

Notas de Aula: Análise de Séries Temporais

André Filipe Bandão Santos

8 de julho de 2025

Resumo

Este resumo está sendo elaborado por André Santos como parte dos estudos para a disciplina. Embora estejam sendo feitos esforços para garantir a precisão, qualquer erro, inconsistência ou omissão neste resumo é de minha exclusiva responsabilidade. Este material é para fins de estudo e não substitui o material didático oficial da disciplina ou a orientação do professor.

SUMÁRIO

1	Módulo 1: Fundamentos da Análise de Séries Temporais	3
1.1	Curso de Séries Temporais e Ferramentas	3
1.2	Conceitos Fundamentais de Séries Temporais	3
1.3	Previsão, Robótica e Inteligência Artificial	4
1.4	Modelos de Previsão e Séries Cíclicas	4
1.5	Modelos para Séries de Tempo: AR vs. MA	4
1.6	Identificação de Modelos e Estacionariedade	4
1.7	Comparação entre Processos AR e MA	5
2	Módulo 2: Avaliação e Comparação de Modelos de Previsão	5
2.1	Fundamentos da Teoria da Decisão e Utilidade	5
2.2	O Papel da Previsão na Tomada de Decisão	5
2.3	Formalização da Teoria da Decisão	5
2.4	Perda de Utilidade e Erro de Previsão	6
2.5	Tipos de Previsão	6
2.6	Avaliação de Modelos e Geração de Erros	6
2.7	Métricas Estatísticas para Comparação de Modelos	6
2.8	Valor Econômico da Previsão	7
3	Módulo 3: Modelos Multivariados e Relações de Longo Prazo	7
3.1	Revisão e Introdução ao Teste de Hipóteses para Comparação de Modelos	7
3.2	Teste de Hipótese para Comparar a Capacidade Preditiva de Dois Modelos	8
3.3	Introdução à Previsão Quantílica	8
4	Módulo 4: Tópicos Avançados e Fronteiras com Machine Learning	9
4.1	Limitações da Previsão pela Média e Introdução à Previsão Quantílica .	9
4.2	Value at Risk (VaR)	9
4.3	CoVaR (Conditional Value at Risk) – Medida de Contágio	10

4.4	Stress Test – Medida de Resiliência da Firma	10
4.5	Previsão de Densidade (Density Forecasting)	10
4.6	Previsão com Big Data e Introdução ao Machine Learning	10
5	Módulo 5: Inferência Causal com Séries Temporais	11
5.1	Motivação e Contexto	11
5.2	Metodologia Proposta (3 Passos)	12
5.3	Análise Empírica: Previsão da Taxa de Câmbio US-Canadá	12
5.4	Conclusões e Pesquisa Futura	13
6	Avaliação de Modelos em R e Introdução ao Controle Sintético	13
6.1	Oficina de Código em R: Geração e Avaliação de Previsões	13
6.2	Introdução ao Método do Controle Sintético	14
7	aula 7 - Aprofundamento em Controle Sintético e a Extensão Quantílica	15
7.1	Aprofundamento Teórico: Método e Extensões	15
7.2	Estudo de Caso: O Impacto do Choque Migratório em Roraima	16
8	aula 9 - Seleção de Modelos, Causalidade e Introdução aos Modelos VAR	17
8.1	CrITÉrios de Seleção de Modelos e Suas Limitações	17
8.2	Introdução à Causalidade em Macroeconomia	17
8.3	O Modelo de Vetores Autorregressivos (VAR)	18
9	aula 10 - Fundamentos de Vetores Autorregressivos (VAR) e o Problema de Identificação	19
9.1	Revisão: As Quatro Representações do VAR e a Função de Impulso-Resposta (FIR)	19
9.2	Identificação do VAR Estrutural: Restrições de Curto Prazo (Decomposição de Cholesky)	20
9.3	Identificação do VAR Estrutural: Restrições de Longo Prazo (Blanchard-Quah)	21
9.4	Causalidade de Granger: Testando a Precedência Temporal	21
10	aula 11 - Identificação de Modelos VAR e Introdução a Séries Não-Estacionárias	22
10.1	Identificação de VAR com Restrições de Longo Prazo: A Abordagem de Blanchard-Quah	22
10.2	Testes de Raiz Unitária: Dickey-Fuller e ADF	23
10.3	Cointegração e Modelos de Correção de Erros (VECM)	23
11	aula 12 - Cointegração: Testes e Modelos de Correção de Erros (VECM)	24
11.1	Regressão Espúria e a Definição de Cointegração	24
11.2	A Intuição da Cointegração: Tendências Estocásticas Comuns	25
11.3	Testes de Cointegração	25
11.3.1	O Teste de Engle-Granger (Baseado nos Resíduos)	26
11.3.2	O Teste de Johansen (Baseado no Sistema)	26
11.4	O Modelo de Correção de Erros (VECM) e a Previsão	26

1. MÓDULO 1: FUNDAMENTOS DA ANÁLISE DE SÉRIES TEMPORAIS

1.1. Curso de Séries Temporais e Ferramentas

- O curso será sobre séries temporais, começando do básico e progredindo para tópicos mais avançados, incluindo machine learning.
- Será explorada a possibilidade de aprender econometria com o ChatGPT, destacando que a parte tecnológica do ChatGPT é muito boa para gerar códigos. A limitação apontada é que a melhor comunicação com o ChatGPT é em inglês, pois a qualidade cai com o português.
- O curso terá uma parte teórica que cobrirá vários tópicos de forma aprofundada, com ênfase no domínio da notação teórica para uma comunicação eficaz.
- Haverá uma parte prática com códigos em R, utilizando muitos "loops" para incluir dinâmica na estimação dos modelos.
- Os livros de referência para a parte teórica serão: Walter Enders (*Applied Econometric Time Series*) e James Hamilton (*Time Series Analysis*).

1.2. Conceitos Fundamentais de Séries Temporais

- **Diferença entre *cross-section* e séries temporais:**
 - Em dados *cross-section*, as observações Y_i para $i = 1, \dots, N$ são I.I.D. (independentes e identicamente distribuídas). Isso significa que Y_i e Y_j são independentes para $i \neq j$, o que implica $\text{Cov}(Y_i, Y_j) = 0$. Adicionalmente, todas as observações Y_i vêm da mesma distribuição populacional, ou seja, possuem a mesma média $E[Y_i] = \mu$ e variância $\text{Var}(Y_i) = \sigma^2$ para todo i .
- **A hipótese de I.I.D.** é crucial para provar propriedades desejáveis de estimadores, como o estimador de média amostral $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N Y_i$. Esta hipótese garante que $\hat{\mu}$ é não-enviesado ($E[\hat{\mu}] = \mu$) e consistente ($\hat{\mu} \xrightarrow{p} \mu$ quando $N \rightarrow \infty$).
- **Séries temporais e memória:** Ao contrário dos dados I.I.D., as séries temporais $\{Y_t\}_{t=1}^T$ não possuem a propriedade de "mostra aleatória", tornando a análise mais complexa, mas também criando oportunidades.
- A principal característica das séries temporais é a **memória temporal**, o que significa que observações passadas Y_{t-k} influenciam as futuras Y_t . Essa memória é "importantíssimo para previsão", pois permite extrapolar para o futuro.
- **Processo de Ruído Branco (ϵ_t):** É definido como um processo I.I.D. e "sem memória". Representa o "building block" (parte básica) de qualquer modelo de série temporal. Retornos de ações em baixa frequência são citados como um exemplo de processo de ruído branco.
- **Operador de Defasagem (L):** Introduzido como um operador que defasa a série temporal por um período. Para uma série Y_t , o operador de defasagem L age como

$LY_t = Y_{t-1}$. É uma ferramenta útil para entender a relação entre Y_t e Y_{t-1} em modelos autorregressivos.

