Formulário de Econometria

Artur Silva Lopes, 28/4/2016[1]

Revisão – Estatística

$\mathrm{Var}(X) = \mathrm{E}(X-\mu)^2 = \mathrm{E}(X^2) - \mu^2$; $\mathrm{Cov}(X, Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$; $\rho_{X,Y} = \mathrm{corr}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\,\sigma_Y}$.

$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|\,X)]$; $\mathrm{E}[g(X)|\,x] = g(x)$; se X e Y são independentes, então $\mathrm{E}(Y|\,X) = \mathrm{E}(Y)$; se $\mathrm{E}(Y|\,X) = E(Y)$, então $\mathrm{Cov}(Y, X) = 0$ (e $\rho_{X,Y} = 0$).

Modelo de regressão linear para amostras aleatórias de dados seccionais

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \;\Leftrightarrow\; y_i = x_i\boldsymbol{\beta} + u_i,\; i = 1, 2, \ldots, n \;\Leftrightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

OLS: $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\widehat{u}_i = y_i - \widehat{y}_i$, $\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k-1)$, $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$.

$\widehat{\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\,\mathbf{X})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$, $\widehat{\mathrm{Var}(\widehat{\beta}_j|\,\mathbf{X})} = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)}, j = 1, 2, \ldots, k$, com $SST_j = \sum(x_{ij} - \bar{x}_j)^2$ e R_j^2 o R^2 da regressão de x_j sobre os restantes regressores.

$H_0 : \beta_j = \beta_j^0$, $t = (\widehat{\beta}_j - \beta_j^0)/se(\widehat{\beta}_j) \sim t_{(n-k-1)}$ ou $\overset{a}{\sim} \mathcal{N}(0, 1)$ sob H_0.

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$, $F = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k} \sim F_{(k,n-k-1)}$ sob H_0.

Quaisquer q restrições (com y inalterado): $F = \frac{SSR_R - SSR_{UR}}{SSR_{UR}} \times \frac{n-k-1}{q} = \frac{R^2_{UR} - R^2_R}{1-R^2_{UR}} \times \frac{n-k-1}{q} \sim F_{(q,n-k-1)}$ sob H_0, com SSR_{UR} a soma dos quadrados dos resíduos do modelo sem restrições e SSR_R a soma dos quadrados dos resíduos do modelo com as restrições.

Inferência robusta à heteroscedasticidade:
a) matriz de White: $\widehat{\mathrm{Var}^*(\widehat{\boldsymbol{\beta}}|\,\mathbf{X})} = (\mathbf{X}'\mathbf{X})^{-1}\sum \hat{u}_i^2\mathbf{x}_i'\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}$;
b) estatísticas$-t$ robustas: $H_0 : \beta_j = \beta_j^0$, $t = (\widehat{\beta}_j - \beta_j^0)/se^*(\widehat{\beta}_j) \overset{a}{\sim} \mathcal{N}(0, 1)$ sob H_0; $se^*(\widehat{\beta}_j) = \sqrt{\widehat{\mathrm{Var}^*(\hat{\beta}_j|\,\mathbf{X})}}$. EViews: ...estimate ... OPTIONS: coefficient covariance matrix -> White.

Estatísticas de teste (LM) de heteroscedasticidade: $LM = n\,R^2_{\hat{u}^2} \overset{a}{\sim} \chi^2_{(p)}$ sob H_0; p = número de regressores (excluindo o termo independente) da regressão auxiliar de teste.

Variáveis *dummy*

Afectando apenas o termo independente. Exemplo: $wage = \beta_0 + \delta_0\,female + \beta_1 educ + erro$, para não cair na armadilha das variáveis artificiais, violando a hipótese de não colinearidade perfeita. $\delta_0 = \mathrm{E}(wage|\,female = 1, educ) - \mathrm{E}(wage|\,female = 0, educ)$.

Se a característica qualitativa faz uma partição em g grupos distintos, introduzem-se no modelo apenas $g-1$ variáveis *dummy*, mantendo o termo independente.

Modelo com $\log(y)$: % Δ exacta $= 100[\exp(\widehat{\beta}_1) - 1]$, com $\widehat{\beta}_1$ o coeficiente da *dummy*.

Afectando também o coeficiente de declive. Exemplo: $wage = \beta_0 + \delta_0\,female + \beta_1\,educ + \delta_1\,female \times educ + erro$.

