance is similar to predictions made by the model trained with the in-sample data (Nau, n.d.). OOS testing is a very large and complex subject. This programme does not intend to cover it in its entirety, but it is very important to understand how the OOS periods are selected.

A model's performance "on data outside that used in its construction remains the touchstone for its utility in all applications" (Fildes & Makridakis, 1995:293). See Figure 3 for an example of OOS selection.

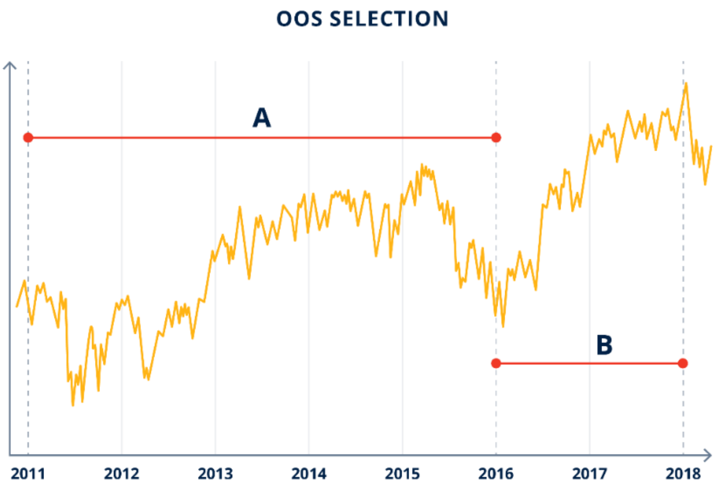

Figure 3: Out-of-sample selection.

Looking at Figure 3, Portion A represents the data in the estimation period used to estimate the parameters of the model. The model is then tested on data in the validation period, marked Portion B. In this example, the OOS years are all at the end of the period. Professor Vulkan recommends that you instead use alternating years. So, for example, if you had a model looking at the years 1990–2002 (13 years), instead of having the years 1999–2002 for the OOS period, you would use 1995, 1997, 1999, and 2001. This is because, if you use just the end of the period as your OOS years, then you may miss some interaction or trend that occurred over this time, which your model would then not have considered.

Perhaps the last few years were the beginning of the 2008 financial crisis. In this case, if the years you looked at were 1999–2009, with the last three years as your OOS period, you would very likely see the model fail OOS. However, would it be fair to deduce from this information that the model is bad? The model would be judged as performing poorly based on something that is not related to its inherent make-up. If a model is bad, it will almost certainly collapse OOS. However, even if the model is good, it still can perform badly OOS for a number of reasons.

Again, it must be stressed that this is not an exact science, as there is no definitive correct answer and the body of knowledge around this subject is immense. This programme does not cover the topic comprehensively. However, it is an extremely important step in the development process, and the concept should be understood.

The data in the estimation is used to estimate the parameters of the model. Parameters defined in this way are called fitted values. This is a direct example of a data-driven model – using given data and attempting to produce a system that can estimate into the future.

4.2 Backtesting

Backtesting is a method of simulating a model on historical market data. It allows for the testing of many versions of the same model or idea accurately and quickly. You can receive information on how your model would perform in the past instantly.

It makes sense to first understand why you would conduct a backtest. The historical data you would use is by definition empirical, so backtesting allows you to carry out empirical research quickly and efficiently. Markets are inherently not determinist, meaning there is no clearly defined cause and effect for all market movements. So, when a trading strategy works, it is not because it has perfectly described a mathematical principle, but rather that it has captured a particular market characteristic that produces a positive return over time. It could be said that there is no absolute certainty that a trade will be the right move, but you can ascertain that, over time, following your set of rules will have you turning a profit most of the time. This is really where backtesting is essential.

There are a few problems with backtesting trading models. This testing is often too strict, which can lead to what is known as a “type II error”, which can be thought of as a false negative. It could lead you to reject a model that does in fact provide positive returns. A simple example of a type II error is if you were driving down the road, heard a strange sound from your engine, and decided that there was no mechanical issue, but there was in fact a mechanical error, and you should have gone straight to the mechanic.

Another issue is that, even if a strategy has undergone significant backtesting, it can often still fail when market conditions change, as they often do.