1.3. Previsão, Robótica e Inteligência Artificial

- A memória temporal é a base para a família de modelos ARIMA, que utiliza essa memória para identificar o modelo que melhor representa os dados.
- A capacidade de memória é fundamental para o funcionamento de robôs e modelos de inteligência artificial como o ChatGPT; sem memória, eles não conseguiriam aprender ou fazer previsões.
- Um robô é descrito como um modelo de série temporal que faz previsões e toma decisões baseadas em teoria econômica. A qualidade da previsão é crucial para o sucesso de um robô.

1.4. Modelos de Previsão e Séries Cíclicas

- A família de modelos ARIMA é um *benchmark* em análise de previsão; qualquer novo modelo proposto deve ter um desempenho superior a eles em termos de precisão.
- A técnica Box-Jenkins é usada para determinar o modelo ARIMA mais adequado.
- Algumas séries de tempo apresentam um comportamento cíclico ou sazonal, enquanto outras podem se comportar como um ruído branco.

1.5. Modelos para Séries de Tempo: AR vs. MA

- **Modelos Autorregressivos (AR):** Usados para capturar componentes cíclicos ou séries com "memória".
- **Modelos de Média Móvel (MA):** Usados para processos que são "puro ruído" mas que ainda possuem memória. São definidos como funções de ruídos brancos passados.

1.6. Identificação de Modelos e Estacionariedade

- **Autocorrelação (ACF):** Mede a correlação de uma série temporal com suas versões passadas (lags). Para processos AR, a autocorrelação **decai lentamente**.
- **Autocorrelação Parcial (PACF):** Mede a correlação entre Y_t e Y_{t-j} após remover o efeito das variáveis intermediárias. Para um processo AR(p), a PACF terá um 'spike' (pico significativo) até o lag p e será zero depois.
- **Estacionariedade Fraca:** Uma série é considerada fracamente estacionária se:
 1. Sua esperança (média) é constante ao longo do tempo (μ).
 2. Sua variância é constante ao longo do tempo (σ^2).
 3. A covariância entre Y_t e Y_{t-j} depende apenas do intervalo j (lag), e não do tempo t .

- A condição de estacionariedade para um modelo AR(1) implica que o valor absoluto de seu coeficiente α deve ser menor que 1 ($|\alpha| < 1$).

1.7. Comparação entre Processos AR e MA

- **Processo AR:** Possui **memória longa**.
- **Processo MA:** Possui **memória curta**.

2. MÓDULO 2: AVALIAÇÃO E COMPARAÇÃO DE MODELOS DE PREVISÃO

2.1. Fundamentos da Teoria da Decisão e Utilidade

- **Introdução:** A aula faz uma introdução à previsão usando a teoria econômica, focando em previsão e teoria da decisão.
- **Base Econômica:** Em economia, uma decisão é avaliada com base na utilidade que ela gera para o indivíduo.
 - Exemplos: Comprar um sorvete ou fazer um doutorado são decisões tomadas porque geram utilidade positiva, mesmo que essa utilidade seja sentida no futuro.
- **Incerteza e Futuro:** A decisão de consumir ou comprar algo envolve incerteza sobre o que vai acontecer no futuro. O objetivo é maximizar a utilidade esperada.

2.2. O Papel da Previsão na Tomada de Decisão

- **Consulta à Previsão:** Antes de tomar uma decisão que depende de um evento futuro incerto, consulta-se uma previsão.
- **Avaliação da Decisão:** Só é possível saber se a decisão foi a melhor após o evento futuro ocorrer, comparando a decisão com o que de fato aconteceu.
- **Proximidade da Previsão à Realidade:** Quanto mais próxima a previsão estiver do que realmente acontecerá, melhores serão as decisões tomadas.

2.3. Formalização da Teoria da Decisão

- **Função de Utilidade:** Um agente possui uma função de utilidade U , que depende de duas variáveis: X e α .
 - **X:** Uma variável incerta no tempo t (hoje), que só será conhecida/observada no futuro $(t + h)$.
 - α : Uma ação ou decisão tomada no tempo t .
- **Escolha da Ação Baseada na Previsão:** O indivíduo escolhe a ação α no tempo T baseando-se numa previsão de X (denotada como X^F).

- X^F é o valor esperado de X em $t + h$, condicionado à informação disponível em t .

$$X^F = E[X_{t+h}|I_t]$$

- **Decisão Ótima:** A decisão ótima α^* é aquela que maximiza a utilidade esperada no tempo t , dada a previsão X^F .
- **Utilidade Ex-Post:** Em $t + h$, após X ser observado, calcula-se a utilidade realizada (ex-post), que é $U(X, \alpha)$.

2.4. Perda de Utilidade e Erro de Previsão

- **Perda de Utilidade (Utility Loss):** É a diferença entre a utilidade que seria obtida com uma previsão perfeita ($X^F = X$) e a utilidade obtida com a previsão X^F que foi de fato utilizada.

$$\text{Utility Loss} = U(X, \alpha(X)) - U(X, \alpha(X^F))$$

- **Decisões Erradas e Previsões Erradas:** A lição principal é que decisões erradas são baseadas em previsões erradas.
- **Erro de Previsão:** Definido como o valor realizado (X) menos o valor previsto (X^F).

2.5. Tipos de Previsão

- **Previsão Pontual:** Inicialmente, a previsão é definida como a média da variável em $t + h$, que representa o valor esperado condicional.
- **Previsão Quantílica (ou de Densidade):** Busca prever a distribuição de densidade completa, permitindo calcular probabilidades de eventos específicos e construir intervalos de confiança.

2.6. Avaliação de Modelos e Geração de Erros

- **Metodologia "Pseudo Out-of-Sample Forecasting":**
 - A série temporal é dividida em uma parte para estimação e outra para avaliação.
 - O modelo é estimado recursivamente: estima-se até t , prevê-se $t+1$; reestima-se até $t+1$, prevê-se $t+2$, e assim por diante.
 - Este processo gera um vetor de erros de previsão fora da amostra para comparar modelos.

2.7. Métricas Estatísticas para Comparação de Modelos

- **Erro Quadrático Médio (MSE):**

$$\text{MSE} = \frac{1}{P} \sum_{t=1}^P (\text{erro}_t)^2$$

O melhor modelo é aquele com o menor MSE.

- **Erro Absoluto Médio (MAE):**

$$\text{MAE} = \frac{1}{P} \sum_{t=1}^P |\text{erro}_t|$$

É mais robusto a outliers que o MSE.

- **R² Fora da Amostra (OOS R²):**

$$\text{OOS R}^2 = 1 - \frac{\sum_{t=1}^P (\text{erro}_{\text{modelo},i,t})^2}{\sum_{t=1}^P (\text{erro}_{\text{benchmark},t})^2}$$

Compara o modelo de interesse com um modelo de referência (benchmark).

2.8. Valor Econômico da Previsão

- Além da comparação estatística, os modelos podem ser avaliados com base no valor econômico que suas previsões geram.
- **Exercício de Riqueza Acumulada:** Simula-se uma estratégia de investimento baseada nas previsões de diferentes modelos para ver qual deles gera maior retorno financeiro.
- **Aplicação em Finanças (Portfólios Ativos):** Utiliza-se previsões para decidir quais ativos comprar ou vender, contrastando com estratégias passivas (que seguem um índice).

3. MÓDULO 3: MODELOS MULTIVARIADOS E RELAÇÕES DE LONGO PRAZO

3.1. Revisão e Introdução ao Teste de Hipóteses para Comparação de Modelos

- Na aula passada, foram construídas medidas de comparação de modelos baseadas no erro de previsão.
- O erro de previsão, $e_{t+h|t} = Y_{t+h} - \hat{Y}_{t+h|t}$ (observado menos previsto), é a variável principal para essa análise.
- Um vetor de erros de previsão fora da amostra é gerado recursivamente, e com ele constroem-se métricas como MSE e MAE.
- **Problema:** Se as métricas de dois modelos são próximas, não se pode afirmar conclusivamente qual é o melhor apenas pelo valor numérico.
- **Objetivo da Aula:** Aprender a realizar testes de hipóteses para verificar se dois modelos têm, estatisticamente, a mesma capacidade preditiva.