Teste de igualdade de equações de regressão. A) Com D a representar a variável *dummy*, $H_0 : \delta_0 = \delta_1 = \ldots = \delta_k = 0$ em $y = \beta_0 + \delta_0\,D + \beta_1\,x_1 + \delta_1\,D\,x_1 + \ldots + \beta_k\,x_k + \delta_k\,D\,x_k + erro$. B) $H_0 : \alpha_0 = \beta_0, \alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k$ em $y = \alpha_0 + \alpha_1\,x_1 + \ldots + \alpha_k\,x_k + u$ para grupo 1 e $y = \beta_0 + \beta_1\,x_1 + \ldots + \beta_k\,x_k + u$ para grupo 2. $F_{CHOW} = \frac{[SSR_p - (SSR_1 + SSR_2)]/(k+1)}{(SSR_1+SSR_2)/[n-2(k+1)]} \sim F_{(k+1,n-2(k+1))}$

[1] Elaborado com base em Wooldridge, J. M. (2009), *Introductory Econometrics*, 4th ed., South-Western, Cengage Learning, e com os comentários e sugestões de Luizete Reis, Ana Regina Pereira e Pierre Hoonout, os quais não têm, no entanto, qualquer responsabilidade nos erros e omissões que possam subsistir.

sob H_0, com SSR_p a SSR obtida do modelo estimado com o conjunto de todas as observações. Se não interessar testar a igualdade dos termos independentes, a SSR_p é calculada com base num modelo que inclui uma *dummy* com efeito sobre esse coeficiente.

Modelos para variáveis dependentes binárias

$y = 0$ ou 1. $y|\,\mathbf{x} \sim B(1, p(\mathbf{x}))$. Revisão: se $X \sim B(1,\theta)$, então $\mathrm{E}(X) = \theta$ e $\mathrm{Var}(X) = \theta(1-\theta)$. $f(x|\,\theta) = \theta^x\,(1-\theta)^{1-x}$, $x = 0, 1$.

Modelo linear de probabilidades: $\mathrm{E}(y|\,\mathbf{x}) = p(\mathbf{x}) = P(y = 1|\,\mathbf{x}) = \beta_0 + \beta_1\,x_1 + \ldots + \beta_k\,x_k$. Logo, $\beta_j = \Delta P(y = 1|\mathbf{x})/\Delta x_j$. Deficiências: i) eventuais $\widehat{y}_i < 0$ ou $\widehat{y}_i > 1$; ii) linearidade (efeitos parciais constantes); iii) heteroscedasticidade.

Previsão dos y_i's: $\tilde{y}_i = 1$, se $\widehat{y}_i \geq 0.5$, e $\tilde{y}_i = 0$, se $\widehat{y}_i < 0.5$.

Modelos Logit e Probit: $P(y = 1|\,\mathbf{x}) = G(\beta_0 + \beta_1\,x_1 + \ldots + \beta_k\,x_k) = G(\mathbf{x}\,\boldsymbol{\beta})$, com $0 < G(z) < 1$. Logit: $G(z) = \Lambda(z) = \frac{\exp(z)}{1+\exp(z)}$; EViews: @clogistic(.). Probit: $G(z) = \Phi(z) = \int_{\infty}^{z} \phi(v)dv$, $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$; EViews: @cnorm(.).

Efeitos parciais das variáveis. **A.** Para pequenas variações: A.1 Se x_j é uma variável aproximadamente contínua, pode fazer-se $\frac{\partial p(\mathbf{x})}{\partial x_j} = g(\mathbf{x}\,\boldsymbol{\beta})\,\beta_j = g(\beta_0 + \beta_1\,x_1 + \ldots + \beta_k x_k)\,\beta_j$, onde $g(z) = \frac{d\,G(z)}{d\,z}$; $g(z) > 0, \forall z$. EViews: @dlogistic(.) e @dnorm(.), respectivamente; A.2 Se x_k é uma variável discreta, quando passa de c_k para $c_k + 1$: $G[(\beta_0 + \beta_1\,x_1 + \ldots + \beta_k(c_k + 1)] - G(\beta_0 + \beta_1\,x_1 + \ldots + \beta_k\,c_k)$. **B.** Outros casos: $\Delta P(y = 1|\,\mathbf{x}) = G(\text{novo ponto}) - G(\text{ponto inicial})$.

