

# 时间序列稳定协和性的数值分析\*

吕纯濂

(南京信息工程大学数学系 南京 210044  
南京金肯学院 南京 211156)

陈舜华

(南京信息工程大学应用气象系 南京 210044)

## NUMERICAL ANALYSIS TO STABLE COINTEGRATION FOR TIME SERIES

Lü Chunlian

(*Dept. of Mathematics of Nanjing University of Information Science & Technology,  
Nanjing 210044; Nanjing Jinken University, Nanjing 211156*)

Chen Shunhua

(*Dept. of Applied Meteorology of Nanjing University of Information Science & Technology,  
Nanjing 210044*)

### Abstract

Cointegration analysis for time series involves the solution of a generalized eigenproblem involving moment matrices and inverted moment matrices. These formulae are unsuitable for actual computations because the condition numbers of the resulting matrices are unnecessarily increased. The special structure of the cointegration procedure is used to achieve numerically stable computations, based on QR and singular value decompositions.

**Key words:** Cointegration, QR decomposition, Reduced rank, Singular-value decomposition

### §1. 引 言

许多观测到的时间序列常显示出非平稳性. 当时间序列的一阶差分为白噪声时, 就产生了非平稳性的简单形式, 在这种情况下, 称该序列, 比如说  $y_t$  为一阶求和的, 记为  $y_t \sim I(1)$ , 而对其一阶差分记为  $\Delta y_t \sim I(0)$ . 在求和序列的统计分析中, 重要的一步是实现对  $I(1)$  变量的线性组合使之成为  $I(0)$  就可能实现了, 此时称这些变量为协和的.

\* 2003 年 7 月 18 日收到.

对于统计分析来说,最简单的是向量自回归 (VAR) 模型. 例如, 对一阶自回归模型:

$$y_t = \pi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N_n(0, \Omega), \quad (1)$$

其中  $y_t$  为  $n \times 1$  向量,  $\pi$  为  $n \times n$  矩阵. 干扰  $\varepsilon_t$  独立同分布 (i.i.d.) 于  $n$  维正态分布  $N_n(0, \Omega)$ . (1) 式写成向量均衡修正模型 (VECM), 用  $y_{t-1}$  减 (1) 式两边, 即  $\Delta y_t = (\pi - I_n)y_{t-1} + \varepsilon_t$ , 其中  $I_n$  为  $n \times n$  单位矩阵. 令  $\Pi = \pi - I_n$ , 可以写成  $\Delta y_t = \Pi y_{t-1} + \varepsilon_t$ . 这说明矩阵  $\Pi$  决定了过程  $y$  以怎样的水准进入系统. 比如当  $\Pi = 0$  时, 过程的动态发展不依赖于任何这些变量的水准. 协和性的统计假设为  $H_0: \text{rank}(\Pi) \leq r$ . 在该假设下,  $\Pi$  可以写为两个矩阵的乘积  $\Pi = \alpha\beta'$ , 其中  $\alpha$  和  $\beta$  皆为  $n \times r$  矩阵. Johansen<sup>[1]</sup> 指出, 这样的约束可以用极大似然方法进行分析.

极大似然估计要求解决一个广义特征问题. 本文的目的是要指出, 解决该特征问题的标准方法, 正如在统计文献以及在下一节将要介绍的, 从数值分析的角度出发, 是不合需要的, 故我们考虑一些新的方法, 它们在数值分析上, 将会更稳定. 下面概括介绍以后将要用到的两种正交分解.

$m \times n$  矩阵  $A$  的奇异值分解 (SVD)<sup>[2]</sup>,  $\text{rank } A = r$ ,  $A = U\Sigma V'$ . 假定  $m \geq n$ , 则给出“瘦”型矩阵  $U$  为  $m \times n$ , 而  $V$  为  $n \times n$ ,  $U'U = V'V = I_n$ , 其中  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ , 奇异值皆为实、非负的, 且排序为  $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0$ , 故  $A$  的秩  $r$  对应非零奇异值的个数.

