

金融工程 2024 年秋 · 时间序列

第 12 讲：向量自回归模型应用

授课人：刘 岩

武汉大学经管学院金融系

2024 年 11 月 18 日

本讲内容

- ① VAR 模型估计
- ② VAR 模型的分析

本节内容

① VAR 模型估计

② VAR 模型的分析

VAR 模型的估计：OLS

- 给定 VAR(p) 模型

$$\mathbf{X}_t = \mathbf{c} + \Phi_1 \mathbf{X}_{t-1} + \cdots + \Phi_p \mathbf{X}_{t-p} + \boldsymbol{\varepsilon}_t$$

- 上述方程组中每一行，都是一个动态回归模型 \Rightarrow 对每一行进行 OLS 估计即可
- 将每行的 OLS 估计系数集合，得到 $\hat{\mathbf{c}}, \hat{\Phi}_i \forall i$ ，进一步可计算样本残差向量

$$\hat{\boldsymbol{\varepsilon}}_t = \mathbf{X}_t - \hat{\mathbf{c}} - \hat{\Phi}_1 \mathbf{X}_{t-1} - \cdots - \hat{\Phi}_p \mathbf{X}_{t-p}$$

- 最后，通过 $\hat{\boldsymbol{\varepsilon}}_t$ 估计残差协方差矩阵

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T - p - (pK + 1)} \sum_{t=p+1}^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t^\top$$

VAR 模型的估计：极大似然

- 假设 $\varepsilon_t \sim N(\mathbf{0}, \mathbf{\Omega})$, 定义 $\mathbf{Y}_t = [1, \mathbf{X}_{t-1}^\top, \dots, \mathbf{X}_{t-p}^\top]^\top$, $\mathbf{\Pi}^\top = [c, \mathbf{\Phi}_1, \dots, \mathbf{\Phi}_p]$, 则 $\mathbf{X}_t | \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p} \sim N(\mathbf{\Pi}^\top \mathbf{Y}_t, \mathbf{\Omega})$
- 给定初值 $\mathbf{X}_0, \dots, \mathbf{X}_{1-p}$, 样本似然函数对数为

$$\begin{aligned} L(\boldsymbol{\theta} | \mathcal{X}) &= -\frac{TK}{2} \log(2\pi) - \frac{T}{2} \log(\det(\mathbf{\Omega})) \\ &= -\frac{1}{2} \sum_{t=1}^T [(\mathbf{X}_t - \mathbf{\Pi}^\top \mathbf{Y}_t)^\top \mathbf{\Omega}^{-1} (\mathbf{X}_t - \mathbf{\Pi}^\top \mathbf{Y}_t)] \end{aligned}$$

- $\mathbf{\Pi}^\top$ 的极大似然估计为 (证明见 Hamilton 教科书 Ch. 11)

$$\hat{\mathbf{\Pi}}^\top = \left[\sum_{t=1}^T \mathbf{X}_t \mathbf{Y}_t^\top \right] \left[\sum_{t=1}^T \mathbf{Y}_t \mathbf{Y}_t^\top \right]^{-1}$$

容易验证上式第 k 行就是 VAR 第 k 行 OLS 估计所得系数

本节内容

① VAR 模型估计

② VAR 模型的分析

VAR 模型的分析：脉冲响应 (impulse response)

- 冲击 ε_{kt} 的变化，如何影响 $X_{\ell t+j}$ 的取值？计算

$$\frac{\partial X_{\ell t+j}}{\partial \varepsilon_{kt}}, \quad j = 0, 1, 2, \dots$$

- VAR 模型下， t 期 k 变量对 $t+j$ 期 ℓ 变量的影响，很容易通过递推模拟得到
- 原理：平稳 VAR(p) 过程有 MA(∞) 表示

$$X_{t+j} = \mu + \sum_{s=0}^{\infty} \Psi_s \varepsilon_{t+j-s}$$

注意，此处的矩阵 Ψ_s 与上一讲中 VAR(p) 变形为 VAR(1) 时所写 Ψ 涵义不同

- $Y_t = \zeta + \Psi Y_{t-1} + e_t$ 中 Ψ 为滞后项 Y_{t-1} 系数矩阵

VAR(p) 过程的 MA(∞) 表示

- 首先将 VAR(p) 过程 $X_t = c + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \varepsilon_t$ 改写为 VAR(1) 过程 $Y_t = \zeta + \Psi Y_{t-1} + e_t$
- 再利用 Y_t 进行 MA(∞) 展开:

$$\begin{aligned}
 Y_t &= \Psi^2 Y_{t-2} + (\Psi + I)\zeta + \Psi e_{t-1} + e_t = \dots \\
 &= \underbrace{\lim_{j \rightarrow \infty} \Psi^j Y_{t-j}}_{\text{平稳性} \Rightarrow \lim_{j \rightarrow \infty} \Psi^j = 0} + \sum_{s=0}^{\infty} \Psi^s \zeta + \sum_{s=0}^{\infty} \Psi^s e_{t-s} \\
 &= (I - \Psi)^{-1} \zeta + \sum_{s=0}^{\infty} \Psi^s e_{t-s}
 \end{aligned}$$

