EAE1223: Econometria III

Aula 6 - Modelos vetoriais autorregressivos

Luis A. F. Alvarez

25 de abril de 2024

Vetores aleatórios e matriz de variância-covariância

- Um vetor (coluna) aleatório $\boldsymbol{X}$, com valores em $\mathbb{R}^d$, é uma função com domínio no espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$ e contradomínio em $\mathbb{R}^d$.
  - Posto de outra forma, um vetor aleatório $\boldsymbol{X}$ é um vetor cujas entradas $\boldsymbol{X}_j$, $j = 1 \ldots, d$, são variáveis aleatórias com valores reais.
- Para um vetor aleatório $\boldsymbol{X}$, a matriz de variância-covariância, $\mathbb{V}[X]$, é definida como:

$$\mathbb{V}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}'] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}']$$

- Matriz de variância-covariância é $d \times d$, simétrica, positiva semidefinida, com entrada (i, j) iguais a:

$$\mathbb{V}[\boldsymbol{X}]_{i,j} = \text{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j)\,.$$

Matriz de covariância entre dois vetores aleatórios

- Seja $\boldsymbol{X}$ um vetor aleatório em $\mathbb{R}^d$, e $\boldsymbol{Y}$ um vetor aleatório em $\mathbb{R}^p$, definimos a matriz de covariância entre $\mathbb{V}[\boldsymbol{X}]$ e $\boldsymbol{Y}$ como:

$$\text{cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathbb{E}[\boldsymbol{X}\boldsymbol{Y}'] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}'] \,.$$

- Matriz $d \times p$ em que a entrada (i, j) é igual a:

$$\text{cov}(\boldsymbol{X}, \boldsymbol{Y})_{i,j} = \text{cov}(\boldsymbol{X}_i, \boldsymbol{Y}_j)$$

Processo vetorial fracamente estacionário

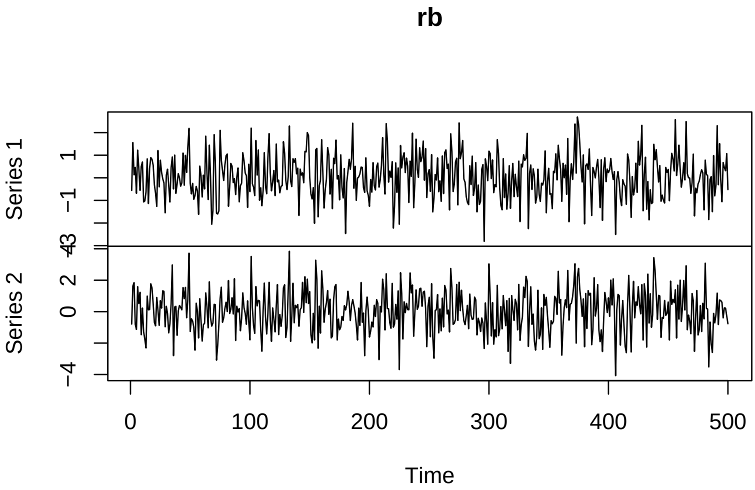
- Um processo estocástico vetorial é uma coleção de vetores aleatórios definidos em $\mathbb{R}^d$, indexados por um conjunto $\mathcal{I}$, isto é $\{\boldsymbol{X}_t : t \in \mathcal{I}\}$, onde cada $\boldsymbol{X}_t$ é vetor aleatório em $\mathbb{R}^d$.
  - Cada entrada $j = 1 \ldots d$ define um processo estocástico com valores reais $\{\boldsymbol{X}_{j,t} : t \in \mathcal{I}\}$ descrevendo a evolução da j-ésima entrada ao longo de $\mathcal{I}$.
- Um processo estocástico vetorial $\{\boldsymbol{X}_t : t \in \mathcal{T}\}$ indexado no tempo $\mathcal{T}$ é dito fracamente estacionário se:
  1. $\mathbb{E}[\boldsymbol{X}_t] = \boldsymbol{\mu}$ para todo $t \in \mathcal{T}$.
  2. $\mathbb{V}[\boldsymbol{X}_t] = \Sigma_0$ para todo $t \in \mathcal{T}$, com $\text{tr}(\Sigma_0) < \infty$.
  3. $\text{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h}) = \Sigma_h$, para todo $t \in \mathcal{T}$, $h \in \mathcal{N}$.
- Extensão do conceito de série de tempo estacionária para o caso vetorial.
- Pedimos estabilidade das covariâncias contemporâneas e extemporâneas entre as entradas dos vetores.
- **Obs:** se processo vetorial é fracamente estacionário, cada uma das entradas $\{\boldsymbol{X}_{j,t} : t \in \mathcal{T}\}$, $j = 1, \ldots, d$, é fracamente estacionária.