3.2. Teste de Hipótese para Comparar a Capacidade Preditiva de Dois Modelos

- Consideram-se dois modelos (1 e 2) que geram previsões $(\hat{Y}_{1t}, \hat{Y}_{2t})$ para a mesma variável Y_t .
- **Função Perda (Loss Function):** A acurácia é medida por uma função perda, $g(e_{it})$, que depende do erro de previsão.
 - **Função Perda Quadrática:** $g(e_{it}) = e_{it}^2$. A média desta função é o MSE. É simétrica.
 - **Função Perda de Valor Absoluto:** $g(e_{it}) = |e_{it}|$. A média desta função é o MAE. É simétrica e mais robusta a outliers.
 - **Função Perda Assimétrica (Check Function):** Penaliza erros positivos e negativos de forma diferente, com inclinações τ e $-(1 - \tau)$, onde $\tau \in (0, 1)$ é o quantil de interesse. Leva a previsões ótimas que são quantis, não a média.
- **Construção do Teste de Comparação (ex: Diebold-Mariano):**
 - Define-se uma série de diferenças das perdas: $D_t = g(e_{1t}) - g(e_{2t})$.
 - **Hipótese Nula (H_0):** Os dois modelos têm a mesma capacidade preditiva, ou seja, $E[D_t] = 0$.
 - **Hipótese Alternativa (H_1):** As capacidades preditivas são diferentes, $E[D_t] \neq 0$.
 - **Estatística de Teste:** Semelhante a um teste t para a média de D_t .

$$T_{\text{stat}} = \frac{\bar{D}}{\text{std_err}(\bar{D})}$$

- **Desafio em Séries Temporais:** A série D_t pode apresentar autocorrelação. O erro padrão de \bar{D} deve usar a **Variância de Longo Prazo (LRV)**, estimada com o estimador de Newey-West:

$$\text{LRV}(\bar{D}) = \frac{1}{P} \left(\hat{\gamma}_0 + 2 \sum_{k=1}^{M_P} w_k \hat{\gamma}_k \right)$$

Onde $\hat{\gamma}_k$ é a autocovariância amostral, w_k são os pesos de Bartlett e M_P é um parâmetro de truncamento.

- **Decisão do Teste:** Compara-se $|T_{\text{stat}}|$ com um valor crítico da normal padrão (ex: 1.96). Se for maior, rejeita-se H_0 .

3.3. Introdução à Previsão Quantílica

- **Nowcasting (H=0):** Previsão para o período corrente usando dados de alta frequência, antes da divulgação do dado oficial (ex: prever o PIB trimestral usando dados mensais de energia).
- **Previsão Ótima e Função Perda (Revisão):**

- Se a função perda é **quadrática**, a previsão ótima é a **esperança condicional**: $Y_{t+h|t}^* = E[Y_{t+h}|I_t]$.
- Se a função perda é a **assimétrica "check function"**, a previsão ótima é o **quantil condicional**: $Y_{t+h|t}^* = Q_\tau(Y_{t+h}|I_t)$.
- **Conceito de Quantil**: O τ -ésimo quantil c_τ de uma variável aleatória com FDA $F(\cdot)$ é o valor tal que $F(c_\tau) = \tau$, ou $c_\tau = F^{-1}(\tau)$.
- **Regressão Quantílica**:
 - **Caso Homoscedástico** ($\text{Var}(Y_i|X_i) = \sigma_\epsilon^2$):

$$Q_\tau(Y_i|X_i) = (\beta_0 + F_\epsilon^{-1}(\tau)) + \beta_1 X_i$$

As retas dos quantis são paralelas à reta da média. O efeito de X_i é constante em toda a distribuição de Y_i .

- **Caso Heteroscedástico** ($\text{Var}(Y_i|X_i)$ **não é constante**):

$$Q_\tau(Y_i|X_i) = \beta_0(\tau) + \beta_1(\tau)X_i$$

O efeito de X_i , dado por $\beta_1(\tau)$, pode variar para diferentes quantis τ . A regressão quantílica torna-se muito mais informativa, capturando a heterogeneidade dos dados.

4. MÓDULO 4: TÓPICOS AVANÇADOS E FRONTEIRAS COM MACHINE LEARNING

4.1. Limitações da Previsão pela Média e Introdução à Previsão Quantílica

- Quando há pouca incerteza (curva de densidade estreita), a média condicional $E[Y_{t+h}|I_t]$ é uma boa representação para a previsão.
- Em cenários de alta incerteza ou volatilidade, a média pode não ser suficiente, e o mercado passa a se interessar por previsões de eventos em outras partes da distribuição (quantis).

4.2. Value at Risk (VaR)

- O **Value at Risk (VaR)** representa a perda potencial máxima de um portfólio para um dado nível de probabilidade (ex: 5%) e um horizonte de tempo.
- **Importância Regulatória**: O Banco Central utiliza o VaR para regular as instituições financeiras.
- **VaR como Previsão Quantílica**: O VaR é uma previsão do τ -ésimo quantil da distribuição de retornos futuros. Para uma probabilidade de 5%, temos $\tau = 0.05$:

$$\text{VaR}_{t+h|t} = Q_{0.05}(Y_{t+h}|I_t)$$

- **Teste de Adequação do VaR (Engel & Manganelli):** Testa-se se o VaR reportado por um banco é adequado através de uma regressão quantílica:

$$Q_{0.05}(R_{t+h}|\text{VaR}_{t+h|t}) = \alpha_{0.05} + \beta_{0.05} \cdot \text{VaR}_{t+h|t}$$

A hipótese nula para um VaR bem especificado é $H_0 : \alpha_{0.05} = 0$ e $\beta_{0.05} = 1$.

4.3. CoVaR (Conditional Value at Risk) – Medida de Contágio

- O CoVaR avalia como o VaR de uma instituição/mercado é afetado quando outra está sob estresse.
- Estima-se uma regressão quantílica, por exemplo, do VaR do Ibovespa (R_M) condicionado ao retorno da Vale (R_i):

$$Q_{\tau}(R_{M,t}|R_{i,t}) = \alpha(\tau) + \beta(\tau)R_{i,t}$$

- Compara-se o VaR do Ibovespa em um cenário normal (ex: $R_{i,t} = \text{Mediana}(R_{i,t})$) com um cenário de estresse (ex: $R_{i,t} = Q_{0.05}(R_{i,t})$). A diferença mede o contágio.

4.4. Stress Test – Medida de Resiliência da Firma

- Avalia a vulnerabilidade de uma firma a uma crise sistêmica, invertendo a regressão anterior. A variável dependente passa a ser o retorno da firma (R_i) e a independente o retorno do mercado (R_M).
- Se o coeficiente $\hat{\beta}(0.05)$ for estatisticamente nulo, a firma é considerada resiliente.

4.5. Previsão de Densidade (Density Forecasting)

- Ao prever múltiplos quantis de uma distribuição futura, pode-se interpolá-los para obter uma aproximação da densidade completa prevista $f(Y_{t+h}|I_t)$.
- Com a densidade, pode-se calcular a probabilidade de qualquer evento, como a chance de uma recessão ou de uma depreciação cambial.

4.6. Previsão com Big Data e Introdução ao Machine Learning

- **Dilema Viés-Variância:** Machine learning busca otimizar o trade-off entre underfitting (alto viés, modelo simples) e overfitting (alta variância, modelo complexo).
- **Desafio com Big Data ($K > T$):** Quando o número de preditores (K) é maior que o de observações (T), MQO é inviável.
- **Elastic Net:** Uma técnica de regularização que minimiza a soma dos quadrados

dos resíduos adicionando duas penalidades aos coeficientes β_j :

$$\min_{\beta_0, \beta} \left(\sum_{t=1}^R (Y_{t+h} - (\beta_0 + \sum_{j=1}^K \beta_j X_{j,t}))^2 \right)$$

sujeito a:

$$\sum_{j=1}^K |\beta_j| \leq s_1 \quad (\text{Penalidade Lasso, L1})$$

$$\sum_{j=1}^K \beta_j^2 \leq s_2 \quad (\text{Penalidade Ridge, L2})$$

- **Elastic Net para Regressão Quantílica:** A função de perda quadrática é substituída pela "check function" (função perda quantílica), permitindo selecionar variáveis preditivas para diferentes quantis da distribuição de interesse.

