EAE1223: Econometria III

Aula 6 - Modelos vetoriais autorregressivos

Luis A. F. Alvarez

3 de junho de 2025

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, $\mathbb{V}[\boldsymbol{X}]$, 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:

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

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 $\boldsymbol{X}$ 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

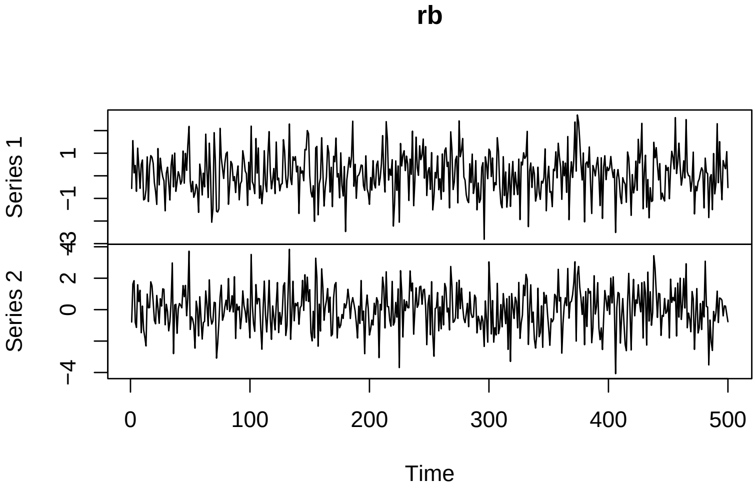
- Um processo estocástico vetorial é uma coleção de vetores aleatórios definidos em $\mathbb{R}^d$, indexados por um conjunto $\mathcal{I}$, isto é $\{\boldsymbol{X}_t : t \in \mathcal{I}\}$, onde cada $\boldsymbol{X}_t$ é vetor aleatório em $\mathbb{R}^d$.
  - Cada entrada $j = 1 \dots d$ define um 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 $\mathrm{tr}(\Sigma_0) < \infty$.
  3. $\mathrm{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h}) = \Sigma_h$, 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 $\{\boldsymbol{X}_{j,t} : t \in \mathcal{T}\}$, $j = 1, \dots, d$, 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 $t \in \mathbb{Z}$:

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

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

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

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, $\{\boldsymbol{X}_{j,t}\}_{t=1}^{T}$, $j = 1, \ldots, d$, 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:

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

- 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 $\mathbb{V}(\boldsymbol{\epsilon}_t)$.
  - 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:

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

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 $\boldsymbol{X}_{p+1}, \ldots \boldsymbol{X}_T$, 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 multivariadas dos critérios de informação vistos para modelos univariados:

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

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

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

  onde $\hat{\Sigma}(p)$ é a matriz de variância-covariância dos resíduos do modelo VAR estimado com p defasagens.
- Assim como no caso univariado, no cálculo dos diferentes $\hat{\Sigma}(p)$, usamos os resíduos dos mesmos $T - \mathrm{p}_{max}$ períodos, onde p_{max} é a ordem máxima a se testar.
- 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.

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

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

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

  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:

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

  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.
  - Preocupação usualmente secundária. Normalidade garante que o MLE condicional domine outros estimadores condicionais, para além do MQO equação a equação e SUR. Também é importante na construção de intervalos de predição, assim como nos ARMA.

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 fora da amostra

- Assim como nos modelos ARMA, a previsão em um modelo VAR se faz de maneira recursiva.
- Dados estimadores dos parâmetros do VAR, previsão um passo à frente de $\boldsymbol{X}_{T+1}$ é dada por:

$$\hat{\boldsymbol{X}}_{T+1} = \hat{\boldsymbol{a}}_0 + \hat{\boldsymbol{A}}_1 \boldsymbol{X}_T + \hat{\boldsymbol{A}}_2 \boldsymbol{X}_{T-1} + \ldots + \hat{\boldsymbol{A}}_p \boldsymbol{X}_{T+1-P}$$