Ruído branco vetorial

- Um processo estocástico vetorial $\{\boldsymbol{X}_t : t \in \mathcal{T}\}$ com valores em $\mathbb{R}^d$ é dito um ruído branco vetorial se:
  - $\mathbb{E}[\boldsymbol{X}_t] = \mathbf{0}_{d\times 1}$, para todo $t \in \mathcal{T}$.
  - $\mathbb{V}[\boldsymbol{X}_t] = \Sigma_0$ para todo $t \in \mathcal{T}$, com $\mathrm{tr}(\Sigma_0) < \infty$.
  - $\mathrm{cov}(\boldsymbol{X}_t, \boldsymbol{X}_l) = \mathbf{0}_{d\times d}$, para todo $t \neq l$.
- No ruído branco vetorial, permitimos associação contemporânea entre as entradas do vetor, mas não há nem autodependência nem dependência cruzada entre as entradas no tempo.

Ruído branco vetorial Gaussiano

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$$



Modelos vetoriais autorregressivos

- Considere d séries de tempo $\{X_{j,t} : t \in \mathbb{Z}\}$, $j = 1, \ldots, d$.
- Dizemos que estas séries definem um processo vetorial autorregressivo de ordem p, ou VAR(p), se, para todo $j = 1, \ldots, d$ e $t \in \mathbb{Z}$:

$$X_{j,t} = a_{j,0} + \sum_{l=1}^{p} a_{j,1,l} X_{1,t-l} + \sum_{l=1}^{p} a_{j,2,l} X_{2,t-l} + \ldots + \sum_{l=1}^{p} a_{j,d,l} X_{d,t-l} + \epsilon_{j,t}$$

$$= a_{j,0} + \sum_{k=1}^{d} \sum_{l=1}^{p} a_{j,k,l} X_{k,t-l} + \epsilon_{j,t} \,,$$

onde $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{d,t})'$ é um ruído branco vetorial.

- No VAR(p), assim como no AR(p), evolução em cada uma das d variáveis depende do que ocorreu nela mesma nos últimos p períodos.
- Mas além disso, trajetória depende do que ocorreu nos últimos p períodos nas demais variáveis.
- Série depende também de uma inovação $\epsilon_{j,t}$, imprevisível com base no passado, mas que pode estar contemporaneamente associada às demais inovações nas outras equações (choques comuns).

VAR(p) em notação vetorial

- Um VAR(p) pode ser escrito, compactamente, em notação vetorial.
- De fato, definindo $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})'$, podemos escrever o sistema como

$$\boldsymbol{X}_t = \boldsymbol{a}_0 + \boldsymbol{A}_1 \boldsymbol{X}_{t-1} + \boldsymbol{A}_2 \boldsymbol{X}_{t-2} + \ldots + \boldsymbol{A}_p \boldsymbol{X}_{t-p} + \epsilon_t \,,$$

onde

$$\boldsymbol{a}_0 = \begin{bmatrix} a_{1,0} \\ a_{2,0} \\ \vdots \\ a_{d,0} \end{bmatrix} \qquad \boldsymbol{A}_l = \begin{bmatrix} a_{1,1,l} & a_{1,2,l} & \cdots & a_{1,d,l} \\ a_{2,1,l} & a_{2,2,l} & \cdots & a_{2,d,l} \\ \vdots & \vdots & \cdots & \vdots \\ a_{d,1,l} & a_{d,2,l} & \cdots & a_{d,d,l} \end{bmatrix}$$

