Mathematisches Forschungsinstitut Oberwolfach

Report No. 2/2004

Statistics in Finance

Organised by
Richard A. Davis (Ft. Collins)
Claudia Klüppelberg (München)

January 11th – January 17th, 2004

Introduction by the Organisers

The Statistics in Finance Workshop, organized by Richard A. Davis (Ft. Collins) and Claudia Klüppelberg (Technische Universität München), was held January 11-17. This meeting was well attended with over 40 participants with broad geographic representation from Europe, England, Australia, the Far East, and the US. This workshop was a nice blend of researchers with various backgrounds including statistics, probability, and econometrics. Approximately 33 talks, of varying lengths, were delivered during the five days. The talks were given by both leading experts in the field as well as by up and coming stars.

There were several major themes in the various sessions. These included, continuous time models, Levy processes, stochastic volatility models, GARCH models, extreme value theory with applications to financial risk, theory of copulas, and option pricing. This meeting generated a great deal of discussion and often smaller groups of people met in the evenings for expanded and detailed lectures. A number of important research contacts were made which we fully expect to stimulate many new collaborative research projects.

In addition to the excellent scientific program, there were two scheduled social activities. The inclement weather cleared up just in time for the traditional Wednesday afternoon hike to Oberwolfach for coffee and Black Forest Cake. The second activity, which most considered the highlight of the week, was a piano recital performed by Peter Brockwell and Gernot Müller.

For many of the participants, this was their first trip to Oberwolfach, and they came away very impressed from the experience. There was a strong consensus that the "Statistics in Finance Workshop" should become a regular Oberwolfach event.

Workshop on Statistics in Finance

Table of Contents

Abstracts

Continuous Time Stochastic Volatility Modelling and Bipower Variation

Ole. E. Barndorff-Nielsen and Neil Shephard

The theory of semi-martingales and stochastic integration constitutes a powerful and natural background for continuous time modelling of stochastic volatility, as observed in financial time series. However, for the models to make financial sense it is necessary to restrict somewhat from the completely general concept of semi-martingales ($\mathcal{SM}$). Recall that $Y \in \mathcal{SM}$ means that Y is of the form $Y = A + M$ where $A \in \mathcal{FV}$ and $M \in \mathcal{M}_{loc}$. We wish to think of Y as the log price process of a financial asset, with the process A expressing potential rewards and M the risk. For this the decomposition of Y into the sum of A and M should be unique. This is achieved by requiring A to be predictable (in which case Y is said to be a special semi-martingale).

If we further assume that M is continuous, $M \in \mathcal{SM}^c$ (below we comment on alternative possibilities), then this has important consequences. First, if the model Y is to be arbitrage free then A has also to be continuous, i.e. $A \in \mathcal{FV}^c$. Further, by the Dambis-Dubins-Schwarz Theorem, M is then representable as a time changed Brownian motion (BM). So $M = B_{[M]}$ and $B = M_T$ where $[M]$ is the quadratic variation of M and the time-change T is the inverse function of $[M]$. (For this it is necessary that $[M]_t \to \infty$ for $t \to \infty$.) Thirdly, supposing that $[M]$ is absolutely continuous, of the form $[M] = \int_0^t \tau_u \mathrm{d}u$, then, again by the no arbitrage requirement, A must also be absolutely continuous, $A = \int_0^t a_u \mathrm{d}u$. The process $\sigma = \sqrt{\tau}$ expresses the volatility, and it is to be noted that τ, which is termed the variance process, may have jumps. Finally, the absolute continuity together with the time-change representation implies that M can be written as $M_t = \int_0^t \sigma_u \mathrm{d}W_u$ for a BM W.

Thus the choice for Y has been narrowed to the continuous stochastic volatility semi-martingale framework ($\mathcal{SVSM}^c$) where

$$Y_t = \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u. \tag{1}$$

This type of process is sometimes called a Brownian semi-martingale; but, having the financial context in mind, we refer to it as a (continuous time) stochastic volatility process.

For more specific modelling, aimed at representing the important, and widely established, stylised features of financial observational series, choices have to be made of the two ingredients a and σ of (1). A simple example for a, of some definite interest, takes $a_t = \mu + \beta\tau_t$. More generally, one may consider $a_t = g(t, \tau_t)$ for some smooth function g. As to σ, a number of points have to be considered: (i) Should σ be a pure diffusion process (or perhaps a superposition of such processes); or should it be a pure diffusion plus a finite activity (FA) process (finite activity

meaning that there are only finitely many jumps in any finite time interval), or should it perhaps be an infinite activity (IA) process (for instance, an inverse Gaussian OU process or a superposition of such processes, or one of the CARMA processes introduced by Brockwell (2001), or one of the Lévy driven long-memory models considered by Anh, Heyde and Leonenko (2002). (ii) Should the model incorporate leverage, in the sense of dependence between σ and W. (iii) Should σ be Markovian. And more specifically, how should the law of τ be chosen so as to capture both the typical 'exbell' shape of log returns and the, generally observed, quasi long range dependence in the log price series. Note that if a is 0, or independent of σ and W, then under (1) the autocorrelations of the returns are necessarily 0, in accordance with the empirical facts.

To account for possible jumps in the price process one possibility is to add an independent FA process to Y. Another is to substitute the $\sigma \bullet W$ term by either $\sigma \bullet L$ or L_T where L denotes a Lévy process and T is a time-change. Note that, except for stable Lévy motions L, these two approaches are not equivalent; each has its advantages and drawbacks. A simpler approach is pure Lévy modelling which replaces Y by L. This already yields significant improvements over the classical Black-Scholes model, but misses, in particular, the key time-wise dependence feature of finance data. A further variant, that does model dependence, is the recently introduced continuous time GARCH model of Klüppelberg, Lindner and Maller (2004). A great amount of interesting work in this area has been carried out in the project led by Ernst Eberlein at Freiburg University. And in monograph form there are now two recent additions to the literature that set out important aspects of the Lévy based methods: Schoutens (2003) and Cont and Tankov (2003).

Our own joint research has fallen within the framework outlined above, and much of this will be described in considerable detail in our forthcoming book Barndorff-Nielsen and Shephard (2005).

The most recent part of this research concerns the new concept of *bipower variation*, that we have introduced and studied in Barndorff-Nielsen and Shephard (2003, 2004). This considers returns over time periods of lengths $\hbar$ and δ, where $n\delta = \hbar$ for some positive integer n and where for concreteness we may think of $\hbar$ as representing a day, with δ corresponding to 5, 10 or 30 minute consecutive intervals during that day. In the simplest 1,1 case, we define the realised bipower variation on the i-th day as the probability limit for $\delta \to 0$ of

$$\{Y_\delta\}_i^{[1,1]} = \sum_{j=2}^{n} |y_{j-1,i}|\,|y_{j,i}|\,;$$

here

$$y_{j,i} = Y_{(i-1)\hbar+j\delta} - Y_{(i-1)\hbar+(j-1)\delta}, \quad j = 1, 2, ..., n.$$

The limit of $\{Y_\delta\}_i^{[1,1]}$ is denoted by $\{Y\}_i^{[1,1]}$. We show that when we add to an SV process a finite activity jump process then, up to proportionality, the probability limit of this object (subject to some weak assumptions) is the quadratic variation of the SV process over the day as $\delta \downarrow 0$. Thus the realised bipower variation process is reasonably robust to jumps.

An asymptotic distribution theory for realised bipower variation can be calculated. Further, the joint distribution of realised bipower variation and the quadratic variation version of this, can be calculated under the assumption that there are no jumps. This allows us to consistently test the hypothesis that the sample path of price processes have jumps.

References

[1] Anh, V.V., Heyde, C.C. and Leonenko, N.N. (2002): Dynamic models of long-memory processes driven by Lévy noise. *J. Appl. Prob.* **39**, 730-747.
[2] Barndorff-Nielsen, O.E. and Shephard, N. (2003): Econometrics of testing for jumps in financial economics using bipower variation. Unpublished paper, Nuffield College, Oxford.
[3] Barndorff-Nielsen, O.E. and Shephard, N. (2004): Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* **2**, 1-48.
[4] Barndorff-Nielsen, O.E. and Shephard, N. (2005): *Continuous Time Approach to Financial Volatility*. Cambridge University Press (to appear).
[5] Brockwell, P.J. (2001): Lévy driven CARMA processes. *Ann. Inst. Statist. Math.* **53**, 113-124.
[6] Cont, R. and Tankov, P. (2003): *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, London.
[7] Klüppelberg, C., Lindner, A. and Maller, R. (2004): A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour. *J. Appl. Prob.* **41** (3) (to appear).
[8] Schoutens, W. (2003): *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, New York.

Fractionally Integrated Continuous Time ARMA Processes

Peter J. Brockwell
(joint work with Tina Marquardt)

Continuous-time models for time series which exhibit both heavy-tailed and long-memory behaviour are of considerable interest, especially for the modelling of financial time series, where such behaviour is frequently observed empirically. A recent paper of Anh, Heyde and Leonenko (2002) develops such models via the Green-function solution of fractional differential equations driven by Lévy processes. A very general class of Gaussian fractionally integrated continuous time models with extensive financial applications has also been introduced by Comte and Renault (1996).

In this paper we consider the class of Lévy-driven continuous-time ARMA (CARMA) processes and the fractionally integrated (FICARMA) processes obtained by fractional integration of the kernel of the CARMA process. For completeness we include a brief account of the derivation of this kernel and indicate its relevance to the stochastic volatility model of Barndorff-Nielsen and Shephard (2001). In the latter paper an Ornstein-Uhlenbeck process driven by a non-decreasing Lévy process was used to model volatility in a stochastic volatility model for log asset prices. The stationary Ornstein-Uhlenbeck process,

$$X(t) = \int_{-\infty}^{t} e^{-c(t-y)} dL(y), \ c > 0,$$

was chosen because it has a non-negative kernel ($g(t) = \exp(-ct)I_{[0,\infty)}(t)$) and consequently, if the driving Lévy process L is non-decreasing, the process X will be non-negative as is necessary if it is to represent volatility. However the use of the Ornstein-Uhlenbeck process restricts the class of volatility autocorrelation functions to functions of the form $\rho(h) = \exp(-ch)$ for some $c > 0$. Barndorff-Nielsen and Shephard suggested extending this class by using linear combinations of independent Ornstein-Uhlenbeck processes with positive coefficients, however the autocorrelation functions are still restricted to be monotone decreasing. If the Ornstein-Uhlenbeck process is replaced by a non-negative Lévy-driven CARMA process, a much larger class of autocorrelations can be modelled, and in particular the monotonicity constraint can be removed (see Brockwell (2003) for further details).

In this paper we derive explicit expressions for the kernel and auto covariance functions of a FICARMA process whose autoregressive polynomial has distinct zeroes. (Corresponding results for multiple zeroes can be obtained by letting distinct roots converge to a common limit.) We also consider the asymptotic behaviour of these functions for large lags. The results are continuous-time analogues of the results of Sowell (1992) for discrete-time fractionally integrated ARMA processes. A comprehensive treatment of the latter processes can be found in the book of Beran (1994). From a second-order point of view, the fractionally integrated CARMA process is a special case of the (Gaussian) fractionally integrated processes defined by Comte and Renault (1996), however the particular form of the kernel of the CARMA process leads to very simple expressions for the kernel and autocovariance functions for the corresponding fractionally integrated process.

If a CARMA(p, q) process is sampled at times $\{0, 1, 2, \ldots\}$, it is well-known that the sampled process is a discrete-time ARMA(p, r) process with $r < p$. It is therefore of interest to compare the behaviour of the fractionally integrated CARMA process sampled at integer times with that of the sampled CARMA process fractionally integrated (in the discrete-time sense). In this paper we make such a comparison for the fractionally integrated Ornstein-Uhlenbeck process.

References

[1] Anh, V.V., Heyde, C.C. and Leonenko, N.N. (2002): Dynamic models of long-memory processes driven by Lévy noise. *J. Appl. Prob.* **39**, 730-747.

[2] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B* **63**, 1-42.

[3] Beran, R. (1994): *Statistics for long-memory processes*. Chapman and Hall, New York.

[4] Brockwell, P.J. (2003): Representations of continuous-time ARMA processes. *J. Appl. Prob.* (to appear).

[5] Comte, F. and Renault, E. (1996): Long memory continuous time models. *J. Econometrics* **73**, 101-149.

[6] Sowell, F. (1992): Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics* **53**, 165-188.

Extremal Behaviour of Fractal Models

Boris Buchmann
(joint work with Claudia Klüppelberg)

A fractional Brownian motion (FBM) is a centred Gaussian process $(B^H{}_t)_{t\in\mathbb{R}}$ with covariance function

$$EB^H{}_tB^H{}_s = \frac{1}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}.$$

The parameter $H \in (0,1)$ is the so-called Hurst coefficient. For $H = 1/2$ FBM is the Wiener process, otherwise, both long memory ($H > 1/2$) and short memory ($H < 1/2$) occurs in the increments and FBM is no longer a semi-martingale. FBM has been studied by Kolmogorov in the fourties as a model for turbulence and by Mandelbrot and van Ness (1968) to describe certain aspects in the Nile data. Recently, it has been proposed as tool for financial applications (e.g. Hu and Øksendal (1999), Brody, Syroka and Zervos (2003)). The fractional Ornstein-Uhlenbeck process $(O_t^{H,\gamma,\sigma})$ (FOUP), i.e., the stationary solution of the Langevin equation

$$O_t^{H,\gamma,\sigma} = O_0^{H,\gamma,\sigma} - \gamma\int_0^t O_s^{H,\gamma,\sigma}\,ds + \sigma B_t^H,$$

where $\gamma > 0$, has been studied by Cheridito, Kawaguchi, Maejima (2003). It is a Gaussian process which again exhibits long memory ($H > 1/2$) and short memory ($H < 1/2$). In contrast to the latter authors we have studied the shape of the covariance function near zero. Combining both results allows us to develop the extreme value theory for FOUP based on classical results on Gaussian processes (Pickands (1969), Berman (1971), Leadbetter, Lindgren, Rootzén (1983)). More precisely, we obtain the norming constant $b_T(H,\gamma,\sigma)$ such that

$$\frac{2}{\Gamma(2H+1)^{1/2}}\frac{\gamma^H}{\sigma}(\log T)^{1/2}\left\{\max_{0\le t\le T} O_t^{H,\gamma,\sigma} - b_T(H,\gamma,\sigma)\right\} \xrightarrow{d} \quad G,$$

where G is a Gumbel distributed random variable. The extreme value theory can be extended to processes $X_t^{H,\gamma,f} := f(O_t^{H,\gamma})$ where f is a state space transform (SST), i.e., a continuous strictly increasing function. Our concept is related to the work of Davis (1982) where the extreme value theory for diffusions is studied by transforms in time and space. The process $X^{H,\gamma,f}$ to be in the maximum domain of a Gumbel distribution is provided by the following condition on the derivative, namely,

$$\lim_{z\to\infty}\frac{f'(z+a(z))}{f'(z)} = 1$$

for all functions $x \mapsto a(x)$ such that $a(x) = O(x^{-1})$ for $x \to \infty$. Generally, if $f'(x) = r(x)\exp(\kappa x^p)$ for sufficiently large x and some regularly varying function r the condition is satisfied whenever $p < 2$. If f is a SST such that for some constants $C_0 > 0$ and $C_1 \in \mathbb{R}$

$$\log f(x) = C_0x^2 + C_1 + o(1) \qquad \text{for } x \to \infty,$$

then $X^{H,f,\gamma}$ is found to be in the maximum domain of attraction of a Frechét distribution. The concept of SSTs can be related with a geometric approach to solve integral equations of the type

$$X_t - X_0 = \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s^H.$$

As $(B^H{}_t)$ is not a semi-martingale for $H \neq 1/2$ Itô integration can not be used to define an integral w.r.t FBM. Different approaches have been discussed in the literature (Hu and Øksendal (1999), Mikosch and Norvaiša (2000), Duncan, Hu and Pazik-Duncan (2000), Mazet, Alós and Nualart (2001)). We follow the approach of Zähle (1998, 2001) which works for $H > 1/2$. FBM is then sufficiently Hölder continuous such that the integral w.r.t FBM is well-defined as Riemann-Stieltjes integral whenever the integrand is Hölder of some order strictly larger $1 - H$. By a law of iterated logarithm (Arcones (1995)) FBM takes values in the weighted function space $\tilde{V}_H$ containing all functions f Hölder of at least any order strictly smaller H such that

$$\sup_t \frac{|f(t)|}{1 + |t|^H \sqrt{(\log\log)^+(|t|)}} < \infty\,.$$

Replacing B^H by any possible function $g \in \tilde{V}_H$ we derive a necessary and sufficient condition on μ and σ for existence and uniqueness of a solution in terms of SSTs and the FOUP.

References

[1] Arcones, M.A. (1995): On the law of iterated logarithm for Gaussian processes. *J. Theoret. Probab.* **8** (4), 877-904.

[2] Alòs, E., Mazet, O. and Nualart, D (2000): Stochastic calculus with respect to Gaussian processes. *Ann. Prob.* **29**, 766-801.

[3] Berman, S.M. (1971): Maxima and high level excursions of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **160**, 65-185.

[4] Brody, D., Syroka, J. and Zervos, M. (2003): Dynamical pricing of weather derivatives. *Quantitive Finance* (to appear).

[5] Buchmann, B. and Klüppelberg, C. (2004): Extreme value theory for stochastic processes driven by fractional Brownian motion. (In preparation.)

[6] Cheridito, P., Kawaguchi, H. and Maejima, M. (2003): Fractional Ornstein Uhlenbeck Processes. *Electron. J. Probab.* **8** (3), 1-14.

[7] Davis, R. A. (1982): Maximum and minimum of one-dimensional diffusions. *Stoch. Proc. Appl.* **13** (1), 1-9.

[8] Decreuesefond, L. and Üstünel, A.S. (1999): Stochastic analysis of the fractional Brownian motion. *Potential Anal.* **10** (2), 177–214.

[9] Duncan, T.E., Hu, Y. and Pasik-Duncan, B. (2000): Stochastic calculus for fractional Brownian motion. *SIAM J. Control Optim.* **38**, 582-612.

[10] Hu, Y. and Øksendal, B. (1999): Fractional white noise calculus and applications to finance. *SIAM J. Control Optim.* **38**, 582-612.

[11] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983): *Extremes and related properties of random sequences and processes.* Springer, New York.

[12] Mikosch, T. and Norvaiša, R. (2000): Stochastic integral equation without probability. *Bernoulli* **6** (3), 401-434.

[13] Pickands, J. III (1969): Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145**, 51-73.
[14] Pickands, J. III (1969): Asymptotic properties of the maximum in a stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145**, 75-86.
[15] Zähle, M. (1998): Integration with respect to fractal functions and stochastic calculus I. *Probab. Theory Relat. Fields* **111**, 333-374.
[16] Zähle, M. (2001): Integration with respect to fractal functions and stochastic calculus II. *Math. Nachr.* **225**, 154-183.

Structural Models for Credit Risk Migration

Ngai Hang CHAN

1. Abstract

A structural model for credit migration is considered in this paper. The proposed model is firm specific and depends on two parameters: the default distance and credit history. The default distance is the standardized logarithmic asset-to-liability ratio modelled by a Brownian motion and the credit history is modelled by an occupation time variable. By examining the properties of this occupation time variable, the credit performance of a given firm can be analyzed. This model not only allows one to derive a closed-form credit transition probability, but also explains default probability overlaps of different ratings. It can be used to back out the subjective thinking of credit performance of rating agencies.

2. Structural Models

Credit risk management is an important tool in finance, especially in the high-yield bond market and the bank loan market. An essential concern of a financial corporation is changes in credit ratings of companies. Nationally recognized statistical rating organizations (NRSRO), like Standard & Poor's and Moody's Investor Services, classify corporate bond issuers into different credit ratings in order to reflect their credit worthiness. Credit risk managers pay serious attentions to the ratings and transition matrices published by NRSRO. Transition matrices in the form of arrays of migrating probabilities constitute the building block of risk management tools, see for example JP Morgan's Credit Metrics and McKinsey's Credit Portfolio View. The Markov model of Jarrow, Lando and Turnbull (1997) uses the transition matrix to generate the term structure of credit spreads.

Predictions of transition probabilities have been receiving considerable amount of attentions recently. Most of the research make use of historical transition matrices and firm ratings to estimate future transition probabilities, see for example Aderson et. al (1991), Altman and Kao (1992), Kavvathas (2000), and Lando and Skodeberg (2002). Two types of default models are structural approach and reduced-form models. Structural type models suggest that a firm defaults when its asset value drops below its liabilities. KMV corporation implements this structural approach and generates expected default probabilities (EDPs) of firms. Details of

the KMV methodology can be found in Crosbe and Bohn (1993). Jarrow and Turnbull (1995) and Duffie and Singleton (1999) propose a second approach called reduced-form models. The time of default in this model is characterized by an exogenously defined intensity process.

This paper proposes a structural model of credit migration. There are at least two reasons to adopt the structural approach. First, it has a solid theoretical basis as it takes into account of the capital structure of a firm. Second, a structural model makes use of the distance-to-default of a firm. The distance-to-default values can be measured from the market, either through KMV or internal models. Finally, the recent acquisition of KMV and Moody's Investor Services provides the market with a possibility of using structural approach to measure credit transition probabilities.

The proposed model is firm specific and depends on distance-to-default and migrating signal duration. Using the idea of Gordy and Heitfield (2001), distance-to-default is mapped into different rating categories by partitioning the distribution of empirical data. The proposed model is able to capture the slow-to-respond features of rating agencies. Such a time-lagged response to new information can be interpreted as an extra rating criterion to reflect the market reputations of a rated company. If the distance-to-default is assumed to follow a symmetric distribution, then migrating probabilities generated from the model can still be skewed on one side. There are several desirable features of the model. Rating agencies can use it to explain changes in firm ratings in relation to current and historical credit performance. The proposed model also allows the overlap of EDPs across different letter grades, and offers a means to reconcile the empirical findings of Kealhofer et. al (1998). Analytical formula for calculating migrating probability can also be obtained.

3. Résumé

Dans cet article, on développe un modèle structurel pour la migration de crédit. On propose un modèle pour chaque entreprise dépendant de deux paramètres: le temps avant la faillite de l'entreprise et l'histoire du crédit. Le temps avant la faillite est défini comme le rapport logarithmique standardisé entre l'actif et la dette, et est modelisé par un mouvement Brownien. L'histoire du crédit est modelisée par une durée variable. En examinant les propriétés de ce temps variable, la performance du crédit d'une entreprise donnée peut être analysée. Ce modèle permet non seulement d'obtenir l'expression de la probabilité de transition du crédit, mais il explique aussi les chevauchements des probabilités de faillite pour differents taux. Le modèle peut aussi être utilisé pour reévaluer l'idée subjective donnée par une agence chargée de l'estimation de la performance d'un crédit.

References

[1] Altman, E., Kao, D. (1992): Rating drift in high yield bonds. *Journal of Fixed Income* **2**, 15-20.

[2] Anderson, P., Hanson, L.S., Keilding, N. (1991): Non-and semi parametric estimation of transition probabilities from censored observation of non-homogeneous Markov processes. *Scand. J. Statist.* **22**, 153-167.
[3] Crosbe, P.J., Bohn, J.R. (1993): Modeling default risk. KMV Corporate.
[4] Duffie, D., Singleton, K. (1999): Modeling term structures of defaultable bonds. *Review of Financial Studies* **12**, 687-720.
[5] Gordy, M., Heitfield, E. (2001): Of Moody's and Merton: A structural model of bond rating transitions. Working Paper, Board of Governors of the Federal Reserve System.
[6] Jarrow, R.A., Lando, D., Turnbull, S. (1997): A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**, 481-523.
[7] Jarrow, R.A., Turnbull, S.M. (1995): Pricing derivatives on financial securities subject to credit risk. *J. Finance* **50**, 53-58.
[8] Kavvathas. D (2000): Estimating credit rating transition probabilities for corporate bonds. Working Paper, Department of Economics, University of Chicago.
[9] Kealhofer, S., Kwok, S., Weng, W. (1998): Uses and abuses of bond default rates, KMV Corporation.
[10] Lando, D., Skodeberg, T.M. (2002): Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance* **26**, 423-444 .

Stochastic Volatility Models for Ordinal Valued Time Series

Claudia Czado
(joint work with Gernot Müller)

Our aim is to model the intraday development of stock prices, in particular the development of the price change process. The price changes have some specific features which we want to be covered by our model. One important feature is that price changes only occur in integer multiples of a certain amount, the so-called tick size. In modelling the price changes we therefore have to take into account that we observe a discrete time series. Also other important features of such time series are covered by the following model:

$$(1) \qquad y_t^{obs} = k \iff y_t \in [c_{k-1}, c_k) \qquad k \in \{1, \ldots, K\}$$

$$(2) \qquad y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \exp(h_t/2)\varepsilon_t \qquad t \in \{1, \ldots, T\}$$

$$(3) \qquad h_t = \mu + \boldsymbol{z}_t'\boldsymbol{\alpha} + \phi(h_{t-1} - \mu - \boldsymbol{z}_{t-1}'\boldsymbol{\alpha}) + \sigma\eta_t$$

A modified version of the underlying stochastic volatility model (2) and (3) for continuous responses was considered by Chib, Nardari and Shephard (2002). Observed are only the variables y_t^{obs}, which are discretized versions of the latent continuous variables y_t. $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are vectors of covariates, ε_t and η_t are assumed to be i.i.d. $N(0,1)$. We fix c_1 and μ for reasons of identifiability.

For the estimation of the parameters in this model we develop a MCMC algorithm, which is based on the algorithm presented in Chib, Nardari and Shephard (2002) for the underying continuous model. However, standard Gibbs MCMC steps for the additional discretization in Equation (1) lead to bad convergence behaviour of the resulting MCMC iterations. Figure 1 shows the cutpoint chains for simulated data, where the dotted lines indicate true values, when starting values are chosen to be $1.5, 3.0, 4.5, 6.0, 7.5$, respectively.

