ISSN 1011-8888

INSTITUTE OF ECONOMIC STUDIES

WORKING PAPER SERIES

W07:06 July 2007

Review of Econometric Modeling Approaches in Finance

Helgi Tomasson

Address: Helgi Tomasson
Faculty of Economics and Business Administration
University of Iceland
Oddi, at Sturlugata, 101 Reykjavik
Iceland

Email: helgito@hi.is

Review of Econometric Modeling Approaches in Finance

Helgi Tómasson

University of Iceland

Faculty of Economics and Business Administration

helgito@hi.is

July 2, 2007

1 Introduction

The aim of this review is to give a brief review of the statistical tools, models and fundamental concepts that are available for financial data analysis. The approach is set up as an index of basic concepts for the quantiatively minded. This review is inevitably very brief as both finance and statistics are large subjects.

Finance is: "the science that describes the management of money, banking, credit, investments, and assets, basically, finance looks at anything that has to do with money and the market".
(http://financial-dictionary.thefreedictionary.com/finance)
Still a slightly different definition is: "A discipline concerned with determining value and making decisions. The finance function allocates resources, including the acquiring, investing, and managing of resources" (Harvey & Morgenson 2002)

One definition of statistics is: "The mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling".
(http://www.thefreedictionary.com/statistics).
At http://dictionary.laborlawtalk.com/Statistics, the definition given is: "Statistics is the science and practice of developing human knowledge through the use of empirical data. It is based on statistical theory which is a branch of applied mathematics". Within statistical theory, randomness and uncertainty are modelled by probability theory. Because one aim of statistics is to produce the "best" information from available data, for eventual policy making, some authors consider statistics a branch of decision theory. Statistical practice includes the planning, summarizing, and interpreting of observations, allowing for variability and uncertainty. Statistics can be seen as a probability application where the main task is data analysis and giving general statements about an unknown reality. Because finance is a branch of economics, the practice of deriving estimators and tests in this context, is frequently referred to as financial econometrics. Analyzing properties of estimators and tests are important subjects in the statistical discipline.

The discipline of finance has been expanding in the direction of mathematics in the second half of the 20th century. Probably the first stochastic modelling approach in finance took place by Bachelier (1900) in an attempt of formalizing unpredictability of price movements. In physics Einstein (1905) used similar ideas for describing unpredictability of movements of particles. It was though not until Wiener (1923) gave a proper mathematical background to the ideas of Bachelier and Einstein and proved that the Wiener-process was a well defined mathematical concept. The ability to forecast in the stock market would be very valuable if it was possible. The efficient market makes the scope for profitability of forecasting very limited. Academically the problem of forecasting prices and poor performance of experts in financial markets is treated at least as early as Cowles (1933). The notion of unpredictability of prices is present in the statistical literature as early as Kendall (1953). The mathematical approach initiated by Bachelier, Einstein and Wiener in the early 20th century, was rediscovered in a famous article by Black & Scholes (1973). The advance culminated in the 1997 Nobel Prize of economics, when Merton and Scholes were awarded the prize "for a new method to determine the value of derivatives".

This development has shifted the focus of finance theory and increased its level of mathematical sophistication. A great number of books has appeared on use of stochastic models in finance, such as Merton (1990), Karatzas & Shreve (1991), Hull (1993), Wilmott et al. (1995), Neftci (1996), Duffie (1996), Shiryaev (1999), Revuz & Yor (1999), Cvitanic & Zapatero (2004),Björk (2004), etc., the list is very long.

Progress in computer technology and telecommunications has made data much more accessible and transferable than before. The combination of mathematics, data and computer power has opened the field of applied statistics to financial applications. Methods that before were considered utopian both for reasons of lack of data and mathematical complexity are now feasible.

The statistical discipline of linking data and models has responded to the availability of new data and new theory. Collections of articles relating statistical methods to finance are e.g., Maddala & Rao (1996), Hand & Jacka (1998), Chan,

Keung & Tong (2000). Examples of recent textbooks on applied data analysis in finance are Tsay (2002) and Zivot & Wang (2003).

At http://en.wikipedia.org/wiki/Finance the following definition is given: "Finance studies and addresses the ways in which individuals, businesses and organizations raise, allocate, and use monetary resources **over time**, taking into account the risks entailed in their projects. Therefore, the statistical methods for analysis of financial data are highly focused on the importance of time. This is also the case in surveillance where the emphasis is on making timely decisions. As the time factor is extremely important in finance, statistical methods involving time dependency are crucial for financial data analysis. Statistical models involving time dependency rely on the probability theory of stochastic processes. Modern finance also relies on the theory of stochastic processes. Therefore basic knowledge of stochastic processes is essential for the quantitative financial analyst, as well as for the theoretical one. Stochastic processes can be classified by the nature of the state space and the nature of the time-index. The state can be continuous or discrete and the time-index can as well be continuous or discrete. For statistical analysis the most common, and best known tool, is time-series theory, which often refers to the case where the state is continuous, but time is discrete.

The organization of this review is as follows. Section 2 gives a brief background of financial theory for financial markets. In Section 3 a brief review of the principal concepts of classical linear equi-spaced time-series is given. For the theory of finance volatility (standard deviation) of price, is important, both for pricing and for risk management. Therefore models for second moments are of interest. The popular discrete-time approach in modelling that is the ARCH-class of models. Since the appearance of the ARCH model in Engle (1982), many models based on the same idea, focusing on second-moments (variances), have been derived. In Section 4 a brief review of some of the ARCH-family models is given. Discrete time-series models for equi-spaced time-series, ARMA, GARCH, etc., are now easily applicable by use of widespread computer software. The ARCH models are a special kind of non-linear/semi-linear models.

The general class of non-linear models is very large so, therefore it is necessary to limit the functional form. The general class of non-linear is simply to large. In Section 5 some simple non-linear or semi-linear models that have been suggested in the literature is reviewed.

The modern finance literature is dominated by continuous-time models. The continuous-time mathematics approach offers a powerful tool for logical reasoning about a dynamic environment. The benefit of an empirical approach to continuous-time statistical models is that the interpretation will correspond to a theoretical model. In practice, however, the approach is problematic. In practice a time-continuous pattern is never observed, there is always some discretization involved. Even if one could obtain a continuous observation, one would have to integrate a continuous pattern to obtain a value of the likelihood function. The continuous-time models can be classified as continuous-path models or models with jumps.

Bergstrom (1988, 1990) gives an historic overview on use of continuous time-models in econometrics. Bergstrom (1990) mentions that the beginning of statistical analysis of continuous-time models might be traced back to Bartlett (1946), but the main-stream econometric literature might have missed this result, perhaps due to the then recent discovery of treatment of simultaneous models by Haavelmo (1943), which dominated econometric methodology for thirty years. In recent years impressive progress has been made in statistical treatment of discretely observed diffusion processes. Recently, statistical approaches to time-continuous diffusion models have also become feasible. Some properties of these models are briefly reviewed as well as approaches for estimating unknown parameters based on real, discretely observed data. An outline of some ideas is given in Section 6.

Traditional duration-data models/transition-data/survival models analyze jump-only processes, i.e., the process moves between a finite number of alternate states. Transition data models and extreme value models arise naturally when statistical aspects of financial data are analyzed. In Section 7 transition data models are briefly reviewed.

The interest of second moments, or equivalently, volatility (standard deviation) is frequently connected with the fashionable term “risk”. The ARCH type models deal with dynamics of second moments. A related type of risk is the “extreme-value” analysis, the probabilistic nature of extremes. Just as in engineering, catastrophes in finance take place, firms go bankrupt etc. In Section 8 some references to financial applications of extreme-value theory are reviewed.

The finance literature has recently taken steps towards memoryless processes that are partly continuous, i.e. the jump-diffusions or Lévy processes. It is in the nature of statistical inference that it will always be very hard to distinguish between a jump, a steep climb and a heavy tailed distribution. Some aspects of this and some references are given in Section 9.

This summary is written from an econometrics/statistics point of view. It does not contain any review of the literature on technical analysis, artificial intelligence or machine learning. Quantitative analysis of financial data is performed in these fields. The author suspects that the calculations in those disciplines are to a degree similar, but the concern about the probabilistic nature of "the model” is less apparent, and the interpretation of results will differ from the statistical way of thinking.

All practical analysis of financial data is evidently computer-dependent. Data are obtained and treated electronically and a vast arsenal of optimization methods is now implemented in the research centers of financial data analysts. The numerical methods are of great variety. There exist deterministic, iterative, methods, such as, Newton-type methods for maximization and solving equations, and some stepwise methods like EM algorithm, auxilary regression, double-length regression etc. Examples of simulation based methods are MCMC(Markov-Chain-Monte-Carlo) and particle filtering. The emphasis in this review is on the categories of practically feasible models and only to a small degree on the technical implementation, such as, how to get output, estimates, tests etc. for the model/data.

2 A brief background of financial markets problems

In theoretical finance several assumptions of a 'perfect' market are necessary (Merton 1990, p. 477). Among these assumptions several are quite unrealistic, like no transactions costs, investor can always buy or sell as much as he wants of any asset (including borrowing or lending money) at any point in time, market is always in equilibrium. A key assumption is the efficient market hypothesis. Roughly expressed, the efficient market hypothesis states that 'all information' is already included in the prices, and therefore that prices should not be predictable. This non-predictability feature is formalized in the mathematical modelling by the inclusion of a non-predictable stochastic process, the Wiener process (the Brownian motion). How this efficiency is implemented in practice is somewhat debated. Blackwell, Griffiths & Winters (2006) give the following classification. First, the strong-form efficiency, which states that all information (private and public) is embedded in the security price. Second, the semi-strong efficiency, that all public information is available in the security price. And third, the weak-form efficiency, that all past information is included in the security price. In academic finance at least the weak-form of efficiency is required for arbitrage-free pricing. The logic is that if prices were predictable, the agents would quickly discover that and give higher bids on securities that are likely to increase in price. That kind of behaviour would eliminate predictability.

It turns out that enforcing arbitrage-free conditions in financial modelling results in pricing functions where the variance function is a fundamental part. Therefore the variance function is of principal interest for investment-strategy, risk-management, etc. Surveillance of the mean-function in finance is though not uninteresting. For example financial-inspection authorities have the role of monitoring the possibility of insider-trading. For surveillance, therefore both surveillance of mean and variance are of interest for surveillance in finance.

3 Linear time-series analysis

A data-set, $(x_1, \ldots, x_T)$, is a realization of set of random-variables, $(X_1, \ldots, X_T)$. The simplest case is when the random-variables are iid. The theory time-series deals with the situation where there is a dependency structure based on the time-sequence of the sampling. In order to obtain consistent estimates it is necessary to assume that dependency fades away in the sense that observations far apart in time are almost independent. The mathematical term for this property is ergodicity. As only one realization is available, some stability property of X_t is also necessary. The theoretical concept for that is stationary, which means that the dependency structure is invariant over time. In practice weak-stationarity is assumed for reasons of convenience. Weak-stationarity is characterized by covariance-stationarity, i.e. the mean and auto-covariance functions are assumed to be constant over time.

$$E(X_t) = \mu,$$
$$E(X_t - \mu)(X_s - \mu) = \gamma(|t-s|) = \gamma(k), \quad k = |t-s|.$$

The ergodicity assumption states that the dependence between X_t and X_s should decrease as $|t-s|$ increases, i.e. that random-variables very far apart in time should be virtually independent. Several definitions of ergodicity exist, but for estimation of mean and auto-covariance it is essential to require mean-ergodicity:

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow[T\to\infty]{a.s} \mu,$$

where $\xrightarrow[T\to\infty]{a.s}$ denotes almost sure convergence, and auto-covariance ergodicity (see, Wei (1990)):

$$\frac{1}{T}\sum_{t=k+1}^{T} (X_t - \mu)(X_{t-k} - \mu) \xrightarrow[T\to\infty]{a.s} \gamma(k).$$

If the data are assumed to be a sample from a stationary normal process all information about the process is expressed by the mean, $\boldsymbol{\mu}$ and the auto-covariance matrix $\boldsymbol{\Gamma}_T$. Optimal prediction is in theory easily derived from the properties of the conditional normal distribution. If the vector $\boldsymbol{X}$ partitioned, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$,