5. Performance measurement

Once you have validated your model, the hard part is perhaps over. You have found that it works, but how well does it work? This section looks at methods you can use to gauge how well a trading strategy is working. All investors look at performance, but more sophisticated investors look at risk-adjusted return. Risk-adjusted return puts the returns on a more common scale and is a fairer way to compare funds. Think back to Anthony Ledford's Module 1 presentation again. He divides price differences by predicted volatility to put them on a common, comparable scale.

There are many forms of risk-adjusted performance measurements, but the Sharpe ratio is perhaps among the best known. As such, this set of notes focuses primarily on the Sharpe ratio; however, it will also look at the Sortino ratio and the returns to drawdown. You also look at how the VAMI and other indices are used to compare performances.

5.1 Sharpe ratio

The Sharpe ratio measures a financial portfolio's risk-adjusted return by using standard deviation. A higher Sharpe ratio is an indication the fund has better returns relative to the amount of risk they are exposed to. Since the Sharpe ratio uses standard deviation as a measure for risk, you can compare risk-adjusted returns across fund categories (*The Economic Times*, 2021).

When the Sharpe ratio was originally developed in 1966, its developer, William Sharpe, proposed that it be named the "reward-to-variability ratio". But, despite the measure gaining significant popularity over time, this name failed to achieve widespread adoption (Sharpe, 1994). The Sharpe ratio, as it is commonly known today, measures the portion of return generated by a portfolio above that of the risk-free rate or that of a benchmark security.

The risk-free rate is the rate of return an investor could make in a riskless asset. The yield on US Treasury bonds is often used as a benchmark for a riskless asset. Sharpe originally used the 10-year US government bond for this. He defined the extra return above the risk-free rate as "the reward provided the investor for bearing risk" and the Sharpe ratio itself as "the reward per unit of variability" (Sharpe, 1966). It is essentially asking: "Per unit of risk, how much return am I getting?"

The inputs required for calculating the Sharpe ratio are the average annual return of a portfolio (denoted as $R(p)$), the risk-free rate of return (denoted as $r(f)$), and the standard deviation of the portfolio (denoted as $\sigma(p)$). Annual figures are typically used for this calculation. The formula reads as follows:

$$\text{Sharpe ratio (SR)} = \frac{\left(R(p) - r(f)\right)}{\sigma(\text{p})}$$

Think of this in the context of performance measurement. Assume two funds are being compared. If they both have the same level of return, the one with the higher standard deviation will have a lower Sharpe ratio, meaning that it returns less for each unit of risk taken when compared to the other fund. In order to be competitive, the fund with the higher standard deviation would need to produce higher returns to compensate (*The Economic Times*, 2021).

Think back to the efficient frontier you explored in Module 1. Can you see how these are linked? The efficient frontier seeks to maximise returns and minimise risk in a comparison of portfolios. In a similar way, the Sharpe ratio is a comparative measure of risk and return. As with the efficient frontier, it should not be viewed in isolation. The Sharpe ratio of a singular portfolio reveals little insight until it is compared to another portfolio. Typically, a "good" Sharpe is above 1, but, again, this measure only makes sense when used as a comparative device. There needs to be something to compare it to, and the context around where it is being used must be considered.

5.2 Sortino ratio

Named after Frank A. Sortino, the Sortino ratio is a modified version of the Sharpe ratio. The key difference is that the Sortino ratio uses downside deviation as the measure of risk, while the Sharpe ratio uses standard deviation (i.e. both upside and downside deviation) (Rollinger & Hoffman, n.d.). Downside deviation looks only at the standard deviation of returns that are below a specified minimum acceptable return (eVestment, n.d.). This specified minimum is also known as the desired target return and can be an investor-specified level of return, or the required rate of return on an investment. Once this level of return is specified, any component of the portfolio that produces a return less than that is considered risky and thus forms part of the Sortino ratio's measure of risk.

The inputs required for the Sortino ratio are the average annual rate of return of a portfolio (denoted as $R(p)$), the desired rate of return for the investment strategy being considered (denoted as $R(t)$), and the downside deviation (denoted as $\sigma(d)$). The formula reads as follows:

$$\text{Sortino ratio (STR)} = \frac{R(p) - R(t)}{\sigma(\text{d})}$$

The Sortino ratio corrects some of the issues in the Sharpe ratio, such as the use of the standard deviation, which penalises investors for taking on good risk associated with high returns. However, neither measure looks at current or future risk; they are based on historical data and are innately backward-looking (Rollinger & Hoffman, n.d.). As the desired rate of return can vary, it is important to apply the same method of calculation throughout when comparing Sortino ratios. As with the Sharpe ratio, a higher value is more desirable.