Therefore we develop additional grouped move (GM) steps to speed up the convergence especially for the chains of the cutpoints c_k. The idea of GM steps is based on a theorem of Liu and Sabatti (2000) which states: If Γ is a locally compact group of transformations defined on the sample space $\boldsymbol{S}$, L its left-Haar measure, $\boldsymbol{w} \in \boldsymbol{S}$ follows a distribution with density π, and $\gamma \in \Gamma$ is drawn from $\pi(\gamma(\boldsymbol{w}))|J_\gamma(\boldsymbol{w})|L(d\gamma)$, with $J_\gamma(\boldsymbol{w}) = \det(\partial\gamma(\boldsymbol{w})/\partial\boldsymbol{w})$, $\partial\gamma(\boldsymbol{w})/\partial\boldsymbol{w}$ the Jacobian matrix, then $\boldsymbol{w}^* = \gamma(\boldsymbol{w})$ also has density π (Liu and Sabatti (2000), Theorem 1).

Commonly π is considered to be the interesting posterior distribution. The difficulty in the choice of a suitable transformation group is to find one where on the one hand the problematic parameters are transformed and on the other hand the distribution
$\pi(\gamma(\boldsymbol{w}))|J_\gamma(\boldsymbol{w})|L(d\gamma)$ allows to draw samples very fast. We apply this theorem only for the conditional distribution of $\boldsymbol{w} := (y_1,\ldots,y_T,c_3,\ldots,c_{K-1},\beta_0,\ldots,\beta_p)$ given all the observations and all the remaining parameters. This conditional distribution can be computed iteratively. In order to get an easy sampling distribution we now use the scale group, $\Gamma = \{\gamma > 0 : \gamma(\boldsymbol{w}) = (\gamma w_1,\ldots,\gamma w_d)\}$, with $\gamma^{-1}d\gamma$ as left-Haar measure. This finally leads to a Gamma distribution for γ^2. Therefore, after each iteration of our MCMC sampler, we insert the corresponding GM-step, which consists of drawing a γ^2 from the resulting Gamma distribution and update $\boldsymbol{w}$ to $\gamma \cdot \boldsymbol{w}$. As Figure 2 shows, this significantly speeds up the convergence of the algorithm. Here we used the same simulated data as in Figure 1 and the same starting values. By using the GM steps the chains reach the area around the true values within about 50 iterations.

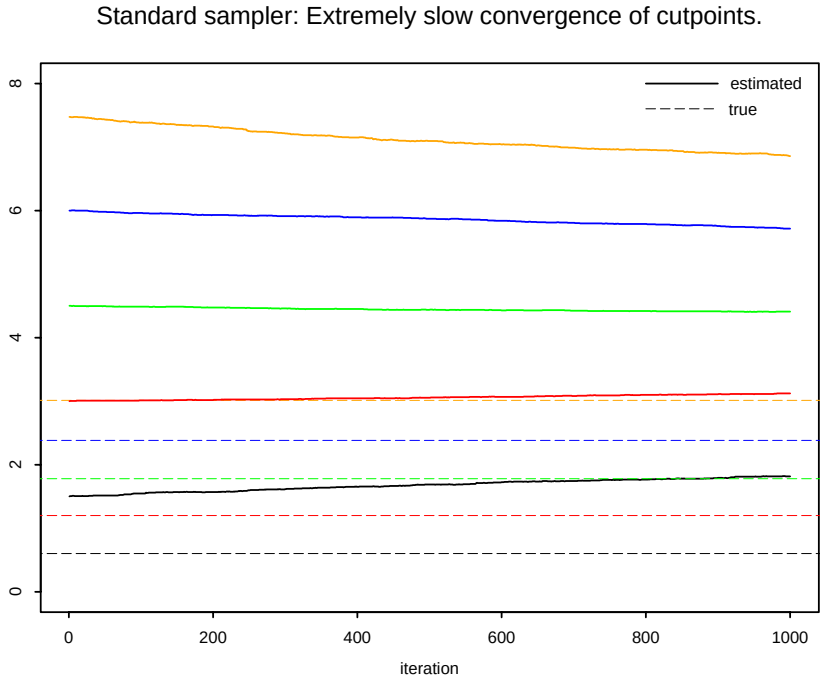

FIGURE 1. First 1000 MCMC iterations for cutpoints produced by standard Gibbs sampler. The dotted lines indicate the true values.

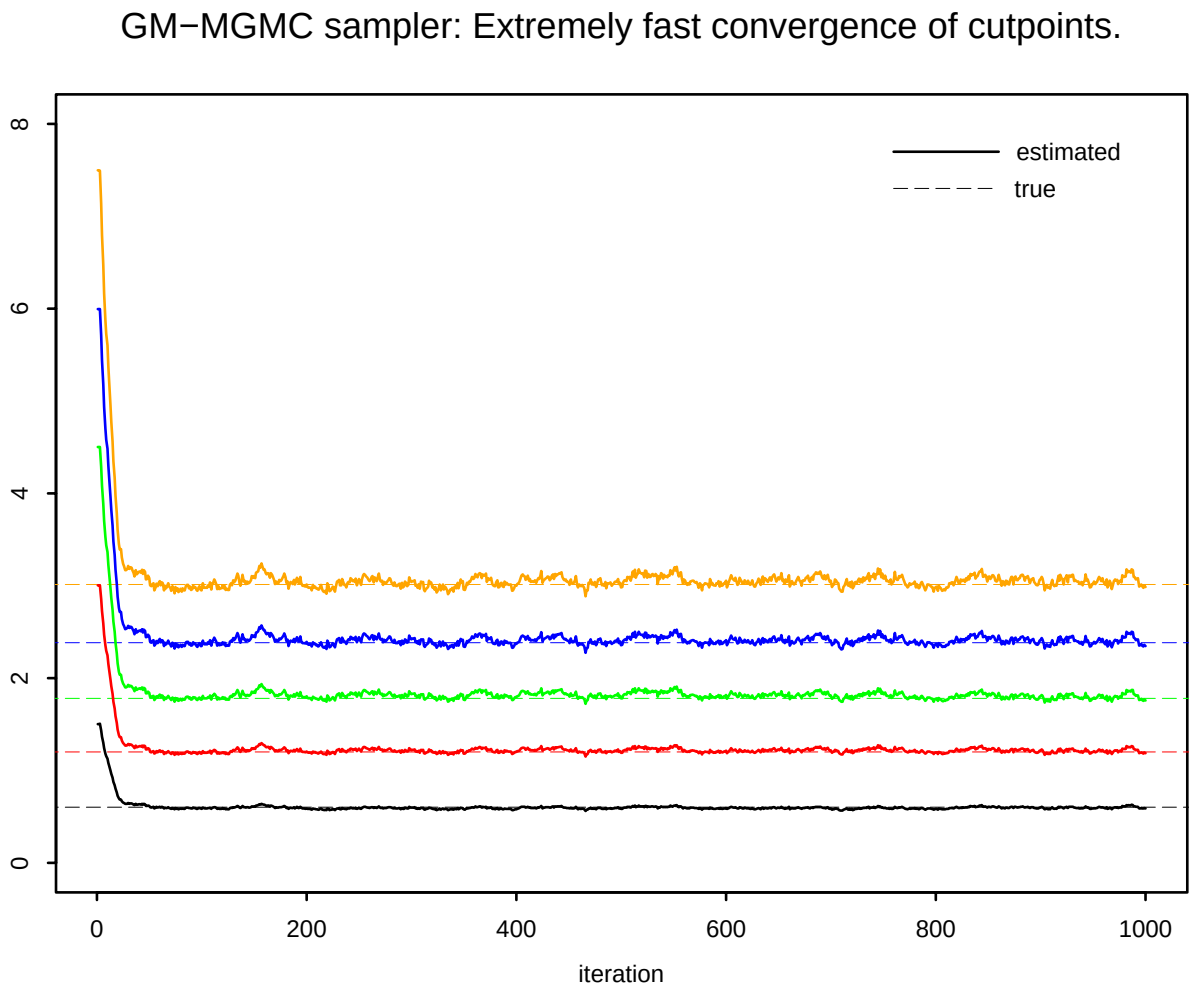

FIGURE 2. First 1000 iterations of chains for cutpoints produced by GM-MGMC sampler. The dotted lines indicate the true values.

Finally we fit the model to IBM intraday data collected in January 2001. We show that a positive price jump increases the probability that the next price jump will be negative and vice versa. Furthermore, the time between transactions has an impact on the log-volatility in Equation (3): The more time elapses between two subsequent transactions, the higher is the probability for a big price jump (upwards or downwards).

REFERENCES

[1] Chib, S., Nardari, F. and Shephard, N. (2002): Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* **108**, 281-316.

[2] Liu, J.S. and C. Sabatti, C. (2000): Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation.*Biometrika* **87**, 353-369.

Data Driven Local Coordinates for Linear State Space Systems

M. Deistler
(joint work with T. Ribarits)

The topic of the lecture is embedded in a larger research program at our department concerning identification of ARMA(X) and state space (SS) systems, in particular stressing the multivariate case. The motivation for this program is that in many applications in econometrics and engineering, AR(X) type models are still preferred in modelling linear systems, despite the fact that ARMA(X) and SS models are more flexible. The reasons for this are that parametrization and estimation is much simpler in the AR(X)case, in particular the (maximum likelihood) estimators are of least squares type, they are explicitly given, numerically fast and have no problems of local optima.

In making ARMA(X) and SS systems more competitive, one direction we follow is to look for better parametrizations, in particular for the SS case. Note that ARMA(X) and SS systems represent the same classes of transfer functions, but SS systems in general have larger classes of observationally equivalent systems. The latter fact should be considered as an advantage, because this allows for selection of more suitable representatives.
We consider a state space system

$$x_{t+1} = Ax_t + B\varepsilon_t$$

$$y_t = Cx_t + \varepsilon_t$$

where ε_t is the s–dimensional white noise innovation, x_t is the n–dimensional state and y_t is the s–dimensional output; (A,B,C) are the system matrices.

We consider two approaches: The first approach, *data driven local coordinates (DDLC)*, has been originally introduced by McKelvey and Helmersson. Here (A,B,C) is embedded in $\mathbb{R}^{n^2+2sn}$. For minimal (A,B,C), the equivalence classes are n^2–dimensional manifolds. Commencing from an initial estimator (A,B,C), the orthocomplement to the tangent space to this manifold at (A,B,C) is taken as a parameter space. The likelihood function then is optimized over this parameter space and the procedure is iterated with the new estimate. We analyse the topological and geometrical properties of this parametrization which are relevant for identification. In particular we show that this parametrization is locally homeomorphic, but globally, problems, e.g. of nonidentifiability, arise.

The second approach is *separable least squares DDLC (`slsDDLC`)*. Here first a least squares step is performed in order to concentrate out B. The concentrated likelihood then only depends on (A,C) and for this reduced parameter space, again DDLC is performed. We again analyse the topological and geometrical properties of this parametrization.

Finally, the numerical properties of the maximum likelihood estimation parametrized by the 'classical' echelon form, by DDLC and by `slsDDLC` are investigated in a simulation study. `slsDDLC` is found to be superior to DDLC and both give much better results than echelon forms.

Estimation in Semi-parametric Volatility Models

Feike C. Drost

The availability of large data sets is rapidly growing, especially in finance. In discrete time models, ARMA models with GARCH type errors are quite suitable to pick up the time-varying nature of the first two conditional moments with only a few parameters. However the implications of parametric volatility models for higher order conditional moments, are not reflected in the data. More precisely formulated, the conditional error distribution cannot be described by just a functional form of the conditional volatility and a fixed nonparametric distribution. To avoid this kind of misspecification we use a semi-parametric model where the conditional error distributions are treated as a nuisance parameter. In continuous time models, stochastic volatility models are used to model similar stylized facts. Since the volatility of volatility functions in these models do not affect the first two conditional moments, a nonparametric approach is advised here as well.

Usually, from a practitioners point of view, some finite dimensional parameter is of interest, for example, the mean or median as a measure of location, the Value at Risk as a measure of risk, etc. The question arises how to efficiently estimate such quantities in general semi- and nonparametric models. To study what is best asymptotically, one needs a bound on the asymptotic performance of estimators in the presence of an infinite dimensional nuisance parameter. For the i.i.d. case, a comprehensive account on the present theory along these lines is given in Bickel, Klaassen, Ritov, and Wellner (1993). In financial data, of course, the time dimension also plays an important role. Drost, Klaassen, and Werker (1997) and Koul and Schick (1997) have developed a unified theory for time series models with independently and identically distributed innovations. This covers, for example, semi-parametric ARMA models (Kreiss (1987)). Recent work in applied financial econometrics shows that the assumption of i.i.d. innovations does not hold when using standard semi-parametric time series models, see Engle and Russell (1998).

Based on the first two conditional moments, a popular method to estimate the parameters is the Quasi Maximum Likelihood (QML) approach. This method applies the Maximum Likelihood (ML) procedure to the data as if the conditional distributions are normal. Under some regularity conditions this approach leads to consistent and asymptotically normal estimators, but the efficiency may be quite low.

An alternative to QML, is the Generalized Method of Moments (GMM). Here the conditional moments are used together with a suitably chosen instrument. It is well-known that the QML estimator can be obtained by a suitable choice of

the instruments. However, these QML instruments are not optimal since they do not use the possible time varying character of the third and fourth moment of the conditional error distribution. Optimal instruments are easily derived along the lines in Wefelmeyer (1996).

Although the GMM estimator is optimal in the class of estimators based on the first two conditional moments, it is not necessarily the optimal estimator. As in the aforementioned literature, additional information can be gained by estimating the conditional error distribution. In several applications it is known that the conditional error distribution (given the total past) equals the conditional error distribution given some restricted information set $\mathcal{H}_t$. Special cases are:

- i.i.d. errors: choose the restricted information set $\mathcal{H}_t$ to be the trivial sigma field.
- Markov errors: choose the restricted information set $\mathcal{H}_t = \sigma(\varepsilon_t)$, the information set generated by the last error.
- general case: do not put any restrictions on $\mathcal{H}_t$, and choose the restricted information set to be the full information set.

Since the first two conditional moments are of particular interest in financial applications, we present the score functions for QML, GMM, and the Semiparametric (SP) approach in the following example. The estimator based on the SP score performs best.

Example 1. *Consider the semi-parametric location-scale model,*

$$Y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}, \qquad E_{G_t}\varepsilon_{t+1} = 0, E_{G_t}\varepsilon_{t+1}^2 = 1, \qquad G_t \equiv \mathcal{L}(\varepsilon_{t+1}|\mathcal{H}_t),$$

with location-scale score

$$\dot{l}(\varepsilon_{t+1}; G_t) = \left[-\frac{g_t'}{g_t}(\varepsilon_{t+1}), -\frac{1}{2}\left\{ 1 + \varepsilon_{t+1}\frac{g_t'}{g_t}(\varepsilon_{t+1}) \right\} \right]^T.$$

Apart from the model assumptions and some regularity conditions, nothing is known about the conditional error distributions. Put $\gamma_t = E_{G_t}\varepsilon_{t+1}^3, \kappa_t = E_{G_t}\varepsilon_{t+1}^4$, *then the score functions of the QML/GMM/SP/ML estimators are given by, respectively,*

$$\begin{aligned}
l_{t+1}^{QML} &= \left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1}^2 - 1 \end{bmatrix}, \\
l_{t+1}^{GMM} &= \left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] \begin{bmatrix} 1 & \gamma_t \\ \gamma_t & \kappa_t - 1 \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1}^2 - 1 \end{bmatrix} \equiv \left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] l^*(\varepsilon_{t+1}; G_t), \\
l_{t+1}^{SP} &= \left\{ \left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] - E\left(\left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] \middle| \mathcal{H}_t \right) \right\} \dot{l}(\varepsilon_{t+1}; G_t) \\
&\qquad + E\left(\left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] \middle| \mathcal{H}_t \right) l^*(\varepsilon_{t+1}; G_t), \\
l_{t+1}^{ML} &= \left[\frac{\dot{\mu}_t}{\sigma_t}, \frac{\dot{\sigma}_t^2}{\sigma_t^2} \right] \dot{l}(\varepsilon_{t+1}; G_t).
\end{aligned}$$

Note that the implied information is strictly increasing: the SP score is closest (in $\mathbb{L}_2$-sense) to the unattainable ML score.

References

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner J.A. (1993): *Efficient and Adaptive Statistical Inference for Semi-parametric Models*. John Hopkins University Press, Baltimore.
[2] Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997): Adaptive estimation in time-series models. *Ann. Statist.* **25**, 786-818.
[3] Engle, R.F. and Russell, J.R. (1998): Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**, 1127-1162.
[4] Koul, H.L. and Schick, A. (1997): Efficient estimation in nonlinear autoregressive timeseries models. *Bernoulli* **3**, 247-277.
[5] Kreiss, J.-P. (1987): On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**, 112-133.
[6] Wefelmeyer, W. (1996): Quasi-likelihood models and optimal inference, *Ann. Statist.* **24**, 405-422.

Modelling Dependence for High–Frequency Data in Finance

Paul Embrechts
(joint work with W. Breymann and A. Dias)

Based on high-frequency data for US$/DM and US$/Yen, stylised facts for extremal dependence in financial data are investigated. Starting with 5' data, through deseasonalisation, data are investigated at the (1 hr, 2 hr, 4 hr, 8 hr, 12 hr, 1 day) frequencies. Dependence is modelled throughout based on the concept of copula. In order to get close to iid bivariate residual vectors, several stochastic models are fitted at the various frequencies. These models include marginal ARMA-GARCH, CCC-GARCH, VECH and DCC-GARCH. The following tests/statistical techniques are performed on the residuals:

- tests for ellipticity
- copula fitting
- dynamic dependence parameters
- comparison of high (low) quantile fitting procedures (leading to the Clayton model)
- spectral measure estimation
- change point analysis.

This work is done jointly with W. Breymann and A. Dias (see [1]) and A. Dias (see [2]). Further references and related work are to be found under www.math.ethz.ch/~embrechts. We also would like to thank Olsen and Associates for providing the data.

References

[1] Breymann, W., Dias, A. and Embrechts P. (2003): Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3** (1), 1-16.
[2] Dias, A. and Embrechts, P.: Dynamic copula models for multivariate high-frequency data in finance. Submitted.

Extremal Behaviour of Continuous-Time Moving Average Processes

Vicky Fasen

We consider a stationary continuous-time moving average (MA) process

$$Y(t) = \int_{-\infty}^{t} f(t-s)\, dL(s) \qquad \text{for } t \geq 0\,,$$

where f is a deterministic kernel function and L is a Lévy process whose increments, represented by $L(1)$, are subexponential and in the domain of attraction of the Gumbel distribution. Examples are Weibull-like distributions with $\alpha \in (0,1)$. The extremal behaviour of subexponential MA processes in the domain of attraction of the Fréchet distribution are well studied Rootzén (1978) and Rosinski and Samorodnitsky (1993). A good overview about subexponential distributions can be found in Embrechts et al. (1997) and about Lévy processes in Sato (1999).

Extremes of $\{Y(t)\}_{t\geq 0}$ are caused by big jumps of the driving Lévy process in combination with large values of the kernel function f. This means that discrete time points $\{t_n\}_{n\in\mathbb{N}}$ chosen properly to incorporate the times where big jumps of the Lévy process and the extremes of the kernel function occur characterise the extremal behaviour of the continuous time process. We restrict ourselves to kernel functions with a finite number of local extremes. Examples for Y include a Weibull-Ornstein-Uhlenbeck process, certain shot noise processes and CARMA processes (Brockwell (2001)).

The extremal behaviour of the discrete-time process $\{Y(t_n)\}_{n\in\mathbb{N}}$ is described by the weak limit of a sequence of marked point processes, i.e.

- by the point processes of exceedances over high thresholds, and
- by marks, which are stochastic processes themselves, and characterize the behaviour of $\{Y(t)\}_{t\geq 0}$, if $Y(t_n)$ exceeds a high threshold.

The limiting distribution of such a sequence of marked point processes is a Poisson process with deterministic marks represented by a scaled version of the kernel function. Further we can compute the normalising constants of the maxima to converge weakly to the Gumbel distribution. The results are similar to the extremal behaviour of discrete MA processes (Davis and Resnick (1988), Rootzén (1986)).

References

[1] Brockwell, P.J. (2001): Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53** (1), 113-123.

[2] Davis, R. and Resnick, S. (1988): Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. *Stoch. Proc. Appl.* **30**, 41-68.

[3] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997): Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

[4] Rootzén, T. (1978): Extremes of moving averages of stable processes. *Ann. Probab.* **6** (5), 847-869.

[5] Rootzén, H. (1986): Extreme value theory for moving average processes. *Ann. Probab.* **14** (2), 612-652.

[6] Rosinski, J. and Samorodnitsky, G. (1993): Distributions of subadditive functionals of sample paths of infinitely divisible processes. *Ann. Probab.* **21** (2), 996-1014.
[7] Sato, K.I. (1999): *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

Nonparametric Value-at-Risk Estimates

Jürgen Franke
(joint work with Mabouba Diagne and Peter Mwita)

We consider a financial time series S_t with returns $R_t = (S_t - S_{t-1})/S_{t-1}$, and we want to estimate the conditional Value-at-Risk, the conditional α-quantile $q_\alpha(r,x) \equiv \mathrm{VaR}(r,x)$ of R_{t+1} given past returns $\overrightarrow{R}_t^{(p)} = (R_t, \ldots, R_{t-p+1}) = r$ and exogeneous market information $X_t = x \in R^d$ (returns of index or other stock, FX or interest rates etc.), i.e. we have

$$\mathrm{pr}(R_{t+1} \le q_\alpha(r,x) | \overrightarrow{R}_t^{(p)} = r, X_t = x) = \alpha.$$

First, we model the returns as a nonlinear ARX-ARCHX-process

$$R_{t+1} = m(R_t, \ldots, R_{t-p+1}, X_t) + \sigma(R_t, \ldots, R_{t-p+1}, X_t) Z_{t+1} \tag{1}$$

with i.i.d. innovations Z_t having mean 0 and variance 1 and a known distribution. We can estimate the local trend and volatility functions m, σ nonparametrically by

- kernel estimates or local polynomials for either **lowdimensional** arguments or under restrictions on the functions m, σ (e.g. additive or generalized additive structure)
- neural networks for **highdimensional** arguments

where, for estimating $\sigma^2(s,x)$, we use the residual-based estimator of Fan and Yao (98). Both nonparametric approaches lead to asymptotically normal and, for tuning parameter (bandwidth for local smoothers and number of neurons for neural networks) changing appropriately for increasing sample size, consistent estimates if the time series (R_t, X_t) satisfies some mixing condition (Franke and Diagne, 2002, Franke, Kreiss and Mammen, 2002, Franke et al., 2002, Franke, Neumann and Stockis, 2004). Using estimates for m, σ, we get as a nonparametric VaR-estimate $\hat{q}_\alpha(r,x) = \hat{m}(r,x) + \hat{\sigma}(s,x) q_\alpha^Z$, where q_α^Z denotes the α - quantile of the law of Z_t. An example for a German stock price illustrates the feasibility of the nonparametric approach and the usefulness of incorporating exogeneous information in the calculation of VaR.

Instead of starting from model (1), we can also estimate the quantile function $q_\alpha(r,x)$ directly, either by a nonparametric version of the Koenker-Bassett (1978) regression quantile approach exploiting that

$$q_\alpha(r,x) = \arg\min_{q \in R} E\left(|R_{t+1} - q|_\alpha \,\middle|\, \overrightarrow{R}_t^{(p)} = r, X_t = x \right)$$

with $|y|_\alpha = (1-\alpha)y^- + \alpha\ y^-$ or by first estimating the conditional distribution function $F(y|r,x) = E(1_{(-\infty,y]}(R_{t+1})|\overrightarrow{R}_t^{(p)} = r, X_t = x)$ nonparametrically and inverting it. Again, we get consistent nonparametric estimates $\hat{q}_\alpha(r,x)$ based on neural networks (Diagne, 2002) or on local smoothing (Abberger, 1996, Franke and Mwita 2003) if we assume the returns to follow a quantile ARX-model

$$R_{t+1} = q_\alpha(R_t, \ldots, R_{t-p+1}, X_t) + \eta_{t+1}. \tag{2}$$

The innovations η_t may depend on the past $R_n, n < t$, and may have infinite variance, and, in contrast to the VaR estimates based on (1), we do not have to assume their distribution to be known. To get a notion of local variability like volatility which does not require moment assumptions we may specify (2) to the following quantile ARX-ARCHX-model

$$R_{t+1} = q_\alpha(R_t, \ldots, R_{t-p+1}, X_t) + \sigma_\alpha(R_t, \ldots, R_{t-p+1}, X_t)W_{t+1}$$

where the α - scale of R_{t+1} given $\overrightarrow{R}_t^{(p)} = r, X_t = x$ is the α - quantile of $|R_{t+1} - q_\alpha(r,s)|_\alpha$. The i.i.d. innovations W_t are standardized to have α - quantile 0 and α - scale 1 (compare also Koenker, 1999). Similar as in the familiar model (1), $q_\alpha(r,x)$ and $\sigma_\alpha(r,x)$ may be estimated simultaneously (Mwita, 2003).

References

[1] Abberger, K. (1996): *Nichtparametrische Schätzung bedingter Quantile in Zeitreihen - Mit Anwendung auf Finanzmarktdaten*. Hartung-Gore Verlag, Konstanz.

[2] Diagne, M. (2002): *Financial Risk Management and Portfolio Optimization Using Artificial Neural Networks*. PhD Thesis. TU Kaiserslautern.

[3] Fan, J. and Yao, Q. (1998): Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645-660.

[4] Franke, J., Kreiss, J.-P. and Mammen, E. (2002): Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8**, 1-37.

[5] Franke, J., Kreiss, J.-P., Mammen, E. and Neumann, M. (2002): Properties of the nonparametric autoregressive bootstrap. *J. Time Ser. Anal.* **23**, 555-585.

[6] Franke, J. and Diagne, M. (2002): Estimating market risk with neural networks. Report in Wirtschaftsmathematik 83, University of Kaiserslautern. Submitted.

