

第十三讲 VAR 模型

概要：本讲主要以存在相互影响的两个时序变量 $\{y_t\}$ 和 $\{z_t\}$ 为例，介绍 VAR 模型的结构及平稳性判别，在此基础上，阐述 VAR 模型变量滞后长度的确定（模型识别），同时，给出了 VAR 模型建构之后，用于分析不同变量之间相互影响的方法——Granger 因果检验、脉冲效应和方差分解。

大家最为熟悉的模型为古典线性回归模型，也称为**单方程模型**，它只能描述经济变量之间的单向关系，即解释变量（自变量）的变化引起的被解释变量（因变量）的变化。但经济现象是错综复杂的，许多经济变量之间存在着交错的双向或多向关系。例如：某种商品的价格影响该商品的需求和供给，同时，该商品的需求和供给又影响该商品的价格，为了描述变量之间的多向关系，就需求建立由多个单方程组成的多方程模型。含有两个以上方程的模型称为**联立方程模型**，VAR 模型就起源于联立方程模型。

一、联立方程模型简介 Multi-equation Model

联立方程中，某些变量可能是一个方程的解释变量（自变量），同时又可能是另一个方程的被解释变量（因变量），这些由模型系统决定其取值的变量称为内生变量。在单方程模型中，内生变量就是被解释变量（因变量），联立方程中，可以把被解释变量（因变量）的全体视为内生变量。由模型系统以外的因素决定其取值的变量称为外生变量，联立方程中，外生变量只能作为解释变量（自变量）存在。内生变量受系统内随机误差项的影响，是随机变量，外生变量不受系统内随机误差项的影响。内生变量的滞后项称为预定内生变量，它不受现期模型系统内的随机误差项的影响。预定内生变量和外生变量统称为预定变量。

联立方程模型中的方程可以分为两类：随机方程（行为方程）和非随机方程（定义方程），以简化凯恩斯宏观经济模型为例：

$$\text{消费方程（行为方程）: } C_t = a_0 + a_1 Y_t + \varepsilon_{1t}$$

$$\text{投资方程（行为方程）: } I_t = b_0 + b_1 Y_t + b_2 Y_{t-1} + \varepsilon_{2t}$$

$$\text{收入方程（定义方程）: } Y_t = C_t + I_t + G_t$$

其中：内生变量为 C_t 、 I_t 和 Y_t ，外生变量为 G_t ，预定内生变量为 Y_{t-1} ，预定变量 G_t 和 Y_{t-1} 。

包括消费方程和投资方程是随机方程（行为方程），式中包含随机项 ε_{1t} 和 ε_{2t} ，需要估计参数 a_0, a_1, b_0, b_1, b_2 。收入方程为非随机方程（定义方程），不需要估计参数。

描述经济变量之间直接影响关系的模型，称为结构模型 Structural Model。它是在对经济变量的影响关系进行经济理论分析基础上建立的，反映了内生变量直接受预定变量、其他内生变量和随机项影响的关系。结构模型中的方程称为结构方程，结构方程中变量的系数称为结构参数，所有结构参数组成的矩阵称为结构参数矩阵。

上述简化凯恩斯宏观模型是根据经济理论建立的，反映了消费、投资、国民收入、政府支出各经济变量之间的直接影响和被影响关系，是联立方程模型中的结构模型。引入虚拟变量 $X_t = 1$ ，上述模型可表示为：

$$\begin{cases} C_t + 0I_t - a_1Y_t + 0Y_{t-1} + 0G_t - a_0X_t = \varepsilon_{1t} \\ 0C_t + I_t - b_1Y_t - b_2Y_{t-1} + 0G_t - b_0X_t = \varepsilon_{2t} \\ -C_t - I_t + Y_t + 0Y_{t-1} - G_t + 0X_t = 0 \end{cases}$$

所以结构参数矩阵为：

$$\begin{matrix} & C_t & I_t & Y_t & Y_{t-1} & G_t & X_t \\ \begin{bmatrix} 1 & 0 & -a_1 & 0 & 0 & -a_0 \\ 0 & 1 & -b_1 & -b_2 & 0 & -b_0 \\ -1 & -1 & 1 & 0 & -1 & 0 \end{bmatrix} \end{matrix}$$

联立方程模型的参数估计是对每个随机方程式的参数估计。在结构方程中，可能有内生变量作为解释变量（自变量），因为它与随机误差项相关，方程存在随机解释变量问题，当解释变量与随机误差项相关时，就使得模型不再符合古典线性回归模型假定中的全部条件，并使参数的最小二乘估计量是有偏且不一致，所以不能直接应用 OLS 估计。上述简化凯恩斯宏观模型的消费方程和投资方程中，内生变量 Y_t 作为解释变量，因此不能用 OLS 估计。

正是存在着上述问题，联立方程模型在模型识别存在着三种情况：恰好识别（参数有唯一解）、过度识别（参数有多个解）和不可识别（参数无解）；参数估计多采用两阶段最小二乘法。相关内容可以参考其他教材，这里不赘述。

二、VAR 模型的结构

20 世纪 70 年代早期，由于石油危机而引发的世界经济衰退和随之而来的通货膨胀，而由联立方程组构成的巨大的模型特别复杂，未能预测到也未能就治理通胀开出有效的“药方”，由此导致了对结构模型的批判。联立方程模型的构建通常需要人为假定某些外生变量，并且是事前给定的，并且不能受到系统内生变量的影响，为了达到模型可识别的目的，还假

