14 Recent and future developments in the modelling of financial time series

14.1 Summary of the book

The purpose of this book was to present and explain, at the introductory level, a variety of techniques that are commonly used for the analysis of financial data, including topics that would usually be treated only in a mathematically advanced way. The book commenced with an outline of some stylised characteristics of financial data and described one econometric software package that is widely employed for the financial data exploration. The techniques and models presented included linear models, univariate linear time series approaches, dealing with non-stationary data and long-run modelling, models for volatility and correlation, limited dependent variable approaches, panel data, regime switching models and simulations methodologies. Along the way, examples were presented in each chapter of relevant financial applications from the published literature, and sample instructions or codes for the software package were also given.

14.2 What was not covered in the book

Although this textbook was intended to offer as broad a set of analytical techniques as possible, this in part conflicts with the twin objective of maintaining the book at a manageable length with all of the material at the introductory level so that it can be followed by students completely new to the subject on a one- or two-semester course. Consequently, some interesting and arguably relevant topics have been omitted owing to space constraints. These topics are discussed (with no equations and in no particular order!) below.

Bayesian statistics

The philosophical approach to model-building adopted in this entire book, as with the majority of others, has been that of 'classical statistics'. Under the classical approach, the researcher postulates a theory and estimates a model to test that theory. Tests of the theory are conducted using the estimated model within the 'classical' hypothesis testing framework developed in chapters 2 and 3. Based on the empirical results, the theory is either *refuted* or *upheld* by the data.

There is, however, an entirely different approach available for model construction, estimation and inference, known as Bayesian statistics. Under a Bayesian approach, the theory and empirical model work more closely together. The researcher would start with an assessment of the existing state of knowledge or beliefs, formulated into a set of probabilities. These prior inputs or priors would then be combined with the observed data via a likelihood function. The beliefs and the probabilities would then be updated as a result of the model estimation, resulting in a set of *posterior probabilities*. Probabilities are thus updated sequentially, as more data become available. The central mechanism, at the most basic level, for combining the priors with the likelihood function, is known as Bayes' theorem.

The Bayesian approach to estimation and inference has found a number of important recent applications in financial econometrics, in particular in the context of GARCH modelling (see Bauwens and Lubrano, 1998, or Vrontos *et al*., 2000 and the references therein for some examples), asset allocation (see, for example, Handa and Tiwari, 2006), portfolio performance evaluation (Baks *et al*., 2001).

The Bayesian setup is an intuitively appealing one, although the resulting mathematics is somewhat complex. Many classical statisticians are unhappy with the Bayesian notion of prior probabilities that are set partially according to judgement. Thus, if the researcher set very strong priors, an awful lot of evidence against them would be required for the notion to be refuted. Contrast this with the classical case, where the data are usually permitted to freely determine whether a theory is upheld or refuted, irrespective of the researcher's judgement.

Chaos in financial markets

Econometricians have searched long and hard for chaos in financial, macroeconomic and microeconomic data, with very limited success to date. *Chaos theory* is a notion taken from the physical sciences that suggests that there could be a deterministic, non-linear set of equations underlying the behaviour of financial series or markets. Such behaviour will appear completely random to the standard statistical tests developed for application to linear models. The motivation behind this endeavour is clear: a positive sighting of chaos implies that while, by definition, long-term forecasting would be futile, short-term forecastability and controllability are possible, at least in theory, since there is some deterministic structure underlying the data. Varying definitions of what actually constitutes chaos can be found in the literature, but a robust definition is that a system is chaotic if it exhibits sensitive dependence on initial conditions (SDIC). The concept of SDIC embodies the fundamental characteristic of chaotic systems that if an infinitesimal change is made to the initial conditions (the initial state of the system), then the corresponding change iterated through the system for some arbitrary length of time will grow exponentially. Although several statistics are commonly used to test for the presence of chaos, only one is arguably a true test for chaos, namely estimation of the largest Lyapunov exponent. The largest Lyapunov exponent measures the rate at which information is lost from a system. A positive largest Lyapunov exponent implies sensitive dependence, and therefore that evidence of chaos has been obtained. This has important implications for the predictability of the underlying system, since the fact that all initial conditions are in practice estimated with some error (owing either to measurement error or exogenous noise), will imply that long-term forecasting of the system is impossible as all useful information is likely to be lost in just a few time steps.