[7] Franke, J. and Mwita, P. (2003): Nonparametric estimates for conditional quantiles of time series. Report in Wirtschaftsmathematik 87, University of Kaiserslautern. Submitted.

[8] Franke, J., Neumann, M. and Stockis, J.P. (2004): Bootstrapping nonparametric estimates of the volatility function. *J. Econometrics* **118**, 189-218.

[9] Koenker, R. und Basset, G. (1978): Regression quantiles. *Econometrica* **46**, 33-50.

[10] Koenker, R. (1999): Galton, Edgeworth, Frisch and prospects for quantile regression in econometrics. *J. Econometrics* **95**, 347-374.

[11] Mwita, P. (2003): *Semi-parametric Estimation of Conditional Quantiles for Time Series with Applications in Finance*. PhD Thesis, University of Kaiserslautern.

Gibbs Sampling for State Space Modelling of Time Series of Counts

Sylvia Frühwirth-Schnatter
(joint work with Helga Wagner)

For applied statisticians it is not unusual to have to deal with time series of counts. As such data are necessarily non-negative integers, it is often appropriate to assume the observed process y_t follows a Poisson distribution: $y_t \sim \text{Poisson}(\lambda_t)$. To capture the effect of exogenous variables z_t, for independent observations a log-linear model could be applied where $\lambda_t = \exp(z_t'\beta)$, with λ_t being the mean of the time series observation y_t given β, and β being a vector of unknown coefficients to be estimated from the data.

To account for the dependency likely to be present in time series data of counts, various extensions of the log-linear model have been suggested which, following Cox (1981), may be classified into parameter-driven and observation-driven models. In an observation driven model the conditional distribution of y_t is specified as a function of the past observations $y_{t-1}, y_{t-2}, \ldots$, see for instance Kaufmann (1987). While observations-driven models are easy to estimate, their theoretical properties can be difficult to derive in comparison to parameter-driven models.

Here we consider parameter-driven models, where the conditional distribution of y_t is allowed to change over time and this change is driven by a latent process. This latent process could be a hidden Markov chain as in Leroux and Puterman (1992), or random effects as in Albert (1992). Smooth changes of the conditional distribution of y_t through state-space models have been considered e.g. in West et al. (1995) and Harvey and Fernandes (1989), whereas a latent stationary autoregressive process has been introduced into the generalized linear model by Zeger (1988).

Estimation of parameter-driven Poisson time series models is known to be a challenging problem. In fact, estimation of these models using maximum likelihood estimation is hampered by the fact that the marginal likelihood, where the latent process is integrated out, is in general not available in closed form. Each evaluation of the likelihood function requires to use some numerical method for solving the necessary high-dimensional integration. One particular useful method in this respect is importance sampling which was applied in Durbin and Koopman (2000) to state space modelling of counts data.

Alternatively, estimation of these models is also feasible within a Bayesian framework using data augmentation as in Tanner and Wong (1987) and Markov chain Monte Carlo (MCMC) methods, as illustrated first by Zeger and Karim (1991). Since this seminal paper, a number of authors have contributed to MCMC estimation of these models. We mention here in particular Shephard and Pitt (1997) and Gamerman (1998) for non-Gaussian time series models based on distributions from the exponential family, and Chib et al. (1998) and Chib and Winkelmann (2001) for more general count data models.

A major difficulties with any of the existing MCMC approaches is that practical implementation requires the use of a Metropolis-Hastings algorithm at least for part of the unknown parameter vector, which in turns make it necessary to define

suitable proposal densities in rather high-dimensional parameter spaces. Single-move sampling for this type of models is known to be potentially very inefficient, see e.g. Shepard and Pitt (1997). The main contribution of this article is to show that *straightforward Gibbs sampling of all parameters*, involving only sampling from simple distributions such as multivariate normal, inverse Gamma, exponential and low-dimensional discrete distributions, is feasible for practical Bayesian estimation of most of the parameter-driven models for time series of counts suggested in the literature so far. This rather unexpected result is achieved by introducing two additional sequences of latent variables through data augmentation. One of these sequences are the unobserved inter arrival times of the events under investigation. The introduction of this first sequence eliminates the non-linearity of the observation equation whereas the non-normality of the error term remains which follows a log exponential distribution. As the mean of the exponential distribution is equal to 1, this distribution is independent of any parameter and may be approximated by a mixture of normal distribution in a similar way as in Kim et al. (1998) who used a mixture approximation to the density of a $\log \chi^2$-distribution in the context of stochastic volatility models. By introducing the component indicator as a second sequence of missing data, the resulting model may be thought of a partially Gaussian state space model as in Shephard (1994). This is particularly useful for state space models for Poisson time series, as multi-move-sampling of the whole state process through forward-filtering backward sampling as in Frühwirt-Schnatter (1994), Carter and Kohn (1994) and de Jong and Shephard (1995) is now possible.

References

[1] Albert, J.H. (1992): A bayesian analysis of a poisson random-effects model for home run hitters. *Amer. Statist.* **46**, 246-253.

[2] Carter, C.K. and R. Kohn, R. (1994): On Gibbs sampling for state space models. *Biometrika* **81**, 541-553.

[3] Chib, S., Greenberg, E. and Winkelmann, R. (1998): Posterior simulation and Bayes factors in panel count data models. *J. Econometrics* **86**, 33-54.

[4] Chib, S. and Winkelmann, R. (2001): Markov Chain Monte Carlo analysis of correlated count data. *J. Bus. Econom. Statist.* **19**, 428-435.

[5] Cox, D.R. (1981): Statistical analysis of time series: Some recent developments. *Scand. J. Statist.* **8**, 93-108.

[6] de Jong, P. and Shephard, N. (1995): The simulation smoother for time series models. *Biometrika* **82**, 339-350.

[7] Durbin, J. and Koopman, S.J. (2000): Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives.*J. R. Stat. Soc. Ser. B* **62** (1), 3-56.

[8] Frühwirth-Schnatter, S. (1994): Data augmentation and dynamic linear models. *J. Time Ser. Anal.* **15**, 183-202.

[9] Gamerman, D. (1998): Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika* **85**, 215-227.

[10] Harvey, A. C. and Fernandes, C. (1989): Time series models for count or qualitative observations. *J. Bus. Econom. Statist.* **7**, 407-417.

[11] Kaufmann, H. (1987): Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Ann. Statist.* **15**, 79-98.

[12] Kim, S., Shephard, N. and Chib, S. (1998): Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**, 361-393.
[13] Leroux, B.G. and Puterman, M.L. (1992): Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* **48**, 545-558.
[14] Shephard, N. (1994): Partial non-Gaussian state space. *Biometrika* **81**, 115-131.
[15] Shephard, N. and Pitt, M.K. (1997): Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653-667.
[16] Tanner, M.A. and Wong, W.H. (1987): The calculation of posterior distributions by data augmentation, *J. Amer. Statist. Assoc.* **82**, 528-540.
[17] West, M., Harrison, Jeff, P., Migon, H.S. (1985): Dynamic generalized linear models and Bayesian forecasting. *J. Amer. Statist. Assoc.* **80**, 73-83.
[18] Zeger, S.L. (1988): A regression model for time series of counts. *Biometrika* **75**, 621-629.
[19] Zeger, S.L. and Karim, M.R. (1991): Generalized linear models with random effects: a Gibbs sampling approach. *J. Amer. Statist. Assoc.* **86**, 79-86.

Estimation and Change Point Detection with a Hidden Markov Model in Finance

X. Guo

Consider a probability space $(\Omega, \mathcal{F}, P)$ and $t \in [0, T]$ for some $T > 0$. Suppose that $\alpha(t)$ is a finite-state continuous time Markov process with state space $M = \{z^1, \ldots, z^m\}$ and generator $Q = (q^{ij}) \in R^{m \times m}$.

Assume that the Markov process $\alpha(t)$ is observed with the process $y(t)$ such that

$$\left\{ \begin{array}{ll} dy(t) = & \mu_{\alpha(t)} dt + \sigma_{\alpha(t)} dw(t), \\ y(0) = & 0 \ \text{ w.p. } 1, \end{array} \right. \tag{1}$$

where $w(\cdot)$ is a standard one-dimensional Brownian motion independent of $\alpha(t)$, and the drift μ and diffusion σ take different values when $\alpha(t)$ is in different states.

Given Eq. (1), we are primarily interested in the parameter estimation problem that is motivated by checking the validation of this Markov modulated (or regime switching) model in the financial time series data. The critical issue is the identification of the Markov chain $\alpha(\cdot)$

The problem is trivial when σ's are all distinct and the observation is continuous: the quadratic variation of Ito's calculus will easily reveal the state of $\alpha(t)$. Therefore, we are mostly interested in two cases: (A) when σ's are independent of $\alpha(t)$ and the observation is continuous, and (B) when the observation is discrete. In a joint work with G. Yin [2], we address (A) under the (more general) framework of Wonham filters; In a joint work with D. Chan (included in the summary report [3]), we address (B) and propose a statistical estimation method for applying this regime switching model to analyze financial time series data. Here, we suggest a notion of "regime shift" and a detection method based on a case study of AT&T stock price.

The optimality of Wonham filter is a direct corollary of a result of independent interest concerning the relationship between choices of error functions and the

optimality of conditional expectations. This is joint work with A. Banerjee and H. Wang [1].

1. Wonham filters

Assume that the Markov process $\alpha(t)$ is observed with the process $y(t)$ such that

$$\begin{cases} dy(t) = & \alpha(t)dt + \sigma(t)dw(t), \\ y(0) = & 0 \ \text{ w.p. } 1, \end{cases} \tag{2}$$

as in Eq. (1) where $\sigma(\cdot) : R \mapsto R$, is a continuously differentiable function with $\sigma(t) \geq c$ for all $t \in [0, T]$ and some $c > 0$.

In this framework, suppose we assume that the values of the states $z^1, \cdots, z^m$ and the generator Q are *known* a priori and fixed. Then, a classical result states that the posterior probability $p(t) = (p^1(t), \ldots, p^m(t)) \in R^{1\times m}$, with $p^i(t) = P(\alpha(t) = z^i | y(s),\ 0 \leq s \leq t), p^i(0) = p_0^i, (i = 1, \ldots, m)$ satisfies the following system of stochastic differential equations:

$$\begin{array}{rl} dp^i(t) = & \sum_{j=1}^m p^j(t)q^{ji}dt - \sigma^{-2}(t)\overline{\alpha}(t)[z^i - \overline{\alpha}(t)]p^i(t)dt \\ & +\sigma^{-2}(t)[z^i - \overline{\alpha}(t)]p^i(t)dy(t),\ i = 1, \ldots, m. \end{array} \tag{3}$$

Here, $\overline{\alpha}(t) = \langle z, p(t)\rangle, z = (z^1, \ldots, z^m)'$, and v' denotes the transpose of v. This is known as the *Wonham filter*, which is the first finite dimensional filter for non-Gaussian processes. It is known to be optimal under the mean square error.

1.1. Optimality of conditional expectation as BLFs.

We first show the optimality of Wonham filter under a general class of loss functions known as *Bregman loss functions* (BLFs) (including L^2-loss functions). This is a direct corollary of our study [1], where we provide necessary and sufficient conditions for general loss functions under which the conditional expectation is the unique optimal predictor.

Theorem 1 (Optimality Property)**.** *Let $\phi : R^d \mapsto R$ be a strictly convex, differentiable function. Let $(\Omega, \mathcal{F}, P)$ be an arbitrary probability space and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Let X be any $\mathcal{F}$-measurable random variable taking values in R^d for which both $E[X]$ and $E[\phi(X)]$ are finite. Then*

$$\arg\min_{Y\in\mathcal{G}} E[D_\phi(X, Y)] = E[X|\mathcal{G}].$$

Theorem 2 (Exhaustiveness of BLFs)**.** *Let $F : R \times R \mapsto R$ be a non-negative function such that $F(x, x) = 0, \forall x \in R$. Assume that F and F_x are both continuous functions. If for all random variables X, $E[X|\mathcal{G}]$ is the unique minimizer for $E[F(X, Y)]$ over random variables $Y \in \mathcal{G}$, i.e., $\arg\min_{Y\in\mathcal{G}} E[F(X, Y)] = E[X|\mathcal{G}]$, then $F(x, y) = D_\phi(x, y)$ for some strictly convex, differentiable function $\phi : R \mapsto R$.*

Here the BLF $D_\phi : R^d \times R^d \mapsto R$ is defined as $D_\phi(x, y) = \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y)\rangle$, for any (strictly) convex and differentiable function $\phi : R^d \mapsto R$.

For further properties of BLFs and corresponding exhaustiveness results for higher dimensions, see [1].

1.2. **Wonham filter with random parameters.** Now, assume that z^i's (or (q^{ij})) are not available, and that only their noisy/corrupted measurements/observations/distributional information are at our disposal. We propose approximated (suboptimal) filters and prove their (exponential rate) of convergence to the desired Wonham filter under simple ergodic conditions.

For instance, if we assume that a sequence of observations of the form $\widehat{z}_n = (\widehat{z}_n^1, \ldots, \widehat{z}_n^m)' \in R^{m\times 1}$ such that $E\widehat{z}_n = z$ can be obtained, then by defining $\overline{z}_n = \frac{1}{n}\sum_{j=1}^n \widehat{z}_j$, we can construct a sequence of approximations $p_n(t)$ by

(4)
$$\begin{cases} dp_n(t) &= p_n(t)Qdt - \sigma^{-2}(t)\overline{\alpha}_n(t)p_n(t)A_n(t)dt + \sigma^{-2}(t)p_n(t)A_n(t)dy(t), \\ p_n(0) &= p_0, \end{cases}$$

where $\overline{\alpha}_n(t) = \langle p_n(t), \overline{z}_n\rangle$, $A_n(t) = \mathrm{diag}(\overline{z}_n^1 - \overline{\alpha}_n(t), \ldots, \overline{z}_n^m - \overline{\alpha}_n(t))$.

Let $e_n(t) = p_n(t) - p(t)$. Now, if we assume that $\{\widehat{z}_n\}$ is a stationary ergodic sequence with $E\widehat{z}_n = z$, uniformly bounded, and that the sequence $\{\widehat{z}_n\}$ is independent of $\alpha(\cdot)$ and the Brownian motion $w(\cdot)$, then we have:

Theorem 3. *As $n \to \infty$, $\sup_{0\le t\le T} E|e_n(t)|^2 \to 0$.*

Theorem 4.

$$\sup_{0\le t\le T} E|e_n^\kappa(t)|^2 = \begin{cases} o(1), & 0 < \kappa < 1/2, \\ O(1), & \kappa = 1/2, \end{cases} \quad \text{as } n \to \infty. \tag{5}$$

Theorem 5. (i) *For any positive integer $\ell > 1$,*

$$\sup_{0\le t\le T} E|e_n^\kappa|^{2\ell} = \begin{cases} o(1), & 0 < \kappa < 1/2, \\ O(1), & \kappa = 1/2, \end{cases} \quad \text{as } n \to \infty. \tag{6}$$

(ii) *As $n \to \infty$, $\sup_{0\le t\le T} E\exp(|e_n^{1/2}(t)|) = O(1)$.*

Similar results are obtained for the error bound estimates in the case when the generator Q is not known a priori. For more details, see [2].

2. Statistical estimation and change point detection in financial time series data

Given discrete feature of financial time series data, a natural statistical problem is the estimation of the states of the Markov chain $\alpha(t)$ when the stock price is observed at discrete time intervals $t = 1, 2, \ldots, n$, i.e.,

$$y_t \;=\; \mu_{\alpha(t)} + \sigma_{\alpha(t)}e_t, \quad e_t \sim \mathrm{N}(0,1), \tag{7}$$

In statistical literature, a model of the above form falls under the umbrella of a more generic class of models called *hidden Markov models* (HMMs). Within a Bayesian framework, we propose a recursive approach for parameter estimation, together with model selection strategies.

A case study of AT&T stock price data indicates that in the financial markets, a given pattern change is more gradual and takes time before its pattern is more sustainable. In this regard, the regime switching model captures this feature well;

our recursive algorithm can be a promising tool in identifying this type of regime change.

For more details of the estimation procedure and on the pros and cons of regime switching models, together with related research problems, see [3].

References

[1] Banerjee, A., Guo, X. and Wang., H. (2003): On the optimality of conditional expectation as a Bregman predictor. http://www.orie.cornell.edu/∼xinguo/papers.html.

[2] Guo, X. and Yin., G. (2003): Wonham filter with random parameters: (exponential) rate of convergence and error bounds. http://www.orie.cornell.edu/∼xinguo/papers.html.

[3] Guo, X.: A regime switching model: estimation, empirical evidence and change point detection. *Proceedings of SIAM-AMS-IMA Research Conference in Mathematical Finance* (to appear). http://www.orie.cornell.edu/∼xinguo/papers.html.

Estimation in Discretely Observed Diffusions: Two Examples of Using Small Δ-Optimality

M. Jacobsen

(joint work with M.L. Østerdal)

Consider a d-dimensional diffusion,

$$dX_t = b_\theta(X_t)\,dt + \sigma_\theta(X_t)\,dB_t$$

driven by a standard d-dimensional Brownian motion and with b_θ a d-dimensional drift function and σ_θ a $d \times d$-matrix valued diffusion function, where both b_θ and $C_\theta := \sigma_\theta \sigma_\theta^T$ are allowed to depend on an unknown p-dimensional parameter $\theta \in \Theta$. It is assumed that for all $\theta \in \Theta$, X has an invariant distribution μ_θ and is ergodic and suitably 'nice'. The task is then to estimate θ based on the observation of $X_{t_1}, \ldots, X_{t_n}$ where $0 < t_1 < \cdots < t_n$. With the likelihood function typically untractable, this may be done using *unbiased estimating* functions $g_{t,\theta}(x,y) = \left(g^k_{t,\theta}(x,y)\right)_{1 \le k \le p}$ where the $g_{t,\theta}$ are given in an explicit analytic form and unbiasedness means that (the μ_θ signifying that X_0 has distribution μ_θ)

$$\mathbb{E}_{\mu_\theta} g^k_{t,\theta'}(X_0, X_t) = 0 \quad \text{iff} \quad \theta = \theta'.$$

The estimator $\hat{\theta}_n$ for θ is now found by solving the equations $G^k_n(\theta) = 0\ (1 \le k \le p)$, where

$$G^k_n(\theta) = \sum_{i=1}^{n} g^k_{\Delta_i,\theta}\left(X_{t_{i-1}}, X_{t_i}\right),$$

writing $\Delta_i = t_i - t_{i-1}$. If $t_i = i\Delta$ for some $\Delta > 0$, it often holds that if θ is the true parameter value, then $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ converges in distribution as $n \to \infty$ to a Gaussian limit $N(0, var_{\Delta,\theta}(g))$. Good choices for $g_{t,\theta}$ are obtained by minimising the asymptotic covariance matrix $var_{\Delta,\theta}(g)$ in a suitable sense.

Some basic examples of unbiased estimating functions are (i) the *simple estimating functions*, see Kessler (2000),

$$g_{t,\theta}^{k}(x,y) = A_\theta h_\theta^k(x),$$

with $A_\theta = \sum_{i=1}^{d} b_\theta^i \partial_{x_i} h + \frac{1}{2}\sum_{i,j=1}^{d} C_\theta^{ij} \partial_{x_i x_j}^2 h$ the infinitesimal generator for X, and (ii) the *martingale estimating functions* introduced by Bibby and Sørensen (1995),

$$g_{t,\theta}^{k}(x,y) = \sum_{\ell=1}^{r} \phi_\theta^{k\ell}(x)\left(f_\theta^\ell(y) - \pi_{t,\theta} f_\theta^\ell(x)\right).$$

where $\pi_{t,\theta} f_\theta^\ell(x) = \mathbb{E}_\theta\left[f_\theta^\ell(X_t)\,|X_0 = x\right]$ is known explicitly. The number r is the dimension of the *base* $\left(f_\theta^\ell\right)$ for the estimating functions.

The concept of *small* Δ*-optimality* ($S\Delta$*-O*), Jacobsen (2001, 2002), aims at minimising $var_{\Delta,\theta}(g)$ as $\Delta \to 0$: (I) if C_θ does not depend on θ, typically

$$var_{\Delta,\theta}(g) = \Delta^{-1} v_{-1,\theta}(g) + O(1)$$

and there is a universal lower bound for $v_{-1,\theta}$ and g is SΔ-O if it achieves this lower bound. With g simple, this is possible only if X is reversible (automatic for $d = 1$); for g a martingale estimating function, SΔ-O may be obtained using a base of dimension $r = d$. (II) By contrast, if C_θ depends on all the p parameters,

$$var_{\Delta,\theta}(g) = \Delta^{-1} v_{-1,\theta}(g) + v_{0,\theta}(g) + O(\Delta)$$

and g is SΔ-O provided $v_{-1,\theta}(g) = 0$ (!) and $v_{0,\theta}(g)$ attains its universal lower bound. In this case (II) it is not possible to find simple g that are SΔ-O and martingale estimating functions that are SΔ-O require a base of dimension $d(d+3)/2$, e.g. $f^\ell(x)$ of the form x_i for $1 \le i \le d$ and $x_i x_j$ for $1 \le i \le j \le d$.

The purpose of the present study is to find SΔ-O estimation functions that combine 'simple' with 'martingale' estimating functions,

$$g_{t,\theta}^{k}(x,y) = tA_\theta h_\theta^k(x) + \sum_{i=1}^{d} \phi_\theta^{ki}(x)(y_i - \pi_{t,\theta} x_i) \tag{1}$$

for models where the first order conditional moments are known explicitly, i.e. typically b_θ an affine function of x. Such g *are S*Δ*-O provided there are functions* Φ_θ^k *such that*

$$\begin{aligned} h_\theta^k &= \Phi_\theta^k, \quad \phi_\theta^{ki} = \partial_{x_i}\Phi_\theta^k, \\ \partial_{x_i,x_j}^2 \Phi_\theta^k &= \sum_{i',j'=1}^{d} \left(\partial_{\theta_k} C_\theta^{i'j'}\right)\left(C_\theta^{i'i}\right)^{(-1)}\left(C_\theta^{j'j}\right)^{(-1)}. \end{aligned} \tag{2}$$

(Warning: for $d \ge 2$, a special structure for C and its inverse is of course required for Φ_θ^k that satisfy the last condition to exist at all!)

SΔ-O estimating functions of the form (1) are simpler in structure and may be easier to find than the pure martingale estimating functions needed for models of type (II). To illustrate this, two examples are considered:

Example 1. *A model suggested in the finance literature as a generalization of the Cox-Ingersoll-Ross process: let* $d = 1$, $p = 4$, *with* $b(x) = a + bx$, $\sigma(x) = \sigma x^{\gamma}$. *Here* $\pi_{t,\theta} x$ *is known but not* $\pi_{t,\theta} x^2$, *which makes it difficult to obtain* $S\Delta$*-O when estimating* σ^2 *and* γ. *But* $\left(g_\theta^1, g_\theta^2\right)$ *of the form (1) with the* h_θ^k *and* ϕ_θ^{ki} *as in (2) is* $S\Delta$*-O provided*

$$\partial_x \Phi_\theta^1(x) = x^{1-2\gamma}, \quad \partial_x \Phi_\theta^1(x) = x^{1-2\gamma}\left((1-2\gamma)\log x - 1\right).$$

Whether this works in practice, is currently being tested! For estimating a *and* b *also (type (I) model), one may combine with a* $S\Delta$*-O martingale estimating function with base* $f^1(x) = x$ *of dimension* 1.

Example 2. *Let* $d \geq 2$ *and consider the* d*-dimensional Ornstein-Uhlenbeck process with* $b(x) = bx$, $C(x) \equiv C$. *Here* $\theta = (b, C)$ *where* $b \in \mathbb{R}^{d\times d}$ *while* $C \in \mathbb{R}^{d\times d}$ *is positive definite. The transition function and therefore the likelihood function is known explicitly, but for* t_i *that are not equidistant becomes most unpleasant to maximize. Again, for estimating* C, *one may use (2) to find* $g^{i_0 j_0}$ *of the form (1) that are* $S\Delta$*-O, viz.*

$$\Phi_\theta^{i_0 j_0}(x) = \sum_{1\leq i\leq j\leq d} x_i x_j \left[C_{i_0 i}^{(-1)} C_{j_0 j}^{(-1)} + C_{i_0 j}^{(-1)} C_{j_0 i}^{(-1)}\right].$$

For estimating b, *combine with the* $S\Delta$*-O martingale estimating function for type (I) models with base* $f^i(x) = x_i$ *of dimension* d: *this still gives quite an unpleasant set of equations for estimating the* b_{ij}, *but it is certainly simpler than the likelihood equations.*

References

[1] Bibby, B.M., Sørensen, M. (1995): Martingale estimating functions for discretely observed diffuions. *Bernoulli* **1**, 17-39.

[2] Jacobsen, M. (2001): Discretely observed diffusions: classes of estimating functions and small Δ-optimality. *Scand. J. Statist.* **28**, 123-149.

[3] Jacobsen, M. (2002): Optimality and small Δ-optimality of martingale estimating functions. *Bernoulli* **8**, 643-668.

[4] Kessler, M. (2000): Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.* **27**, 65-82.

Lévy Copulas for General Lévy Processes

Jan Kallsen

(joint work with Peter Tankov)

Copulas constitute a popular tool to model the dependence of multivariate random variables e.g. in financial and actuarial applications. By virtue of Sklar's theorem, the dependence structure can be considered completely separately from the marginal laws. Various parametric families of Archimedean copulas allow for flexible and parsimoneous modelling (cf. e.g. Nelsen 1999).

In a continuous-time setup, Lévy processes are often applied successfully in order to describe in particular univariate data in finance and insurance. Parametric multivariate Lévy models, however, are scarce and typically very limited as far as the dependence between the components is concerned.

This suggests to transfer the notion of copulas to Lévy processes. In order to obtain a time-independent concept one works with the Lévy-Khinchine triplet. Since the correlation structure of the Brownian motion part is completely determined by the covariance matrix, it remains to consider the Lévy measure.