(future and past) as:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim N(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}'_{12} & \boldsymbol{\Gamma}_2 \end{bmatrix}). \tag{1}$$

Then the conditional distribution of the future, $\boldsymbol{X}_1$, given the past, $\boldsymbol{x}_2$, is:

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Gamma}'_{12}\boldsymbol{\Gamma}_2^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}'_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{12}). \tag{2}$$

Equation (2) is easily interpretable but there are technical difficulties concerning representation and computations for a stationary process. Even if the auto-covariance function, $\gamma(k)$, was known the formula involves inverting a large matrix. A popular representation of a stationary process is by approximating it with an ARMA(p,q) process:

$$\begin{aligned}(X_t - \mu) = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \\ \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}.\end{aligned} \tag{3}$$

A more compact notation by use of the backward operator, B, $BX_t = X_{t-1}$, is frequently used. Sometimes the backward operator is called the lag-operator, L. Equation (3) is written more compactly in polynomials of the B operator.

$$\begin{aligned}&\Phi(B)(X_t - \mu) = \Theta(B)\varepsilon_t, \\ &\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p, \\ &\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q.\end{aligned}$$

This parameterization is not unique, in the sense that another set of ϕ's and θ could generate the same auto-covariance function. Requiring *invertibility*, i.e. that the roots of the polynomial $\Theta(z)$ lie outside the unit circle solves that issue. Requiring that the roots of $\Phi(z)$ lie outside the unit circle guarantees stationarity. It is also required that the polynomials $\Phi(z)$ and $\Theta(z)$ do not have any common factors. The auto-covariance function, $\gamma(k)$, for a particular ARMA(p,q) process, is a complicated function of the parameters $(\phi_1, \ldots, \phi_p, \theta_1, \theta_q, \sigma)$. It can be calculated for example by the Durbin-Levinson algorithm, see, e.g. Brockwell & Davis (1991). The process X_t is a *filtered* version of the ε_t process. Sometimes the properties of an ARMA process are better visualized by means

of spectral methods. The spectrum is defined as the Fourier transform of the auto-covariance function,

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-ik\lambda} \gamma(k).$$

The spectral density function $f(\lambda)$ is an excellent tool for describing the cyclical properties of X_t. A peak in $f(\lambda)$ at λ_0 indicates a cycle of length $2\pi/\lambda_0$. Something that is much harder to visualize in terms of the auto-covariance function. The ARMA(p,q) representation is a flexible way of parameterizing a stationary process. This flexibility has made ARMA models very popular in applied work, based on the idea of approximating stationary processes by ARMA(p,q). The seminal book by Box & Jenkins (1970) is essentially a cookbook, based on statistical principles, on how to proceed from data to forecast. Box & Jenkins (1970) extended the idea to be applicable to processes that could be transformed into ARMA(p,q). Their recommendations were to transform the process by a variance stabilizing transform, e.g. taking logarithm, and the date differences. The idea is to approximate a transformed version of a process by an ARIMA(p,d,q) process:

$$(1 - B)^d \Phi(B)(X_t - \mu) = \Theta(B)\varepsilon_t.$$

Their approach was pragmatic rather than theoretical. It consists of steps, the first step which they called *identification*, consisted of choosing a proper variance stabilizing transform, the number of differences, d, to take, and values of p, and q. The second step they named *estimation* and consisted of obtaining estimates of the ϕ's and θ's. The third step was called *diagnostics* and consisted of analyzing $\hat{\varepsilon}_t = \hat{\varepsilon}_t(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$. The modelling process was considered a success if the properties of empirical residuals seemed similar to what was assumed about the theoretical residuals, i.e. white-noise. If the diagnostic step was passed one proceeded to the fourth and final step, the *forecasting* step. The forecasting step consisted of calculating point forecast and corresponding interval forecasts.

Both the minimum-mean-square-error-prediction, X_{T+h} and its conditional variance, the variance of e_{T+h} is in theory easily derived from the fundamental results on normal models given by equation (1). Numerically it is however more

practical to use some recursive methods, the Durbin-Levinson algorithm, the innovation algorithm, or use the Kalman-filter recursions or use the Cholesky decomposition of the variance-covariance matrix (Brockwell & Davis 1991).

Since the early BJ practice progress has taken place in the numerical procedures. For the normal model an algorithm for calculating the likelihood of an ARMA process was given by Gailbraith & Gailbraith (1974) and shortly after a computationally efficient algorithms were given by e.g, Ansley (1979) and Melard (1983). The Kalman-filter algorithm also offers an easy way of calculating the likelihood value.

The approach of approximating non-stationary series with an ARIMA process suggested the idea of a fractionally-integrated process:

$$(1 - B)^d \Phi(B) X_t) = \Theta(B) \varepsilon_t$$

where d need not be an integer, see e.g. Granger & Joyeux (1980). If d is in the interval $(-0.5, 0.5)$ the process is stationary. The behaviour of the autocorrelation function differs from the usual, $d = 0$, case where the autocorrelation function decays exponentially:

$$|\rho(k)| < r^{-k},$$

and is instead the autocorrelation function decays polynomially:

$$|\rho(k)| < k^{2d-1}.$$

One may distinguish between the case $d < 0$ where $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ and the case $0 < d < 0.5$ where $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$. Brockwell & Davis (1991) label the former case as an intermediate-memory process and the latter as a long-memory process. The maximum-likelihood estimation is treated in Sowell (1992) and Beran (1995). Beran (1994) has published a detailed book on long-memory modelling. A recent empirical investigation of the usefulness of ARFIMA, the AR-fractionally-IMA, for macroeconomics and finance is given by Bhardwaj & Swanson (2006).

The availability of cheap computing power and efficient algorithms for calculating the likelihood function has made exact maximum-likelihood estimation

feasible and to a degree made the BJ scheme less formal, i.e. the identification step, the estimation step and the principle of parsimony have essentially merged into one.

One of the virtues of the BJ scheme was simplicity. When the BJ scheme is generalized to the multivariate case the simplicity disappears. One representation of a multivariate ARMA model is:

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_1 \boldsymbol{X}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{X}_{t-p} + \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1} - \cdots - \boldsymbol{\varepsilon}_{t-q}, \tag{4}$$

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0} \quad t \neq s. \tag{5}$$

The auto-correlation function $\boldsymbol{\Gamma}(k) = E(\boldsymbol{X}_t \boldsymbol{X}_{t-k}')$ is now a sequence of matrices so the plotting of auto-correlation function is a non-trivial matter. Equation (4) in polynomials of the backward operator B is:

$$\boldsymbol{\Phi}(B)\boldsymbol{X}_t = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t.$$

Requiring that the roots of the polynomials $|\boldsymbol{\Phi}(z)|$ and $|\boldsymbol{\Theta}(z)|$ lie outside the unit circle and that the polynomials $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$ have no common left factors are conditions that are inherited from the univariate case. A way to get a minimal multivariate ARMA representation is to minimize the McMillan-degree (Hannan & Deistler 1988).

The computational complexity of multivariate ARMA models is evidently much higher than that of the univariate case, but calculation of likelihood, predictors, etc., is just a technicality. Multivariate extension of the Durbin-Levinson algorithm and the innovation algorithms exist and can be programmed in a modern programming language. State-space representation and the Kalman-filter provide an approach that is perhaps the easiest to implement, (Brockwell & Davis 1991; Harvey 1989, 1993; Lütkepol 1991). Another possible approach is to consider the multivariate system as a periodic-ARMA, PARMA process, (Lund & Basawa 1999; McLeod 1993; Pagano 1978).

The non-stationary case for multivariate is more complicated than in the univariate case. The degree of non-stationarity can vary across coordinates. In the univariate case the class of $ARIMA(p, d, q)$ has proven to be class of non-stationary models that is large enough to be interesting. BJ suggested a very

crude method of estimating an integer value of d. Since the article by Dickey & Fuller (1979) there has been a huge development of unit-root tests. Granger was awarded the Nobel Prize in economics in 2003, "for methods of analyzing economic time series with common trends (cointegration)", which is a way of analyzing the relations between non-stationary time-series. The literature on cointegration is now huge.

Classical time-series analysis in the ARMA-spirit is essentially about linear filters,

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k}. \tag{6}$$

The case were the input, ε_t, is iid-normal is completely treated in textbooks. The literature on deviations from iid-normal is basically in two directions, a) ε_t uncorrelated, but somehow dependent and b) ε_t independent, but the distribution not normal. One of the stylized facts on financial time series is that their tails are heavier than would be allowed by a normal model. A simple way to incorporate that into (6) is to assume that ε_t is heavy-tailed, e.g. some t-distribution. Then X_t will be distributed as a weighted sum of independent t-distributions, which is certainly not a t-distribution, but a rather messy compound. Inference in such model is though relatively straightforward, because the likelihood can be calculated recursively. If the input series, ε_t, has finite variance, then the output series X_t will consist of a weighted sum of independent finite variance components and might therefore look "more normal" than the input due to central-limit-theorem arguments.

In the case of normal input, the filtered process, X_t, is also normal because the normal family is closed under addition. For continuous random variables this property defines the family of stable-distributions, i.e., the sum of iid variables from the family belongs also to the family. The Cauchy distribution has also this property. The normal distribution is the only continuous stable distribution which has finite variance. Therefore if the input in (6) is iid stable, but not normal, X_t does not have finite variance. Obviously, a criterion like MMSEP (minimum-mean-square-error-of-prediction) will not make sense in such cases. The density of the stable distributions is not available in closed form, but the

logarithm of the characteristic function is of the form:

$$log(E(e^{iuX})) = \begin{cases} iu\beta - d|u|^{\alpha}(1 - i\theta\,\mathrm{sgn}(u)\tan(\pi\alpha/2)) & \alpha \neq 1, \\ iu\beta - d|u|(1 + 2i\theta/\pi)\,\mathrm{sgn}(u)\log(|u|) & \alpha = 1. \end{cases}$$

The interpretation of the parameters is, β indicates location, $d^{1/\alpha}$ indicates scale, θ indicates symmetry and α indicates tail behaviour. Recently algorithms for numerically calculating the density, generating random numbers have become available. (Lambert & Lindsey 1999). The method of (Lambert & Lindsey 1999) has been implemented in the R-package ((R Development Core Team 2005)) stable. Figure 1 shows some examples of density functions for stable dis-

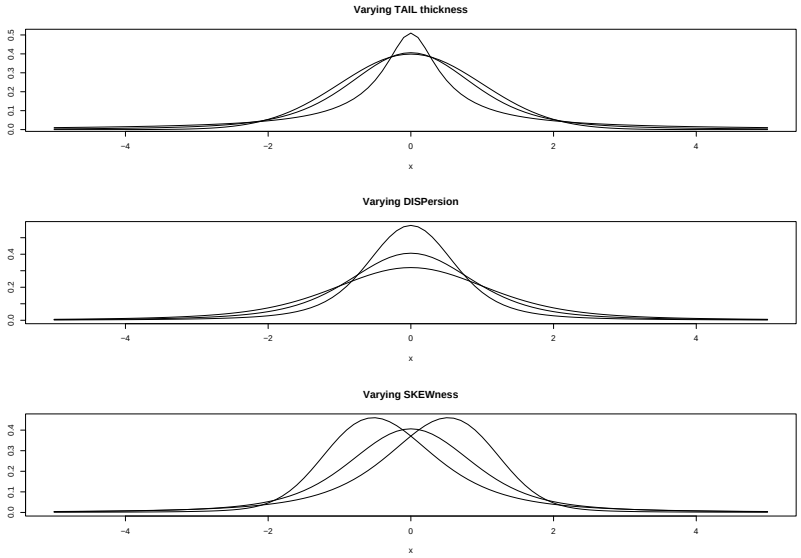

Figure 1: Examples of stable density functions.

tributions. A textbook treatment of linear-filters with infinite variance input is given in Brockwell & Davis (1991), chapter 13.3. For more advanced treatment, see, e.g., Hall, Peng & Yao (2002).

