INTRODUCTION TO QUANTITATIVE INVESTMENT MANAGEMENT

John Nuttall

February 2007

This report consists in Chapter 1 of a brief survey of equity portfolio management, followed by Chapter 2, which contains by an overview of quantitative investing based on the techniques of Machine Learning.

Return to home page

CHAPTER 1

SURVEY OF THE EQUITY MANAGEMENT PROCESS

An investment manager is entrusted with capital from investors that he uses to acquire a portfolio of assets, in this case equities. From time to time the assets may be bought and sold on a market where such assets are traded. The manager chooses the particular assets to hold at any given time from a set of assets known as the universe. The universe is usually specified in advance by the manager. The aim of the manager is to increase the value of the portfolio, and often to meet other performance conditions on the portfolio, such as return relative to a benchmark, volatility, drawdown, etc., as agreed between investors and the manager.

A high proportion of investment managers follow a process that consists of something along the lines of the following steps.

1 INFORMATION COLLECTION

- 1.1 Raw Data
- 1.2 Derived Data Used In Information Processing

2 INFORMATION PROCESSING

- 2.1 Return Prediction - Alternative Methods
  - 2.1.1 Extreme Theoretical
  - 2.1.2 Intermediate
  - 2.1.3 Extreme Empirical
- 2.2 Portfolio Formation
  - 2.2.1 Covariance Matrix
  - 2.2.2 Optimization
  - 2.2.3 Other Approaches to Portfolio Formation
- 2.3 Trading Algorithm

In this report we give only brief comments on these items, except for Sec. 2.1.3, which we discuss in more detail in Chapter 2 "Quantitative Equity Management via Machine Learning".

1 INFORMATION COLLECTION

Presumably every method of investment management depends on knowing certain data before each trading decision is made. The nature of that data will vary widely from manager to manager. Two broad types of data are

- Data that are publicly available, often at a cost from a commercial source. Some examples are given below;
- Proprietary data collected by the manager, such as information about a company's business obtained by interviews with customers, suppliers and competitors.

We cannot say anything useful about proprietary data, and our analysis proceeds on the assumption that the data in question is of the public type.

For actual portfolio management we need to obtain real-time data. In many cases it is also useful, and in some cases essential, to have historical data for back testing simulations. For both these uses we distinguish between

- Raw data that comes directly from the source;
- Derived data, which we define as information formed as combinations or transformations of raw data.

1.1 Raw Data

The raw data in common use fall into the four categories listed below, with a few specific examples.

Market - related information

- Individual company
    - Share price - common
    - Total return
    - Market cap
    - Volume
    - Volume value
    - Option implied standard deviation
- Collective
    - S&P500 price
        - Total return

Financial statement information

- Income statement
    - Net income
    - Income
        - Sales
    - Expenses

        - Interest charge
        - Dividend
- Balance sheet
    - Debt book value
    - Common equity book value
    - Shares outstanding
- Cash flow

External Views about Companies

- Analyst expectations
- Debt rating
- Solvency rating

Data external to the stock market

- T-bill rate
- Industrial production
- Rate of inflation
- Government bonds rate
    - Various maturities

1.2 Derived Data Used In Information Processing

In historical form some of the above data can be extensive. For example it is easy to obtain stock prices and volumes daily, and even higher frequency data are available. To reduce the complication caused by so much data, it is convenient and common to base predictions of stock return on a set of quantities that we refer to as 'predictors', often called factors or sometimes attributes. A predictor is defined at a given time as a function of the raw data available to the manager at that time. We have compiled a list of predictors used by a small number of different managers to predict returns. Two sets of predictors may be found in [Ha96, Ba01].

Note that some predictors are stock specific. In this case, for a given predictor and time, it is common to use the same function of the raw data values to calculate the predictor for all stocks - only the data values change from stock to stock.

Other predictors may be independent of the stock. Examples are market indices and economic data.

2 INFORMATION PROCESSING

The second major component of the investment management process is the transformation of the derived data into stock trading decisions. It is common that this component is split into three stages

- Return prediction - At each time when the portfolio is being revised, we predict the return of each stock in the universe for one or more future periods starting now;
- Portfolio formation - Using the predicted returns we decide how the current portfolio should be modified by changing stock holdings;
- Trading algorithm - Some managers use algorithms to assist in performing the required transactions most effectively.