Estimação dos efeitos parciais médios. i) Para variáveis aproximadamente contínuas: $[\frac{1}{n}\sum_{i=1}^{n} g(\widehat{\beta}_0 + \widehat{\beta}_1\,x_{i1} + \ldots + \widehat{\beta}_k\,x_{ik})]\,\widehat{\beta}_j = [\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}_i\,\widehat{\boldsymbol{\beta}})]\,\widehat{\beta}_j$. Exemplo EViews: series facesc = @dnorm(c(1) + c(2)*x2) series efeipar = facesc*c(2) scalar epm=@mean(efeipar).

ii) Para variáveis discretas. x_k passa de c_k para $c_k + 1$:

$$\frac{1}{n}\sum_{i=1}^{n}\left[G\left(\widehat{\beta}_0 + \widehat{\beta}_1\,x_{i1} + \ldots + \widehat{\beta}_k\,(c_k + 1)\right) - G(\widehat{\beta}_0 + \widehat{\beta}_1\,x_{i1} + \ldots + \widehat{\beta}_k\,c_k)\right].$$

Exemplo de EViews, com x_2 =D(*ummy*): series a= @cnorm(c(1)+c(2)*x1+c(3)) series b=@cnorm (c(1)+c(2)*x1) scalar epmd=@mean(a-b).

Estimação de MV: $\widehat{\boldsymbol{\beta}} = \arg\max \mathcal{L}(\boldsymbol{\beta})$. $\mathcal{L}(\boldsymbol{\beta}) = \sum l_i(\boldsymbol{\beta}) = \sum y_i \log[G(\mathbf{x_i}\boldsymbol{\beta})] + (1 - y_i)\log[1 - G(\mathbf{x_i}\,\boldsymbol{\beta})] < 0, \forall \boldsymbol{\beta}$. EViews: quick -> estimate equation ...-> METHOD: binary -> logit ou probit. Propriedades: a) $\mathrm{plim}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$; b) $\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}(.,.)$; c) $\widehat{\boldsymbol{\beta}}$ é assintoticamente eficiente; d) $se(\widehat{\beta}_j)$, $j = 0, 1, \ldots, k$, são válidos assintoticamente.

Inferência. $H_0 : \beta_j = 0$, $\widehat{\beta}_j / se(\widehat{\beta}_j) \overset{a}{\sim} \mathcal{N}(0,1)$ sob H_0. Quaisquer q restrições de exclusão: $LR = 2\,(\mathcal{L}_{UR} - \mathcal{L}_R) \overset{a}{\sim} \chi^2_{(q)}$ sob H_0. EViews: view -> coefficient tests -> redundant variables – Likelihood ratio -> ... (nomes das variáveis).

Bondade do ajustamento. % previsões correctas = $\frac{npc}{n} \times 100$, npc =número de previsões correctas. EViews: view -> prediction-expectation evaluation -> valor de limiar = 0.5.

Para comparações grosseiras: coef. probit $\overset{\times 1.6}{\longrightarrow}$coef. logit; coef. logit $\overset{\times 0.625}{\longrightarrow}$coef. probit; coef. probit $\overset{\times 0.4}{\longrightarrow}$coef. MLP; coef. logit $\overset{\times 0.25}{\longrightarrow}$coef. MLP.

Regressão básica com séries temporais

Série temporal, x_t, é uma realização de um processo estocástico, $\{x_t\}$, sucessão de variáveis aleatórias indexadas pelo tempo.

Exemplo de modelo estático: $y_t = \beta_0 + \beta_1 z_t + u_t, t = 1, 2, \ldots, n$. Exemplo de modelo dinâmico, FDL(2): $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$; MCP $= \delta_0$; MLP$= \delta_0 + \delta_1 + \delta_2$. Variação transitória unitária em t (e $u_t \equiv 0, \forall t$): $\delta_0 = y_t - y_{t-1}; \delta_1 = y_{t+1} - y_{t-1}; \delta_2 = y_{t+2} - y_{t-1}$. Variação permanente unitária em t (e $u_t \equiv 0, \forall t$): $y_t - y_{t-1} = \delta_0; y_{t+1} - y_{t-1} = \delta_0 + \delta_1; y_{t+2} - y_{t-1} = \delta_0 + \delta_1 + \delta_2, \ldots$.