4 Conditional Heteroskedacity

In 2003 R.F. Engle was awarded the Nobel Prize in economics "for methods of analyzing economic time series with time-varying volatility (ARCH)". It became apparent that the dependency structure of a process was more than

just the auto-covariance function. Granger (1983) enlightened that some properties of a white-noise process could be predicted. Engle (1982) analysed a particular case and introduced the concept of Auto-Regressive-Conditional-Heteroskedacity (ARCH) models. A simple version of ARCH is:

$$\varepsilon_t = v_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2}, \tag{7}$$
$$\text{with } E(v_t) = 0, \quad E(v_t^2) = 1, \quad v_t \sim N(0,1).$$

By some rearranging,

$$\varepsilon_t|\varepsilon_{t-1},\ldots \sim N(0,h_t), \quad h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2,$$
$$\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \underbrace{(\varepsilon_t^2 - h_t)}_{\text{residual}}. \tag{8}$$

Equation (8) shows the similarity between ARCH models an AR models. The distribution of the residual in (8) is restricted by the dynamic structure and the fact that $\varepsilon_t^2 > 0$. It is therefore, necessarily non-normal. The second moments of ε_t have a non-zero autocorrelation structure but the first moments do not. A straightforward extension of (7) is the ARCH(p):

$$\varepsilon_t = v_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2}.$$

The ARCH(p) process is easily interpreted as an AR(p) process for second moments an ARMA version of that is GARCH(p,q), generalized-ARCH:

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 + \beta_1 h_{t-1} + \cdots + \beta_q h_{t-q}. \tag{9}$$

Equation (8) can be generalized and the GARCH(p,q) process, (9), can be written as:

$$\Phi^*(B)(\varepsilon_t^2 - \alpha_0) = \Theta^*(B)\text{residual}_t$$

where Φ^* and Θ^* are polynomials. So the development of the GARCH models can easily be interpreted as a spin of from ARMA modelling for squared processes. An $AR(p)$ process with p large can be well approximated by an $ARMA(p,q)$ process with $p+q$ small. The BJ principle of parsimony also applies to $GARCH$ processes and practitioners have therefore often preferred

$GARCH(1,1)$ because a reasonable fit to real data by $ARCH(p)$ usually requires a large p. Figure 2 illustrates the features of models from the ARCH/GARCH family. The process generates volatility clusters, the auto-correlation structure in X_t is weak whereas it is clear in X_t^2. Other ideas like the persistence prop-

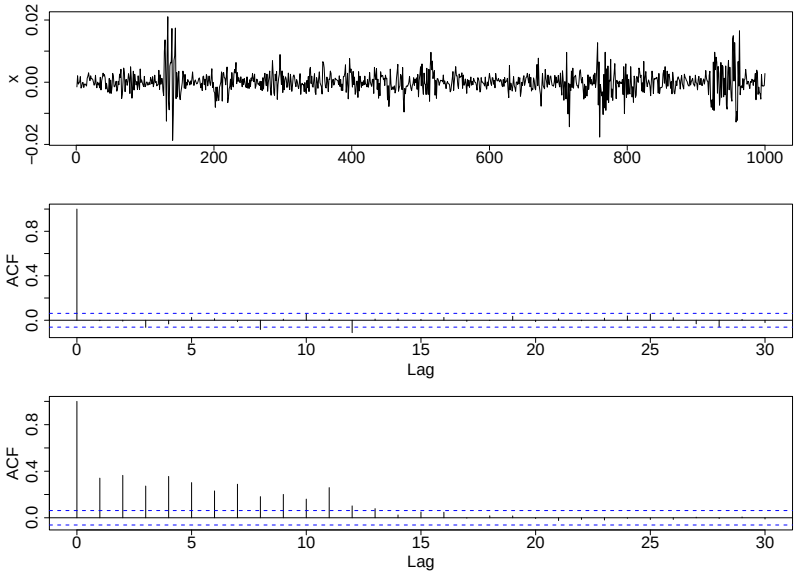

Figure 2: A simulated GARCH(1,1) process (top), its auto-correlation (middle) and the auto-correlation of the series squared (bottom).

erty, i.e. unit root, a basic idea of ARIMA modelling can be incorporated in the ARCH framework such as the integrated-GARCH, IGARCH, (Engle & Bollerslev 1986). The simplest form is the IGARCH(1,1), which is a constrained version of (9):

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1) h_{t-1}.$$

The ARIMA contains a unit root and is non-stationary. The IGARCH contains a unit-root but can be strictly stationary, (Nelson 1990). The long-memory idea of ARIMA, the ARFIMA has also found its way into the volatility modelling and has been termed FIGARCH (fractionally-integrated-GARCH), (Baille, Bollerslev & Mikkelsen 1996).

The step from a univariate model to multivariate model for a GARCH is badly hit by the curse of dimensionality. The multivariate generalization ARCH

refers to a n-dimensional vector, $\boldsymbol{\varepsilon}_t$, which is conditionally heteroskedastic:

$$H_t = V(\boldsymbol{\varepsilon}_t | \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots).$$

Defining the dynamics of the sequence of matrices H_t, analogously to the one dimensional case results in the following recursions for the (k,l) element of H_t:

$$h_{k,l,t} = c_{k,l} + \sum_{i=1,q} \left[\sum_{m=1}^{n} \sum_{s=1}^{n} \alpha_{k,l,m,s,i} \varepsilon_{m,t-i} \varepsilon_{s,t-i} \right] + \sum_{i=1}^{p} \left[\sum_{m=1}^{n} \sum_{s=1}^{n} \beta_{k,l,m,s,i} h_{m,s,t-i} \right]. \tag{10}$$

There are some natural constraints on the parameters in equation (10) due to the fact that H_t is a covariance matrix and therefore symmetric positive-definite. But nevertheless, the number of parameters grows dramatically with the number of dimensions. For a multivariate ARCH(1) model the number of parameters is $1 + n(n+1)/2$, (Gourieroux 1997). Therefore more parsimonious versions have been used. A variant suggested by Bollerslev (1990) is the constant-conditional-correlation-GARCH, CCC-GARCH:

$$h_{i,i,t} = c_{i,i} + \alpha_{ii} \varepsilon_{i,t-1}^2 + \beta_{i,i} h_{i,i,t-1},$$
$$h_{i,j,t} = \rho_{i,j} h_{i,i,t}^{1/2} h_{j,j,t}^{1/2}, \quad \text{for } i \neq j.$$

More constrained version exist, e.g. DVEC-GARCH, (Yang & Allen 2005).

5 Non-linear time-series models

The ARMA model is a linear-filter, i.e. the current value, X_t, is a linear function, a weighted sum, of past values of the series, $X_{t-1}, X_{t-2}, \ldots$ and current and past noise-values, $\varepsilon_t, \varepsilon_{t-1}, \ldots$. A generalization of this would be a non-linear filter, so that a non-linear univariate time-series, X_t can be represented as e.g.:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, \varepsilon_t, \ldots), \quad \text{where } \varepsilon_t \text{ is some noise or,} \tag{11}$$

$$\varepsilon_t = g(X_t, X_{t-1}, \ldots), \quad \text{or} \tag{12}$$

$$X_t = h(\varepsilon_t, \varepsilon_{t-1}, \ldots). \tag{13}$$

The function in (11) represents a mixed non-linear ARMA representation, in a kind of non-linear AR is represented by (12) a non-linear MA in equation

(13). All three are in a non-anticipative form, i.e. they represent X_t only as a function of its past values and past and current ε_t's. It is evidently quite hopeless to estimate a very general function of the above form from a single realization of a time series. Some intelligent starting point is needed as well as sensible bounds on how complicated the function can be. A starting point is to use h in (13) and assume that a Taylor expansion is allowed. Following a classic monograph on the fundamentals of non-linear time series, Priestley (1991), h in (13) is Taylor-expanded:

$$X_t = \mu + \sum_{i_1=0}^{\infty} g_i \varepsilon_{t-i} + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} g_{i_1,i_2}\varepsilon_{t-i_1}\varepsilon_{t-i_2} + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} g_{i_1,i_2,i_3}\varepsilon_{t-i_1}\varepsilon_{t-i_2}\varepsilon_{t-i_3} + \cdots \quad . \tag{14}$$

The Volterra expansion, (14) suggests that a reasonable starting point could be a kind of bilinear model:

$$X_t = \mu + \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_{12}\varepsilon_{t-1}\varepsilon_{t-2} \tag{15}$$

or as Granger & Andersen (1978) suggest:

$$X_t = \varepsilon_t + \alpha\varepsilon_{t-1}X_{t-2},$$

which have zero-autocorrelation and are therefore, not linearly predictable, but might be non-linearly predictable. The optimal predictor for (15) is $\alpha_{12}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-2}$ if $\hat{\varepsilon}_t$ is constructed recursively by:

$$\hat{\varepsilon}_t = X_t - \alpha_1\hat{\varepsilon}_{t-1} - \alpha_{12}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-2}.$$

From the look of (14) it is clear that search for a parsimonious non-linear representation of a process in the spirit of BJ for linear processes will be complicated, as well as problematic, in terms of identification. Therefore the design of non-linear time-series modelling has developed towards models that have the ability of capturing particular stylized features of data. One such example is the threshold-auto-regressive (TAR) model. The idea is that the dynamics of a process X_t is different when the level of X_t is high than when it is low. A

simple first order TAR(1) model is:

$$X_t = \begin{cases} \phi_{H,1} X_{t-1} + \varepsilon_{H,t} & \text{if } X_{t-1} \geq d, \\ \phi_{L,1} X_{t-1} + \varepsilon_{L,t} & \text{if } X_{t-1} < d \end{cases} .$$

Where $\phi_{H,1}, \phi_{L,1}, \varepsilon_{H,t}, \varepsilon_{L,t}$ refer to the high-level and the low-level case, respectively. Obviously generalizations to more complicated TAR are theoretically straightforward. The practical implementation gets quickly difficult, choosing the number of thresholds, lag-length, etc. Tong (1983) gives strategies for practical modelling, such as limiting the decision to jump to a single specific time-lag, common for all thresholds. Tong (1983) calls this, SETAR (self-excited-threshold-autoregressive) (Priestley 1991). Properties such as stationary distributions, auto-covariance function, properties of estimators are non-trivial, (Jones 1978; Klimko & Nelson 1978; Tong 1983). Many modern researchers would use computer intensive methods, such as bootstrap or simulation. The TAR model is varying-parameter model, i.e., the parameter jumps if the series passes a certain value. Some may find that feature unwanted and therefore some alternatives have been developed. One version is the EAR (exponential-autoregressive) model, a simple version of EAR(2) is:

$$\begin{aligned} X_t &= \phi_1(X_{t-1}) + \phi_2(X_{t-2}) + \varepsilon_t, \\ \phi_1(X_{t-1}) &= \alpha_{1,1} + \alpha_{2,1} \exp(-\alpha_{3,1} X_{t-1}^2), \\ \phi_2(X_{t-2}) &= \alpha_{2,1} + \alpha_{2,2} \exp(-\alpha_{3,2} X_{t-2}^2). \end{aligned}$$

The EAR(2) can behave very similar to an AR(2). When the characteristic polynomial of the AR(2) has complex roots, the process tends to show cyclical behavior. The EAR is to a degree similar to TAR but the coefficients, $\phi_1, \phi_2, \ldots$ evolve smoothly between the minimum ($\alpha_{1,i}$) and the maximum value ($\alpha_{1,i} + \alpha_{2,i}$). Therefore, the EAR can generate amplitude-dependent cycles, jump-like behavior, limit-cycles. For more details see, e.g., Ozaki (1982) and Ozaki (1985). Identifiability is obviously an issue, e.g. it will be difficult to estimate $\alpha_{2,i}$ and $\alpha_{3,i}$ when $\alpha_{2,i}$ is small and $\alpha_{3,i}$ is large. The EAR is likely to capture similar features of a series, such as the TAR. Many practitioners use AIC/BIC criteria, or to a degree, common sense, in choosing between models. Still a

slightly modified version is the STAR (smooth-transition-autoregressive) model. A simple form is:

$$X_t = \alpha_1 X_{t-1} + X_{t-1} F(X_{t-1}) + \varepsilon_t \tag{16}$$

where F is a suitable function. A possible choice of F is the logistic function and then the model (16) is labelled LSTAR (logistic-STAR).

