

EAE1223: ECONOMETRIA III

AULA 6 - MODELOS VETORIAIS AUTORREGRESSIVOS

Luis A. F. Alvarez

10 de junho de 2025

VETORES ALEATÓRIOS E MATRIZ DE VARIÂNCIA-COVARIÂNCIA

- Um vetor (coluna) aleatório \mathbf{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 \mathbf{X} é um vetor cujas entradas $\mathbf{X}_j, j = 1 \dots, d$, são variáveis aleatórias com valores reais.
- Para um vetor aleatório \mathbf{X} , a matriz de variância-covariância, $\mathbb{V}[\mathbf{X}]$, é definida como:

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

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

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

MATRIZ DE COVARIÂNCIA ENTRE DOIS VETORES ALEATÓRIOS

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

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

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

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

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 $\{\mathbf{X}_t : t \in \mathcal{I}\}$, onde cada \mathbf{X}_t é vetor aleatório em \mathbb{R}^d .
 - Cada entrada $j = 1 \dots d$ define um processo estocástico com valores reais $\{\mathbf{X}_{j,t} : t \in \mathcal{I}\}$ descrevendo a evolução da j -ésima entrada ao longo de \mathcal{I} .
- Um processo estocástico vetorial $\{\mathbf{X}_t : t \in \mathcal{T}\}$ indexado no tempo \mathcal{T} é dito fracamente estacionário se:
 1. $\mathbb{E}[\mathbf{X}_t] = \boldsymbol{\mu}$ para todo $t \in \mathcal{T}$.
 2. $\mathbb{V}[\mathbf{X}_t] = \boldsymbol{\Sigma}_0$ para todo $t \in \mathcal{T}$, com $\text{tr}(\boldsymbol{\Sigma}_0) < \infty$.
 3. $\text{cov}(\mathbf{X}_t, \mathbf{X}_{t-h}) = \boldsymbol{\Sigma}_h$, 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 $\{\mathbf{X}_{j,t} : t \in \mathcal{T}\}$, $j = 1, \dots, d$, é fracamente estacionária.

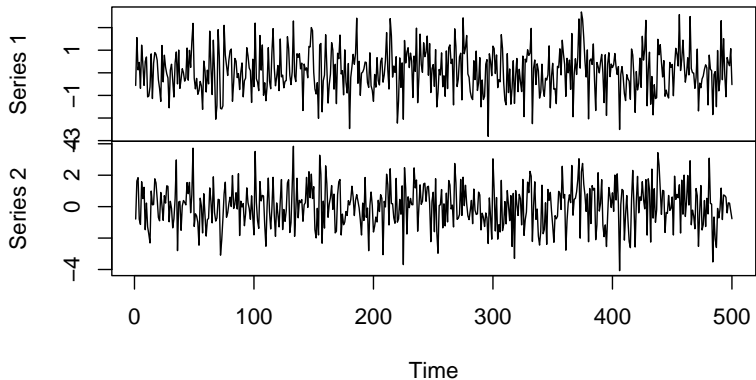
RUÍDO BRANCO VETORIAL

- Um processo estocástico vetorial $\{\mathbf{X}_t : t \in \mathcal{T}\}$ com valores em \mathbb{R}^d é dito um ruído branco vetorial se:
 - $\mathbb{E}[\mathbf{X}_t] = \mathbf{0}_{d \times 1}$, para todo $t \in \mathcal{T}$.
 - $\mathbb{V}[\mathbf{X}_t] = \Sigma_0$ para todo $t \in \mathcal{T}$, com $\text{tr}(\Sigma_0) < \infty$.
 - $\text{cov}(\mathbf{X}_t, \mathbf{X}_l) = \mathbf{0}_{d \times d}$, 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

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

rb



MODELOS VETORIAIS AUTORREGRESSIVOS

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

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

onde $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{d,t})'$ 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 $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{d,t})'$, podemos escrever o sistema como

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

onde

$$\mathbf{a}_0 = \begin{bmatrix} a_{1,0} \\ a_{2,0} \\ \vdots \\ a_{d,0} \end{bmatrix} \quad \mathbf{A}_l = \begin{bmatrix} a_{1,1,l} & a_{1,2,l} & \dots & a_{1,d,l} \\ a_{2,1,l} & a_{2,2,l} & \dots & a_{2,d,l} \\ \vdots & \vdots & \dots & \vdots \\ a_{d,1,l} & a_{d,2,l} & \dots & a_{d,d,l} \end{bmatrix}$$

VAR(p) ESTACIONÁRIO

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

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

- Todas as raízes do polinômio $\phi(z) = \det(\mathbb{I}_{d \times d} - \mathbf{A}_1 z - \mathbf{A}_2 z^2 \dots - \mathbf{A}_p z^p)$ 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, $\{\mathbf{X}_{j,t}\}_{t=1}^T$, $j = 1, \dots, d$, 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:

