STATISTICAL ASPECTS OF MODELLING AND PREDICTION FOR FINANCIAL TIME SERIES

N. H. BINGHAM

Essen, Wednesday 21 November, 2012

Aims:
(a) to unify "BFK" and "BIK" [below];
(b) to provide a statistical complement to BIK.
Sources:

BFK

[BK1] NHB, RK: Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance* **2** (2002), 241-250, MR1922404.
[BKS] NHB, RK & R. Schmidt): A semi-parametric approach to risk management. *Quantitative Finance* **3** (2003), 426-441, MR2026570.
[BFK] B, J. M. FRY & K: Multivariate elliptic processes. *Stat. Neerl.* **64** (2010), 352-366;

BIK

[B1] NHB, Szegö's theorem and its probabilistic descendants. *Probability Surveys* **9** (2012), 287-324;

[B2] NHB, Multivariate prediction and matrix Szegö theory. *Probability Surveys* **9** (2012), 325-339.

[B3] NHB, Modelling and prediction of financial time series. *Comm. Stat.: Theory and Methods*, 2012+.

[BIK] NHB, A. Inoue & Y. Kasahara: An explicit representation of Verblunsky coefficients. *Statistics and Probability Letters* **82**.2 (2012), 403-410, MR2875229.

[KB] Y. Kasahara & NHB: Verblunsky coefficients and Nehari sequences. *TAMS*, to appear.

1. Semi-parametric models

$S = (S_t)$, discrete time t, a d-vector of discounted prices $S_i(t)$ of risky assets.

Discounting:

(a) to achieve stationarity;

(b) in math. finance, discount everything, and take conditional expectations under the equivalent martingale measure (EMM – or risk-neutral measure). See e.g. [BK], Preface.

Markowitz (1952):

(i) think of risk (covariance matrix Σ) and return (mean vector μ together, not separately;

(ii) diversify: hold a large number d of assets, with lots of negative correlation.

Thus any model for asset prices needs (μ, Σ) – a parametric component.

We restrict to Σ positive definite (so invertible) – the generic case.

Standardisation: $X_t := \Sigma^{-\frac{1}{2}}(S_t - \mu)$: $X = (X_t)$ has mean 0 and cov. I.

§2. Multivariate elliptic processes (MEP)

X above is spherically symmetric. Then we can reduce to the quadratic form

$$Q := \|X_t\|^2 = X_t^T X_t = (S_t - \mu)^T \Sigma^{-1} (S_t - \mu),$$

which is in the *elliptical family* [BFK]. We assume X is of the form

$$X_t - \mu := R_t A^T U_t = R_t \Sigma^{\frac{1}{2}} U_t, \qquad (MEP)$$

where Σ has Cholesky decomposition $\Sigma = A^T A$ (so $A = \Sigma^{\frac{1}{2}}$, the usual matrix square root of the positive definite matrix Σ), $U = (U_t)$ is Brownian motion on the d-dimensional sphere, and $R = (R_t)$ is the *risk driver* (one-dimensional). X is a MEP [BFK]. From (MEP),

$$var(X_t | R_t) = R_t^2 \Sigma, \quad var(X_t) = E[R_t^2] \Sigma.$$

This gives a simple *stochastic volatility (SV)* model! As large or small values of R tend to be followed by large or small values of R, this gives *volatility clustering* – one of the stylized

facts of mathematical finance.
Estimation of parametric part (μ, Σ).
μ: imprecise – subject to *mean blur* (Merton, 1980; Luenberger, 1998, §8.5). Work robustly (e.g., Oja median).
Σ: robustness; affine equivariance; Lopuhaä & Rousseeuw, AS 1991.
Estimation of non-parametric part.
(i) MEP, R an ergodic diffusion [BFK]. Estimate the stationary density from

$$R_t^2 = Q$$

and density estimation. Cf.
[Kut] Yu. A. Kutoyants, *Statistical inference for ergodic diffusion processes*. Springer, 2004.
(ii) MEP, $R \in SD$, the class of *self-decomposable* laws. These are the limit laws as $t \to \infty$ of solutions of SDEs

$$dR_t = -cR_t dt + dZ_t, \qquad (OU)$$

of Ornstein-Uhlenbeck (OU) type, with driving noise $Z = (Z_t)$ a subordinator (positive Lévy process), $c > 0$ ($c = 1$ if convenient). Theory: see Sato §15-17 and §33, [BFK] §3:

[Sat] K.-I. Sato, *Lévy processes and infinitely divisible distributions*. CUP, 1999.

Estimation: see

[JonMV] G. Jongbloed and F. H. van der Meulen, Parametric estimation for subordinators and induced OU processes. *Scand. J. Stat.* **33** (2006), 825-847.