满秩  $m \times n$  矩阵  $A$  的 QR 分解<sup>[3]</sup>, 假定  $m \geq n$ , 有  $A = QR$ , 这里  $Q$  为  $m \times m$ ,  $Q'Q = I_m$ , 而  $R$  为具有正对角元素的  $m \times n$  上三角阵, 即  $R$  的下面  $m - n$  行全为零, 而上面  $n$  行构成上三角阵  $R_n$ ,  $Q_n$  由  $Q$  的左  $n$  列构成, 即  $Q = (Q_n : *)$ ,  $R' = (R'_n : 0)$ , 且  $A = Q_n R_n$ . 不保存  $Q$ , 仅保存 Householder 变换的  $m \times n$  矩阵;  $R'_n R_n$  是  $A'A$  的 Choleski 分解.

## §2. 协和性分析的标准算法

现考虑更高阶的自回归模型, 且具有确定性变量, 比如常数项和趋势项. 从统计学的观点来看, 常数项和趋势项的处理是非常重要的. 如果出现趋势项, 通常限制在协和性空间中, 得到一般的向量均衡修正模型 (VECM):

$$\Delta x_t = \Pi x_{t-1}^* + \sum_{j=1}^{k-1} \Gamma_j \Delta x_{t-j} + \Phi q_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

其中  $\{x_t\}$  为  $n$  维向量时间序列, 初始值  $(x_{1-k}, \dots, x_0)$  固定,  $\{\varepsilon_t\}$  独立同分布 (i.i.d.) 于  $N_n(0, \Omega)$ , 而  $x_{t-1}^* = (x'_{t-1}, d'_t)'$ , 其中  $d_t$  和  $q_t$  为确定性回归量,  $T$  为允许滞后与差分后的样本容量.

估计协和性空间的极大似然方法由 Johansen<sup>[1]</sup> 提出,  $\Pi$  可以写为两个矩阵的乘积  $\Pi = \alpha\beta'$ , 其中  $\alpha$  皆为  $n \times r$  矩阵,  $\beta$  为  $n_1 \times r$  矩阵. 在这种情况下, 有  $n_1 - n$  个变量  $d_t$  限制在协和性空间, 而非限制的确定性变量  $q_t$  和滞后的差分 (以及可能的附加非模型变量) 可以

从  $\Delta x_t$  和  $x_{t-1}^*$  中分离出去 (即预先从回归中删除). 只要给定  $y_{0t}$  和  $y_{1t}$

$$y_{0t} = \alpha\beta'y_{1t} + \varepsilon_t, \quad t = 1, \dots, T \quad (3)$$

将被分离出去的数据用向量形式表示为  $Y_0' = (y_{01}, \dots, y_{0T})$  和  $Y_1' = (y_{11}, \dots, y_{1T})$ , 则给出 (3) 式的矩阵等价式

$$Y_0 = Y_1\beta\alpha' + E. \quad (4)$$

似然函数可由求解以下广义特征问题而达极大,

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0, \quad (5)$$

其中  $S_{ij} = T^{-1}Y_i'Y_j$ ,  $i, j = 0, 1$ . 还可以通过用  $S_{11}^{-1/2}$  分别左乘和右乘而转化为标准特征问题<sup>[1]</sup>, 只要给定  $n$  个特征值  $1 > \hat{\lambda}_1 > \dots > \hat{\lambda}_n > 0$ , 剩下的任意  $n_1 - n$  个特征值为零,  $\lambda$  上面的  $\hat{\lambda}$  表示这些是估计量. 用“迹”统计量<sup>[1]</sup> 可以检验  $H_0: \text{rank}(\Pi) \leq r$ .

形如  $\alpha = B\theta$  和  $\beta = H\phi$  的简单约束允许以较小的校正而合并, 令  $B = Q_bR_b$  为  $B$  的 QR 分解, 则 (3) 式可写为  $Q_b'y_{0t} = R_b\theta\phi'H'y_{1t} + Q_b'\varepsilon_t$ .

常用以下算法来执行 Johansen 程序<sup>[1]</sup>:

**标准算法.** 若  $Y_0$  和  $Y_1$  具有列满秩, 则 (4) 式的极大似然估计 (ML) 可按以下步骤获得:

(1) 二阶矩矩阵  $S_{00}, S_{01}, S_{10}$ :  $S_{ij} \leftarrow T^{-1}Y_i'Y_j$ .

(2)  $S_{11}$  的 Choleski 分解:  $S_{11} = PP'$ .

(3) 解对称特征问题:  $P^{-1}S_{01}'S_{00}^{-1}S_{01}P^{-1} = H\Lambda H'$ , 其中  $\Lambda$  为具有特征值  $\hat{\lambda}_i$  的对角矩阵, 而  $\hat{\beta} = P^{-1}H$ ,  $\hat{\alpha} = S_{01}\hat{\beta}$ .