Modelo clássico. TS.1 Linearidade nos parâmetros: $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t)\} : y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, t = 1, 2, \ldots, n$; $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tk})$, $\mathbf{X}$ é $n \times (k+1)$; notação matricial: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Importante: a hipótese de amostragem aleatória não é assumida. **TS.2** Exogeneidade estrita: $\mathrm{E}(u_t | \mathbf{X}) = 0, \forall t \Rightarrow \mathrm{Cov}(u_t, x_{sj}) = 0, \forall t, s, j$. Mais forte que exogeneidade contemporânea: $\mathrm{E}(u_t | \mathbf{x}_t) = \mathrm{E}(u_t | x_{t1}, x_{t2}, \ldots, x_{tk}) = 0, \forall t \Rightarrow \mathrm{Cov}(u_t, x_{tj}) = 0, \forall t, j$. Violações de TS.2: forma funcional incorrecta; variáveis omitidas correlacionadas com as incluídas; alguns casos de erros de medição de regressores; variações dos erros que provocam variações futuras dos regressores (*feedback*). **TS.3** Não colinearidade perfeita: $r(\mathbf{X}) = k + 1$. **Teor.**: TS.1 + TS.2 + TS.3 $\Rightarrow \mathrm{E}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}(\widehat{\boldsymbol{\beta}} | \mathbf{X}) = \boldsymbol{\beta}$.

TS.4 Homoscedasticidade: $\mathrm{Var}(u_t | \mathbf{X}) = \sigma^2, t = 1, 2, \ldots, n$. **TS.5** Ausência de autocorrelação: $\mathrm{corr}(u_t, u_s) = 0, \forall t, s, t \neq s$. **Teor.**: TS.1 a TS.5 (hipóteses de Gauss-Markov) $\Rightarrow \mathrm{Var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)}, j = 1, 2, \ldots, k$. **Teor.**: TS.1 a TS.5 $\Rightarrow$ com $\hat{\sigma}^2 = SSR/(n-k-1) = SSR/DF$, $\mathrm{E}(\hat{\sigma}^2) = \sigma^2$. **Teor.** (Gauss-Markov): TS.1 a TS.5 $\Rightarrow$ condicional em $\mathbf{X}$, o estimador OLS, $\widehat{\boldsymbol{\beta}}$, é o *BLUE* para $\boldsymbol{\beta}$. **TS.6** Normalidade: u_t's independentes de $\mathbf{X}$ e $u_t \sim iid\mathcal{N}(0, \sigma^2)$. **Teor.:** TS.1 a TS.6 (hipóteses CLM) $\Rightarrow$ a) $\widehat{\boldsymbol{\beta}} | \mathbf{X} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$; b) sob as respectivas H_0, t's $\sim t_{(.)}$ e F's $\sim F_{(.,.)}$; c) a construção usual de ICs é válida.

Forma funcional: tudo o que foi estudado para dados seccionais mantém-se válido. Idem para variáveis *dummy*. Exemplos de EViews: pill=year>1962; variáveis desfasadas: z(-1), z(-2), ...

Modelo simples de tendência linear (determinística): $y_t = \alpha_0 + \alpha_1 t + e_t$, com $e_t \sim iid(0, \sigma^2)$, processo *ruído branco*, e $t = 1, 2, \ldots, n$; $\Delta y_t \approx \alpha_1$; EViews: (*generate series*) $t =$ @*trend*$+1$. Modelo de tendência exponencial: $\log(y_t) = \beta_0 + \beta_1 t + e_t$; $\Delta \log(y_t) \approx \frac{y_t - y_{t-1}}{y_{t-1}} \approx \beta_1$. Modelo de tendência quadrática: $y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t$.

Variáveis com tendência na regressão. Modelo adequado: $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + v_t$; os estimadores $\widehat{\beta}_1$ e $\widehat{\beta}_2$ do modelo $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ são enviesados se x_{t1} e/ou x_{t2} "tiverem tendência"; alguns dos resultados obtidos poderão ser espúrios.

Teor. FWL I. As estimativas OLS $\widehat{\beta}_1$ e $\widehat{\beta}_2$ de $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t$ podem obter-se de: i) remoção da tendência de y_t, x_{t1} e x_{t2}; por exemplo, $\ddot{y}_t = y_t - \widehat{y}_t = y_t - (\widehat{\alpha}_0 + \widehat{\alpha}_1 t)$; ii) regressão de $\ddot{y}_t$ sobre $\ddot{x}_{t1}$ e $\ddot{x}_{t2}$. Interpretação dos coeficientes: em termos de variações em torno da tendência. Nota: pode ser útil incluir t numa regressão em que y_t "não tem tendência" mas um regressor tem.

