smfd13.tex

Day 13. 28.11.2014.

Szegö's theorem.

By *Szegö's theorem* (Gabor SZEGÖ (1895-1985) in 1915), the deterministic component in the Wold decomposition is absent (the 'nice case') iff

$$\int_0^{2\pi} \log w(\theta) d\theta > -\infty,$$

where w is the density of the spectral measure μ of the process (the logarithm of the density enters here in connection with the concept of *entropy*, which arises in Statistical Mechanics and Thermodynamics).

Hidden frequencies.

Spectral (or Fourier) methods are specially well adapted for searching for hidden frequencies. They can be traced back to work of Lord Kelvin on tides, and to work of Sir Arthur SCHUSTER (1851-1934) of 1897 and 1906 on sunspots (which show a periodicity of around 11 years). They are widely used for detecting the chemical composition of stars from analysing the frequencies found in starlight. An obviously relevant area for Math. Finance is analysing the *business cycle*. Under normal economic conditions (pre-Crash of 2007-8), economic life showed a natural rhythm, in which business activity tended to increase, leading to expansion (and eventually overheating) of the economy (employment increasing, and wages increasing as employers competed for labour), followed by contraction (and eventually depression) od the economy (employment and wages falling). The authorities would try to control this by increasing interest rates to slow the economy down (making it more expensive for firms to borrow to invest), and decreasing interest rates to stimulate the economy. Note that this has not applied since the Crash: we have had long periods of near-zero interest rates, combined with economic depression. The Japanese had even worse experiences, in the 1990s and later (the 'lost decade', or decades).

Wavelets.

One can combine time domain and frequency domain ('time-frequency analysis') by using *wavelets* (1980s on)[1]; we must omit details.

[1] Wavelets are a speciality of the Statistics Section here at Imperial College

10. ARCH and GARCH; Econometrics ([BF, 9.4.1, 220-222)

There are a number of *stylised facts* in mathematical finance. E.g.:
(i). Financial data show *skewness*. This is a result of the asymmetry between profit and loss (large losses are lethal!)
(ii). Financial data have much *fatter tails* than the normal/Gaussian (I.5).
(iii) Financial data show *volatility clustering*. This is a result of the economic and financial environment, which is extremely complex, and which moves between good times/booms/upswings and bad times/slumps/downswings. Typically, the market 'gets stuck', staying in its current state for longer than is objectively justified, and then over-correcting. As investors are highly sensitive to losses (see (i) above), downturns cause widespread nervousness, which is reflected in higher volatility. The upshot is that good times are associated with periods of growth but low volatility; downturns spark extended periods of high volatility (and stagnation, or shrinkage, of the economy).

ARCH and GARCH.

We turn to models that can incorporate such features (volatility clustering, etc.).

The model equations are (with Z_t ind. $N(0,1)$)

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \alpha_0 + \sum_1^p \alpha_i X_{i-1}^2, \qquad (ARCH(p))$$

while in $GARCH(p,q)$ the σ_t^2 term becomes

$$\sigma_t^2 = \alpha_0 + \sum_1^p \alpha_i X_{i-1}^2 + \sum_1^q \beta_j \sigma_{t-j}^2. \qquad (GARCH(p,q))$$

The names stand for (generalised) autoregressive conditionally heteroscedastic (= variable variance). These are widely used in Econometrics, to model *volatility clustering* – the common tendency for periods of high volatility, or variability, to cluster together in time. See e.g. Harvey 8.3, [BF] 9.4, [BFK].

Integrated processes.

One standard technique used to reduce non-stationary processes to the stationary case is to *difference* them repeatedly (one differencing operation replaces X_t by $X_t - X_{t-1}$). If the series of dth differences in stationary but that of $(d-1)$th differences is not, the original series is said to be *integrated of order d*; one writes $(X_t) \sim I(d)$.

Co-integration.

If $(X_t) \sim I(d)$, we say that (X_t) is *cointegrated* with *cointegration vector* α if $\alpha^T X_t)$ is (integrated of) order less than d.

A simple example arises in random walks. If $X_n = \sum_{i=1}^n \xi_i$ with ξ_i iid random variables, $Y_n = X_n + \epsilon_n$ is a noisy observation of X_n, then $(X,Y) = (X_n, Y_n)$ is cointegrated of order 1, with coint. vector $(-1,1)^T$.