给定  $\beta$ , 则  $\alpha$  可由  $\hat{\alpha}(\beta) = S_{01}\beta(\beta'S_{11}\beta)^{-1}$  计算<sup>[1]</sup>, 因为  $\hat{\beta}'S_{11}\hat{\beta} = I$ , 故得  $\hat{\alpha} = S_{01}\hat{\beta}$ . 注意,  $\alpha$  和  $\beta$  并非唯一确定: 对任意非奇异矩阵  $V$  有  $\alpha\beta' = (\alpha V^{-1})(V\beta')$ .

求解具有对称、正定矩阵  $A$  和  $B$  的广义特征问题  $|\lambda A - B| = 0$ , 在一些参考书中可以找到<sup>[4,5,6]</sup>. 在协和性情况下, 矩阵  $A$  和  $B$  有更多的结构, 因此从数值计算上考虑, 标准算法是不合适的. 同样从统计学上考虑, 回归也不一定能由二阶矩矩阵求逆得到. 本文将在下面两节介绍在数值计算上更稳定的两种新的算法.

### §3. QR 分解算法

首先考虑对 Johansen 程序与数据矩阵联系在一起, 且在 QR 分解基础上进行的一种程序:

**QR 分解算法.** 若  $Y_0$  和  $Y_1$  具有列满秩, 则 (4) 式的极大似然估计 (ML) 可按以下步骤获得:

(1)  $T \times (n_1 + n)$  矩阵  $Y = (Y_1 \ Y_0)$  的 QR 分解:  $Y = QR = Q \begin{pmatrix} R_{11} & R_{10} \\ 0 & R_{00} \end{pmatrix}$ .

(2)  $n_1 \times n$  矩阵  $Z$  的计算:  $Z \leftarrow R_{10}R_{00}^{-1}$ .

(3)  $Z$  的瘦奇异值分解 (SVD):  $Z = U\Sigma V'$ .

所求的估计量为

$$\hat{\lambda}_i = \frac{\sigma_i^2}{1 + \sigma_i^2}, \quad R_{11}\hat{\beta} = T^{1/2}U, \quad \hat{\alpha} = T^{-1/2}R'_{10}U,$$

其中  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ ;  $Q$  和  $V$  不需要.

证明. 在第 (1) 步中用 QR 分解作  $Y$  的二阶矩阵且用  $S_{ij}$  表示

$$T^{-1}Y'Y = T^{-1} \begin{pmatrix} R'_{11}R_{11} & R'_{11}R_{10} \\ R'_{10}R_{11} & R'_{10}R_{10} + R'_{00}R_{00} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{10} \\ S_{01} & S_{00} \end{pmatrix}.$$

由对称分块矩阵的结论, 定义  $\begin{pmatrix} A & B \\ B' & D \end{pmatrix}$ , 其中  $A$  和  $D$  满秩, 则有  $(A - BD^{-1}B')^{-1} = A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}$ . 当  $A = I$ ,  $D = -R'_{00}R_{00}$  和  $B = R_{10}$  时, 有  $[I + R_{10}(R'_{00}R_{00})^{-1}R'_{10}]^{-1} = I - R_{10}(R'_{10}R_{10} + R'_{00}R_{00})^{-1}R'_{10}$ .

用上式代替到 (5) 中, 则给出新的对称特征问题:

$$\begin{aligned} 0 &= |\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| \\ &= |T^{-1/2}R'_{11}||\lambda I - R_{10}(R'_{10}R_{10} + R'_{00}R_{00})^{-1}R'_{10}||T^{-1/2}R_{11}| \\ &= |T^{-1/2}R'_{11}| |(\lambda - 1)I + [I + R_{10}(R'_{00}R_{00})^{-1}R'_{10}]^{-1}| |T^{-1/2}R_{11}| \\ &= |T^{-1/2}R'_{11}| |(\lambda - 1)^{-1}I + I + R_{10}(R'_{00}R_{00})^{-1}R'_{10}| |T^{-1/2}R_{11}| \\ &= |T^{-1/2}R'_{11}| |\lambda(\lambda - 1)^{-1}I + R_{10}(R'_{00}R_{00})^{-1}R'_{10}| |T^{-1/2}R_{11}| \\ &= (-1)^{n_1} |T^{-1/2}R'_{11}| |-\lambda(\lambda - 1)^{-1}I - R_{10}(R'_{00}R_{00})^{-1}R'_{10}| |T^{-1/2}R_{11}|. \end{aligned}$$

在第四行中, 用到逆对称矩阵的特征值为原矩阵特征值的倒数.  $R_{11}$  是上三角阵, 而  $R_{10}$  不是.

