Econometria II, Lic. em Economia

Luís Filipe Martins	Dpt. de Métodos Quantitativos	ISCTE - EG
luis.martins@iscte.pt	http://home.iscte.pt/~lfsm	Lisboa, 20/09/2005

INTRODUÇÃO À REGRESSÃO COM TIME SERIES E AUTOCORRELAÇÃO DOS ERROS

Wooldridge §10, 11, 12

0.1 Alguns Conceitos Introdutórios em Time Series

Para dados seccionais é (normalmente) consensual admitir-se que a amostra é aleatória no sentido em que não existe depêndencia entre as observações. O valor registado para o agente i não é influenciado pelo do agente $j, i \neq j$. Veja-se as hipóteses $M1$ e $M4.2$.

Na presença de time series - dados de natureza cronológica - esta propriedade raramente se verifica. Sendo uma sequência de valores que está indexada por t, uma determinada time series apresentada correlações significativas em momentos próximos na amostra. Tomemos os exemplos da taxa de desemprego, inflação, índice PSI-20, entre muitas outras time series. Para qualquer destes casos, a obsevação num periodo t, diga-se 2005, depende (está correlacionada) dos valores registados no passado $t-1, t-2, ...$ e de uma forma mais acentuada em relação aos mais recentes. Alguns conceitos são apresentados em seguida.

Definition 1 *Uma time series (ou processo estocástico) em $\Re$ é uma função real $Y(t, \omega)$ definida em $\Im \times \Omega$ em que para cada t fixo, $Y(t, \omega)$ é uma variável aleatória no espaço de probabilidade* (Ω, A, P).

A função $Y(t, \omega)$ é normalmente representada por $Y_t(\omega)$ ou Y_t e pode ser interpretada como uma sequência $\{Y_t : t \in \Im\}$ de variáveis aleatórias onde $\Im$ é um conjunto de indices, normalmente $\Im = \{0, 1, ...\}$ ou $\Im = \{..., -1, 0, 1, ...\}$. Se $\Im$ contém apenas um elemento, então Y_t é simplesmente uma única variável aleatória. Para $\omega \in \Omega$ fixo, $Y(t, \omega)$ é uma função real de t (sequência de números reais). Esta função é denominada por realização. Conceptualmente, o gráfico de uma variável time series corresponde ao gráfico de $Y(t, \omega), \omega$ fixo. A colecção de todas as possíveis realizações é denominada por *ensemble* de realizações.

Em termos de notação, não confundir os conceitos de Y, Ω com os apresentados no contexto do MRLM!

Definition 2 *A função distribuição conjunta de um conjunto finito de variáveis aleatórias* $\{Y_{t_1}, Y_{t_2}, ..., Y_{t_m}\}$ *da colecção* $\{Y_t : t \in T\}$ *é definido como*

$$F_{Y_{t_1},Y_{t_2},...,Y_{t_m}}\left(y_{t_1}, y_{t_2}, ..., y_{t_m}\right) = P\left\{\omega : Y(t_1, \omega) \leq y_{t_1}, ..., Y(t_m, \omega) \leq y_{t_m}\right\}.$$

Definition 3 *O processo estocástico (time series)* Y_t *é estritamente estacionário se para arbitrários* $t_1 < ... < t_m$*, a função distribuição conjunta de* $\{Y_{t_1}, ..., Y_{t_m}\}$ *é igual à função distribuição conjunta de* $\{Y_{t_1+h}, ..., Y_{t_m+h}\}$ *com* $h \geq 1$.

Como o processo estritamente estacionário tem identica função distribuição conjunta para periodos adjacentes, isso implica que também tem momentos iguais para periodos adjacentes, nomeadamente correlações. Veja-se o caso de $m = 2$. Um conceito mais lato é o de processo estacionário em covariância em que os primeiros e segundos momentos são constantes para todo t e as covariâncias entre duas observações dependem apenas do desfasamento e não de t.

Definition 4 *O processo estocástico* Y_t *é estacionário em covariância se para todo* t *e* $l \geq 0, E\left(Y_t\right) = \mu$ *e* $Cov\left(Y_t, Y_{t+l}\right) = \Gamma\left(l\right).$

Em certas variáveis macroeconómicas e financeiras que possuem a propriedade de estacionaridade e em que existe correlação entre periodos diferentes, é fundamental garantir que essa dependência vá-se dissipando à medida que as observações estão mais distantes no tempo. Neste sentido, definimos da seguinte forma um processo estacionário cuja memória vai desaparecendo à medida que o desfasamento temporal aumenta indefinidadmente (processo assimptóticamente não correlacionado/independente).

Definition 5 *O processo estocástico estacionário* Y_t *possui memória que se dissipa se* $\Gamma\left(l\right) \to 0$ *quando* $l \to \infty$.

Esta é apenas uma forma de definir um processo dito fracamente dependente (*weakly dependent*) em que se caracteriza por independência assimptótica. Existem outras definições para processos *weakly dependent* nas quais existem (U)LLN e (F)CLT que podem ser invocados (por exemplo, $\alpha-$mixing e $m-$dependent).