5. MÓDULO 5: INFERÊNCIA CAUSAL COM SÉRIES TEMPORAIS

5.1. Motivação e Contexto

- **Limitações da Previsão Pontual:** Em cenários de alta incerteza e volatilidade, a previsão da média é insuficiente. O mercado tem interesse em prever a probabilidade de eventos extremos.
- **Previsão Quantílica em Finanças e Economia:** Várias aplicações usam a previsão quantílica para analisar riscos, como:
 - **Growth at Risk:** Prever a probabilidade de uma recessão.
 - **Gerenciamento de Risco:** Medir a perda potencial de um portfólio (Value at Risk).
 - **Comércio Internacional:** Analisar a volatilidade do comércio.
- **O Papel dos Dados Textuais:**
 - Dados textuais, como notícias, contêm informações não capturadas por indicadores tradicionais.
 - Muitos textos em jornais financeiros referem-se a riscos de cauda, justificando o uso de dados textuais para previsão quantílica.
- **Desafios dos Dados Textuais:**
 - **Alta Dimensionalidade:** O número de palavras (variáveis) é muito maior que o número de observações (tempo) - $N \gg T$.
 - **Construção de Índices:** Dicionários predefinidos podem ser subjetivos e instáveis.
 - **Esparsidade (Sparsity):** Nem todas as palavras têm poder preditivo.

- **Pergunta da Pesquisa:** Como construir um dicionário de palavras que seja relevante para prever especificamente os quantis (caudas da distribuição) de uma variável?

5.2. Metodologia Proposta (3 Passos)

A metodologia proposta visa criar um modelo de previsão quantílica usando dados textuais.

- **Passo 1: Pré-processamento dos Dados Textuais**
 - Artigos de jornais publicados em um período t são agregados.
 - Cria-se um vetor Ω_t onde cada elemento Ω_{it} representa a frequência da palavra i .
 - As contagens são normalizadas para criar variáveis Z_{it} com média zero e desvio padrão um.
- **Passo 2: Seleção de Palavras com Elastic Net Quantílico**
 - Para cada quantil τ de interesse, utiliza-se o **Elastic Net para Regressão Quantílica** para selecionar as palavras úteis para prever o τ -ésimo quantil da variável Y_{t+1} .
 - A função perda quadrática é substituída pela função perda quantílica (ρ_τ), com penalidades Lasso (L1) e Ridge (L2).
 - Este processo cria dicionários Ω_t^τ específicos para cada quantil.
- **Passo 3: Extração de Fatores e Previsão Quantílica**
 - As palavras selecionadas no passo 2 são muitas vezes correlacionadas.
 - Utiliza-se a **Análise de Componentes Principais (PCA)** para extrair fatores latentes $F_{t,\tau}$ desse conjunto de palavras.
 - A equação final de previsão quantílica usa este fator como preditor:

$$Q_\tau(Y_{t+1}|F_{t,\tau}) = \alpha(\tau) + \beta(\tau)F_{t,\tau}$$

5.3. Análise Empírica: Previsão da Taxa de Câmbio US-Canadá

- **Objetivo:** Testar se a metodologia consegue prever a variação da taxa de câmbio (Δs_{t+1}), que, segundo a Hipótese de Mercados Eficientes, segue um **Random Walk**.
- **Dados:** Dados textuais mensais do NYT e WSJ (1980-2022) e taxa de câmbio do FRED.
- **Resultados:**
 - **Previsão de Densidade e Intervalo:** O modelo textual produziu previsões significativamente melhores que o Random Walk, com intervalos de confiança que se adaptam à volatilidade.

- **Análise de Portfólio (Valor Econômico):** Uma estratégia de investimento baseada no modelo textual gerou uma riqueza acumulada muito superior às alternativas.
- **Dicionários Seleccionados:** Palavras ligadas a conflitos geopolíticos foram preditivas para o risco de **alta** do dólar (upside risk), enquanto palavras ligadas a recessões foram preditivas para o risco de **baixa** do dólar (downside risk).

5.4. Conclusões e Pesquisa Futura

- **Palavras importam:** Dados textuais contêm informação valiosa para prever os riscos de cauda de variáveis financeiras.
- O método proposto permite que os dicionários de palavras relevantes variem com o tempo e entre diferentes quantis.
- **Pesquisa Futura:** A metodologia pode ser estendida para prever todo o cross-section de retornos de ações, construindo portfólios baseados não apenas no retorno esperado, mas também no risco (volatilidade).

6. AVALIAÇÃO DE MODELOS EM R E INTRODUÇÃO AO CONTROLE SINTÉTICO

Esta aula aborda a implementação prática da avaliação de modelos de previsão em R e introduz o método do Controle Sintético para análise de impacto de políticas.

6.1. Oficina de Código em R: Geração e Avaliação de Previsões

Esta parte é dedicada a uma demonstração prática, mostrando o processo de geração de previsões fora da amostra de forma recursiva e a subsequente comparação de modelos.

Configuração do Exercício e Dados

- **Geração de Dados:** Para o exercício, os dados são gerados por computador a partir de um processo autorregressivo de ordem 2 (AR(2)).
- **Divisão da Amostra (In-Sample e Out-of-Sample):** A amostra total (tamanho T) é dividida em uma parte para estimação (R observações) e uma para avaliação (P observações), onde $T = R + P$.

O Loop Recursivo para Previsão

- A estimação recursiva é implementada com um **loop**. O processo utiliza uma **janela recursiva**, que aumenta a cada passo: estima-se com dados até t , prevê-se $t + 1$; reestima-se com dados até $t + 1$, prevê-se $t + 2$, e assim por diante.

Geração de Previsões e Avaliação

- **Modelos Comparados:** Vários modelos ARIMA e um modelo Naive (apenas intercepto) são estimados.

- **Função forecast:** Após estimar cada modelo, a função **forecast** é usada para gerar a previsão pontual, que corresponde à média condicional $E[Y_{t+h}|I_t]$.
- **Métricas de Avaliação:** Funções customizadas em R são criadas para calcular métricas como RMSE e MAE a partir dos erros de previsão. O OOS R^2 também é calculado para comparar os modelos a um benchmark:

$$\text{OOS } R^2 = 1 - \frac{\sum_{t=1}^P (\text{erro}_{\text{modelo } J, t})^2}{\sum_{t=1}^P (\text{erro}_{\text{benchmark}, t})^2}$$

- **Combinação de Previsões:** É apresentada como uma estratégia para diversificar e reduzir o risco, criando um novo "modelo combinado" a partir da média das previsões individuais.

6.2. Introdução ao Método do Controle Sintético

Este método é utilizado para avaliação de políticas públicas em cenários onde apenas uma unidade agregada (ex: um estado) sofre uma intervenção.

Motivação e Estrutura dos Dados

- **Problema Central:** Para avaliar o efeito de uma política, é necessário prever o ****contrafactual****: o que teria acontecido com a unidade tratada se a política não tivesse sido implementada.
- **O Efeito da Política:** É a diferença entre o valor observado e o contrafactual previsto.

$$\text{Efeito}_{\text{politica}} = Y_{\text{observado}} - Y_{\text{contrafactual_previsto}}$$

- **Estrutura de Dados:** O método requer dados em painel, com uma única **unidade tratada** e múltiplas **unidades de controle** ("doadoras"), observadas antes e depois da intervenção.

Formalização e Aplicações

- **Variáveis:** A análise utiliza uma variável de interesse (Y) e um conjunto de preditores (X) que ajudam a prever Y.
- **Exemplos de Aplicação:**
 - Avaliar o impacto de leis de porte de arma na Califórnia.
 - Analisar os efeitos do choque migratório venezuelano em Roraima.
 - Estimar qual seria o cenário socioeconômico de Cuba sem a Revolução de 1959.
 - Medir o impacto de controles de capital na Argentina.