Tankov (2003) introduced a notion of copulas on the level of Lévy measures for multivariate processes with only positive jumps. In Kallsen and Tankov (2004) this concept of *Lévy copulas* is generalized to general Lévy processes X. Similarly as for random vectors, they are defined as tail integrals of measures with uniform marginals. An analogue of Sklar's theorem states that the Lévy measure can be recovered from the Lévy copula and the marginal Lévy measures. Conversely, any Lévy copula and any univariate Lévy measures can be combined to yield a Lévy measure. Archimedean Lévy copulas as e.g. the Clayton family are defined similarly as in the case of random vectors.

Finally, two limit theorems are discussed which show how to obtain the Lévy copula and also the Gaussian copula corresponding to the Brownian motion part of X as a limit of properly rescaled copulas of the random vectors X_t for $t \to 0$. The proof of these results relies on a characterization of weak convergence in terms of copula convergence by Lindner and Szimayer (2004).

References

[1] Kallsen, J. and Tankov, P. (2004): Lévy copulas for general Lévy processes. Preprint.
[2] Lindner, A. and Szimayer, A. (2004): A limit theorem for copulas. Preprint.
[3] Nelsen, R. (1999): *An Introduction to Copulas*. Springer, New York.
[4] Tankov, P. (2003): Dependence structure of spectrally positive Lévy processes. Preprint.

Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements

Siem Jan Koopman
(joint work with Borus Jungbacker and Eugenie Hol)

ABSTRACT

The increasing availability of financial market data at intraday frequencies has not only led to the development of improved volatility measurements but has also inspired research into their potential value as an information source for volatility forecasting. In this paper we explore the forecasting value of historical volatility (extracted from daily return series), of implied volatility (extracted from option pricing data) and of realised volatility (computed as the sum of squared high frequency returns within a day). First we consider unobserved components and long

memory models for realised volatility which is regarded as an accurate estimator of volatility. The predictive abilities of realised volatility models are compared with those of stochastic volatility models and generalised autoregressive conditional heteroskedasticity models for daily return series. These historical volatility models are extended to include realised and implied volatility measures as explanatory variables for volatility. The main focus is on forecasting the daily variability of the Standard & Poor's 100 stock index series for which trading data (tick by tick) of almost seven years is analysed. The forecast assessment is based on the hypothesis of whether a forecast model is outperformed by alternative models. In particular, we will use superior predictive ability tests to investigate the relative forecast performances of some models. Since volatilities are not observed, realised volatility is taken as a proxy for actual volatility and is used for computing the forecast error. A stationary bootstrap procedure is required for computing the test statistic and its p-value. The empirical results show convincingly that realised volatility models produce far more accurate volatility forecasts compared to models based on daily returns. Long memory models seem to provide the most accurate forecasts.

Description of paper

Modelling and forecasting volatility in financial markets has gained much interest in the financial and economic literature. The seminal paper of Engle (1982) has started the development of a large number of so-called historical volatility models in which a time-varying volatility process is extracted from financial returns data. Most volatility models can be regarded as variants of the generalised autoregressive conditional heteroskedasticity (GARCH) models of Bollerslev (1986), see Bollerslev et al. (1994) for a review. A rival class of volatility models is associated with the stochastic volatility (SV) model, see Taylor (1986) and Harvey et al. (1994). The overviews presented in Shephard (1996) and Ghysels et al. (1996) provide an excellent introduction to historical volatility models. A more recent review of volatility models together with an assessment of their forecasting performances is given by Poon and Granger (2003).

Both GARCH and SV models are regularly used for the analysis of daily, weekly and monthly returns. ¿From a theoretical perspective these models can also be applied to returns data measured at higher frequencies (intraday). However, it is learned from empirical studies that these models can not accomodate all information in high frequency returns. The initial work of Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001) show that realised volatility (a daily volatility measure) as computed by the cumulative sum of squared intraday returns is less subject to measurement error and therefore less noisy. This empirical fact is supported by the theory that the measurement noise contained in daily squared returns prevents the observation of the volatility process while it is reduced as the sampling frequency of the return series from which volatility is calculated is increased, see Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2001, 2002). These results also justify the earlier work of French et al. (1987), amongst others. Andersen and Bollerslev (1998)

show that daily forecasts of exchange rates based on GARCH models, when evaluated against realised volatility, are far more accurate than had been previously assumed. These findings were subsequently confirmed with regards to stock index data by Blair et al. (2001) and Hansen and Lunde (2003) who examined the predictive accuracy of volatility forecasts based on GARCH models.

Volatility can be extracted from returns data but it can also be derived from option pricing data in combination with an option pricing model. Early empirical studies have indicated that implied volatility, when compared with historical standard deviations, can be regarded as a good predictor of future volatility. Implied volatility is often referred to as the market's volatility forecast and is said to be forward looking as opposed to historical based methods which are by definition backward looking. Recent study by Blair et al. (2001) shows that accurate volatility forecasts for returns on stock indices are often based on implied volatility. Moreover, their research strongly suggests that daily returns contain little or no incremental information about future volatility.

In this paper we investigate the potential gains of different measures of volatility and different ways of modelling these data for the purpose of volatility forecasting. For example, it is suggested to incorporate realised volatility as an explanatory variable in the variance equation of a daily GARCH model. They found a considerable improvement in the forecasting performance in this way. Another possible explanatory variable for volatility is implied volatility. We will explore this option further by incorporating such explanatory variables in both GARCH and SV models.

Realised volatility can also be modelled directly which is reminiscent of the methods adopted for monthly volatility in a number of earlier studies. The forecasting performance of realised volatility models has been studied, amongst others, by Andersen, Bollerslev, Diebold and Ebens (2001) and Barndorff-Nielsen and Shephard (2004). In the first paper, it is stressed that long memory features are present in the logarithms of realised volatility and that the autoregressive fractionally integrated moving average (ARFIMA-RV) model is effective in empirical modelling. The second paper builds on Barndorff-Nielsen and Shephard (2002) where volatility is represented as a continuous time series process, the sum of independent Lévy driven Ornstein-Uhlenbeck (OU) processes. This approach forms the basis of an unobserved components (UC-RV) model for realised volatility that consists of independent ARMA components with restricted parameters.

The empirical investigation is for the Standard & Poor's 100 (S&P 100) stock index series over the period 6 January 1997 to 14 November 2003 with 1725 trading days. Opening and closure prices for all trading days in the sample are available in this period together with all price quotes within the days (tick by tick). Further we have obtained the S&P 100 implied volatility index from the Chicago Board Options Exchange Market volatility index (VIX) which is known to be a highly liquid options market. The forecasting performance of various volatility models for the last 525 days of the data set is the focus of the empirical study. We compare the forecasts of ARFIMA-RV, UC-RV, SV and GARCH volatility models;

the latter two models are considered with and without explanatory variables. The forecasts are generated by a rolling-window of 1200 observations through the last 525 daily observations. Forecast comparison is based on four different loss functions including the mean squared error and the mean absolute error statistics. The fact that a particular loss criterion is smallest for a particular model does not provide any information about its forecast superiority in other samples of the data set and in future samples of the data. The results in White (2000) and the important refinements in Hansen (2001) constitute a framework that constructs a formal test for superior prediction ability (SPA) of a benchmark or base model relative to a set of rival models. Since volatility can never be observed, realised volatility is taken as a proxy for actual volatility and used for determining the forecast error. This may introduce inconsistencies in the ranking of forecast models but it is argued that the occurrence of such inconsistencies are unlikely in our study. The method of computing the SPA test statistic and its p-value requires bootstrap samples obtained by, for example, the stationary bootstrap procedure of Politis and Romano (1994). The construction of the test and some details of implementation are discussed.

The findings of this extensive empirical study are presented by reporting a selection of the most interesting results. The maximum likelihood estimates for the coefficients of the considered models are reported for the full sample. Although these estimates are not used for forecasting since all models are re-estimated for each rolling window sample (starting from 17 October 2001), the reported estimation results provide insights about the S&P 100 data set and the effectiveness of models to capture volatility information from the data. A selection of the forecasting results is also presented but most attention is paid to the SPA results. It has become clear that the realised volatility models are overwhelmingly superior and therefore making comparisons between, say, GARCH and ARFIMA-RV is not useful. We therefore concentrate on the comparison of models within the two classes of realised volatility models and historical volatility models. It will be concluded that both the ARFIMA-RV and the SV model with realised volatility as the explanatory variable are superior within their classes for the forecasting of S&P 100 volatility. To get some insight in how forecasts evolve over time in our study, in Figure 1 we present one-step ahead forecasts for the S&P 100 volatility between 9 September 2002 and 18 November 2002 (51 trading days).

References

[1] Andersen, T.G. and Bollerslev, T. (1998): Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Internat. Econom. Rev.* **39**, 885–905.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X., and Ebens, H. (2001): The distribution of realized stock return volatility. *Journal of Financial Economics* **61**, 43–76.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001): The distribution of exchange rate volatility. *J. Amer. Statist. Assoc.* **96**, 42–55.

[4] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian OU based models and some of their uses in financial economics (with discussion). *J. R. Stat. Soc. Ser. B* **63**, 167–241.

[5] Barndorff-Nielsen, O.E. and Shephard, N. (2002): Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B* **64**, 253–280.

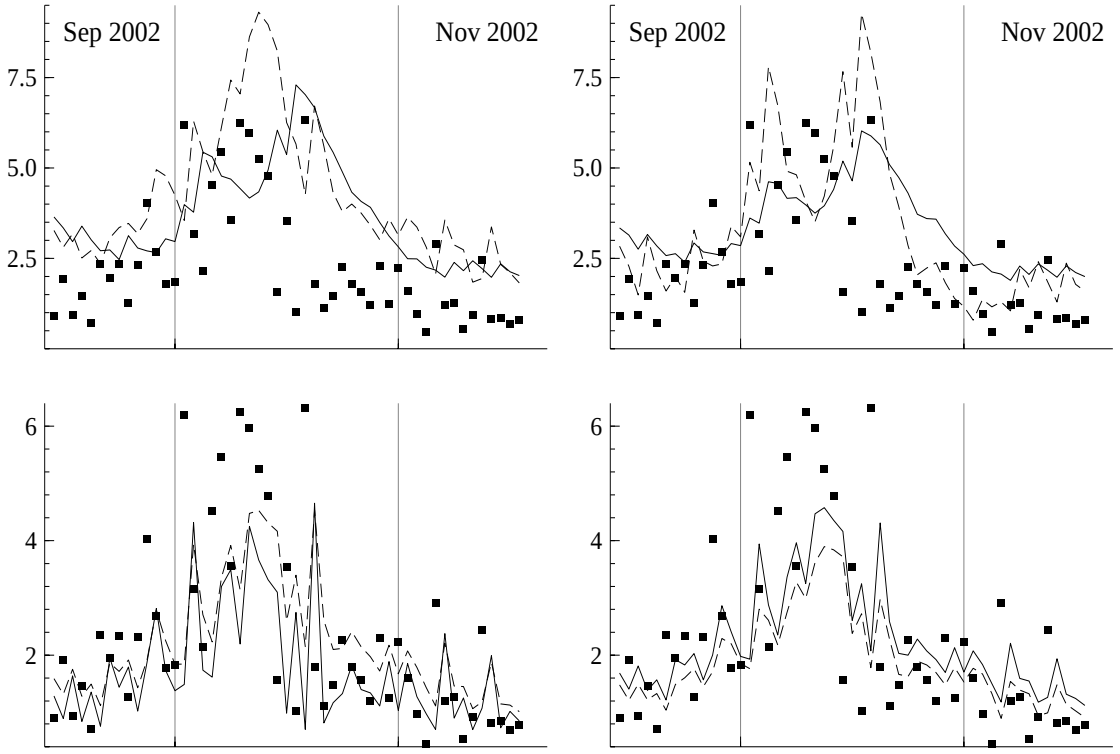

FIGURE 1. Realised volatility (as dots) and one-day ahead volatility forecasts from (i) GARCH (solid) and GARCH with RV (dashed), (ii) SV (solid) and SV with RV (dashed), (iii) UC-RV1 (solid) and UC-RV2 (dashed) and (iv) ARFIMA-RV (solid) and log ARFIMA-RV (dashed) models for the period between 9 September 2002 and 18 November 2002 (day 225 to 275).

[6] Barndorff-Nielsen, O.E. and Shephard, N. (2004): Measuring and forecasting financial variability using realised variance. In: Harvey, A.C., Koopman, S.J. and Shephard, N. (Eds.), *State Space and Unobserved Components Models: Theory and Applications*, Cambridge University Press, Cambridge.

[7] Beckers, S. (1981): Standard deviations implied in options prices as predictors of future stock price variability. *Journal of Banking and Finance* **5**, 363–381.

[8] Blair, B.J., Poon, S.H. and Taylor, S.J. (2001): Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high frequency returns. *J. Econometrics* **105**, 5–26.

[9] Bollerslev, T. (1986): Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[10] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994): ARCH models. In: Engle, R.F. and McFadden, D.L. (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2959–3038, Elsevier Science, Amsterdam.

[11] Chiras, D.P. and Manaster, S. (1978): The information content of options prices and a test of market efficiency. *Journal of Financial Economics* **6**, 213–234.

[12] Christensen, B.J. and Prabhala, N.R. (1998): The relation between implied and realized volatility. *Journal of Financial Economics* **50**, 125–150.

[13] Engle, R.F. (1982): Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom Inflation. *Econometrica* **50**, 987–1007.
[14] Fleming, J. (1998): The quality of market volatility forecast implied by S&P 100 index option prices. *Journal of Empirical Finance* **5**, 317–345.
[15] French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987): Expected stock returns and volatility. *Journal of Financial Economics* **19**, 3–29.
[16] Ghysels, E., Harvey, A.C. and Renault, E. (1996): Stochastic Volatility. In: Maddala, G.S. and Rao, C.R. (Eds.), *Handbook of Statistics*, Volume 14, Statistical Methods in Finance, pp. 119–191, North-Holland, Amsterdam.
[17] Giot, P. (2003). The information content of implied volatility in agricultural commodity markets. *Journal of Futures Markets* **23**, 441–454.
[18] Hansen, P.R. (2001): A test for superior predictive ability. Discussion Paper, Brown University Working Paper 2001-06.
[19] Hansen, P.R. and Lunde, A. (2003): A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Discussion Paper, Brown University Working Paper.
[20] Harvey, A.C., Ruiz, E. and Shephard, N. (1994): Multivariate Stochastic Variance models. *Rev. Econom. Stud.* **61**, 247–264.
[21] Latané, H.A. and Rendleman, R.J. (1976): Standard deviations of stock price ratios implied in option prices. *J. Finance* **31**, 369–381.
[22] Martens, M. (2001): Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance* **20**, 1–23.
[23] Politis, D.N. and Romano, J.P. (1994): The stationary bootstrap. *J. Amer. Statist. Assoc.* **89**, 1303–1313.
[24] Poon, S.H. Granger, C. (2003): Forecasting volatility in financial markets: a review. *Journal of Economic Literature* **41**, 478–539.
[25] Poon, S.H. and Taylor, S.J. (1992): Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance* **16**, 37–59.
[26] Shephard, N. (1996): Statistical aspects of ARCH and Stochastic Volatility. In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, Number 65 in Monographs on Statistics and Applied Probability, pp. 1–67, Chapman and Hall, London.
[27] Taylor, S.J. (1986): *Modelling financial time series*. Wiley, Chichester.
[28] White, H. (2000): A reality check for data snooping. *Econometrica* **68**, 1097–1126.

Leroux's method for General Hidden Markov Models and Stochastic Volatility Models

Catherine Larédo
(joint work with Valentine Genon-Catalot)

Parametric inference for Hidden Markov Models (H.M.M.) has been widely investigated, especially in the last decade. The observed process (Z_n) is modelled via an unobserved Markov chain (U_n). When studying the statistical properties of H.M.M.s, a difficulty arises since the exact likelihood cannot be explicitly calculated. As a consequence, many authors have studied approximations by means of numerical and simulation techniques (see for instance Del Moral et al., 2001; Pitt and Shephard, 1999; Durbin and Koopman, 1997).

The theoretical study of the exact maximum likelihood has been investigated for finite state space (see Leroux, 1992; Bickel and Ritov, 1996; Bickel et al., 1998) and for compact state space (see Jensen and Petersen, 1999; Douc and Matias,

2001). In previous papers (Genon-Catalot et al., 1998, 1999, 2000, 2003), we have investigated some statistical properties of discretely observed Stochastic Volatility models (S.V.). When the sampling interval is fixed, stochastic volatility models are H.M.M.s, for which the hidden chain has non-compact state space.

We extend here a method of Leroux (1992) to study the likelihood and related contrast processes for general hidden Markov models. We define the entropy associated to these models and characterize the limit of the loglikelihood and related processes, under specific assumptions.

Generic examples of such processes are obtained setting $Z_n = G(U_n, \epsilon_n)$, where $G : \mathcal{U} \times \mathbb{R}^l \to \mathbb{R}$ is a known function, (U_n) is a strictly stationary Markov chain on $\mathcal{U}$, and (ϵ_n) a sequence of i.i.d random variables on $\mathbb{R}^l$, independent of (U_n) with known density. These methods are applied to the Kalman filter ($G(u, v) = u + v$ and (U_n) is AR(1)), to stochastic volatility models ($G(u, v) = \sqrt{u} \times v$ and (U_n) a Markov chain in $\mathbb{R}^2$), and to the multiplicative explicit filter proposed by Genon-Catalot and Kessler (2004).

This research was supported in part by Dynstoch European Network.

References

[1] Bickel, P.J. and Ritov, V. (1996): Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli* **2** (3), 199-228.

[2] Bickel, P.J., Ritov, Y. and Ryden, T. (1998): Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614-1635.

[3] Del Moral, P., Jacod, J. and Protter, Ph. (2001): The Monte-Carlo method for filtering with discrete time observations. *Probab. Theory Related Fields* **120** (3), 346-368.

[4] Douc, P. and Matias, L. (2001): Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7** (3), 381-420.

[5] Durbin, J. and Koopman, S.J. (1997): Monte-Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84** (3), 669-684.

[6] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998): Limit theorems for discretely observed stochastic volatility models. *Bernoulli* **4** (3), 283-303.

[7] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999): Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5** (5), 855-872.

[8] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000): Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** (6), 1051-1079.

[9] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2003): Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scand. J. Statist.* **30** (2), 297-316.

[10] Genon-Catalot, V. and Kessler, M. (2004): Random scale perturbation of an AR(1) and its properties as a non-linear explicit filter. *Bernoulli* (to appear).

[11] Jensen, J.L. and Petersen, N.V. (1999): Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27** (2), 514-535.

[12] Leroux, B.G. (1992): Maximum likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127-143.

[13] Pitt, M.K. and Shephard, N. (1999): Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** (446), 590-599.

A Continuous Time GARCH(1,1) Process

Alexander Lindner
(joint work with Claudia Klüppelberg and Ross Maller)

1. Introduction

Discrete time GARCH(1,1) models are commonly used to model financial time series like asset prices and exchange rates. They capture many of the so-called stylized features such as heavy tails and uncorrelatedness without being independent. The latter is e.g. manifested in the nonzero autocorrelation of the squared sequence. Various attempts have been made to capture these features in a *continuous time* model such as diffusion approximations (see e.g. Duan (1996) or Nelson (1990)) and other stochastic volatility models, as e.g. in Anh et al. (2002) or Barndorff-Nielsen and Shephard (2001). These models have in common that they are driven by two random processes. Here, we propose a continuous time GARCH(1,1) model with only one source of randomness, capturing the stylized features by the dependence structure alone. The talk is based on results of Klüppelberg et al. (2004).

2. From discrete to continuous GARCH

The discrete time GARCH(1,1) process is given by

$$Y_n = \sigma_n \varepsilon_n,\ n \in \mathbb{N}_0,\ \text{where } \sigma_n^2 = \beta + \lambda Y_{n-1}^2 + \delta \sigma_{n-1}^2$$

with constants $\beta, \delta > 0$, $\lambda \geq 0$ and an iid sequence $(\varepsilon_n)_{n\in\mathbb{N}_0}$, independent of σ_0^2. Then σ_n can be written as

$$\begin{aligned}\sigma_n^2 &= \beta \sum_{i=0}^{n-1} \prod_{j=i+1}^{n-1} (\delta + \lambda \varepsilon_j^2) + \sigma_0^2 \prod_{j=0}^{n-1} (\delta + \lambda \varepsilon_j^2) \\ (1) &= \beta \int_0^n \exp\left(\sum_{j=\lfloor s \rfloor + 1}^{n-1} \log(\delta + \lambda \varepsilon_j^2) \right) ds + \sigma_0^2 \exp\left(\sum_{j=0}^{n-1} \log(\delta + \lambda \varepsilon_j^2) \right),\ n \in \mathbb{N}.\end{aligned}$$

This suggests, in continuous time, to replace the noise variables ε_n by the increments $\Delta L_t = L_t - L_{t-}$ of a Lévy process $(L_t)_{t\geq 0}$. Keep $\beta, \delta > 0$, $\lambda \geq 0$, and define the process $(X_t)_{t\geq 0}$ by

$$X_t = -t \log \delta - \sum_{0<s\leq t} \log(1 + \frac{\lambda}{\delta} (\Delta L_s)^2), \quad t \geq 0.$$

Then, in analogy with (1), for a finite random variable $\sigma_0 \geq 0$, independent of $(L_t)_{t\geq 0}$, define the left-continuous *volatility process* $(\sigma_t)_{t\geq 0}$ by

$$\sigma_t^2 = \left(\beta \int_0^t e^{X_s} ds + \sigma_0^2 \right) e^{-X_t}, \quad t \geq 0, \tag{2}$$

and the *continuous time GARCH* (*"COGARCH"*) process $(G_t)_{t\geq 0}$ as the cádlág process satisfying

$$dG_t = \sigma_t \, dL_t, \quad t \geq 0, \quad G_0 = 0. \tag{3}$$

Thus G jumps as the same times as L does, and has jumps of size $\Delta G_t = \sigma_t \Delta L_t$.

3. Properties of the model

In this section we give some of the properties of model (2), (3). First, we note that $(X_t)_{t\geq 0}$ defines a spectrally negative Lévy process of bounded variation with drift $\gamma_{X,0} = -\log \delta$. For the volatility process, it holds:

Proposition 1. *The process $(\sigma_t^2)_{t\geq 0}$ satisfies the stochastic differential equation*

$$d\sigma_{t+}^2 = \beta dt + \sigma_t^2 e^{X_{t-}} d(e^{-X_t}), \quad t > 0,$$

and we have

$$\sigma_t^2 = \beta t + \log \delta \int_0^t \sigma_s^2 ds + (\lambda/\delta) \sum_{0<s<t} \sigma_s^2 (\Delta L_s)^2 + \sigma_0^2, \quad t \geq 0.$$

Denote by Π_L the Lévy measure of $(L_t)_{t\geq 0}$, and assume that it is nonzero. Then we can give necessary and sufficient conditions for strict stationarity of the volatility process $(\sigma_t^2)_{t\geq 0}$.

Theorem 2. *The volatility process $(\sigma_t^2)_{t\geq 0}$ is a time homogeneous Markov process. The random variable σ_0 can be chosen such that $(\sigma_t^2)_{t\geq 0}$ is strictly stationary, if and only if*

$$\int_{-\infty}^{\infty} \log(1 + \frac{\lambda}{\delta} y^2) \Pi_L(dy) < -\log \delta.$$

In that case, for any $k \in \mathbb{N}$, σ_t^2 has finite k'th moment if and only if $EL_1^{2k} < \infty$ and

$$\Psi(k) := k \log \delta + \int_{-\infty}^{\infty} \left((1 + \frac{\lambda}{\delta} y^2)^k - 1 \right) \Pi_L(dy) < 0.$$

If $EL_1^4 < \infty$ and $\Psi(2) < 0$, then the autocovariance function of σ_t^2 decreases exponentially with the lag.

Using Theorem 2, it can be shown that for any Lévy process $(L_t)_{t\geq 0}$ the stationary version of the volatility process $(\sigma_t^2)_{t\geq 0}$ has certain infinite moments. In that sense, the volatility process is heavy tailed. For the COGARCH process $(G_t)_{t\geq 0}$ itself, we have:

Theorem 3. *Assume $(\sigma_t^2)_{t\geq 0}$ is the strictly stationary volatility process. Then the integrated GARCH(1,1) process $(G_t)_{t\geq 0}$ has stationary increments. Assume further that $EL_1^8 < \infty$ and $\Psi(4) < 0$, that $(L_t)_{t\geq 0}$ is a quadratic pure jump process*

(i.e. has no Gaussian component) and that $EL_1 = 0$, $\int_{\mathbb{R}} y^3 \Pi_L(dy) = 0$. *Let* $r > 0$ *be fixed. Then there exists a positive constant* C_r *such that for any* $t \geq 0$ *and* $h \geq r$*:*

$$\begin{aligned} \mathrm{Cov}(G_{t+r} - G_t, G_{t+r+h} - G_{t+h}) &= 0, \\ \mathrm{Cov}((G_{t+r} - G_t)^2, (G_{t+r+h} - G_{t+h})^2) &= C_r e^{h\Psi(1)}. \end{aligned}$$

Theorem 3 shows, in analogy with the discrete time GARCH model, that the increments of $(G_t)_{t\geq 0}$ are uncorrelated, but that their squares are not.

References

[1] Anh, V.V., Heyde, C.C. and Leonenko, N.N. (2002): Dynamic models of long–memory processes driven by Lévy noise. *J. Appl. Prob.* **39**, 730-747.

[2] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian Ornstein–Uhlenbeckbased models and some of their uses in financial economics (with discussion). *J. R. Stat. Soc. Ser. B* **63**, 1-42.

[3] Duan, J.C. (1996): A unified theory of option pricing under stochastic volatility – from GARCH to diffusion. Working paper, available at http://www.rotman.utoronto.ca/∼jcduan/.

[4] Klüppelberg, C., Lindner, A. and Maller, R. (2004): A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour. *J. Appl. Prob.* **41** (3) (to appear).

[5] Nelson, D.B. (1990): ARCH models as diffusion approximations. *J. Econometrics* **45**, 7-38.