- (Plots ...)

Voltemos ao MRLM. Como, ao contrário dos dados seccionais, a ordem da observação é importante no estudo de variáveis time series, utilizamos a seguinte notação para o MRLM (incluindo o termo independente):

$$\begin{aligned}
y_t &= \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t = x_t^{'} \beta + u_t, t = 1, ..., T \qquad (1) \\
y &= X\beta + u \\
x_{t_{k\times 1}} &= (1, x_{t2}, ..., x_{tk})^{'}, t = 1, ..., T; \beta_{k\times 1} = (\beta_1, \beta_2, ..., \beta_k)^{'}; \\
y_{T\times 1} &= \begin{pmatrix} y_1 \\ ... \\ y_T \end{pmatrix}; X_{T\times k} = \begin{pmatrix} x_1^{'} \\ ... \\ x_T^{'} \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & ... & x_{1k} \\ ... & ... & & ... \\ 1 & x_{T2} & ... & x_{Tk} \end{pmatrix}; \\
u_{T\times 1} &= (u_1, ..., u_T)^{'}.
\end{aligned}$$

Em time series, os regressores podem relacionar-se contemporaneamente com a variável dependente ou de uma forma desfasada. É a teoria económica ou o próprio investigador que estabelece o tipo de especificação no modelo. Entre muitas outras formas funcionais (e nem todas serão alvo de estudo na disciplina) salientamos as seguintes:

Example 1 $y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t, t = 1, ..., T$

Example 2 $y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + ... + \beta_{l+2} x_{t-l} + u_t, t = 1, ..., T$

Example 3 $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + ... + \beta_{l+2} y_{t-l} + u_t, t = 1, ..., T$

Example 4 $y_t = \beta_1 + \beta_2 y_{t-1} + u_t, t = 1, ..., T$

Example 5 $y_t = \beta_1 + \beta_2 y_{t-4} + \beta_3 y_{t-8} + u_t, t = 1, ..., T$

Example 6 $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \beta_4 x_{t-1} + u_t, t = 1, ..., T$

Example 7 $y_t = \beta_1 + \beta_2 t + u_t, t = 1, ..., T$

Example 8 $y_t = u_t, t = 1, ..., T$

Example 9 $y_t = \beta_1 + \beta_2 u_{t-1} + u_t, t = 1, ..., T$

Example 10 $y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 u_{t-1} + u_t, t = 1, ..., T$

Devido à sua importância, estudamos em seguida as propriedades dos processos nos exemplos 8, 9 e 4. Denomine-se ε_t por ruído branco (*white noise*) o processo em que $E(\varepsilon_t) = 0, V(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$ e $\Gamma(l) = 0$ para todo $l : \varepsilon_t \sim wn\left(0, \sigma_\varepsilon^2\right)$. É raro encontrar-se uma variável económica que seja ruído branco - *i.i.d.* - mas o seu estudo justifica-se pela sua simplicidade e base de construção de modelos mais sofisticados.

O processo médias móveis de primeira ordem, $MA(1)$, y_t satisfaz

$$y_t = \beta_1 + \beta_2 \varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim wn\left(0, \sigma_\varepsilon^2\right), t = 1, ..., T. \qquad (2)$$

É fácil de concluir que o processo $MA(1)$ é estacionário e de memória que se dissipa, para qualquer valor real de β_1, β_2, pois

$$E(y_t) = \beta_1; V(y_t) = \left(\beta_2^2 + 1\right)\sigma_\varepsilon^2; \Gamma(l) = \begin{cases} \beta_2 \sigma_\varepsilon^2, l = 1 \\ 0, l > 1 \end{cases}. \tag{3}$$

O processo autoregressivo de ordem um, $AR(1)$, sem termo independente y_t satisfaz

$$y_t = \rho y_{t-1} + \varepsilon_t, \varepsilon_t \sim wn\left(0, \sigma_\varepsilon^2\right), t = 1, ..., T. \tag{4}$$

$$y_t = \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} = \rho^t y_0 + \sum_{i=1}^{t} \rho^{t-i} \varepsilon_i. \tag{5}$$

Se considerarmos que o processo teve inicio no passado mais distante, $\Im_t = \{-\infty, ..., t-2, t-1, t\}$,

$$y_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j} = \sum_{i=1}^{\infty} \rho^{t-i} \varepsilon_i. \tag{6}$$

Para que este processo seja estacionário e de memória que se dissipa, $|\rho| < 1$. Neste caso,

$$E(y_t) = 0; V(y_t) = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}; \Gamma(l) = \rho^l V(y_t); Cor(Y_t, Y_{t+l}) = \rho^l. \tag{7}$$

- (Demonstrações, plots)

Por simplicidade, neste capitulo vamos estudar o MRLM em que não existe qualquer variável desfasada (y nem x) no conjunto dos regressores mas onde introduzimos autocorrelação nos erros (e consequentemente em y_t). O objectivo é por isso similar ao do capítulo da heterocedasticidade: estudar o MRLM quando a hipótese $M4.2$ não está presente (além de existir dependência na time series y_t).

