AULA 16 - Autocorrelação

Econometria I

Área quantitativa - IE/UFRJ

Autocorrelação

- O mais comum problema de séries de tempo é a autocorrelação ou correlação serial dos distúrbios entre os períodos.
- Os problemas de estimação e inferência são semelhantes a heterocedasticidade. Sendo os mínimos quadrados ineficientes, e a inferência é afetada negativamente.
- Apesar de serem ineficientes, são não viesados.
- A forma de como o problema será atacado pode diferenciar da heterocedasticidade, pois em muitos casos o problema de autocorrelação está associado a um problema de má especificação do modelo.

Autocorrelação

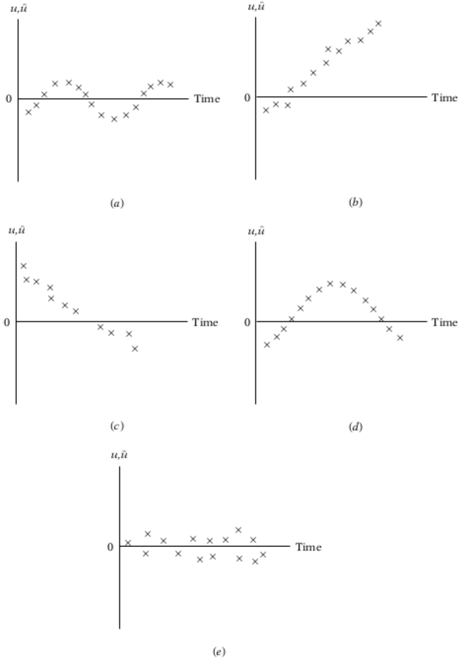
O que vamos abordar:

- Qual é a natureza da autocorrelação?
- Quais são as consequências teóricas e práticas da autocorrelação?
- Uma vez que a hipótese de não haver autocorrelação diz respeito ao distúrbios não observados u_t, como se sabe que há autocorrelação em qualquer dada situação? Faremos testes
- Observe que agora usamos o subescrito t para enfatizar que estamos lidando com dados de séries temporais.
- Como se resolve o problema da autocorrelação?

A natureza da autocorrelação

- Na configuração de séries temporais, os distúrbios são assumidos homoscedásticos, mas correlacionados entre as observações, de modo que
$$E[uu'|X] = \sigma^2\Omega$$
em que $\sigma^2\Omega$ é uma matriz positiva definida de posto total
$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$

Comportamentos típicos de auto- e não autocorrelação



A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Inércia**. A maioria das séries temporais econômicas possuem um comportamento inercial, tais como: PIB, índices de preços, ciclos de produção, emprego.
  Quando a recuperação econômica começa, a maioria dessas séries começam a se mover para cima. Nesta subida, o valor de uma série em um ponto no tempo é maior que seu valor anterior.

A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Viés de Especificação: Variáveis Excluídas**. Na análise empírica o pesquisador geralmente começa com um modelo de regressão plausível que pode não ser o mais "perfeito".
  Exemplo: Seja o modelo correto
  $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$
  onde Y = quantidade de carne demandada, X_2 = preço da carne, X_3 = renda do consumidor, X_4 = preço do frango
  Entretanto, estimamos: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t$ assim
  $v_t = \beta_4 X_{4t} + u_t$
  Na medida em que o preço do frango afeta o consumo de carne bovina, o erro v refletirá um padrão sistemático, criando assim (falsa) autocorrelação.

A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Viés de Especificação: Forma Funcional Incorreta**.
  Exemplo: Seja o modelo correto
  Custo marginal$_t = \beta_1 + \beta_2 produto_t + \beta_3 produto_t^2 + u_t$
  Entretanto, estimamos:
  Custo marginal$_t = \beta_1 + \beta_2 produto_t + u_t$

A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Fenômeno de teia de aranha**. A oferta de muitas commodities agrícolas reage ao preço com uma defasagem de um período de tempo porque as decisões de fornecimento levam tempo para serem implementadas. Exemplo: Seja o modelo correto $\text{Oferta}_t = \beta_1 + \beta_2 P_{t-1} + u_t$
- **Defasagem** A variável dependente é explicada por sua variável defasada.
  Exemplo: $\text{Consumo}_t = \beta_1 + \beta_2 \text{renda}_t + \beta_3 \text{Consumo}_{t-1} + u_t$

A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Manipulação dos dados**.
  Por exemplo, em regressões envolvendo dados trimestrais, esses são geralmente derivados dos dados mensais simplesmente somando e divindo por 3. Tornando assim a série mais suave.
  Outra fonte de manipulação é a interpolação ou extrapolação de dados. Por exemplo, o Censo da População é realizado a cada 10 anos em um país, sendo o último em 2000 e o anterior em 1990. Agora, se existe uma necessidade de obter dados por algum ano dentro do período 1990-2000, a prática comum é interpolar com base em algumas suposições subjetivas.

A natureza da autocorrelação

Questão: por que ocorre a correlação serial?

- **Transformação dos dados**.
  Considere o modelo $Y_t = \beta_1 + \beta_2 X_t + u_t$
  Se isto é verdade, então para o período anterior será verdade também $Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$ que são os valores defasados de Y, X e u.
  Agora de subtrairmos os dois modelos temos $\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t$ onde o operado $\Delta Y = (Y_t - Y_{t-1})$ e o mesmo para as demais
  Para a modelagem empírica usamos $\Delta Y_t = \beta_2 \Delta X_t + v_t$ onde $v_t = \Delta u_t = (u_t - u_{t-1})$ Chamamos de forma na diferença, o que significa a variação da variável.
  Muitas vezes temos usar a modelagem na diferença, porque as séries são não estacionárias (tendência estocástica).