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

- Para cada equação, rodamos uma regressão com $T - p$ 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:

$$\min_{b_0, B_1, \dots, B_p} \sum_{t=p+1}^T \left(\mathbf{X}_t - \mathbf{b}_0 - \sum_{l=1}^p \mathbf{B}_l \mathbf{X}_{t-l} \right)' \hat{\Sigma}_0^{-1} \left(\mathbf{X}_t - \mathbf{b}_0 - \sum_{l=1}^p \mathbf{B}_l \mathbf{X}_{t-l} \right)$$

MÁXIMA VEROSSIMILHANÇA CONDICIONAL

- Sob a hipótese auxiliar:

$$\epsilon_t \stackrel{iid}{\sim} N(\mathbf{0}_{d \times 1}, \Sigma_0)$$

podemos calcular a verossimilhança de $\mathbf{X}_{p+1}, \dots, \mathbf{X}_T$, condicional a $\mathbf{X}_1, \dots, \mathbf{X}_p$.

- Estimador de máxima verossimilhança condicional estima **simultaneamente** \mathbf{a}_0 , os \mathbf{A}_l e Σ_0 .

RELAÇÃO ENTRE OS MÉTODOS DE ESTIMAÇÃO

- Se o espaço de parâmetros é irrestrito, no sentido de que os parâmetros \mathbf{a}_0 e os \mathbf{A}_I 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 \mathbf{a}_0 e os \mathbf{A}_I , 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:

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

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

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

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 $T - p_{max}$ 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.

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

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

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

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:

$$H_0 : \text{cov}(\epsilon_t, \epsilon_{t-j}) = \mathbf{0}_{d \times d}, \quad j = 1, \dots, h.$$

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:

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

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

$$\mathbf{X}_t = \mathbf{a}_0 + \gamma \mathbf{d}_t + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2} + \dots + \mathbf{A}_p \mathbf{X}_{t-p} + \epsilon_t,$$

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 \mathbf{X}_{T+1} é dada por:

$$\hat{\mathbf{X}}_{T+1} = \hat{\mathbf{a}}_0 + \hat{\mathbf{A}}_1 \mathbf{X}_T + \hat{\mathbf{A}}_2 \mathbf{X}_{T-1} + \dots + \hat{\mathbf{A}}_p \mathbf{X}_{T+1-p}$$

- A previsão dois passos à frente é dada por:

$$\hat{\mathbf{X}}_{T+2} = \hat{\mathbf{a}}_0 + \hat{\mathbf{A}}_1 \hat{\mathbf{X}}_{T+1} + \hat{\mathbf{A}}_2 \mathbf{X}_T + \dots + \hat{\mathbf{A}}_p \mathbf{X}_{T+2-p},$$

e assim recursivamente.

- Previsões do VAR estacionário convergem para média incondicional estimada do processo:

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

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 \mathbf{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 $\mathbf{X}_t = [\mathbf{X}_t^e, \mathbf{X}_t^I]$.
 - Para as variáveis em \mathbf{X}_t^e , observamos os dados até $T + 1$.
 - Para as variáveis em \mathbf{X}_t^I , observamos os dados até T .
- Sejam $\hat{\mathbf{a}}_0$, $\{\hat{\mathbf{A}}_k\}_{k=1}^p$ 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 \mathbf{X}_t^I nossa melhor projeção para ϵ_{T+1}^I .
 - 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 \mathbf{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 ϵ_{T+1}^l com base no conhecimento do passado e \mathbf{X}_{T+1}^e é:

$$\mathbf{s}_{T+1}^l = \Sigma_{l,e} \Sigma_{e,e}^{-1} \epsilon_{T+1}^e$$

onde $\Sigma_{e,e} = \mathbb{V}[\epsilon_t^e]$ e $\Sigma_{l,e} = \text{cov}[\epsilon_t^l, \epsilon_t^e]$.

- Ideia é estimar \mathbf{s}_{T+1}^l como:

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

onde $\hat{\mathbf{X}}_{T+1|T}^e$ é a projeção usual do VAR com dados até T , para $T+1$.

- Projetamos \mathbf{X}_{T+1}^l como:

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

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

Conceitos de interrelação para processos multivariados

DISTRIBUIÇÕES CONJUNTAS DE PROCESSOS VETORIAIS

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

$$f[(\mathbf{Z}_t)_{t=1}^T | \mathbf{Z}_0] = \prod_{t=1}^T f[\mathbf{Z}_t | \mathcal{Z}_{t-1}],$$

onde $\mathcal{Z}_{t-1} = (\mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1})$.

CAUSALIDADE DE GRANGER

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

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

onde $\mathcal{Z}_{t-1}^1 = (\mathbf{Z}_0^1, \mathbf{Z}_1^1, \dots, \mathbf{Z}_{t-1}^1)$.

