金融学 2022 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

2023 年 11 月 27 日

本讲内容

1. VAR 模型估计
2. VAR 模型的分析

本节内容

VAR 模型的估计：OLS

- 给定 VAR(p) 模型

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

- 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 $\hat{\boldsymbol{c}}$, $\widehat{\boldsymbol{\Phi}}_i\ \forall i$，进一步可计算样本残差向量

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

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

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

VAR 模型的估计：极大似然

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

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

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

$$\widehat{\boldsymbol{\Pi}}^{\mathsf{T}} = \left[\sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{Y}_t^{\mathsf{T}} \right] \left[\sum_{t=1}^{T} \boldsymbol{Y}_t \boldsymbol{Y}_t^{\mathsf{T}} \right]^{-1}$$

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

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

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

- 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

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

注意，此处的矩阵 $\boldsymbol{\Psi}_s$ 与上一讲中 VAR(p) 变形为 VAR(1) 时所写 $\boldsymbol{\Psi}$ 涵义不同

  - $\boldsymbol{Y}_t = \boldsymbol{\zeta} + \boldsymbol{\Psi}\boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$ 中 $\boldsymbol{\Psi}$ 为滞后项 $\boldsymbol{Y}_{t-1}$ 系数矩阵

VAR(p) 过程的 MA(∞) 表示

- 首先将 VAR(p) 过程 $\boldsymbol{X}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1\boldsymbol{X}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{X}_{t-p} + \boldsymbol{\varepsilon}_t$ 改写为 VAR(1) 过程 $\boldsymbol{Y}_t = \boldsymbol{\zeta} + \boldsymbol{\Psi}\boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$
- 再利用 $\boldsymbol{Y}_t$ 进行 MA(∞) 展开：

$$
\begin{aligned}
\boldsymbol{Y}_t &= \boldsymbol{\Psi}^2\boldsymbol{Y}_{t-2} + (\boldsymbol{\Psi} + \boldsymbol{I})\boldsymbol{\zeta} + \boldsymbol{\Psi}\boldsymbol{e}_{t-1} + \boldsymbol{e}_t = \cdots \\
&= \underbrace{\lim_{j\to\infty} \boldsymbol{\Psi}^j\boldsymbol{Y}_{t-j}}_{\text{平稳性}\Rightarrow\lim_{j\to\infty}\boldsymbol{\Psi}^j=\boldsymbol{0}} + \sum_{s=0}^{\infty}\boldsymbol{\Psi}^s\boldsymbol{\zeta} + \sum_{s=0}^{\infty}\boldsymbol{\Psi}^s\boldsymbol{e}_{t-s} \\
&= (\boldsymbol{I} - \boldsymbol{\Psi})^{-1}\boldsymbol{\zeta} + + \sum_{s=0}^{\infty}\boldsymbol{\Psi}^s\boldsymbol{e}_{t-s}
\end{aligned}
$$

故上式中的 $\boldsymbol{\Psi}^s$ 即对应前页 MA(∞) 中的 $\boldsymbol{\Psi}_s$ 矩阵

脉冲响应与因果性

- 通常情况下，$\mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\mathsf{T}} = \boldsymbol{\Omega}$ 不是对角矩阵
  - ε_{kt} 与 $\varepsilon_{\ell t}$ 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 $\ell \neq k$ 的信息，相应的脉冲响应 $\frac{\partial X_{\ell t+j}}{\partial \varepsilon_{kt}}$ 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 $X_{\ell t+j}$ 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 $\varepsilon_{kt}, \varepsilon_{\ell t}$ 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 $\boldsymbol{\Omega}$，存在下三角阵 $\boldsymbol{A}$ 及对角矩阵 $\boldsymbol{D} = \mathrm{diag}(d_{11}, \ldots, d_{KK})$ 使得

$$\boldsymbol{\Omega} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^{\top}$$

  且满足 $d_{kk} > 0\ \forall k = 1, \ldots, K$，以及 $\boldsymbol{A}$ 对角线均为 1
  - 构造性证明：对 $\boldsymbol{\Omega}$ 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 $\boldsymbol{D}^{\frac{1}{2}} = \mathrm{diag}\left(d_{11}^{1/2}, \ldots, d_{KK}^{1/2}\right)$ 以及 $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{D}^{\frac{1}{2}}$，则可得到

$$\boldsymbol{\Omega} = \boldsymbol{P}\boldsymbol{P}^{\top}$$

  其中 $\boldsymbol{P}$ 为下三角阵，称为正定阵矩阵 $\boldsymbol{\Omega}$ 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 $\boldsymbol{\Omega} = \boldsymbol{P}\boldsymbol{P}^\top$ 定义新的冲击向量

$$\boldsymbol{v}_t = \boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_t$$