The Large-Sample Distribution of the Sharpe Ratio

R. A. Maller

1. Introduction and Summary

In the Markowitz efficient portfolio paradigm, we maximise the expected return on a portfolio of assets for a given level of "risk", as measured by the standard deviation of the portfolio return. Among the set of portfolios derived in this way, we can select the one which has the *maximum* return to risk tradeoff, as measured by the ratio of expected return (excess over the risk-free rate) to standard deviation of return, that is, the portfolio with maximum *Sharpe ratio*. This portfolio has desirable optimality properties and is important both for purposes of allocation of resources and for the performance evaluation of portfolios.

Given sample estimates of the mean vector and covariance matrix of the excess returns which are asymptotically normally distributed, we might expect to get asymptotic normality of the maximised Sharpe ratio. But because the maximisation procedure means that we are not dealing with just a simple ratio of mean to standard deviation, this is not true in general, though it is in some cases. We are able to give a complete description of the large-sample behaviour of the Sharpe ratio for

a wide class of portfolios, and (when there are restrictions on short-selling), a partial solution which still covers some useful situations – but we merely summarise the results here. (For details see Maller, 2004.)

Although not always asymptotically normal, the Sharpe ratio is in the domain of attraction of the normal in the cases we study, so the usual kinds of statistical analyses which are applied to the Sharpe ratio are valid, at least in large samples.

2. Background – the Markowitz paradigm

We are given a d–vector $\tilde{\mu}$ of expected asset returns and an associated $d \times d$ positive definite covariance matrix Σ. The excess returns are:

$$\mu = \tilde{\mu} - ri,$$

where r is the risk-free rate and i is a d-vector each of whose elements is 1. The optimisation problem is to choose a d–vector x of asset weights such that the portfolio standard deviation

$$\sigma_p = \sqrt{x^T \Sigma x}$$

is minimised for a specified expected return,

$$\mu_p = x^T \mu$$

(or, equivalently, μ_p is maximised for a specified level of risk, σ_p.) ¿From pairs (μ_p, σ_p) constructed in this way we can trace out an efficient frontier, representing portfolios whose return/risk tradeoff is optimal in the mean–variance sense.

The vector x will be further restricted to a class C, say, which must include $\{i^T x = 1\}$ (the "total allocation constraint"). We only consider C of the form

$$C_A = \mathbb{R}^d \cap \left\{x : i^T x = 1\right\},$$

or

$$C_+ = \mathbb{R}^d \cap \left\{x : x \geq 0, i^T x = 1\right\}.$$

In C_A, the components of x may be negative – short sales of assets are allowed. In C_+, the components of x are non-negative – short sales of assets are not allowed.

The *Sharpe ratio* (SR) of a portfolio (or a single asset) is its expected (excess) return divided by its standard deviation. We ask for the portfolio with the *maximum SR* along the efficient frontier. This is the portfolio with the highest return/risk tradeoff achievable from the assets: the *optimal risky portfolio*.

The basics of the optimisation problem have been well understood since the seminal work of Markowitz (1952, 1991). When short sales are allowed and

$$i^T \Sigma^{-1} \mu > 0, \tag{1}$$

the optimal risky portfolio is located at the point of tangency of a line from the origin (since we have excess returns) to the efficient frontier. But when

$$i^T \Sigma^{-1} \mu < 0 \tag{2}$$

following this method gives a portfolio with the *minimum* SR. Maller and Turkington (2002) showed how to find the portfolio with the maximum SR achievable in this case. The case in (2) is not by any means pathological.

Suppose first that $C = C_A$, so we wish to maximise the function

$$f(x) = \frac{x^T \mu}{\sqrt{x^T \Sigma x}}$$

for variations in x, under the sole constraint that $i^T x = 1$. An easy analysis gives

$$\sup_{i^T x = 1} |f(x)| \leq \sqrt{\mu^T \Sigma^{-1} \mu},$$

and, supposing now that (1) holds, we get

$$\sup_{i^T x = 1} f(x) = +\sqrt{\mu^T \Sigma^{-1} \mu}.$$

This is achieved for the allocation

$$x_{\max} = \frac{\Sigma^{-1} \mu}{i^T \Sigma^{-1} \mu}.$$

This is a textbook solution (e.g., Elton and Gruber, 1995).

By contrast, when (2) holds, the maximum of $f(x)$ occurs at infinite values of x, having value (Maller and Turkington, 2002)

$$+\sqrt{\mu^T \Sigma^{-1} \mu - (i^T \Sigma^{-1} \mu)^2 / i^T \Sigma^{-1} i}. \tag{3}$$

The term under the square root sign of (3) is non-negative, and is zero if and only if μ is proportional to i, i.e., if the excess returns of all N assets are equal. Thus we can expect to achieve a positive SR regardless of the value of $i^T \Sigma^{-1} \mu$.

3. Sample Statistics

In practise we will have estimates

$$\widehat{\mu}_n = (\widehat{\mu}_{n1}, \ldots, \widehat{\mu}_{nd}),$$

of the mean (excess) returns calculated from a sample of size n, and an estimate $\widehat{\Sigma}_n = (\widehat{\sigma}_{nij})$ of a positive definite matrix. We carry out a Markowitz (1952) optimal allocation of funds among the securities. For our analysis $\widehat{\Sigma}_n$ need not be related to the covariance matrix of the returns, though in practise it usually is. (We keep $d \geq 2$ from now on.)

The sample Sharpe ratio is defined as

$$\widehat{SR}_n = \sup_{x \in C} \left(\frac{x^T \widehat{\mu}_n}{\sqrt{x^T \widehat{\Sigma}_n x}} \right).$$

Note that we maximise the ratio with regard to sign, as advocated, eg. by Sharpe (1994), rather than taking the absolute value or square, as is occasionally done. The statistic $\widehat{SR}_n$ provides one way of summarising the risk/return tradeoff of the optimal portfolio. Comparisons between portfolios can be made by comparing their Sharpe ratios. So it's natural to ask how the precision of estimation of $\widehat{\mu}_n$ and $\widehat{\Sigma}_n$ is transferred to $\widehat{SR}_n$.

It is relatively easy to show that the estimator is consistent for any choice of C if based on consistent estimators of μ and Σ. (This is not quite trivial to prove since the supremum can occur at infinite values of x. But we omit details here.) We might further guess that asymptotic normality of $\widehat{\mu}_n$ and $\widehat{\Sigma}_n$, i.e., assuming

$$\sqrt{n}\left(\widehat{\mu}_n - \mu, \mathrm{vech}(\widehat{\Sigma}_n - \Sigma)\right) \xrightarrow{D} N(0, \zeta),$$

where ζ is a positive definite matrix, will imply asymptotic normality of $\widehat{SR}_n$. This is not the case in general, even for the class C_A, though it is sometimes. Specifically, when $C = C_A = \mathbb{R}^d \cap \{i^T x = 1\}$, and $\mu \neq 0$, then we can show that, as $n \to \infty$,

$$\sqrt{n}(\widehat{SR}_n - SR) \xrightarrow{D} N(0, \sigma^2_{C_A})$$

for some $\sigma^2_{C_A} > 1$ (depending on μ, Σ, and ζ), where SR is the population Sharpe ratio. When $\mu = 0$ the limit of $\sqrt{n}(\widehat{SR}_n - SR)$ can be explicitly worked out for C_A, and is not normal (in fact it is a non-negative random variable and depends on the unknown Σ). Heuristically, what happens when $\mu = 0$ is that the sample estimate $\widehat{\mu}$ can oscillate around zero, alternately bringing into play the situations in (1) and (2).

Finally, when $C = C_+$, and further assuming that Σ is diagonal, the limit of $\sqrt{n}(\widehat{SR}_n - SR)$ can again be worked out, and again is not normal in all situations. Especially, the case $\mu = 0$ leads to non-normality, but so do some other values of μ, in the C_+ case.

Although not always asymptotically normal, the Sharpe ratio is in the domain of attraction of the normal in the cases we study. Details of these results, together with some practical implications of the analyses, are in Maller (2004).

References

[1] Elton, E.J. and Gruber, M.J. (1995): *Modern Portfolio Theory and Investment Analysis*, 5th Ed., Wiley, New York.

[2] Maller, R.A. (2004): The Large-Sample Distribution of the Sharpe Ratio. Preprint.

[3] Maller, R.A. and Turkington, D.A. (2002): New light on the portfolio allocation problem. *Math. Methods Oper. Res.* **56** (3), 501-511.

[4] Markowitz, H. (1952): Portfolio Selection. *J. Finance* **7**, 77-91.

[5] Markowitz, H. (1991): *Portfolio Selection: Efficient Diversification of Investment*. Blackwell, Cambridge, Mass.

[6] Sharpe, W.F. (1994): The Sharpe Ratio. *The Journal of Portfolio Management* (Fall), 49-58.

The t Copula and Related Copulas

Alexander J. McNeil
(joint work with Stefano Demarta)

The t copula (see for example Embrechts et al. (2001) or Fang and Fang (2002)) can be thought of as representing the dependence structure implicit in a multivariate t distribution. It is a model which has received much recent attention, particularly in the context of modelling multivariate financial return data (for example daily relative or logarithmic price changes on a number stocks). A number of recent papers such as Mashal and Zeevi (2002) and Breymann et al. (2003) have shown that the empirical fit of the t copula is generally superior to that of the so-called Gaussian copula, the dependence structure of the multivariate normal distribution. One reason for this is the ability of the t copula to capture better the phenomenon of dependent extreme values, which is often observed in financial return data.

The objective of this talk is to bring together what is known about the t copula, particularly with regard to its extremal properties, to present some extensions of the t copula, and to describe copulas that are related to the t copula through extreme value theory.

The two new extensions of the t copula are known respectively as the skewed (or asymmetric) t copula and the grouped t copula. Both are constructed by generalising the Gaussian mixture construction of the multivariate t distribution. The skewed t copula is obtained as the copula of a mean-variance mixture of multivariate normals using an inverse gamma mixing distribution, and is a member of the family of generalised hyperbolic copulas. The grouped t copula is the copula of a distribution that is obtained by mixing different subvectors of a Gaussian vector with different inverse-gamma distributed mixing variables, all of which are perfectly positively dependent. Both copulas are interesting for applied work as they suggest ways of incoporating more heterogeneity into the modelling of tail-dependent risks.

The two new copulas arising from extreme value theory are known as the t extreme value (t-EV) copula and the t lower tail limit copula. The former is the limiting copula of componentwise maxima of t distributed random vectors; the latter is the limiting copula of bivariate observations from a t distribution that are conditioned to lie below some joint threshold that is progressively lowered. Both these copulas may be approximated for practical purposes by simpler, better-known copulas, these being the Gumbel and Clayton copulas respectively. They are thus of more theoretical than practical interest.

The finding that the Clayton copula may successfully approximate the t lower tail copula provides some support for the empirical finding by Breymann et al. (2003) that bivariate exchange rate return data are consistent with a t copula as overall model and a Clayton copula for the most extreme negative returns at many different sampling frequencies.

References

[1] Breymann, W., Dias, A. and Embrechts, P. (2003): Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3**, 1-14.

[2] Embrechts, P., McNeil, A. and Straumann, D. (2001): Correlation and dependency in risk management: properties and pitfalls. In: *Risk Management: Value at Risk and Beyond*, M. Dempster and H. Moffatt, eds., Cambridge University Press.

[3] Fang, H. and Fang, K. (2002): The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* **82**, 1-16.

[4] Mashal, R. and Zeevi, A. (2002): Beyond correlation: extreme co–movements between financial assets. Unpublished, Columbia University.

Stable Limits for GARCH Parameter Estimation

Thomas Mikosch

(joint work with Daniel Straumann)

This talk is based on joint work with Daniel Straumann (ETH Zurich); see [9].

We consider a GARCH(p, q) (generalized autoregressive conditionally heteroscedactic process of order (p, q)) given by the equations

$$X_t = \sigma_t Z_t\,, \quad X_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2\,, \quad t \in \mathbb{Z}\,, \tag{1}$$

for non-negative coefficients α_j and β_k. This process is one of the standard models for returns of speculative prices. It is a well-known empirical fact that returns are heavy-tailed. The GARCH model allows for modeling those tails either by heavy tails of the σ- or Z-processes.

Regular variation and stochastic recurrence equations

A theoretical means to describe heavy tails in the univariate and multivariate cases is *regular variaton*: a random vector $\mathbf{X} \in \mathbb{R}^d$ and its distribution are *regularly varying with index* $\alpha \geq 0$ if there exists $\mathbf{\Theta} \in \mathbb{S}^{d-1}$ such that for any $t > 0$, $S \subset \mathbb{S}^{d-1}$ with $P(\mathbf{\Theta} \in \partial S) = 0$,

$$\lim_{x\to\infty} \frac{P\left(|\mathbf{X}| > tx\,, \widetilde{\mathbf{X}} \in S\right)}{P(|\mathbf{X}| > x)} = t^{-\alpha}\, P(\mathbf{\Theta} \in S)\,,$$

where $\widetilde{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$. The limiting distribution $P_{\mathbf{\Theta}}$ is the *spectral measure* of $\mathbf{X}$.

The notion of multivariate regular variation is a very natural one. It is used as necessary and sufficient domain of attraction condition for partial sums of iid random vectors with infinite variance stable weak limits ([12]) and for componentwise maxima of iid random vectors ([11]). Moreover, under mild conditions on the sequence of iid non-negative random vectors $((\mathbf{A}_i\,, \mathbf{B}_i))$, the stationary solution $(\mathbf{X}_t)$ to the stochastic recurrence equation

$$\mathbf{X}_t = \mathbf{A}_t\, \mathbf{X}_{t-1} + \mathbf{B}_t\,, \quad t \in \mathbb{Z}\,. \tag{2}$$

is regularly varying in the sense that

$$P\left((\widetilde{\mathbf{x}}\,,\mathbf{X})>x\right)\sim c(\widetilde{\mathbf{x}})\,x^{-\alpha}\,,\quad x\to\infty,,\quad \widetilde{\mathbf{x}}\in\mathbb{S}^{d-1}\,, \tag{3}$$

for some $\alpha>0$ ([7]). It is not difficult to verify that the vector

$$\mathbf{X}_t=\left(\sigma_{t+1}^2,\ldots,\sigma_{t-q+2}^2,X_t^2,\ldots,X_{t-p+2}^2\right)'\,,$$

which is constructed from the GARCH($p.q$) process (1) satisfies (2) and, hence, (3) applies. See [8] for a review on GARCH models, regular variation and stochastic recurrence equations.

Gaussian maximum likelihood estimation with heavy-tailed innovations

Gaussian maximum likelihood for the GARCH parameters α_i and β_j is based on the maximization of the log-likelihood function of a sample $X_1,\ldots,X_n$ (assuming the Z_t's iid standard normal)

$$L_n(\theta)=-\frac{1}{n}\,\sum_{t=1}^{n}[\log(\sigma_t^2(\theta))+\frac{\sigma_t^2(\theta_0)Z_t^2}{\sigma_t^2(\theta)}]\,,$$

with respect to the GARCH parameter θ, where θ_0 is the true parameter of the GARCH model, underlying the observations, and $\widehat{\theta}_n$ is the resulting *Gaussian maximum likelihood estimator*. Taylor expansion of $L_n'(\widehat{\theta}_n)$ at θ_0 yields

$$\widehat{\theta}_n-\theta_0=-(L_n''(\theta_n))^{-1}\,L_n'(\theta_0)\,,$$

for some θ_n with $|\theta_0-\theta_n|\le|\theta_0-\widehat{\theta}_n|$. By the ergodic theorem, $L_n''(\theta_n)\to\mathbf{B}_0$ a.s. for some deterministic matrix $\mathbf{B}_0$, and therefore weak limit theory for $\widehat{\theta}_n$ reduces to

$$L_n'(\theta_0)=\frac{1}{n}\sum_{t=1}^{n}\frac{(\sigma_t^2(\theta_0))'}{\sigma_t^2(\theta_0)}\,(Z_t^2-1)=\frac{1}{n}\sum_{t=1}^{n}\mathbf{G}_t\;Y_t\,.$$

If $EZ_1^4<\infty$ the CLT for stationary ergodic martingale differences ([3]) gives asymptotic normality for $\widehat{\theta}_n$

$$\sqrt{n}\,(\widehat{\theta}_n-\theta_0)\xrightarrow{d}\mathrm{N}(\mathbf{0},-E(Z_1^4-1)\,\mathbf{B}_0^{-1})\,. \tag{4}$$

This was proved in [2]. An interesting observation as regards (4) is that the GARCH structure is not essential for the limit theorem (4): as long as $\mathbf{G}_t=(\sigma_t^2(\theta_0))'/\sigma_t^2(\theta_0)$ is stationary ergodic and predictable, and $E|\mathbf{G}_0Y_0|^2<\infty$ the CLT applies. In the GARCH context it is remarkable, that $\mathbf{G}_0$ has finite moments of all orders (see [2]) and therefore the regular variation of the X_t's (see the previous section) is not essential for the asymptotic theory of $\widehat{\theta}_n$, even if $\mathrm{var}(X_0)=\infty$.

Recently, [6] have extended (4) to the case when $EZ_1^4=\infty$. Assuming that Z_1 is regularly varying with index $\alpha\in(2,4)$, they show that infinite variance stable limits appear in (4). In the paper [9] it is shown that such limits appear for general models $X_t=\sigma_tZ_t$, if (σ_t) is predictable, stationary ergodic, β-mixing with geometric rate, (Z_t) is an iid sequence and regularly varying with index $\alpha\in(2,4)$

and if $E|\mathbf{G}_0|^{(\alpha/2)+\delta} < \infty$ for some $\delta > 0$. The latter conditions are satisfied for the GARCH model.

The results of [9] are based on an analogue to the CLT for stationary ergodic sequences in the case of infinite variance summands $\mathbf{X}_t$ which was proved in [4, 5]. Assuming that $(\mathbf{X}_t)$ satisfies a mild mixing condition (milder than strong mixing) and that its finite-dimensional distributions are regularly varying with index $\kappa \in (0, 2)$, $a_n^{-1} \sum_{t=1}^{n} \mathbf{X}_t$ (suitably centered) weakly converges to a stable limit, where $P(|\mathbf{X}_0| > a_n) \sim n^{-1}$.

In particular, it applies to summands of the form $\mathbf{X}_t = \mathbf{G}_t Y_t$ for vector-valued predictable stationary ergodic $\mathbf{G}_t$ with $E|\mathbf{G}_0|^{\kappa+\delta} < \infty$, some $\delta > 0$, and regularly varying Z_t with index $\alpha = 2\kappa \in (2, 4)$. Indeed, then regular variation of the finite-dimensional distributions is conveniently verified. If one has a particular structure such as GARCH, the verification of the β-mixing condition for $\mathbf{G}_t = ((\sigma_t^2(\theta_0))'/\sigma_t^2(\theta_0))$ can be derived from β-mixing for $((\sigma_t^2)', \sigma_t^2)$. In the GARCH case, this condition can be verified by applying a result of [10] on mixing properties of solutions to stochastic recurrence equations (2): then $(\sigma_t^2, (\sigma_t^2)')$ can be embedded in such a stochastic recurrence equation.

The CLT for the GARCH Gaussian maximum likelihood estimator of θ_0 when $EZ_1^4 < \infty$ has $\sqrt{n}$-rates of convergence. This is in contrast to the case when Z_1 is regularly varying with index $\alpha \in (2, 4)$, where the rate of convergence is of the order $n^{1-2/\alpha}$. This means that slow rates of convergence and unusually wide confidence bands for the parameter estimators appear.

References

[1] Basrak, B., Davis, R.A. and Mikosch. T. (2002): Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95-116 .

[2] Berkes, I., Horváth, L. and Kokoszka, P. (2003): GARCH processes: structure and estimation. *Bernoulli* **9**, 201-228.

[3] Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley, New York.

[4] Davis, R.A. and Hsing, T. (1995): Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879-917.

[5] Davis, R.A. and Mikosch, T. (1998): The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049-2080.

[6] Hall, P.G. and Yao, G. (2003): *Econometrica* **71**, 285-317.

[7] Kesten, H. (1973): Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207-248.

[8] Mikosch, T. (2003): Modeling dependence and tails of financial time series. In: Finkenstädt, B. and Rootén, H. (Eds.) *Extreme Values in Finance, Telecommunications, and the Environment*, pp. 185-286. Chapman and Hall, Boca Raton.

[9] Mikosch, T., Straumann, D. (2003): Stable limits of martingale transforms with application to the estimation of GARCH parameters. Technical report. Available under www.math.ku.dk/∼mikosch.

[10] Mokkadem, A. (1990): Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.* **26**, 219-260.

[11] Resnick, S.I. (1987): *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

[12] Rvačeva, E. L. (1962): On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. and Probability*, American Mathematical Society, Providence, R.I. **2**, 183-205.

The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions

Per Mykland
(joint work with Yacine Aït-Sahalia)

Diffusion models, and their extensions such as jump-diffusions and Markov models driven by Lévy processes, are essential tools for much of theoretical asset pricing. Estimating these models from discrete time observations has become in recent years an active area of research in econometrics and statistics. Beyond the choice of inference strategy, an important debate in this area concerns the question of what sampling scheme to use, if a choice is available, and in any event what to do with the sampling times. The most straightforward thing to do, in accordance with the usual low-frequency data collection procedures in finance, is to view the sampling as occurring at fixed discrete time intervals, such as a day or a week. In many circumstances, however, this is not realistic. In fact, all transaction-level data are available at irregularly and randomly spaced times.

Not only are the data randomly spaced in time, but whenever a theoretical model is spelled out in continuous time, its estimation necessarily relies on discretely sampled data. By now, there is a good understanding in the literature of the implications of sampling discreteness, and how to design estimation methods that correct for it. The objective in this work is to understand the additional effect that the randomness of the sampling intervals might have when estimating a continuous-time model with discrete data. Specifically, we seek to disentangle the effect of the sampling randomness from the effect of the sampling discreteness, and to compare their relative magnitudes. We also examine the effect of simply ignoring the sampling randomness. We achieve this by comparing the properties of three likelihood-based estimators, which make different use of the observations on the state process and the times at which these observations have been recorded. We design these estimators in such a way that each one of them is subject to a specific subset of the effects we wish to measure. As a result, the differences in their properties allow us to zero in and isolate these different effects.

Our main conclusion is that the loss from not observing, or not using, the sampling intervals, will be at least as great, and often substantially greater, than the loss due to the fact that the data are discrete rather than continuous. While correcting for the latter effect has been the main focus of the literature in recent years, our results suggest however that empirical researchers using randomly spaced data should pay as much attention, if not more, to sampling randomness as they do to sampling discreteness.

The second paper develops tools for analyzing similar problems in the context of non-likelihood inference (estimating or moment equations), and studies specifically the effect of using approximations such as the Euler scheme.

A further contribution of the work is the development of a set of tools that allows these calculations to be performed in closed form.

References

[1] Aït-Sahalia, Y. and Mykland, P.A. (2003): The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71**, 483-549.

[2] Aït-Sahalia, Y. and Mykland, P.A. (2004?): Estimators of diffusions with discrete observations: a general theory. Technical report 539, Dept. of Statistics, The University of Chicago (to appear in *Ann. Statist.*).

Tail Behaviour of the Stationary Distribution of a Random Coefficient Autoregressive Model

Serguei Pergamenchtchikov

(joint work with Claudia Klüppelberg)

We consider the following autoregressive process with ARCH errors.

$$x_n = a_1 x_{n-1} + \cdots + a_q x_{n-q} + \sqrt{1 + \sigma_1^2 x_{n-1}^2 + \cdots + \sigma_q^2 x_{n-q}^2}\,\varepsilon_n\,, \quad n \in \mathbb{N}\,, \tag{1}$$

where (ε_n) are i.i.d. $\mathcal{N}(0,1)$. We are interested in the existence of a stationary version of the process $(x_n)_{n\in\mathbb{N}}$, represented by a random variable (rv) x_∞ and its properties. We investigate the tail behaviour

$$\mathbf{P}(x_\infty > t) \quad \text{as} \quad t \to \infty\,. \tag{2}$$

This is, in particular, the first step for an investigation of the extremal behaviour of the corresponding stationary process. For $q = 1$ the model (1) was investigated in Borkovec and Klüppelberg [3] by direct analytic methods. For the general case $q > 1$ it is not possible to apply this approach since in this case the model (1) is a non-linear equation with respect to x_n. One can, however, show (see Lemma 2.7 in [12]) that this model is in distribution equivalent to a random coefficient autoregressive process

$$y_n = \alpha_{1n} y_{n-1} + \cdots + \alpha_{qn} y_{n-q} + \xi_n\,, \quad n \in \mathbb{N}\,, \tag{3}$$

where the independent coefficient sequences $(\alpha_{in}\,,\, n \geq 1)$ are i.i.d. and $\alpha_{in} \sim \mathcal{N}(a_i, \sigma_i^2)$ for each $1 \leq i \leq q$. Moreover the noise variables $(\xi_n)_{n\in\mathbb{N}}$ are an i.i.d. $\mathcal{N}(0,1)$ sequence independent of $(\alpha_{in}\,,\, n \geq 1)_{1\leq i\leq q}$. Consequently, the problem (2) is equivalent to the investigation of the tail behaviour of a stationary version of the process (3) represented by a random variable y_∞.

To obtain the asymptotic behaviour of the tail of y_∞ we embed $(y_n)_{n\in\mathbb{N}}$ into a multivariate set-up.

Set $Y_n = (y_n, \ldots, y_{n-q+1})'$. Then the multivariate process (Y_n) can be considered in the much wider context of random recurrence equations of the type

$$Y_n = A_n Y_{n-1} + \zeta_n\,, \quad n \in \mathbb{N}\,, \tag{4}$$

where $\zeta_n = (\xi_n, 0, \ldots, 0)'$ and

$$A_n = \begin{pmatrix} \alpha_{1n} & \cdots & \alpha_{qn} \\ I_{q-1} & & 0 \end{pmatrix}, \quad n \in \mathbb{N}\,, \tag{5}$$

where I_{q-1} denotes the identity matrix of order $q-1$.