定某些前定变量仅仅出现在某些方程中。这些假定都受到了 Christopher A. Sims 的严厉批评，他认为，如果在—组变量之间有真实的联立性，那么就应该对这些变量平等地加以对待，而不是事先区分内生变量和外生变量。由此，Sims 就提出了向量自回归模型，即 VAR Model。

以两个变量 $\{y_t\}$ 和 $\{z_t\}$ 为例，这两个变量均受到自身和对方的当期和滞后一期的影响，写成方程为：

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} \quad 1.1$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} \quad 1.2$$

前提假设：1) $\{y_t\}$ 和 $\{z_t\}$ 序列都是平稳的；2) ε_{yt} 和 ε_{zt} 都是白噪声序列，其均值均为零，标准差为 σ_y 和 σ_z ；3) ε_{yt} 和 ε_{zt} 不相关。

式 1.1 和式 1.2 中最长的滞后期数为 1，因此构成了 1 阶向量自回归 first-order vector autoregression (VAR(1))。在这两个式子构成的模型中， y_t 和 z_t 相互影响，存在着反馈机制 Feedback。 $-b_{12}$ 度量了当期 z_t 变动一个单位对 y_t 的影响程度， γ_{12} 度量了 z_{t-1} 变动一个单位对 y_t 的影响程度。这里要注意， ε_{yt} 和 ε_{zt} 分别是 $\{y_t\}$ 和 $\{z_t\}$ 内部的新息 innovations(或冲击 shocks)，当然， b_{21} 不等于零的话， ε_{yt} 对当期的 z_t 存在非直接的影响， b_{12} 不等于零的话， ε_{zt} 对当期的 y_t 也存在着非直接的影响。

式 1.1 和式 1.2 不能采用 OLS 估计，因为当期的 y_t 和 z_t 存在相互影响，造成自变量和误差项相关。有必要对式 1.1 和式 1.2 进行变换，形成易于实际应用的形式。将此两式写成矩阵形式，即：

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \quad 1.3$$

简写成：

$$Bx_t = \Gamma_0 + \Gamma_1x_{t-1} + \varepsilon_t$$

$$\text{其中，} B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, X_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

上式两边同乘以 B^{-1} ，则形成了 VAR 模型的标准形式：

$$X_t = A_0 + A_1X_{t-1} + e_t \quad 1.4$$

其中， $A_0 = \Gamma_0B^{-1}$ ， $A_1 = \Gamma_1B^{-1}$ ，以及 $e_t = \varepsilon_tB^{-1}$ ，为了便于比较，也可以写成：

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \quad 1.5$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \quad 1.6$$

为了区别式 1.1、式 1.2 与式 1.5、式 1.6，前者称为结构向量自回归模型 structural VAR，后者称为标准向量自回归模型 standard VAR。通常所讲的 VAR 模型是 standard VAR，即式

1.5、式 1.6，除非有特别说明。这里要注意的是， e_{1t} 和 e_{2t} 都是由 ε_{yt} 和 ε_{zt} 的构成的，可以推导：

$$e_{1t} = \frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}} \quad 1.7$$

$$e_{2t} = \frac{\varepsilon_{zt} - b_{12}\varepsilon_{yt}}{1 - b_{12}b_{21}} \quad 1.8$$

由于 ε_{yt} 和 ε_{zt} 都是白噪声序列，可以推知 e_{1t} 和 e_{2t} 为零均值和常数方差，且二者之间存在着相关关系。以下以 e_{1t} 为例，推导 $\{e_{1t}\}$ 的统计性质。

对式 1.7 两边取期望，得：

$$E(e_{1t}) = E\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right) = 0 \quad 1.9$$

其方差为：

$$Var(e_{1t}) = E(e_{1t}^2) = E\left[\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right]^2 = \frac{\sigma_y^2 + b_{12}^2\sigma_z^2}{(1 - b_{12}b_{21})^2} \quad 1.10$$

式 1.10 可以看出，方差与时间 t 无关。 e_{1t} 和 e_{1t-i} 的自相关系数为：

$$E(e_{1t}e_{1t-i}) = E\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right)\left(\frac{\varepsilon_{y,t-i} - b_{12}\varepsilon_{z,t-i}}{1 - b_{12}b_{21}}\right) = 0 \quad i \neq 0 \quad 1.11$$

要特别注意的是， e_{1t} 和 e_{2t} 之间是存在相关关系，其协方差为：

$$E(e_{1t}e_{2t}) = E\left(\frac{\varepsilon_{yt} - b_{12}\varepsilon_{zt}}{1 - b_{12}b_{21}}\right)\left(\frac{\varepsilon_{zt} - b_{12}\varepsilon_{yt}}{1 - b_{12}b_{21}}\right) = \frac{-(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)}{(1 - b_{12}b_{21})^2} \quad 1.12$$

当 $b_{12} = b_{21} = 0$ 时，式 1.12 等于 0，即当 y_t 和 z_t 之间不存在当期相互影响时， e_{1t} 和 e_{2t} （冲击 shocks）之间是不存在相关关系的。以下给出 e_{1t} 和 e_{2t} 的方差/协方差矩阵定义式：

$$\Sigma = \begin{bmatrix} var(e_{1t}) & cov(e_{1t}, e_{2t}) \\ cov(e_{1t}, e_{2t}) & var(e_{2t}) \end{bmatrix} \quad 1.13$$

