Cointegration Primer

Jacques-Olivier Moussafir

May 11th, 2009

1. Introduction

Roughly speaking, two random processes x_t and y_t are cointegrated when some linear combination of them is stationary. One easily constructs examples of such processes, and there are many multi-dimensional financial time series that match more or less that definition. The purpose of this document is to describe precisely all of it in the simplest possible case. There are some technical details usually not covered in textbooks and papers and that are included here. Section 2 reviews some basic results about multi-dimensional least squares. Section 3 describes the reduced rank regression which is the technical central component of cointegration estimation. Section 4 introduces the VAR and VECM models, explains why these models can be used to model cointegrated financial time-series, and explain precisely the estimation process.

2. Least Squares

In ordinary least squares, one considers $n+1$ observed time series: y_t, x_t^1, $\ldots$, x_t^n, with $t = 1, \ldots, T$, and one looks for n coefficients β_1, $\ldots$, β_n that solve

$$\min_{\beta} \sum_t ||y_t - \beta' x_t||^2$$

with obvious notations. One doesn't need any underlying probabilistic context. The natural extension of this problem to multi-dimensional time series works like this: let $y_t \in \mathbb{R}^p$ and $x_t \in \mathbb{R}^q$ be two multi-dimensional time series, solve

$$\min_{A} \sum_t ||y_t - A x_t||^2$$

where A is a matrix of size $p \times q$. This problem has many equivalent forms, for instance if X denotes the $q \times T$ matrix whose columns are x_t and Y, the

$p \times T$ matrix whose columns are y_t, then the least square problem becomes

$$\min_A \operatorname{Tr}\,(Y - AX)'(Y - AX), \tag{1}$$

where Tr denotes the trace operator, and X' is X transposed.

Proposition 1. *Let y_t, x_t, $t = 1, \dots T$ be two p-dimensional time series, and let X and Y denote the $p \times T$ matrices whose columns are x_t and y. If (XX') is invertible, the solution of problem (1) is*

$$A^* = (YX')(XX')^{-1}.$$

This is straightforward: the minimization of convex quadratic problems is easily solved analytically using first order conditions. There is another way to set the least squares problem. Assume for instance that $p = 2$. The matrices X and Y correspond to two sets of T points in $\mathbb{R}^2$, and the least square problem looks for the linear transformation that approximately maps x_t to y_t. It is natural to consider residuals

$$\epsilon_t = y_t - Ax_t.$$

A good matrix A would lead to small ϵ_t and one may look at the empirical covariance Σ of these residuals or its eigenvalues since small residuals correspond to small eigenvalues. One could for instance want to minimize the determinant of Σ. The least square problem should therefore be equivalent to

$$\min_A |(Y - AX)(Y - AX)'|. \tag{2}$$

This happens to be the case.

Proposition 2. *Let $y_t \in \mathbb{R}^p$, $x_t \in \mathbb{R}^q$, $t = 1, \dots T$ be two multi-dimensional time series, and let X and Y denote the matrices whose columns are x_t and y. Let*

$$f(A) = |(Y - AX)(Y - AX)'|.$$

If (XX') is invertible, and $f(A) > 0$ for all A, then problems (1) and (2) have the same solution.

Proof. Let H be an arbitrary $p \times q$ matrix,

$$\begin{aligned} f(A+H) &= f(A)\,|\,(AXX'H' + HXX'A' - YX'H' - HXY' + HXX'H')\,| \\ &= f(A) + f(A)\operatorname{Tr}\left(\alpha(A)^{-1}\left((AX - Y)X'H' + HX(X'A' - Y')\right)\right) + o(H) \end{aligned}$$

where $\alpha(A) = (Y - AX)(Y - AX)'$. Hence

$$\mathrm{d}f_A(H) = \mathrm{Tr}\left(\alpha(A)^{-1}((AX - Y)X'H' + HX(X'A' - Y'))\right).$$

If A is a critical point for f, then $\mathrm{d}f_A = 0$ and for every H, and

$$\mathrm{Tr}\,\alpha(A)^{-1}(\Gamma' H' + H\Gamma) = 0,$$

where $\Gamma = X(X'A' - Y')$, hence $\Gamma = 0$ and $A = (YX')(XX')^{-1}$. The reader should check that the critical point is a a global minimum. □

Remark 1. Usually the dimension p is much smaller than T and the technical conditions of this proposition are met: if XX' was not invertible, then all x_t would lie in a hyperplane which would be surprising. Similarly if $(Y - AX)(Y - AX)'$ were degenerate, all residuals would lie in a hyperplane. Actually, the invertibility condition for XX' seems mandatory: without it we'd had to solve a degenerate problem. However, unfortunately, this kind of degeneracy is exactly what we'll meet in the following sections.