Such equations play an important role in many applications as e.g. in queueing; see Brandt, Franken and Lisek [4] and in financial time series; see Engle [7]. See also Diaconis and Freedman [5] for an interesting review article with a wealth of examples.

If the Markov process defined in (4) has a stationary distribution and Y has this stationary distribution, then certain results are known on the tail behaviour of Y. In the one-dimensional case ($q = 1$) Goldie [8] has derived the tail behaviour of Y in a very elegant way by a renewal type argument: the tail decreases like a power-law. For the multivariate model for the matrix A_n with positive elements Kesten [9] shows that for each non-zero vector $x \in \mathbb{R}^q$ there exists some $\lambda > 0$ such that $\lim_{t\to\infty} t^\lambda \mathbf{P}(x'Y > t) < \infty$.

However, our model (4) does not satisfy the positivity condition on the matrices A_n. Consequently, we derived a new limiting theorem for the model (4) with the matrix of special form (5) in the spirit of Kesten's results. The following is our main result.

Theorem 1. *We assume that the eigenvalues of the matrix* $\mathbf{E}\, A_1 \otimes A_1$ *have moduli less than one and* $a_q^2 + \sigma_q^2 > 0$. *Then the stationary distribution* Y *of the process (4) satisfies*

$$\lim_{t\to\infty} t^\lambda \mathbf{P}(x'Y > t) = h(x)\,, \quad x \in S = \{z \in \mathbb{R}^q \,:\, |z| = 1\}\,.$$

The function $h(\cdot)$ *is strictly positive and continuous on* S *and the parameter* λ *is given as the unique positive solution of*

$$\kappa(\lambda) = 1\,, \tag{6}$$

where for some probability measure ν *on* S

$$\kappa(\lambda) := \lim_{n\to\infty} \left(\mathbf{E}|A_1 \cdots A_n|^\lambda\right)^{1/n} = \int_S \mathbf{E}\,|x'A_1|^\lambda\, \nu(x)\,,$$

and the solution of (6) satisfies $\lambda > 2$.

References

[1] Basrak, B., Davis, R.D. and Mikosch, T. (2002): Regular variation of GARCH processes. *Stoch. Proc. Appl.* **99**, 95-115.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1989): *Regular Variation.* Revised paperback edition. Cambridge University Press, Cambridge.

[3] Borkovec, M. and Klüppelberg, C. (2001): The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220-1241.

[4] Brandt, A., Franken, P. and Lisek, B. (1990): *Stationary Stochastic Models.* Wiley, Chichester.

[5] Diaconis, P. and Freedman, D. (1999): Iterated random functions. *SIAM Review* **41**, 45-76.

[6] Diebolt, J. and Guegan, D. (1993): Tail behaviour of the stationary density of general non-linear autoregressive processes of order 1. *J. Appl. Prob.* **30**, 315-329.

[7] Engle, R.F. (1995) *ARCH. Selected Readings.* Oxford University Press, Oxford.

[8] Goldie, C.M. (1991): Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126-166.

[9] Kesten, H. (1973): Random difference equations and renewal theory for products of random matrixes. *Acta Math.* **131**, 207-248.

[10] Kesten, H. (1974): Renewal theory for functional of a Markov chain with general state space. *Ann. Probab.* **2**, 355-386.
[11] Klüppelberg, C. and Pergamenchtchikov, S. (2003): Renewal theory for functionals of a Markov chain with compact state space. *Ann. Probab.* **31** (4), 2270 - 2300
[12] Klüppelberg, C. and Pergamenchtchikov, S. (2004): The tail of the stationary distribution of a random coefficient AR(q) model. *Ann. Applied Probab.* (to appear). Available under www.ma.tum.de/stat/.

Multivariate Extremes, Max-Stable Processes and Financial Risk

Richard L. Smith

1. Introduction

Extreme value theory has become increasingly applied in mathematical finance, especially in conjunction with "Value at Risk"calculations (Embrechts *et al.* 1997, Finkenstädt and Rootzén, 2003). Univariate extreme value theory is fairly well understood by now, with extensive development of the threshold approach, which is based on the Generalized Pareto distribution (GPD) fitted to exceedances over some high threshold (Davison and Smith 1991, Coles 2001). These methods are applicable to financial time series provided some account is taken of volatility. One approach to that is due to McNeil and Frey (1999), who proposed fitting a GARCH model to financial time series with residuals from an unknown distribution, whose tail was analyzed using threshold methods.

However, there has been relatively less work on dependence in the extremes, by which we mean both temporal dependence in a single time series, and cross-dependence between time series. Multivariate extreme value theory and its generalization, the theory of max-stable processes, are natural candidates to model the joint extremal behaviour of several financial time series. This is the subject of the present paper.

2. Multivariate Extreme Value Theory

Suppose $\mathbf{Y}_i = (Y_{i1}, ..., Y_{iD}),\ i = 1, 2, ...$ is an i.i.d. sequence of D-dimensional random vectors. For each $d \in \{1, ..., D\}$, let $M_{nd} = \max\{Y_{id},\ 1 \le i \le n\}$.

If normalizing constants a_{nd}, b_{nd} and a D-dimensional distribution function G exist such that as $n \to \infty$,

$$\Pr\left\{\frac{M_{nd} - b_{nd}}{a_{nd}} \le x_d,\ 1 \le d \le D\right\} \to G(x_1, ..., x_D)$$

then G is called a *multivariate extreme value distribution.*

There are various representations of multivariate extreme value distributions due to Pickands, de Haan and Resnick, Deheuvels, etc. (Resnick (1987) has a comprehensive account) but these are too general to be directly applicable to statistics. Some authors (e.g. Tawn, Coles) have used parametric subfamilies while others (e.g. de Haan) used nonparametric approaches, but it is not easy to apply any of the existing methods to series in very high dimensions. This motivates an alternative approach.

3. Max-Stable Processes

Max-stable processes are the infinite-dimensional generalization of multivariate extreme value distributions. They are a natural framework within which to study extremal properties of multivariate time series.

Suppose $\{Y_{id},\ i = 0, \pm 1, \pm 2,\ d = 1, ..., D\}$ is a D-dimensional time series with discrete time index i. Without loss of generality, we may assume $\Pr\{Y_{id} \le y\} = e^{-1/y}$ for $0 < y < \infty$ (the unit Fréchet distribution). In practice, this would be achieved only above a given threshold, by first fitting a univariate threshold model to the marginal distributions.

The process is *max-stable* if for any $n \ge 1,\ N \ge 1,\ y_{id} \ge 0$ for $i = 1, ..., n,\ d = 1, ..., D$,

$$\Pr^N \{Y_{id} \le N y_{id},\ 1 \le i \le n,\ 1 \le d \le D\} = \Pr\{Y_{id} \le y_{id},\ 1 \le i \le n,\ 1 \le d \le D\}.$$

A subclass of max-stable consists of *multivariate maxima of moving maxima* (M4 for short) defined by

$$Y_{id} = \max_{\ell=1}^{\infty} \max_{k=-\infty}^{\infty} a_{\ell,k,d} Z_{\ell,i-k},$$

where $Z_{\ell,i}$ are independent unit Fréchet for all ℓ, i; $a_{\ell,k,d} \ge 0$; and $\sum_{\ell=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{\ell,k,d} = 1, \quad d = 1, ..., D$. For this process,

$$\Pr\{Y_{id} \le y_{id},\ i = 1, ..., n,\ d = 1, ..., D\} = \exp\left(- \sum_{\ell=1}^{\infty} \sum_{m=-\infty}^{\infty} \max_{k=1-m}^{n-m} \max_{d=1}^{D} \frac{a_{\ell,k,d}}{y_{m+k,d}} \right).$$

Smith and Weissman (1996), generalizing Deheuvels (1983), showed that subject to some non-degeneracy conditions, any max-stable process may be approximated arbitrarily closely by an M4 processes.

Statistically, however, these processes are hard to estimate, because of the presence of "signature patterns" of the form

$$Y_{id} = a_{\ell^*,i-m^*,d} Z_{\ell^*,m^*},\ i = 1, ..., n,\ d = 1, ..., D,$$

which arise when a single very large value Z_{ℓ^*,m^*} dominates all its neighbors. If these relations hold, it is possible to derive very precise estimates of the coefficients (Zhang and Smith, 2004a) but this approach is not robust against even tiny deviations from the model. For this reason, it is not a practical approach with real data.

Some alternative estimation strategies include

- **(a):** estimation based on the empirical distribution function (Hall, Peng and Yao, 2002; Zhang and Smith, 2004b);
- **(b):** assuming the observed process is of the form $X_{id} = Y_{id} + \epsilon_{id}$ with Y an M4 process and $\{\epsilon_{id}\}$ random noise; it may then be possible to filter out the noise by Monte Carlo methods. In ongoing PhD research, Francisco Chamú of the University of North Carolina has been exploring this approach;

(c): a more *ad hoc* method in which observed signature patterns are grouped into clusters and the coefficients of the M4 process inferred from the cluster centers (Smith 2003).

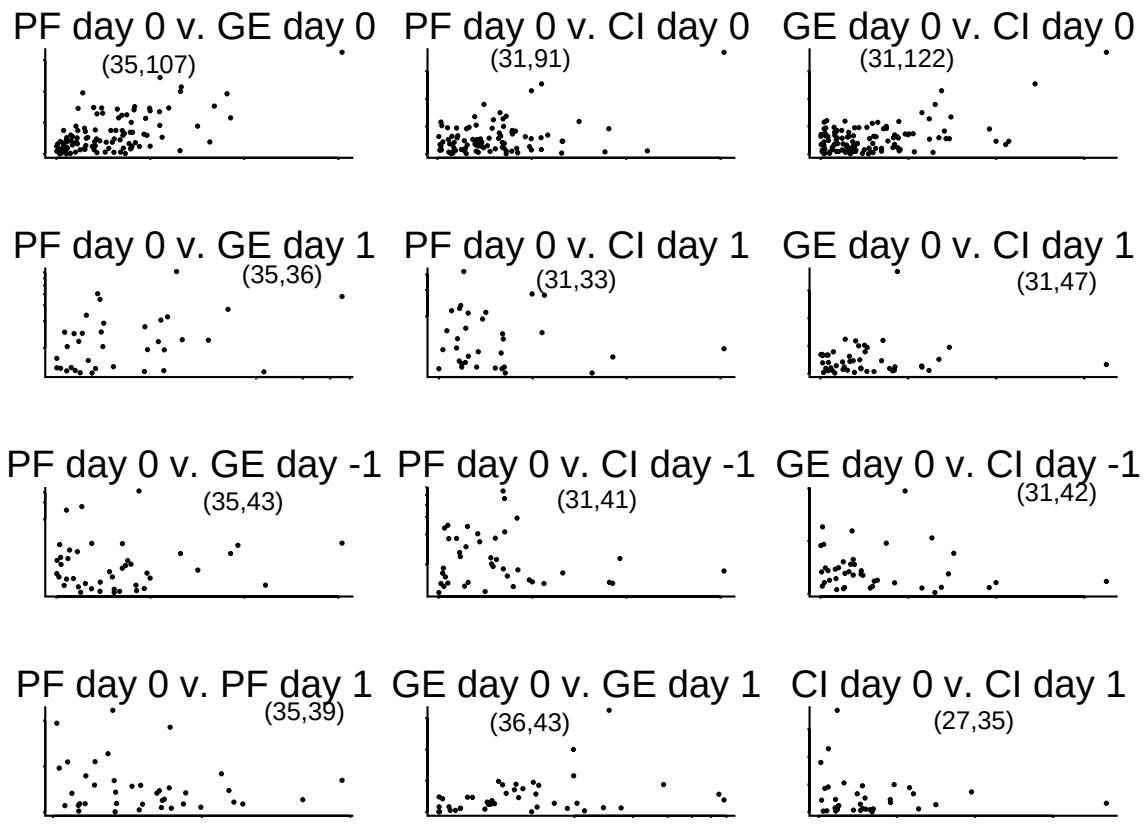

FIGURE 1. Scatterplots of standardized exceedances on unit Fréchet scale. The three stocks are Pfizer (PF), General Electric (GE) and Citibank (CI); plotted are the values on the current day (day 0) versus current day, following day (day 1) and previous day (day –1). The two numbers displayed on each plot are the actual number of joint exceedances (second number), and the expected value of the number of joint exceedances if the two variables were independent (first number). It can be seen that for all three "day 0 v.day 0" plots, there is substantial dependence between the extremes of the two series. For plots of day 0 against day 1 or day –1, however, the evidence for dependence is much less clear-cut.

4. Application to Financial Time Series

We consider 20 years of financial returns data from three stocks (Pfizer, GE and Citibank). For each series, the GARCH(1,1) model is used to estimate volatility,

and the process standardized by dividing returns by the estimated volatility. They are then transformed to unit Fréchet margins (above a threshold) by applying univariate extreme value technology. Pairwise correlation plots of the Fréchet standardized exceedances (Fig. 1) show substantial dependence among the three series, which we model using an M4 process. In this application, we estimated coefficients $a_{\ell,k,d}$ which are assumed non-zero for $\ell = 1, 2, ..., 25$ and $k = -2, -1, 0, 1, 2$.

Finally, a cross-validation exercise shows that the model provides good estimation of some simple functionals of the joint extremes. One possible functional is the probability that, over a window of 10 trading days (a typical time window in Value at Risk calculations), at least one of the three series crosses a given target value. Fig. 2 shows both an empirical crossing rate and a cross-validated model-generated crossing rate (Smith, 2003) as the target value increases; the agreement is excellent until the very highest target values, where neither method can be expected to give accurate results.

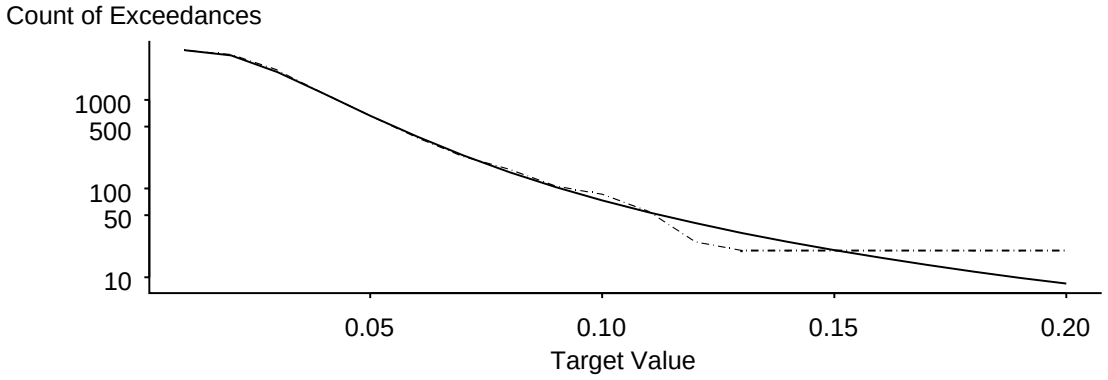

FIGURE 2. Plot of estimated number of expected exceedances of a given target value for the maximumof the daily returns over all 3 stocks over a 10-day window. Solid curve: cross-validated model-based estimate. Dashed curve: empirical value.

5. Acknowledgements

Portions of this work are collaborative with Ishay Weissman (Technion), Zhengjun Zhang (Washington University, St.Louis) and Francisco Chamú (University of North Carolina). Financial support is acknowledged from NSF grants DMS-0084375 and DMS-9971980.

References

[1] Coles, S.G. (2001): *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, New York.

[2] Davison, A.C. and Smith, R.L. (1990): Models for exceedances over high thresholds (with discussion). *J.R. Statist. Soc. Ser. B* **52**, 393-442.

[3] Deheuvels, P. (1983): Point processes and multivariate extreme values. *J. Multivar. Anal.* **13**, 257-272.
[4] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, New York.
[5] Finkenstädt, B. and H. Rootzén, H. (eds.) (2003): *Extreme Values in Finance, Telecommunications and the Environment*. Chapman and Hall/CRC Press, London.
[6] Hall, P., Peng, L. and Yao, Q. (2002): Moving-maximum models for extrema of time series. *J. Stat. Plann. Inference* **103**, 51-63.
[7] McNeil, A. and Frey, R. (1999): Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empirical Finance* **7**, 271-300.
[8] Resnick, S. (1987): *Extreme Values, Point Processes and Regular Variation*. Springer Verlag, New York.
[9] Smith, R.L. (2003): Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of Finkenstädt and Rootzén, 1-78.
[10] Smith, R.L. and Weissman, I. (1996): Characterization and estimation of the multivariate extremal index. http://www.stat.unc.edu/postscript/rs/extremal.pdf (under revision.)
[11] Zhang, Z. and Smith, R.L. (2004a): The behaviour of multivariate maxima of moving maxima processes. Tentatively accepted for *J. Appl. Prob.*
[12] Zhang, Z. and Smith, R.L. (2004b): On the estimation and application of max-stable processes. Submitted for publication.

A Flexible Class of Stochastic Volatility Models of the Diffusion-Type

Michael Sørensen

Stochastic volatility models of the type

$$dX_t = (\alpha + \beta V_t)dt + \sqrt{V_t}dW_t,$$

where V_t is a suitable positive stochastic process, are widely used in finance to model the logarithm of the price of an asset. Several possible specifications of the process V have been proposed. Barndorff-Nielsen and Shephard (2001) proposed to model V as an Ornstein-Uhlenbeck process driven by a Lévy process or a sum of such processes, which is a very flexible class of models. Here we present a class of models with a similar flexibility where the volatility process is a sum of mean-reverting processes driven by Wiener processes. The results given here are based on the paper Bibby, Skovgaard and Sørensen (2003).

Let f be a given continuous, bounded, and strictly positive probability density on $(0, \infty)$ that is zero when $x \leq 0$ and has finite variance. Define a function v by

$$v(x) = \frac{2\theta \int_l^x (\mu - y) f(y) dy}{f(x)}, \quad x > 0,$$

where μ denotes the expectation of f. It is not difficult to see that $v(x) > 0$ for $x > 0$. The stochastic differential equation

$$dV_t = -\theta(V_t - \mu)dt + \sqrt{v(V_t)}dW_t, \quad t \geq 0,$$

where W denotes a standard Wiener process, has a unique weak solution that is ergodic with invariant density f. If $V_0 \sim f$, V is stationary, and its autocorrelation

function is given by

$$\mathrm{Corr}(V_{s+t}, V_s) = e^{-\theta t}, \quad s, t \geq 0.$$

The diffusion V is the only ergodic mean-reverting diffusion with invariant density f.

The function v can be found explicitly for a number of standard distributions on $(0, \infty)$. For the gamma-distribution with shape-parameter α and scale-parameter β

$$v(x) = 2\theta\beta x.$$

In this case the volatility model is the square root process, and the corresponding volatility model is the Heston (1993) model. For the inverse Gaussian distribution with density

$$f(x) = \left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}} x^{-\frac{3}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0,$$

we find that

$$v(x) = 4\theta\mu\sqrt{\frac{2\pi}{\lambda}} e^{\lambda/\mu} x^{3/2} \exp\left(\frac{\lambda}{2\mu^2}x + \frac{\lambda}{2}x^{-1}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}} - \sqrt{\frac{\lambda x}{\mu^2}}\right),$$

where Φ is the standard normal distribution function.

Usually, the correlation function $e^{-\theta t}$ is too simple to fit the autocorrelation of the volatility observed in financial time series. Therefore the following construction is useful. Let f be a strictly positive, infinitely divisible probability density on $(0, \infty)$ that is zero when $x \leq 0$, and let $C(t)$ denote the characteristic function of f. Suppose the positive real numbers φ_i, $i = 1, \ldots, m$, satisfy that $\varphi_1 + \cdots + \varphi_m = 1$. Then the functions $C(t)^{\varphi_i}$, $i = 1, \ldots, m$, are characteristic functions too. Assume that the corresponding density functions f_i, $i = 1, \ldots, m$, satisfy the conditions imposed on f earlier, and define

$$v_i(x) = \frac{2\theta_i}{f_i(x)} \int_0^x (\varphi_i \mu - y) f_i(y) dy.$$

Then the process

$$V_t = V_t^{(1)} + \cdots + V_t^{(m)},$$

where

$$dV_t^{(i)} = -\theta_i \left(V_t^{(i)} - \varphi_i \mu\right) dt + \sqrt{v_i\left(V_t^{(i)}\right)} dB_t^{(i)},$$

with $B^{(1)}, \ldots, B^{(m)}$ denoting independent standard Wiener processes, has marginal density f, provided that $V_0^{(i)} \sim f_i$, $i = 1, \ldots, m$. The autocorrelation function of V is given by

$$\mathrm{Corr}(V_{s+t}, V_s) = \varphi_1 \exp(-\theta_1 u) + \cdots + \varphi_m \exp(-\theta_m u).$$

For the gamma-distribution with shape-parameter α and scale-parameter β

$$v_i(x) = 2\beta\theta_i x,$$

and for the inverse Gaussian distribution

$$v_i(x) = 4\theta_i\mu\sqrt{\frac{2\pi}{\lambda}}e^{\varphi_i\lambda/\mu}x^{3/2}\exp\left(\frac{\lambda}{2\mu^2}x + \frac{\varphi_i^2\lambda}{2}x^{-1}\right)\Phi\left(-\varphi_i\sqrt{\frac{\lambda}{x}} - \sqrt{\frac{\lambda x}{\mu^2}}\right).$$

For distributions where the f_i cannot be found explicitly, an approximation to v_i can be found in Bibby, Skovgaard and Sørensen (2003).

References

[1] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econometrics (with discussion). *J. R. Stat. Soc. Ser. B* **63**, 167-241.

[2] Bibby, B.M., Skovgaard, I.M. and Sørensen, M. (2003): Diffusion-type models with given marginals and autocorrelation function. Preprint No. 5, Department of Applied Mathematics and Statistics, University of Copenhagen.

[3] Heston, S.L. (1993): A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327-343.

Adaptive Estimation for a Varying Coefficient GARCH Model

Vladimir Spokoiny
(joint work with Jörg Polzehl)

Financial time series are often modelled by parametric ARCH or GARCH models under the assumption of stationarity. This approach is not flexible enough to incorporate models with structural breaks and time varying parameters. This paper presents a unified approach for modeling non (local) stationary time series including change point and smooth transition models. The procedure is based on the *Adaptive Weights* idea from Polzehl and Spokoiny (2000, 2002, 2003). The paper discusses important theoretical properties of the method and illustrates its numerical performance by mean of simulated examples and applications to real data.

References

[1] Polzehl, J. and Spokoiny, V. (2000): Adaptive weights smoothing with applications to image segmentation. *J. R. Stat. Soc. Ser. B* **62**, 335-354.

[2] Polzehl, J. and Spokoiny, V. (2002): Local likelihood modeling by adaptive weights smoothing. Preprint 787. WIAS 2002. Available under http://www.wias-berlin.de/publications/preprints/787.

[3] Polzehl, J. and Spokoiny, V. (2003): Varying coefficient regression modeling by adaptive weights smoothing. Preprint 818, WIAS 2003.

Is GARCH(1,1) as good a model as the Nobel prize accolades would imply

Cătălin Stărică

1. Abstract

This paper investigates the relevance of the stationary, conditional, parametric ARCH modeling paradigm as embodied by the GARCH(1,1) process to describing and forecasting the dynamics of returns of the Standard & Poors 500 (S&P 500) stock market index.

A detailed analysis of the series of S&P 500 returns featured in Section 3.2 of the Advanced Information note on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel reveals that during the period under discussion, there were no (statistically significant) differences between GARCH(1,1) modeling and a simple non-stationary, non-parametric regression approach to next-day volatility forecasting.

A second finding is that the GARCH(1,1) model severely over-estimated the unconditional variance of returns during the period under study. For example, the annualized implied GARCH(1,1) unconditional standard deviation of the sample is 35% while the sample standard deviation estimate is a mere 19%. Over-estimation of the unconditional variance leads to poor volatility forecasts during the period under discussion with the MSE of GARCH(1,1) 1-year ahead volatility more than 4 times bigger than the MSE of a forecast based on historical volatility.

We test and reject the hypothesis that a GARCH(1,1) process is the true data generating process of the longer sample of returns of the S&P 500 stock market index between March 4, 1957 and October 9, 2003. We investigate then the alternative use of the GARCH(1,1) process as a local, stationary approximation of the data and find that the GARCH(1,1) model fails during significantly long periods to provide a good local description to the time series of returns on the S&P 500 and Dow Jones Industrial Average indexes.

Since the estimated coefficients of the GARCH model change significantly through time, it is not clear how the GARCH(1,1) model can be used for volatility forecasting over longer horizons. A comparison between the GARCH(1,1) volatility forecasts and a simple approach based on historical volatility questions the relevance of the GARCH(1,1) dynamics for longer horizon volatility forecasting for both the S&P 500 and Dow Jones Industrial Average indexes.

2. Figures

Figure 1 displays the estimated $\alpha_1 + \beta_1$ under the assumption of non-stationary data. The Garch(1,1) model has been initially estimated on the first 2000 observations of the sample corresponding roughly to the period 1957-1964, then re-estimated every 50 observations on a sample containing 2000 past observations.

The graph shows that the IGARCH effect significantly[1] affects the GARCH(1,1) models (estimated on a sample that ends) during the period 1997-2003[2]. This fact at its turn, is likely to cause the explosion of the estimated unconditional variance of the GARCH(1,1) processes fitted on samples that end during this period.

