PREFACE - Modeling Financial Markets with Kernel Regression

The use of computers for modeling financial markets has been growing exponentially. Huge sums of money are being managed based purely upon computer models. Data is fed into the models and directives are issued as output. The directives advise the user regarding suggested changes in positions of financial instruments (e.g., stocks, bonds and commodities). The directives are based upon the results of models that predict future price or volatility changes.

This book is directed towards people in the financial community who are involved in evaluating, developing or using market models and trading systems. Kernel Regression (KR) is a modeling technique that is particularly attractive for financial market applications. Although the primary emphasis of this book is on financial modeling, the methodology can easily be applied to any high dimensional modeling problem. KR has many attractive features that make it a modeling technique that should be in the arsenal of any data mining software suite.

Attempts to model time series date back to the 1920's when Yule invented the autoregressive technique for predicting the annual number of sunspots. Since then the general subject of time series analysis has become well established. There are a number of techniques available for modeling time series in general and financial markets in particular. Financial market modeling is typically associated with large amounts of high dimensional multivariate data. Furthermore, the data typically has a low signal to noise ratio and the signals are usually nonlinear. These problems make financial market modeling particularly challenging.

Another major problem associated with financial market modeling is that one really doesn't know which if any related time series are relevant. For example, assume we are trying to develop a model to predict changes in the S&P price index. One could list a variety of series that might affect the S&P index (e.g., short and long term interest rates, commodity prices, etc.). For each related series a number of different indicators can be proposed as candidate predictors for the S&P model (e.g., the one-day fractional changes in the short term interest rate). Other candidate predictors can be suggested based upon several related series (e.g., changes in the ratio of long term to short term interest rates). One can easily develop a set of hundreds of candidate predictors that might or might not be included in the resulting model (or models).

To develop models for financial markets, methodologies are required which allow rapid analysis even though the number of candidate predictors is large. If we start with several hundred candidate predictors we will certainly eliminate most of these prior to completing the modeling task. The methodology should be geared towards finding a subset or subsets of the candidate predictor space that have an acceptable level of predictive power. KR, if applied properly, is an excellent method for developing such models. It can be extremely fast, a basic requirement when one considers the number of subsets of the candidate predictor space which might be examined as part of the modeling process. KR can be used as a stand-alone modeling technique or as a preprocessor for slower techniques such as Neural Networks. (A comparison of KR and Neural Networks is included in Appendix D. The results illustrate the complementary nature of these two modeling methodologies.)

This book is geared to three types of readers. The first group includes those who are interested in modeling in general and desire an overview of the KR technique. The second group includes those involved in the development and/or usage of KR software. The third group includes readers primarily interested in the development of computerized trading systems. The first two chapters include an introduction to the subject and an overview of the modeling process. The next three chapters include the technical details of the KR method. In Chapter 3 the mathematical basis of the KR method is developed. Chapter 4 introduces a data structure that permits an efficient implementation of KR. Chapter 5 provides information regarding the effect of the various parameters upon performance. This chapter is particularly useful for users of KR software. The methodology used in Chapter 5 utilizes artificial data and introduces a general approach to evaluating modeling software that can also be applied to other modeling techniques. Chapter 6 discusses modeling strategies and is relevant to analysts and managers interested in planning a modeling project. Some of the ideas introduced in Chapter 6 can also be applied to other modeling techniques. The final chapter discusses the application of predictions from the computer models to the development of trading systems.

The technical chapters are written for readers who are familiar with college level mathematics but are not necessarily mathematical statisticians. Many of the books on time series are directed towards the statistical community and as a result, the mathematics and notation schemes are difficult to follow for the non-statistician. The emphasis in this book is on application, evaluation and implementation rather than topics of concern primarily to statisticians. Numerical examples rather than theorems and lemmas are used to help the reader understand the details.