Cointegrated series are series that move together, and commonly occur in economics. These concepts arose in econometrics, in the work of R. F. ENGLE (1942-) and C. W. J. (Sir Clive) GRANGER (1934-2009) in 1987. Engle and Granger gave (in 1991) an illustrative example – the price of tomatoes in North Carolina and South Carolina. These states are close enough for a significant price differential between the two to encourage sellers to transfer tomatoes to the state with currently higher prices to cash in; this movement would increase supply there and reduce it in the other state, so supply and demand would move the prices towards each other.

Engle and Granger received the Nobel Prize in Economics in 2003. The citation included the following: "Most macroecomomic time series follow a stochastic trend, so that a temporary disturbance in, say, GDP has a long-lasting effect. These time-series are called non-stationary; they differ from stationary series which do not grow over time, but fluctuate around a given value. Clive Granger demonstrated that the statistical methods used for stationary time series could yield wholly misleading results when applied to the analysis of nonstationary data. His significant discovery was that specific combinations of nonstationary time series may exhibit stationarity, thereby allowing for correct statistical inference. Granger called this phenomenon cointegration. He developed methods that have become invaluable in systems where short-run dynamics are affected by large random disturbances and long-run dynamics are restricted to economic equilibrium relationships. Examples include the relations between wealth and consumption, exchange rates and price levels, and short- and long-term interest rates."

Spurious regression.

Standard least-squares method work perfectly well if they are applied to *stationary* time series. But if they are applied to *non-stationary* time series, they can lead to spurious or nonsensical results. One can give examples of two time series that clearly have nothing to do with each other, because they come from quite unrelated contexts, but nevertheless have a high value of R^2. This would normally suggest that a correspondingly high propertion of the variability in one is accounted for by variability in the other – while in fact *none* of the variability is accounted for. This is the phenomenon of *spurious regression*, first identified by G. U. YULE (1871-1851) in 1927, and later studied by Granger and Newbold in 1974. We can largely avoid such pitfalls by restricting attention to stationary time series, as above.

From Granger's obituary (*The Times*, 1.6.2009): "Following Granger's arrival at UCSD in La Jolla, he began the work with his colleague Robert F. Engle for which he is most famous, and for which they received the Bank of Sweden Nobel Memorial Prize in Economic Sciences in 2003. They developed in 1987 the concept of cointegration. Cointegrated series are series that tend to move together, and commonly occur in economics. Engle and Granger gave the example of the price of tomatoes in North and South Carolina .... Cointegration may be used to reduce non-stationary situations to stationary ones, which are much easier to handle statistically and so to make predictions for. This is a matter of great economic importance, as most macroeconomic time series are non-stationary, so temporary disturbances in, say, GDP may have a long-lasting effect, and so a permanent economic cost. The Engle-Granger approach helps to separate out short-term effects, which are random and unpredictable, from long-term effects, which reflect the underlying economics. This is invaluable for macroeconomic policy formulation, on matters such as interest rates, exchange rates, and the relationship between incomes and consumption."

Endogenous and exogenous variables.

The term 'endogenous' means 'generated within'. The $ARCH$ and $GARCH$ models above show how variable variance (or volatility) can arise in such a way. By contrast, 'exogenous' means 'generated outside'. Exogenous variables might be the effect in a national economy of international factors, or of the national economy on a specific firm or industrial sector, for example. Often, one has a vector autoregressive (VAR) model, where the vector of variables is partitioned into two components, representing the endogenous and exogenous variables. For monograph treatments in the econometric setting, see e.g. [G], [GM].

Discrete and continuous time.

While econometric data arrives discretely (monthly trade figures, daily closing prices for stocks, etc.), continuous time is more convenient for dynamic models of the economy. See e.g.

A. R. BERGSTROM: *Continuous-time econometric modelling*, OUP, 1990.

11. State-space models and the Kalman filter

State-space models originate in Control Engineering. This field goes back to the governor on a steam engine (James WATT (1736-1819) in 1788): to prevent a locomotive going too fast, the governor (a rotating device mounted on top of the engine) rose under centrifugal force as the speed increased, thus operating a valve to reduce the steam entering the cylinders. This was an early form of *feedback control*.