3. Reduced Rank Regression

The reduced rank regression is one of the main technical components for the estimation of cointegration relations. We'll see how it appears in the cointegration framework. The reduced rank regression deals with least square when the dimension of y_t and x_t are different, or when one wants to force some degeneracy in the linear dependence between x_t and y_t.

Let x_t, y_t, $t = 1, \ldots, T$ be two time series, $x_t \in \mathbb{R}^p$ and $y_t \in \mathbb{R}^q$. Here is the reduced rank regression problem statement:

$$\min_{\Pi,\ \mathrm{rank}\Pi = r} |\sum_t y_t - \Pi x_t|$$

or equivalently

$$\min_{\Pi,\ \mathrm{rank}\Pi = r} |(Y - \Pi X)(Y - \Pi X)'| \tag{3}$$

using notations introduced in previous section. Any $q \times p$ matrix Π of rank r may be written

$$\Pi = \alpha\beta'$$

where α is a $q \times r$ matrix and β a $r \times p$ matrix. Solving problem (3) requires a couple of propositions.

Proposition 3. *Let Q be a $p \times p$ symmetric positive matrix, and let*

$$\varphi(x) = \frac{|x'Qx|}{|x'x|}.$$

where x is a $p \times r$ matrix. Let $v_1, \ldots, v_p$, $\lambda_1 \geq \ldots \geq \lambda_p$ denote the eigenvectors and eigenvalues of Q, and let x_* and x^* denote the $p \times r$ matrices

$$x^* = (v_1, \ldots, v_r)\,, \ x_* = (v_{p-r+1}, \ldots, v_p)\,.$$

Then

$$\varphi(x_*) \leq \varphi(x) \leq \varphi(x^*)$$

and

$$\varphi(x_*) = \Pi_{i=p-r+1}^{p} \lambda_i, \ \varphi(x^*) = \Pi_{i=1}^{r} \lambda_i.$$

Sketch of proof. This result has a very simple and nice geometric interpretation. We only give it an leave some work to the reader. First, remark that if Q is the matrix of a quadratic form q in the usual basis of $\mathbb{R}^p$, and x denotes another basis of $\mathbb{R}^p$, the matrix of q with respect to x is $x'Qx$. This result extends to the case where $x = (x_1, \ldots, x_r)$ spans a sub-vector space of dimension r. Second, recall that the determinant of a symmetric matrix is the product of its eigenvalues. Finally, recall the simple relation between eigenvalues of a positive symmetric matrix, and the shape of the ellipsoid $\mathcal{E}$ defined by $u'Qu = 1$, where here $u \in \mathbb{R}^p$: $\mathcal{E}$ has p symmetry axis and the intersection of $\mathcal{E}$ with them defines p segments whose lengths are $2/\sqrt{\lambda_1}$, ..., $2/\sqrt{\lambda_p}$. We call them the lengths of the ellipsoid axis.

The maximization of φ therefore corresponds to search of a sub-vector space of dimension r whose intersection with $\mathcal{E}$ gives an ellipsoid $\mathcal{E}'$ whose axis are as short as possible. More precisely we want the product of these lengths to be as small as possible. Similarly, the minimization of φ corresponds to searching a sub-vector space of dimension r whose axis are as large as possible.

Solving this is therefore straightforward: for the maximization, pick the r eigenvectors of Q corresponding to the r largest eigenvalues values (and smallest symmetry axis), and for the minimization, pick the r eigenvectors of the r smallest eigenvalues. $\square$

Remark 2. This proposition gives a global maximum and a global minimum for φ. The proof suggests that they need not be unique: all basis of the same sub-vector space should lead to the same value for φ. Hence argmin and argmax for φ are defined up to a change of basis. This remark also applies to the next proposition.

The following proposition is a simple consequence of the previous one.