The above models, TAR/EAR/STAR, etc., have in common, that they switch regimes depending on the observed time-series X_t. A related idea is to let another process rule the regime switching. An idea is the Markov switching regime, Hamilton (1989):

$$X_t = \begin{cases} \text{model 1} & S_t = 1 \\ \text{model 2} & S_t = 2 \end{cases},$$

where the state S_t is ruled by a Markov-chain with transition probabilities:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}.$$

The modelling consists of estimating the number of states, the transition matrix as well as the dynamic model in each state. Hamilton (1994) reviews some aspects of a practical approach, estimating parameters, singularities in the likelihood etc.

The usual GARCH model can be thought of as a kind of linear filter of second moments. The non-linear ideas above have been brought over to volatility modelling. Nelson (1991) suggests that:

$$\log(h_t) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(\varepsilon_{t-k}),$$

in order to accommodate for the asymmetric relation between volatility, $\sigma_t = \sqrt{h_t}$, and prices, i.e. markets react differently to negative shocks than to positive shocks. Nelson (1991) suggest that by choosing:

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma(|\varepsilon_t| - E(|\varepsilon_t|)),$$

the series σ_t can be a well-behaved process, depending on the choice of g. This turns out to be very similar to e.g. the TAR process, (Tsay 2002). A related

version is the PGARCH (power-GARCH) mixing the long-memory property with non-linearity:

$$\sigma_t^d = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i})^d + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^d.$$

The PGARCH form includes many members of the GARCH family as special cases.

The tremendous creativity in the models in this category has been driven by the wish to represent the stylized facts of a financial market in simple ARMA-type formulas. First and second moments of series are modelled and of course the two are mixed, e.g., AR-GARCH, ARCH-M etc. A brief list of the ARCH family is shown in table 1. A recent text on nonlinear time-series is Fan & Yao (2003).

PGARCH	AGARCH	TGARCH	IGARCH
EGARCH	AR-ARCH	ARCH-M	BEKK
CCC	FIGARCH	PSD-VECH	DVEC
FIEGARCH	NGARCH	VGARCH	QGARCH

Table 1: A subset of the ARCH family .

6 Continuous time diffusions

The field of finance has evolved strongly in the past decades. A corner-stone element of the theory is the Wiener-process, or Brownian-motion. The Wiener-process is a continuous-path, memoryless process. If the time-horizon is $t \in [0, 1]$:

$$W(0) = 0,$$
$$W(t_4) - W(t_3) \text{ indpendent of } W(t_2) - W(t_1), \quad t_1 < t_2 < t_3 < t_4,$$
$$V(W(t)|W(0)) = t,$$
$$W(t) \text{ continous with probability 1.}$$

The process can be defined similarly for any time-interval. According to the functional-central-limit-theorem $W(t)$ is normally distributed. The Wiener process is in the literature often called Brownian motion after the biologist Brown (1827), who was describing movements he observed in his microscope. The term Wiener-process is due to the mathematician Wiener (1923) who proved that the process was mathematically well defined. The mathematical literature on Wiener/Brownian process is huge.

The theory of differential equations is aimed at describing dynamics, i.e. movement of a particle in time. Time-series models in discrete time can be thought of as difference equations containing a stochastic term, i.e. the input noise. Following the same spirit in the continuous-time case gives rise to the need to define the concept stochastic-differential-equation (SDE). This is done by defining SDE through the concept of stochastic integral. The SDE is written as:

$$\mathrm{d}X(t) = \mu(X(t), t)\,\mathrm{d}t + \sigma(X(t), t)\,\mathrm{d}W(t). \tag{17}$$

The interpretation of (17) is that it has a solution on the form:

$$X(t) = X(t_0) + \underbrace{\int_{t_0}^{t} \mu(X(s), s)\,\mathrm{d}s}_{\text{term 1}} + \underbrace{\int_{t_0}^{t} \sigma(x(s), s)\,\mathrm{d}W(s)}_{\text{term 2}}. \tag{18}$$

The functions μ and σ are called trend and diffusion, respectively. The first term in equation (18) is an ordinary Riemann-integral. The second term is a

stochastic integral. The most commonly used concept for a stochastic integral is the Ito integral. If the function σ was a step function which only jumped in $t_1, \ldots, t_n = t$, the Ito-integral is defined as:

$$\int_{t_0}^{t} \sigma(s) dW(s) = \sum_{k=1}^{n} \sigma(t_{k-1})(W(t_k) - W(t_{k-1})).$$

The key issue here is to define the integral based on the value of σ the left end of the interval. That way the independent increment property of W makes formulas simpler, such as the variance:

$$V(\int_{t_0}^{t} \sigma(s) dW(s)) = \sum_{k=1}^{n} [\sigma(t_{k-1})]^2 (t_k - t_{k-1}),$$

all covariance terms disappear due to the independent increments of $W(t)$ and the forward increment feature of the definition of the integral. The definition is then extended to functions that can be approximated by a step function so that the Ito-integral is defined for a class of "well-behaved" functions. This mathematical background means that (17) is just another way of writing a stochastic integral. Another definition of a stochastic integral is the Stratonovic-integral. Having a working definition of the stochastic integral activates a vast mathematical machinery. In many ways the continuous-time approach is more tractable than the discrete-time one. The $\mathrm{d}W(t)$-term in (18) plays the role of ε_t, i.e. the white-noise in the discrete-time models. The term $\mathrm{d}W(t)$ is often called the continuous-time white-noise, even it does not exist mathematically, it is still a useful form which refers directly to the integral. A useful property of the Ito-integral is that it is a martingale. Another virtue of using the Ito-integral as a definition of the stochastic integral is the practically of the Ito-lemma, if the dynamics of $X(t)$ is given by (18) then the dynamics of $Y(t) = g(X(t), t)$ is given by:

$$\begin{aligned}
&\mathrm{d}Y(t) = \mu^*(X(t), t)\,\mathrm{d}t + \sigma^*(X(t), t)\,\mathrm{d}W(t), \\
&\mu^*(X(t), t) = \\
&\left[\frac{\partial g}{\partial t}(X(t), t) + \mu(X(t), t)\frac{\partial g}{\partial x}(X(t), t) + \frac{1}{2}\sigma^2(X(t), t)\frac{\partial^2 g}{\partial x^2}(X(t), t)\right] \\
&\sigma^*(X(t), t) = \sigma(X(t), t)\frac{\partial g}{\partial x}(X(t), t).
\end{aligned}$$

If g is a well-behaving process the dynamics of transformed version of the process is also driven by, a "normal white-noise", $\mathrm{d}W(t)$. This is very different from the discrete-time case. If a discrete-time model is driven by white-noise, a function of it is in general not driven by white noise.

Having defined the white-noise term, $\mathrm{d}W(t)$, in terms of the Ito-integral it is possible to define a continuous-time linear filter, i.e., a continuous-time ARMA, as:

$\Phi(D)X(t) = \sigma\Theta(D)DW(t)$ where D is a differential operator

but because $W(t)$ is nowhere differentiable, a convenient representation for calculations is the state-space form

$X(t) = \sigma\boldsymbol{\theta}Z(t)$

$$\mathrm{d}X(t) = AX(t)\,\mathrm{d}t + \mathbf{1}\,\mathrm{d}W(t), \quad t > 0$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\phi_p & -\phi_{p-1} & -\phi_{p-2} & \cdots & -\phi_1 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{p-1} \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

The dynamics of the state vector is given a linear SDE which has the solution:

$$Z(t) = e^{At}Z(0) + \int_0^t e^{A(t-s)}\mathbf{1}\,\mathrm{d}W(s)$$

$$e^A = I + \sum_{k=1}^{\infty} \frac{A^k}{k!}$$

Standard rules for variance gives the variances of the state-vector, and as the state-vector is the sum (integral) of normal components it is normally distributed (given $Z(0) = 0$):

$$Z(t) \sim N(0, \int_0^t e^{As}\mathbf{1}\mathbf{1}'e^{A's}\,\mathrm{d}s) \tag{23}$$

If the integral converges, the process is stationary, and we have an invariant distribution. Sufficient conditions are that the real-part of the eigenvalues of A are all negative.

As a contrast to the discrete-time case, the step from linear dynamics to non-linear dynamics is more manageable in the continuous-time case. Many

one-dimensional non-linear SDE are tractable. The mathematical literature on SDE and its application to finance is huge. The mathematical conditions on the existence of a solution of a SDE depend on the nature of the functions μ and σ. One distinguishes between strong and weak solutions. Good mathematical references are, Karatzas & Shreve (1991); Øksendal (1998); Revuz & Yor (1999).

Visualization of a diffusion-process can be done by simulation. The simulated process is essentially, a step function with frequent jumps, i.e. the diffusion-process is approximated by a process that jumps at discrete time-points, $t_1, \ldots, t_n$. A well known simulation scheme is the Euler scheme, which is simply based on substituting independent standard pseudo-normal random variables, Z_i, into to equation (18),

$$X(t_i) = X(t_{i-1}) + \mu(X(t_{i-1}, t_{i-1})\Delta_i + \sigma(X(t_{i-1}, t_{i-1})\underbrace{Z_i\sqrt{\Delta_i}}_{\widehat{\mathrm{d}W(t_i)}}, \tag{24}$$

$$\Delta_i = t_i - t_{i-1}, \quad V(\widehat{\mathrm{d}W(t_i)}) = \Delta_i. \tag{25}$$

The feature $V(\mathrm{d}W(t)) = \mathrm{d}t$ is reflected in the simulation by equation (25). The quality of the simulation depends on how fine the mesh, Δ_i is, as well the complexity of the process, i.e. μ and σ. An improved version, which is based on Taylor approximations, is the Milstein-scheme:

$$\begin{aligned} X(t_i) &= X(t_{i-1}) + \mu(X(t_{i-1}, t_{i-1})\Delta_i + \sigma(X(t_{i-1}, t_{i-1})Z_i\sqrt{\Delta_i} \\ &+ \frac{1}{2}\sigma(X(t), t)\frac{\partial\sigma}{\partial x}(X(t_{i-1}, t)((Z_i\sqrt{\Delta_i} - Z_{i-1}\sqrt{\Delta_{i-1}})^2 - \Delta_i). \end{aligned}$$

Kloeden & Platen (1992) give higher order approximations as well as multivariate version of the approximations. These are of course only approximations. The simulated process is constant or somehow interpolated between the time-points. In the general case simulation is hindered due to the fact, that the transition density is generally unknown, and therefore direct sampling from the transition density is impossible. Beskos & Roberts (2005) and Beskos., Papaspiliopoulos, Roberts & Fearnhead (2006) give an algorithm for sampling from the exact transition density.

The practical situations of estimation is more complicated than in the discrete-time case. The process is continuous, but in practical cases only discrete obser-

vations are available. A continuous-time approach based on deriving estimators for case that data consists of an entire path is shown in Kutoyants (1984). The log likelihood-function of a continuously observed process is given by (26):

$$log(L(\boldsymbol{\theta}|t_0,t)) = c + \int_{t_0}^{t} \frac{\mu(X(s),s)}{\sigma^2(X(s),s)}\,\mathrm{d}X - \frac{1}{2}\int_{t_0}^{t} \frac{\mu^2(X(s),s)}{\sigma^2(X(s),s)}\,\mathrm{d}s. \tag{26}$$

Implementation of methods such as maximizing (26) would require dense observations. In the (usual) case of discrete observations:

$$x(t_1),\ldots,x(t_n), \quad t_1 < t_2 < \cdots < t_n,$$

an idea could be to simulate the pattern between observations and replace the integrals with sums. The traditional statistical approaches for seeking estimators, method-of-moments, least-squares, maximum-likelihood and Bayesian methods are all complicated. The existence of moments and which moments to match is complicated. Calculation of the likelihood, $L(\boldsymbol{\theta})$, by recursively calculating the transition density $f(x(t_i)|x(t_{i-1})$, using:

$$L(\boldsymbol{\theta}|x(t_1),\ldots,x(t_n)) = f(x(t_1),\boldsymbol{\theta})\prod_{i=2}^{n} f(x(t_i)|x(t_{i-1}),\boldsymbol{\theta}),$$

is possible due to the nature of the process. But as closed form for the transition density are only available for some particular processes, in general, some approximations are necessary. Some popular models are shown in table 2.