VAR(p) estacionário

- Um VAR(p) é dito estacionário se o processo vetorial resultante é estacionário.
- Condição para estacionariedade do VAR(p) é que:

$$|z| \leq 1 \implies \det(\mathbb{I}_{d \times d} - \boldsymbol{A}_1 z - \boldsymbol{A}_2 z^2 \ldots - \boldsymbol{A}_p z^p) \neq 0$$

- Todas as raízes do polinômio $\phi(z) = \det(\mathbb{I}_{d \times d} - \boldsymbol{A}_1 z - \boldsymbol{A}_2 z^2 \ldots - \boldsymbol{A}_p z^p)$ devem se encontrar fora do círculo unitário.
- Nesta aula, focaremos na estimação de VAR(p) estacionários.
    - Portanto, cada uma das séries deverá estar devidamente estacionarizada.

Estimação do VAR(p)

- Dado um painel com observações de d séries durante T períodos, $\{\boldsymbol{X}_{j,t}\}_{t=1}^{T}$, como estimar os parâmetros de um VAR(p)?
- Maneira mais simples é estimar os parâmetros através de MQO, equação a equação.
- Isto é, estimamos os parâmetros da j-ésima equação resolvendo:

$$\min_{b_{0,j},\{b_{j,k,l}\}_{k,l}} \sum_{t=p+1}^{T} \left(\boldsymbol{X}_{j,t} - b_{0,j} - \sum_{k=1}^{d} \sum_{l=1}^{p} b_{j,k,l} X_{j,t-l} \right)^2$$

- Para cada equação, rodamos uma regressão com T observações e $p \times d + 1$ parâmetros.

SUR

- A estimação dos parâmetros de um VAR por MQO, equação a equação, é potencialmente ineficiente.
- Isso se deve ao fato de que os choques em uma equação j, por serem (potencialmente) contemporaneamente correlacionados com os choques das demais equações, podem conter informação relevante para estimar os parâmetros de outras equações
  - Erro em uma equação é informativo sobre a outra equação.
- Seja $\hat{\Sigma}_0$ um estimador preliminar de $\mathbb{V}(\epsilon_t)$.
  - Por exemplo, estimador da variância-covariância com base nos resíduos do estimador de MQO equação a equação.
- O estimador de *seemingly unrelated regression* (SUR) dos parâmetros de um VAR propõe-se a estimar os parâmetros do sistema conjuntamente, minimizando:

$$\min_{\boldsymbol{b}_0, \boldsymbol{B}_1, \ldots \boldsymbol{B}_p} \sum_{t=p+1}^{T} \left(\boldsymbol{X}_t - \boldsymbol{b}_0 - \sum_{l=1}^{p} \boldsymbol{B}_l \boldsymbol{X}_{t-l} \right)' \hat{\Sigma}^{-1} \left(\boldsymbol{X}_t - \boldsymbol{b}_0 - \sum_{l=1}^{p} \boldsymbol{B}_l \boldsymbol{X}_{t-l} \right)$$

Máxima verossimilhança condicional

- Sob a hipótese auxiliar:

$$\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{d\times 1}, \Sigma_0)$$

  podemos calcular a verossimilhança de $\boldsymbol{X}_{t+p+1}, \ldots \boldsymbol{X}_T$, condicional a $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p$.
- Estimador de máxima verossimilhança condicional estima simultaneamente $\boldsymbol{a}_0$, os $\boldsymbol{A}_l$ e Σ_0.

Relação entre os métodos de estimação

- Se o espaço de parâmetros é irrestrito, no sentido de que os parâmetros $\boldsymbol{a}_0$ e os $\boldsymbol{A}_l$ podem tomar qualquer valor real, os estimadores de MQO equação a equação, SUR, e máxima verossimilhança condicional são numericamente iguais.
- Se impomos restrições nos parâmetros $\boldsymbol{a}_0$ e os $\boldsymbol{A}_l$, estimador de MQO equação a equação é consistente, embora SUR e máxima verossimilhança condicional sejam mais eficientes.
- Se, além disso, impomos restrições na matriz Σ_0, os três estimadores são consistentes, mas máxima verossimilhança condicional é o mais eficiente.