0.2 Autocorrelação nos Erros e Consequências para o OLS

Em modelos time series onde os erros estão autocorrelacionados, a hipótese clássica $M4.2$ não é válida. Devido a efeitos de sazonalidade, persistência ou inércia, os dados time series e correspondentes erros estão autocorrelacionados. Este fenómeno também pode ser consequência de uma má especificação (por exemplo, a omissão de y_{t-1} como regressor ou lags nos $x's$). Neste sentido, consideramos um esquema de autocorrelação nos erros tal que

$$Cov(u_t, u_s|X) = Cov(u_t, u_s|x_1, ..., x_T) \neq 0, t, s = 1, ..., T, \text{ para algum } t \neq s. \tag{8}$$

Identificamos esta hipótese como $TS4.2$. Convém salientar algumas observações. Em primeiro lugar, a autocorrelação não tem necessáriamente de ser para todo $t \neq s$ (veja-se o caso do MA(1)) nem tem de ser uma função dos regressores, com valores diferente de zero. Em segundo lugar, a covariância de $u_t, u_s, t \neq s$ condicionada aos regressores não é apenas contemporânea

(x_t, x_s) mas também com o passado e presente, isto é, é estrita. Claro que este conceito implica covariância condicional contemporânea

$$Cov(u_t, u_s|X) = Cov(u_t, u_s|x_t, x_s) \neq 0, t, s = 1, ..., T, \text{ para algum } t \neq s. \tag{9}$$

Na prática, temos de definir um esquema (especificação) de autocorrelação para os erros: Em geral, onde

$$\Omega_{T\times T} = V(u|X) = E(\mathrm{u}\,\mathrm{u}^{'}|X) = \begin{pmatrix} E(u_1^2|X) & ... & E(u_1 u_T|X) \\ ... & & ... \\ E(u_T u_1|X) & ... & E(u_T^2|X) \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & a_{12} & ... & a_{1T} \\ & 1 & ... & a_{2T} \\ ... & ... & ... & ... \\ & & ... & 1 \end{pmatrix}, \tag{10}$$

é uma matriz simétrica, existem no máximo $\frac{T^2-T}{2}+1$ parâmetros para estimar. Esta poderá ser uma tarefa difícil quando dispomos de T observações (para além de termos de estimar β)!

Suponhamos que u_t é gerado por um $AR(1)$ estacionário, $|\rho| < 1$. A escolha deste processo é justificada pela sua simplicidade mas também porque o $AR(1)$ possui um esquema de autocorrelação que tem muito fundamento na teoria económica ou empírica - $Cor\,(u_t, u_{t+l}) = \rho^l$ tem decaimento exponencial com l. Se para além de seguir uma lei $AR(1)$, u_t fôr homocedástico (ver $M4.1$), então $TS4$ é definida à custa da matriz simétrica positiva definida

$$\begin{aligned} \Omega_{T\times T} &= V(u|X) = E(\mathrm{u}\,\mathrm{u}^{'}|X) = \begin{pmatrix} E(u_1^2|X) & ... & E(u_1 u_T|X) \\ ... & & ... \\ E(u_T u_1|X) & ... & E(u_T^2|X) \end{pmatrix} \qquad (11) \\ &= \begin{pmatrix} \sigma_u^2 & \rho\sigma_u^2 & ... & \rho^{T-1}\sigma_u^2 \\ \rho\sigma_u^2 & \sigma_u^2 & ... & \rho^{T-2}\sigma_u^2 \\ ... & ... & ... & ... \\ \rho^{T-1}\sigma_u^2 & \rho^{T-2}\sigma_u^2 & ... & \sigma_u^2 \end{pmatrix} = \sigma_u^2 \begin{pmatrix} 1 & \rho & ... & \rho^{T-1} \\ \rho & 1 & ... & \rho^{T-2} \\ ... & ... & ... & ... \\ \rho^{T-1} & \rho^{T-2} & ... & 1 \end{pmatrix} \neq \sigma_u^2 I_T, \end{aligned}$$

onde $\rho \neq 0, \sigma_u^2 = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}$ e $\Gamma(l) = \rho^l \sigma_u^2$. Implicitamente, assumimos que exogeneidade estrita $TS2$

$$E(u_t|X) = E(u_t|x_1, ..., x_T) = 0, t = 1, ..., T \Leftrightarrow E(u|X) = 0_{T\times 1} \tag{12}$$

se verifica para que $Cov(u_t, u_s|X) = E(u_t u_s|X)$. Claramente, exogeneidade estrita implica o conceito menos restrictivo de exogeneidade contemporânea[1]

$$E(u_t|X) = E(u_t|x_t) = E(u_t|x_{t2}, ..., x_{tk}) = 0, t = 1, ..., T \tag{13}$$

[1] Consequentemente, **não** contemporânea exogeneidade implica **não** exogeneidade estrita!