5.3 Returns to drawdown

There are a number of measures that look at returns related to drawdown. The drawdown is the largest loss of value in a fund's movement over the specified time, from the highest peak to the lowest valley before a new high is reached. This is expressed as a percentage of the highest peak. For example, looking at Figure 4, Points A, C, E, G, and I are peaks, and Points B, D, F, and H are valleys. The drawdown over this period would be (E − H) ÷ E. Note that it is not E – F, as it does not have to be a consecutive movement; it is instead the lowest point before a new high point is obtained. In this case, I is the new high point. However, note that the new high point does not form part of the calculation (YCharts, n.d.).

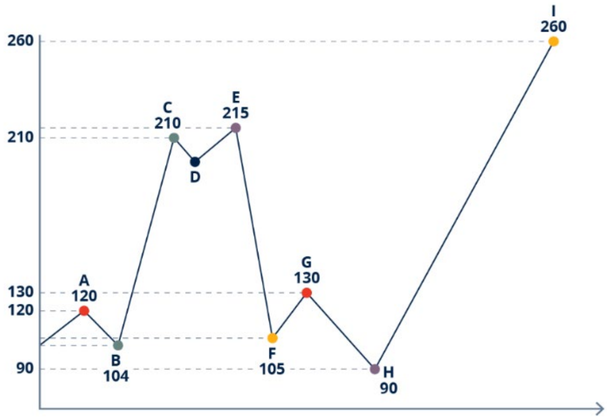

Figure 4: Maximum drawdown.

One measure using drawdowns is the Calmar ratio, which looks at a fund's drawdown as a key measure of performance. Three years is the typical time frame, but this could be amended depending on what makes sense for a particular investment. A higher Calmar ratio indicates better performance, while a lower one indicates higher drawdown risks.

A measure suggested by Professor Vulkan is to use the average of the last three drawdowns instead of only the largest. This reduces the measure's sensitivity to one large drawdown and will thus give you a more rounded measure of the fund's performance relative to its drawdowns. The advantage of this measure is that it diminishes the overall impact of periods with extremely high or low returns that would skew other measures such as the Sharpe ratio.

5.4 Value-Added Monthly Index (VAMI)

The Value-Added Monthly Index (VAMI) is used by investors to compare funds. It is calculated by starting with a base value of $1000 and adding the fund's returns as a percentage. For example, if the fund made a 4% return and the VAMI was at $1000, it would increase to $1040. VAMI is calculated as follows:

Current VAMI = *Initial value* × *(*1 + *Rate of return)*

From then on:

Current VAMI = *Previous VAMI* × *(*1 + *Rate of return)*

VAMI assumes the reinvestment of all profits and interest income. Incentive and management fees are deducted from the calculation of rate of return, meaning that the VAMI calculation is net of all fees.

VAMI is one of the tools most commonly used by investors to assess fund and fund manager performance, and to compare performance across multiple funds. Software is often used to calculate and chart VAMI calculations, allowing investors to assess performance at a glance, and enabling comparison across funds and with specified benchmarks.

5.5 M^2 measure

Also known as the Modigliani risk-adjusted performance measure, the M^2 measure is a variant of the Sharpe ratio that aims to simplify interpretation by focusing on returns. Like the Sharpe ratio, the M^2 measure considers total risk. The difference is that the M^2 measure creates an adjusted portfolio in such a way that the volatility of the adjusted portfolio matches that of a market index (such as the FTSE 100).

The adjusted portfolio has the same standard deviation as the index, which enables the two to be compared simply by comparing returns. The adjusted portfolio's return can be calculated by subtracting the Sharpe ratio of the market index from the Sharpe ratio of the portfolio and then multiplying the answer by the standard deviation of the market (Hsieh & Hodnett, 2013).

The M^2 measure can be written as:

$$M^2 = (Sharpe_P - Sharpe_M) \times \delta_M$$

Where $Sharpe_p$ is the Sharpe ratio of the portfolio, $Sharpe_M$ is the Sharpe ratio of the market index, and δ_M is the standard deviation of the market index.

As with most of the performance measures, the M^2 measure has the disadvantage that it is backward-looking. However, it has the distinct advantage of easy interpretation, and the fact that it is not a ratio means that comparison across different M^2 calculations is simple. The M^2 measure is a particularly good one to use in the case of negative returns, where the ratio measures are difficult to interpret.

