smfd11.tex

Day 11. 13.6.2012

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). We have discussed this in I.5.

3(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 (as well as stagnation, or shrinkage, of the economy).

$ARCH$ and $GARCH$. We turn to models that can incorporate such features.

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

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

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

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

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].

Cointegration and spurious regression.

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 of cointegration arises in random walks. Suppose $X_n = \sum_{i=1}^{n} \xi_i$ with ξ_i iid random variables, and $Y_n = X_n + \epsilon_n$, with the ϵ_n iid errors as above, is a noisy observation of X_n. Then the bivariate process $(X, Y) = (X_n, Y_n)$ 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 Ecomomics 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 or 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 d 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.

C. GOURIÉROUX, *ARCH models and financial applications*, Springer, 1997,
C. GOURÉROUX and A. MONFORT, *Time series and dynamic modles*, CUP, 1990.

VII. MULTIVARIATE ANALYSIS

1. PRELIMINARIES: MATRIX THEORY.

Modern Algebra splits into two main parts: Groups, Rings and Fields on the one hand, and Linear Algebra on the other. Linear Algebra deals with *linear transformations* between *vector spaces*. We confine attention here to the *finite-dimensional* case; the infinite-dimensional case needs Functional Analysis and is harder. Broadly, Parametric Statistics can be handled in finitely many dimensions, Non-Parametric Statistics needs infinitely many.

Given a finite-dimensional vector space V, we can always choose a *basis* (a maximal set of linearly independent vectors). All such bases contain the same number of vectors; if this is n, the vector space has *dimension* n.

Given two finite-dimensional vector spaces and a linear transformation α between the two, choice of bases $(e_1, \ldots, e_m)$ and $(f_1, \ldots, f_n)$ determines a *matrix* $A = (a_{ij})$ by

$$e_i\alpha = \sum_{j=1}^{n} a_{ij} f_j \qquad (i = 1, \ldots, m).$$

We write

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix},$$

or $A = (a_{ij})$ more briefly. The a_{ij} are called the *elements* of the matrix; we write A $(m \times n)$ for A (m rows, n columns).

Matrices may be subjected to various operations:

1. *Matrix addition.* If $A = (a_{ij})$, $B = (b_{ij})$ have the same size, then

$$A \pm B := (a_{ij} \pm b_{ij})$$

(this represents $\alpha \pm \beta$ if α, β are the underlying linear transformations).

2. *Scalar multiplication.* If $A = (a_{ij})$ and c is a scalar (real, unless we specify complex), then the matrix

$$cA := (ca_{ij})$$

represents $c\alpha$.

3. *Matrix multiplication.* If A is $m \times n$, B is $n \times p$, then $C := AB$ is $n \times p$,

where $C = (c_{ij})$ and

$$c_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}$$

(this represents the product, or composition, $\alpha\beta$ or $x \mapsto x\alpha\beta$).
Note. Matrix multiplication is non-commutative! – $AB \neq BA$ in general, even when both are defined (which can only happen for A, B square of the same size).
Partitioning.

We may *partition* a matrix A in various ways. for instance, A as above partitions as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} is $r \times s$, A_{12} is $r \times (n-s)$, A_{21} is $(m-r) \times s$, A_{22} is $(m-r) \times (n-s)$, etc. In the same way, A may be partitioned as
(i) a column of its rows; (ii) a row of its columns.
Rank.

The maximal number of linearly independent *rows* of A is always the same as the maximal number of independent *columns*. This number, r, is called the *rank* of A. When $r = \min(m, n)$ is as big as it could be, the matrix A has *full rank.*
Inverses.

When a square matrix A $(n \times n)$ has full rank n, the linear transformation $\alpha : V \to V$ that it represents is *invertible*, and so has an inverse map $\alpha^{-1} : V \to V$ such that $\alpha\alpha^{-1} = \alpha^{-1}\alpha = i$, the identity map, and α^{-1} is also a linear transformation. The matrix representing α^{-1} is called A^{-1}, the *inverse matrix* of A:

$$AA^{-1} = A^{-1}A = I,$$

the *identity matrix* of size n: $I = (\delta_{ij})$ ($\delta_{ij} = 1$ if $i = j$, 0 otherwise – the *Kronecker delta*).
Transpose.