写成一般形式为：

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \quad 1.14$$

其中， $var(e_{it}) = \sigma_i^2$ ， $cov(e_{1t}, e_{2t}) = \sigma_{21} = \sigma_{12}$

三、VAR 模型的稳定性/平稳性 Stability or Stationary

VAR 模型平稳性或稳定性度量了多变量之间是否存在稳定的长期关系。

对于 AR(1)模型： $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$ ，平稳的充要条件是 $|\phi_1| < 1$ ，式 1.4 的 VAR 模型也有类似的结果。将 X_{t-1} 以式 1.4 的形式展开并代入到式 1.4 中，得：

$$X_t = A_0 + A_1(A_0 + A_1 X_{t-2} + e_{t-1}) + e_t = (I + A_1)A_0 + A_1^2 X_{t-2} + A_1 e_{t-1} + e_t$$

其中, $I = 2 \times 2$ 的单位阵 identity matrix。接下来, 同样方式, 经过 n 次迭代, 得:

$$X_t = (I + A_1 + \cdots + A_1^n)A_0 + A_1^{n+1}X_{t-n-1} + \sum_{i=0}^n A_1^i e_{t-i} \quad 1.15$$

这样一直迭代下去, 式 1.15 要收敛的话, 随着 $n \rightarrow \infty$, $A_1^n \rightarrow 0$ 。

以下证明平稳性条件。

将式 1.5 和式 1.6 改写成延迟算子形式, 即:

$$y_t = a_{10} + a_{11}By_t + a_{12}Bz_t + e_{1t} \quad 1.16$$

$$z_t = a_{20} + a_{21}By_t + a_{22}Bz_t + e_{2t} \quad 1.17$$

或

$$(1 - a_{11}B)y_t = a_{10} + a_{12}Bz_t + e_{1t} \quad 1.16$$

$$(1 - a_{22}B)z_t = a_{20} + a_{21}By_t + e_{2t} \quad 1.17$$

式 1.17 解出 z_t 得:

$$Bz_t = \frac{B(a_{20} + a_{21}By_t + e_{2t})}{(1 - a_{22}B)} \quad 1.18$$

代入式 1.16 得:

$$(1 - a_{11}B)y_t = a_{10} + a_{12} \frac{B(a_{20} + a_{21}By_t + e_{2t})}{(1 - a_{22}B)} + e_{1t} \quad 1.19$$

即将 $\{y_t\}$ 和 $\{z_t\}$ 的 VAR(1) 模型转换为 $\{y_t\}$ 的 2 阶差分方程, 解出 y_t 得:

$$y_t = \frac{a_{10}(1 - a_{22}) + a_{12}a_{20} + (1 - a_{22}B)e_{1t} + a_{12}e_{2t-1}}{(1 - a_{11}B)(1 - a_{22}B) - a_{12}a_{21}B^2} \quad 1.20$$

采用同样的方法, 可以解出 z_t 得:

$$z_t = \frac{a_{20}(1 - a_{11}) + a_{21}a_{10} + (1 - a_{11}B)e_{2t} + a_{21}e_{1t-1}}{(1 - a_{11}B)(1 - a_{22}B) - a_{12}a_{21}B^2} \quad 1.21$$

式 1.20 和式 1.21 具有相同的特征方程, 要 $\{y_t\}$ 和 $\{z_t\}$ 收敛, 则多项式 $(1 - a_{11}B)(1 - a_{22}B) - a_{12}a_{21}B^2 = 0$ 的根要在单位圆外。注意, 二阶差分方程的根要么是实根, 要么是复根, 要么收敛, 要么发散, 由于 $\{y_t\}$ 和 $\{z_t\}$ 具有相同的特征方程, 只要 a_{21} 和 a_{12} 不同时为零, 则两个特征方程就具有相同的特征根, $\{y_t\}$ 和 $\{z_t\}$ 的图像显示出近似的波动形状。

图 1.1 给出了四种典型的 VAR 模型的图像, 每张图的 $\{y_t\}$ 和 $\{z_t\}$ 的初始值 y_0 和 z_0 都设定为零, 模型采用的是式 1.5 和式 1.6。