Chaos theory was hyped and embraced in both the academic literature and in financial markets worldwide in the 1980s. However, almost without exception, applications of chaos theory to financial markets have been unsuccessful. Consequently, although the ideas generate continued interest owing to the interesting mathematical properties and the possibility of finding a prediction holy grail, academic and practitioner interest in chaotic models for financial markets has arguably almost disappeared. The primary reason for the failure of the chaos theory approach appears to be the fact that financial markets are extremely complex, involving a very large number of different participants, each with different objectives and different sets of information – and, above all, each of whom are human with human emotions and irrationalities. The consequence of this is that financial and economic data are usually far noisier and 'more random' than data from other disciplines, making the specification of a deterministic model very much harder and possibly even futile.

Neural network models

Artificial neural networks (ANNs) are a class of models whose structure is broadly motivated by the way that *the brain performs computation*. ANNs have been widely employed in finance for tackling time series and classification problems. Recent applications have included forecasting financial asset returns, volatility, bankruptcy and takeover prediction. Applications are contained in the books by Trippi and Turban (1993), Van Eyden (1996) and Refenes (1995). A technical collection of papers on the econometric aspects of neural networks is given by White (1992), while an excellent general introduction and a description of the issues surrounding neural network model estimation and analysis is contained in Franses and van Dijk (2000).

Neural networks have virtually no theoretical motivation in finance (they are often termed a 'black box' technology), but owe their popularity to their ability to fit any functional relationship in the data to an arbitrary degree of accuracy. The most common class of ANN models in finance are known as *feedforward network models*. These have a set of inputs (akin to regressors) linked to one or more outputs (akin to the regressand) via one or more 'hidden' or intermediate layers. The size and number of hidden layers can be modified to give a closer or less close fit to the data sample, while a feedforward network with no hidden layers is simply a standard linear regression model.

Neural network models are likely to work best in situations where financial theory has virtually nothing to say about the likely functional form for the relationship between a set of variables. However, their popularity has arguably waned over the past five years or so as a consequence of several perceived problems with their employment. First, the coefficient estimates from neural networks do not have any real theoretical interpretation. Second, virtually no diagnostic or specification tests are available for estimated models to determine whether the model under consideration is adequate. Third, ANN models can provide excellent fits in-sample to a given set of 'training' data, but typically provide poor out-of-sample forecast accuracy. The latter result usually arises from the tendency of neural networks to fit closely to sample-specific data features and 'noise', and therefore their inability to generalise. Various methods of resolving this problem exist, including 'pruning' (removing some parts of the network) or the use of information criteria to guide the network size. Finally, the non-linear estimation of neural network models can be cumbersome and computationally time-intensive, particularly, for example, if the model must be estimated rolling through a sample to produce a series of one-step-ahead forecasts.

Long-memory models

It is widely believed that (the logs of) asset prices contain a unit root. However, asset return series evidently do not possess a further unit root, although this does not imply that the returns are independent. In particular, it is possible (and indeed, it has been found to be the case with some financial and economic data) that observations from a given series taken some distance apart, show signs of dependence. Such series are argued to possess *long memory*. One way to represent this phenomenon is using a 'fractionally integrated' model. In simple terms, a series is integrated of a given order d if it becomes stationary on differencing a minimum of d times. In the fractionally integrated framework, d is allowed to take on non-integer values. This framework has been applied to the estimation of ARMA models (see, for example, Mills, 1999). Under fractionally integrated models, the corresponding autocorrelation function (ACF) will decline hyperbolically, rather than exponentially to zero. Thus, the ACF for a fractionally integrated model dies away considerably more slowly than that of an ARMA model with $d = 0$. The notion of long memory has also been applied to GARCH models, where volatility has been found to exhibit long-range dependence. A new class of models known as fractionally integrated GARCH (FIGARCH) have been proposed to allow for this phenomenon (see Ding, Granger and Engle, 1993 or Bollerslev and Mikkelsen, 1996).

14.3 Financial econometrics: the future?

It is of course, difficult to predict with accuracy what will be the new and important econometric models of tomorrow. However, there are of course topics that are currently 'hot' and which are likely to see continued interest in the future. A non-exhaustive selection of these is discussed below. There are also several survey papers published in academic journals that discuss recent and possible future developments in financial econometrics. Surveys of a technical nature, which are now slightly dated, include those of Pagan (1996) and Tsay (2000). An excellent overview of the state of the art in a vast array of areas in econometrics is provided by Mills and Patterson (2006).