该算法在第一步用到 QR 分解, 而用奇异值分解 (SVD) 去求对称特征值问题.

#### §4. 奇异值分解 (SVD) 算法

若不论  $Y_0$  还是  $Y_1$  非列满秩, 则上述算法就会失败. 更可靠的解决办法就是用通过以下步骤的奇异值分解 (SVD):

**奇异值分解 (SVD) 算法.** (4) 式的极大似然估计 (ML) 可按以下步骤获得:

(1)  $Y_i (i=0,1)$  的瘦奇异值分解 (SVD):  $Y_i = U_i \Sigma_i V_i'$ .

(2)  $n_1 \times n$  矩阵  $Z$  的计算:  $Z \leftarrow U_1' U_0$ .

(3)  $Z$  的瘦奇异值分解 (SVD):  $Z = U_Z \Sigma_Z V_Z'$ .

所求的估计量为  $\hat{\lambda}_i = \sigma_i^2$ ,  $\hat{\beta} = T^{1/2}V_1 \tilde{\Sigma}_1^{-1} U_Z$ ,  $\hat{\alpha} = T^{-1/2}V_0 \Sigma_0 Z' U_Z (\equiv T^{-1/2}Y_0' U_1 U_Z)$ , 其中  $\tilde{\Sigma}_1^{-1} = \text{diag}(\tilde{\sigma}_1^{-1}, \dots, \tilde{\sigma}_n^{-1})$ , 而

$$\tilde{\sigma}_i^{-1} = \begin{cases} \sigma_i^{-1}, & \sigma_i > \varepsilon_y, \\ 0, & \sigma_i \leq \varepsilon_y. \end{cases}$$

证明. 与 QR 分解算法类似, 有  $|T^{-1/2}V_1 \Sigma_1| |\lambda I - U_1' U_0 U_0' U_1| |T^{-1/2} \Sigma_1 V_1'| = 0$ .

$\varepsilon_y$  的选取依赖于机器的精度, 建议用  $\varepsilon_y = 10^{-9} \|Y_1\|_\infty$ , 这对于双精度算法是合适的.

## §5. 运算计数

根据 Golub and Van Loan 所定义的浮点运算计数法<sup>[6]</sup>, 对一个数值加、乘或除作为一次运算.

Golub and Van Loan 给出了  $m \times n$  矩阵的瘦 Householder QR 分解的运算次数为  $2mn^2 - 2n^3/3$ . 在本文中,  $m$  对应样本容量  $T$ , 且假设显著大于  $n$ , 故 QR 分解的运算次数取为  $2Tn^2$ . 当  $A$  为  $T \times q$  矩阵时, 用 Householder 变换计算  $Q'A$  的运算次数为  $4Tnq$ <sup>[6]</sup>. 在某些情况下还需要  $Q_q$  ( $Q$  的前  $q$  列), 运算次数近似为  $4qmn - 2(m+n)q^2 + 4q^3/3$ . 当  $m = T$  显著大, 且  $q = n$ , 运算次数减少为  $2Tn^2$ , 这与 QR 分解的运算次数相同.

假设  $T$  显著大, 瘦奇异值分解 (SVD) 的运算次数为  $6Tn^2$ <sup>[6]</sup>. 当  $A$  和  $B$  皆为  $T \times n$  矩阵时,  $A'B$  运算次数为  $2Tn^2$ ,  $A'A$  的为其一半是  $Tn^2$ .

标准算法的运算次数. 假设  $n_1 = n$  且样本容量  $T$  足够大, 以至  $Tn^2$  占优势, 由于标准算法要形成 4 个二阶矩阵  $S_{ij}$  ( $i, j = 0, 1$ ), 故共需  $4Tn^2$  次运算.

QR 分解算法的运算次数. 一个  $T \times 2n$  矩阵的 QR 分解, 需  $2T(2n)^2 = 8Tn^2$  次运算.