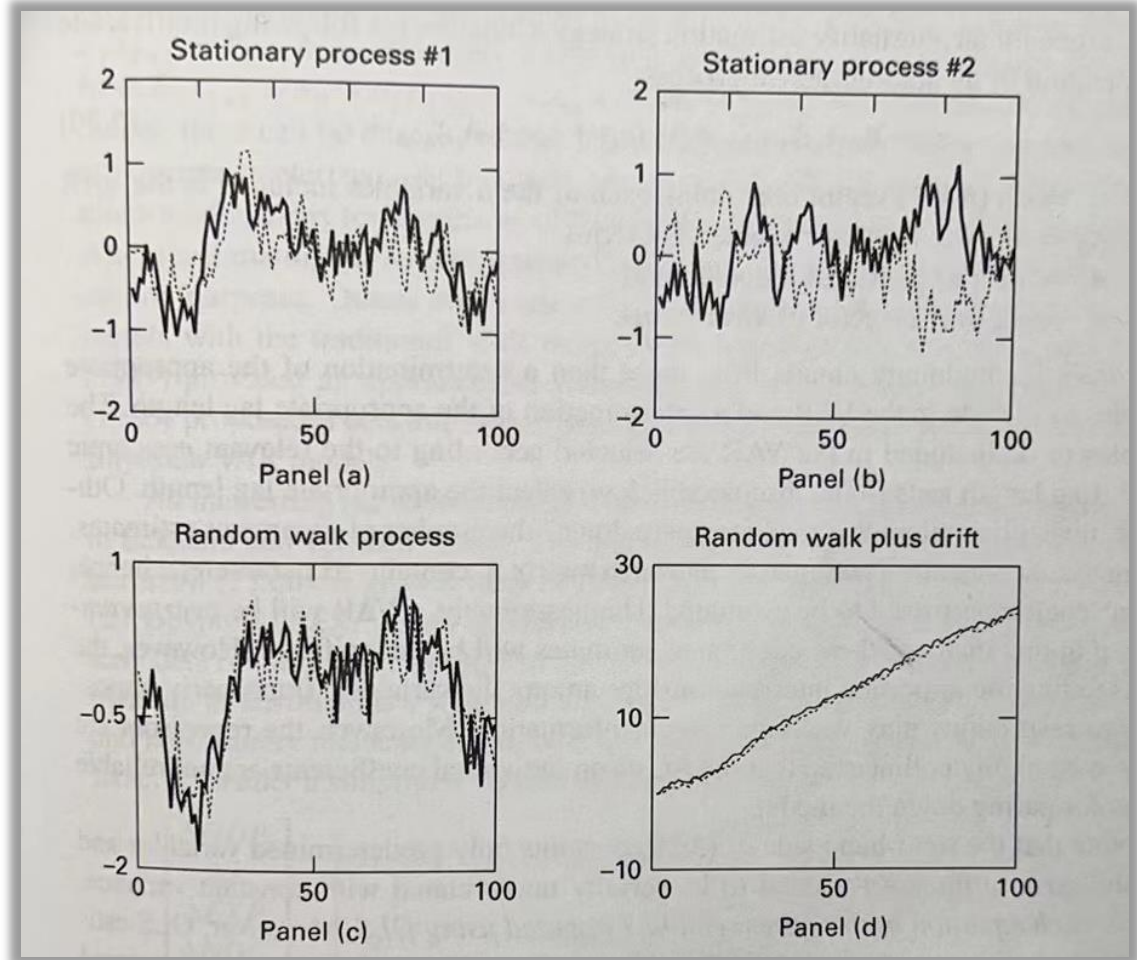


图 1.1 四种典型的 VAR 模型的图形

Panel(a)采用以下数据:

$$a_{10} = a_{20} = 0, \quad a_{11} = a_{22} = 0.7, \quad a_{12} = a_{21} = 0.2$$

对应的式 1.20 或式 1.21 的多项式 $(1 - a_{11}B)(1 - a_{22}B) - a_{12}a_{21}B^2 = 0$ 的根为 1.111 和 2.0, 都在单位圆外, $\{y_t\}$ 和 $\{z_t\}$ 是平稳序列, 解得 y_t 和 z_t 对应的特征根为 0.9 和 0.5。由于这些根是正的、实数的、小于 1 的, 所以收敛是直接的, 两个序列有一起移动的趋势。由于 a_{21} 是正的, y_t 的观测值比较大时, 就会导致 z_{t+1} 的观测值也比较大; 由于 a_{12} 也是正的, z_t 的大观测值也会导致 y_{t+1} 的观测值比较大。说明两个序列之间的协相关系数 Cross-Correlations Coefficient 是正的。

Panel(b)采用以下数据:

$$a_{10} = a_{20} = 0, \quad a_{11} = a_{22} = 0.5, \quad a_{12} = a_{21} = -0.2$$

也都是平稳序列, 其 y_t 和 z_t 对应的特征根为 0.7 和 0.3。由于 a_{12} 和 a_{21} 都是负数, z_t 的观测值和 y_{t+1} 的观测值呈现反向波动, y_t 的观测值和 z_{t+1} 的观测值也呈现反向波动, 从图中直

接观测到这种情形，说明两个序列之间的协相关系数 Cross-Correlations Coefficient 是负的。

Panel(c)采用的数据为： $a_{10} = a_{20} = 0$ ， $a_{11} = a_{22} = a_{12} = a_{21} = 0.5$ ，可以解出 y_t 和 z_t 对应的特征根均为 1，两个序列几乎没有恢复到恒定长期值的趋势。截距项 a_{10} 和 a_{20} 等于零，因此这张图表示多元随机游走模型，图形特征是 $\{y_t\}$ 和 $\{z_t\}$ 似乎是蜿蜒在一起的。

Panel(d)采用的数据是 $a_{10} = 0.5$ ， $a_{20} = 0$ ，其余同 Panel(c)，即在 Panel(c) 的 VAR 过程中包含一个非零的截距项 $a_{10} = 0.5$ ，该截距项起着“漂移 drift”的作用，图形特征是， $\{y_t\}$ 和 $\{z_t\}$ 似乎紧密地在一起移动，漂移项为两个系列增加了一个向右上方移动的趋势，同时也主导了两个序列的长期波动。

