Quantitative trader roles within large quant funds are often perceived to be one of the most prestigious and lucrative positions in the quantitative finance employment landscape. Trading careers in a "parent" fund are often seen as a springboard towards eventually allowing one to form their own fund, with an initial capital allocation from the parent employer and a list of early investors to bring on board.

Competition for quantitative trading positions is intense and thus a significant investment of time and effort is necessary to obtain a career in quant trading. In this article I will outline the common career paths, routes in to the field, the required background and a self-study plan to help both retail traders and would-be professionals gain skills in quantitative trading.

Setting Expectations

Before we delve into the lists of textbooks and other resources, I will attempt to set some expectations about what the role involves. Quantitative trading research is much more closely aligned with scientific hypothesis testing and academic rigour than the "usual" perception of investment bank traders and the associated bravado. There is very little (or non-existent) discretionary input when carrying out quantitative trading as the processes are almost universally automated.

The scientific method and hypothesis testing are highly-valued processes within the quant finance community and as such anybody wishing to enter the field will need to have been trained in scientific methodology. This often, but not exclusively, means training to a doctoral research level - usually via having taken a PhD or graduate level Masters in a quantitative field. Although one can break into quantitative trading at a professional level via alternate means, it is not common.

The skills required by a sophisticated quantitative trading researcher are diverse. An extensive background in mathematics, probability and statistical testing provide the quantitative base on which to build. An understanding of the components of quantitative trading is essential, including forecasting, signal generation, backtesting, data cleansing, portfolio management and execution methods. More advanced knowledge is required for time series analysis, statistical/machine learning (including non-linear methods), optimisation and exchange/market microstructure. Coupled with this is a good knowledge of programming, including how to take academic models and implement them rapidly.

This is a significant apprenticeship and should not be entered into lightly. It is often said that it takes 5-10 years to learn sufficient material to be consistently profitable at quantitative trading in a professional firm. However the rewards are significant. It is a highly intellectual environment.
with a very smart peer group. It will provide continuous challenges at a fast pace. It is extremely well remunerated and provides many career options, including the ability to become an entrepreneur by starting your own fund after demonstrating a long-term track record.

**Necessary Background**

It is common to consider a career in quantitative finance (and ultimately quantitative trading research) while studying on a numerate undergraduate degree or within a specialised technical doctorate. However, the following advice is applicable to those who may wish to transition into a quant trading career from another, albeit with the caveat that it will take somewhat longer and will involve extensive networking and a lot of self-study.

At the most basic level, professional quantitative trading research requires a solid understanding of mathematics and statistical hypothesis testing. The usual suspects of multivariate calculus, linear algebra and probability theory are all required. A good class-mark in an undergraduate course of mathematics or physics from a well-regarded school will usually provide you with the necessary background.

If you do not have a background in mathematics or physics then I would suggest that you should pursue a degree course from a top school in one of those fields. You will be competing with individuals who do have such knowledge and thus it will be highly challenging to gain a position at a fund without some definitive academic credentials.

In addition to having a solid mathematical understanding it is necessary to be adept at implementation of models, via computer programming. The common choices of modelling languages these days include R, the open-source statistical language; Python, with its extensive data analysis libraries; or MatLab. Gaining extensive familiarity with one of these packages is a necessary prerequisite to becoming a quantitative trader. If you have an extensive background in computer programming, you may wish to consider gaining entry into a fund via the Quantitative Developer route.

The final major skill needed by quantitative trading researchers is that of being able to objectively interpret new research and then implement it rapidly. This is a skill learned via doctoral training and one of the reasons why PhD candidates from top schools are often the first to be picked for quantitative trading positions. Gaining a PhD in one of the following areas (particularly machine learning or optimisation) is a good way into a sophisticated quant fund.

**Econometrics and Time Series Analysis**

Fundamentally the majority of quantitative trading is about time series analysis. This predominantly includes asset price series as a function of time, but might include derivative series in some form. Thus time series analysis is an essential topic for the quantitative trading researcher. I've written about how to get started in the article on Top 10 Essential Resources for Learning Financial Econometrics. That article includes basic guides to probability and beginning programming in R.