If $A = (a_{ij})$, the *transpose* is A', or $A^T := (a_{ji})$.

Note that, when all the matrices are defined,

$$(AB)^{-1} = B^{-1}A^{-1}$$

(as this gives $(AB)(AB)^{-1} = ABB^{-1}A^{-1} = AA^{-1} = I$, and similarly $(AB)^{-1}(AB) = I$, as required), and

$$(AB)^T = B^T A^T$$

(as this has (i,j) element $\sum_k (B^T)_{ik}(A^T)_{kj} = \sum_k b_{ki}a_{jk} = \sum_k a_{jk}b_{ki} = (AB)_{ji}$, as required).

Determinants.

There are $n!$ *permutations* σ of the set

$$N_n := \{1, 2, \ldots, n\}$$

– bijections $\sigma : N_n \to N_n$. Each permutation may be decomposed into a product of *transpositions* (interchanges of two elements), and the *parity* of the number of transpositions in any such decomposition is always the same. Call σ *odd* or *even* according as this number is odd or even. Write

$$\text{sgn } \sigma := 1 \quad \text{if } \sigma \text{ is even}, \quad -1 \quad \text{if } \sigma \text{ is odd}$$

for the *sign* or *signum* of σ. For A a square matrix of size n, the function

$$\det A, \text{ or } |A|, := \sum_\sigma (-1)^{\text{ sgn } \sigma} a_{1,\sigma(1)} a_{2,\sigma(2)} \ldots a_{n,\sigma(n)},$$

where the summation extends over all $n!$ permutations, is called the *determinant* of A, $\det A$ or $|A|$.

Properties.

1. $|A^T| = |A|$.

Proof. If σ^{-1} is the inverse permutation to σ, σ and σ^{-1} have the same parity, so the sums for their determinants have the same terms, but in a different order.

2. If two rows (or columns) of A coincide, $|A| = 0$.

Proof. Interchanging two rows changes the sign of $|A|$ (extra transposition, which changes the parity), but leaves A and so $|A|$ unaltered (as the two rows coincide). So $|A| = -|A|$, giving $|A| = 0$.

3. $|A|$ depends linearly on each row (or column) (det is a *multilinear* function, and this area is called Multilinear Algebra).

4. If A is $n \times n$, $|A| = 0$ iff A has rank $r < n$.

5. *Multiplication Theorem for Determinants.* If A, B are $n \times n$ (so AB, and BA, are defined),

$$|AB| = |A|.|B|.$$

Proof. We can display a matrix A as a row of its columns, $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ (or as a column of its rows). The kth column of the matrix product $C = AB$ is then

$$\mathbf{c}_k = b_{1k}\mathbf{a}_1 + \ldots + b_{nk}\mathbf{a}_n.$$

For, the ith element of the kth column of C is

$$c_{ik} = \sum_j a_{ij} b_{jk} = \sum_j b_{jk} [\mathbf{a}_j]_i = [\sum_j b_{jk} \mathbf{a}_j]_i.$$

This is the ith element of the above vector equation, on both sides. Then

$$detC = detAB = det[b_{11}\mathbf{a}_1 + \ldots + b_{n1}\mathbf{a}_n, \ldots, b_{1n}\mathbf{a}_1 + \ldots + b_{nn}\mathbf{a}_n].$$

Expand the RHS by the first column. We get a sum of the form

$$\sum_{j_1} b_{j_1,1} det[\ldots].$$

Expand each det here by the second column. We get a double sum, of the form

$$\sum_{j_1,j_2} b_{j_1,1} b_{j_2,1} det[\ldots],$$

and so on, finally getting

$$\sum_{j_1,\ldots,j_n} b_{j_1,1} \ldots b_{j_n,1} det[\ldots].$$

Each matrix whose det we are taking here is a row of columns of A. Each such det with two columns the *same* vanishes. So we can reduce the 'big' sum (n^n terms) to a smaller sum with all columns *different* ($n!$ terms). Then we have a *permutation* of the columns, σ say, giving

$$detC = \sum_\sigma b_{\sigma(1),1} \ldots b_{\sigma(n),n} det[\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(n)}].$$