故上式中的 Ψ^s 即对应前页 MA(∞) 中的 Ψ_s 矩阵

脉冲响应与因果性

- 通常情况下, $\mathbb{E}\varepsilon_t\varepsilon_t^\top = \mathbf{\Omega}$ 不是对角矩阵
 - ε_{kt} 与 ε_{lt} 有相关性
- 这意味着 ε_{kt} 的变动不仅包括变量 k 的信息, 还包括影响其他变量 $l \neq k$ 的信息, 相应的脉冲响应 $\frac{\partial X_{lt+j}}{\partial \varepsilon_{kt}}$ 缺乏明确的“因果性”涵义
 - 因果性: 假设其他变量都不变, 只改变 X_{kt} 时 X_{lt+j} 如何变化? 此时 X_{kt} 的变化全部归结为 t 期外生冲击
- VAR 变量的递归排序 (recursive ordering) 提供了一种最简单、直观的方法, 将 $\varepsilon_{kt}, \varepsilon_{lt}$ 等的变化分解为互不相关的部分

Cholesky 分解

- 对 $K \times K$ 实对称正定矩阵 Ω , 存在下三角阵 A 及对角矩阵 $D = \text{diag}(d_{11}, \dots, d_{KK})$ 使得

$$\Omega = ADA^T$$

且满足 $d_{kk} > 0 \forall k = 1, \dots, K$, 以及 A 对角线均为 1

- 构造性证明: 对 Ω 不断使用行列消元, 直到只剩对角线为止
 - Wikipedia、网络查询或 Hamilton/Hayashi 之类教科书
- 进一步定义 $D^{\frac{1}{2}} = \text{diag}(d_{11}^{1/2}, \dots, d_{KK}^{1/2})$ 以及 $P = AD^{\frac{1}{2}}$, 则可得到

$$\Omega = PP^T$$

其中 P 为下三角阵, 称为正定阵矩阵 Ω 的 Cholesky 分解

递归排序与冲击的 Cholesky 分解

- 固定 VAR 变量的一个排序, 由原冲击向量协方差矩阵 Cholesky 分解 $\Omega = PP^T$ 定义新的冲击向量

$$v_t = P^{-1} \varepsilon_t$$

则 $\mathbb{E}v_t v_t^T = \mathbb{E}P^{-1} \varepsilon_t \varepsilon_t^T (P^T)^{-1} = P^{-1} \mathbb{E}[\varepsilon_t \varepsilon_t^T] (P^T)^{-1} = P^{-1} \Omega (P^T)^{-1} = I_K$, 故 v_t 的各个分量无相关性

- 具体而言, $Pv_t = \varepsilon_t$ 有如下形式

$$\begin{bmatrix} d_{11}^{1/2} & 0 & 0 & \cdots & 0 \\ p_{21} & d_{22}^{1/2} & 0 & \cdots & 0 \\ p_{31} & p_{32} & d_{33}^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & p_{K3} & \cdots & d_{KK}^{1/2} \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ \vdots \\ v_{Kt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \vdots \\ \varepsilon_{Kt} \end{bmatrix}$$

递归排序下与 Cholesky 分解下冲击的涵义

- 给定 VAR 观测 (observable) 冲击 ε_t (在给定排序下的) Cholesky 分解 $v_t = P^{-1}\varepsilon_t$, 则 v_t 各分量彼此正交 (零相关)
- 称 v_t 各分量为该 VAR 在变量递归排序下识别出的结构性 (structural) 冲击
 - 此种结构性冲击的识别方法, 称为递归识别 (recursive identification)
- 故 $\varepsilon_{1t} = d_{11}^{1/2}v_{1t}$, $\varepsilon_{Kt} = p_{K1}v_{1t} + \dots + p_{K,K-1}v_{K-1,t} + d_{KK}^{1/2}v_{Kt}$
 - 即排序最后的变量 K 的观测冲击 ε_{Kt} , 可能受到同期所有结构性冲击的影响
 - 换言之, 此时对排序最后变量对应的结构性冲击 v_{Kt} 的识别是最保守的 (conservative): 剔除所有其他变量对应结构性冲击后, 观测冲击 ε_{Kt} 中剩下的变动部分

递归排序与脉冲响应

- 利用变量的递归排序与冲击的 Cholesky 分解, VAR 过程的 MA(∞) 表示可写为

$$\mathbf{X}_{t+j} = \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t+j-s} = \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \mathbf{P} \mathbf{v}_{t+j-s}$$