Recently I came across a fantastic resource called OTexts, which provides open access textbooks.
The following book is especially useful for forecasting:

- **Forecasting: Principles and Practice** by Hyndman and Athanasopoulos - This free book is an excellent way to begin learning about statistical forecasting via the R programming environment. It covers simple and multivariate regression, exponential smoothing and ARIMA techniques as well as more advanced forecasting models. The book is originally pitched at business/commerce degrees but is sufficiently technical to be of interest to beginning quants.

With the basics of time series under your belt the next step is to begin studying statistical/machine learning techniques, which are the current "state of the art" within quantitative finance.

**Statistical Machine Learning**

Modern quantitative trading research relies on extensive statistical learning techniques. Up until relatively recently, the only place to learn such techniques as applied to quantitative finance was in the literature. Thankfully well-established textbooks now exist which bridge the gap between theory and practice. It is the next logical follow-on from econometrics and time series forecasting techniques although there is significant overlap in the two areas.

The main techniques of interest include **Multivariate Linear Regression**, **Logistic Regression**, **Resampling Techniques**, **Tree-Based Methods** (including **Random Forests**), **Support Vector Machines** (SVM), **Principal Component Analysis** (PCA), **Clustering** (K-Means, Hierarchical), **Kernal Methods** and **Neural Networks**. Each of these topics is a significant learning exercise in itself, although the above two texts will cover the necessary introductory material, providing further references for deeper study.

A particularly useful (and free!) set of web courses on Machine Learning/AI are provided by Coursera:

- **Machine Learning** by Andrew Ng - This course covers the basics of the methods I have briefly mentioned above. It has received high praise from individuals who have participated. It is probably best watched as a companion to reading ISL or ESL, which are two books mentioned in the Essential Algorithmic Trading Reading List PDF.

- **Neural Networks for Machine Learning** by Geoffrey Hinton - This course focuses primarily on neural networks, which have a long history of association with quantitative finance. If you wish to specifically concentrate on this area, then this course is worth taking a look at, in conjunction with a solid textbook.

Statistical learning is extremely important in quant trading research. We can bring to bear the entire weight of the scientific method and hypothesis testing in order to rigourously assess the quant trading research process. For quantitative trading we are interested in testable, repeatable results that are subject to constant scrutiny. This allows easy replacement of trading strategies as and when performance degrades. Note that this is in stark contrast to the approach taken in "discretionary" trading where performance and risk are not often assessed in this manner.
Why Should We Use The Scientific Method In Quantitative Trading?

The statistical approach to quant trading is designed to eliminate issues that surround discretionary methods. A great deal of discretionary technical trading is rife with cognitive biases, including loss aversion, confirmation bias and the bandwagon effect. Quant trading research uses alternative mathematical methods to mitigate such behaviours and thus enhance trading performance.

In order to carry out such a methodical process quant trading researchers possess a continuously skeptical mindset and any strategy ideas or hypotheses about market behaviour are subject to continual scrutiny. A strategy idea will only be put into a "production" environment after extensive statistical analysis, testing and refinement. This is necessary because the market has a rather low signal-to-noise ratio. This creates difficulties in forecasting and thus leads to a challenging trading environment.

What Modelling Problems Do We Encounter In Quantitative Finance?

The goal of quantitative trading research is to produce algorithms and technology that can satisfy a certain investment mandate. In practice this translates into creating trading strategies (and related infrastructure) that produce consistent returns above a certain pre-determined benchmark, net of costs associated with the trading transactions, while minimising "risk". Hence there are a few levers that can be pulled to enhance the financial objectives.

A great deal of attention is often given to the signal/alpha generator, i.e. "the strategy". The best funds and retail quants will spend a significant amount of time modelling/reducing transaction costs, effectively managing risk and determining the optimal portfolio. This PDF is primarily aimed at the alpha generator component of the stack, but please be aware that the other components are of equal importance if successful long-term strategies are to be carried out.