在实际分析过程中，我们的分析重点不是 VAR 模型的每个方程的系数参数，因为，每个方程的系数只是反映了一个局部的动态关系，不能捕捉全面复杂的互动过程。而是要关注某个给定变量的变化，对系统内其他变量产生的影响是正向的还是负向的，以及这个变量的变化在系统内会产生的影响会持续多久。这些信息，就可以通过考察 VAR 模型的关注脉冲效应函数和预测方差分解。

四、脉冲效应

脉冲效应是描绘 $\{y_t\}$ 和 $\{z_t\}$ 之间相互关系的一种方法，它指系统对其中某一种变量的一个随机冲击所做出的反应以及这种反应能持续多久。

对于 AR(p)模型，可以通过 Green 函数，写出对应的 MA(∞)形式，同样，对于 VAR(p)模型，也可以通过迭代变换，写出对应的 VMA(∞)形式，即 Vector Moving Average 向量移动平均模型，VMA 形式的参数构成了两个冲击序列 $\{\varepsilon_{yt}\}$ 和 $\{\varepsilon_{zt}\}$ 对两个序列 $\{y_t\}$ 和 $\{z_t\}$ 的影响模型，这就是脉冲效应函数（模型）Impulse Response Function。

以下以 VAR(1)模型为例，推导脉冲效应函数。

式 1.16 和式 1.17，可以写成矩阵形式，即：

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad 1.22$$

类似 AR(1)模型的迭代过程，将 y_{t-1} 和 z_{t-1} 写成式 1.22 的形式，即写成 $(t-1)$ 时刻的 VAR(1)模型并代入到式 1.22 中（具体迭代过程，这里省略），依次迭代，就可以得到：

$$\begin{aligned} \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_{1t-1} \\ e_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^2 \begin{bmatrix} e_{1t-2} \\ e_{2t-2} \end{bmatrix} \dots + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^k \begin{bmatrix} e_{1t-k} \\ e_{2t-k} \end{bmatrix} \\ &+ \dots \end{aligned} \quad 1.23$$

式 1.23 可以与 AR(1)模型的 MA(∞)形式进行比较，非常类似。其中：

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \quad 1.24$$

$$\bar{y} = \frac{a_{10}(1 - a_{22}) + a_{12}a_{20}}{(1 - a_{11})(1 - a_{22}) - a_{12}a_{21}} \quad 1.25$$

$$\bar{z} = \frac{a_{20}(1 - a_{11}) + a_{21}a_{10}}{(1 - a_{11})(1 - a_{22}) - a_{12}a_{21}} \quad 1.26$$

式 1.24 就是式 1.7 和式 1.8 的矩阵形式。将式 1.24 代入式 1.23，得：

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix} \quad 1.27$$

式 1.27 看上去太复杂，可以进行表达式简化，记：

$$\phi_i = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \quad 1.28$$

则式 1.27 可以简化成：

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix} \quad 1.29$$

式 1.29 就是 VMA(∞)形式，有助于观察两个序列 $\{y_t\}$ 和 $\{z_t\}$ 之间的相互作用，这里的 $\phi_{11}(i), \phi_{12}(i), \phi_{21}(i)$ 和 $\phi_{22}(i)$ 就是脉冲效应函数 Impulse Response Functions，即 ϕ_i 描绘了 ε_{yt} 和 ε_{zt} 冲击（注意：这里的冲击不是 e_{1t} 和 e_{2t} ）对两个序列 $\{y_t\}$ 和 $\{z_t\}$ 在整个时间路径的影响。例如：系数 $\phi_{12}(0)$ 度量 ε_{zt} 单位冲击（脉冲）对 y_t 的瞬时影响； $\phi_{11}(1)$ 和 $\phi_{12}(1)$ 分别度量了 ε_{yt-1} 和 ε_{zt-1} 单位冲击（脉冲）对 y_t 的一期影响，当然，也反应了 ε_{yt} 和 ε_{zt} 单位变化对 y_{t+1} 的一期影响。通过不同期数脉冲响应函数系数求和，可以得到 ε_{yt} 和 ε_{zt} 中单位冲击（脉冲）的累积效应。例如，在 n 期之后， ε_{zt} 对 y_{t+n} 值的影响为 $\phi_{12}(n)$ ，则在 n 个周期之后， ε_{zt} 对序列 $\{y_t\}$ 的累积影响为：

$$\sum_{i=0}^n \phi_{12}(i) \quad 1.30$$

通常将 ϕ_i 的值与 i 画在图上，就形成了脉冲效应图，按影响对象的不同，一共有四张，即 ε_{yt} 分别对 y_t 和 z_t 的影响， ε_{zt} 分别对 z_t 和 y_t 的影响。

五、预测方差分解

方差分解是描绘 $\{y_t\}$ 和 $\{z_t\}$ 之间相互关系的另一种方法，它的实质是将系统的预测均方差分解成系统中各变量冲击所做的贡献。

以 VAR(1)模型为例，进行方差分解分析。式 1.22 写成单方程形式，即：

$$\mathbf{X}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{e}_t \quad 1.31$$

其中：

$$\mathbf{X}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \mathbf{A}_0 = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix}, \mathbf{A}_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \mathbf{X}_{t-1} = \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix}, \mathbf{e}_t = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

现对 $t+1$ 时刻的值进行预测，即以 t 时刻为基点，进行向前一步预测 one-step-ahead forecasting，即：

$$\hat{\mathbf{X}}_t(1) = E(\mathbf{X}_{t+1}) = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{X}_t \quad 1.32$$