As pointed out in Duffie (1996), page 132, some of these models are special cases of equation (27).

$$dX(t) = [\alpha_1(t) + \alpha_2(t)X(t) + \alpha_3(t)X(t)\log(X(t))]\,dt + [\beta_1(t) + \beta_t(t)X(t)]^{\nu}\,\mathrm{d}W(t) \tag{27}$$

Typically the parameter-space is restricted. Some of the model in table (2) have an invariant distribution, i.e. they are stationary. It is usually easy to decide whether a one-dimensional diffusion process has a stationary distribution. When modelling dynamic phenomena, that are believed to have some stationary features, some simple stationary diffusions could be used as starting points.

In some cases calculations of moments are easy. The form (18) is essentially a way of writing the integral:

$$X(t) = X(t_0) + \int_{t_0}^{t} \mu(X(s),s)\,\mathrm{d}s + \int_{t_0}^{t} \sigma(X(s),s)\,\mathrm{d}W(s) \tag{28}$$

Model	$\mu(x)$	$\sigma(x)$	support
Wiener-process with drift	α	σ	$-\infty < x < \infty$
Ornstein-Uhlenbeck (Vasicek)	$\alpha(\beta - x)$	σ	$-\infty < x < \infty$
Cox-Ingersoll-Ross (Square-root)	$\alpha(\beta - x)$	$\sigma\sqrt{x}$	$0 < x < \infty$
CEV	$\alpha(\beta - x)$	σx	$0 < x < \infty$
CKLS	$\alpha(\beta - x)$	σx^{ρ}	$0 < x < \infty$
Duffie Kan (1996)	$\alpha(\beta - x)$	$\sqrt{\sigma + \rho x}$	$0 < x < \infty$
Brennan Schwartz (1979)	$\alpha x(\beta - \log(x))$	σx	$0 < x < \infty$
Marsch Rosenfeld (1983)	$\alpha x + \beta x^{-(1-\rho)}$	$\sigma x^{\rho/2}$	$0 < x < \infty$
Constantinides (1982)	$\alpha + \beta x + \rho x^2$	$\sigma + \rho x$	$0 < x < \infty$
Double well	$\alpha x - \beta x^3$	σ	$-\infty < x < \infty$
Generalized logistic	$\big[(\alpha - \beta)\cosh(\frac{x}{2})$ $-(\alpha + \beta)\sinh(\frac{x}{2})\big]\cosh(\frac{x}{2})$	$2\cosh(\frac{x}{2})$	$-\infty < x < \infty$

Table 2: Some popular diffusion processes.

Using the forward nature of the Ito-integral, the conditional expected value of $X(t)|X(t_0)$ can be calculated by taking expectation through (28). So:

$$E(X(t)|X(t_0)) = X(t_0) + \int_{t_0}^{t} E(\mu(X(s), s))\,\mathrm{d}s \tag{29}$$

Ito-lemma give the dynamics of $Y(t) = X(t)^2$ and taking expectation again gives:

$$E(X(t)^2|X(t_0)) = \\ X(t_0)^2 + \int_{t_0}^{t} \left[2E(X(s)\mu(X(s), s) + E(\sigma^2(X(s), s))\right]\mathrm{d}s$$

In the case when $\mu(x)$ is linear in x the first conditional moment is derived by using Ito-lemma and solving a differential equations, e.g. if $\mu(x) = \alpha(\beta - x)$ then (29) becomes:

$$m(t) = E(X(t)|X(t_0)) = \\ X(t_0) + \int_{t_0}^{t} \alpha(\beta - E(X(s)|X(t_0))\,\mathrm{d}s = X(t_0) + \int_{t_0}^{t} \alpha(\beta - m(s))\,\mathrm{d}s,$$

$m'(t) = \alpha(\beta - m(t)),$ i.e. $m(t) = X(t_0)\mathrm{e}^{-\alpha(t-t_0)} + \beta(1 - \mathrm{e}^{-\alpha(t-t_0)})$.

For some particular $\sigma(x)$ the derivation of conditional second moments can be just as simple, i.e. a question of solving a differential equation. In the general case this approach is not feasible.

For a one-dimensional diffusion the invariant distribution is (in the case that it exists) on the form:

$$f(x) \propto \frac{1}{\sigma^2(x)} \mathrm{e}^{\int_c^x \frac{2\mu(s)}{\sigma^2(s)}\,\mathrm{d}s}.$$

Again this might be hard to evaluate in some cases. Mao, Yuan & Yin (2005) give some numerical methods. When using stationary diffusion processes in modelling the stationary distribution should reflect features to be matched with the scientific phenomenon of interest.

Moment-based methods have their appeal because for some models moments can be calculated in closed form, even if transition probabilities cannot. Moment-based versions are e.g. GMM, EMM, SMM and some methods based on estimating functions. Bibby, Jacobsen & Sørensen (2004) give a review on the use of estimating functions.

Some non-parametric, semi-parametric, partly parametric approaches have been tried. If the existence of an invariant distribution $f(x)$ is assumed, then the relation between the invariant distribution, the drift function and the diffusion function is:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sigma^2(x))f(x) = 2\mu(x)f(x),$$
$$\sigma^2(x) = \frac{1}{f(x)} \int_{-\infty}^{x} 2\mu(s)f(s)\,\mathrm{d}s,$$

Aït-Sahalia (1996) suggests estimating the drift parametrically, estimating the invariant distribution with a kernel method and then plug $\hat{\mu}(x)$ and $\hat{f}(x)$ into (30) to get a non-parametric estimate of the diffusion function $\sigma(x)$. Aït-Sahalia (1999, 2002) approximates the transition density by Taylor expanding the Kolmogorov-forward equation, and substitutes the approximation for the true likelihood and maximizes it numerically. For affine diffusions the characteristic function is often manageable. It is possible to use numerical Fourier inversion to derive the likelihood. An example of an estimation procedure based

on using the characteristic function for affine diffusions is given by Singleton (2001). Nonparametric approaches based on kernel smoothing are also conceivable. Fan (2005) gives an overview of nonparametric methods in financial econometrics.

7 Analysis of duration and transition data

A lot of financial activity consists of waiting for a particular event to occur such as a shift of state. A principle statistical discipline of analyzing waiting time is survival analysis. The survival time, T, the waiting time for a shift in the state of survival is a non-negative random variable. A standard form for describing the risk for a change of state is the hazard function, $\lambda(t)$:

$$\lambda(t) = P(t < T < t + \mathrm{d}t | T > t) = \frac{f(t)}{1 - F(t)},$$
$$F(t) = P(T \leq t) = 1 - \mathrm{e}^{\int_0^t \lambda(s) ds}, \quad f(t) = F'(t).$$

The hazard function denotes the instantaneous risk of change of state conditional on the state up to time t. There exists extensive literature on survival analysis where the focus is on studying the hazard of the one-way transition from life to death, (Andersen, Borgan, Gill & Keiding 1993; Fleming & Harrington 1991).

The hazard function can in principle be any non-negative function. It should be noted that if the integral

$$\int_0^\infty \lambda(s) ds$$

is finite, then that means that there is positive probability of eternal life. When dealing with biological data it is usually not realistic to analyse a sequence of survival times for an individual. In sociological data it is conceivable to observe e.g. the sequences between jail sentences for criminals but, in general there are so few time spells that it is not fruitful to derive a dynamic structure.

Engle & Russell (1998) defined the ACD, autoregressive conditional duration model for explaining dynamics of waiting times between transactions in a financial market. If transactions take place at time-points $t_1, t_2, \ldots$, the ACD

approach is basically:

$$x_i = t_i - t_{i-1},$$
$$E(x_i|x_{i-1}, x_{i-2}, \ldots) = \psi_i(\boldsymbol{\theta}, x_{i-1}, x_{i-2}, \ldots). \tag{30}$$

Equation (30) denotes the conditional expectation of duration number i and is a function of past durations and a parameter vector $\boldsymbol{\theta}$. The conditional probability model for X_i is supposed to be of the form:

$$X_i = \psi_i \varepsilon_i,$$

where ε_i is a sequence of iid variables with a parametric distribution and parameter $\boldsymbol{\phi}$, e.g., an exponential or a Weibull distribution. The dynamic functional form of ψ_i is given by:

$$\psi_i = \omega + \sum_{j=1}^{m} \alpha_j x_{i-j} + \sum_{j=0}^{q} \beta_j \psi_{i-j}.$$

The abbreviation for this type of model is ACD(m,q). There is a striking resemblance in the derivation of the ACD model and the derivation of the ARCH model. As in the case of ARCH models many new abbreviations have been generated for describing the various forms of auto-correlated durations. Examples are AACD of Fernandes & Grammig (2006) which is based on applying the idea of Box-Cox transformation (flexible forms), to the ACD dynamics:

$$\psi_i^{\lambda} = \omega + \alpha \psi_{i-1}^{\lambda} \left[|\varepsilon_{i-1} - b| + c(\varepsilon_{i-1} - b)\right]^{\nu} + \beta \psi_{i-1}^{\lambda}. \tag{31}$$

Equation (31) is an example of AACD(1,1). The idea is that the AACD should contain a lot of other ACD-variants as special cases. The form is decided by the parameters, (b, c, λ, ν). Still another way of relaxing the functional form of the duration dynamics is e.g., the semi-parametric autoregressive conditional proportional hazard (SACPH) model in Gerhard & Hautcsh (2002).

Lancaster (1990) gives a review of methods for duration/transition-data analysis. The standard survival model generalizes to models with competing risks or multiple hazards, i.e. the situation where exit from a state can occur due to various reasons. In the competing risk model there are many waiting-time variables, $T_1, \ldots, T_k$, but there is only opportunity to observe one:

$$T = \min(T_1, \ldots, T_n).$$

The case where the competing durations are independent is relatively easy, because then the total hazard is just the sum of the competing hazard. A financial study of lender/borrower waiting times is given by Lambrecth, Perraudin & Satchell (2003). Some effort in exploring the possibility of dependent competing risks is given in Lindeboom & Van den Berg (1994). A version of the competing risk model with possible exit to many destinations, is given a set of hazard functions, λ_{ij} describing the hazard for transition from state i to state j, (Lancaster 1990). A typical multi-state phenomenon in finance is a rating system, AAA, Aaa, etc. Examples of analysis of transition rates between ratings are in Lando & Skødeberg (2002) and Bladt & Sørensen (2006). Observed multivariate durations in multiple states, e.g. bivariate a distribution (T_1, T_2), refer to the process of waiting for many events. An example application to financial data is by An, Christensen & Gupta (2003) where the pension and retirement of spouses are analyzed. An application to financial market data is given by Quoreshi (2006) using the analogy of count-processes and durations.

A typical characteristic of duration/transition data is censoring. The type of censoring, functional form of the impact of regressors, the choice of probability model, all these issues affect the choice of estimation strategy.