The Kalman filter (Rudolf KALMAN (1930-) in 1960) was a device for online (or real-time) control, suitable for use with linear systems, quadratic loss and Gaussian errors (LQG) (the term *filter* is used because one 'filters out' the noise from the signal to reveal the best estimate of the state). This appeared just when it was needed, for online control of manned spacecraft during the 60s. We shall not develop the control aspects here; see e.g.
M. H. A. DAVIS, *Linear estimation and stochastic control*, Chapman & Hall, 1977,
M. H. A. DAVIS & R. B. VINTER, *Stochastic modelling and control*, Chapman & Hall, 1985.
But the power of the method even without control can be seen in applications such as to *mortar-locating radar*[2]. We follow Whittle ([W], Th. 12.5.2); cf. [BD1] Ch. 12, [BD2] Ch. 8.

The Kalman filter has been extensively applied in Time Series, financial and otherwise. We cited Harvey's *Time series models* in D0; see also
A. C. HARVEY, *Forecasting, structural time series models and the Kalman filter*, CUP, 1991.

Before proceeding to technicalities, we stress one great advantage of the Kalman filter and state-space methods: *they do not depend on stationarity*. Most processes encountered in economics and finance – and indeed, in life generally – are *not* stationary. One can induce stationarity by two main methods: *differencing* (as in the Box-Jenkins ARMA/ARIMA approach – rather brutal), or *discounting* (as in the Black-Scholes approach to option pricing, to get the EMM – but interest rates vary!). State-space models are more direct.

With the engineering example in mind for definiteness, suppose that the *state* of the system at time n is represented by some p-vector $x(n)$. Although the state x is what we are interested in, we cannot observe it directly; what we can observe is a *signal*, or *output* y, or $y(n)$ at time n, a q-vector. We apply a *control* $u(n-1)$, based on information $\mathcal{F}_{n-1}$ available at time $n-1$. The dynamics are represented by the following two equations, the *state equation*

[2] Used in, e.g., the Siege of Sarajevo, 1992-96.

(SE) and the *observation equation* (OE):

$$x(n) = A(n-1)x(n-1) + B(n-1)u(n-1)\epsilon(n-1), \qquad (SE)$$

$$y(n) = C(n)x(n) + \eta(n). \qquad (OE)$$

Here $A(.), B(.), C(.)$ are known matrices. The errors $\epsilon(.)$, $\eta(.)$ are p- and q-vectors respectively, with means 0; the errors at different times are all uncorrelated (= independent, if the errors are Gaussian, as we may assume here); the covariance matrices are known matrices

$$cov(\epsilon(n)) = N(n), \qquad cov(\eta(n)) = M(n), \qquad cov(\epsilon(n), \eta(n)) = L(n),$$

In the motivating trajectory example, $A(.)$ comes from the dynamics of the vehicle being tracked, $C(.)$ from the properties of the tracking equipment, $B(.)$ from the control mechanism.

For simplicity, we restrict to the case where $A(n) = A$ for all n, and similarly for B and C; there is no difficulty in extending to the general case.

We write $\hat{x}(n)$ for the best linear predictor (in the sense of minimising expected squared error) of $x(n)$ given the information $\mathcal{F}(n)$ available at time n.

Theorem (Kalman filter). (i) The conditional distribution of $x(n)$ given $\mathcal{F}(n)$ is $N(\hat{x}(n), V(n))$, where the covariance matrix $V(n)$ is given by the Kalman recursion in (iii) below.
(ii) (*Kalman filter*). The estimate $\hat{x}(n)$ is given by the recursion (updating relation)

$$\hat{x}(n) = A\hat{x}(n-1) + Bu(n-1) + H(n)(y(n) - C\hat{x}(n-1)),$$

where

$$H(n) := (L + AV(n-1)C^T)(M + CV(n-1)C^T)^{-1}.$$

(iii) (*Kalman recursion*). The covariance matrix $V(n)$ is given by the recursion (updating relation)

$$V(n) = N+AV(n-1)A^T-(L+AV(n-1)C^T)(M+CV(n-1)C^T)^{-1}(L^T+CV(n-1)A^T).$$

Proof. (i) We start with $x(0) \sim N(\hat{x}(0), V(0))$. That $x(n)|\mathcal{F}(n) \sim N(\hat{x}(n), V(n))$ is clear from IV.6 on conditioning and regression for the multinormal, and is also proved by induction from the recursions (ii), (iii) below.