当时间到达 $t+1$ 时刻时，可以得到 \mathbf{X}_{t+1} 的观测值，此时向前一步预测误差为：

$$\mathbf{X}_{t+1} - E(\mathbf{X}_{t+1}) = \mathbf{e}_t \quad 1.32$$

同样， $t+2$ 时刻观测值的表达式为：

$$\mathbf{X}_{t+2} = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{X}_{t+1} + \mathbf{e}_{t+2} = \mathbf{A}_0 + \mathbf{A}_1(\mathbf{A}_0 + \mathbf{A}_1 \mathbf{X}_t + \mathbf{e}_{t+1}) + \mathbf{e}_{t+2} \quad 1.33$$

而以 t 时刻为基点，向前两步预测 two-step ahead forecasting 的表达式为：

$$\hat{\mathbf{X}}_t(2) = E(\mathbf{X}_{t+2}) = (\mathbf{I} + \mathbf{A}_1)\mathbf{A}_0 + \mathbf{A}_1^2 \mathbf{X}_t \quad 1.34$$

则向前两步预测的误差为：

$$\mathbf{X}_{t+2} - E(\mathbf{X}_{t+2}) = \mathbf{e}_{t+2} + \mathbf{A}_1 \mathbf{e}_{t+1} \quad 1.35$$

推广到向前 n 步预测的表达式为：

$$E(\mathbf{X}_{t+n}) = (\mathbf{I} + \mathbf{A}_1 + \mathbf{A}_1^2 + \cdots + \mathbf{A}_1^{n-1})\mathbf{A}_0 + \mathbf{A}_1^n \mathbf{X}_t \quad 1.36$$

则向前 n 步预测误差为：

$$\mathbf{X}_{t+n} - E(\mathbf{X}_{t+n}) = \mathbf{e}_{t+n} + \mathbf{A}_1 \mathbf{e}_{t+n-1} + \mathbf{A}_1^2 \mathbf{e}_{t+n-2} + \cdots + \mathbf{A}_1^{n-1} \mathbf{e}_{t+1} \quad 1.37$$

另一种表达误差的形式是上一节推导的 VMA(∞)形式，即式 1.29，写成单方程形式为：

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \boldsymbol{\varepsilon}_{t-i} \quad 1.29$$

其中：

$$\boldsymbol{\mu} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix}, \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}, \boldsymbol{\Phi}_i = \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix}$$

采用式 1.29，向前一步预测误差为：

$$\mathbf{X}_{t+1} - E(\mathbf{X}_{t+1}) = \Phi_0 \boldsymbol{\varepsilon}_{t+1} \quad 1.30$$

拓展到 $t+n$ 时刻的观测值为：

$$\mathbf{X}_{t+n} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \Phi_i \boldsymbol{\varepsilon}_{t+n-i} \quad 1.31$$

则向前 n 步预测误差为：

$$\mathbf{X}_{t+n} - E(\mathbf{X}_{t+n}) = \sum_{i=0}^{n-1} \Phi_i \boldsymbol{\varepsilon}_{t+n-i} \quad 1.32$$

注意： $\mathbf{X}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}$, $\Phi_i = \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix}$, $\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$ 。将式 1.32 的结果只应用在 y_t 上，则可得 y_t 到向前 n 步预测误差，即：

$$\begin{aligned} y_{t+n} - E(y_{t+n}) &= \phi_{11}(0)\varepsilon_{yt+n} + \phi_{11}(1)\varepsilon_{yt+n-1} + \cdots + \phi_{11}(n-1)\varepsilon_{yt+1} + \phi_{12}(0)\varepsilon_{zt+n} \\ &\quad + \phi_{12}(1)\varepsilon_{zt+n-1} + \cdots + \phi_{12}(n-1)\varepsilon_{zt+1} \end{aligned} \quad 1.33$$

由此可以得到 y_t 向前 n 步预测误差的方差 $\sigma_y^2(n)$ 为：

$$\begin{aligned} \sigma_y^2(n) &= \sigma_y^2(\phi_{11}(0)^2 + \phi_{11}(1)^2 + \cdots + \phi_{11}(n-1)^2) \\ &\quad + \sigma_z^2(\phi_{12}(0)^2 + \phi_{12}(1)^2 + \cdots + \phi_{12}(n-1)^2) \end{aligned} \quad 1.34$$

从式 1.34 可以看出，随着 n 的增大，预测误差的方差也逐渐增大，同时， $\sigma_y^2(n)$ 是由两部分构成，一部分是由 y_t 序列的冲击 ε_{yt} 产生的，另一部分是由 z_t 序列的冲击 ε_{zt} 产生的，这两部分的对总的预测误差方差的贡献分别为：

$$\frac{\sigma_y^2(\phi_{11}(0)^2 + \phi_{11}(1)^2 + \cdots + \phi_{11}(n-1)^2)}{\sigma_y^2(n)} \quad \text{和} \quad \frac{\sigma_z^2(\phi_{12}(0)^2 + \phi_{12}(1)^2 + \cdots + \phi_{12}(n-1)^2)}{\sigma_y^2(n)}$$