奇异值分解 (SVD) 算法的运算次数. 需 2 个奇异值分解 (SVD) 和一个  $Z$  的计算, 故共需  $14Tn^2$  次运算.

仅从运算速度来看, 标准算法最快, 为  $4Tn^2$  次; QR 分解算法是其两倍, 为  $8Tn^2$  次; SVD 分解算法是其 3.5 倍, 为  $14Tn^2$  次.

在目前计算机运算速度不断加快的情况下, 计算速度不是主要问题, 主要是运算的稳定性和正确性. 下一节就来说明三种算法的运算稳定性和正确性.

## §6. 说明和结论

为了说明以上算法的不同的稳定性和正确性, 我们估计二维向量自回归 VAR(2) 模型:

$$y_t = \begin{pmatrix} x_t \\ x_t + u_t 10^{-m} \end{pmatrix}, u_t \sim N(0, 1),$$

其向量均衡修正模型 (VECM) 没有确定性项, 为  $\Delta y_t = \alpha\beta'y_{t-1} + \Delta y_{t-1} + \varepsilon_t$ .

我们利用 Ox version 2.20 中提供的训练数据<sup>[7]</sup> 作为  $x_t$ ,  $u_t$  由默认标准正态随机数发生器产生, 当  $m = 0$  时, 得到  $(x_t, x_t + u_t)$  的特征值为  $(0.35091938503, 0.00450356266)$ , 随着  $m$  的增大, 系统将约化为  $(x_t, x_t)$ , 成为一维自回归模型, 其特征值为  $0.00487801748$ . 为了说明三种算法的不同稳定性和正确性, 计算了不同  $m$  值. 三种算法的最大特征值, 关键数据如表 1 所示.

表 1 VAR(2) 模型的最大特征值

算法	$m=1$	$m=3$	$m=5$	$m=10$
标准	0.35091940 <b>555</b>	0.35089 <b>570155</b>	<b>0.05949553867</b>	失败
QR 分解	0.35091938503	0.35091938503	0.350919385 <b>40</b>	失败
SVD 分解	0.35091938503	0.35091938503	0.350919384 <b>94</b>	0.00487801748*

在表 1 中不正确的数字用黑体并下加线标出; \* 此时已成为一维自回归模型的特征值.

选择不同  $m$  值得到的结果列在表 1 中, 显然标准算法比 QR 分解算法和 SVD 分解算法的稳定性要差.

标准算法在  $m=5$  时, 得到错误的特征值, 就已经几乎失败了, 此时还没有完全失败的信息; 而从  $m=6$  开始, 其中的 Choleski 分解开始崩溃, 致使运算失败.

QR 分解算法从  $m=6$  才开始发出列秩减少的信息, 而不能提供问题的解决, 但在此之前, 都是正确的.

SVD 分解算法可以处理奇异性. 在  $m=6$  之前都是正确的, 当  $m=6$  时, 最大特征值为 0.350917.  $m=7$  时, 就变为 0.004914, 已接近一维自回归模型, 此后就发出奇异性的信息. 从  $m=10$  开始就已完全约化为一维自回归模型, 而得到正确的答案, 说明该算法既稳定又正确.

QR 分解算法和 SVD 分解算法明显地优于标准算法, 直至  $m=5$  时一直给出正确的答案, 说明相当稳定. 一般来说, 最好用 SVD 分解算法, 但由于 QR 分解算法较 SVD 分解算法更快, 对于 Monte Carlo 随机模拟试验, 建议用 QR 分解算法.

### 参 考 文 献

- [1] Johansen, S., Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford, 1995.
- [2] 吕纯濂, 用奇异值分解求解回归方程的迭代加细, 数值计算与计算机应用, 19:4 (1998), 265~270.
- [3] 吕纯濂, 用 QR 分解拟合回归方程参数估计和剩余的迭代加细, 数值计算与计算机应用, 21:4 (2000), 308~313.
- [4] 南京大学数学系计算数学专业, 线性代数. 科学出版社, 北京, 1978, 151~158.
- [5] 曹志浩等, 矩阵计算和方程求根. 人民教育出版社, 北京, 1979, 172~180.
- [6] Golub, G.H., Van Loan, C.F., Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore. 1996.
- [7] Doornik, J.A., Object-Oriented Matrix Programming Using Ox, 4<sup>th</sup> Edition, Timberlake Consultants Press, London. 2001.