Write the estimation error as $\Delta(n) := x(n) - \hat{x}(n)$; then $V(n) = cov(\Delta(n))$. Now

$$\begin{aligned}
\Delta(n-1) &= x(n-1) - \hat{x}(n-1) \\
\epsilon(n-1) &= x(n) - Ax(n-1) - Bu(n-1) \\
\eta(n) &= y(n) - Cx(n-1)
\end{aligned}$$

are jointly normal with mean 0 and covariance matrix

$$\begin{pmatrix} V & 0 & 0 \\ 0 & N & L \\ 0 & L^T & M \end{pmatrix},$$

where for convenience we write V for $V(n-1)$. We now replace $x(n-1)$ (unobservable) by $\hat{x}(n-1) + \Delta(n-1)$ (we know the first, and know the covariance V of the second), and define

$$\begin{aligned}
\zeta^*(n) &:= x(n) - A\hat{x}(n-1) - Bu(n-1) \\
&= x(n) - Ax(n-1) - Bu(n-1) + A(x(n-1) - \hat{x}(n-1)) \\
&= \epsilon(n) + A\Delta(n-1),
\end{aligned}$$

$$\begin{aligned}
\zeta(n) &:= y(n) - C\hat{x}(n-1) \\
&= y(n) - Cx(n-1) + C(x(n-1) - \hat{x}(n-1)) \\
&= \eta(n) + C\Delta(n-1).
\end{aligned}$$

Then

$$\begin{pmatrix} \zeta^*(n) \\ \zeta(n) \end{pmatrix} = \begin{pmatrix} A\Delta(n-1) + \epsilon(n) \\ C\Delta(n-1) + \eta(n) \end{pmatrix} = \begin{pmatrix} A & 1 & 0 \\ C & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta(n-1) \\ \epsilon(n) \\ \eta(n) \end{pmatrix} \sim N(0, \Sigma),$$

where the covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} A & 1 & 0 \\ C & 0 & 1 \end{pmatrix} \begin{pmatrix} V & 0 & 0 \\ 0 & N & L \\ 0 & L^T & M \end{pmatrix} \begin{pmatrix} A^T & C^T \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A & 1 & 0 \\ C & 0 & 1 \end{pmatrix} \begin{pmatrix} VA^T & VC^T \\ N & L \\ L^T & M \end{pmatrix}:$$

$$\Sigma = \begin{pmatrix} N + AVA^T & L + AVC^T \\ L^T + CVA^T & M + CVC^T \end{pmatrix}.$$

Both (ii) (conditional means) and (iii) (conditional variances) now follow from the normal Conditioning Theorem of IV.6, D9. //

It is difficult to overestimate the practical importance of this key result. It has proved invaluable in many areas since its introduction in 1960.

Extensions.

1. *Non-Gaussian errors.*

The result extends beyond the context of Gaussian errors (multivariate normal distribution) above. One does not obtain the full distribution, but works instead with means and variances. See e.g. Whittle [W], 12.8 and Th. 12.9.4; see also the Bayes linear estimate of VII.7.8 D19.

2. *Prediction further into the future.*

The method above can be readily adapted to prediction k time-steps into the future. This is done in detail in [BD2], 12.3.

3. *Smoothing.*

Instead of predicting the future, one can instead seek to get the best fit we can to the data. The mathematics is very similar; see e.g. [BD2], Prop. 12.2,3, 4.

4. *Riccati equation.*

The non-linear recursion (iii) is a matrix *Riccati equation*, and this name is often used instead of Kalman recursion.

5. *Off-line calibration.*

To use a Kalman filter, one needs the relevant matrices, A, B, C, L, M, N. In practice, these will have to be estimated numerically. This can be done off-line, 'at leisure'. Once accurate (enough) numerical estimates of these matrices are known, and the recursions (ii) and (iii) programmed, the filter can be used online (in real time).

6. *Innovations.*

The *innovations* are $I(n) := y(n) - C\hat{x}(n-1)$. These are the differences between an observation $y(n)$ and the prediction $C\hat{x}(n-1)$ we would have made for it at time $n-1$ (from the observation equation (OE)). This is the new information at time n – beyond what we could have predicted. They are mutually uncorrelated (independent, in the Gaussian case, as here). One can base the theory on them ([W], 12.7).

7. *Continuous time.*

One can work instead in continuous time, where the recurrence (or difference) equations above are replaced by differential equations. See e.g. [W], Ch. 20. Hence the name Riccati – Riccati's differential equation.