em que algum regressor x_{tj} pode estar correlacionado com passado ou futuro de $u_t, u_{t\pm l}$. Um exemplo desta situação ocorre no modelo dinâmico

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t; E(u_t|y_{t-1}) = 0, t = 2, ..., T \tag{14}$$

em que y_{t-1} tem o papel de regressor $(x_t \equiv y_{t-1})$. Aqui, y_{t-1} e u_{t-1} estão autocorrelacionados pois como $y_{t-1} = \beta_1 + \beta_2 y_{t-2} + u_{t-1}$,

$$E(u_t|x_t) \equiv E(u_t|y_{t-1}) = 0 \textbf{mas } E(u_{t-1}|x_t) \equiv E(u_{t-1}|y_{t-1}) \neq 0. \tag{15}$$

Em termos da hipótese $TS1$, e para que seja possível invocar (U)LLN e (F)CLT, consideramos que no MRLM

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t = x_t^{'}\beta + u_t, t = 1, ..., T, \tag{16}$$

$\{(x_{t1}, ..., x_{tk}, y_t) : t = 1, ..., T\}$ é um processo estacionário e *weakly dependent*. Finalmente, mantemos as hipóteses $M3$ e $M5$ às quais denominaremos $TS3$ e $TS5$: não existe perfeita colinearidade, $rank(X) = k$, e $u|X \sim Normal$.

Para o modelo 16, sob as hipóteses $TS1$ de uma amostra estacionária e *weakly dependent;* $TS2$ de exogeneidade estrita entre erros e regressores e $TS3$ de não perfeita colinearidade, o estimador OLS

$$\begin{aligned}
\widehat{\beta}_{k\times 1} &= \left(X^{'}X\right)^{-1}X^{'}y = \left(\sum_{t=1}^{T} x_t x_t^{'}\right)^{-1}\sum_{t=1}^{T} x_t y_t && (17)\\
&= \begin{pmatrix} T & \sum_{t=1}^{T} x_{t2} & ... & \sum_{t=1}^{T} x_{tk} \\ & \sum_{t=1}^{T} x_{t2}^2 & ... & \sum_{t=1}^{T} x_{t2}x_{tk} \\ ... & ... & & ... \\ & & ... & \sum_{t=1}^{T} x_{tk}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_{t2}y_t \\ ... \\ \sum_{t=1}^{T} x_{tk}y_t \end{pmatrix} && (18)\\
&= \beta + \left(X^{'}X\right)^{-1}X^{'}u = \beta + \left(\sum_{t=1}^{T} x_t x_t^{'}\right)^{-1}\sum_{t=1}^{T} x_t u_t && (19)
\end{aligned}$$

é centrado e, consequentemente, consistente, $E\left(\widehat{\beta}|X\right) = \beta$ e $p\lim_{T\to\infty}\widehat{\beta} = \beta$. Quando a exogeneidade estrita é substituida por exogeneidade contemporânea (menos restritiva) o estimador é consistente (mas não necessariamente centrado)[2]. De salientar mais uma vez que neste capitulo, apesar de algumas referências, **não** estamos a estudar MRLM's em que os regressores incluem variáveis desfasadas ($x's, y's$ ou ambos)[3]. Esse será alvo de estudo num outro capitulo.

Similar ao caso da heterocedasticidade, quando os erros estão autocorrelacionados, o OLS não é BLUE e os testes habituais não são válidos, mesmo assimptóticamente. No MRLS e sob

[2]Veja-se na revisão das propriedades assimptóticas a relação entre $Cov(u, X)$ e $E(u|X)$.

[3]No modelo dinâmico $y_t = \beta_1 + \beta_2 y_{t-1} + u_t$, com $u_t \sim AR(1)$, o OLS é inconsistente pois $Cov(y_{t-1}, u_t) \neq 0$.

u_t gerado por um $AR(1)$ estacionário, $|\rho| < 1$,

$$\begin{aligned}
V\left(\widehat{\beta}_2|\mathbf{x}\right) &= \frac{V\left(\sum_{t=1}^{T} u_t\left(x_t - \overline{x}\right)|\mathbf{x}\right)}{\left[\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2\right]^2} \qquad (20)\\
&= \frac{\sum_{t=1}^{T} V\left(u_t\left(x_t - \overline{x}\right)|\mathbf{x}\right) + 2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} Cov\left(u_t\left(x_t - \overline{x}\right), u_{t+j}\left(x_{t+j} - \overline{x}\right)|\mathbf{x}\right)}{\left[\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2\right]^2}\\
&= \frac{\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2 V\left(u_t|\mathbf{x}\right) + 2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\left(x_t - \overline{x}\right)\left(x_{t+j} - \overline{x}\right) E\left(u_t, u_{t+j}|\mathbf{x}\right)}{\left[\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2\right]^2}\\
&= \frac{\sigma_u^2\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2 + 2\sigma_u^2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\left(x_t - \overline{x}\right)\left(x_{t+j} - \overline{x}\right)\rho^j}{\left[\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2\right]^2}\\
&= \frac{\sigma_u^2}{\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2} + \frac{2\sigma_u^2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\rho^j\left(x_t - \overline{x}\right)\left(x_{t+j} - \overline{x}\right)}{\left[\sum_{t=1}^{T}\left(x_t - \overline{x}\right)^2\right]^2},
\end{aligned}$$