5.6 Indices

This section starts with a short explanation of what stock market indices are. If you are already familiar with this concept, feel free to skip ahead.

A stock market index represents the value of a specific section of the stock market. The price of stocks thought to best represent a particular section are weighted to form the index. Their common purpose is to describe the stock market and to be used for comparison purposes between different sectors of the market. It is a theoretical construction and cannot be traded directly like common stocks. Instead, you need to invest in exchange-traded funds (ETFs) that are designed to track the index, or use derivatives (Sharp Trader, n.d.).

Once you have validated your model and found it has held OOS and seems to have a decent performance, you need to know how well it performs compared to others. Your algo system will be judged on the relevant criteria; these criteria are typically the markets you trade and the type of strategies your employ. An example of this classification, taken from HSBC's Alternate Investment Group, is shown in Figure 5.

Figure 5: Classification by strategy. (Source: HSBC)

For a meaningful comparison, a fund manager would need to ensure they compare their fund with the appropriate index – an index that has the same underlying assets and is traded in the same market. For example, a large fund invested in stocks might compare itself to the S&P 500.

As Professor Vulkan mentioned, when considering indices, survivorship bias is an important consideration. In the context of this programme, this refers to the fact that funds are removed from a performance index when they fail, which leads to a situation where the overall index is not an accurate depiction of performance, as it only contains funds that have been successful.

Further reading:

The impact of survivorship bias on index performance can often deceptively amplify certain fund returns while hiding poor performances.

The key takeaway here is that performance must be comparable. If your system has a Sharpe of 1.1 in a market where other players have a Sharpe of 0.95, then it is probably safe to say that you are doing well. If you have a Sharpe of 1.5 in a market where the other players have a Sharpe of 1.9, then perhaps you are not.

Your model failing in-sample and out-of-sample would depend on the metric that you are attempting to maximise when optimising your model. Typically, you would allow for a degradation of up to 20% of this performance metric. For example, if you have an in-sample Sharpe ratio of 2.0 and an out-of-sample Sharpe ratio of less than 1.6, then this model could be considered to have failed out-of-sample. Again, it must be stressed that this is not an exact science. Market conditions and the selection of out-of-sample years must still be considered when making a final decision.

***

Based on what you have learnt thus far, try to answer the following quick questions.

Quick questions:

Question 1: When it comes to confirming the validity of a model, what is the difference between backtesting and out-of-sample testing?

Question 2: How do the Sharpe and Sortino ratios differ in relation to measuring model performance?

Question 3: What is returns to drawdown, and what is the typical time period over which this measure is calculated?

Note: Refer to the end of this set of notes to view the answers.

6. Risk management

At this stage, you have estimated your model, validated it, measured its performance, and you have found that it is worth implementing. You now need to consider how to manage risk factors relating to your model. Perhaps this model is just one of many you have created, and together they make up your portfolio. This section looks at tools for measuring risk, which can give you a better understanding of the nature of your risk and how to manage it better.

6.1 Value at Risk (VaR)

Value at Risk (VaR) is a statistical tool that is commonly used to assess financial risk associated with individual firms, specific assets, or whole portfolios. VaR is measured by identifying a specific time frame and then calculating the amount of potential value loss of the security under consideration. For example, if a portfolio had a 4% one-month VaR of 3%, then there is a 4% chance of the portfolio's value falling by 3% over that month. Another way this is sometimes presented is as the probability of occurrence being reflected in terms of a confidence level – for example, "With a 95% level of confidence, what is the most I can expect to lose in the next month?"

A standard issue with the use of VaR is that it is assumed that future price movements will follow those of the past. This problem is typically addressed by making use of stress testing or a sensitivity analysis (Farid, 2010).

Generally speaking, these are the steps that are followed when calculating VaR:

- Specify portfolio positions (what shares and bonds are held, etc.).
- Ascertain exposures that could impact these positions (risk).
- Give probabilities to the risk factors (likelihood of the risk factor impacting the portfolio's position).
- Calculate the price functions of the portfolio's position as a function of risk factors.

(Farid, 2010)

Within this structure, there are three main methods of calculating VaR: the variance-covariance method, historical returns, and Monte Carlo simulation. These are expanded on briefly in the following subsections.