Putting the columns here in their natural order,

$$detC = \sum_\sigma b_{\sigma(1),1} \ldots b_{\sigma(n),n}.(-1)^{sgn(\sigma)} det[\mathbf{a}_1, \ldots, \mathbf{a}_n].$$

The determinant here is $detA$, so we can take it out. This leaves $detB$, so

$$detC = det(AB) = detA.detB. \qquad //$$

6. *Inverses again.*

If A is $n \times n$, the (i,j) *minor* is the determinant of the $(n-1) \times (n-1)$ submatrix obtained by deleting the ith row and jth column. The (i,j)

cofactor, or *signed minor* A_{ij}, is the (i, j) minor times $(-)^{i+j}$ (the signs follow a chessboard or chequerboard pattern, with $+$ in the top left-hand corner),

The matrix $B = (b_{ij})$, where

$$b_{ij} := A_{ji}/|A|,$$

ia the *inverse matrix* A^{-1} of A, defined iff $|A| \neq 0$ (A is called *singular* if $|A| = 0$, *non-singular* otherwise (thus a square matrix has a determinant iff it is non-singular), and

$$AA^{-1} = A^{-1}A = I$$

as before:

inverse = transposed matrix of cofactors over determinant.

Proof. With B as above, $C := AB = (c_{ij})$,

$$c_{ij} = \sum_k a_{ik}b_{kj} = \sum_k a_{ik}.A_{jk}/|A|.$$

If $i = j$, the RHS is 1 (expansion of $|A|$ by its ith row). If not, the RHS is 0 (expansion of the determinant of a matrix with two identical rows). So $c_{ij} = \delta_{ij}$, so $C = AB = I$. Similarly, $BA = I$.

Solution of linear equations.

If A is $n \times n$, the linear equations

$$Ax = b$$

possess a unique solution x iff A is non-singular (A^{-1} exists), and then

$$x = A^{-1}b.$$

If A is singular (A has rank $r < n$), then *either* there is *no* solution (the equations are *inconsistent*), *or* there are *infinitely many* solutions (some equations are *redundant*, and one can give some of the elements x_i *arbitrary* values and solve for the rest). What decides between these two cases is the rank of the *augmented* matrix (A, b) obtained by adjoining the vector b as a final column.

Orthogonal Matrices.

A square matrix A is *orthogonal* if

$$A^T = A^{-1},$$

or equivalently, if

$$A^T A = A A^T = I.$$

Then $|A^T A| = |A^T||A| = |A|.|A| = |I| = 1$, $|A|^2 = 1$, $|A| = \pm 1$ (we take the $+$ sign).

If $A = (a_1, \ldots, a_n)$ (row of column vectors, so A^T is the column of row-vectors a_i^T) is orthogonal, $A^T A = I$, i.e.

$$\begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} (a_1, \ldots, a_n) = I,$$

$a_i^T a_j = \delta_{ij}$: the columns of A are orthogonal to each other, and similarly the rows are orthogonal to each other.

Note. If A, B are orthogonal, so is AB, since $(AB)^T AB = B^T A^T AB = B^T B = I$.

Generalised inverses.

The theory above partially extends to non-square matrices, and matrices not of full rank. For A $m \times n$, call A^- a *generalised inverse* if

$$AA^- A = A.$$

We quote:

1. Generalised inverses always exist (but need not be unique),
2. If the linear equation

$$Ax = b$$

is consistent (has at least one solution), then a particular solution is given by

$$x = A^- b.$$

Eigenvalues and eigenvectors.

If A is square, and

$$Ax = \lambda x \qquad (x \neq 0),$$

λ is called an *eigenvalue* (latent value, characteristic value, e-value) of A, x an *eigenvector* (latent vector, characteristic vector, e-vector) (determined only to within a non-zero scalar factor c, as $A(cx) = \lambda(cx)$). Then

$$(A - \lambda I)x = 0$$

has non-zero solutions x, so infinitely many solutions cx, so $A - \lambda I$ is singular:

$$|A - \lambda I| = 0.$$

If A is $n \times n$, this is a polynomial equation of degree n in λ. By the Fundamental Theorem of Algebra (see e.g. M2PM3 L19-L20), there are n roots $\lambda_1, \ldots, \lambda_n$ (possibly complex, counted according to multiplicity).