7. AULA 7 - APROFUNDAMENTO EM CONTROLE SINTÉTICO E A EXTENSÃO QUANTÍLICA

Esta aula aprofunda o método do Controle Sintético, abordando suas limitações e extensões, e o aplica em um estudo de caso sobre o impacto de um choque migratório.

7.1. Aprofundamento Teórico: Método e Extensões

Revisão do Método do Controle Sintético

- **Cenário de Aplicação:** Ideal para situações com apenas uma unidade agregada "tratada" (ex: um estado) e múltiplas unidades de controle.
- **Objetivo Central:** Prever o **contrafactual** — o que teria acontecido com a unidade tratada se a política não tivesse sido implementada.
- **Dados Necessários:** Requer dados de séries temporais para a variável de interesse (Y) e seus preditores (X) em um período pré-intervenção suficientemente longo.

A Estimação dos Pesos

- O desafio do método é estimar os pesos (W) para as unidades de controle. As abordagens incluem:
 - **Método Naive 1:** Média simples, com pesos iguais para todas as unidades de controle.
 - **Método Naive 2:** Ponderado pela população.
 - **Método Sofisticado:** Minimizar a distância entre os preditores da unidade tratada e a combinação ponderada dos preditores das unidades de controle no período pré-intervenção.

Limitações e a Extensão Quantílica

- **Limitações do Método Tradicional:**
 - Dificuldade em realizar inferência estatística (obter intervalos de confiança).
 - Foco apenas no efeito médio, escondendo como diferentes partes da distribuição são afetadas.
- **Extensão com Controle Sintético Quantílico:**
 - **Motivação:** Toda política gera "vencedores e perdedores", e o efeito agregado pode ser enganoso.
 - **Ideia Central:** Estima o contrafactual para cada quantil da distribuição, não apenas a média.
 - **Requisito de Dados:** Exige microdados (dados granulares) para calcular os quantis em cada unidade.
 - **Vantagens:** Permite analisar o impacto da política sobre a desigualdade (ex: Índice de Gini) e sobre diferentes faixas de renda.

7.2. Estudo de Caso: O Impacto do Choque Migratório em Roraima

Esta seção aplica os conceitos de controle sintético para analisar os efeitos do choque migratório venezuelano.

Contexto da Pesquisa

- **Choque Migratório Exógeno:** A partir de 2013-2015, uma crise na Venezuela gerou uma emigração em massa para o estado de Roraima.
- **Isolamento de Roraima:** O isolamento geográfico do estado concentrou o impacto inicial, tornando-o um caso ideal para análise de choque exógeno.
- **Dados:** Utiliza dados da PNAD Contínua para analisar salário e emprego nos setores formal e informal.

Resultados do Estudo

- **Resultados Pontuais (Na Média):**
 - **Salário Agregado:** Queda estatisticamente significativa.
 - **Emprego Formal:** Nenhum impacto estatisticamente significativo.
 - **Emprego Informal:** Aumento estatisticamente significativo.
 - **Conclusão Preliminar:** O ajuste do mercado de trabalho ocorreu via setor informal.
- **Resultados Distribucionais (Nos Quantis):**
 - Os efeitos adversos se concentraram nos **trabalhadores de baixa renda**, que tiveram uma queda de salário mais acentuada.
 - Os trabalhadores de **alta renda** não tiveram mudança significativa no salário.
 - Como resultado, o choque migratório **aumentou a desigualdade** de renda em Roraima.

Análise sobre Criminalidade e Conclusões Gerais

- **Criminalidade:** A análise também mostrou que o choque migratório esteve associado a um aumento nas taxas de crimes como homicídios e roubos. O método, contudo, não permite identificar a autoria (imigrantes ou nativos).
- **Contribuição do Estudo:** O trabalho unifica um debate na literatura ao mostrar que o efeito da imigração no salário não é homogêneo, sendo negativo para os de baixa renda e nulo para os de alta renda.

8. AULA 9 - SELEÇÃO DE MODELOS, CAUSALIDADE E INTRODUÇÃO AOS MODELOS VAR

Esta aula aborda critérios para seleção de modelos, estabelece a ponte entre a análise de causalidade em micro e macroeconomia, e introduz formalmente os modelos de vetores autorregressivos (VAR).

8.1. Critérios de Seleção de Modelos e Suas Limitações

Seleção de Modelos Aninhados: AIC vs. BIC Para comparar modelos aninhados (onde um é um caso especial do outro), como na família ARIMA, utilizam-se critérios de informação.

- **Akaike Information Criterion (AIC):**

$$AIC_M = \log(\hat{\sigma}_M^2) + \frac{2K_M}{N}$$

O AIC representa um trade-off entre o bom ajuste do modelo (menor variância residual $\hat{\sigma}_M^2$) e uma penalidade por complexidade (aumentar o número de parâmetros K_M). Ele não é um critério consistente.

- **Bayesian Information Criterion (BIC):**

$$BIC_M = \log(\hat{\sigma}_M^2) + \frac{\log(N) \cdot K_M}{N}$$

A penalidade do BIC é mais rigorosa que a do AIC para amostras com $N \geq 8$, tendendo a escolher modelos mais parcimoniosos. O BIC é um critério consistente.

A Limitação dos Critérios e a Motivação para Machine Learning

- Para modelos não aninhados, o número de combinações possíveis de regressores cresce exponencialmente (2^K), tornando a comparação via AIC/BIC computacionalmente inviável.
- Essa limitação motiva o uso de métodos de *supervised machine learning* (Lasso, Elastic Net), que selecionam variáveis de forma eficiente em cenários de alta dimensionalidade.

8.2. Introdução à Causalidade em Macroeconomia

Causalidade: Da Microeconomia (RCT) à Macroeconomia (Choques)

- Em microeconomia, a causalidade é estudada pelo efeito de um programa ou política. Em macroeconomia, o análogo é um "**choque**": um evento novo e inesperado que afeta a economia (ex: um aumento inesperado na taxa de juros).
- O padrão-ouro para identificar causalidade é o **Randomized Control Trial (RCT)**, onde a alocação aleatória garante que os grupos de tratamento e controle sejam idênticos em média, exceto pelo tratamento.

- A **Análise de Impulso-Resposta** em macroeconomia é o equivalente ao Efeito Médio do Tratamento (ATE) dos RCTs, medindo como as variáveis respondem a um choque ao longo do tempo.

O Desafio em Macroeconomia: Estimação de Choques

- Diferente de um RCT, o choque macroeconômico não é diretamente observado e precisa ser estimado.
- Duas abordagens principais para isso são os **Vetores Autorregressivos Estruturais (SVAR)** e as **Projeções Locais**.

8.3. O Modelo de Vetores Autorregressivos (VAR)

O VAR Estrutural (SVAR): Contemporaneidade e Endogeneidade Um SVAR representa a teoria econômica, modelando múltiplas variáveis simultaneamente e incluindo efeitos contemporâneos, o que gera endogeneidade.

$$BY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \epsilon_t$$

Os erros ϵ_t são os choques estruturais "puros" e não correlacionados (matriz de covariância Σ_ϵ diagonal).

O VAR na Forma Reduzida: A Solução para Estimação Para estimar o modelo, o SVAR é transformado em sua forma reduzida, que não tem endogeneidade e pode ser estimada por MQO.

$$Y_t = A_0 + A_1 Y_{t-1} + e_t$$

Os erros da forma reduzida, $e_t = B^{-1}\epsilon_t$, são correlacionados entre si.

O Problema de Identificação: Contando Parâmetros O modelo estrutural possui mais parâmetros (10, no exemplo da aula) do que os que podem ser estimados a partir da forma reduzida (9). O sistema é sub-identificado e precisa de restrições teóricas para ser resolvido.