6.1.1 Variance-covariance method

Here it is assumed that price data follows a normal distribution (Farid, 2010). This is already problematic, as price data has empirically shown strong evidence of non-normality (Richardson & Smith, 1993). It is something that should be kept in mind when choosing a method that best suits you.

In this method, the portfolio's overall standard deviation is derived from the standard deviations of the individual securities and the correlation between them. Therefore, you need to start by calculating the values for these individual securities.

Further reading:

The variance-covariance method can be complex, but it can be useful to understand the primary technical points.

6.1.2 Historical returns method

The historical returns method is a non-parametric method of calculation whereby the market data is not assumed to fit any specific distribution, which means it has fewer assumptions to satisfy. The VaR calculation is done directly on the data, but it still has to satisfy the assumption that future price movements will follow past ones (Farid, 2010). This historical method uses the 100-day change in price levels to estimate a hypothetical data set on which the calculation is performed.

Further reading:

The historical returns calculation assumes that past performance is a good predicter of near-future performance.

6.1.3 Monte Carlo simulation method

Monte Carlo simulation is a method for computerised risk analysis that models possible distributions through the process of random number generation. This method is similar to the historical returns method, except that the hypothetical data set is approximated by a statistical distribution. The main assumption here is that the approximated distribution correctly estimates the price movements of the portfolio's securities (Farid, 2010).

Further reading:

The Monte Carlo simulation provides a range of different outcomes that can occur from various decisions. Calculating the Monte Carlo simulation is similar to the historical returns method, except that it uses random numbers instead of past performance.

7. Conclusion

This set of notes took you on the journey from basic mathematical modelling to producing a financial model that works under market conditions. This programme is a non-technical journey by design; it is important to understand the process from this perspective, as you can only then begin to look at the individual technicalities of each step. Also, keep the process in mind for the next module when you will get to see parts of it in action.

8. Bibliography

Downs, E. 2000. *7 Chart patterns that consistently make money.* United States of America: Marketplace Books.

eVestment. n.d. *Downside deviation.* Available: https://www.evestment.com/resources/investment-statistics-guide/downside-deviation/ [2018, May 3].

Farid, J. 2010. *VaR methods – calculating value at risk.* Available: https://financetrainingcourse.com/education/2010/03/master-class-calculating-value-at-risk-var-var-methods/ [2018, May 14].

Fildes, R. & Makridakis, S. 1995. The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*. 63(3):289-308.

Hsieh, H.H. & Hodnett, K. 2013. A review of performance evaluation measures for actively-managed portfolios. *Journal of Economics and Behavioral Studies.* 5(12):815-824.

Kidd, D. 2011. *Measures of risk-adjusted return: let's not forget Treynor and Jensen.* Available: https://www.cfapubs.org/doi/pdf/10.2469/ipmn.v2011.n1.12 [2018, May 3].

Nau, R. n.d. *Three types of forecasts: estimation, validation, and the future*. Available: https://people.duke.edu/~rnau/three.htm [2021, August 26].

Research Desk. 2003. *Treynor ratio.* Available: https://www.valueresearchonline.com/story/h2_storyview.asp?str=3940 [2018, May 3].

Richardson, M. & Smith, T. 1993. A test for multivariate normality in stock returns. *The Journal of Business.* 66(2):295-321.

Rollinger, T.N. & Hoffman, S.T. n.d. *Sortino: a 'Sharper' ratio.* Available: http://redrockcapital.homestead.com/Sortino__A__Sharper__Ratio_Red_Rock_Capital.pdf [2018, May 3].

Sharp Trader. n.d. *Overview of stock market indices.* Available: http://www.sharptrader.com/new-to-trading/stock-indices/overview-of-stock-market-indices/ [2018, May 14].

Sharpe, W.F. 1966. Mutual fund performance. *The Journal of Business*. 39(1):119-138.

Sharpe, W.F. 1994. The Sharpe ratio. *The Journal of Portfolio Management.* 21(1):49-58.

The Economic Times. 2021. Definition of 'Sharpe ratio'. *The Economic Times.* 24 August. Available: https://economictimes.indiatimes.com/definition/Sharpe-ratio [2021, August 24].

TradingView. n.d. *FTSE 100, D, TVC.* Available: https://www.tradingview.com/chart/?symbol=UKX [2018, May 2].

Wolberg, J.R. 2000. *Expert trading systems: modeling financial markets with kernel regression.* Wiley. USA: Michigan.