2.1 Return Prediction - Alternative Methods

Among managers there is a great variety of both the predictor sets and the methods employed to make predictions of the returns of stocks. We categorize prediction methods on the basis of the balance between theory and empiricism, with a broad middle ground between the two extremes.

2.1.1 Extreme Theoretical

In the extreme theoretical approach the manager presumes that he has a theory that describes how the financial economy works, or at least the relevant part of it. This theory must be more than the assumption of many economists that the economy is governed by a stable (stationary) process, so that behavior in the future can be deduced (at least in a statistical sense) from behavior in the past. Rather, the theory tells the manager what will happen in the future if conditions now are different from anything seen before.

It is not possible to test the accuracy of the theory by studying how it might have worked in the past.

2.1.2 Intermediate

Everything not in the other two categories is included here.

2.1.3 Extreme Empirical

In contrast to the theoretical approach, the empiricist believes that markets are too complicated to understand or describe by means of a theory. In this view economics is fundamentally different from physics. However, due to the persistence of the behavior of market participants, patterns may arise in the relation between predictors and returns. These patterns must be discovered by analyzing past data, without any preconceptions as

to the form of the relations. We call this approach "Quantitative Equity Management" (QEM).

QEM may be thought of as an extension of what is often called Technical Analysis (TA) as applied to equities. Usually TA [Pr02] makes predictions and generates trading rules that depend on past market data (mostly prices).

In Chapter 2 we describe the systematic version of QEM that follows from the application of the principles of Machine Learning (ML), a discipline that has been used with success in a wide variety of fields.

2.2 Portfolio Formation

On a given date, let us suppose that we wish to hold each stock in the portfolio for a fixed length of time, and that we have available predictions of the return of all stocks in the universe for that period, beginning now. Following Markowitz [Ma59] it is now commonplace for managers to choose the weights of stocks in their portfolio so that

- Predicted portfolio return is suitably high;
- Predicted portfolio variance, as calculated with the help of the covariance matrix of stocks in the universe, is suitably low;
- Other constraints on the portfolio are also met.

Below we mention

- The meaning of the covariance matrix and how it is calculated;
- How to meet the above conditions on the portfolio - the optimization problem.

We also note alternative approaches to mean-variance optimization (Sec. 2.2.3), and efforts to improve trading effectiveness (Sec. 2.3).

2.2.1 Covariance Matrix

There is an extensive literature on this subject and this area contains material on the following topics.

- The meaning of the covariance matrix. The term implies the existence of a joint probability distribution of the stock returns, but it is rare to find any discussion of how this is defined, or how to check whether a supposed covariance matrix is correct;
- Predicting variance via time series. There is a vast literature on the use of models such as GARCH to predict volatility of an individual asset return. This approach does not seem to have progressed far when applied to the covariance matrix of a lot of stocks;

- Factor and principal component methods for estimating the covariance matrix. These methods aim to find an approximation of small rank to the covariance matrix. In order to make the result non-singular, another matrix is added in an ad hoc manner (e.g. by shrinking [Le04]or adding a diagonal matrix);
- The work of Stanley et al [Pl99] involving random matrices. It tends to support the preceding method, but suggests that the added matrix should be random;
- The view of Michaud [Mi98] on the accuracy of predicted covariance matrices - not very high. He advocates the use of something like the averaging philosophy of Granger [Gr04] and Breiman [Br96];
- A tentative alternative approach. We take the view that anyone who believes that he can predict with useful accuracy the 500,000+ independent elements of the covariance matrix for 1000 stocks is deluding himself. It may be better to face up to the reality that at best just a few of the largest eigenvectors can be predicted, and work with a low rank, singular matrix.

2.2.2 Optimization

Commercial suppliers of estimated covariance matrices, such as Barra, Northfield, APT, Quantal, etc. include an optimization software package. It is also possible to obtain the necessary constrained quadratic programming software from other sources, or write your own, for example using the method described by Collobert/Bengio for SVM [Co01].

2.2.3 Other Approaches to Portfolio Formation

A modest start has been made to the notion of forming the portfolio directly from the predictors. Examples that may touch on this problem are [Ba00, De02, Br05].

2.3 Trading Algorithm

Two recent references on the subject of algorithmic trading are to be found in the Northfield web site [Ba05, In05].

CHAPTER 2

QEM VIA MACHINE LEARNING

1 BASIC STRUCTURE

Here we describe the basic structure of a general approach to equity return prediction based on the formalism and methods of Machine Learning [Mi97, Ha01]. There is room for modification, with the possibility of improvement, of several aspects of the approach. We divide the prediction process into several steps, described in more detail below.