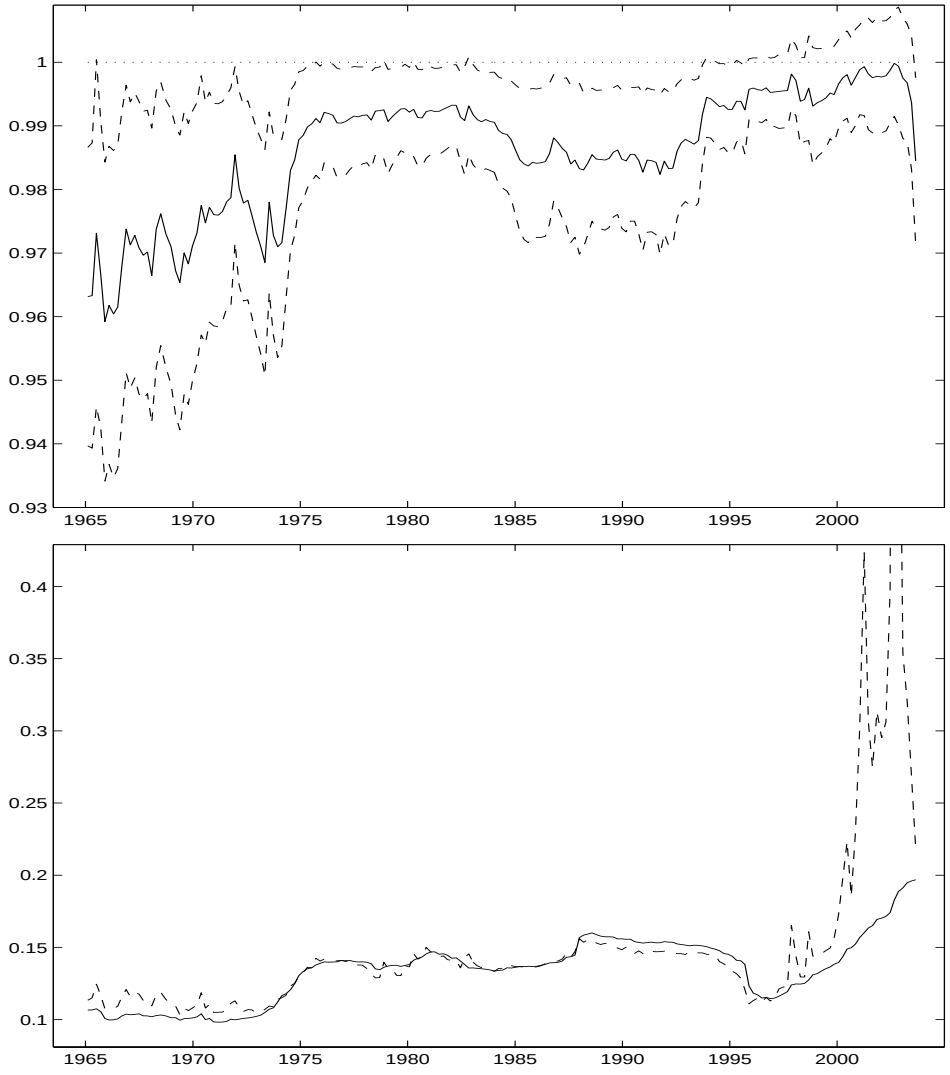

FIGURE 1. Top: Estimated $\alpha_1 + \beta_1$. Bottom: Estimated GARCH (1,1) sd (dotted line) together with sample sd (both estimates are annualized) (full line) for the S&P 500 log-returns. The time mark corresponds to the end of the sub-sample that yields the two standard deviation estimates. While most of the time the two curves in the bottom graph are remarkably close to each other, the GARCH(1,1) variance seems to explode towards the end of the sample.

[1]The point estimate is close to 1 and, more importantly, 1 belong to the 95% confidence interval.

[2]During the interval 1994-1996, the value 1 is the upper bound of the confidence interval.

To see that indeed this is the case, let us take look at the bottom graph of the same Figure 1 where the GARCH(1,1) unconditional sd (broken line) and the corresponding sample sd (full line) are displayed. The GARCH(1,1) unconditional sd is obtained from the values of the parameters estimated on a window of size 2000 moving through the data. The graph shows a good agreement between the two estimates at all times except during the period when the IGARCH effect becomes strongly statistically significant, i.e. samples that end in the interval 1997-2003[3],[4].

The bottom graph in Figure 1 show that the GARCH(1,1) model fails to provide a local stationary approximation to the time series of returns on the S&P 500 during significantly long periods.

An explanation for the strong IGARCH effect in the second half of the 90's can be the sharp change in the unconditional variance (see Mikosch and Starica [1]). There it is proved, both theoretically and empirically, that sharp changes in the unconditional variance can cause the IGARCH effect. Figure 2 displays non-parametric estimates of the unconditional sd together with the 95% confidence intervals[5] for the S&P 500 returns (top) and the Dow Jones industrial index returns (bottom). The two graphs show a pronounced increase of the volatility from around 5% in 1993-1994 to three times as much (around 15%) in the period 2000-2003.

References

[1] Mikosch, T. and Stărică, C. (2004): Non-stationarities in financial time series, the long-range dependence and IGARCH effects. *Rev. Econom. Statist.* (to appear). Available under www.math.ku.dk/∼mikosch.

[2] Mikosch, T. and Stărică, C. (2003) Stock market risk-return inference. An unconditional non-parametric approach. Technical Report. Available under www.math.ku.dk/∼mikosch.

Pricing of Contingent Claims When Prices Are Perturbed: An Elementary Example for Discussion

J. Michael Steele

The basic aim of this talk was to suggest consideration of a class of models that one may view as perturbations of another (unobserved) price processes which is either well-understood or which may be blessed with some special theoretical

[3]The analysis was also performed with smaller sample sizes of 1500, 1250 and 1000. As expected, the confidence intervals in Figures 1 get wider and hence less meaningful. However, for every sample sized mentioned, there is always a period between 1997 and 2003 where the unconditional variance of the estimated model explodes. Estimation based on samples smaller than 1000 observations is infeasible as it produces extremely unstable coefficients and renders problematic the use of any asymptotic result.

[4]Contrast this finding with the statement on page 16 of the Advanced Information note: "Condition $\alpha_1+\beta_1 < 1$ is necessary and sufficient for the first-order GARCH process to be weakly stationary, and the estimated model (on the short S&P 500, n.n.) satisfies this condition."

[5]The method used to obtain the estimates is that of kernel smoothing in the framework of non-parametric regression with non-random equi-distant design points. For more details on the performance of this method on financial data see Mikosch and Starica [2].

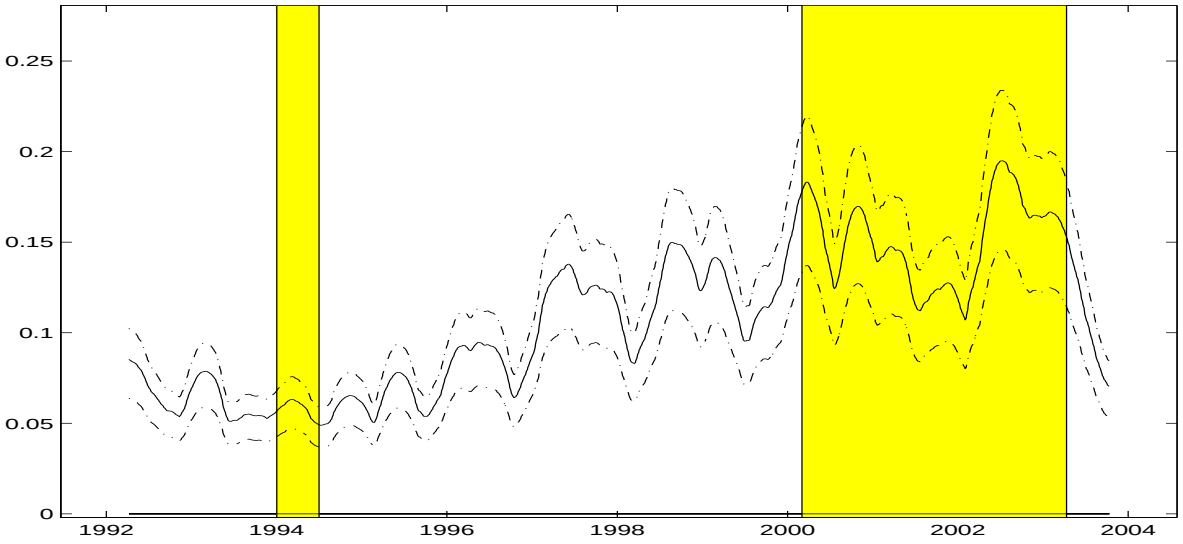

FIGURE 2. Estimated unconditional standard deviation (annualized) with 95% confidence intervals for the S&P 500 returns. The shaded areas correspond to bear market periods.

appeal. This discussion is part of a larger program which hopes to explain more fully the costs and benefits of relying on misspecified models.

The example used in the talk was simply the Black-Scholes model where the observed log prices are perturbed by a mean-zero mean-reverting process, and, for specificity, we took the perturbing process to be an independent Ornstein-Uhlenbeck process. Formally, we considered processes S_t and O_t which satisfy

$$dS_t = \mu dt + \sigma dW_t \quad \text{and} \quad dO_t = -\alpha O_t dt + \epsilon d\widetilde{W}_t$$

where the process $(W_t, \widetilde{W}_t)$ is an uncorrelated Brownian motion in $\mathbb{R}^2$, and we then considered a price process $\{P_t\}$ which is specified by setting

$$P_t = P_0 \exp(Y_t) \quad \text{and} \quad Y_t = S_t + O_t. \tag{1}$$

One reason to consider this model is that it contains as special cases both the Black-Scholes model and the model of Lo and Wang (1995). Like the Lo and Wang (1995) model, $\{P_t\}$ exhibits aspects of predictability, but here it also captures additional elements of economic reality. Specifically, we view $P_t^T \equiv \exp S_t$ as a "true" (but unobserved) price process, and we posit that market forces will drive the observed price P_t back to P_t^T after any random deviations from P_t^T. The model offers a practical compromise between a theoretically appealing model, and one which manifests some modest predictability.

On interesting feature of $\{P_t\}$ is that it is not a Markov process, so *a priori* one might not expect that the PDE methods for pricing contingent claims would apply. Nevertheless, in the case of European call options, easy calculations and ancient recipes quickly bring one to an almost exact replicate of the Black-Scholes PDE — only the volatility parameter is changed. Moreover, this heuristic derivation

turns out to be theoretically justified; risk-neutral pricing theory leads one to the same valuation formula.

A further instructive feature of the price process $\{P_t\}$ is its relation to the notion of viability which Bick (1990) introduced to addresses the consistency of a price process with a certain economic equilibrium. He and Leland (1993) later developed a PDE based criterion for viability, and, although it is not strictly applicable here, one can check that the process $\{P_t\}$ does not pass the He-Leland test (where, with eyes closed, we pretend for a moment that $\{P_t\}$ is Markovian!). It remains to be seen if $\{P_t\}$ is viable in the more general framework of Decamps and Lazrak (2000), this also seems doubtful. Nevertheless, the practical motivation underlying consideration of the process $\{P_t\}$ remains in tact; it is, after all, a perturbation of a process $\{P_t^T\}$ that passes anyone's test of viability.

References

[1] Bick, A. (1990): On viable diffusion price processes of the market portfolio. *J. Finance* **45**, 673-689.

[2] Decamps, J.P. and Lazrak, A. (2000): A martingale characterization of equilibrium asset prices processes. *Econom. Theory* **15**, 207-213.

[3] He, H. and Leland, H. (1993): On equilibrium asset price processes. *Review of Financial Studies* **6**, 593-617.

[4] Lo, A.W. and Wang, J. (1995): implementing option pricing models when asset returns are predictable. *J. Finance* **50**, 87-129.

Quasi-Maximum Likelihood Estimation and Conditional Heteroskedastic Time Series

Daniel Straumann

By exploiting the techniques of stochastic recurrence equations, we develop a general and unifying limit theory for the maximum likelihood estimator (MLE) and quasi maximum likelihood estimator (QMLE) in a certain parametric class of conditionally heteroscedastic processes, which contains widely used financial time series models: (asymmetric) GARCH(1,1) and EGARCH. Our approach generalizes and clarifies work of Lumsdaine (1996) and Berkes et al. (2003). We furthermore discuss the issue of misspecification in the MLE and the behaviour of the QMLE in the presence of a heavy-tailed noise distribution. This complements work by Newey and Steigerwald (1997) and Hall and Yao (2003).

A Multinomial Approximation of American Option Prices in a Lévy Process Model

Alex Szimayer
(joint work with Ross A. Maller and David H. Soloman)

This paper examines the pricing of American options in models where the stock price follows an exponential Lévy process. We propose a multinomial model approximating the stock price process which can be viewed as a generalisation of the binomial model of Cox et al. (1979) adapted from Brownian motion to the broader class of Lévy processes. Under mild conditions, it is proved that American option prices obtained under the multinomial model converge to the corresponding prices under the continuous time Lévy process model. Further, explicit schemes are given for the jump diffusion model, the variance gamma model.

The Distribution of the LR Test for a Nonlinear Latent Variable Model of Equity Returns

Mark Van De Vyver
(joint work with Ross A. Maller)

1. Abstract

This paper is devoted to deriving, under quite general conditions, the distribution of a likelihood ratio statistic for testing whether several versions of a generalized autoregressive conditional heteroscedasticity (GARCH) model are superior to a general random walk model, in depicting the true (unknown) data generating process for the natural log of an equity price or their continuously compounded returns. This is the statistic which the one sided LM test statistic approximates, and provides a first check as to whether GARCH effects are in fact present in the data. The application of these results is illustrated using equity market data (contained in the full paper).

2. Introduction

This paper extends the results of [7] and provides new results in the subject of latent variable model specification testing in the discrete time, continuous state setting. Specifically, we consider models that naturally arise in the context of financial modelling, and derive the distribution of the deviance statistic (negative two times the quasi log–likelihood ratio), which is of use in testing for the reduction of the alternative model to a more parsimonious null model. The deviance statistic is that which the more commonly used Lagrange Multiplier (LM) statistic approximates ([5]). Initially the alternative model is specified quite generally, and the parametrization we consider includes the nonlinear GARCH model. The NGARCH-M model we consider is suitable for option pricing, and belongs to a class for which [9] have developed an accurate and parsimonious option pricing algorithm, capable of pricing American and exotic options. While less general, the

null model is of special interest. This is a random walk with innovations that are independently and identically distributed. When the innovations have the normal distribution this model converges to the geometric Brownian motion model that lies behind the celebrated Black-Scholes and Merton (BSM) option pricing model ([1] and [8]). The following is the model of an asset price series, which incorporates volatility clustering as well as an asymmetric correlation between returns and volatility innovations:

$$\begin{aligned} X_i &= \phi X_{i-1} + \left(\mu - \tfrac{1}{2}\sigma_i^2 + \lambda\sigma_i\right) + \varepsilon_i\sigma_i \\ \sigma_i &= \sqrt{\omega + \alpha\sigma_{i-1}^2\left(\varepsilon_{i-1} - c\right)^2 + \beta\sigma_{i-1}^2}. \end{aligned} \tag{1}$$

Here X_i is the natural logarithm of the stock price, μ is a drift parameter that frequently is interpreted as the risk free rate; ϕ is an autoregression coefficient (and we are particularly interested in testing the hypothesis that ϕ is unity); λ is another drift term interpreted as the market price of risk; ω is the instantaneous variance of a Gaussian random walk when there are in fact no ARCH or GARCH effects present; α and β are, respectively, the ARCH and GARCH terms that feedback the effects of past observations into the variance equation; c is a generic parameter in the variance equation which permits an asymmetric correlation between the returns and volatility process (see [6] and [4]). We refer to the first two equations as the price (or, when $\phi = 1$, return) equation and volatility equation. We wish to test two null hypotheses of interest, against different alternatives. Both null models will have a common feature $\phi = 1$, and $\alpha = \beta = 0$. Then $\sigma_i = \sigma = \omega$ does not depend on i and we can write (1) as:

$$X_i = X_{i-1} + \mu - \frac{1}{2}\omega + \lambda\sqrt{\omega} + \varepsilon_i\sqrt{\omega}, \quad i = 1, 2, \ldots, n \tag{2}$$

Notice that the parameter c is not present under the null (the term containing c disappears from (1) when $\alpha = 0$), and two other parameters, μ and λ, combine into a single drift parameter and thus cannot be uniquely determined under the null. To reflect this we introduce a drift parameter, ψ, in a simple reparameterization, as $\psi = \mu + \lambda\sqrt{\omega} - \frac{1}{2}\omega$. Applying this to (1) introduces an additional parameter that disappears under the null hypothesis. We handle these parameters using the methods of [2] and [3].

3. Results

3.1. **Testing for GARCH effects alone.**

Theorem 1. *Assume X_i satisfies* (1) *and that $\phi = 1$, for i.i.d ε_i with expectation 0, variance 1 and finite third and fourth moments, μ_3 and μ_4. Suppose that the null hypothesis holds; $\phi = 1$, and $\alpha = \beta = 0$. When evaluating the null and alternate models, the deviance statistic, d_n, has asymptotic distribution:*

$$d_n^{(2)}(\tau) \xrightarrow{D} kN^2(0,1)I\big(N \geqslant 0\big),$$

where N is standard normal and k is

$$k = 1 + \frac{\mu_4 + 2\mu_3\left(\lambda_0 - \sqrt{\omega_0}\right) - 3}{2 + \left(\lambda_0 - \sqrt{\omega_0}\right)^2}. \tag{3}$$

$I\big(Z \geqslant 0\big)$ is 1 if $Z \geqslant 0$ and 0 otherwise.

3.2. **Testing for GARCH effects and a unit root.**

Theorem 2. *Assume X_i satisfies (1) for i.i.d ε_i with expectation 0, variance 1 and finite third and fourth moments, μ_3 and μ_4. Suppose that the null hypothesis holds; $\phi = 1$, and $\alpha = \beta = 0$. When evaluating the null and alternate models, the deviance statistic, d_n, has asymptotic distribution:*

$$d_n^{(1)}(\tau) \xrightarrow{D} N_1^2 + N_2^2 I\big(N_2 \geqslant 0\big).$$

where N_i are normal random variables with some variance-covariance given in the full paper.

4. Empirical Application

For each company in the S&P 500 index as at 8/8/2003, we select those which have at least 1000 observations, leaving 481 firms. Using individual company returns over this period we fit (1) and (2), and calculate the robust deviance as set out in Theorem 1. The first box plot summarizes the lower range of values of the robust deviances calculated using the 95% CI of the moments of the QML estimated resdiuals. The second and third box plots summarize the range of values of the robust deviance, and it's upper range of values. The last box plot shows the two sided lagrange multiplier test statistic's empirical values. In each plot the box spans the 25%-75% quantiles, the whiskers cover 3/2 of the interquantile range from the edges of the box, individual points represent outlying observations and the dashed line indicates the median value. The two horizontal lines indicate the one-sided chi-square critical values, with one degree of freedom, at the 5% and 1% levels of significance. The upper and lower robust deviance values reflect the range of values the deviance may take, using the 95% CI surrounding the null model residuals' moments, calculated using QML estimated residuals.

References

[1] Black, F. and Scholes, M. (1973): The pricing of options and corporate liabilities. *Journal of Political Economy* **81** (3), 637-54.

[2] Davies, R.B. (1977): Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** (2), 247-254.

[3] Davies, R.B. (1987): Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** (1), 33-43.

[4] Duan, J.-C. (1995): The GARCH option pricing model. *Math. Finance* **5**, 13-32.

[5] Engle, R. F.(1983): Estimates of the variance of U.S. inflation based upon the ARCH model. *Journal of Money, Credit and Banking* **15** (3), 286-301.

[6] Engle, R.F. and Ng, V. K. (1993): Measuring and testing the impact of news on volatility. *J. Finance* **48** (5), 1749-78.

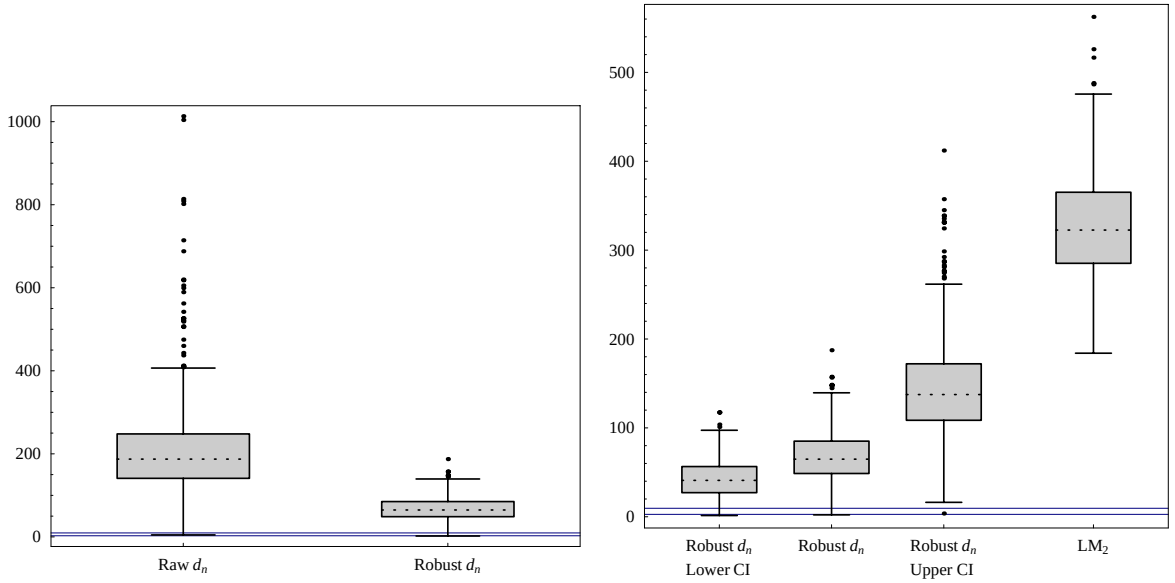

FIGURE 1. GARCH vs. random walk hypothesis raw and robust deviances of 481 companies in the S&P 500 index: 18/8/1999 to 8/8/2003

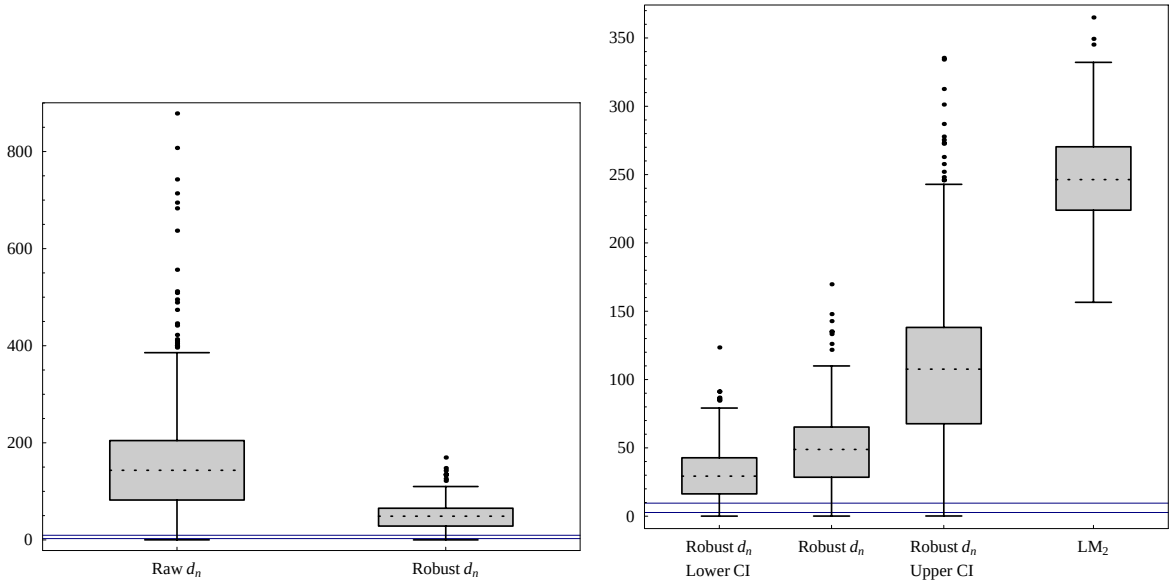

FIGURE 2. ARCH vs. random walk hypothesis raw and robust deviances of 481 companies in the S&P 500 index: 18/8/1999 to 8/8/2003

[7] Klüppelberg, C., Maller, R., Van De Vyver, M. and Wee, D. (2002): Testing for Reduction to Random Walk in Autoregressive Conditional Heteroscedasticity Models. *Econom. J.* **5** (2), 387-416.

[8] Merton, R.C. (1973): Theory of rational option pricing. *The Bell Journal of Economics* **4** (1), 141-83.

[9] Ritchken, P. and Trevor, R. (1999): pricing options under generalized GARCH and stochastic volatility processes. *J. Finance* **54**, 377-402.

Option Pricing and Statistics Inference for GARCH Models and Diffusions

Yazhen Wang

Stock market modeling has two types of approaches. One is continuous-time modeling that assumes a stock price to change with time continuously and obey a continuous-time stochastic process. Historically continuous-time models based on stochastic differential equations have been developed in financial economics, and modern finance theory is much based on the continuous-time modeling. However, in reality all data are recorded only at discrete intervals. Unknown parameters in the continuous-time models need to be estimated and tested from the observed discrete-time data. Due to the difficulty in statistical inference for the continuous time model based on the discrete data, the validity of the continuous-time modeling is not straightforward to check. Another approach is discrete-time modeling of available discrete data. Successful discrete-time models are the autoregressive conditionally heteroscedastic (ARCH) models. These discrete-time models often provide parsimonious representations for the observed discrete-time data, and their statistical inference is relatively much easier. The weak convergence of the discrete-time ARCH model to continuous-time diffusion established first by D. Nelson in early 1990 has generated a general belief that the ARCH model and diffusions are more or less equivalent.

This talk presents asymptotic equivalence of the Garch, discrete stochastic volatility (SV), and diffusion models with respect to option pricing, implied volatility, and statistical inferences based on option data (or implied volatility). As discrete observation intervals shrink to zero, the GARCH and SV models weakly converge to a bivariate diffusion. First we prove that the GARCH option price converges to diffusion price at the speed near to the square root of the observation interval length. Second we show that under the three models, the prices of a European option and their corresponding implied volatilites are equal up to the order near to the square root of the observation interval length, and asymptotically option based statistical inferences under the three models are statistically equivalent. This shows that asymptotically the three models are equivalent in all aspects regarding to option pricing, implied volatility and statistical inference for option data. It presents a sharp contrast with nonequivalence of the GARCH and its diffusion limit regarding to statistical inferences for historical time series price data.