- \mathbf{Z}_t^2 não Granger causa \mathbf{Z}_t^1 se defasagens de \mathbf{Z}_t^2 não são informativas sobre \mathbf{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 \mathbf{Z}_t , onde $\mathbf{A}_l = \begin{bmatrix} \mathbf{A}_{l,1,1} & \mathbf{A}_{l,1,2} \\ \mathbf{A}_{l,2,1} & \mathbf{A}_{l,2,2} \end{bmatrix}$, nula de que \mathbf{Z}_t^2 não Granger causa \mathbf{Z}_t^1 é equivalente a:

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

CAUSALIDADE DE GRANGER, CAUSALIDADE CONTEMPORÂNEA E EXOGENEIDADE FRACA

- Conceito de causalidade de Granger nos diz que conhecimento do passado de \mathbf{Z}_t^2 não é informativo sobre \mathbf{Z}_t^1 contemporâneo.
- Note que não falamos nada sobre conhecimento **contemporâneo** de \mathbf{Z}_t^2 ser informativo sobre \mathbf{Z}_t^1 contemporâneo.
 - De fato, num VAR, \mathbf{Z}_t^2 é no geral contemporaneamente informativo sobre \mathbf{Z}_t^1 por conta das correlações em $\mathbb{V}(\epsilon_t)$ diferentes de zero.
 - No VAR, é possível testar essas relações contemporâneas olhando-se para as entradas de $\mathbb{V}(\epsilon_t)$ (“**causalidade**” **contemporânea**; também um conceito preditivo).
 - Nula de não causalidade contemporânea é que coeficiente do melhor preditor linear de ϵ_t^1 em ϵ_t^2 é zero, i.e. $\text{cov}(\epsilon_t^1, \epsilon_t^2) \mathbb{V}[\epsilon_t^2]^{-1} = \mathbf{0}_{d_1 \times d_2}$.
- Um conceito relacionado, embora diferente da causalidade de Granger, diz respeito à modelagem da distribuição de \mathbf{Z}_t^2 ser importante, do ponto de vista estatístico, na estimação do modelo para \mathbf{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 $f[\mathbf{Z}_t|\mathcal{Z}_{t-1}] = f[\mathbf{Z}_t|\mathcal{Z}_{t-1}; \theta_0]$, 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:

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

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

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

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 $\mathbf{Z}_t^1|\mathbf{Z}_t^2, \mathcal{Z}_{t-1}$, tratando \mathbf{Z}_t^2 como fixo.

EXOGENEIDADE FRACA NO VAR(p)

- Se \mathbf{Z}_t segue um VAR(p) e nosso interesse consiste em estudar a evolução de \mathbf{Z}_t^1 no tempo, exogeneidade fraca implica que não há perda de eficiência estatística em estimar o modelo que trata \mathbf{Z}_t^2 como uma **variável externa** no modelo:

$$\mathbf{Z}_t^1 = \mathbf{b}_0 + \sum_{j=1}^p \mathbf{B}_j \mathbf{Z}_{t-j}^1 + \sum_{l=0}^p \mathbf{C}_l \mathbf{Z}_{t-l}^2 + \mathbf{v}_t^1$$

- Se os coeficientes do modelo VAR(p) original para \mathbf{Z}_t são irrestritos, então um subconjunto de variáveis \mathbf{Z}_t^2 sempre será fracamente exógeno.
- Note, entretanto, que se formos usar a equação acima para previsão, precisaríamos ter projeções para trajetória de \mathbf{Z}_t^2 fora da amostra.
 - Será que poderíamos usar projeções externas para \mathbf{Z}_t^2 , por exemplo aquelas obtidas com base na estimação de um modelo com dados históricos de \mathbf{Z}_t^2 ou mesmo aquelas divulgadas por um órgão externo, de maneira estatisticamente consistente?
 - Esse é o conceito de **exogeneidade forte** que estudaremos agora.

EXOGENEIDADE FORTE

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

ANÁLISE CONTRAFACTUAL E SUPER EXOGENEIDADE

- Retorne à fatoração da densidade conjunta:

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

- Essa fatoração, nos sugere um método para avaliar intervenções alternativas em \mathbf{Z}_t^2 , que aqui são medidas por mudanças na distribuição $\mathbf{Z}_t^2 | \mathcal{Z}_{t-1}$.
 - Se \mathbf{Z}_t^2 é a taxa de juros e \mathbf{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**.
 - $\mathbf{Z}_t^1 | \mathbf{Z}_t^2, \mathcal{Z}_{t-1}$ é invariante se modelo e seu parâmetro para $f[\mathbf{Z}_t^1 | \mathbf{Z}_t^2, \mathcal{Z}_{t-1}; \lambda_1]$ não mudam sob distribuições alternativas para $\mathbf{Z}_t^2 | \mathcal{Z}_{t-1}$.
- 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 alternativos 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

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

onde ξ são os parâmetros fundamentais (preferências, tecnologia, política fiscal e monetária etc.), e tanto os choques como \mathbf{A}_i 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.