onde $\sigma_u^2 = \frac{\sigma_\varepsilon^2}{(1-\rho^2)}$, que apenas coincide com a expressão standard $\frac{\sigma_u^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$ se $\rho = 0$, ou seja, $E\left(u_t, u_{t+j}|\mathbf{x}\right) = 0$ para todo $t, j > 0$ - não autocorrelação (white noise)! Portanto, o estimador usual para $V\left(\widehat{\beta}_2|\mathbf{x}\right)$ é enviesado sempre que $\rho \neq 0$. Se $\rho > 0$ e x_t está positivamente autocorrelacionado com o tempo então $V\left(\widehat{\beta}_2|\mathbf{x}\right) > \frac{\sigma_u^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$ o que significa que a expressão habitual da variância do estimador OLS para o declive está sub-avaliada em relação à verdadeira. Consequentemente, o estimador não está a ser estimado com tanta precisão quanto pensamos e os rácios-t estão sobre-avaliados e portanto a tendência é de aceitar regressores que podem não ser estatisticamente significativos. Inferência com testes F e construção de intervalos de confiança também são erróneos. Outra explicação para o enviesamento de do estimador para $V\left(\widehat{\beta}_2|\mathbf{x}\right)$ sob autocorrelação é o facto de $\widehat{\sigma}_u^2$ ser enviesado, $E\left(\widehat{\sigma}_u^2\right) \neq \sigma_u^2$.

No MRLM,

$$\Sigma_{k\times k} = V\left(\widehat{\beta}|X\right) = \left(X'X\right)^{-1} X'\Omega X\left(X'X\right)^{-1} \qquad (21)$$

que, em geral, é diferente de $\sigma_u^2\left(X'X\right)^{-1}$ porque $\Omega \neq \sigma_u^2 I_T$ e onde

$$p\lim_{n\to\infty}\left(\frac{1}{n}X'\Omega X\right) = D,$$

uma determinada matriz positiva definida.

Como vimos no capitulo da heterocedasticidade, na prática, temos de primeiro testar a hipótese de que os erros não estão autocorrelacionados. Se não aceitarmos essa premissa, então devemos utilizar uma duas técnicas - estimação robusta de $V(\widehat{\beta}_{OLS})$ ou estimação do modelo por (F)GLS. Estes serão os tópicos a desenvolver em seguida.

0.3 Testes para a sua Detecção

Nesta secção, apresentamos alguns testes à autocorrelação dos erros u_t no MRLM (16). Numa fase inicial, vamos ignorar modelos dinâmicos do tipo (14) em que a hipótese de exogeneidade estrita não se verifica[4]. A análise gráfica de $\widehat{u}_t$ ou $\widehat{u}_t/\sigma_{\widehat{u}}$ ao longo de t ou o plot $(\widehat{u}_t, \widehat{u}_{t-1})$ podem ser uma primeira abordagem de natureza intuitiva à autocorrelação dos erros.

Testar a hipótese de erros u_t não autocorrelacionados sujeitos à lei AR(1),

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{22}$$

em que $|\rho| < 1$, ε_t é um processo *white noise* $wn\left(0, \sigma_\varepsilon^2\right)$, e $E(\varepsilon_t | u_{t-1}, u_{t-2}, ...) = 0$, no MRLM (16) com exogeneidade estrita (12) $E(u_t|X) = 0, t = 1, ..., T$, é equivalente ao ensaio $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. Nestas condições, o rácio-t no modelo (22) é uma estatística natural a ser implementada. Como os erros u_t não são observáveis, o teste t é aplicado ao modelo

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + e_t, \tag{23}$$

onde $\widehat{u}_t$ são os (observáveis) resíduos OLS e

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}. \tag{24}$$

Apesar de $\widehat{u}_t$ depender de $\widehat{\beta}$, sob a hipótese de exogeneidade estrita pode-se demostrar que o rácio-t mantém as propriedades assimptóticas usuais.