A matrix A is singular iff the linear equation $Ax = 0$ has some non-zero solution x. This is the condition for 0 to be an eigenvalue:

a matrix is singular iff 0 is an eigenvalue.

Since the coefficient of λ^n in the polynomial $p(\lambda) := |A - \lambda I|$ is $(-)^n$, $p(\lambda)$ factorises as

$$p(\lambda) := |A - \lambda I| = \prod_1^n (\lambda - \lambda_i).$$

Put $\lambda = 0$:

$$|A| = \prod_1^n \lambda_i :$$

the determinant is the product of the eigenvalues.

Match the coefficients of $(-\lambda)^{n-1}$: in the RHS, we get a λ_i term for each i, so the coefficient is $\sum_i \lambda_i$, the sum of the eigenvalues. In the LHS, we get an a_{ii} term for each i, so the coefficient is $\sum a_{ii}$, the sum of the diagonal elements of A, which is called the *trace* of A. Comparing:

$$\text{tr } A = \sum_i \lambda_i :$$

the trace is the sum of the eigenvalues.

Properties.
1. If A is *symmetric*, eigenvectors x_i, x_j corresponding to distinct eigenvalues λ_i, λ_j are *orthogonal.*
Proof. $Ax_i = \lambda_i x_i$, so $x_i^T A^T = \lambda_i x_i^T$, or $x_i^T A = \lambda_i x_i^T$ as A is symmetric. So $x_i^T A x_j = \lambda_i x_i^T x_j$. Interchanging i and j and transposing (or arguing as above), $x_i^T A x_j = \lambda_j x_i^T x_j$. Subtract: $(\lambda_i - \lambda_j) x_i^T x_j = 0$, giving $x_i^T x_j = 0$ as $\lambda_i \neq \lambda_j$. //
2. If A is *real and symmetric*, its eigenvalues are *real.* For $Ax = \lambda x$; taking complex conjugates gives $A\overline{x} = \overline{\lambda}\overline{x}$ as A is real. Transposing, as A is

symmetric, this gives $\overline{x}^T A = \overline{\lambda}\overline{x}^T$. So $\overline{x}^T Ax = \overline{\lambda}\overline{x}^T x$. Also $Ax = \lambda x$, so $\overline{x}^T Ax = \lambda\overline{x}^T x$. Subtract: $0 = (\overline{\lambda} - \lambda)\overline{x}^T x$. But if x has jth element $x_j + iy_j$, $\overline{x}^T x = \sum_j (x_j^2 + y_j^2)$, which is non-zero as x is non-zero. So $\overline{\lambda}^T = \lambda$, and λ is real. //
Note. The same proof shows that if A is *anti-symmetric* – $A^T = -A$ – the eigenvalues are purely imaginary.
3. If A is *real and orthogonal*, its eigenvalues are of *unit modulus*: $|\lambda| = 1$.
Proof. If $Ax = \lambda x$, $A\overline{x} = \overline{\lambda}\overline{x}$ as A is real, so $\overline{x}^T A^T = \overline{x}^T\overline{\lambda}$. So $\overline{x}^T A^T Ax = \overline{x}^T\overline{\lambda}.\lambda x$, which as A is orthogonal is $\overline{x}^T x = \overline{\lambda}\lambda.\overline{x}^T x$. Divide by $\overline{x}^T x = \sum_i x_i^2 > 0$ (as $x \neq 0$): $\overline{\lambda}.\lambda = |\lambda|^2 = 1$. //
4. If C, A are similar ($C = B^{-1}AB$), A has eigenvalues λ and eigenvectors x – then C has eigenvalues λ and eigenvectors $B^{-1}x$.
Proof. $|A - \lambda I| = 0$, so $|C - \lambda I| = |B^{-1}AB - \lambda B^{-1}IB| = |B^{-1}||A - \lambda I||B| = 0$. So C has eigenvalues λ. $C(B^{-1}x) = (B^{-1}AB)(B^{-1}x) = B^{-1}Ax = B^{-1}\lambda x = \lambda(B^{-1}x)$, so C has eigenvectors $B^{-1}x$. //
Corollary. Similar matrices have the same determinant and trace.
Proof. These are the product and sum of the eigenvalues. //
5. If A is non-singular, the eigenvalues of A^{-1} are the reciprocals λ^{-1} of the eigenvalues λ of A, and the eigenvectors are the same.
Proof. $Ax = \lambda x$, so $x = A^{-1}\lambda x$, so $A^{-1}x = \lambda^{-1}x$. //