As Quatro Representações do VAR e a Função de Impulso-Resposta O caminho da estimação para a análise de política passa por quatro representações matemáticas:

1. **SVAR (Estrutural):** O modelo teórico de interesse.
2. **VAR (Forma Reduzida):** O modelo estimável.
3. **VMA (Vetorial de Médias Móveis):** Uma representação que expressa Y_t como função dos erros da forma reduzida passados (e_t).

$$Y_t = \mu + \sum_{k=0}^{\infty} A_1^k e_{t-k}$$

4. **SVMA (VMA Estrutural):** A representação final, que expressa Y_t como função dos choques estruturais puros (ϵ_t). Esta equação define a **Função de Impulso-Resposta**.

$$Y_t = \mu + \sum_{k=0}^{\infty} \underbrace{(A_1^k B^{-1})}_{\Theta_k} \epsilon_{t-k}$$

9. AULA 10 - FUNDAMENTOS DE VETORES AUTORREGRESSIVOS (VAR) E O PROBLEMA DE IDENTIFICAÇÃO

9.1. Revisão: As Quatro Representações do VAR e a Função de Impulso-Resposta (FIR)

A aula começa com uma revisão das quatro representações de um modelo de vetores autorregressivos, que são formas diferentes de escrever o mesmo sistema econômico.

- **VAR Estrutural (SVAR):** É o modelo de interesse teórico, pois seus erros são os **choques estruturais** (ϵ_t) não correlacionados, com uma matriz de covariância diagonal. No entanto, ele sofre de endogeneidade e não pode ser estimado diretamente por MQO. Em forma matricial, é representado como:

$$BY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \epsilon_t, \quad \text{onde } E[\epsilon_t \epsilon_t'] = \Sigma_\epsilon = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 \\ 0 & \sigma_{\epsilon_2}^2 \end{pmatrix}$$

- **VAR na Forma Reduzida:** É uma representação que não tem problema de endogeneidade, pois contém apenas variáveis defasadas no lado direito. Seus parâmetros podem ser estimados por MQO, mas seus erros (e_t) são correlacionados entre si. Os parâmetros da forma reduzida (A_0, A_1) são combinações dos parâmetros estruturais (B, Γ_0, Γ_1). Sua forma é:

$$Y_t = A_0 + A_1 Y_{t-1} + e_t, \quad \text{onde } A_i = B^{-1} \Gamma_i \text{ e } e_t = B^{-1} \epsilon_t$$

- **Representação Vetorial de Médias Móveis (VMA):** Derivada da forma reduzida, expressa as variáveis do sistema como uma soma infinita dos erros da forma reduzida (e_t). Matematicamente, para um VAR(1) estacionário:

$$(I - A_1 L) Y_t = A_0 + e_t \implies Y_t = (I - A_1 L)^{-1} A_0 + (I - A_1 L)^{-1} e_t = \mu + \sum_{k=0}^{\infty} A_1^k e_{t-k}$$

- **VMA Estrutural (SVMA):** É a representação mais importante para a análise de política, pois expressa as variáveis como uma função dos **choques estruturais** (ϵ_t). Ela é obtida substituindo a definição de e_t na representação VMA:

$$Y_t = \mu + \sum_{k=0}^{\infty} (A_1^k B^{-1}) \epsilon_{t-k} = \mu + \sum_{k=0}^{\infty} \Theta_k \epsilon_{t-k}$$

O objetivo final é estimar a **Função de Impulso-Resposta (FIR)**, que são os coeficientes da matriz $\Theta_k = A_1^k B^{-1}$. A FIR mede o efeito de um choque estrutural (ϵ_t) na trajetória futura das variáveis do sistema (Y_{t+S}). Essa análise é o análogo macroeconômico do *Average Treatment Effect* (ATE).

9.2. Identificação do VAR Estrutural: Restrições de Curto Prazo (Decomposição de Cholesky)

Como o VAR Estrutural tem mais parâmetros (10) do que os que podem ser estimados a partir da forma reduzida (9), é necessário impor restrições para identificar o modelo. A primeira abordagem discutida é a **Decomposição de Cholesky**, que se baseia em uma restrição de curto prazo.

- **A Restrição Matemática:** Impõe-se a hipótese de que um dos efeitos contemporâneos é nulo. No exemplo da aula, assume-se que $\beta_{1,2} = 0$. Isso significa que a matriz B de efeitos contemporâneos se torna triangular inferior:

$$B = \begin{pmatrix} 1 & 0 \\ \beta_{21} & 1 \end{pmatrix}$$

Essa restrição implica que a segunda variável (y_2) não tem impacto instantâneo sobre a primeira (y_1). A escolha de qual parâmetro restringir é, a princípio, arbitrária e não baseada em teoria econômica.

- **O Processo de Identificação em 4 Passos:**

1. **Estimar o VAR na Forma Reduzida:** Utiliza-se MQO para obter as estimativas dos parâmetros \hat{A}_1 e da matriz de covariância dos resíduos, $\hat{\Omega}$.

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T \hat{e}_t \hat{e}_t' = \begin{pmatrix} \hat{\omega}_{11} & \hat{\omega}_{12} \\ \hat{\omega}_{21} & \hat{\omega}_{22} \end{pmatrix}$$

2. **Estimar a Matriz B:** A restrição $\beta_{1,2} = 0$ permite resolver o sistema de equações que liga os parâmetros, $\Omega = B^{-1} \Sigma_{\epsilon} (B^{-1})'$. Isso resulta nas seguintes relações que permitem recuperar os parâmetros estruturais a partir dos parâmetros estimados da forma reduzida:

$$\hat{\sigma}_{\epsilon_1}^2 = \hat{\omega}_{11} \quad \hat{\beta}_{21} = -\frac{\hat{\omega}_{21}}{\hat{\sigma}_{\epsilon_1}^2} = -\frac{\hat{\omega}_{21}}{\hat{\omega}_{11}} \quad \hat{\sigma}_{\epsilon_2}^2 = \hat{\omega}_{22} - \hat{\beta}_{21}^2 \hat{\sigma}_{\epsilon_1}^2$$

Com isso, estima-se a matriz \hat{B} .

3. **Estimar os Coeficientes da FIR ($\hat{\Theta}_S$):** Os parâmetros da função de impulso-resposta são calculados usando a relação $\hat{\Theta}_S = \hat{A}_1^S (\hat{B})^{-1}$. Como \hat{A}_1 foi estimado no passo 1 e \hat{B} no passo 2, $\hat{\Theta}_S$ pode ser calculado para qualquer horizonte S .
4. **Plotar as Funções de Impulso-Resposta:** Os elementos da matriz $\hat{\Theta}_S$ são plotados ao longo do tempo (S) para visualizar a resposta dinâmica de cada variável a cada choque estrutural.

9.3. Identificação do VAR Estrutural: Restrições de Longo Prazo (Blanchard-Quah)

Como a restrição de Cholesky pode ser arbitrária, uma abordagem alternativa usa a teoria econômica para impor **restrições de longo prazo**.

- **A Matriz de Efeitos de Longo Prazo:** Define-se uma matriz de multiplicadores de longo prazo, $\Theta(1)$, cujos elementos são a soma (a integral) de cada função de impulso-resposta. Ela representa o efeito total e acumulado de cada choque sobre cada variável:

$$\Theta(1) = \sum_{S=0}^{\infty} \Theta_S = (I - A_1)^{-1} B^{-1}$$

- **Teoria Econômica como Restrição:** A teoria econômica é usada para justificar que alguns desses efeitos de longo prazo sejam nulos.
 - **Choques de Oferta e Demanda:** Distinguem-se dois tipos de choques estruturais: **choques de oferta** (ex: mudança tecnológica), que podem afetar permanentemente a capacidade produtiva, e **choques de demanda** (ex: aumento de gastos do governo), cujos efeitos sobre o produto real são considerados temporários.
 - **A Restrição de Longo Prazo:** Com base na teoria de que um choque de demanda (ex: ϵ_2) não afeta o produto (ex: y_1) no longo prazo, impõe-se a restrição de que o elemento correspondente na matriz de longo prazo seja zero. Matematicamente, $[\Theta(1)]_{1,2} = 0$. Esta equação fornece a restrição necessária para identificar o sistema.
- **Vantagem:** Essa abordagem fornece uma fundamentação econômica para a restrição de identificação, tornando-a menos arbitrária do que a de Cholesky.