O teste de Durbin-Watson é um procedimento alternativo. A estatística de teste é

$$DW = \frac{\sum_{t=2}^{T} \left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2}{\sum_{t=1}^{T} \widehat{u}_t^2} = \frac{\sum_{t=2}^{T} \triangle \widehat{u}_t^2}{SSR} \approx 2\left(1 - \widehat{\rho}\right), \tag{25}$$

e, sob as hipóteses clássicas do MRLM[5], a sua distribuição condicional a X não é standard e depende de T, k (com ou sem termo independente). A distribuição admite uma região do seu domínio em que o teste é inconclusivo. Esta é delimitada por d_L e d_U, valores que têm de ser consultados numa tabela. A estatística DW varia entre 0 e 4 quando $\widehat{\rho} \in (-1, 1)$. Sob $H_0 : \rho = 0, DW \approx 2$; sob autocorrelação positiva $H_1 : \rho > 0, DW \in (0, 2)$ e $DW \approx 0$ para $\widehat{\rho} \approx 1$; sob autocorrelação negativa $H_1 : \rho < 0, DW \in (2, 4)$ e $DW \approx 4$ para $\widehat{\rho} \approx -1$. Portanto, no ensaio $H_0 : \rho = 0, H_1 : \rho > 0$, não se aceita H_0 para $0 < DW < d_L$; nada se conclui para $d_L < DW < d_U$; e aceita-se H_0 para $d_U < DW < 2$. No ensaio $H_0 : \rho = 0, H_1 : \rho < 0$, aceita-se H_0 para $2 < DW < 4 - d_U$; nada se conclui para $4 - d_U < DW < 4 - d_L$; e não se aceita H_0 para $4 - d_L < DW < 4$. Quando nada se conclui para o caso de AR(1) com o teste DW, há quem utilize o teste AR(1) assimptótico, $\sqrt{T}\widehat{\rho} \xrightarrow{d} N\left(0, 1\right)$ sob $H_0 : \rho = 0$. Este teste, no entanto, é menos potente do que o DW.

[4] Exogeneidade estrita está normalmente associada a regressores deterministicos.

[5] Sob $H_0 : \rho = 0$ não há autocorrelação.

Os testes do rácio-t e da DW não devem ser aplicados em modelos cuja hipótese de exogeneidade estrita não se verifica (por exemplo, os dinâmicos do tipo (14)) porque as suas distribuições não são válidas, mesmo assimptóticamente. Para modelos dinâmicos de exogeneidade contemporanea, é frequente utilizar-se os testes $h-$Durbin e $h-alt$ quando a amostra é de grande dimensão (as distribuições são assimptóticas). A estatística $h-$Durbin tem a forma

$$h = \widehat{\rho}\sqrt{\frac{T}{1 - T.V\left(\widehat{\beta}_2\right)}}, \tag{26}$$

onde $\widehat{\beta}_2$ é o estimador do coeficiente associado ao regressor y_{t-1}, e sob a hipótese nula de $\rho = 0$, é assimptóticamente distribuido $N(0,1)$. Como conhecemos a relação entre $\widehat{\rho}$ e $DW, (25)$, e a DW é obtida na maior parte dos programas econométricos, podemos obter a estatística $h - alt$ sem ser necessário estimar $\widehat{\rho}$:

$$h - alt = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{T}{1 - T.V\left(\widehat{\beta}_2\right)}}. \tag{27}$$

Como foi dito anteriormente, $h \xrightarrow{d} N(0,1)$ e $h - alt \xrightarrow{d} N(0,1)$.

Para modelos mais gerais de exogeneidade contemporanea (ver (16) e (13)), pode-se usar o racio-t do coeficiente ρ da seguinte regressão auxiliar:

$$\widehat{u}_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \rho \widehat{u}_{t-1} + e_t, t = 2, ..., T, \tag{28}$$

onde $\widehat{u}_t$ são os resíduos OLS do modelo (16). A distribuição para grandes amostras é a usual (t) porque a inclusão de $x_{t2}, ..., x_{tk}$ permite que estes sejam correlacionados com u_{t-1} como é próprio da exogeneidade contemporânea - não é necessário impôr exogeneidade estrita (compare este método com (23))

Para finalizar esta secção enunciamos testes à autocorrelação dos erros de ordem superior (não restrita a $AR(1)$). Para erros $AR(p)$, e sob homocedasticidade, uma estatística é a F (significância conjunta de $\widehat{u}_{t-1}, ..., \widehat{u}_{t-p}$) da regressão auxiliar[6]

$$\widehat{u}_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \rho_1 \widehat{u}_{t-1} + ... + \rho_p \widehat{u}_{t-p} + e_t, t = p + 1, ..., T. \tag{29}$$

Alternativamente pode-se usar a estatística de Breusch-Godfrey $(T-p) R^2$, onde o R^2 é da regressão anterior, e que assimptóticante é distribuída como χ^2_p. O teste Box-Pierce e Lyung-Box são procedimentos alternativos aos anteriores.

0.4 Estimação Robusta de $V(\widehat{\beta}_{OLS})$

Admitamos que o MRLM tem erros autocorrelacionados. À semelhança do que foi discutido no capítulo da heterocedasticidade, uma das soluções passa por manter o estimador OLS (consistente) mas corrigir as expressões para $V(\widehat{\beta}_{OLS})$ de forma a tornar a inferência válida. A

[6] Pode-se omitir os $x's$ sob exogeneidade estrita.

correção faz-se por escalar os desvios padrões do estimador devido a autocorrelação. Neste tipo de procedimento não é necessário impôr exogeneidade estrita e a forma de autocorrelação pode ser muito diversa.