Theorem (Spectral Decomposition, or Jordan Decomposition). A symmetric matrix A can be decomposed as

$$A = \Gamma\Lambda\Gamma^T = \sum \lambda_i \gamma_i \gamma_i^T,$$

where $\Lambda = diag(\lambda_i)$ is the diagonal matrix of eigenvalues λ_i and $\Gamma = (\gamma_1, \ldots, \gamma_n)$ is an orthogonal matrix whose columns γ_i are standardised eigenvectors ($\gamma_i^T\gamma_i = 1$).

We shall prove a more general result (SVD) in Day 12. As a corollary, one can show that for A symmetric, its rank $r(A)$ is the number of non-zero eigenvalues.
Square root of a matrix.
If A is symmetric, with decomposition as above, and we define $\Lambda^{1/2} := diag(\lambda_i^{1/2})$, then putting

$$A^{1/2} := \Gamma\Lambda^{1/2}\Gamma^T,$$

$$A^{1/2}A^{1/2} = \Gamma\Lambda^{1/2}\Gamma^T\Gamma\Lambda^{1/2}\Gamma^T$$

$$\begin{aligned} &= \Gamma\Lambda^{1/2}\Lambda^{1/2}\Gamma^T \qquad (\Lambda \text{ is orthogonal}) \\ &= \Gamma\Lambda\Gamma^T \qquad (\Lambda = diag(\lambda_i)) \\ &= A. \end{aligned}$$

We call $A^{1/2}$ the *square root* of A. If also A is non-singular (so no eigenvalue is 0, so each λ_i^{-1} is defined), write

$$A^{-1/2} := \Gamma\Lambda^{-1/2}\Gamma^T.$$

A similar argument shows that

$$A^{-1/2}A^{-1/2} = A^{-1},$$

so we call $A^{-1/2}$ the square root of A^{-1}, and the inverse square root of A.
Positive definite matrices.

If A $(n \times n)$ is real and symmetric, A is *positive definite* (respectively *non-negative definite*) if

$$x^T Ax > 0 \qquad (\text{respectively } \geq 0) \qquad \text{for all non-zero } x.$$

Here $x^T Ax = \sum_{i,j=1}^n x_i a_{ij} x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i \neq j} a_{ij} x_i x_j$ is a *quadratic form* in the n variables $x_1, \ldots, x_n$ (one can replace $\sum_{i \neq j}$ by $2\sum_{i<j}$).

By the Spectral Decomposition Theorem,

$$\begin{aligned} x^T Ax &= x^T \Gamma\Lambda\Gamma^T x = y^T \Lambda y \qquad (y := \Gamma^T x) \\ &= \sum \lambda_i y_i^2. \end{aligned}$$

So A is non-negative definite (positive definite) iff $\sum_i \lambda_i y_i^2 \geq 0$ for all y (> 0 for all non-zero y) iff all $\lambda_i \geq 0$ (> 0):
Proposition. A real symmetric matrix A is non-negative definite (positive definite) iff all its eigenvalues are non-negative (positive).

Matrices of the form $A^T A$ are common in Statistics (e.g., in Regression).
1. $A^T A$ is always non-negative definite, since $x^T A^T Ax = (Ax)^T(Ax) = y^T y = \sum y_i^2 \geq 0$, with $y := Ax$. So all eigenvalues of $A^T A$ are non-negative.
2. $A^T A$ is positive definite iff all eigenvalues are positive iff $A^T A$ is non-singular, and one can show this happens iff A has full rank.
3. If $N(A)$ is the null space of A (the vector space of all x with $Ax = 0$), $N(A) = N(A^T A)$.
4. $A^T A$ and A have the same rank.