A Glimpse at Quantitative Finance

Patrick Burns*

2nd October 2007

1 Introduction

A common perception of quantitative finance is that it is options pricing and related tasks. While options pricing is one form of quantitative finance, it is not the only slice. And it is not the slice I will talk about.

My background is in applying statistics to equities, especially in regard to fund management. That is the piece of quantitative finance on which I will focus.

Statistics is about extracting information from data that have randomness. That is a description of pretty much everyone who works in finance. If markets don't produce data with randomness, then such a thing doesn't exist. Yet people trained in statistics are a rarity in the field. I'm not suggesting that formal training in statistics should be mandatory. But I do suggest that the tools of statistics and the mindset of statisticians are under-utilized.

2 Statistics Means Models

Proposing models is central to statistics. Typically a model is hypothesized, data are fit to the model, and then the results are examined to see if there is evidence that the model is incorrect. Often this leads to a new model being proposed, which starts the cycle anew.

Once upon a time George E. P. Box made a statement about models:

`All models are wrong, some models are useful.`

That is the most important sentence in this document—remember it.

*This document is available in the finance section of http://www.burns-stat.com.

3 The Efficient Market Hypothesis

Perhaps the Efficient Market Hypothesis is not the concept in finance that has had the most nonsense written about it. If it isn't, it has to be close.

There has been an avalanche of academic papers purporting to prove the Efficient Market Hypothesis. Nonsense. The best that can be done is to put a limit on how far a market is from efficient.

There have been practioner papers stating that the Efficient Market Hypothesis is totally false. Nonsense.

The Efficient Market Hypothesis is a model. All models are wrong. Some models are useful. Some models are useful sometimes.

Index funds are a very useful instrument. They have been a very helpful tool for many investors. Index funds assume the Efficient Market Hypothesis is true and act accordingly.

Hedge funds can be a useful addition to a portfolio. Successful hedge funds have found one or more ways that the Efficient Market Hypothesis is not true. Skeptics will suggest that "good" hedge funds are either lucky or they are bearing some form of extra risk. Yes, that is a possibility—Section 7 will discuss a quite powerful way to test if a fund is really exhibiting skill.

4 Return Models

If the Efficient Market Hypothesis were true, then predicting returns would be impossible. Impossible in the sense that whatever prediction you came up with would not allow you to outperform. The Efficient Market Hypothesis is very close to true, so producing useful predictions is very hard.

Return models have an extremely low signal to noise ratio. This causes serious complications in the statistical practice of developing the models. Overfitting and data snooping are common problems that should be addressed. The Bayesian concept of shrinkage is often useful.

5 Risk Models

While return models are heavily impacted by market efficiency, risk models are not. There is no reason to suppose that risk is especially hard to predict. We can predict risk fairly well. But risk does have its quirks.

A key quirk is the phenomenon of volatility clustering. There are times that the returns of an asset exhibit high volatility and other times of low volatility. Volatility will typically spike up, trend lower, and then spike again. The spikes will be of various heights. Robert Engle shared the 2003 Nobel Prize in economics for creating the ARCH family of models (which includes GARCH) that mimics this behavior and predicts future volatility.

The variance matrix of the returns of a large number of assets is often desired. In this situation we generally want a good estimate (that is, at least positive definite) of the variance matrix when we have many more variables (assets) than observations (dates). This is impossible when proceeding naively. Using a statistical factor model solves the problem. The statistical factor model only uses the history of returns to produce a variance matrix. Factor models can also be built using other data. There are fundamental models that use firm-specific data such as book-to-price. There are also macroeconomic models that use interest rates, exchange rates and so on.

Factor models are not the only possibility for estimating a variance matrix. Another solution is to shrink the sample variance matrix (the one that is not going to be positive definite) towards some other matrix that involves only a few parameters. A choice of the latter that seems good is the equal correlation matrix. The variances (diagonal) in this matrix are the same as in the sample variance, but the covariances are all computed based on the assumption that all of the correlations between assets are equal. The correlation that is used is the average correlation from the sample variance matrix.

6 Other Models

Trading costs are a very important topic for fund management. Market impact models can provide clues about how much a proposed trade is likely to cost.

Consider the case of a company that has two classes of stock. The classes will have somewhat different rights attached to them, but they are both pieces of ownership in the same entity. As the value of the company changes, the value of the stocks will change but there should be a constant offset between the prices of the two classes. If one class becomes cheap relative to the other—perhaps because of a large sell order—then the relative prices of the classes are likely to snap back towards normal. The prices of the two classes of stock are cointegrated. Clive Granger won the other half of the 2003 Nobel Prize for cointegration.

Algorithmic trading involves the creation of return models, but much more as well—such as disguising intention and taking counter-measures against other algorithms.

7 Modern Statistics Means Randomness

Actually modern statistics includes computer-intense estimation schemes like smoothing and Bayesian analyses, but techniques involving randomness are arguably the most striking development.

The key word is Monte Carlo. Monte Carlo creates a random process to answer some question.

The statistical bootstrap is the most common form of Monte Carlo in statistics. The purpose of the bootstrap is to determine the variability of a statistic from a dataset. The bootstrap looks at the value of the statistic of interest on a large number of artificial datasets. The artificial datasets are created by sampling the original dataset with replacement the same number of times as there are observations in the original dataset.

A statistical analysis that merely estimates some quantities is incomplete. The variability of those estimates should also be known. The bootstrap is an easy, and generally good way of finding the variability.

Cross validation is a Monte Carlo technique for evalating the quality of a fitted model. This can ease the problem of overfitting alluded to in Section 4.

A form of Monte Carlo specific to fund management is random portfolios. Given a universe of assets and a set of constraints, random portfolios are a sample from the set of portfolios that satisfy all of the constraints. There are many applications of random portfolios. Some of the most important are: performance measurement, evaluating trading strategies, and setting constraint bounds.

Performance of a fund is often measured relative to either a benchmark (such as the S&P 500) or a peer group of funds. It takes many years to get a significant measure relative to a benchmark, assuming a simple model. The reality is much worse than that—changes in the correlation between returns and benchmark weights makes the measurement much noisier than in the simple model. Sometimes the benchmark is easy to beat, other times it is hard to beat.

There are also problems with peer groups. Selecting a peer group can be quite arbitrary. Even assuming a quality peer group we don't know what the results mean: is being at the 90^{th} percentile of the group 90^{th} percentile in skill or 90^{th} percentile in luck?

Random portfolios solve these problems. They provide a means of rigorously stating the skill that has been exhibited over a given period. Random portfolios are a much more powerful means of performance measurement.

But wait, there's more—yet more power can be extracted with random portfolios by using the initial holdings of a fund in the analysis. This is not possible with the other techniques of performance measurement.

Evaluating trading strategies is really merely performance measurement before a fund is created rather than after. Using random portfolios to test a trading strategy can provide a much clearer picture of how good a strategy actually is.

Constraints on funds are commonly set via guesswork. Random portfolios allow a much more studied approach to setting constraints—the effect of constraints can be examined over whatever time frame data are available.

8 Summary

The continuing surge of computing power means that quantitative finance has an expanding future. There are significant challenges as well. The collapse of Long Term Capital Management in 1998 and the volatility spike in August 2007 highlight the fact that trading on a model changes the behavior of the model. In both these cases observations were made that would be extremely unlikely based on historical data. We need to be able to model the changes in our models induced by using them.