则 $\mathbb{E}\boldsymbol{v}_t\boldsymbol{v}_t^\top = \mathbb{E}\boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top(\boldsymbol{P}^\top)^{-1} = \boldsymbol{P}^{-1}\mathbb{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top](\boldsymbol{P}^\top)^{-1} = \boldsymbol{P}^{-1}\boldsymbol{\Omega}(\boldsymbol{P}^\top)^{-1} = \boldsymbol{I}_K$，故 $\boldsymbol{v}_t$ 的各个分量无相关性

- 具体而言，$\boldsymbol{P}\boldsymbol{v}_t = \boldsymbol{\varepsilon}_t$ 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 $\boldsymbol{\varepsilon}_t$ （在给定排序下的）Cholesky 分解 $\boldsymbol{v}_t = \boldsymbol{P}^{-1}\boldsymbol{\varepsilon}_t$，则 $\boldsymbol{v}_t$ 各分量彼此正交（零相关）
- 称 $\boldsymbol{v}_t$ 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 $\varepsilon_{1t} = d_{11}^{1/2} v_{1t}$， $\varepsilon_{Kt} = p_{K1} v_{1t} + \cdots + p_{K,K-1} v_{K-1,t} + d_{KK}^{1/2} v_{Kt}$
  - 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)：提出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

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

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

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

  其中 $\boldsymbol{p}_k$ 为 $\boldsymbol{P}$ 的第 k 列

- 与之对比，$\partial \boldsymbol{X}_{t+j} / \partial \varepsilon_{kt} = \boldsymbol{\Psi}_j \boldsymbol{e}_k$，其中 $\boldsymbol{e}_k$ 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

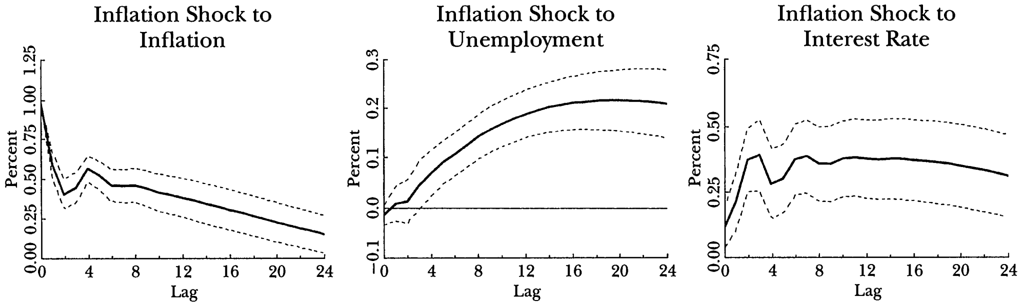

图: 通货膨胀冲击

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

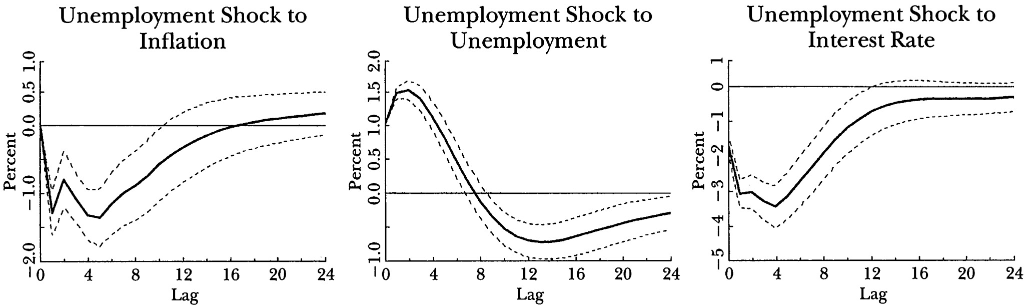

图: 失业率冲击

脉冲响应示例：Stock and Waston 2001

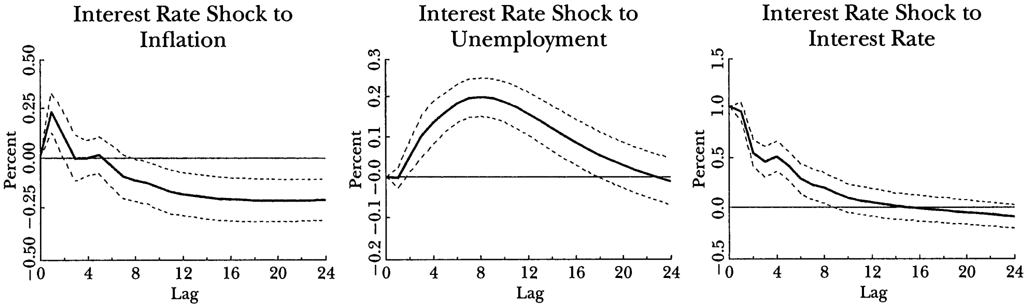

图: 利率冲击

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 $\varepsilon_{k,t+s}$，$s=1,\ldots,j$ 所带来的贡献：由