Valuation of American Options via Basis Functions

Samuel Po-Shing Wong
(joint work with Tze Leung Lai)

Using the methodology of pricing and hedging American options proposed by AitSahlia and Lai (2001), we apply the idea of neuro-dynamic programming to develop

(1) nonparametric pricing formulas for actively traded American options, and
(2) simulation-based optimization strategies for complex over-the-counter options, whose optimal stopping problems are prohibitively difficult to solve numerically by standard backward induction algorithms because of the curse of dimensionality.

An important issue in this approach is the choice of basis functions, for which some guidelines and their underlying theory are provided.

This paper is going to be published by *IEEE Transactions in Automatic Control* in 2004.

References

[1] AitSahlia, F. and Lai, T.L. (2001): Exercise boundaries and efficient approximations to American option prices and hedge parameters. *J. Computational Finance* **4**, 85-103.

Approximating Volatilities by Asymmetric Power GARCH Functions

Qiwei Yao
(joint work with Jeremy Penzer and Mingjin Wang)

Let $\{X_t\}$ be a strictly stationary process defined by the volatility model

$$X_t = \sigma_t \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of independent random variables with mean 0, $\sigma_t \geq 0$ is $\mathcal{F}_{t-1}$-measurable, and $\mathcal{F}_{t-1}$ is the σ-algebra generated by $\{X_{t-k}, k \geq 1\}$. Furthermore, ε_t is independent of $\mathcal{F}_{t-1}$. The conventional ARCH/GARCH formulation assumes that the conditional standard deviation σ_t is of the form

$$\begin{aligned} \sigma_t^2 &= \mathrm{Var}\,(X_t|\mathcal{F}_{t-1}) = E(X_t^2|X_{t-1}, X_{t-2}, \cdots) \\ &= c + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2. \end{aligned} \tag{2}$$

where $c > 0$ and b_i, a_j are non-negative. The above model also implies $\mathrm{Var}\,(\varepsilon_t) = 1$. Under the condition $\sum_j a_j < 1$, (2) admits the representation

$$\sigma_t^2 = E(X_t^2|X_{t-1}^2, X_{t-2}^2, \cdots) = d_0 + \sum_{j=1}^{\infty} d_j X_{t-j}^2, \tag{3}$$

where $d_i \geq 0$ are some constants. This suggests that σ_t^2 is the autoregressive function of X_t^2 on its lagged values $X_{t-1}^2, X_{t-2}^2, \cdots$.

On the other hand, there exists the abundance of empirical evidence indicating that for some financial returns the autocorrelation of the squared returns $\{X_t^2\}$, although significant, is often not as strong as, for example the autocorrelation of the absolute returns $\{|X_t|\}$. See, for example, Granger et al (1999) and Rydberg (2000) and the references within. Therefore, instead of modelling the conditional second moments as in (3), Ding, Granger and Engle (1993) proposed to model the conditional γ-th absolute moment of X_t given $\mathcal{F}_{t-1}$ by an asymmetric power GARCH formula, with $\gamma \in (0, 2]$ determined by the data; see (4) below.

In this paper, we do not impose any explicit form on σ_t which is merely assumed to be $\mathcal{F}_{t-1}$-measurable. Instead we seek for an index $\gamma \in (0, 2]$ such that a GARCH-like model for $\{|X_t|^\gamma\}$ provides the *best* approximation for σ_t^γ. More specifically, we approximate σ_t^γ by an asymmetric GARCH function

$$\begin{aligned} \xi_{t,\gamma} &\equiv c + \sum_{i=1}^{p} b_i \{|X_{t-i}| - d_i X_{t-i}\}^\gamma + \sum_{j=1}^{q} a_j \xi_{t-j,\gamma} \\ &= c + \sum_{i=1}^{p} b_i |X_{t-i}|^\gamma \{1 - d_i \operatorname{sgn}(\varepsilon_{t-i})\}^\gamma + \sum_{j=1}^{q} a_j \xi_{t-j,\gamma}, \end{aligned} \tag{4}$$

for any $\gamma \in (0, 2]$, where the parameters c, b_i, a_j are non-negative, and $d_i \in (-1, 1)$. We then choose the γ such that the approximation is *optimum* in certain sense. Equation (4) admits a unique strictly stationary solution

$$\xi_{t,\gamma} = \frac{c}{1 - \sum_{j=1}^{q} a_j} + \sum_{i=1}^{p} b_i |X_{t-i}|^\gamma \{1 - d_i \operatorname{sgn}(\varepsilon_{t-i})\}^\gamma \tag{5}$$

$$+ \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} |X_{t-i-j_1-\cdots-j_k}|^\gamma \{1 - d_i \operatorname{sgn}(\varepsilon_{t-i-j_1-\cdots-j_k})\}^\gamma \tag{6}$$

with $E(\xi_{t,\gamma}) < \infty$, provided that $\{X_t\}$ is strictly stationary with $E|X_t|^\gamma < \infty$, and $\boldsymbol{\theta} \equiv (c, b_1, \cdots, b_p, a_1, \cdots, a_q, d_1, \cdots, d_p)^\tau \in \Theta$, where

$$\Theta = \Big\{ (c, b_1, ..., b_p, a_1, ..., a_q, d_1, ..., d_p) \;\Big|\; c, b_i, a_j > 0,\ d_i \in [-1+\delta_0, 1-\delta_0],\ \sum_{j=1}^{q} a_j < 1 \Big\}, \tag{7}$$

where $\delta_0 > 0$ is a small constant. We restrict d_i in a closed interval contained in $(-1, 1)$ for some technical convenience.

First we consider how to estimate $\boldsymbol{\theta}$, for a given γ. To make σ_t uniquely defined in (1), we always assume that the median of $|\varepsilon_t|$ is equal to 1, unless specified otherwise. Now $\log(|\varepsilon_t|) = \log(|X_t|) - \gamma^{-1} \log(\sigma_t^\gamma)$ are i.i.d. with median 0. Therefore it holds that

$$\sigma_t^\gamma = \arg\min_{a>0} E\big\{ \big| \log|X_t| - \frac{1}{\gamma} \log a \big| \,\big|\, \mathcal{F}_{t-1} \big\}.$$

This leads to an L_1 estimator

$$\widehat{\boldsymbol{\theta}}_1 \equiv \widehat{\boldsymbol{\theta}}_1^{(\gamma)} = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu}^{n} \Big| \log|X_t| - \frac{1}{\gamma}\log\{\xi_{t,\gamma}(\boldsymbol{\theta})\}\Big|.$$

An *approximate* (conditional) Gaussian MLEs may also be entertained based on an additional assumption that ε_t in (1) are independent $N(0,1)$ random variables. This condition implies a different parametrisation since now the median of $|\varepsilon_t|$ is not 1. Note σ_t defined in (1) differs under the two parameterisations by a constant independent of t. This impacts on the parameters in $\xi_{t,\gamma}$ as follows; c and all b_i differ by a common constant under the two parametrisation while d_i and a_j remain unchanged. The resulting estimator is

$$\widehat{\boldsymbol{\theta}}_2 \equiv \widehat{\boldsymbol{\theta}}_2^{(\gamma)} = \arg\min_{\boldsymbol{\theta}} \sum_{t=\nu}^{n} \big[X_t^2/\{\xi_{t,\gamma}(\boldsymbol{\theta})\}^{2/\gamma} + 2\gamma^{-1}\log\{\xi_{t,\gamma}(\boldsymbol{\theta})\}\big]. \tag{8}$$

We note that the method is based on approximating σ_t by $\xi_{t,\gamma}^{1/\gamma}$.

Now we consider the problem of estimating the power index γ. Since our goal is to estimate volatility function σ_t, a good estimation should ensure the residuals $\widehat{\varepsilon}_t = X_t/\widehat{\sigma}_t$ behave like an i.i.d. sequence, or, contain little information on $\mathcal{F}_{t-1}$, where $\widehat{\sigma}_t$ denotes an estimator for σ_t. Let $\widehat{\boldsymbol{\theta}}^{(\gamma)}$ be a reasonable estimator for the parameter $\boldsymbol{\theta} \equiv \boldsymbol{\theta}_\gamma$ of $\xi_{t,\gamma}$. Define residuals

$$\widehat{\varepsilon}_t^{(\gamma)} = X_t/\{\xi_{t,\gamma}(\widehat{\boldsymbol{\theta}}^{(\gamma)})\}^{1/\gamma}, \quad t = \nu, \cdots, n. \tag{9}$$

If $\widehat{\varepsilon}_t^{(\gamma)}$ is a good estimator for ε_t, $E\{\widehat{\varepsilon}_t^{(\gamma)} I(X_{t-j} \le x)\} \approx 0$ for any $j \ge 1$ and x. This suggests to choose $\widehat{\gamma} \in [u_0, 2]$ which minimises

$$R(\gamma) \equiv \sum_{j=1}^{k} \sup_x \frac{1}{n}\Big|\sum_{t=\nu}^{n} \widehat{\varepsilon}_t^{(\gamma)} I(X_{t-j} \le x)\Big|, \tag{10}$$

where $k \ge 1$ is an integer, $u_0 > 0$ is a small constant. We restrict $\widehat{\gamma}$ to be bounded away from 0 for technical convenience. The statistics of this type have been used for model checking by, for example, Stute (1997) and Koul and Stute (1999). In practice, we may use either the least absolute deviations estimator $\widehat{\boldsymbol{\theta}}_1^{(\gamma)}$ or the Gaussian MLE $\widehat{\boldsymbol{\theta}}_2^{(\gamma)}$ as $\widehat{\boldsymbol{\theta}}^{(\gamma)}$ in (9), and we may also standardise $\widehat{\varepsilon}_t^{(\gamma)}$ such that the sample mean and variance are, respectively, 0 and 1.

Under some regularity conditions, we have establish the asymptotic normality for the estimators $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$, and the weak consistency for the estimator $\widehat{\gamma}$. The method has also been illustrated with four sets of financial return data. It is interesting to see that the estimated power index $\widehat{\gamma}$ is often around 1 for those real data sets, leading to better estimation for the volatility function σ_t in comparison with a conventional GARCH fitting.

A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data

Lan Zhang
(joint work with Per A. Mykland and Yacine Aït-Sahalia)

In the analysis of high frequency financial data, a major problem concerns the nonparametric determination of the volatility of an asset return process. A common practice is to estimate volatility from the sum of the frequently-sampled squared returns. Though this approach is justified under the assumption of a continuous stochastic model in an idealized world, it meets the challenge from market microstructure in real applications. We argue that this customary way of estimating volatility is flawed in that it overlooks observation error. The usual mechanism for dealing with the problem is to throw away some data, by sampling less frequently or constructing "time-aggregated" returns from the underlying high frequency asset prices. We propose here a statistically sounder device. Our device is model-free, it takes advantage of the rich sources in tick-by-tick data, and to a great extend it corrects the effect of the microstructure noise on volatility estimation. In the course of constructing our estimator, it becomes clear why and where the "usual" volatility estimator fails when the returns are sampled at high frequency.

Our interest lies in using high frequency intraday data to estimate the integrated volatility over some time periods. To fix the ideas, let $\{S_t\}$ denote the price process of a security, and suppose the log-return process $\{X_t\}$, where $X_t = \log S_t$, follows an Itô process

$$X_t = \mu_t dt + \sigma_t dB_t \tag{1}$$

where B_t is a standard Brownian motion. Typically, σ_t^2, the instantaneous variance (or diffusion coefficient) of the return process $\{X_t\}$, will be stochastic. The parameter of interest is the integrated (cumulative) volatility over one or successive time periods, $\int_0^{T_1} \sigma_t^2 dt$, $\int_{T_1}^{T_2} \sigma_t^2 dt$, ... A natural way to estimate the cumulative volatility over, say, a single time interval from 0 to T, is to use the sum of squared incremental returns,

$$\sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 \approx \int_0^T \sigma_t^2 dt, \tag{2}$$

where the X_{t_i}'s are all the observations of the return process in $[0, T]$. The estimator $\sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2$ is commonly used and generally called "realized volatility" or "realized variance." For a sample of the recent literature in integrated volatility, see Hull and White (1987), Jacod and Protter (1998), Gallant et al. (1999), Chernov and Ghysels (2000), Gloter (2000), Andersen et. al. (2001), Barndorff-Nielsen and Shephard (2001), Mykland and Zhang (2002) and others.

Under model (1), the approximation in (2) is justified by theoretical results in stochastic processes which state that

$$\text{plim} \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 = \int_0^T \sigma_t^2 dt, \tag{3}$$

as the sampling frequency increases. In other words, the estimation error of the realized volatility diminishes. According to (3), realized volatility computed from the highest frequency data ought to provide the best possible estimate for $\int_0^T \sigma_t^2 dt$ the integrated volatility.

However, this is not the general viewpoint from the finance literature. It is generally held there that the returns process X_t should not be sampled too often, regardless of the fact that the asset prices can often be observed with extremely high frequency, such as several times per second. It has been found empirically that the estimator is not robust when the sampling interval is quite small. Issues including bigger bias in the estimate and non-robustness to changes in sampling interval have been reported (see e.g., Brown (1990), Campell et al. (1997), Bai et al. (2000)). The main explanation for this phenomenon is a vast array of issues collectively known as market microstructure, such as, but not limited to, the existence of the bid-ask spread: see Aït-Sahalia et al. (2003) for a description of these phenomena and their grounding in the vast theoretical literature describing the frictions inherent in the trading process. When prices are sampled at finer intervals, microstructure issues become more pronounced. It is then suggested that the bias induced by market microstructure effects makes the most finely sampled data unusable for the calculation, and many authors prefer to sample over longer time horizons to obtain reasonable estimates. The length of the typical choices in the literature is ad hoc and ranges from 5 to 30 minutes for exchange rate data, for instance.

This approach to handling the data poses a conundrum from the statistical point of view. We argue here that sampling over longer horizon merely reduces the impact of microstructure, rather than quantifying and correcting its effect for volatility estimation. And it goes against the grain to throw away data. On the other hand, market microstructure may pose so many problems that subsampling is the only way out.

In this paper we analyze the trade-offs involved in the choice of sampling frequency and develop a method to estimate integrated volatility in such a way as to lessen this conflict. Our contention in the following is that the contamination due to market microstructure is, to first order, the same as what statisticians usually call "observation error". We shall incorporate the observation error into the estimating procedure for integrated volatility. In other words, we shall suppose that the return process as observed at the sampling times is of the form

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{4}$$

Here X_t is a latent true, or efficient, return process, and the ϵ'_{t_i}s are independent noise around the true return. A similar structure was used in the parametric

context where $\sigma_t = \sigma$ is constant by Aït-Sahalia et al. (2003). In that paper, due to the parametric nature of volatility, we proposed likelihood-based corrections for market microstructure.

We show in the paper that, if the data have a structure of the form (4), ignoring microstructure noise would have a devastating effect on the use of the realized volatility. Instead of (2), one gets

$$\sum_{t_i, t_{i+1} \in [0,T]} (Y_{t_{i+1}} - Y_{t_i})^2 = 2nVar(\epsilon) + O_p(n^{1/2}) \tag{5}$$

where the errors ϵ_{t_i}'s are i.i.d. with mean 0, and n is the number of sampling intervals over $[0, T]$. As we will show, ignoring market microstructure noise in the context of stochastic volatility leads to an even more dangerous situation than when σ is constant and $T \to \infty$. The results from equation (5) suggest that the realized volatility does not estimate the true integrated volatility, but rather the variance of the contamination noise. In fact, we will show that the true integrated volatility, which is $O_p(1)$, is even dwarfed by the magnitude of the asymptotically Gaussian $O_p(n^{1/2})$ term in (5).

Of course, the model (4) may also not be correct. When made the basis of inference, it could still occur that one does not wish to sample as frequently as the data would permit. It may, however, make it possible to use substantially larger amounts of data than what would be possible under (2).

In seeking to create an inference procedure under measurement error, we have sought to draw some lessons from the empirical practice that one should not use all the data, while at the same time not violating basic statistical principles. Our approach is built on separating the observations into multiple "grids". We found that the best results can be obtained by combining the usual ("single grid") realized volatility with the multiple grid based device. This gives an estimator which is approximately unbiased. We have also shown how to assess the (random) variance of this estimator, and how to balance the effect in (5) and an effect due to the sampling frequencies.

The theory, including asymptotic distributions, is developed mainly in the context of finding the integrated volatility over one time period; at the end, we extend this to multiple periods. Also, in the case where the noise can be taken to be almost negligible, we provide a way of optimizing the sampling frequency if one wishes to use the classical "realized volatility" or its multi-grid extension.

One important message of the paper: Any time one has an impulse to sample sparsely, one can always do better with a multi-grid method. No matter what the model is, no matter what quantity is being estimated.

References

[1] Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2003): How often to sample a continuous-time process in the presence of market microstructure noise. Technical Report, Princeton University.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001): The distribution of exchange rate realized Volatility. *J. Amer. Statist. Assoc.* **96**, 42-55.

[3] Bai, X., Russell, J.R. and Tiao G.C. (2000): Beyond Merton's utopia I: effects of non-normality and dependence on the precision of variance estimates using high frequency financial data. Technical Report, University of Chicago.
[4] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B* **63**, 167-241.
[5] Brown, S.J. (1990): Estimating volatility. In: *Financial Options: From Theory to Practice*. Eds: S. Figlewski, W.L. Silber and M.G. Subrahmanyam, pp. 516-537. Business One-Irwin, Homewood, IL,
[6] Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997): *The Econometrics of Financial Markets*. Princeton University Press. Princeton, NJ.
[7] Chernov, M. and Ghysels, E. (2000): A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics* **57**, 407-458.
[8] Gallant, A.R., Chien-Te Hsu and Tauchen, G.T. (1999): Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Rev. Econom. Statist.* **81**, 617-631.
[9] A. Gloter (2000): *Estimation des Paramètres d'une Diffusion Cachée*. PhD Thesis Université de Marne-la-Vallée.
[10] Hull, J. and White, A. (1987): The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281-300.
[11] J. Jacod and P. Protter (1998): Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26**, 267-307.
[12] Mykland, P.A. and Zhang, L. (2002): ANOVA for diffusions. Technical Report, University of Chicago, Department of Statistics.

Reporter: Alexander Lindner

Participants

Prof. Dr. Ole E. Barndorff-Nielsen
oebn@imf.au.dk
Dept. of Mathematical Sciences
Aarhus University
Ny Munkegade
DK-8000 Aarhus C

Prof. Dr. Peter J. Brockwell
pjbrock@stat.colostate.edu
Department of Statistics
Colorado State University
Fort Collins CO 80523-1877 – USA

Dr. Boris Buchmann
bbuch@mathematik.tu-muenchen.de
bbuch@ma.tum.de
Zentrum Mathematik
Technische Universität München
Boltzmannstr. 3
D–85747 Garching bei München

Prof. Dr. Ngai Hang Chan
nhchan@sta.cuhk.edu.hk
Department of Statistics
Chinese University of Hong Kong
Risk Management Science Program
Shatin N.T. – Hong Kong

Prof. Dr. Claudia Czado
cczado@ma.tum.de
Lehrstuhl für Mathematische Statistik
Technische Universität München
D–85747 Garching bei München

Prof. Dr. Rainer Dahlhaus
dahlhaus@statlab.uni-heidelberg.de
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
D–69120 Heidelberg

Prof. Dr. Richard A. Davis
rdavis@stat.colostate.edu
Department of Statistics
Colorado State University
Fort Collins CO 80523-1877 – USA

Prof. Dr. Manfred Deistler
Manfred.Deistler@tuwien.ac.at
Institut für Ökonometrie und Operations Research
Technische Universität
Argentinierstraße 8
A-1040 Wien

Dr. Feike C. Drost
F.C.Drost@TilburgUniversity.nl
Department of Econometrics
Tilburg University
P. O. Box 90153
NL-5000 LE Tilburg

Prof. Dr. Paul Embrechts
embrechts@math.ethz.ch
Departement Mathematik
ETH-Zentrum
Rämistr. 101
CH-8092 Zürich

Vicky Fasen
fasen@ma.tum.de
Zentrum Mathematik
TU München
Boltzmannstr. 3
D–85748 Garching bei München

Prof. Dr. Jürgen Franke
franke@mathematik.uni-kl.de
Fachbereich Mathematik
Universität Kaiserslautern
Erwin-Schrödinger-Straße
D–67653 Kaiserslautern

Prof. Dr. Rüdiger Frey
frey@mathematik.uni-leipzig.de
Mathematisches Institut
Universität Leipzig
Augustusplatz 10/11
D–04109 Leipzig

Prof. Dr. Sylvia Frühwirth-Schnatter
sfruehwi@wv-wien.ac.at
Sylvia.Fruehwirth-schnatter@jku.at
Institut für Angewandte Statistik
Johannes Kepler Universität
A-4040 Linz

Prof. Dr. Xin Guo
xinguo@orie.cornell.edu
School of Operations Research and
Industrial Engineering
Cornell University
Rhodes Hall
Ithaca, NY 14853 – USA

Dipl.-Math. Ralph Högn
hoegn@ma.tum.de
Zentrum Mathematik
Technische Universität München
Boltzmannstr. 3
D–85747 Garching bei München

Prof. Dr. Martin Jacobsen
martin@math.ku.dk
Department of Applied Mathematics
and Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen

Dr. Jan Kallsen
kallsen@stochastik.uni-freiburg.de
Kallsen@ma.tum.de
HVB-Institute for Mathematical
Finance
Munich University of Technology
Boltzmannstr. 3
D–85747 Garching

Prof. Dr. Claudia Klüppelberg
cklu@ma.tum.de
cklu@mathematik.tu-muenchen.de
Zentrum Mathematik
TU München
Boltzmannstr. 3
D–85748 Garching bei München

Prof. Dr. Siem Jan Koopman
s.j.koopman@feweb.vu.nl
Department of Econometrics
Vrije University
De Boelelaan 1105
NL-1081 HV Amsterdam

Prof. Dr. Jens-Peter Kreiß
j.kreiss@tu-bs.de
Institut für Mathematische
Stochastik der TU Braunschweig
Pockelsstr. 14
D–38106 Braunschweig

Prof. Dr. Catherine Laredo
catherine.laredo@jouy.inra.fr
Laboratoire de Biometrie
Departement de Mathematiques
I.N.R.A.
Domaine de Velvert
F-78352 Jouy-en-Josas Cedex

Dr. Alexander Lindner
lindner@ma.tum.de
Lehrstuhl für Mathematische
Statistik
Technische Universität München
D–85747 Garching bei München

Prof. Dr. Ross A. Maller
Ross.Maller@anu.edu.au
School of Finance and Applied
Statistics
Australian National University
Acton 0200 – Australia

Prof. Dr. Alexander McNeil
mcneil@math.ethz.ch
Departement Mathematik
ETH-Zentrum
Rämistr. 101
CH-8092 Zürich

Prof. Dr. Thomas Mikosch
mikosch@math.ku.dk
Laboratory of Actuarial Mathematics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen

Dipl.-Math. Gernot Müller
mueller@ma.tum.de
Zentrum Mathematik
Technische Universität München
Boltzmannstr. 3
D–85747 Garching bei München

Prof. Dr. Per Mykland
mykland@galton.uchicago.edu
Department of Statistics
The University of Chicago
5734 University Avenue
Chicago, IL 60637 – USA

Prof. Dr. Michael H. Neumann
mi.neumann@tu-bs.de
Institut für Mathematische
Stochastik der TU Braunschweig
Pockelsstr. 14
D–38106 Braunschweig

Prof. Dr. Serguei Pergamenchtchikov
Serge.Pergamenchtchikov@univ-rouen.fr
Laboratoire de Mathematiques
Raphael Salem
UMR CNRS 6085, UFR Sciences
Univerisite de Rouen
F-76821 Mont Saint Aignan Cedex

Mark Podolskij
podolski@cityweb.de
Fakultät für Mathematik
Ruhr-Universität Bochum
D–44780 Bochum

Dr. Martin Schlather
Martin.Schlather@uni-bayreuth.de
schlather@cu.lu
Laboratoire de Mathematique
Universite du Luxembourg
162 A, avenue de la Faiencerie
L-1511 Luxembourg

Prof. Dr. Neil Shephard
neil.shephard@nuffield.ox.ac.uk
Nuffield College
Oxford University
GB-Oxford OX1 1NF

Prof. Dr. Richard L. Smith
rls@email.unc.edu
Statistics Department
University of North Carolina
Chapel Hill, NC 27599-3260 – USA

Prof. Dr. Michael Sorensen
michael@math.ku.dk
Department of Applied Mathematics
and Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen

Prof. Dr. Vladimir Spokoiny
spokoiny@wias-berlin.de
WIAS
Mohrenstr. 39
D–10117 Berlin

Prof. Dr. Catalin Starica
starica@math.chalmers.se
Department of Mathematics
Chalmers University of Technology
and University of Göteborg
Eklandag. 86
S-412 96 Göteborg

Prof. Dr. J.Michael Steele
steele@wharton.upenn.edu
Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6302 – USA

Dr. Daniel Straumann
strauman@math.ethz.ch
Mathematik Departement
RiskLab, H6F 42.3
ETH-Zentrum
Rämistr. 101
CH-8092 Zürich

Dr. Alex Szimayer
aszimaye@ecel.uwa.edu.au
The University of Western Australia
Accounting & Finance
UWA Business School
35 Stirling Highway
Crawley WA 6009 – Australia

Mark Van de Vyver
mvdv@spamcop.net
School of Business (H69)
The University of Sydney
Cnr Rose and Codrington Ave.
Sydney 2006 – Australia

Prof. Dr. Yazhen Wang
yzwang@stat.uconn.edu
Department of Statistics
University of Connecticut
Box U-4120
Storrs, CT 06269-3120 – USA

Prof. Dr. Samuel P. Wong
imsam@ust.hk
Department of Information and
Systems Management
School of Business and Management
Clear Water Bay
Kowloon – Hong Kong

Prof. Dr. Qiwei Yao
q.yao@lse.ac.uk
Department of Statistics
London School of Economics
Houghton Street
GB-London,WC2A 2AE

Prof. Dr. Lan Zhang
lzhang@stat.cmu.edu
Dept. of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213 – USA