- Sample formation - We collect and organize the information necessary to make the predictions;
- Predictor function - We assume that there is a function of the predictors that provides the 'expected' return (i.e. the return we expect) of all stocks at each time. The predictor function is chosen from a family of possible functions determined by the particular method of ML we employ;
- Training - The process of choosing the parameters describing the predictor function by requiring that it would have worked well in the past;
- Choices - A discussion of some choices that must be made to produce a specific QEM method.

1.1 Sample Formation

First we must describe the structure of the information needed for the approach.

Periods There is a set of times $\{t(j),\ j=1,\ldots\}$ in the past, present and future when we will form a new portfolio. The subscript j labels the date, with higher values corresponding to later periods. The period between successive times might, for example, be days, weeks or months, and we will call them weeks for simplicity.

Stocks At each time $t(j)$ there is a set of stocks $U(j)$ that we are allowed to buy or sell short, chosen from the overall universe. For simplicity we shall assume that the number of stocks in $U(j)$ is independent of j. Call that number N.

Hold Return Suppose we wish to hold stocks in the portfolio for a holding period of length H. Denote the return of stock k for period H starting at time $t(j)$ by R(j,k).

Predictors For each stock label k, at each time $t(j)$, we choose M quantities called predictors $P(i,j,k), i=1,\ldots,M$ that can be calculated in terms of data available at that time. The M predictors for stock k at time $t(j)$ are denoted by the vector $P(j,k)=\{P(i,j,k),\ i=1,\ldots,M\}$.

Samples It is convenient to collect the information that will be used to make predictions into a set of samples. For each stock k and time $t(j)$ we define the sample

$$S(j,k) = \{P(j,k), R(j,k)\}. \quad (1)$$

1.2 Prediction Function

1.2.1 Definition

Our task is to predict the hold return starting at time $t(j)$ of all stocks in the current universe $U(j)$, using only information in those samples that can be calculated at time $t(j)$ or earlier. From the point of view of predicting stock returns, we assume that all significant similarities and differences between stocks at a given time are described by the predictor vector. Thus we make the following important assumption

Assumption There is a prediction function $G[X,t]$, where $X = \{X(i), i = 1,\ldots,M\}$, that gives the predicted hold return starting at time t for any stock that has predictor vector X at that time.

Thus the predicted hold return of stock k at time $t(j)$ is

$$RPRED(j,k) = G[P(j,k), t(j)]. \quad (2)$$

1.2.2 Determination of Prediction Function

The prediction function $G[X,t]$ at any given time t is found by a process known as 'learning' that is a result of 'training'.. That process may use information in any samples such that both the predictor vector $P(j,k)$ and the return $R(j,k)$ are known at time t. That means that the samples must satisfy the requirement

$$t(j) + H \le t. \quad (3)$$

We define $SPRED(j)$ to be the set of samples chosen for use in the learning process at time $t(j)$.

In essence the idea is to find the acceptable function $G[X,t]$ that comes closest to satisfying the relation

$$G[P(ja,k), t(j)] = R(ja,k) \quad (4)$$

for all samples $S(ja,k)$ that we decide to use in the learning process at time $t(j)$, i.e. the set $SPRED(j)$.

The standard approach is to choose a family of possible forms for the prediction function, $GFAM[X,t,W]$, where $W=\{W(i), i=1,\ldots,MW\}$ is a vector of MW parameters $W(i)$ that specify the function. The problem of finding $G[X,t]$ is reduced to the question of finding the particular parameter vector so that $GFAM[X,t(j),W]$ is the acceptable function of X that is nearest to satisfying

GFAM (P(ja,k),t(j),W) = R(ja,k), for all ja, k with $S(ja,k)$ in $SPRED(j)$ (5)

1.2.3 Acceptable Prediction Functions

Before going into details, it is important to remember that the actual return is not likely to be the same as the predicted return. We expect that there will be a distribution of actual returns about those predicted. In other words there is noise in the process, i.e. an element of the return that we cannot predict.

We could certainly choose a form for the prediction function $G[X,t]$ such that (5) was satisfied exactly for all the past samples in $SPRED(j)$. Thus, if MW is large enough, at least as large as the number of samples in $SPRED(j)$, then we can (except in special cases) find a form of $GFAM[X,t,W]$ such that (4) is satisfied exactly for all samples in $SPRED(j)$, but that form is unlikely to be useful in predicting future returns.