Sazonalidade regular é modelada com *dummies* sazonais. EViews: @seas(.). **Teor. FWL II**: as estimativas OLS $\widehat{\beta}_1$ e $\widehat{\beta}_2$ de $y_t = \alpha_0 + \delta_1 Q_{t1} + \delta_2 Q_{t2} + \delta_3 Q_{t3} + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$ (dados trimestrais) podem obter-se de: i) remoção da sazonalidade de y_t, x_{t1} e x_{t2}; por exemplo, $\ddot{y}_t = y_t - \widehat{y}_t = y_t - (\widehat{\gamma}_0 + \widehat{\gamma}_1 Q_{t1} + \widehat{\gamma}_2 Q_{t2} + \widehat{\gamma}_3 Q_{t3})$; ii) regressão de $\ddot{y}_t$ sobre $\ddot{x}_{t1}$ e $\ddot{x}_{t2}$. Interpretação dos coeficientes: em termos de variações após remoção da sazonalidade (dessazonalização).

Tópicos adicionais sobre a utilização do OLS

O processo estocástico $\{x_t\}$, com $\mathrm{E}(x_t^2) < \infty$ é **estacionário** (em covariância) se: i) $\mathrm{E}(x_t)$ é constante em t; ii) $\mathrm{Var}(x_t)$ é constante em t; iii) $\forall t, h \geq 1$, $\mathrm{Cov}(x_t, x_{t+h})$ só depende de h e não de t. Em termos grosseiros, um processo estacionário $\{x_t\}$ é fracamente dependente se $\lim_{h \to \infty} \mathrm{corr}(x_t, x_{t+h}) = 0$ (assintoticamente não autocorrelacionado).

Processo MA(1): $x_t = e_t + \alpha\, e_{t-1}, e_t \sim iid(0, \sigma_e^2)$. $\mathrm{E}(x_t) = 0, \forall t$; $\mathrm{Var}(x_t) = \sigma_e^2(1+\alpha^2)$; $\mathrm{Cov}(x_t, x_{t+1}) = \alpha\, \sigma_e^2$ e $\mathrm{corr}(x, x_{t+1}) = \frac{\alpha}{1+\alpha^2}$; $\mathrm{Cov}(x_t, x_{t+h})$ =$\mathrm{corr}(x_t, x_{t+h}) = 0, \forall h \geq 2$. Os processos MA são sempre estacionários e fracamente dependentes.

Processo AR(1): $y_t = \rho\, y_{t-1} + e_t$. É estacionário (e fracamente dependente) se $|\rho| < 1$. Nesse caso, $\forall t$, $\mathrm{E}(y_t) = 0, Var(y_t) = \sigma_y^2 = \frac{\sigma_e^2}{1-\rho^2}$, $\mathrm{Cov}(y_t, y_{t+1}) = \rho\, \sigma_y^2$, $\mathrm{corr}(y_t, y_{t+1}) = \rho$, $\mathrm{corr}(y_t, y_{t+h}) = \rho^h, \forall h \geq 0$.

Exemplo de processo estacionário em tendência: $y_t = \alpha_0 + \alpha_1\, t + e_t$. Mas também $y_t = \alpha_0 + \alpha_1\, t + u_t$, com $u_t = e_t + \gamma e_{t-1}$, ou $u_t = \rho\, u_{t-1} + e_t$, com $|\rho| < 1$, etc. . Embora não estacionários, estes processos não levantam problemas na análise de regressão.

Propriedades assintóticas do OLS. TS.1' Linearidade e dependência fraca: $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t)\} : y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k\, x_{tk} + u_t, t = 1, 2, \ldots, n$, e é estacionário e fracamente dependente. **TS.2'** Exogeneidade contemporânea: $\mathrm{E}(u_t|\, x_{t1}, \ldots, x_{tk}) = \mathrm{E}(u_t|\, \mathbf{x}_t) = 0, \forall t$. Exemplos: y_{t-1} não é estritamente exógeno mas é, em princípio, contemporaneamente exógeno; pode existir efeito de *feedback*, desde que não seja contemporâneo. **TS.3'** = TS.3 Não colinearidade perfeita: $r(\mathbf{X}) = k+1$. **Teor.:** TS.1' + TS.2' + TS.3' $\Rightarrow \mathrm{plim}(\widehat{\beta}_j) = \beta_j, j = 0, 1, \ldots, k$.