Yau, C. n.d. *Non-parametric methods*. Available: http://www.r-tutor.com/elementary-statistics/non-parametric-methods [2018, May 3].

YCharts. n.d. *Max drawdown.* Available: https://ycharts.com/glossary/terms/max_drawdown [2018, May 3].

Young, T.W. 1991. Calmar ratio: a smoother tool. *Futures.* 20(1):40.

9. Appendix

9.1 Treynor ratio

The Treynor ratio, also known as the Treynor measure or index, is very similar to the Sharpe ratio, as it looks at the difference between a fund's average returns and the risk-free rate. However, instead of dividing by total risk (captured by the portfolio's standard deviation), it focuses only on systematic, or market, risk (captured by beta). It produces the amount of return over the risk-free rate per unit of market risk (Research Desk, 2003).

The Sharpe ratio uses standard deviation, which is a measure of variation. Hence, the Sharpe ratio is also known as the reward-to-variability ratio. The Treynor ratio ignores risk that can be diversified away and only uses market risk, which is basically volatility. Hence, the Treynor ratio is also known as the reward-to-volatility ratio.

The inputs required for the Treynor ratio are the average return of a portfolio (denoted as $R(p)$), the risk-free rate of return (denoted as $r(f)$), and the portfolio's beta (denoted as $\beta(p)$). The formula reads as follows:

$$\text{Treynor ratio (TR)} = \frac{R(p) - r(f)}{\beta(p)}$$

When a portfolio is well diversified, total risk will approach market risk, and the Sharpe and Treynor ratios will produce similar results. In this instance, any difference between variability and the volatility in the market is likely due to transitory effects. As such, the Treynor measure is likely to be the better measure to use, as the denominator will not be swayed by these transitory effects, and the measure may produce good indications about future performance. If, however, a fund is not well diversified, the Treynor index will not be a good measure of past performance, as it will not capture the risk exposure resulting from lack of diversification (Sharpe, 1966).

The Treynor measure is most useful when an asset or portfolio is part of a larger, well-diversified investment portfolio. In this case, the measure allows you to evaluate the asset or portfolio's excess return against its systematic risk, rather than total risk, thereby assessing that particular asset or portfolio's contribution to performance. In contrast, the Sharpe ratio is more appropriate when the portfolio represents the whole investment.

9.2 Jensen's alpha

Jensen's alpha is used to evaluate a fund manager's performance in absolute terms, as opposed to relative terms (Kidd, 2011). Jensen's alpha is the difference between the return produced by the fund and the return necessary to reward an investor for taking on a certain degree of market risk.

The inputs required for Jenson's alpha are the portfolio's return (denoted as $R(p)$), the risk-free rate (denoted as $r(f)$), the portfolio's beta (denoted as $\beta(p)$), and the market return (denoted as $R(m)$) (Kidd, 2011). The formula reads as follows:

$$\text{Jensen's alpha (JA)} = (R(p) - r(f)) - \beta(\mathrm{p}) \times (\mathrm{R(m)} - \mathrm{r(f)})$$

Like the Treynor ratio, Jensen's alpha is best used when the portfolio is just one sub-portfolio of many, as it focuses only on beta and thus assumes the larger investment portfolio is well diversified.

10. Answers

Quick questions:

Question 1: Backtesting is an empirical method of model simulation using historical market data. It allows you to test many versions of the same model accurately and quickly. Out-of-sample (OOS) testing refers to withholding a portion of market data when training an algo model. This helps to simulate what would happen if you executed your strategy on live market data. Backtesting and OOS testing often work together to confirm that the model is valid. While you can backtest the model any number of times using the in-sample data, you should ideally only test the final model on the OOS data once to determine if the model is valid.

Question 2: The Sharpe and Sortino ratios are both model performance measures that give an indication of a financial portfolio's risk-adjusted returns. The Sharpe ratio incorporates the standard deviation of the portfolio, while the Sortino ratio uses the portfolio's downside deviation. The downside deviation only incorporates the standard deviation of returns that are below a specified threshold and thus the Sortino ratio does not penalise investors for taking on good risk associated with high returns.

Question 3: Returns to drawdown is a performance measure that calculates the largest loss of value in a fund's movement over a specified period. In other words, it is the difference between the highest peak and lowest valley before a new high is reached. The typical period over which this measure is calculated is three years. However, this can differ depending on the investment.