8 Extreme value analysis

The term risk has been frequently used in financial context in recent years. The term "risk" can reflect many different apsects. Sometimes it seems that risk refers to lack of certainty, sometimes volatility in a price process, sometimes it seems to be risk for a particular event e.g., bankruptcy. The formal notion of the statistic of interest is:

$$X_{max,n} = \max_{0 \leq i \leq n} X(i) \quad \text{or} \quad X_{min,n} = \min_{0 \leq i \leq n} X(i).$$

For the discrete-time case of iid data the finite-sample maximum is simply:

$$F_{max}(x_{max}) = P(X_i \leq x_{max} | \text{ for } i \leq n) = F(x_{max})^n.$$

The classical central-limit-theorem can be interpreted as a large sample result for a sum of iid random variables. If a sequence of iid random variables, $X_1, \ldots, X_n$

with $E(X_i^2) < \infty$, the sum converges in distribution to a normal distribution, i.e.:

$$S_n = \sum_{i=1}^{n} X_i \xrightarrow{d} N(a_n, b_n), \quad \text{with } a_n = nE(X_i), \quad b_n = nV(X_i).$$

In many statistical textbooks, in chapters on order statistics there are similar results for the maximum (and minimum) of iid random variables. The question is whether we have limit result for the function maximum (or minimum), just as we have a limit result for function sum. If there exists sequences of constants, a_n, b_n, such that $(X_{max,n} - a_n)/b_n$ converges in distribution to a non-trivial limit, then the form of limiting distribution has to one of the following, (Embrechts, Klüppelberg & Mikosch 1997; Mood, Graybill & Boes 1974):

$$F_1(x) = I_{(0,\infty)}(x)\mathrm{e}^{-x^{-\gamma}}, \quad \gamma > 0, \tag{32}$$

$$F_2(x) = I_{(-\infty,0)}(x)\mathrm{e}^{-|x|^{\gamma}} + I_{[0,\infty)}(x), \tag{33}$$

$$F_3(x) = \mathrm{e}^{-\mathrm{e}^{-x}}. \tag{34}$$

The limiting distribution of (32) is obtained if (and only if)

$$\frac{1 - F(x)}{1 - F(\tau x)} \xrightarrow[x\to\infty]{} \tau^{\gamma}.$$

The limit distribution of (33) is obtained if and only if

$$F(x_0) = 1 \text{ for some } x_0 \text{ and } F(x_0 - \varepsilon) < 1, \text{ for all } \varepsilon > 0.$$

The limit distribution is of (34) if and only if

$$n(1 - F(b_n x + a_n)) \xrightarrow[n\to\infty]{} \mathrm{e}^{-x}.$$

The limit distribution is called extreme-value distribution and its properties are decided by the behavior of the distribution function F in the tails. The result in equations (32,33,34) is sometime referred to as Fisher-Tippett theorem for limit laws of maxima and the distributions are called Frechet, Weibull and Gumbell, respectively.

If the X_i's form a dependent sequence results similar to CLT are available, i.e. the dependency has to fade away with increasing time lag, A formal way of expressing such a fadeout is the concept m-dependence. Some examples are in Mcneil (1997), Resnick (1997) and Johansson (2003).

Multivariate generalizations are complicated and perhaps only practical in special situations. Starica (1999) treats the multivariate case of constant conditional correlations. It is in the nature of extremes that multivariate extremes are hard to deal with. Typical, practical statistical problems in extreme-value theory are quantile estimation and tail-index estimation. Extreme values are rare so getting an accurate estimate of a high quantile is not possible. A direct citation from Embrechts, Klüppelberg & Mikosch (1997) is: "There is no free lunch when it comes to high quantile estimation"; they also give sample properties of some estimators. Mikosch (2004) gives some guidelines. For more advanced treatment of extreme-value theory see, e.g., Embrechts, Klüppelberg & Mikosch (1997) and Resnick (1987). In the statistical package R (R Development Core Team 2005) it is possible to do some univariate and bivariate extreme-value calculations using the package

`evd`.

In recent years the approach of using copulas has become a popular tool for representing multivariate distributions. The idea is to get a formal tool of defining dependence between two variables. A copula is a multivariate density function with uniform margins. Any random variable can in theory be transformed to a univariate random variable by applying the univariate distribution function to it. If a bivariate normal random variable. (X_1, X_2) was transformed to have support on $[0,1] \times [0,1]$ by applying the univariate inverse distribution to X_1 and X_2, respectively, the result consists of two uniform $U(0,1)$ random-variable with a particular dependency structure. Conversely, if we have two dependent $U(0,1)$ random variables, U and V, with cumulative-distribution-function (cdf) $F(u,v)$, and apply the inverse of a univariate normal cdf to U and V, respectively, then we have a bivariate random variable with normal margins, but in general this bivariate random variable is not bivariate normal. The motive is that in finance, variables may seem weekly correlated most of the time, but when something serious happens, the catastrophy happens to both variables. A textbook on copula methods in finance is Cherubini, Luciano & Vecchiato (2004). Statistically speaking, when it comes to estimating copulas

from data, it is a question of estimating a multivariate distribution. Estimation takes either place through parsimonious parameterization or through (kernel) smoothing where accuracy is decided by the choice of bandwidth. In the case of multivariate extremes there is *no free lunch*, data will be thin and estimates will be inaccurate. A sober view of the importance of copulas in statistical analysis of multivariate extremes is given by Mikosch (2005)

9 Jump processes, and further topics

The popularity of continuous-time finance has generated an understanding of the nature of diffusion processes among data-analysts. They have realized that some of the movements that are seen in the financial markets are not likely to be outcome of a continuous-state process. Therefore the diffusion models have been modified to allow jumps. This calls for a definition of the nature of jumps. In the words of Merton (1990) this is *continuous-path with rare events*. The idea is to add a weighted Poisson-type process, $N(t)$ to the diffusion.

$$\mathrm{d}X(t) = \mu(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}W(t) + J(t)\,\mathrm{d}N(t).$$

The function $J(t)$ denotes the size of the jump. To be operational, it is necessary, in addition to the Wiener process (normal-distribution), to define the probability distribution of $N(t)$, the event of jump and that of the jump size, $J(t)$. For maximum-likelihood estimation there will be identification problems that will require restrictions on the parameter space, (Honoré 1998). Evidently the urge for tractability will influence the model choice. Kou (1999) gives a model that should be able to capture stylized facts, being analytically well-behaved at the same time: "A Jump Diffusion Model for Option Pricing with Three Properties: Leptokurtic Feature, Volatility Smile, and Analytical Tractability". The model is defined in equations (2.1) and (2.2) in his paper:

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) + \mathrm{d}\left(\sum_{i=1}^{N(t)}(V_i - 1)\right), \tag{35}$$

$N(t)$ Poisson $E(N(t)) = \lambda t$,

V_i iid double exponential.

The solution of (35). conditioned on $S(0)$:

$$S(t) = S(0)\mathrm{e}^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W(t)} \prod_{i=1}^{N(t)} V_i,$$

contains three different probability distributions, the normal, the Poisson and the double exponential. The memoryless property of the Wiener process, the Poisson process, and the double-exponential are helpful in deriving closed form solution of some option pricing problems. The spirit of this model is similar as in Merton (1976) and Merton (1990). There are many more examples of applications of jump diffusions in finance, (Wong & Li 2006).

Another modern approach of incorporating discontinuity in continuous-time process is the use of Lévy processes. Sato (2001) defines, $L(t)$, the Lévy process as:

$L(t_2) - L(t_1)$ and $L(t_4) - L(t_3)$ are independent if $t_1 < t_2 < t_3 < t_4$,

$L(0) = 0$,

The distribution of $L_{t+s} - L_t$ is not af function of t,

$P(|L_{s+t} - L_t| > \varepsilon) \xrightarrow[s\to 0]{} 0, \quad \forall \varepsilon > 0$,

$L(t)$ is right continuous with left limits.

The class of Lévy processes is very large. The Poisson process and Wiener process are both Lévy processes. In the Poisson process all movements occur in jumps of size 1, in the Wiener process all movement is along a continuous pattern. Eberlein (2001) gives a simple formula for representing a Lévy process which is generated with a distribution that has finite first moments:

$$X(t) = \sigma W(t) + Z(t) + \alpha t,$$

where $W(t)$ is standard Wiener process and $Z(t)$ is a purely discontinuous martingale independent of $W(t)$. The formal notion of a Lévy process allows for a formal definition of a stochastic differential equation that is driven by a Lévy process instead of the more traditional Wiener process, e.g. an Ornstein-Uhlenbeck type process:

$$\mathrm{d}X(t) = \alpha(\beta - X(t))\,\mathrm{d}t + \sigma\,\mathrm{d}L(t). \tag{36}$$

Equation (36) represents a dynamic system that in finance is called the Vasicek model. It is conceivable to have $L(t)$ as purely discontinuous, such that $X(t)$ will remain positive. That way it can be a plausible model for a dynamic variance (or volatility). Barndorff-Nielsen & Shephard (2001) give a thorough treatment of ideas in that spirit. There is a substantial interest in creating a positive process for modelling the volatility process, Tsai & Chan (2005) give a comment on how to ensure non-negativity. Marquardt & Stelzer (2005) discuss a multivariate CARMA driven with Lévy Process. Many of the other models can be nested within the formal framework of working with Lévy processes. Brockwell & Marquardt (2005) treat the fractionally integrated CARMA. Brockwell, Chadraa & Lindner (2006) derive a continuous-time GARCH process.

10 A comment on model-building

A key issue in statistical analysis is the "model". Data is interpreted through a model. The model is a kind of mathematical idealization of some real-world phenomena, and the general conclusion to be drawn from data refers to that particular model. There are several principles in designing an interesting model. Draper & Smith (1966) give three principal categories, a) the functional model, which explains the nature of the underlying process, b) the control model, i.e., a model where control does not rely on full understanding of the underlying process and c) the predictive model, which aims at giving a method for statements about future observations. A famous phrase is "all models are wrong but some are useful" (Box 1979), describes the situation facing the applied data analyst. Practical choice of statistical model is mainly based on a combination of the following, a) mathematical tractability b) some theoretical basis and c) some functions that are likely to fit stylized features of data. A model should be plausible, e.g. we do not like to get negative values from something that should be positive. It should also be possible to interpret the model. It is preferable that the model can explain something, and last but not least it should be possible to reject a model if data look very incompatible to possible output of the model.

A frequent approach in practical analysis is to employ some kind of pre-testing. The estimation-procedure is then essentially a two-step procedure, consisting of a test of a particular hypothesis and if the hypothesis is rejected, then a particular estimation procedure is performed. For example:

$$\boldsymbol{\theta}_{Pre-test} = I_{H_0}(\boldsymbol{X})\boldsymbol{\theta}_0 + I_{H_1}(\boldsymbol{X})\hat{\boldsymbol{\theta}}_{ML}. \tag{37}$$

In equation (37), $I_{H_0}(\boldsymbol{X})$ and $I_{H_1}(\boldsymbol{X})$ are indicator-functions of data taking the value 1 or 0, depending on which hypothesis is supported. An early literature review is given by Judge & Bock (1978). A recent Ph.D. thesis on the subject is Danilov (2003). Extensive pre-testing schemes have been developed. One scheme is based on starting with a big model and try to test away model components of minor importance. Sometimes this is labeled stepwise-backward or general-to-specific. The system PCGETS described in Hendry & Krolzig (2001) is an example of this. Another approach is sometimes called stepwise-forward or specific-to-general is based on starting with a simple model and try to include 'significant' variables. The system RETINA Perez-Amaral, Gallo & White (2003) is an example of this approach.

The pre-test procedure is biased, where there is a bias towards H_0. It is to be expected that this bias is beneficial, relative to say least-squares estimation, when the truth is close to H_0. It is also to be expected that biased estimation is harmful if the truth is far from H_0. It turns out that for some loss-functions it is possible to dominate the ordinary ordinary-least-squares estimators in linear models. Examples of such estimators are the Stein-rule estimators. An early review is given by Judge & Bock (1978). The idea is to shrink the ML estimates towards an a priori defined subspace of the parameter space. The pre-test strategy is a kind of jump-shrinking strategy, whereas the Stein-family estimators and ridge-regression type estimators are examples of continuous shrinkage. Tómasson (1986) applies these ideas to ARMA models. It is clear that how to shrink depends on the characteristics of the application. Stein-estimators can be derived as a kind of empirical Bayes procedure. An empirical Bayes procedure is based on a Bayes-estimator where parts of the prior are estimated from data. A later review is given by Saleh (2006). Many practitioners now do this

by minimizing the AIC or the BIC.

$$AIC = -Tl(\hat{\boldsymbol{\theta}}) + k,$$
$$BIC = -Tl(\hat{\boldsymbol{\theta}}) + \frac{k}{2}\log(T).$$

Where T is the number of observations, $l(\hat{\boldsymbol{\theta}})$ is the log-likelihood evaluated at the ML-estimate and k is the number of estimated parameters. For linear models the AIC behaves similar to a pre-test estimator with a fixed rejection level, whereas the BIC behaves similar to pre-test estimator where the rejection level depends on the sample size.