Proposition 4. *Let M be a $p \times p$ symmetric positive matrix, and let N be a $p \times p$ symmetric positive definite matrix. Let*

$$\varphi(x) = \frac{|x'Mx|}{|x'Nx|},$$

where x is a $p\times r$ matrix. Let $v_1, \dots, v_p$, $\lambda_1 \geq \dots \geq \lambda_p$ denote the eigenvectors and eigenvalues of $N^{-1/2}MN^{-1/2}$, and let x_* and x^* denote the $p\times r$ matrices

$$x^* = N^{-1/2}\left(v_1, \dots, v_r\right), \; x_* = N^{-1/2}\left(v_{p-r+1}, \dots, v_p\right).$$

Then

$$\varphi(x_*) \leq \varphi(x) \leq \varphi(x^*)$$

and

$$\varphi(x_*) = \Pi_{i=p-r+1}^{p}\lambda_i, \; \varphi(x^*) = \Pi_{i=1}^{r}\lambda_i.$$

The resolution of problem (3) works like this. First we write the matrix of rank r, Π as

$$\Pi = \alpha\beta'.$$

Problem (3) becomes

$$\min_{\alpha,\beta} |(Y - \alpha\beta'X)(Y - \alpha\beta'X)'|. \tag{4}$$

We first fix β and use least square to get α as a function of β. Then we solve

$$\min_{\beta} |(Y - \alpha(\beta)\beta'X)(Y - \alpha(\beta)\beta'X)'|$$

using previous propositions and get β. This algorithms relies on

$$\min_{\beta}\min_{\alpha} |(Y - \alpha\beta'X)(Y - \alpha\beta'X)'| = \min_{\alpha,\beta} |(Y - \alpha\beta'X)(Y - \alpha\beta'X)'|,$$

which happens to be true.

Proposition 5. *Let y_t, x_t, $t = 1, \ldots T$ be two p-dimensional time series, and let X and Y denote the $p \times T$ matrices whose columns are x_t and y. If XX' and YY' are invertible, the solutions of*

$$\min_{\alpha,\beta} |(Y - \alpha\beta' X)(Y - \alpha\beta' X)'|$$

are given by

$$\beta^* = (XX')^{-1/2}\,(v_{p-r+1}, \ldots, v_p)\,\phi$$
$$\alpha^* = YX'\beta\,(\beta' XX'\beta)$$

where $v_1, \ldots, v_p$ and $\lambda_1 \geq \ldots \geq \lambda_r$ are the eigenvectors and eigenvalues of

$$Q = \mathrm{I} - (XX')^{-1/2} XY'(YY')^{-1} YX'(XX')^{-1/2},$$

ϕ is any invertible $r \times r$ matrix, and I *the identity matrix.*

Proof. First remark that if XX' is invertible and β has full rank, then $\beta' XX'\beta$ is also invertible. We first perform the optimization for full rank β.

For full rank fixed β, least squares give the solution of

$$\min_{\alpha} |(Y - \alpha\beta' X)(Y - \alpha\beta' X)'|,$$

$$\alpha^* = YX'\beta(\beta' XX'\beta)^{-1}.$$

Minimizing with respect to β leads to

$$\min_{\beta} |YY' - YX'\beta(\beta' XX'\beta)^{-1}\beta' XY'|.$$

Remark that for any square matrix of same size, A, B, C, D, with A and C invertible,

$$|A||C - DA^{-1}B| = |C||A - BC^{-1}D|.$$

This identity gives

$$|YY' - YX'\beta(\beta' XX'\beta)^{-1}\beta' XY'| = |YY'|\frac{|\beta'(XX' - XY'(YY')^{-1}YX')\beta|}{|\beta' XX'\beta|},$$

and using invertibility of YY' we are left with

$$\min_{\beta} \frac{|\beta'(XX' - XY'(YY')^{-1}YX')\beta|}{|\beta' XX' \beta|},$$

which corresponds exactly to the problem solved in proposition 4. Finally, let

$$\varphi(\alpha, \beta) = |(Y - \alpha\beta' X)(Y - \alpha\beta' X)'|.$$

One checks easily that

$$\min_{\alpha,\beta} \varphi(\alpha, \beta) \leq \min_{\alpha,\beta,\operatorname{rank}\beta<r} \varphi(\alpha, \beta),$$

which ensures that the assumption we made about the rank of β is actually harmless. $\square$

4. Cointegrated VAR and VECM

VAR processes stands for vector auto regressive process. They extend naturally the usual auto-regressive process. A fair introduction can be found in [4]. The basic definitions are directly inspired by the one dimensional case. We quote them here.