Thus we restrict our search for prediction functions to what we loosely call 'acceptable' functions. An acceptable function

- does not have a lot of short scale fluctuations in predictor space;
- it is derived from a $GFAM[X,t(j),W]$ where MW, the number of parameters, is small compared to the number of samples in $SPRED(j)$;

1.3 Training

In order to construct a training process a popular method is to define a quantity that measures the extent to which (5) is not satisfied, and then attempt to choose parameter vector W that minimizes that quantity. For example the mean squared error, MSE, is often used, where

$$MSE(W)=\sum\left(GFAM[Pja,k),t(j),W]-R(ja,k)\right)^2, \qquad (6)$$

the sum running over all ja,k with $S(ja,k)$ in $SPRED(j)$.

The prediction function would then be

$$G[X,t] = GFAM[X,t(j),WMIN], \qquad (7)$$

where $WMIN$ is the value of W that minimizes $MSE(W)$. Note that $WMIN$ will depend on j. In some cases the minimization has to be done numerically (e.g. neural networks), and it can be difficult to find the global minimum of $MSE(W)$.

1.4 Representative Choices

At several points in the general approach we need to make choices, such as

- The length of the holding period H, and whether the return is relative to some collective variable, such as an index, or not;
- The form and number of the predictors $P(j,k)$;
- The set of samples $SPRED(j)$ to be used each period;
- The family of possible prediction functions $GFAM[X,t,W]$;
- The method used in training to minimize the mean squared error MSE.

In the next section we describe some examples of these choices, taken from the academic literature in financial economics and machine learning. Taking one choice from each category gives a quantitative method of return prediction. There are a lot of such combinations.

2 EXAMPLES

Investment managers using quantitative equity management (QEM) techniques usually are inclined to keep them secret. This may be to protect valuable ideas from falling into the hands of their competitors, but there may be cases when secrecy also prevents clients from discovering how flawed the manager's method might be. There may even be cases when the manager does not know what technique he is using, having inherited a computer program from a predecessor.

It is therefore not surprising that we are not aware of any published work that gives details of an operational QEM method that makes serious use of the ideas of Machine Learning (ML). For now we will describe the contents of two instructive sources that utilize crude versions of ML. The authors probably do not think of their methods as having anything to do with ML.

2.1 Haugen

As Breiman [01] suggests, people in many fields, not just financial economics, when faced with the problem of fitting empirical data, automatically use a linear model, since they are not aware of any alternatives. We suspect that this applies to a considerable majority of QEM practitioners, who base their return predictions on a linear function of predictors [Ba01].

2.1.1 Predictors

Haugen describes such a model in some detail in [Ha96]. Returns of a universe of US stocks are predicted every month for a one-month holding period. There appear to be 54 predictors listed on page 436 - Haugen calls the predictor "exposure to a factor". Of these 44 can have values in a continuous range, while 10 have only the values 0, 1 to denote the principal line of business. Before use the predictors are each normalized by a linear Box-Cox transformation, and outliers are removed.

2.1.2 Prediction Function

The chosen form of the family of prediction functions is

$$GFAM[X,t,W] = \sum_{i=1}^{M} W(i)X(i) = W.X \ , \tag{8}$$

where *W*, *X* are *M*-component vectors, and $W \,.\, X$ denotes the scalar product. Haugen calls the parameters *W* "payoffs".

Suppose that we knew the values of parameters $W = WMIN(j)$ for month *j*. Then the predicted return of stock *k* for month *j* would be

$$R(j,k) = WMIN(j).P(j,k) \,. \tag{9}$$

Haugen's procedure for calculating $WMIN(j)$ is as follows.

- For each month prior to month *j* use multi-linear regression to approximate $R(j,k)$ by $R(j,k) = WR(j).P(j,k)$. That is find the vector $WR(j)$ such that the mean square error over all stocks between the two sides of this relation is minimum. To include the possibility of a constant term in (9), an extra predictor always equal to 1 may be included;
- Determine the vector $WMIN(j)$ by extrapolation from $WR(ja),\ ja = j - MS, \ldots, j-1$, where $MS = 12$ in [Ha96]. In that paper the extrapolation is performed by averaging over the values of $WR(ja)$. In his investment advisory firm [Ha05] that appears to be based on the paper, Haugen uses an undisclosed method of extrapolation.