We will now investigate problems encountered in signal generation and how to solve them. The following is a basic list of such methods (which clearly overlap) that are often encountered in signal generation problems:

- **Forecasting/Prediction** - The most common technique is direct forecasting of a financial asset price/direction based on prior prices (or fundamental factors). This usually involves detection of an underlying signal in the "noise" of the market that can be predicted and thus traded upon. It might also involve regressing against other factors (including lags in the original time series) in order to assess the future response against future predictors.

- **Clustering/Classification** - Clustering or classification techniques are methods designed to group data into certain classes. These can be binary in nature, e.g. "up" or "down", or multiply-grouped, e.g. "weak volatility", "strong volatility", "medium volatility".

- **Sentiment Analysis** - More recent innovations in natural language processing and computational speed have lead to sophisticated "sentiment analysis" techniques, which are essentially a classification method, designed to group data based on some underlying sentiment factors. These could be directional in nature, e.g. "bullish", "bearish", "neutral" or emotional such as "happy", "sad", "positive" or "negative". Ultimately this will lead to a
trading signal of some form.

- **Big Data** - Alternative sources of data, such as consumer social media activities, often lead to terabytes (or greater) of data that requires more novel software/hardware in order to interpret. New algorithm implementations have been created in order to handle such "big data".

**Modelling Methodology**

I've provided some key Machine Learning textbooks in the accompanying PDF – The Essential Algorithmic Trading Reading List – and they will discuss the following topics and models, which are necessary for a beginning quant trader to know:

- **Statistical Modelling and Limitations** - The books will outline what statistical learning is and isn't capable of along with the tradeoffs that are necessary when carrying out such research. The difference between prediction and inference is outlined as well as the difference between supervised and unsupervised learning. The bias-variance tradeoff is also explained in detail.

- **Linear Regression** - Linear regression (LR) is one of the simplest supervised learning techniques. It assumes a model where the predicted values are a linear function of the predictor variable(s). While this may seem simplistic compared to the remaining methods in this list, linear regression is still widely utilised in the financial industry. Being aware of LR is important in order to grasp the later methods, some of which are generalisations of LR.

- **Supervised Classification: Logistic Regression, LDA, QDA, KNN** - Supervised classification techniques such as Logistic Regression, Linear/Quadratic Discriminant Analysis and K-Nearest Neighbours are techniques for modelling qualitative classification situations, such as prediction of whether a stock index will move up or down (i.e. a binary value) in the next time period.

- **Resampling Techniques: Bootstrapping, Cross-Validation** - Resampling techniques are necessary in quantitative finance (and statistics in general) because of the dangers of model-fitting. Such techniques are used to ascertain how a model behaves over different training sets and how to minimise the problem of "overfitting" models.

- **Decision Tree Methods: Bagging, Random Forests** - Decision trees are a type of graph that are often employed in classification settings. Bagging and Random Forest techniques are ensemble methods making use of such trees to reduce overfitting and reduce variance in individually fitted supervised learning methods.

- **Neural Networks** - Artificial Neural Networks (ANN) are a machine learning technique often employed in a supervised manner to find non-linear relationships between predictors and responses. In the financial domain they are often used for time series prediction and forecasting.

- **Support Vector Machines** - SVMs are also classification or regression tools, which work by constructing a hyperplane in high or infinite dimensional spaces. The kernel trick allows
non-linear classification to occur by a mapping of the original space into an inner-product space.

- **Unsupervised Methods**: PCA, K-Means, Hierarchical Clustering, NNMF - Unsupervised learning techniques are designed to find hidden structure in data, without the use of an objective or reward function to "train" on. Additionally, unsupervised techniques are often used to pre-process data.

- **Ensemble Methods** - Ensemble methods make use of multiple separate statistical learning models in order to achieve greater predictive capability than could be achieved from any of the individual models.