[JonM] G. Jongbloed, F. H. van der Meulen and A. W. van der Vaart, Non-parametric inference for Lévy-driven OU processes. *Bernoulli* **11** (2005), 759-791.

[BKRW] P. J. BICKEL, C. A. J. KLAASSEN, Y. RITOV & J. WELLNER, *Efficient and adaptive estimation for semiparametric models*, 2nd ed., Springer, 1998.

3. Prediction in general: Szegö theory.
The basis of the prediction theory of stationary time-series is the *Kolmogorov Isomorphism Theorem (KIT)* ([Kol]; see e.g. [B1], §2, scalar case, [B2], §2, vector case). There is a random measure Y with orthogonal increments, the *Cramér process* or *spectral process* (Cramér 1942, Cramér & Leadbetter 1967, §7.5) and a probability measure m on the unit circle T, the *spectral measure*, plus an isomorphism

$$X_n \leftrightarrow e^{in.}$$

between the Hilbert spaces $\mathcal{H}$ (the L_2-space of the process $X = (X_t)$) and $L_2(m)$, which maps between the *time domain* on the left and the *frequency domain* on the right. One has the *Cramér (spectral) representation*

$$X_n = \int_T e^{in\theta} dY(\theta), \qquad (CR)$$

$$E[(dY(\theta))^2] = dm(\theta).$$

4. ACF and PACF

Also from KIT: taking $E[X_n] = 0$, $var(X_n) = 1$ for simplicity, the autocorrelation function (ACF) $\gamma = (\gamma_n)$ is given by

$$\gamma_n := E[X_n \bar{X}_0] = \int_T e^{-in\theta} dm(\theta).$$

Partial autocorrelation function (PACF): $\alpha = (\alpha_n)$, where α_n is the correlation between the residuals at times 0, n regressed on the intermediate values.

ACF: cut-off for $MA(q)$

PACF: cut-off for $AR(p)$.

The PACF gives an *unrestricted parametrization*: all values α_n in the unit disc D are possible, and

$$\alpha \leftrightarrow m$$

is a bijection between D^∞ and $P(T)$, the space of probability measures on T. This is *Verblunsky's theorem* of 1935-6 (rediscovered in statistics, by Barndorff-Nielsen & Schou 1973,

F. L. Ramsey, AS 1974). The PACF (matrix-valued in the vector case) is the sequence of diagonals in the infinite triangular matrix of finite-predictor coefficients (Levinson-Durbin algorithm).

Theory: orthogonal polynomials on the unit circle (OPUC, [B1]); matrix orthogonal polynomials on the unit circle (MOPUC, [B2]).

[Sim] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory. Part 2: Spectral theory*. AMS Colloq. Publ. 54.1, 54.2, AMS, 2005.

The Levinson-Durbin algorithm is the three-term recurrence relation in OPUC/MOPUC.

Estimation of PACF: see e.g. Dégerine, IEEE 1993, J. Multiv. Anal. 1994.

Estimation of m: frequency-domain or spectral methods in Time Series: C. W. J. Granger & M. Hatanaka; E. J. Hannan; M. B. Priestley; B. G. Quinn.

By Verblunsky's theorem, we have a choice here!

5. Szegö's theorem

The *one-step prediction error*

$$\sigma^2 := E[(X_0 - P_{(-\infty,-1]}X_0)^2]$$

has $\sigma > 0$ in the *non-deterministic* ('good') case, $\sigma = 0$ in the *deterministic* ('bad') case. The *Wold decomposition* $X = U + V$ gives X as the sum of a non-deterministic U and a deterministic V:

$$X_n = U_n + V_n;$$

U is a moving average,

$$U_n = \sum_{j=-\infty}^{n} m_{n-j}\xi_j = \sum_{k=0}^{\infty} m_k \xi_{n-k},$$

ξ_j zero-mean and uncorrelated, with each other and with V; $E[\xi_n] = 0$, $var(\xi_n) = E[\xi_n^2] = \sigma^2$. So when $\sigma = 0$ $\xi_n = 0$, $U = 0$ and X is deterministic. When $\sigma > 0$, the spectral measures of U_n, V_n are μ_{ac} and μ_s, the absolutely continuous and singular components of μ (again, the 'good' and 'bad' parts). Think of ξ_n as the 'innovation' at time n – the new random input, a measure of the unpredictability of the present from the past. This is only present when $\sigma > 0$; when $\sigma = 0$, the present is determined by the past – even by the remote past.

Szegö's Theorem.

(i) $\sigma > 0$ iff $\log w \in L_1$, that is,

$$\int -\log w(\theta) d\theta > -\infty. \qquad (Sz)$$

(ii) $\sigma > 0$ iff $\alpha \in \ell_2$.