$$\boldsymbol{X}_{t+j}-\hat{\boldsymbol{X}}_{t+j|t}=\boldsymbol{\varepsilon}_{t+j}+\boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t+j-1}+\cdots+\boldsymbol{\Psi}_{j-1}\boldsymbol{\varepsilon}_{t+1}$$

得到

$$\begin{aligned}MSE\left(\hat{\boldsymbol{X}}_{t+j|t}\right)&=\mathbb{E}\left[\left(\boldsymbol{X}_{t+j}-\hat{\boldsymbol{X}}_{t+j|t}\right)\left(\boldsymbol{X}_{t+j}-\hat{\boldsymbol{X}}_{t+j|t}\right)^{\mathsf{T}}\right]\\&=\boldsymbol{\Omega}+\boldsymbol{\Psi}_1\boldsymbol{\Omega}\boldsymbol{\Psi}_1^{\mathsf{T}}+\cdots+\boldsymbol{\Psi}_{j-1}\boldsymbol{\Omega}\boldsymbol{\Psi}_{j-1}^{\mathsf{T}}\end{aligned}$$

- 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

$$\begin{aligned} \boldsymbol{\varepsilon}_t &= \boldsymbol{P}\boldsymbol{v}_t = \boldsymbol{p}_1 v_{1t} + \boldsymbol{p}_2 v_{2t} + \cdots + \boldsymbol{p}_K v_{Kt} \\ \Rightarrow \boldsymbol{\Omega} &= \boldsymbol{p}_1\boldsymbol{p}_1^\top \mathrm{var}(v_{1t}) + \boldsymbol{p}_2\boldsymbol{p}_2^\top \mathrm{var}(v_{2t}) + \cdots + \boldsymbol{p}_K\boldsymbol{p}_K^\top \mathrm{var}(v_{Kt}) = \boldsymbol{p}_1\boldsymbol{p}_1^\top + \cdots + \boldsymbol{p}_K\boldsymbol{p}_K^\top \end{aligned}$$

  其中 $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K$ 为矩阵 $\boldsymbol{P}$ 的各个列向量

- 相应的均方预测误差可分解为

$$MSE\left(\hat{\boldsymbol{X}}_{t+j|t}\right) = \sum_{k=1}^{K} \Big[\underbrace{\boldsymbol{p}_k\boldsymbol{p}_k^\top + \boldsymbol{\Psi}_1\boldsymbol{p}_k\boldsymbol{p}_k^\top\boldsymbol{\Psi}_1^\top + \cdots + \boldsymbol{\Psi}_{j-1}\boldsymbol{p}_k\boldsymbol{p}_k^\top\boldsymbol{\Psi}_{j-1}^\top}_{\text{第}k\text{项冲击}v_{k,t+s},\ s=1,\ldots,j\text{的贡献}} \Big]$$

- $j \to \infty$ 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 $\boldsymbol{X}_t$ 协方差（矩阵）的贡献

方差分解示例：Stock and Waston 2001

B.iii. Variance Decomposition of R

Forecast Horizon	Forecast Standard Error	Variance Decomposition (Percentage Points)		
		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 $X_{\ell t}$ 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 $\boldsymbol{X}_t = [\boldsymbol{X}_{1t}^\top, \boldsymbol{X}_{1t}^\top]^\top$ 的两组变量间的 Granger 因果关系：对下述回归中的系数矩阵 $\boldsymbol{A}_2 = \boldsymbol{0}$ 与 $\boldsymbol{B}_1 = \boldsymbol{0}$ 进行检验

$$
\begin{aligned}
\boldsymbol{X}_{1t} &= \boldsymbol{c}_1 + \boldsymbol{A}_1^\top \boldsymbol{Y}_{1t} + \boldsymbol{A}_2^\top \boldsymbol{Y}_{2t} + \boldsymbol{\varepsilon}_{1t} \\
\boldsymbol{X}_{2t} &= \boldsymbol{c}_2 + \boldsymbol{B}_1^\top \boldsymbol{Y}_{1t} + \boldsymbol{B}_2^\top \boldsymbol{Y}_{2t} + \boldsymbol{\varepsilon}_{2t}
\end{aligned}
$$

其中 $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{Y}$ 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		
Regressor	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 $\chi^2(\ell)$ 分布