- 相应的, v_{kt} 对 \mathbf{X}_{t+j} 的边际影响即脉冲响应为

$$\partial \mathbf{X}_{t+j} / \partial v_{kt} = \boldsymbol{\Psi}_j \mathbf{p}_k$$

其中 \mathbf{p}_k 为 \mathbf{P} 的第 k 列

- 与之对比, $\partial \mathbf{X}_{t+j} / \partial \varepsilon_{kt} = \boldsymbol{\Psi}_j \mathbf{e}_k$, 其中 \mathbf{e}_k 为第 k 个位置为 1、其余均为 0 的单位列向量

脉冲响应示例：Stock and Waston 2001

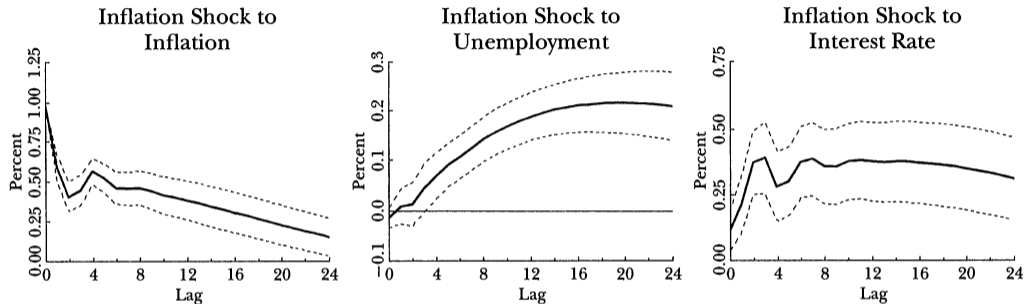
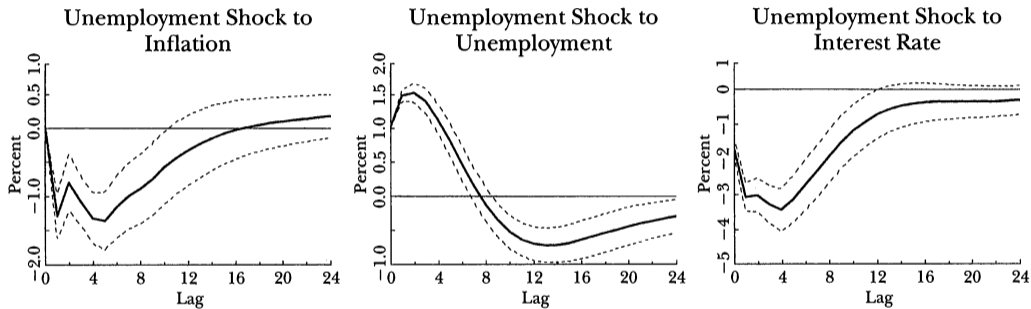


图: 通货膨胀冲击

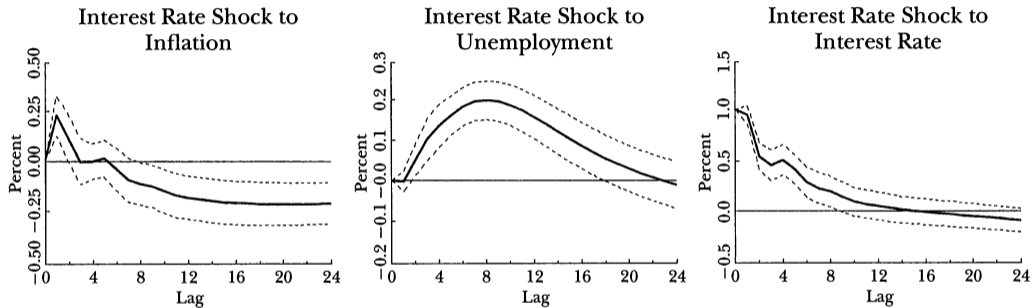
虚线表示 ± 1 标准差的置信区间，显著性水平约 66%；常用的置信区间计算使用自助法 (bootstrap)

脉冲响应示例：Stock and Waston 2001



图：失业率冲击

脉冲响应示例：Stock and Waston 2001



图：利率冲击

货币政策 VAR 分析中的递归识别

- 此类 VAR 分析中，利率变量作为货币政策代理变量，通常排序最后：允许利率观测冲击 ε_t^i 对利率 i_t 的影响中，包括通货膨胀、失业率的结构性冲击部分 v_t^π, v_t^u
- 潜在的货币政策时点 (timing) 假设：货币政策观测冲击 ε_t^i 不但受到结构性冲击 v_t^i 的影响，还取决于当期通胀、失业率（产出）冲击的影响，即当期货币政策的制定，会考虑当期通胀、失业率（产出）的变化
- 同时，通胀排序第一，则通胀的当期变动 ε_t^π ，即观测冲击，只受到通胀本身的当期结构性冲击 v_t^π 的影响，不受失业率（产出）与货币政策的影响：这反映了短期价格刚性这一凯恩斯经济学的常规假定