9.4. Causalidade de Granger: Testando a Precedência Temporal

- **Conceito:** A Causalidade de Granger não testa a causalidade no sentido filosófico, mas sim a **precedência temporal**. A questão é "quem veio primeiro, o ovo ou a galinha?", ou seja, o passado de uma variável ajuda a prever o futuro de outra?
- **Relevância:** É uma ferramenta importante para testar a hipótese de **causalidade reversa**. Muitos modelos microeconômicos assumem uma direção causal (X causa Y) sem testá-la. Se a causalidade reversa existir, os estimadores podem ser viesados.
- **O Teste Matemático:** O teste é realizado dentro de um VAR na forma reduzida. Para testar se y_1 não causa y_2 no sentido de Granger, olhamos para a equação de y_2 em um VAR com k defasagens:

$$y_{2,t} = a_{20} + \sum_{i=1}^k a_{21}^{(i)} y_{1,t-i} + \sum_{i=1}^k a_{22}^{(i)} y_{2,t-i} + e_{2,t}$$

Testa-se a hipótese nula conjunta de que todos os coeficientes associados aos *lags* de y_1 são iguais a zero:

$$H_0 : a_{21}^{(1)} = a_{21}^{(2)} = \dots = a_{21}^{(k)} = 0$$

O teste estatístico utilizado para esta hipótese conjunta é o **Teste F**.

10. AULA 11 - IDENTIFICAÇÃO DE MODELOS VAR E INTRODUÇÃO A SÉRIES NÃO-ESTACIONÁRIAS

10.1. Identificação de VAR com Restrições de Longo Prazo: A Abordagem de Blanchard-Quah

Esta seção continua a discussão sobre a identificação de VARs estruturais, focando na abordagem que utiliza restrições de longo prazo baseadas na teoria econômica.

- **A Matriz de Covariância de Longo Prazo (Λ):** Um resultado fundamental da teoria de séries temporais é que a matriz de covariância de longo prazo de um vetor de variáveis Y_t pode ser expressa em função da matriz de impacto de longo prazo ($\Theta(1)$) e da matriz de covariância dos choques estruturais (Σ_ϵ). A relação é:

$$\Lambda = \Theta(1)\Sigma_\epsilon\Theta(1)'$$

- **A Estratégia de Identificação:** A identificação é alcançada combinando resultados da teoria econômica e da álgebra linear.

1. **Restrição 1 (Normalização):** Assume-se, sem perda de generalidade, que os choques estruturais têm variância unitária, de modo que sua matriz de covariância é a matriz identidade: $\Sigma_\epsilon = I$. Com isso, a equação se simplifica para $\Lambda = \Theta(1)\Theta(1)'$.
2. **Restrição 2 (Teoria Econômica):** Conforme discutido na aula anterior, a teoria econômica (modelo AD-AS) sugere que choques de demanda não têm efeito de longo prazo sobre o produto. Isso impõe uma restrição de zero em um dos elementos da matriz de impacto de longo prazo (ex: $[\Theta(1)]_{1,2} = 0$), o que torna a matriz $\Theta(1)$ **triangular inferior** (*lower triangular*).
3. **Solução via Decomposição de Matriz:** O problema se resume a encontrar uma matriz triangular inferior $\Theta(1)$ cujo produto com sua transposta seja igual à matriz Λ . Por definição, esta matriz $\Theta(1)$ é a **decomposição de Cholesky** (ou a "raiz quadrada") da matriz Λ .

- **O Processo Final:**

1. Estima-se o VAR na forma reduzida e calcula-se uma estimativa da matriz de covariância de longo prazo, $\hat{\Lambda}$.
2. Aplica-se a decomposição de Cholesky em $\hat{\Lambda}$ para obter $\hat{\Theta}(1)$.
3. Usa-se a relação $\hat{\Theta}(1) = (I - \hat{A}_1)^{-1}\hat{B}^{-1}$ para encontrar a matriz de efeitos contemporâneos \hat{B} .
4. Com \hat{B} e os parâmetros da forma reduzida, calculam-se as Funções de Impulso-Resposta ($\hat{\Theta}_S = \hat{A}_1^S \hat{B}^{-1}$).

10.2. Testes de Raiz Unitária: Dickey-Fuller e ADF

Esta seção aborda como testar formalmente se uma série temporal é não estacionária (possui uma raiz unitária).

- **O Problema com o Teste-t Padrão:** Ao testar a hipótese nula de raiz unitária ($H_0 : \phi = 1$) em um processo AR(1) como $y_t = \phi y_{t-1} + \epsilon_t$, a estatística-t usual não segue a distribuição normal padrão.
- **A Distribuição Dickey-Fuller (DF):** Sob a hipótese nula de raiz unitária, a estatística-t converge para uma distribuição não padrão, conhecida como distribuição Dickey-Fuller, que é derivada de funcionais do movimento Browniano.
 - **Movimento Browniano:** É um conceito de um passeio aleatório em tempo contínuo, geralmente definido no intervalo $[0, 1]$.
 - **Valores Críticos:** A consequência mais importante é que os valores críticos da distribuição DF são diferentes (mais negativos) que os da normal. Por exemplo, para um teste de 5% à esquerda, o valor crítico da normal é -1.64, enquanto o da DF é aproximadamente -1.95. Usar o valor crítico errado pode levar a rejeitar incorretamente a hipótese nula.
- **Componentes Determinísticos:** A especificação da regressão usada para o teste afeta a distribuição e os valores críticos. A inspeção visual do gráfico da série ajuda a decidir qual caso usar:
 - **Sem intercepto nem tendência:** O valor crítico é ~ -1.95 .
 - **Com intercepto:** A distribuição muda para uma versão "de-meaned" (média removida), e o valor crítico se torna mais negativo (~ -2.9).
 - **Com intercepto e tendência:** A distribuição muda para uma versão "detrended" (tendência removida), e o valor crítico é ainda mais negativo (~ -3.4).
- **Teste Augmented Dickey-Fuller (ADF):** O teste DF padrão assume que os erros são ruído branco. Se houver correlação serial nos erros, o teste ADF deve ser usado. Ele "aumenta" a regressão adicionando defasagens da variável em primeira diferença (Δy_{t-p}) para capturar essa correlação serial e garantir que o resíduo final seja ruído branco. Os valores críticos do teste ADF são os mesmos do teste DF correspondente.

10.3. Cointegração e Modelos de Correção de Erros (VECM)

- **Regressão Espúria:** Regredir uma série não estacionária em outra série não estacionária independente geralmente resulta em um coeficiente estatisticamente significativo e um R^2 alto, mesmo que não haja relação verdadeira entre elas. Isso é conhecido como **regressão espúria**. Uma solução simples é rodar a regressão com as variáveis em primeira diferença.
- **Conceito de Cointegração:** Um conjunto de séries não estacionárias ($I(1)$) é dito **cointegrado** se existe uma combinação linear entre elas que é estacionária ($I(0)$).
 - **Intuição Econômica:** A cointegração representa uma **relação de equilíbrio de longo prazo** estável entre as variáveis. Embora as séries possam se desviar

aleatoriamente no curto prazo, elas estão "amarradas" por essa relação de longo prazo.