2.1.3 Portfolio Formation

To test the out-of-sample accuracy of the predicted returns, Haugen employs a simple method of forming portfolios each month. Using the formula (9) the predicted return for each stock is calculated and the stocks are divided into ten deciles according to predicted return, with decile 10 being the highest. The returns of equally weighted portfolios formed from each decile are then determined. In Table 2 Haugen displays the annual return for each decile portfolio for the years 1979 - 1993. There is a clear trend each year for the higher deciles to have higher returns. In each of those years a neutral portfolio formed with decile 10 long and decile 1 short would have had a positive return, usually substantial, as a percentage of the amount invested long.

2.1.4 Discussion

There are several points to note about Haugen's results.

- **Trading costs** The returns are calculated on the assumption that stocks could have been bought and sold at the closing price on the last day of each month. This is unrealistic, as there would have been commissions, bid-ask spreads and market impact effects. This issue has been studied by Hanna and Ready [Ha05a].
- **Prediction Errors** There are two sources of error in Haugen's return prediction process, namely the error in approximating a month's stock returns by the linear regression $R(j,k) = WR(j).P(j,k)$, and the error in estimating $WMIN(j)$ by extrapolation from the regression coefficients $WR(ja)$. Among the 44 continuous predictors are 7 'Technical factors' that are defined as the excess return (relative to the S&P 500) of the stock for the previous 1, 2, 3, 6, 12, 24 and 60 months. We have found that, if we use only these 7 predictors, but replace $WMIN$ by WR (unknown when the portfolio is formed) the portfolio returns are much higher. This suggests that the extrapolation process of [Ha96] to find $WMIN$ is not very accurate, and a check of how it worked in the past confirms that view. Presumably Haugen has found a better method of extrapolation at his advisory firm [Ha05].

2.2 Cooper

In Section 3 of a recent article Cooper [Co99] describes the simulation of a method of portfolio management that contains some of the elements of ML. The holding period is one calendar week, and a new portfolio is formed every week. The universe consists of 300 large US stocks. Below we outline Cooper's approach, at times simplified for ease of explanation. In a recent working paper [Ya06] we have demonstrated how the techniques of ML may be used to improve Cooper's method.

2.2.1 Predictors

For week j and stock k there are three predictors $P(j,k)$ (a three component vector), which are

- Return of the stock for the previous week;
- Return of the stock for the week before last;
- The growth in volume from the week before last to the previous week.

2.2.2 Prediction Function

To find the return prediction function for week j we first, for each predictor, divide into deciles the historical distribution of predictor values for the previous 15 years. Using the decile boundary values, the three-dimensional space of predictor values is partitioned into 1000 cells, with each cell corresponding to a choice of three decile numbers. Let us call the cells $CL(i), i = 1,\ldots,1000$.

We define $RCL(j,i)$ to be the average of the one-week hold returns of all stocks in cell $CL(i)$ in the weeks j in the above 15 years. Effectively, although he does not state it explicitly, Cooper assumes that the predicted return of stock *k* for week j is

$$R(j,k) = RCL(j,ia), \tag{10}$$

where $ia = LAB(P(j,k))$, i.e. the label of the cell that contains the predictor vector $P(j,k)$.

This result may be derived in the framework of ML Sec.1.2.2. Thus we choose the family of prediction functions to be

$$GFAM[X,t,W] = \sum_{i=1}^{1000} W(i)STP(X,i), \tag{11}$$

where the step function *STP* is defined as

$$\begin{aligned} STP(X,i) &= 1 \qquad \text{if} \quad i = LAB(X) \\ STP(X,i) &= 0 \qquad \text{otherwise.} \end{aligned} \tag{12}$$

If we apply the procedure of Sec. 1.3 using all the samples in the 15-year period, the result will be

$$W(i) = RCL(j,i), \tag{13}$$

so that (10) is equivalent to (7).

2.2.3 Portfolio Formation

In order to form a portfolio that will test the accuracy of the predicted return Cooper [C099] uses a procedure that is an elaboration of the following method.

- Each week j choose with equal weight all stocks with predictor vector $P(j,k)$ in the 10% of cells *i* with the highest average return $RCL(j,i)$.

The equivalent can be done with the lowest 10% and a neutral portfolio formed. As with Haugen [Ha96] good results are produced before costs.

2.2.4 Discussion

It is difficult to determine just how much the effect of costs would have been. However, we can use the Cooper example to study whether a more sophisticated form of ML can improve predicted return, as measured by portfolio returns before costs.