方差分解

- VAR 模型方差分解：将 $t + j$ 期变量 l 预测均方误差分解为不同变量冲击项 $\varepsilon_{k,t+s}$, $s = 1, \dots, j$ 所带来的贡献：由

$$\mathbf{X}_{t+j} - \hat{\mathbf{X}}_{t+j|t} = \varepsilon_{t+j} + \Psi_1 \varepsilon_{t+j-1} + \dots + \Psi_{j-1} \varepsilon_{t+1}$$

得到

$$\begin{aligned} \text{MSE}(\hat{\mathbf{X}}_{t+j|t}) &= \mathbb{E} \left[(\mathbf{X}_{t+j} - \hat{\mathbf{X}}_{t+j|t})(\mathbf{X}_{t+j} - \hat{\mathbf{X}}_{t+j|t})^\top \right] \\ &= \mathbf{\Omega} + \Psi_1 \mathbf{\Omega} \Psi_1^\top + \dots + \Psi_{j-1} \mathbf{\Omega} \Psi_{j-1}^\top \end{aligned}$$

- 当不同变量冲击项 ε_{kt} 存在相关性时，上述表达式无法将均方预测误差分解到各个冲击项上 \Rightarrow 先将冲击项拆分为互不相关的部分
- 常用方法仍然是 Cholesky 分解

递归排序与方差分解

- 给定一组递归变量排序，则有相应的冲击项 Cholesky 分解

$$\varepsilon_t = P v_t = p_1 v_{1t} + p_2 v_{2t} + \cdots + p_K v_{Kt}$$

$$\Rightarrow \Omega = p_1 p_1^T \text{var}(v_{1t}) + p_2 p_2^T \text{var}(v_{2t}) + \cdots + p_K p_K^T \text{var}(v_{Kt}) = p_1 p_1^T + \cdots + p_K p_K^T$$

其中 p_1, \dots, p_K 为矩阵 P 的各个列向量

- 相应的均方预测误差可分解为

$$MSE(\hat{X}_{t+j|t}) = \sum_{k=1}^K \underbrace{\left[p_k p_k^T + \Psi_1 p_k p_k^T \Psi_1^T + \cdots + \Psi_{j-1} p_k p_k^T \Psi_{j-1}^T \right]}_{\text{第 } k \text{ 项冲击 } v_{k,t+s}, s = 1, \dots, j \text{ 的贡献}}$$

- $j \rightarrow \infty$ 时，方差分解收敛到一个常数：此时可以衡量变量 k 的冲击项对 X_t 协方差（矩阵）的贡献

方差分解示例: Stock and Waston 2001

B.iii. Variance Decomposition of R

<i>Forecast Horizon</i>	<i>Forecast Standard Error</i>	<i>Variance Decomposition (Percentage Points)</i>		
		π	u	R
1	0.85	2	19	79
4	1.84	9	50	41
8	2.44	12	60	28
12	2.63	16	59	25

Granger 因果检验

- Granger 因果检验：变量 k 的滞后项 $1, \dots, p$ ，是否对变量 l 的当期值，有显著的影响
 - 在 X_{lt} 的回归中，检验 $X_{kt-1}, \dots, X_{kt-p}$ 的系数是否同时为 0 \Rightarrow Wald 或 F 检验
 - Granger 因果性不是真的因果性，而主要是时间上的领先滞后关系
- 也可以检验 $X_t = [X_{1t}^T, X_{2t}^T]^T$ 的两组变量间的 Granger 因果关系：对下述回归中的系数矩阵 $A_2 = 0$ 与 $B_1 = 0$ 进行检验

$$X_{1t} = c_1 + A_1^T Y_{1t} + A_2^T Y_{2t} + \varepsilon_{1t}$$

$$X_{2t} = c_2 + B_1^T Y_{1t} + B_2^T Y_{2t} + \varepsilon_{2t}$$

其中 A, B, Y 的定义与第 5 页类似

Granger 检验示例: Stock and Waston 2001

A. Granger-Causality Tests

Dependent Variable in Regression

<i>Regressor</i>	π	u	R
π	0.00	0.31	0.00
u	0.02	0.00	0.00
R	0.27	0.01	0.00

表中数字为 F -检验的 p -值; 小样本 + 正态假设下, Wald 统计量/参数约束个数 l 等于 F 统计量; 大样本下, F 统计量乘以 l 服从 $\chi^2(l)$ 分布