14.3.1 Tail models

It is widely known that financial asset returns do not follow a normal distribution, but rather they are almost always *leptokurtic*, or *fat-tailed*. This observation has several implications for econometric modelling. First, models and inference procedures are required that are robust to non-normal error distributions. Second, the riskiness of holding a particular security is probably no longer appropriately measured by its variance alone. In a risk management context, assuming normality when returns are fat-tailed will result in a systematic underestimation of the riskiness of the portfolio. Consequently, several approaches have been employed to systematically allow for the leptokurtosis in financial data, including the use of a Student's t distribution.

Arguably the simplest approach is the use of a mixture of normal distributions. It can be seen that a mixture of normal distributions with different variances will lead to an overall series that is leptokurtic. Second, a Student's t distribution can be used, with the usual degrees of freedom parameter estimated using maximum likelihood along with other parameters of the model. The degrees of freedom estimate will control the fatness of the tails fitted from the model. Other probability distributions can also be employed, such as the 'stable' distributions that fall under the general umbrella of extreme value theory (see Brooks, Clare, Dalle Molle and Persand, 2005 for an application of this technique to value at risk modelling).

14.3.2 Copulas and quantile regressions

As discussed in chapter 2, covariance and correlation provide simple measures of association between series. However, as is well known, they are very limited measures in the sense that they are linear and are not sufficiently flexible to provide full descriptions of the relationship between financial series in reality. In particular, new types of assets and structures in finance have led to increasingly complex dependencies that cannot be satisfactorily modelled in the classical framework. *Copulas* provide an alternative way to link together the individual (*marginal*) distributions of series to model their joint distribution. One attractive feature of copulas is that they can be applied to link together any marginal distributions that are proposed for the individual series. The most commonly used copulas are the Gaussian and Clayton copulas. They are particularly useful for modelling the relationships between the tails of series, and find applications in stress testing and simulation analysis. For introductions to this area and applications in finance and risk management, see Nelsen (2006), Alexander (2008, chapter 4) and Embrechts *et al*. (2003).

The possibility of application in the risk management arena has also stimulated renewed interest in another rather old technique, which has now become fashionable, known as *quantile regression*. Dating back to Koenker and Bassett (1978), quantile regression involves constructing a set of regression curves each for different quantiles of the conditional distribution of the dependent variable. So, for example, we could look at the dependency of y on x in the tails of y's distribution. This set of regression estimates will provide a more detailed analysis of the entire relationship between the dependent and independent variables than a standard regression model would (see Koenker, 2005). The latter would only be sufficient in the context that the dependent and independent variables followed a bivariate normal distribution. Taylor (1999) and Engle and Manganelli (2004) use quantile regression for value at risk estimation, while Alexander (2008) provides a novel application to hedging.[1]

14.3.3 Market microstructure

One of the most rapidly evolving areas of financial application of statistical tools is in the modelling of market microstructure problems. 'Market microstructure' may broadly be defined as the process whereby *investors' preferences and desires are translated into financial market transactions*. A comprehensive survey is given by Madhavan (2000). He identifies several aspects of the market microstructure literature, including price formation and price discovery, issues relating to market structure and design, information and disclosure. There are also relevant books by O'Hara (1995), Harris (2002) and Hasbrouck (2007).

Research efforts in this area have been motivated by enhancements in computer technology, which have improved the quality and quantity of available data. Trends towards 'globalisation' have implied that investors are increasingly looking beyond their own shores in the search for higher returns or more efficient diversification. It is also likely that the number of exchanges will reduce considerably over the next decade or two, so it is therefore essential that the new exchanges be organised optimally.

At the same time, there has been considerable advancement in the sophistication of econometric models applied to microstructure problems. An important innovation was the Autoregressive Conditional Duration (ACD) model due to Engle and Russell (1998). An interesting application can be found in Dufour and Engle (2000), who examine the effect of the time between trades on the price-impact of the trade and the speed of price adjustment.