The ML method described above produces a piecewise constant approximation to the presumably continuous 'correct' predicted return function of the predictor vector. There is a considerable array of ML methods to choose from that do produce continuous functions. They differ from each other in several respects, and in particular in the degree of locality involved. By this we mean the extent to which the value of the predicted return depends on samples that are near to the predictor vector where we require the prediction. Thus the method above may be called pseudo-local in that the value of predicted return depends only on samples in the cell in question. On the other hand, a method such as neural networks is completely global because all samples affect the predicted return at all points.

In practice the highest (and lowest) predicted returns will come from regions of predictor space with a low density of samples, and it is important in the present application that the predicted return is well approximated in these regions. We are not so concerned with having a good approximation in high density regions. Thus a global method that will be dominated by high density regions is probably not what we need. Some sort of pseudo-local method, preferably continuous or nearly so, is likely to be more successful in this application.

We have developed some methods along these lines [Ya06], and indeed they do seem to produce results that are an improvement on the corresponding Cooper method.

REFERENCES

Ba00 Barberis, N., "Investing for the Long Run when Returns Are Predictable," Journal of Finance, **55**, 225 (2000)

Ba01 di Bartolomeo, D. and Warrick, S., "Estimating Nonlinear Effects of Management Styles in the US Equity Market," http://www.northinfo.com/documents/43.pdf (2001?)

Ba05 di Bartolomeo, D., "Optimal Algorithmic Trading", http://www.northinfo.com/documents/173.pdf (2005)

Br96 Breiman, L., "Bagging predictors," Machine Learning, **24**, 123 (1996)

Br01 Breiman, L., "Statistical modeling: The two cultures," Statistical Science, **16**, 199 (2001).

Br05 Brandt, M. W., Santa-Clara, P. and Valkanov, R., "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns," http://faculty.fuqua.duke.edu/~mbrandt/papers/working/paramport.pdf (September 2005)

Co99 Cooper, M., "Filter Rules Based on Price and Volume in Individual Security Overreaction," Review of Financial Studies, **12**, 901 (1999)

Co01 Collobert, R. and Bengio, S., "SVMTorch: Support Vector Machines for Large-Scale Regression problems," J. Machine Learning Research, **1,** 143 (2001)

De02 Dempster, M.A.H. and Romahi, Y.S. "Intraday FX trading: an evolutionary reinforcement learning approach," http://www.jbs.cam.ac.uk/research/working_papers/wp2002.html (2002)

Gr04 Granger, C. W. J. and Jeon, Y., "Thick modeling," Economic Modeling, **21**, 323 (2004)

Ha96 Haugen, R. A., and N. L. Baker, "Commonality in the Determinants of Expected Stock Returns," Journal of Financial Economics, **41**, 401 (1996)

Ha01 Hastie, T., Tibshirani, R. and Friedman, J., "The elements of statistical learning : data mining, inference, and prediction," Springer, New York (2001)

Ha05 Haugen Custom Financial Systems http://www.haugensystems.com/index.html

Ha05a Hanna, J. D. and Ready, M. J., "Profitable predictability in the cross section of stock returns," Journal of Financial Economics **78**, 2005, 463 (2005)

In05 Instinet, "Algorithmic Trading An Overview", http://www.northinfo.com/documents/172.pdf (2005)

Le04 Ledoit, O. and Wolf, M., "Honey, I Shrunk the Sample Covariance Matrix," Journal of Portfolio Management. **30**, 110 (2004)

Ma59 Markowitz, H. M., "Portfolio Selection: Efficient Diversification of Investments," Wiley, New York (1959)

Mi98 Michaud, R. O., "Efficient asset management : a practical guide to stock portfolio optimization and asset allocation," Harvard Business School Press, Boston (1998)

Mi97 Mitchell, T. M., "Machine Learning," McGraw-Hill, New York (1997)

Pl99 Plerou, V., et al, "Universal and Nonuniversal properties of Cross Correlations in Financial Time Series," Physical Review Letters, **83**, 1471 (1999)

Pr02 Pring, M. J., "Technical Analysis Explained," fourth edition, McGraw-Hill New York (2002)

Ya06 Yan, R. J., Nuttall, J., and Ling, C. X., "Application Of Machine Learning To Short-Term Equity Return Prediction," Working paper (2006) available at http://ssrn.com/abstract=888778
or this web site http://publish.uwo.ca/~jnuttall/cooper.pdf