Definition 1. *A VAR(k) process y_t is a multi-dimensional random process that satisfies*

$$y_t = A_1 y_{t-1} + \cdots + A_k y_{t-k} + u_t, \ \textit{for } t \ \in \mathbb{Z}$$

where u_t is a second order white noise, i.e. a process with moments of order 2 such that $\mathbb{E}[\,u_t\,] = 0$, $\mathbb{E}[\,u_t u'_{t+h}\,] = 0$ for $h \neq 0$. A VAR(k) process is said to be stationary if $\mathbb{E}[\,y_t\,]$ and $\mathbb{E}[\,y_t\ y'_{t+h}\,]$ do not depend on t. It is said to be stable if the roots of

$$|\mathrm{I} - A_1 z - \cdots A_k z^k|$$

lie outside of the unit circle.

Definition 2. *A random multivariate process $y_t \in \mathbb{R}^p$ is said to be cointegrated of order r, $1 \leq r \leq p$ when r independent linear combinations of y_t^i*

are stationary. The process y_t is said to be cointegrated if it is cointegrated of order r for some r, $1 \leq r \leq p$.

The processes we're interested in have the following properties: they are not stationary, cointegrated, and become stationary after differentiation. If we also assume that they are VAR(1), which is quite general, we get a model for the (log-) prices we want to model.

If y_t is a p-dimensional VAR(1) and Δy_t stationnary, then

$$\Delta y_t = (A - I)y_{t-1} + \epsilon_t$$

and

$$(A - I)y_t$$

must be stationnary. Two possibilities occur: if $A = I$, then y_t is a random walk. If $A \neq I$, its rank cannot be full, and

$$A - I = \alpha\beta'$$

where α and β have dimensions $p \times r$ with $r \leq p - 1$.

Multi-variate processes that satisfy such a a relation are called Vector Error Correction Models (VECM). Because of last proposition we get that any price process that is mild enough to fit into a VAR(1) and exhibits cointegration relations can be modeled as a VECM. Moreover the β that appears in the VECM model contains the cointegration relations.

The methods presented in section 3 are obviously useful for the estimation of VECM models.

Proposition 6. *Let $y_1, \ldots y_T$ be the sample of a p-dimensional VECM*

$$\Delta y_t = \alpha\beta' y_{t-1} + u_t.$$

where α, β denote two $p \times r$ matrices and u_t a white noise. Let ΔY and Y denote the $p \times T$ matrices whose columns are Δy_{t-1} and y_t. The maximum likelihood estimators of α and β are

$$\beta^* = (v_{p-r+1}, \ldots, v_p)\,\phi$$

$$\alpha^* = \Delta Y Y'\beta\,(\beta' Y Y'\beta)$$

where $v_1, \ldots, v_p$ and $\lambda_1 \geq \ldots \geq \lambda_r$ are the eigenvectors and eigenvalues of .

$$Q = \mathrm{I} - Y\Delta Y'(\Delta Y \Delta Y')^{-1}\Delta Y Y'(YY')^{-1},$$

ϕ is any invertible $r \times r$ matrix, and I *the identity matrix.*

There is also a least square estimator of VECM, that results from the direct application of formula given in 2, using notations of previous property

$$\alpha\beta' = \Delta Y Y'(YY')^{-1}$$

This method has an apparent default since it uses least squares for non-stationary data. However one can prove, see [4] that the resulting estimators are consistent. There are two other advantages for the likelihood estimators. First they directly give access to α and β'. Second they allow the development of an estimation framework where one can also estimate likelihood ratio, and compare the hypothesis rank $A = r_0$ with the hypothesis rank $A = r_0 + 1$ for instance.

References

[1] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods. 2nd ed.* Springer-Verlag, 1991.

[2] Søren Johansen. Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control*, 12(2-3):231–254, 1988.

[3] Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

[4] Helmut Lütkepohl. *New introduction to multiple time series analysis.* Springer-Verlag, 2006.