11 Summary and discussion

The aim of financial data analysis is to make inference based on data. Typically data consist of a single realization of a time-series. Stationarity and ergodicity are necessary for obtaining consistent estimates. The linear-normal discrete-time model is the most widely used and best understood model, Univariate ARMA, or multivariate, VARMA, regression based-variants, ARMAX, where it is allowed to condition on some explanatory variables all belong to that class. Common to those is that, prediction, missing data, systematic sampling, calculation of likelihood, etc., most treatments of interest are possible by using recursive algorithms like the Kalman-filter. The ARMA is a linear filter. An important feature of the linear filter is, that if the input is normal, then the output is also normal. If the input to a MA process is a sequence of iid stable distributions then the output is also stable. The normal distribution is the only stable distribution with finite variance. In general the closed form of the density of a stable distribution is unknown, but in some cases the characteristic function can be written down. There exist computer programs for simulating a sample from stable distributions. Numerical methods of the inverting the Fourier transform can be used to obtain the likelihood function, that then can be numerically maximized. Even quantile-regression and simulation based methods could be used. In many dimensions the situation is more complicated. There exist methods for multivariate stable distributions, but the computational aspect is

difficult. When the input to a linear filter is a finite variance non-normal noise, then the output is not normal, but due to central-limit-theorem type arguments, the output is "more normal" than the input. In many applications, it is natural to assume that the level of a series is much more normal, than the series of innovation. Assuming some parametric form of the distribution of innovations makes it possible to calculate the likelihood function recursively, estimate parameters, perform tests, etc. The usual ARMA model has an exponential decaying autocorrelation function. The ARFIMA is a way of allowing slower decay of the autocorrelation without abandoning the stationarity assumption.

Doing constructive non-linear modelling requires a firm idea of the form of nonlinearity, e.g., a process jumping in first or second moments. The ARCH-family aims at modelling second moments of a process by applying the auto-regressive concept to the second moments of a measured process. The distributional properties are somewhat complicated. The general multivariate ARCH is extremely complicated. Many ideas of ordinary time-series, threshold models, long memory etc. have found their way to the ARCH literature. As an example, in ARCH-M and AR-ARCH models both the dynamics of first and second moments are modelled.

The continuous-time models based on the Wiener-process have a mathematical appeal, and many models can be motivated by theory, e.g., pricing methods based on no arbitrage. The univariate diffusions may also have a stationary invariant distribution that can be interpreted in a real world context. A feature of the univariate diffusion model is that it can combine short-term dynamics and long-term equilibrium into one formula. High dimensional diffusion process have in common with the multivariate non-linear discrete-time models that they are hard to visualize and in practical cases it is necessary to have a good idea about the nature of the functional relationship between variables. For a high-dimensional diffusion system the issue of an invariant stationary distribution is no longer simple, both on its existence and on how to find it in case of existence.

In the continuous-time discrete-state-space, any positive random-variable can play the role of a waiting time (duration) for a shift between states. The in-

terpretable form of the distribution of the waiting-time is the hazard function. Monitoring the hazard-function over time mounts to monitoring the risk for change of state. The hazard rate itself can be thought of as a positive stochastic process, with a dependency structure over time and possibly a function of external regressors.

The extreme-value theory is evolving, but even if multivariate models become computationally tractable, their usability will be limited by lack of data.

Some data-analysts have found that the elegant mathematical idealization of the diffusion process does not give a realistic picture of what is happening in financial markets. Stylized facts as the volatility smile, etc. have led researches to work with modifications such as jump diffusions and Lévy-driven stochastic differential equations.

When time-series are analysed it is of importance to understand the source of the data, e.g. whether we have stock-data or flow-data. A lot of the models in the literature focus on the return data in a financial market. Trading data consist of time, price and volume.

In the author's mind, the following are most important building blocks for surveillance of time-dependent data. First the linear-filter with normal input. Then an understanding of the importance of the distribution. The diffusion models are an elegant approach of linking theory and data. The theory of diffusion models can give good motivation for a choice of a linear model. The univariate extreme-value theory is reasonably simple and will give good approximation. Multivariate extreme-value model will require judgment and parsimonious parameterization. The continuous-time models driven with Lévy process have not yet reached the practitioners, so for a while they will be doing ad-hoc jump diffusion models. The ARCH models have been around for some time. Academics and practioners have built up experience on their benefits and drawbacks. Their future is dependent on the emerge of realistic alternatives. Starica (2004) has raised some questions about their ability to cope with short term properties and long term properties of financial time series. His criticism is roughly that even though the ARCH-type model can capture some short term volatility dynamics,

it comes at the price of overestimating long term variance. The practitioner has always to choose a model that serves a certain aim. The choice of model is a compromise of its simplicity and its ability to capture important features of real life.

References

Aït-Sahalia, Y. (1996). Non-parametric pricing of interest rate derivative securities. *Econometrica*, *64*, 527–560.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusion. *Journal of Finance*, *54*(4), 1361–1395.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, *70*(1).

An, M. Y., Christensen, B. J., & Gupta, N. D. (2003). On pensions and retirement: Bivariate mixes proportional hazard modelling of joint retirement. Working paper no 163. Center of analytical finance, University of Aarhus, Aarhus Business School.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process. *Biometrika*, *66*, 59–65.

Bachelier, L. (1900). Theorie de la speculation. *Annales de l'Ecole Normale Superiore*, (17), 21–86.

Baille, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedacity. *Journal of Econometrics*, *74*, 3–30.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 167–241.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *Journal of the Statistical Society Supplement*, *8*, 27–41.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall.

Beran, J. (1995). Maximum-likelihood estimation of the differencing parameter for invertible short and long ARIMA models. *Journal of the Royal Statistical Society Series B*, *57*(4), 659–672.

Bergstrom, A. R. (1988). The history of continuous-time econometric models. *Econometric theory*, *4*, 365–383.

Bergstrom, A. R. (1990). *Continuous-Time Econometric Modelling*. Oxford University Press.

Beskos., A., Papaspiliopoulos, O., Roberts, G. O., & Fearnhead, P. (2006). Exact and efficient likelihood-based inference for discretely observed diffusion processes (with discussion. *Journal of the Royal Statistical Society, series B*, *68*, 333–382.

Beskos, A. & Roberts, G. O. (2005). Exact simulation of diffusions. *Annals of Applied Probability*, *15*, 2422–2444.

Bhardwaj, G. & Swanson, N. R. (2006). An empirical investigation of the usefulness of arfima models for predicting macroecnomic and financial time series. *Journal of Econometrics*, *131*, 539–578.

Bibby, B. M., Jacobsen, M., & Sørensen, M. (2004). Estimating functions for discretely sampled diffusion-type models. Preprint 2004-4, Department of Applied Mathematics and Statistics, University of Copenhagen.

Björk, T. (2004). *Arbitrage Theory in Continuous Time*. Oxford University Press.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, (81), 635–654.

Blackwell, D., Griffiths, M., & Winters, D. (2006). *Modern Financial Marktes*. John Wiley & Sons.

Bladt, M. & Sørensen, M. (2006). Efficient estimation of transition rates between credit ratings from observations at discrete time points. Preprint No. 2006-2, Department of Applied Mathematics and Statistics, University of Copenhagen.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, *72*, 498–505.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & E. G N. Wilkinson (Eds.), *Robustness in Statistics*. Akademic Press: New York.

Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Fransisco, Holden Day.

Brockwell, P., Chadraa, E., & Lindner, A. (2006). Continuous-time GARCH processes. *The Annals of Probability*, *16*(2), 790–826.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag.

Brockwell, P. J. & Marquardt, T. (2005). Levy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statistica Sinica*, *15*, 477–494.

Brown, R. (1827). *A brief account of microscopical observations*. Unpublished, London.

Chan, W.-S., Keung, W., & Tong, H. (Eds.). (2000). *Statistics and Finance: An Interface*. Imperial College Press.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*. John Wiley & sons.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, *1*(3), 309–324.

Cvitanic, J. & Zapatero, F. (2004). *Introduction to the Economics and Mathematics of Financial Markets*. MIT press.

Danilov, D. (2003). *The Effects of Pretesting in Econometrics with Applications in Finance*. PhD thesis, Tilburg University.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association*, *74*, 427–431.

Draper, N. & Smith, H. (1966). *Applied Regression Analysis*. John Wiley & Sons.

Duffie, D. (1996). *Dynamic Asset Pricing Theory*. Princeton University Press.

Eberlein, E. (2001). Application of generalized hyperbolic levy motions to finance. In O. E. Barndorff-Nielsen, T. Mikosch, & S. I. Resnick (Eds.), *Lévy Processes, Theory and Applications*. Birkhäuser, Boston.

Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid by the molecular-kinetic theory of heat. *Annalen der Physik*, (17), 549–560.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Heidelberg.

Engle, R. F. (1982). Autoregressive conditional heteroskedacity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Engle, R. F. & Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, *5*, 1–50.

Engle, R. F. & Russell, J. R. (1998). Autoregressive conditional duration a new model for irregularly-spaced transaction data. *Econometrica*, *66*(5), 1127–1162.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, *20*(4).

Fan, J. & Yao, Q. (2003). *Nonlinear Time Series*. Springer-Verlag, New York.

Fernandes, M. & Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, *130*(1), 1–23.

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons.

Gailbraith, R. F. & Gailbraith, J. I. (1974). On the inverse of some patterned matrices arising in the theory of stationary time series. *Journal of applied probability*, *11*, 63–71.

Gerhard, F. & Hautcsh, N. (2002). Semiparametric autoregressive conditional proportional hazard models. No 2002-W2, Economics Papers from Economics Group, Nuffield College, University of Oxford.

Gourieroux, C. (1997). *ARCH models and Financial Applications*. Springer-Verlag, New York.

Granger, C. (1983). *Forecasting white noise, in, Applied Time Series Analysis of Economic Data, Proceedings of the Conference on Applied Time Series Analysis of Economic Data (October 1981), Editor. A. Zellner*. U.S. Government Printing Office.

Granger, C. W. J. & Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenchoeck & Ruprect, Göttingen.

Granger, C. W. J. & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series*, *1*, 15–29.

Haavelmo, T. (1943). The implications of a system of simultaneous equations. *Econometrica*, *11*, 1–12.

Hall, P., Peng, L., & Yao, Q. (2002). Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series*, *23*(3), 313–331.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hand, D. J. & Jacka, S. D. (Eds.). (1998). *Statistics in Finance*. Arnold.

Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley & Sons.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. (1993). *Time Series Models*. Harverster Wheatsheaf.

Harvey, C. R. & Morgenson, G. (2002). *The New York Times Dictionary of Money and Investing: The Essential A-to-Z Guide to the Language of the New Market*. Times Books.

Hendry, D. F. & Krolzig, H.-M. (2001). *Automatic Econometric Model Selection Using PcGets*. Timberlake Consultants Press.

Honoré, P. (1998). Pitfalls in estimating jump-diffusion models. Working paper, Unversity of Aarhus.

Hull, J. (1993). *Options, Futures and other Derivatives* (second ed.). Prentice-Hall.

Johansson, N. C. J. (2003). *Moment Estimation Using Extreme Value Techniques*. PhD thesis, Chalmers University of Technology and Göteborg University.

Jones, D. A. (1978). Non-linear autoregressive processes, series a. *Proc. Roy. Soc. London*, *360*, 71–95.

Judge, G. G. & Bock, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-rule Estimators in Econometrics*. North-Holland.

Karatzas, I. & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus, second edition*. Springer-Verlag.

Kendall, M. G. (1953). The analysis of economic time-series part I: Prices,. *Journal of the Royal Statistical Society. Series A (General)*, (1), 11–25.

Klimko, L. A. & Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics*, *6*, 629–642.

Kloeden, P. E. & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin.

Kou, S. (1999). A jump diffusion model for option pricing with three properties: Leptokurtic feature, volatility smile, and analytical tractability.

Kutoyants, Y. A. (1984). *Parameter Estimation for Stochasic Processes*. Helderman, Berlin.

Lambert, P. & Lindsey, J. K. (1999). Analysing financial returns using regression models based on non-symmetric stable distributions. *Applied Statistics*, *48*, 409–424.

Lambrecth, B. M., Perraudin, W. R. M., & Satchell, S. (2003). Mortgage default and possession under recourse: A competing hazard approach. *Journal of Money, Credit and Banking*, *35*(3), 425–442.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press.