TS.4' Homoscedasticidade: $\mathrm{Var}(u_t|\, \mathbf{x_t}) = \boldsymbol{\sigma^2}, \forall \mathbf{t}$. **TS.5'** Ausência de autocorrelação: $\mathrm{Cov}(u_t, u_s) = 0, \forall t \neq s$. **Teor.:** TS.1' a TS.5' $\Rightarrow \widehat{\beta}_j \overset{a}{\sim} \mathcal{N}(.,.)$; os usuais erros-padrão (se's), estatísticas-t, $-F$ e LM são válidos assintoticamente. Exemplo: segundo a HEM, com y_t um rendimento (percentual) de um activo financeiro, $\mathrm{E}(y_t|\, y_{t-1}, y_{t-2}, \ldots) = \mathrm{E}(y_t), \forall t$.

Séries temporais **altamente persistentes**. Passeio aleatório sem deriva: $y_t = y_{t-1} + e_t$ (AR(1) com $\rho = 1$, processo de **raiz unitária**); $y_t = y_0 + \sum_{i=1}^t e_i = y_0 +$ tend. estocástica; $\mathrm{Var}(y_t) = t\, \sigma_e^2, \forall t$; $\mathrm{E}(y_{t+h}|\, I_t) = y_t, \forall h \geq 1$; $\mathrm{corr}(y_t, y_{t+h}) = \sqrt{\frac{t}{t+h}}$. Se $y_t = \rho y_{t-1} + e_t$, com $|\rho| < 1$, $\mathrm{E}(y_{t+h}|\, I_t) = \rho^h\, y_t, \forall h \geq 1$.

Passeio aleatório com deriva: $y_t = \alpha_0 + y_{t-1} + e_t$ (também AR(1) com $\rho = 1$); $y_t = y_0 + \alpha_0\, t + \sum_{i=1}^t e_i = y_0 +$ tend. determinística + tend. estocástica; não é estacionário em tendência.

Processo fracamente dependente = $\mathbf{I(0)}$, não é necessário diferenciar nenhuma vez para poder usar na regressão. Processo fortemente dependente = altamente persistente = de raiz unitária = $\mathbf{I(1)}$: necessário diferenciar uma vez para ficar (estacionário e) fracamente dependente (I(0)). Exemplos: passeios aleatórios; mais geralmente: $y_t = (\alpha_0 +)\, y_{t-1} + u_t$, com $u_t \sim I(0)$ qualquer. Nota: diferenciar uma série também reduz numa unidade o grau do polinómio em t que ela possa ter; Ex.: $y_t = \alpha_0 + \alpha_1\, t + v_t \Rightarrow \Delta\, y_t = \alpha_1 + \Delta\, v_t$. Não nos preocuparemos com a possibilidade I(2).

As séries I(0) têm um comportamento de reversão ou regressão para a média que as séries I(1) não têm.

Decisão sobre I(1) ou I(0): i) estimar ρ_1 (ou ρ) em $y_t = \alpha + \rho_1\, y_{t-1} + u_t$; ii) se $\widehat{\rho}_1 > 0.9$ ou $0.8 \Rightarrow$ considerar que $y_t \sim I(1)$. Atenção: se y_t "tem tendência", esta necessita ser removida. Suspeita de regressão espúria "em níveis" $\Rightarrow$ regressão em primeiras diferenças.

Se a igualdade $\mathrm{E}(u_t|\, \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \ldots) = 0$, for satisfeita, o modelo $y_t = \beta_0 + \beta_1\, x_{t1} + \ldots + \beta_k\, x_{tk} + u_t$ diz-se **dinamicamente completo**. A condição anterior é equivalente à condição $\mathrm{E}(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots) = \mathrm{E}(y_t|\, \mathbf{x}_t)$. Se um modelo é dinamicamente completo, então satisfaz TS.5'.