预测误差方差分解告诉我们序列中由于“自身”冲击与其他变量的冲击所造成的波动的比例。如果 ε_{zt} 冲击无法解释 $\{y_t\}$ 在所有预测水平上的预测误差方差，可以说 $\{y_t\}$ 序列完全是外生的 exogenous。在这种情况下， $\{y_t\}$ 的变动将与 ε_{zt} 冲击和 $\{z_t\}$ 序列无关。另一种极端情况， ε_{zt} 冲击可以解释 $\{y_t\}$ 序列在所有预测水平上的所有预测误差方差，则 $\{y_t\}$ 序列完全是内生的 endogenous。在应用研究中，一个变量在短期内几乎可以解释其所有的预测误差方差，而在较长期内所占比例较小。

六、滞后期数的确定

原则上，可以在 VAR 模型中加入大量变量。可以构造一个由 n 个方程的 VAR 模型，每个方程包含所有 n 个变量的 p 个滞后项。但随着变量的增加，自由度很快就被削弱了。例

如，使用 12 个滞后项的月数据，在每个方程中包含一个额外的变量就会减少额外的 12 个自由度，直观理解：自由度减少，估计的方差会变大。因此，在 VAR 模型中增加变量和滞后阶数时，要有所考量。

包含 n 个方程的 VAR 模型可以写成下式：

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ \vdots \\ A_{n0} \end{bmatrix} + \begin{bmatrix} A_{11}(B) & A_{12}(B) & \cdots & A_{1n}(B) \\ A_{21}(B) & A_{22}(B) & \cdots & A_{2n}(B) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}(B) & A_{n2}(B) & \cdots & A_{nn}(B) \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{nt} \end{bmatrix} \quad 1.35$$

其中， A_{i0} 代表第 i 个方程的截距项； $A_{ij}(B)$ 代表滞后算子 B 的多项式，例如： $A_{11}(B)$ 就是 $a_{11}(1), a_{11}(2), \dots$ ，括号中的数字代表滞后期数。 n 个方程具有同样的滞后期数，因此 $A_{ij}(B)$ 具有同样的自由度。 e_{it} 代表第 i 个方程的扰动项，通常是白噪声，且不同方程的扰动项之间存在相关关系，其方差/协方差可以写成 $n \times n$ 的矩阵形式，这里用 Σ 表示。

通常是根据相关的经济理论来确定 VAR 模型中所包含的变量个数，而确定适当的滞后阶数（长度）就需要依据一定的检验方法。每个方程中的不同变量允许有不同的滞后长度，但为了保持模型的对称性及有效使用 OLS 估计，建议对所有方程的不同变量采用相同的滞后长度，因为只要每个方程中有相同的自变量，OLS 估计才是一致且渐近有效的。在 VAR 模型中，大的滞后长度会快速消耗自由度，如果滞后期数为 p ，则 n 个方程中的每个方程都包含 np 个系数参数加上截距项。适当的滞后长度选择是至关重要的。如果 p 太小，则模型有可能识别错误；如果 p 太大，自由度就消耗了。检验滞后长度，要从最长的合理长度或给定自由度的最长可行长度开始，进行 VAR 模型参数估计，并形成残差的方差/协方差矩阵 Σ 。比如：如使用季度数据，通常认为 3 年的数据能有效捕捉系统的动态特征，滞后长度的检验就从 12 开始。把具有 12 期滞后长度的模型所形成的残差的方差/协方差矩阵记作 Σ_{12} 。

接下来考虑滞后长度是否可以从 12 减少到 8，这样可以使每个方程的待估参数减少 $4n$ 个。目标是确定滞后长度 8 是否适用于所有方程，没必要对滞后长度 9 到 12 的 VAR 模型进行逐一的 F 检验。交叉限制的正确检验方法是似然比检验 likelihood ratio test¹，即采用相同的观测值，使用 8 个滞后长度的 VAR 模型进行重新估计，并获得残差的方差/协方差矩阵 Σ_8 。注意： Σ_8 的结果呈现的是整个系统（VAR 模型），该系统有 n 个方程，相对于无限制条件的 Σ_{12} 所对应的 VAR 模型， Σ_8 的每个方程中有 $4n$ 个限制， n 个方程总共有 $4n^2$ 个限制。似然比统计量为：

$$N(\ln|\Sigma_8| - \ln|\Sigma_{12}|)$$

¹ 似然比检验是指如果约束条件成立，则相应约束模型与非约束模型的极大似然函数值应该是近似相等。

在实际的经济研究中，依据常见经济数据的样本容量，建议用调整过的似然比统计量，即：

$$(N - c)(\ln|\Sigma_8| - \ln|\Sigma_{12}|)$$

其中： N 代表观测值个数（样本容量）， c 代表无限制条件的每个方程待估参数个数， $\ln|\Sigma_n|$ 代表行列式的自然对数。在上述例子中， $c = 1 + 12n$ ，无限制模型（即滞后长度为 12 的模型）每个方程对每个变量都有 12 个滞后长度，再加上一个截距项。