It is also evident that microstructure is important since it potentially impacts on many other areas of finance. For example, market rigidities or frictions can imply that current asset prices do not fully reflect future expected cashflows (see the discussion in chapter 9 of this book). Also, investors are likely to require compensation for holding securities that are illiquid, and therefore embody a risk that they will be difficult to sell owing to the relatively high probability of a lack of willing purchasers at the time of desired sale. Measures such as volume or the time between trades are sometimes used as proxies for market liquidity.

[1] Quantile regression is available in EViews version 6 – see *EViews User's Guide II*, chapter 31.

14.3.4 *Computational techniques for options pricing and other uses*

The number and complexity of available derivative securities has increased enormously over the past decade, and this expansion continues today. There are now many examples of financial options, for example, whose payoffs are so complex that an analytical formula for valuing the option is not available. Consequently, alongside developments in the mathematics of option pricing formulas, interest in new computational techniques, for example based on lattice or simulations methods, has surged. New theoretical models have been proposed, such as those including 'jumps' in the data generating process for the underlying asset (see, for example, Amin, 1993 or Naik, 1993).

Computational speed and power continues to increase rapidly, such that problems which were previously infeasible even with a supercomputer can now be accomplished using a desktop PC. This augurs well for the continued expansion of the application of *simulation methods* in economics and finance. Researchers' understanding of the properties of simulations-based estimators is also improving as the body of knowledge and cumulated experience in this area grows. In econometrics, the simulation of large multivariate GARCH or switching models is now within the realms of possibility. Similarly in finance, real-time Monte Carlo scenario analysis for risk management models could now be conducted.

Computational advancements have also led to enhancements in the quality and quantity of databases that can be used in financial econometrics. For example, just a few years ago, the notion of holding a large database of high frequency financial data covering tick-by-tick observations on thousands of companies would have been unthinkable. Such large data sources are becoming more and more readily available as the costs of obtaining, storing and retrieving the information falls. This is likely to lead to significant new contributions in the area of real-time analysis, market microstructure, examination of technical trading rules, and so on.

14.3.5 *Higher moment models*

Research over the past two decades has moved from examination purely of the first moment of financial time series (i.e. estimating models for the returns themselves), to consideration of the *second moment* (models for the variance). While this clearly represents a large step forward in the analysis of financial data, it is also evident that conditional variance specifications are not able to fully capture all of the relevant time series properties. For example, GARCH models with normal (0,1) standardised disturbances cannot generate sufficiently fat tails to model the leptokurtosis that is actually observed in financial asset returns series. One proposed approach to this issue has been to suggest that the standardised disturbances are drawn from a Student's t distribution rather than a normal. However, there is also no reason to suppose that the fatness of tails should be constant over time, which it is forced to be by the GARCH-t model.

Another possible extension would be to use a conditional model for the third or fourth moments of the distribution of returns (i.e. the skewness and kurtosis, respectively). Under such a specification, the conditional skewness or kurtosis of the returns could follow a GARCH-type process that allows it to vary through time. Harvey and Siddique (1999, 2000) have developed an autoregressive conditional skewness model, while a conditional kurtosis model was proposed in Brooks, Burke, Heravi and Persand (2005). Such models could have many other applications in finance, including asset allocation (portfolio selection), option pricing, estimation of risk premia, and so on.

An extension of the analysis to moments of the return distribution higher than the second has also been undertaken in the context of the capital asset pricing model, where the conditional co-skewness and co-kurtosis of the asset's returns with the market's are accounted for (e.g., Hung *et al*., 2004). A recent study by Brooks *et al*. (2006) proposed a utility-based framework for the determination of optimal hedge ratios that can allow for the impact of higher moments on the hedging decision in the context of hedging commodity exposures with futures contracts.

14.4 The final word

I wrote in the previous edition of this book that it was probably fair to say that there had been a hiatus in the development of new econometric techniques for the analysis of financial data over the past decade; seven years on, I still believe this is true. Arguably, the majority of recent developments in financial econometrics have involved improvements in both the quantity and quality of applications, rather than the development of entirely new techniques. The last decade has not, for example, seen the development of new classes of models on the grand scale of those for cointegration or ARCH.

It is clear that an ideal model for asset returns, which is intuitive to interpret and easy to estimate yet which is able to adequately describe all of the stylised features of the data at hand, has yet to be discovered. Maybe you will find it!