Se assumirmos *(i)* MRLS (um único regressor); *(ii)* homocedasticidade dos erros; e *(iii)* erros AR(1), então apenas temos que substituir σ_u^2 e ρ por $\widehat{\sigma}_u^2$ e $\widehat{\rho}$ na expressão (20) . A raíz quadrada de $\widehat{V\left(\widehat{\beta}_2|\mathbf{x}\right)}$ denomina-se desvio padrão de $\widehat{\beta}_j$ *robusto* a autocorrelação. Para o MRLM (e homocedasticidade e erros AR(1)), substitui-se σ_u^2 e ρ por $\widehat{\sigma}_u^2$ e $\widehat{\rho}$ em $\Omega = \sigma_u^2 \begin{pmatrix} 1 & \rho & ... & \rho^{T-1} \\ \rho & 1 & ... & \rho^{T-2} \\ ... & ... & ... & ... \\ \rho^{T-1} & \rho^{T-2} & ... & 1 \end{pmatrix}$.
Mantendo a hipótese de homocedasticidade, se a especificação da autocorrelação dos erros fosse conhecida mas distinta do AR(1) (por exemplo, MA(1)) então as expressões para $V\left(\widehat{\beta}_2|\mathbf{x}\right)$ e Ω seriam outras mas o desvio padrão de $\widehat{\beta}_j$ *robusto* a autocorrelação seria obtido de uma forma análoga: substituir os parametros em $V\left(\widehat{\beta}_2|\mathbf{x}\right)$ e Ω pelas suas estimativas.

Consideremos o MRLM devido à sua generalidade relativamente ao MRLS. A estimação robusta de $V(\widehat{\beta}_{OLS})$ a heterocedasticiade e uma forma geral de autocorrelação (não apenas AR(1)) será estudada mais adiante na secção "Estimador (F)GLS e Estimação HAC de $V(\widehat{\beta}_{OLS})$". Wooldridge (2003) propõe um método para se obter o desvio padrão de $\widehat{\beta}_j$ *robusto* a autocorrelação apenas (não tão geral como a HAC que também admite heterocedasticidade). Ver Wooldridge (2003) para detalhes. Para $j = 2, ..., k$, a expressão é

$$se\left(\widehat{\beta}_j|\mathbf{x}\right) = \left(\frac{se_0\left(\widehat{\beta}_j|\mathbf{x}\right)}{\widehat{\sigma}_u}\right)^2 \sqrt{\widehat{v}}, \tag{30}$$

onde $se_0\left(\widehat{\beta}_j|\mathbf{x}\right)$ é o desvio padrão de $\widehat{\beta}_j$ **não** robusto a autocorrelação (fórmula standard do OLS), $\widehat{\sigma}_u$ é o desvio padrão da regressão (16) e

$$\begin{aligned} \widehat{v} &= \sum_{t=1}^{T} \widehat{a}_t^2 + 2\sum_{l=1}^{m} \kappa_{l,m} \left(\sum_{t=l+1}^{T} \widehat{a}_t \widehat{a}_{t-l}\right) = \sum_{t=1}^{T} \widehat{a}_t^2 + 2\sum_{l=1}^{m}\sum_{t=l+1}^{T} \kappa_{l,m} \widehat{a}_t \widehat{a}_{t-l}; & (31) \\ \kappa_{l,m} &= 1 - \frac{l}{m+1}; & (32) \\ \widehat{a}_t &= \widehat{r}_t \widehat{u}_t; & (33) \end{aligned}$$

onde $\widehat{u}_t$ é o resíduo OLS do modelo (16) e $\widehat{r}_t$ é o resíduo OLS da regressão auxiliar que tem x_{tj} como variável dependente e $x_{ti}, i = 1, ..., k, i \neq j$ como regressores (inclui termo independente). A escolha para a "janela" $\kappa_{l,m}$ poderia ser outra e m é previamente escolhido (representa o número de lags para o qual se admite a existência de autocorrelação significativa).

Newey and West (1987) por sua vez derivam o seguinte estimador que é *robusto* apenas a autocorrelação:

$$\widehat{\Sigma}=\widehat{S}_0+\frac{1}{T}\sum_{l=1}^{m}\sum_{t=l+1}^{T}\kappa_{l,m}\widehat{u}_t\widehat{u}_{t-l}\left(x_tx_{t-l}^{'}+x_{t-l}x_t^{'}\right), \tag{34}$$

onde $\widehat{S}_0=\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2x_tx_t^{'}$ (ver a matriz de White no capítulo de heterocedasticidade). A raíz quadrada de $\widehat{\Sigma}_{jj}$ é o desvio padrão de $\widehat{\beta}_j$ *robusto* a autocorrelação. A matriz $\widehat{\Sigma}$ é o estimador consistente para Σ de Newey-West robusto a autocorrelação.