该统计量具有渐近的 χ^2 分布，其自由度等于系统中限制的数量。在上述例子中，每个方程有 $4n$ 个限制，系统中总共有 $4n^2$ 个限制。如果样本似然比统计量的值大于显著性水平（通常为 95%），则滞后长度的减少（即从 12 减少到 8）不具有约束力，此时，滞后 12 长度的模型优于滞后 8 长度的模型。如果样本似然比统计量的值小于显著性水平（通常为 95%），也就是说， $\ln|\Sigma_8|$ 和 $\ln|\Sigma_{12}|$ 近似相等，则滞后长度的减少（即从 12 减少到 8）具有约束力，或者说滞后长度 8 的模型优于滞后长度 12 的模型。进一步，可以以滞后长度 8 的模型为无约束模型，滞后长度 4 的模型为有约束模型，再进行似然比检验，统计量如下：

$$(N - c)(\ln|\Sigma_4| - \ln|\Sigma_8|)$$

如此进行顺序检验 sequential testing procedure，最终确定合适的滞后长度。

似然比检验用于大样本条件下，在小样本条件下，可以使用信息准则法，常用的统计量是 AIC 和 SBC，即：

$$AIC = N\ln|\Sigma| + 2(n^2p + n) \quad 1.36$$

$$SBC = N\ln|\Sigma| + (n^2p + n)\ln N \quad 1.37$$

其中， $|\Sigma|$ 代表残差方差/协方差矩阵的行列式 determinant， N 代表观测值个数（样本容量）， n 代表方程个数或变量个数， p 代表滞后长度。一方面参数增加会带来好处，即会提高拟合度，即式 1.36 和式 1.37 的第一项所表达的，增加参数提高了 R^2 （拟合优度），降低残差平方和，也就是使得第一项变小；另一方面参数增加也会带来坏处，即使得自由度下降，影响估计精度，也就是式 1.36 和式 1.37 的第二项所表达的，增加参数，第二项会变大。好处大于坏处的反映标志就是整体变小，也就是说选择最优滞后期数应使得 AIC 和 SBC 值达到最小。

七、格兰杰因果 Granger Causality

计量经济模型的建立过程，本质上是利用回归分析工具处理一个经济变量对其他经济变量的依存性问题（即数量变动关系），但这并不意味着这个经济变量和其他经济变量之间必

定存在因果关系（即谁影响谁的问题）。由于没有因果关系的变量之间往往有很好的拟合效果，把回归模型的解释变量和被解释变量倒过来也能有很好的拟合效果，因此，回归分析不能检验因果关系的存在性，更不用说识别因果关系的方向。Granger 因果只是表示了时间序列之间领先和滞后关系，如果 A 对于预测 B 是有用的，则 A 是 B 的 Granger 原因。而真正的因果关系表示的是一种逻辑上的顺序，通过逻辑思考得到 A 是 B 发生的原因。

检验 Granger 因果关系，就是在方程中一个变量的滞后项是否会影响到另一个变量。在有 p 个滞后长度的双方程模型中（式 1.36），当且仅当 $A_{21}(B)$ 的所有系数都等于零时，即：

$$a_{21}(1) = a_{21}(2) = a_{21}(3) = \cdots = a_{21}(p) = 0$$

$\{y_t\}$ 不是 $\{z_t\}$ 的格兰杰原因，也就是说，如果 $\{y_t\}$ 滞后项的加入不能提高 $\{z_t\}$ 的预测效果。

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \end{bmatrix} + \begin{bmatrix} A_{11}(B) & A_{12}(B) \\ A_{21}(B) & A_{22}(B) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad 1.36$$

如果 VAR 模型中的所有变量都是平稳的，那么 Granger 因果检验就是构造的 F 统计量，即：

$$F = \frac{(SSR_r - SSR_u)/p}{SSR_u/(N - 2p)} \quad 1.37$$

$$H_0: a_{21}(1) = a_{21}(2) = a_{21}(3) = \cdots = a_{21}(p) = 0$$

$$H_1: a_{21}(1), a_{21}(2), a_{21}(3), \dots, a_{21}(p) \text{ 至少一个不为零}$$

其中， SSR_r 表示施加了约束条件（即 H_0 成立）后 VAR 模型的残差平方和， SSR_u 表示未施加约束条件下 VAR 模型的残差平方和， p 代表最大的滞后长度， N 代表观测值个数（样本容量）。在 H_0 成立的条件下，式 1.37 渐近服从自由度为 $(p, N - 2p)$ 的 F 分布。

可以将上述结果推广到具有 n 个变量（方程）的 VAR 模型（式 1.35）。 $A_{ij}(B)$ 代表了变量 j 的各期滞后值（各期历史值）对变量 i 的影响系数，当 $A_{ij}(B)$ 的值都为零时，变量 j 就不是变量 i 的格兰杰原因。检验时，滞后长度 p 的选取是任意的，如果 $p = 1$ 就存在显著影响，则不必再做滞后长度大于 1 的检验，如果 $p = 1$ 不存在显著影响，则应该再做滞后长度更长的检验。一般来说，要检验若干不同滞后长度的格兰杰因果检验，且结论相同时，才能下最终的结论。这里要强调，不存在协整关系的非平稳变量之间不能进行格兰杰因果检验。

最后，区别格兰杰原因和外生性 exogeneity 这两个概念。 $\{y_t\}$ 不是 $\{z_t\}$ 的格兰杰原因，代表的是 $\{y_t\}$ 的历史值对 $\{z_t\}$ 的现值不会产生影响，即 $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ 对 z_t 不会产生影响。而外生性指的是 $\{y_t\}$ 的现值对 $\{z_t\}$ 的现值不产生影响，即 y_t 对 z_t 不会产生影响。