- A previsão dois passos à frente é dada por:

$$\hat{\boldsymbol{X}}_{T+2} = \hat{\boldsymbol{a}}_0 + \hat{\boldsymbol{A}}_1 \hat{\boldsymbol{X}}_{T+1} + \hat{\boldsymbol{A}}_2 \boldsymbol{X}_T + \ldots + \hat{\boldsymbol{A}}_p \boldsymbol{X}_{T+2-P},$$

  e assim recursivamente.
- Previsões do VAR estacionário convergem para média incondicional estimada do processo:

$$\lim_{h \to \infty} \hat{\boldsymbol{X}}_{T+h} = (\mathbb{I} - \hat{\boldsymbol{A}}_1 - \hat{\boldsymbol{A}}_2 \ldots - \hat{\boldsymbol{A}}_p)^{-1} \hat{\boldsymbol{a}}_0$$

Previsão *online*

- Em diversas situações, a previsão em tempo real ocorre num cenário em que as divulgações dos valores mais recentes das variáveis ocorrem com defasagem no tempo.
  - Dados para taxa inflação em $T+1$ são divulgados em um certo dia do mês $T+2$, enquanto dados da SELIC média de um mês podem ser obtidas antes, a depender do calendário do COPOM.
- Nessas janelas intermediárias, em que alguns dados de $T+1$ já foram divulgados, mas faltam outros, é interessante entender como calcular as previsões para as variáveis que ainda não foram divulgadas para $T+1$, mas incorporando toda a informação disponível
  - Se as defasagens entre divulgações forem muito longas e com um padrão bem-definido, uma possível solução, do ponto de vista preditivo, é trabalhar com as variáveis que saem muito antes de modo adiantado, i.e. incluímos a variável adiantada (em $t+1$) entre os $\boldsymbol{X}_t$ no modelo VAR.
    - No entanto, essa estratégia é menos útil quando o calendário de divulgações é mais irregular e com menos intervalos entre divulgações.
    - Além disso, defasar algumas séries pode complicar a interpretação dos resultados das estimações.

Previsão *online* e *nowcasting*

- A alternativa para os cenários com divulgação irregular, que mantém a interpretabilidade dos coeficientes, consiste em realizar um procedimento conhecido como *nowcasting*.
- Formalmente, suponha que podemos dividir o VAR em dois conjuntos de variáveis $\boldsymbol{X}_t = [\boldsymbol{X}_t^e, \boldsymbol{X}_t^l]$.
  - Para as variáveis em $\boldsymbol{X}_t^e$, observamos os dados até $T+1$.
  - Para as variáveis em $\boldsymbol{X}_t^l$, observamos os dados até T.
- Sejam $\hat{\boldsymbol{a}}_0$, $\{\hat{\boldsymbol{A}}_k\}_{k=1}^p$ e $\hat{\Sigma}_0$ estimadores dos parâmetros do VAR com base em dados **até** T.
  - Para esse conjunto de dados, não faltam informações para resolver o problema de otimização que produz o estimador.
- Ideia é incorporar na projeção de $\boldsymbol{X}_t^l$ nossa melhor projeção para $\boldsymbol{\epsilon}_{T+1}^l$.
  - Quando não temos dados de $T+1$, sabemos que nossa melhor projeção para esse termo é $\mathbf{0}$, visto que se trata de um ruído branco.
  - No entanto, conhecimento de $\boldsymbol{X}_{T+1}^e$ é possivelmente informativo sobre ruído branco, visto que há potencial correlação contemporânea nestes.

Nowcasting

- Possível mostrar que melhor projeção (linear) para $\boldsymbol{\epsilon}^{l}_{T+1}$ com base no conhecimento do passado e $\boldsymbol{X}^{e}_{T+1}$ é:

$$\boldsymbol{s}^{l}_{T+1} = \Sigma_{l,e}\Sigma^{-1}_{e,e}\boldsymbol{\epsilon}^{e}_{T+1}$$

  onde $\Sigma_{e,e} = \mathbb{V}[\boldsymbol{\epsilon}^{e}_{t}]$ e $\Sigma_{l,e} = \text{cov}[\boldsymbol{\epsilon}^{l}_{t}, \boldsymbol{\epsilon}^{e}_{t}]$.
- Ideia é estimar $\boldsymbol{s}^{l}_{T+1}$ como:

$$\hat{\boldsymbol{s}}^{l}_{T+1} = \hat{\Sigma}_{0,l,e}\hat{\Sigma}^{-1}_{0,e,e}(\boldsymbol{X}^{e}_{T+1} - \hat{\boldsymbol{X}}^{e}{}_{T+1|T}),$$

  onde $\hat{\boldsymbol{X}}^{e}{}_{T+1|T}$ é a projeção usual do VAR com dados até T, para $T+1$.
- Projetamos $\boldsymbol{X}^{l}_{T+1}$ como:

$$\hat{\boldsymbol{X}}^{l}_{T+1} = \hat{\boldsymbol{X}}^{l}{}_{T+1|T} + \hat{\boldsymbol{s}}^{l}_{T+1}$$

- $\hat{\boldsymbol{s}}^{l}_{T+1}$ é o ganho informacional da divulgação adiantada de e.
- Projeções para o horizonte longo são feitas de forma recursiva, usando $\hat{\boldsymbol{X}}^{l}_{T+1}$ e $\boldsymbol{X}^{e}_{T+1}$ para $T+1$.

Conceitos de interrelação para processos multivariados

Distribuições conjuntas de processos vetoriais

- Seja $\{\boldsymbol{Z}_t\}_{t \in \mathbb{Z}}$ um processo vetorial de interesse.
- A densidade conjunta numa janela $\{\boldsymbol{Z}_t\}_{t=1}^{T}$, condicional a um valor inicial $\boldsymbol{Z}_0$, fatora-se como:

$$f[(\boldsymbol{Z}_t)_{t=1}^{T} | Z_0] = \prod_{t=1}^{T} f[\boldsymbol{Z}_t | \mathscr{Z}_{t-1}]\,,$$

onde $\mathscr{Z}_{t-1} = (\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{t-1})$.

Causalidade de Granger

- Particione $\boldsymbol{Z}_t$ em dois conjuntos de variáveis, $\boldsymbol{Z}_t^1$ e $\boldsymbol{Z}_t^2$, com $d_1 + d_2 = d$.
- Dizemos que $\boldsymbol{Z}_t^2$ não Granger causa $\boldsymbol{Z}_t^1$ se:

$$f(\boldsymbol{Z}_t^1 | \mathcal{Z}_{t-1}) = f(\boldsymbol{Z}_t^1 | \mathcal{Z}_{t-1}^1)\,,$$

  onde $\mathcal{Z}_{t-1}^1 = (\boldsymbol{Z}_0^1, \boldsymbol{Z}_1^1, \ldots, \boldsymbol{Z}_{t-1}^1)$.
- $\boldsymbol{Z}_t^2$ não Granger causa $\boldsymbol{Z}_t^1$ se defasagens de $\boldsymbol{Z}_t^2$ não são informativas sobre $\boldsymbol{Z}_t^1$.
  - Conceito preditivo, e não causal (no sentido econométrico do termo), embora o termo causalidade esteja no nome.
- Em um modelo VAR(p) para $\boldsymbol{Z}_t$, onde $\boldsymbol{A}_l = \begin{bmatrix} \boldsymbol{A}_{l,1,1} & \boldsymbol{A}_{l,1,2} \\ \boldsymbol{A}_{l,2,1} & \boldsymbol{A}_{l,2,2} \end{bmatrix}$, nula de que $\boldsymbol{Z}_t^2$ não Granger causa $\boldsymbol{Z}_t^1$ é equivalente a:

$$H_0 : \boldsymbol{A}_{l,1,2} = \boldsymbol{0}_{d_1 \times d_2}\,, \quad l = 1, \ldots p\,.$$