(iii)

$$\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2),$$

so $\sigma > 0$ iff the product converges, i.e. iff

$$\sum |\alpha_n|^2 < \infty : \qquad \alpha \in \ell_2;$$

(iv) σ^2 is the geometric mean $G(\mu)$ of μ:

$$\sigma^2 = \exp(\frac{1}{2\pi} \int \log w(\theta) d\theta) =: G(\mu) > 0. \quad (K)$$

((i)-(iii): Szegö, 1915, 1920, 1921; (iv): Kolmogorov, 1941).

Under (Sz), the *Szegö function*

$$h(z) := \exp(\frac{1}{4\pi} \int (\frac{e^{i\theta} + z}{e^{i\theta} - z}) \log w(\theta) d\theta) \qquad (z \in D) \qquad (OF)$$

has $h \in H_2$ (Hardy space of order 2);
h is an *outer function*;

$$|h(e^{i\theta})|^2 = w(\theta)$$

(h is an 'analytic square root' of w).
We usually assume not only (Sz) ('nice component present'), but also that the remote past is trivial:

$$\mathcal{H}_{-\infty} := \bigcap_{n=-\infty}^{\infty} \mathcal{H}_n = \{0\} \qquad (PND)$$

('nasty component absent'). The process is then called *purely non-deterministic (PND)*:

$$(PND) = (ND) + (\mu_s = 0) = (Sz) + (\mu_s = 0)$$

$$= (\sigma > 0) + (\mu_s = 0) \qquad (PND)$$

6. Discrete and continuous time

In (CR), the process (X_n) in discrete time corresponds to the Cramér process Y. Replacing integer time n by continuous time t in (CR),

$$X_t := \int_T e^{it\theta} dY(\theta), \qquad (CR)$$

defines a process $X = (X_t)$ in continuous time, interpolating (X_n) at integer times. This (X_t) is very smooth: it is a random entire function of exponential type π, by the Paley-Wiener theorem. This is an instance of the *sampling theorem*: under suitable conditions, we can recover a continuous-time signal from a discrete-time signal, sampled frequently enough (at at least the *Nyquist rate*). The Nyquist rate is attained here (rate 1: integers 1 apart, circle has length 2π).

The familiar ARMA (Box-Jenkins) models in discrete time have counterparts in CARMA models in continuous time (see e.g. P. J. Brockwell and co-workers). Similarly, the GARCH processes in discrete time have COGARCH analogues (see e.g. C. Klüppelberg and co-workers). Econometric data is usually gathered in discrete time. But there is an extensive theory in continuous time; see e.g.

[Berg] A. R. Bergstrom, *Continuous-time econometric modelling*. Oxford University Press, 1990.

The BFK approach via MEP is in continuous time, and gives stochastic volatility (SV) – volatility clustering. The BIK approach using (CR) takes continuous time in its stride, but not volatility clustering. By contrast, COGARCH enables one to model SV explicitly, but is more complicated than its discrete-time counterpart, GARCH.

7. Stationarity v. non-stationarity

All three models above ('MEP-Lévy, MEP-diffusion and Szegö') depend on stationarity. This is a strong assumption! One of the great themes of the Nobel Prize winner Sir Clive Granger was to warn one not to use methods based on stationarity in non-stationary situations. This can lead, via *spurious regression*, to misleading expert advice to politicians, hence to mistaken macroeconomic policies, and hence to massive and irreversible losses in GDP!

Recall also (§1; [BK], Preface) that one *discounts* to use the standard risk-neutral valuation theory of mathematical finance.

But, the risk-free interest rate r that one discounts by varies over time; there are several relevant rates (Bank rate, Libor rate, ...), etc.

So: discounting, though mathematically trivial and convenient, is problematic in practice on real data, particularly econometric or financial data over long time periods.

One has several choices:
(i) Discount anyway, as best one can.
(ii) Avoid discounting, by using a non-stationary extension of the theory above. E.g., KIT extends, but now with a spectral *bimeasure* in place of a spectral measure (two arguments: we now need two time arguments, rather than one).
(iii) 'Split the difference': use *local stationarity*. See e.g. R. Dahlhaus and co-workers.
(iv) Use time-frequency methods. See e.g. R. CARMONA, W.-L. HIANG & B. TORRÉSANI: *Practical time-frequency analysis: Gabor and wavelet transforms, with an implementation in S*. Acad. Press, 1998.

Comparison of methods; data analysis.
Work in progress! The aim is to compare how well the various approaches fit real data. This would even be interesting in one dimension - but much more so in c. 50, say: c. 12 economic sectors, and c. 4 firms per sector. NHB