- **Exemplos:** Renda e consumo (hipótese da renda permanente); taxas de juros de curto e longo prazo (arbitragem); preços de ações e dividendos; taxa de câmbio à vista e a termo.
- **Modelo de Correção de Erros (VECM):** Se as séries são cointegradas, usar um VAR apenas em primeiras diferenças descarta a valiosa informação sobre a relação de longo prazo. O VECM é um VAR que incorpora essa informação.
 - **Estrutura:** É um VAR com as variáveis em primeira diferença (ΔY_t), mas que inclui um termo adicional: o **termo de correção de erro**, que é o desvio do equilíbrio de longo prazo no período anterior (ex: $c_{t-1} - y_{t-1}$).
 - **Mecanismo (Termostato):** Se no período $t - 1$ as variáveis se desviam do equilíbrio, o termo de correção de erro se torna diferente de zero. No período t , o modelo prevê uma mudança em ΔY_t que "corrige" parte desse desequilíbrio, empurrando o sistema de volta para a relação de longo prazo. O coeficiente α mede a velocidade desse ajuste.

11. AULA 12 - COINTEGRAÇÃO: TESTES E MODELOS DE CORREÇÃO DE ERROS (VECM)

11.1. Regressão Espúria e a Definição de Cointegração

Esta seção formaliza o conceito de cointegração como a solução para o problema da regressão espúria quando se trabalha com múltiplas séries temporais não estacionárias.

- **Problema e Solução Padrão:** O problema da **regressão espúria** ocorre quando se roda uma regressão em nível com variáveis não estacionárias ($I(1)$), o que pode gerar resultados estatisticamente significativos falsos. A solução mais simples é rodar um Vetor Autorregressivo (VAR) com as variáveis em **primeira diferença** (ΔY_t).
- **A Limitação do VAR em Diferenças:** Se as séries não estacionárias possuem uma relação de equilíbrio de longo prazo, um VAR em primeiras diferenças é um modelo mal especificado, pois omite uma variável preditora importante. O modelo correto nesses casos é o **Modelo de Correção de Erros (VECM)**.
- **Definição Formal de Cointegração:** Um vetor de N séries temporais Y_t , onde cada série é não estacionária ($I(1)$), é dito **cointegrado** se existe um vetor de cointegração β tal que a combinação linear $\beta'Y_t$ é estacionária ($I(0)$).
 - **Normalização:** O vetor β não é único. Para garantir a unicidade, aplica-se uma normalização, usualmente fixando o primeiro coeficiente como 1. Isso permite escrever a relação de cointegração como uma equação de regressão de longo prazo:

$$y_{1t} = \alpha + \beta_2 y_{2t} + \dots + \beta_n y_{nt} + u_t$$

Onde $u_t = \beta'Y_t$ é o termo de erro estacionário que representa os desvios do equilíbrio de longo prazo. Uma regressão em nível só faz sentido se as variáveis

forem cointegradas.

11.2. A Intuição da Cointegração: Tendências Estocásticas Comuns

A cointegração é possível porque as séries não estacionárias compartilham a mesma fonte de não estacionariedade.

- **Representação de um Passeio Aleatório:** Um processo com raiz unitária (passeio aleatório), $y_t = y_{t-1} + \epsilon_t$, pode ser reescrito como uma acumulação de choques passados:

$$y_t = \sum_{s=1}^t \epsilon_s$$

Este termo de somatório, $\sum \epsilon_s$, é a **tendência estocástica** que torna a série não estacionária.

- **Tendência Estocástica Comum:** Séries temporais são cointegradas quando são geradas pela mesma tendência estocástica comum. O vetor de cointegração funciona ao criar uma combinação linear que cancela essa tendência comum, resultando em uma série estacionária.

- **Exemplo:** Se temos duas séries y_1 e y_2 geradas por uma mesma tendência estocástica $TS_t = \sum v_s$:

$$* y_{2t} = TS_t + \text{ruído estacionário}_2$$

$$* y_{1t} = \beta_2 \cdot TS_t + \text{ruído estacionário}_1$$

A combinação linear $y_{1t} - \beta_2 y_{2t}$ irá eliminar o termo TS_t , resultando em uma série puramente estacionária.

- **Número de Vetores e Tendências:** Para um sistema com N variáveis, existe uma relação fundamental: se há **R vetores de cointegração**, então deve haver **$N-R$ tendências estocásticas comuns**.

- Se $N = 2$ e $R = 1 \implies$ há $2 - 1 = 1$ tendência estocástica comum.
- Se $N = 3$ e $R = 1 \implies$ há $3 - 1 = 2$ tendências estocásticas comuns.
- Se $N = 3$ e $R = 2 \implies$ há $3 - 2 = 1$ tendência estocástica comum.
- O número de vetores de cointegração R não pode ser igual a N , pois isso implicaria $N - R = 0$ tendências estocásticas, significando que as séries já eram estacionárias para começar.

11.3. Testes de Cointegração

Existem duas abordagens principais para testar a existência e o número de relações de cointegração.

11.3.1. O Teste de Engle-Granger (Baseado nos Resíduos)

Este teste é mais simples e aplicável quando se suspeita que há no máximo **um** vetor de cointegração ($R = 1$).

- **Procedimento em 2 Passos:**

1. **Estimar a Regressão de Longo Prazo:** Roda-se uma regressão em nível de uma variável sobre as outras por MQO e salvam-se os resíduos, \hat{u}_t .
2. **Testar os Resíduos:** Aplica-se um teste de raiz unitária (ADF) sobre os resíduos \hat{u}_t .

- **Hipóteses:**

- H_0 : O resíduo possui raiz unitária (não é estacionário).
- H_A : O resíduo é estacionário.

- **Conclusão:** A hipótese nula de **não cointegração** é rejeitada se a hipótese nula do teste de raiz unitária for rejeitada (ou seja, se os resíduos forem estacionários).
- **Valores Críticos:** Os valores críticos para o teste de raiz unitária sobre resíduos são ligeiramente diferentes dos valores críticos do ADF padrão, mas os pacotes estatísticos fazem este ajuste automaticamente.

11.3.2. O Teste de Johansen (Baseado no Sistema)

Este é o teste mais popular e geral, pois permite identificar múltiplos vetores de cointegração ($R \geq 1$).

- **Procedimento Sequencial:** O teste determina o número de vetores de cointegração (R) através de uma sequência de testes de hipóteses.

1. **Etapa 1:** Testa-se $H_0 : R = 0$ (não há cointegração) contra $H_A : R > 0$. Se H_0 não for rejeitada, o teste para e conclui-se que não há cointegração. Se for rejeitada, passa-se para a próxima etapa.
2. **Etapa 2:** Testa-se $H_0 : R = 1$ contra $H_A : R > 1$. Se H_0 não for rejeitada, o teste para e conclui-se que existe um vetor de cointegração. Se for rejeitada, continua-se.
3. **Continuação:** O processo continua, testando $H_0 : R = k$ contra $H_A : R > k$, até que uma hipótese nula não seja rejeitada.

- **Importância dos Componentes Determinísticos:** A escolha correta do caso para os componentes determinísticos (constante, tendência) é crucial, pois afeta os valores críticos do teste. A inspeção visual do gráfico das séries é o primeiro passo para determinar qual dos cinco casos possíveis deve ser utilizado no teste de Johansen.

11.4. O Modelo de Correção de Erros (VECM) e a Previsão

- **Estrutura:** O VECM é um VAR em primeiras diferenças que incorpora a relação de equilíbrio de longo prazo. Para um sistema com um vetor de cointegração, sua

forma é:

$$\Delta Y_t = \alpha(\beta'Y_{t-1}) + \text{lags de } \Delta Y_t + \dots$$

Onde $\beta'Y_{t-1}$ é o **termo de correção de erro**, que representa o desequilíbrio do período anterior.

- **Mecanismo (Termostato):** Se no período $t - 1$ as variáveis se desviam do equilíbrio, o termo $(\beta'Y_{t-1})$ se torna diferente de zero. O coeficiente α mede a **velocidade de ajuste**, ou seja, quanto de ΔY_t mudará no período t para "corrigir" o erro e levar o sistema de volta ao equilíbrio.
- **Poder de Previsão:** O VECM é um modelo de previsão poderoso porque combina a dinâmica de curto prazo (os lags de ΔY_t) com a teoria econômica de longo prazo (a relação de cointegração). Isso o torna superior a um simples VAR em diferenças quando a cointegração está presente.