Autocorrelação e heteroscedasticidade

Se $\exists t, s, t \neq s : \mathrm{Cov}(u_t, u_s) \neq 0$, quais as implicações negativas? No modelo $y_t = \beta_0 + \beta_1\, x_t + u_t$, com $\bar{x} = 0$, $u_t = \rho\, u_{t-1} + e_t, |\rho| < 1, e_t \sim iid(0, \sigma^2)$, tem-se

$$\mathrm{Var}(\widehat{\beta}_1 |\, \mathbf{X}) = \frac{\sigma^2}{SST_x} + 2\,\frac{\sigma^2}{SST_x^2}\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^j\, x_t x_{t+j},$$

ou seja, em geral, a precisão do estimador OLS é sobre-avaliada. Daqui resulta que os métodos usuais de inferência passem a ser inválidos, mesmo assintoticamente. Nalguns modelos dinâmicos passa a ter-se $\mathrm{plim}(\widehat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$.

Teste-t para autocorrelação AR(1) com regressores estritamente exógenos. $H_0 : \rho = 0$ em $u_t = \rho\, u_{t-1} + e_t$. Na regressão de $\widehat{u}_t$ sobre $\widehat{u}_{t-1}$, sem constante, $t_\rho = \frac{\widehat{\rho}}{se(\widehat{\rho})} \overset{a}{\sim} \mathcal{N}(0,1)$ sob H_0.

Teste $h - alt$ de Durbin para autocorrelação AR(1) sem exogeneidade estrita dos regressores. Na regressão auxiliar de $\widehat{u}_t$ sobre $x_{t1}, x_{t2}, \ldots, x_{tk}, \widehat{u}_{t-1}$, $t_\rho = \frac{\widehat{\rho}}{se(\widehat{\rho})} \overset{a}{\sim} \mathcal{N}(0,1)$ sob H_0.

Testes de Breusch-Godfrey para autocorrelações de ordens mais elevadas, sem exogeneidade estrita dos regressores. $H_0 : \rho_1 = \rho_2 = \ldots \rho_q = 0$ em $u_t = \rho_1\, u_{t-1} + \rho_2\, u_{t-2} + \ldots + \rho_q\, u_{t-q} + e_t$. Da regressão auxiliar de $\widehat{u}_t$ sobre $x_{t1}, x_{t2}, \ldots, x_{tk}, \widehat{u}_{t-1}, \widehat{u}_{t-2}, \ldots, \widehat{u}_{t-q}$, $BG(q) = LM(q) = (n-q)\, R^2_{\widehat{u}} \overset{a}{\sim} \chi^2_{(q)}$ sob H_0. Versão alternativa: estatística-F para testar a significância conjunta de $\widehat{u}_{t-1}, \widehat{u}_{t-2}, \ldots, \widehat{u}_{t-q}$.

A diferenciação do modelo pode ser útil para eliminar os sintomas de autocorrelação.

As consequências da heteroscedasticidade são semelhantes às de autocorrelação dos erros. Testes de heteroscedasticidade: de White e de Breusch-Pagan, com $\widehat{u}_t^2$ como variável dependente das regressões auxiliares de teste.

"Tópicos Avançados"

Testes de Dickey-Fuller (DF) de raiz unitária. Caso base: $y_t = \alpha + \rho\, y_{t-1} + e_t$. $H_0 : \rho = 1$ *vs.* $H_1 : \rho < 1 \;\Leftrightarrow\; H_0 : y_t \sim I(1)$ *vs.* $H_1 : y_t \sim I(0) \Leftrightarrow H_0 : \theta = 0$ *vs.* $H_1 : \theta < 0$ em $\Delta\, y_t = \alpha + \theta\, y_{t-1} + e_t$, $\theta = \rho - 1$. DF $= t_\theta = \widehat{\theta}/se(\widehat{\theta}) \overset{a}{\sim} DF$ sob H_0; $\mathrm{vc}_{0.05} = -2.86$. Complicações: a) o AR(1) pode não ter dinâmica suficiente; b) a série pode ter tendência. Regressão auxiliar no caso mais geral:

$$\Delta\, y_t = \alpha + \beta\, t + \theta\, y_{t-1} + \sum_{i=1}^{p}\gamma_i\, \Delta y_{t-i} + e_t, \quad e_t \sim iid(0, \sigma^2).$$

$H_0 : \theta = 0$ *vs.* $H_1 : \theta < 0$, (A)DF$= t_\theta = \widehat{\theta}/se(\widehat{\theta}) \overset{a}{\sim} DF$ sob H_0, com a distribuição a deslocar-se para a esquerda(direita) com a inclusão(exclusão) de regressores determinísticos.