MQO na presença de autocorrelação

- Como mostramos acima
$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$
- Definimos as autocovariâncias:
$$Cov[u_t, u_{t-s}|X] = Cov[u_{t+s}, u_t|X] = \sigma^2\Omega_{t,t-s} = \gamma_s$$
- A correlação entre u_t e u_{t-s} é a sua autocorrelação
$$Corr[u_t, u_{t-s}|X] = \frac{Cov[u_t, u_{t-s}|X]}{\sqrt{Var[u_t|X]Var[u_{t-s}|X]}} = \frac{\gamma_s}{\gamma_0} = \rho_s$$

MQO na presença de autocorrelação

- Para derivarmos a matriz var-covar quando as observações são autocorrelacionadas, mostraremos o processo AR(1) - modelos com ordem maior são construídos como um refinamento.
- O processo AR(1) é representado na forma autoregressiva como

$$u_t = \rho u_{t-1} + \epsilon_t, \quad -1 < \rho < 1$$

  em que $E[\epsilon_t] = 0$, $var[\epsilon_t] = \sigma^2_\epsilon$ e $Cov[\epsilon_t, \epsilon_{t+s}] = 0$ se $s \neq 0$
- Dividing por γ_0 a autocorrelação é definida por:

$$Corr[u_t, u_{t-1}] = \rho_1$$

MQO na presença de autocorrelação

- Voltando ao modelo de regressão simples $Y_t = \beta_1 + \beta_2 X_t + u_t$ temos que o coeficiente do estimador MQO é

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2}$$

e sua variância:

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$$

para simplificar a notação dizemos que a letra em minúscula é o desvio da média.

- Agora, sob o esquema AR(1), pode-se mostrar que a variância deste estimador é:

$$var(\hat{\beta}_2)_{AR1} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + ... + 2\rho^{n-1}\frac{\sum x_t x_n}{\sum x_t^2}\right]$$

Testes para autocorrelação

- **Teste LM**: é o teste do multiplicador de Lagrange sendo
  H_0: não autocorrelação
  H_1: $\epsilon_t = AR(P)$ ou $\epsilon_t = MA(P)$
  O teste faz a regressão por MQO dos resíduos e_t em x_t, $e_{t-1}...e_{t-p}$ e TR^2 o valor crítico para uma distribuição qui-quadrado com P graus de liberdade.
  Rejeita H_0 se $LM > \chi^2_*[P]$
- **Teste Q** Examina as correlações entre os resíduos e os P valores defasados dos resíduos.
  O teste sugerido por Ljung e Box é:
  $$Q = T(T+2)\sum_{j=1}^{P} \frac{r_j^2}{T-j}$$
  Rejeita H_0 se $Q > \chi^2_*[P]$

Testes para autocorrelação

- **Teste Durbin-Watson**: a estatística do teste é:
  $$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = 2(1-r) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2}$$
  em que r é a autocorrelação de primeira ordem. Se a amostra é grande o último termo pode ser negligenciado e temos
  $d \approx 2(1-r)$
  H_0: $\rho = 0$
  H_1: $\rho > 0$
  A distribuição amostral de d depende dos valores das variáveis explicativas e assim o teste tem limite superior (du) e inferior (dL) para o nível de significância.
  Se $d < dL$ rejeitamos a H_0 Se $d > du$ não rejeitamos a H_0 Se $dL < d < du$ o teste é inconclusivo
- Este teste é limitado pois considera somente autocorrelação de primeira ordem e é inconclusivo se está entre dL e du.
- Existe uma versão para a presença de defasagem desenvolvida por Durbin.

Estimação eficiente com Ω conhecido

- Sendo Ω conhecido, então o estimador GLS
$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$$
e a estimativa da variância amostral
$$Est.Var[\hat{\beta}] = \hat{\sigma}_{\epsilon}^{2}[X'\Omega^{-1}X]^{-1}$$
em que
$$\hat{\sigma}_{\epsilon}^{2} = \frac{(y-X\hat{\beta})'\Omega^{-1}(y-X\hat{\beta})}{T}$$
pode ser computada em um passo.

Estimação eficiente com Ω conhecido

- Os dados para o modelo são transformados. No caso AR(1) temos
$$y_* = \begin{bmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}, \quad X_* = \begin{bmatrix} \sqrt{1-\rho^2} x_1 \\ x_2 - \rho x_1 \\ x_3 - \rho x_2 \\ \vdots \\ x_T - \rho x_{T-1} \end{bmatrix}$$
- A variância dos erros transformados são: $Var[\epsilon_t - \rho\epsilon_{t-1}] = Var[u_t] = \sigma_u^2$
- A variância dos erro pode ser estimada usando $(1-\rho^2)\hat{\sigma}_\epsilon^2$

Estimação quando Ω é desconhecido

- É preciso achar um estimador consistente de $\Omega(\rho)$
- O modelo AR(1) é o mais amplamente utilizado e estudado.
- Podemos usar o FGLS com um estimador natural de ρ.
- Como b é consistente, nós podemos usar r.
- O segundo passo para o FGLS pode seguir a estimativa de Prais and Winsten ou omitir a primeira observação basado na estimativa de Cochrane e Orcutt.
- Diferentemente do modelo com heterocedasticidade, interar quando tem autocorrelação não produz o estimador por máxima verossimilhança.
- Os estimadores de máxima verossimilhança podem ser obtidos maximizando a probabilidade de logaritmo em relação a β, σ_u^2 e ρ.