Causalidade de Granger, causalidade contemporânea e exogeneidade Fraca

- Conceito de causalidade de Granger nos diz que conhecimento do passado de $\boldsymbol{Z}_t^2$ não é informativo sobre $\boldsymbol{Z}_t^1$ contemporâneo.
- Note que não falamos nada sobre conhecimento **contemporâneo** de $\boldsymbol{Z}_t^2$ ser informativo sobre $\boldsymbol{Z}_t^1$ contemporâneo.
  - De fato, num VAR, $\boldsymbol{Z}_t^2$ é no geral contemporanemanete informativo sobre $\boldsymbol{Z}_t^1$ por conta das correlações em $\mathbb{V}(\boldsymbol{\epsilon}_t)$ diferentes de zero.
  - No VAR, é possível testar essas relações contemporâneas olhando-se para as entradas de $\mathbb{V}(\boldsymbol{\epsilon}_t)$ ("causalidade" contemporânea; também um conceito preditivo).
    - Nula de não causalidade contemporânea é que coeficiente do melhor preditor linear de $\boldsymbol{\epsilon}_t^1$ em $\boldsymbol{\epsilon}_t^2$ é zero, i.e. $\text{cov}(\boldsymbol{\epsilon}_t^1, \boldsymbol{\epsilon}_t^2)\mathbb{V}[\boldsymbol{\epsilon}_t^2]^{-1} = \mathbf{0}_{d_1 \times d_2}$.
- Um conceito relacionado, embora diferente da causalidade de Granger, diz respeito à modelagem da distribuição de $\boldsymbol{Z}_t^2$ ser importante, do ponto de vista estatístico, na estimação do modelo para $\boldsymbol{Z}_t^1$.
- A esse conceito damos o nome de exogeneidade fraca.

Modelo condicional e exogeneidade fraca

- Formalmente, um modelo é uma parametrização da distribuição condicional $f[\boldsymbol{Z}_t|\mathcal{Z}_{t-1}] = f[\boldsymbol{Z}_t|\mathcal{Z}_{t-1};\theta_0]$, onde $\theta_0 \in \Theta$ é um parâmetro desconhecido, que toma algum valor em Θ.
- Note que sempre podemos escrever, usando a regra de Bayes:

$$f[(\boldsymbol{Z}_t)_{t=1}^{T}|Z_0;\theta_0] = \prod_{t=1}^{T} f[\boldsymbol{Z}_t|\mathcal{Z}_{t-1};\theta_0] = \prod_{t=1}^{T} f[\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1};\theta_0] f[\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1};\theta_0]\,,$$

- Nesse contexto, um subconjunto ψ dos parâmetros θ_0 é dito fracamente exógeno com respeito a $\boldsymbol{Z}_t^2$ se podemos reescrever:

$$f[\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1};\theta_0] f[\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1};\theta_0] = f[\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1};\lambda_1] f[\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1};\lambda_2]$$

  onde ψ é função somente de λ_1, e λ_1 e λ_2 podem variar livremente.
  - No caso, não há perda de informação relevante em estimar ψ somente considerando o modelo $\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1}$, tratando $\boldsymbol{Z}_t^2$ como fixo.

Exogeneidade forte

- O conceito de exogeneidade fraca nos diz que, para estimar ψ, não há perda de informação em considerar o modelo para $\boldsymbol{Z}_t^1 | \boldsymbol{Z}_t^2, \mathcal{Z}_{t-1}$.
- Se formos realizar previsões para $\boldsymbol{Z}_{T+h}^1$ com base nesse modelo, no entanto, vamos precisar de projeções para $\boldsymbol{Z}_{T+h}^2$.
- Podemos fazer isso imputando $\boldsymbol{Z}_{T+h}^2$ com base nas projeções de um modelo para $\boldsymbol{Z}_{T+h}^2$ com base somente no seu passado?
  - Isso só faz sentido em termos estatístico-preditivos se $\boldsymbol{Z}_{T+h}^1$ **não Granger causa** $\boldsymbol{Z}_{T+h}^2$.
  - Nesse caso, podemos modelar $\boldsymbol{Z}_{T+h}^2$ externamente e somente incluir suas projeções futuras no modelo para $\boldsymbol{Z}_t^1 | \boldsymbol{Z}_t^2, \mathcal{Z}_{t-1}$.
- Exogeneidade fraca de $\boldsymbol{Z}_t^2$ com respeito a ψ + não causalidade de Granger $\boldsymbol{Z}_t^1$ sobre $\boldsymbol{Z}_t^2$: $\boldsymbol{Z}_t^2$ é fortemente exógeno com respeito a ψ.