Escolha de p: com método "*general-to-specific*" (GTS) $t - sig$, com testes com nível de 10%. **i.** Fixar p_{MAX}; **ii.** estimar regressão com $p = p_{MAX}$ e testar significância de γ_p com teste assintótico com $\alpha = 0.10$. **iii.** Se $\widehat{\gamma}_p$ for estatisticamente significativo, fazer $\widehat{p} = p_{MAX}$ e $t_\theta = ADF(\widehat{p})$; no caso contrário, repetir ii. e iii. com $p = p_{MAX} - 1$. Paragem: quando $\widehat{\gamma}_p$ significativo (a 10%) ou $\widehat{p} = 0$.

Que regressores determinísticos? Em excesso => testes perdem potência; a menos => testes perdem potência. Critérios: i) razoabilidade económica; a série tem relação com a de produtividade, ou com a da população ou com a de preços? ii) análise gráfica.

Regressões espúrias com processos I(1) independentes (passeios aleatórios, para simplificar). Sejam $x_t = x_{t-1} + a_t, a_t \sim iid(0, \sigma_a^2)$ e $y = y_{t-1} + e_t, e_t \sim iid(0, \sigma_e^2)$, com a_t e e_t independentes. Na regressão $y_t = \alpha + \beta\, x_t + u_t$, tem-se: a) $\mathrm{plim}(\widehat{\beta}) \neq \beta = 0$; b) $|t_\beta| \to +\infty$ (i.e., o resultado $t_\beta \overset{a}{\sim} \mathcal{N}(0,1)$ sob $H_0 : \beta = 0$ não é válido); c) $R^2 \to$ variável aleatória.

Propriedades de combinações lineares de séries I(0) e I(1). Com a e b constantes não nulas: 1) se $y_t \sim I(0)[I(1)] \Rightarrow (a + b\,y_t) \sim I(0)[I(1)]$; 2) se $y_t \sim I(0)$ e $x_t \sim I(0) \Rightarrow (a\,y_t + b\,x_t) \sim I(0)$; 3) se $y_t \sim I(1)$ e $x_t \sim I(0) \Rightarrow (a\,y_t + b\,x_t) \sim I(1)$; 4) *em geral*, se $y_t \sim I(1)$ e $x_t \sim I(1) \Rightarrow (a\,y_t + b\,x_t) \sim I(1)$. Todavia, se $\exists \beta \neq 0 : y_t - \beta\,x_t = u_t \sim I(0)$ (com média nula), y_t e x_t dizem-se cointegradas, $(y_t, x_t) \sim CI(1,1)$, e β é o parâmetro de cointegração. Neste caso, na regressão $y_t = \alpha + \beta\,x_t + u_t$, $\text{plim}(\widehat{\beta}) = \beta$.

Testes de cointegração. A) se β for conhecido: testes (A)DF sobre u_t; B) se β for desconhecido: testes de Engle-Granger (EG) sobre $\widehat{u}_t$. Nota: se y_t e/ou x_t tem(têm) deriva, deve incluir-se um termo em t na potencial equação de cointegração.

Teorema de representação de Granger (versão *"light"*). Sejam $y_t \sim I(1)$ e $x_t \sim I(1)$, mas $y_t(-\alpha-\gamma\,t)-\beta\,x_t = u_t \sim I(0)$ (com média nula), isto é, $(y_t, x_t) \sim CI(1,1)$; então, existe representação em modelo de correcção de erros (MCE). Por exemplo:

$$\Delta\,y_t = \alpha_0 + \alpha_1\,\Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1\,\Delta x_{t-1} + \delta\,u_{t-1} + e_t, \quad \text{com} \;\; \delta < 0.$$

A recíproca também é, em geral, verdadeira: se existe representação MCE para $y_t \sim I(1)$ e $x_t \sim I(1)$, então $(y_t, x_t) \sim CI(1,1)$.

Estimação do MCE. **A)** Se β conhecido: imediata. **B)** Se β desconhecido: método de Engle-Granger em 2 passos. i) estimação OLS de β e de u_t; ii) emprego de $\widehat{u}_{t-1}$ no MCE e estimação deste com o OLS.