Selecionando a ordem p de um VAR

- Para selecionar a ordem p de um VAR, podemos adotar generalizações dos critérios de informação vistos para modelos multivariados:

$$\text{AIC}(p) = \ln(\det(\hat{\Sigma}(p))) + \frac{2}{T}(d^2 p + d)\,,$$

$$\text{HQ}(p) = \ln(\det(\hat{\Sigma}(p))) + \frac{2\log\log(T)}{T}(d^2 p + d)\,,$$

$$\text{SC}(p) = \ln(\det(\hat{\Sigma}(p))) + \frac{\log(T)}{T}(d^2 p + d)\,,$$

  onde $\hat{\Sigma}(p)$ é a matriz de variância-covariância dos resíduos do modelo VAR estimado com p defasagens.

- Métodos oferecem aproximações ao erro quadrático médio de previsão um passo à frente, corrigidas do viés induzido por *overfitting*.

Testando a inclusão de defasagens

- Partindo de um VAR(p). Podemos testar se há necessidade de incluir a p-ésima defasagem.
- Especificamente, gostaríamos de testar.

$$H_0 : \boldsymbol{A}_p = \boldsymbol{0}_{d\times d} \quad H_1 : \boldsymbol{A}_p \neq \boldsymbol{0}_{d\times d} \tag{1}$$

- Teste pode ser conduzido através da estatística de razão de verossimilhança.

$$\text{LR} = 2 \cdot (\hat{L}_p - \hat{L}_{p-1}) \tag{2}$$

onde $\hat{L}_j$ é a log-verossimilhança maximizada do estimador de máxima verossimilhança condicional, do modelo que inclui j defasagens.

- Sob hipótese nula, com T grande, estatística segue distribuição qui-quadrado com d^2 graus de liberdade.

Diagnóstico dos resíduos

- Escolhida uma ordem p do VAR, podemos avaliar o comportamento dos erros.
- Em particular, para $h \in \mathbb{N}$, podemos testar a hipótese nula:

$$H_0 : \mathrm{cov}(\boldsymbol{\epsilon}_t, \epsilon_{t-j}) = \mathbf{0}_{d \times d}, \quad j = 1, \ldots, h\,.$$

  usando um análogo vetorial do teste de Ljung-Box.
- A esse teste, damos o nome de Portmanteau.
- Também é possível estudar a normalidade conjunta dos erros, através de um análogo vetorial do teste de Jarque-Bera.

Inclusão de componentes determinísticos

- No modelo VAR apresentado, incluímos tão somente um intercepto entre os componentes determinísticos.
- Para séries *trend stationary*, devemos fazer o *detrending* dos dados para a estimação do VAR estacionário.
- Se todas as séries apresentam tendência, um jeito mais eficiente de fazer isso é estimar conjuntamente a tendência ao VAR, ajustando o modelo:

$$\boldsymbol{X}_t = \boldsymbol{a}_0 + \boldsymbol{a}_1 t + \boldsymbol{A}_1 \boldsymbol{X}_{t-1} + \boldsymbol{A}_2 \boldsymbol{X}_{t-2} + \ldots + \boldsymbol{A}_p \boldsymbol{X}_{t-p} + \boldsymbol{\epsilon}_t\,,$$

- Se há evidência de sazonalidade nas séries, podemos ajustar o VAR incluindo um conjunto de *dummies* sazonais.

$$\boldsymbol{X}_t = \boldsymbol{a}_0 + \gamma \boldsymbol{d}_t + \boldsymbol{A}_1 \boldsymbol{X}_{t-1} + \boldsymbol{A}_2 \boldsymbol{X}_{t-2} + \ldots + \boldsymbol{A}_p \boldsymbol{X}_{t-p} + \boldsymbol{\epsilon}_t\,,$$

Previsão

- Assim como nos modelos ARMA, a previsão em um modelo VAR se faz de maneira recursiva

Referências