Análise contrafactual e super exogeneidade

- Retorne à fatoração da densidade conjunta:

$$f[\boldsymbol{Z}_t|\mathcal{Z}_{t-1};\theta_0] = f[\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1};\lambda_1]f[\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1};\lambda_2]$$

- Essa fatoração, nos sugere um método para avaliar intervenções alternativas em $\boldsymbol{Z}_2^t$, que aqui são medidas por mudanças na distribuição $\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1}$.
  - Se $\boldsymbol{Z}_t^2$ é a taxa de juros e $\boldsymbol{Z}_t^1$ compila inflação e taxa de desemprego, qual seria o comportamento da taxa de juros sobre uma regra de decisão alternativa?
- Formalmente, a avaliação de políticas alternativas (contrafactuais) com base na estimação de $\hat{\lambda}_1$ só será válida sobre um requerimento adicional à exogeneidade fraca, denominado **invariância**.
  - $\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1}$ é invariante se modelo e seu parâmetro para $f[\boldsymbol{Z}_t^1|\boldsymbol{Z}_t^2,\mathcal{Z}_{t-1};\lambda_1]$ não mudam sob distribuições alternativas para $\boldsymbol{Z}_t^2|\mathcal{Z}_{t-1}$.
- Exogeneidade fraca + invariância: superexogeneidade com respeito a um parâmetro.

Invariância e crítica de Lucas

- Num VAR(p) de inflação, desemprego e taxa de juros, a hipótese de invariância para avaliação do efeito contrafactual de regra de juros alternativas parece razoável?
- A crítica clássica de Lucas sugere que, na medida em que as decisões de produção e preço de uma economia são tomadas por agentes racionais que levam em conta expectativas acerca do regime em questão, a resposta é **não**.
- Nesse caso, os parâmetros do VAR(p) em questão são funções complicadas dos parâmetros fundamentais da Economia (preferências, parâmetros da função de produção e política econômica), de modo que existe uma relação complicada entre os parâmetros de equações diferentes de um VAR(p).
  - Não podemos simplesmente alterar uma das equações do modelo, mantendo as demais fixas nos valores estimados.
  - Além disso, note que as equações do modelo, a princípio, também não são necessariamente funções política passíveis de modificação, e sim objetos de equilíbrio.

Alternativas à crítica de Lucas

- A crítica de Lucas ensejou duas tradições na macroeconometria.
- A tradição de DSGE busca partir de modelos econômicos estocásticos microfundamentados e completamente especificados, e derivar um VAR(p) a partir das primitivas do modelo econômico e de hipóteses de equilíbrio do sistema.
- Formalmente, o VAR(p) resultante se escreve nesses casos como

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

  onde ξ são os parâmetros fundamentais (preferências, tecnologia, política fiscal e monetária etc.), e tanto os choques como $\boldsymbol{A}_l$ são funções complicadas desses parâmetros que emergem do equilíbrio do sistema.
- Ideia é estimar ξ indiretamente a partir do VAR(p), e realizar os contrafactuais variando-se ξ.
  - Dificuldades de identificação, estimação e inferência.
  - Além disso, método depende muito da correta especificação do modelo.

VAR (semi)estrutural

- Uma segunda tradição busca postular modelos que, partindo da teoria econômica, explicitem as relações causais entre as variáveis, embora sejam agnósticos sobre os microfundamentos específicos dos problemas.
  - Diversos modelos econômicos postulam relações contemporâneas e extemporâneas entre desemprego, juros e inflação.
- Definido o modelo, analisamos condições que garantam que sejamos capazes de avaliar o efeito de choques (surpresas) nas variáveis de política econômica sobre as demais variáveis, **no regime em questão**.
  - Análise de natureza causal, embora não contrafactual (visto que mantemos o regime atual).
  - Para análise contrafactual, precisamos de mais hipóteses, visto que não postulamos todos os microfundamentos como na abordagem DSGE.
- Abordagem de VAR (semi)estrutural, que veremos mais à frente no curso.