Lando, D. & Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance,*, *26*(2-3), 423–444.

Lindeboom, M. & Van den Berg, G. J. (1994). Heterogeneity in bivariate duration models: The importance of mixing distribution. *Journal of the Royal Statistical Society; B*, *56*, 49–60.

Lund, R. & Basawa, I. (1999). Recursive prediction and likelihood evaluation for periodic ARMA models.

Lütkepol, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Maddala, G. S. & Rao, C. R. (Eds.). (1996). *Handbook of Statistics, Volume 14: Statistical Methods in Finance*. Elsevier.

Mao, X., Yuan, C., & Yin, G. (2005). Numerical method for stationary distribution of stochastic differential equations with Markovian switching. *J. Comput. Appl. Math.*, *174*(1), 1–27.

Marquardt, T. & Stelzer, R. (2005). Multivariate CARMA processes.

McLeod, A. (1993). Parsimony, model adequacy and periodic correlation in forecasting time series. *International Statistical Review*, *61*.

Mcneil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin*, *27*(1), 117–137.

Melard, G. (1983). A fast algorithm for the exact likelihood of moving average models. *Applied Statistics*, *33*, 104–114.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Merton, R. C. (1990). *Continuous-Time Finance*. Blackwell.

Mikosch, T. (2004). How to model multivariate extremes if one must? Maphysto, Research report no. 21.

Mikosch, T. (2005). Copulas tales and facts. Discussion paper, International Conference on Extreme Value Analysis in Gothenburg.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (Third ed.). McGraw-Hill.

Neftci, S. N. (1996). *An Introduction to the Mathematics of Financial Derivatives*. Academic Press, San Diego.

Nelson, D. (1991). Conditional heteroskedacity in asset returns. *Econometrica*, *59*, 347–370.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric theory*, *6*, 318–344.

Øksendal, B. (1998). *Stochastic Differential Equations: An Introduction with Applications* (5 ed.). Springer-Verlag.

Ozaki, T. (1982). The statistical analysis of perturbed limit cycle processes using nonlinear time series models. *Journal of Time Series*, *3*, 29–41.

Ozaki, T. (1985). Nonlinear time series models and dynamical systems. In E. Hannan & P. Krishnaiah (Eds.), *Handbook of Statistics*, volume 5. North-Holland, Amsterdam.

Pagano, M. (1978). On the periodic and multiple autoregressions. *Annals of Statistics*, *6*, 1310–1317.

Perez-Amaral, T., Gallo, G., & White, H. (2003). A flexible tool for model building: The relevant transformation of the inputs network approach (retina). *Oxford Bulleting of Economics and Statistics*, (65), 821–838.

Priestley, M. B. (1991). *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press.

Quoreshi, A. M. M. S. (2006). Bivariate time series modeling of financial count data. *Communications in Statistics: Theory and Methods*, *35*(7), 1343–1358.

R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Resnick, S. I. (1987). *Extreme Values Regular Variations, and Point Processes*. Springer-Verlag.

Resnick, S. I. (1997). Discussion on the data on large fire insurance losses. *Astin Bulletin*, *27*(1), 139–151.

Revuz, D. & Yor, M. (1999). *Continuous Martingales and Brownian Motion* (3 ed.). Springer-Verlag Berlin Heidelberg.

Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*. John Wiley & Sons.

Sato, K. (2001). Basic results on Lévy processes. In O. E. Barndorff-Nielsen, T. Mikosch, & S. I. Resnick (Eds.), *Lévy processes: Theory and Application*. Birkhäuser, Boston.

Shiryaev, A. N. (1999). *Essentials of Stochastic Finance*. World Scientific Publishing.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, *102*(2), 111–141.

Sowell, F. (1992). Maximum-likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, *53*, 165–188.

Starica, C. (1999). Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, *6*, 515–553.

Starica, C. (2004). Is GARCH(1,1) as good a model as the nobel prize accolades would imply? Econometrics 0411015, EconWPA. available at http://ideas.repec.org/p/wpa/wuwpem/0411015.html.

Tómasson, H. (1986). *Prediction and Estimation in ARMA Models*. PhD thesis, University of Gothenburg.

Tong, H. (1983). *Threshold Models in Non-Linear Time-Series Analysis*. Springer-Verlag, New York.

Tsai, H. & Chan, K. S. (2005). A note on non-negative continuous time processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(4), 589–597.

Tsay, R. S. (2002). *Analysis of Financial Time Series*. John Wiley & Sons.

Wei, W. W. S. (1990). *Time Series Analysis - Univariate and Multivariate Methods*. Addison-Wesley, Inc.

Wiener, N. (1923). Differential space. *Journal of Mathematical Physics*, *2*, 131–174.

Wilmott, P., Howison, S., & Dewynne, J. (1995). *The Mathematics of Financial Derivatives*. Cambridge University Press.

Wong, H. Y. & Li, C. P. (2006). Estimating jump diffusion structural credit risk models. Working Paper, Chinese Univeristy, Hong Kong.

Yang, W. & Allen, D. E. (2005). Multivariate GARCH hedge ratios and hedging effectiveness in Australian futures markets. *Accounting and Finance*, *45*(2), 301–321.

Zivot, E. & Wang, J. (2003). *Modeling Financial Time Series with S-plus*. Springer-Verlag.

INSTITUTE OF ECONOMIC STUDIES WORKING PAPERS 1987-2006
Formerly Iceland Economic Papers Series

Editor Sveinn Agnarsson

A complete list of IoES Working Papers and Reprints can be accessed on our World Wide Web site at http://www.ioes.hi.is

W07:06 Helgi Tomasson: Review of Econometric Modeling Approaches in Finance.

W07:05 Anna Gunnthorsdottir, Roumen Vragov, Kevin McCabe and Stefan Seifert: The meritocracy as a mechanism to overcome social dilemmas.

W07:04 Gylfi Zoega: Endogenous employment cycles in Euroland.

W07:03 Thorvaldur Gylfason: The international economics of natural resources and growth.

W07:02 Helga Kristjansdottir: Talking trade or talking aid? Does investment substitute for aid in the developing countries?

W07:01 Ali Choudhary, Thorlakur Karlsson and Gylfi Zoega: Testing for customer markets using survey data.

W06:13 Brynhildur Davidsdottir: Sustainable energy development.

W06:12 Helga Kristjansdottir: Substitution between inward and outward foreign direct investment.

W06:11 Helga Kristjansdottir: Evaluation of Icelandic trade flows, the gravity model approach.

W06:10 Brynhildur Davidsdottir: Capital constraints and the effectiveness of environmental policy.

W06:09 Gylfi Zoega: Market forces and the continent's growth problem.

W06:08 Fridrik M Baldursson and Nils-Henrik M von der Fehr: Vertical integration and long-term contracts in risky markets.

W06:07 Ragnar Arnason: Conflicting uses of marine resources: Can ITQ's promote an efficient solution?

W06:06 Thorvaldur Gylfason and Gylfi Zoega: A golden rule of depreciation.

W06:05 Ron Smith and Gylfi Zoega: Global factor, capital adjustment and the natural rate.

W06:04 Thorolfur Matthiasson: To whom should the rent accrue?

W06:03 Tryggvi Thor Herbertsson and Gylfi Zoega: Iceland's Currency Dilemma.

W06:02 Thorolfur Matthiasson: Possible stakeholder conflicts in quota regulated fisheries,contribution to the political economics of fisheries.

W06:01: Eyjolfur Sigurdsson, Kristin Siggeirsdottir, Halldor Jonsson jr , Vilmundur Gudnason, Thorolfur Matthiasson, Brynjolfur Y Jonsson:Early discharge and home intervention reduces unit costs after total hip replacement: Results of a cost analysis in a randomized study.

W05:14 Gylfi Zoega and J Michael Orszag: Are Risky Workers More Valuable to Firms?

W05:13 Friðrik Már Baldursson: Fairness and pressure group competition.

W05:12 Marias H. Gestsson and Tryggvi Thor Herbertsson: Fiscal Policy as a Stabilizing Tool.

W05:11 Tryggvi Thor Herbertsson and Gylfi Zoega: On the Adverse Effects of Developement Aid.

W05:10 Thráinn Eggertsson and Tryggvi Thor Herbertsson: Evolution of Financial Institutions: Iceland's Path from Repression to Eruption.

W05:09 Tryggvi Thor Herbertsson: The Icelandic Pension System in 2004.

W05:08 Ron Smith and Gylfi Zoega: Unemployment, investment and global expected returns: A panel FAVAR approach.

W05:07 Gylfi Zoega and Thorlakur Karlsson: Does Wage Compression Explain Rigid Money Wages?

W05:06 Thorvaldur Gylfason: India and China

W05:05 Edmund S. Phelps: Can Capitalism Survive?

W05:04 Thorvaldur Gylfason: Institutions, Human Capital, and Diversification of Rentier Economies

W05:03 Jón Daníelsson and Ásgeir Jónsson: Countercyclical Capital and Currency Dependence

W05:02 Alison L. Booth and Gylfi Zoega: Worker Heterogeneity, Intra-firm Externalities and Wage Compression

W05:01 Tryggvi Thor Herbertsson and Martin Paldam: Does development aid help poor countries catch up?

W04:12 Tryggvi Thor Herbertsson: Personal Pensions and Markets

W04:11 Fridrik M. Baldursson and Sigurdur Johannesson: Countervailing Power in the Icelandic Cement Industry

W04:10 Fridrik M. Baldursson: Property by ultimatum: The case of the Reykjavik Savings Bank

W04:09 Ingólfur Arnarson: Analyzing Behavior of Agents of Economic Processes in Time

W04:08 Otto Biering Ottosson and Thorolfur Matthiasson: Subsidizing the Icelandic Fisheries

W04:07 Niels Vestergaard and Ragnar Arnason: On the Relationship between Greenland's Gross Domestic Product and her Fish Exports: Empirical Estimates

W04:06 Ingolfur Arnarson: Modelling Fishery Management Schemes with an Olympic System Application

W04:05 Ingolfur Arnarson and Pall Jensson: Adding the Sales Markets Dimension to Bio-Economic Models. The Case of Fishery Management

W04:04 Edmund S. Phelps: Changing Prospects, Speculative Swings: Structuralist Links through Real Asset Prices and Exchange Rates

W04:03 Ingolfur Arnarson: Analysing Behavior of Agents of Economic Processes in Time

W04:02 Ron Smith and Gylfi Zoega: Global Shocks and Unemployment Adjustment

W04:01 Fridrik M. Baldursson and Nils-Henrik M von der Fehr: Prices vs. quantities: public finance and the choice of regulatory instruments

W03:07 Sveinn Agnarsson and Ragnar Arnason: The Role of the Fishing Industry in the Icelandic Economy. A historical Examination

W03:06 Thorolfur Matthiasson: Paying paper by paper, the wage system of Icelandic University teachers explained

W03:05 Gur Ofur and Ilana Grau: Bringing the Government hospitals into line: The next step of reform in the healthcare sector

W03:04 Ingolfur Arnarson and Pall Jensson: The Impact of the Cost of the Time Resource on the Efficiency of Economic Processes

W03:03 Torben M. Andersen and Tryggvi Thor Herbertsson: Measuring Globalization

W03:02 Tryggvi Thor Herbertsson and J. Michael Orszag: The Early Retirement Burden: Assessing the Costs of the Continued Prevalence of Early Retirement in OECD Countries

W03:01 Eirik S. Amundsen, Fridrik M. Baldursson and Jørgen Birk Mortensen: Price Volatility and Banking in Green Certificate Markets

W02:10 Tryggvi Thor Herbertsson and Gylfi Zoega: A Microstate with Scale Economies: The Case of Iceland

W02:09 Alison, L. Booth and Gylfi Zoega: Is Wage Compression a Necessary Condition for Firm-Financed General Training

W02:08 Asgeir Jonsson: Exchange rate interventions in centralized labor markets

W02:07 Alison, L. Booth, Marco Francesconi and Gylfi Zoega: Oligopsony, Institutions and the Efficiency of General Training

W02:06 Alison L. Booth and Gylfi Zoega: If you’re so smart, why aren’t you rich? Wage inequality with heterogeneous workers