0.5 Estimação Eficiente e Procedimentos Iterativos

Quando o MRLM tem erros autocorrelacionados, existe um estimador alternativo ao OLS que é BLUE. Este é um estimador GLS $\widehat{\beta}=\left(X^{'}\Omega^{-1}X\right)^{-1}X^{'}\Omega^{-1}y; V\left(\widehat{\beta}|X\right)=\left(X^{'}\Omega^{-1}X\right)^{-1}$ (ver o capitulo da heterocedasticidade) ou FGLS consoante ele seja viável ou não. Semelhante ao método apresentado no capítulo da heterocedasticidade, o objectivo é transformar a equação (16) com erros autocorrelacionados de tal forma que o novo modelo transformado tenha erros não autocorrelacionados. Vamos restringir a análise a modelos com exogeneidade estrita (12) $E(u_t|X)=E(u_t|x_1,...,x_T)=0,t=1,...,T.$

Para o MRLM (16) ,

$$y_t=\beta_1+\beta_2x_{t2}+...+\beta_kx_{tk}+u_t=x_t^{'}\beta+u_t,t=1,...,T, \tag{35}$$

consideremos o caso de erros AR(1),

$$u_t=\rho u_{t-1}+\varepsilon_t, \tag{36}$$

em que $|\rho|<1,\varepsilon_t\sim wn\left(0,\sigma_\varepsilon^2\right),E(\varepsilon_t|u_{t-1},u_{t-2},...)=0$ e $V(u_t|X)=\frac{\sigma_\varepsilon^2}{(1-\rho^2)}$. A transformação adequada é a de quase-diferenças, isto é, subtrair

$$\rho y_{t-1}=\rho\left(\beta_1+\beta_2x_{t-1,2}+...+\beta_kx_{t-1,k}+u_{t-1}\right)=\rho\left(x_{t-1}^{'}\beta+u_{t-1}\right),t=2,...,T \tag{37}$$

a ambos os membros da equação (16) . Desta forma, o modelo transformado é

$$\begin{aligned} y_t-\rho y_{t-1} &= (1-\rho)\beta_1+\beta_2(x_{t2}-\rho x_{t-1,2})+...+\beta_k(x_{tk}-\rho x_{t-1,k})+(u_t-\rho u_{t-1}) && (38)\\ &= \left(x_t^{'}-\rho x_{t-1}^{'}\right)\beta+(u_t-\rho u_{t-1}),t=2,...,T, && (39)\end{aligned}$$

$$y_t^*=(1-\rho)\beta_1+\beta_2x_{t2}^*+...+\beta_kx_{tk}^*+\varepsilon_t=x_t^{*'}\beta+\varepsilon_t,t=2,...,T. \tag{40}$$

onde ε_t é não autocorrelacionado. Para que o OLS do modelo transformado - GLS com ρ conhecido - seja BLUE é necessário definir o modelo também para $t=1:$

$$\begin{aligned} y_1^* &= \beta_1\sqrt{1-\rho^2}+\beta_2x_{12}^*+...+\beta_kx_{1k}^*+u_1^*=x_1^{*'}\beta+u_1^* && (41)\\ y_1^* &= y_1\sqrt{1-\rho^2};x_{1j}^*=x_{1j}\sqrt{1-\rho^2};u_1^*=u_1\sqrt{1-\rho^2}. \end{aligned}$$

Deste modo, $V(u_1^*|X) = \sigma_\varepsilon^2 < \frac{\sigma_\varepsilon^2}{(1-\rho^2)} = V(u_1|X)$. Neste modelo transformado, as hipóteses Gauss-Markov são válidas e por isso este estimador GLS, onde ρ é conhecido, é BLUE. Adicionalmente, a inferência é válida assimptóticamente (também para amostras de reduzida dimensão se admitir-mos como verdadeira a hipótese $M5$ de erros normalmente distribuidos).

Quando ρ não é conhecido, o GLS não é viável e por isso deve-se procurar usar o FGLS onde um estimador para ρ é utilizado. Este é obtido pela fórmula (24) que resulta da regressão (23). O estimador $\widehat{\rho}$ é consistente e por isso, apesar de poder ser enviesado (o que implica que pode não ser BLUE), ao substituir-se ρ por $\widehat{\rho}$ em (40) o estimador FGLS $\widehat{\beta}$ é consistente (assumimos exogeneidade estrita e dados *weakly dependent*), assimptóticamente eficiente e a inferência é válida para amostras de grande dimensão. Se a primeira observação (41) não é utilizada, denominamos este procedimento como método de Cochrane-Orcutt, caso contrário é o de Prais-Winsten (assimptóticamente, a utilização da primeira observação não é necessária).

Na prática, e porque este é um procedimento a dois passos no qual as propriedades do FGLS para amostras finitas não são fáceis de se obter, estes dois métodos são aplicados de uma forma iterativa. A ideia é estimar sucessivamente ρ (e β) com novas séries de resíduos até à iteração na qual a variação de $\widehat{\rho}$ é negligenciável. O estimador FGLS $\widehat{\beta}$ é pois obtido à custa da última iteração para $\widehat{\rho}$. Eis como o processo iterativo funciona: *(i)* usar os resíduos OLS $u_t^{(1)}$ da regressão (16) e com isso obter $\widehat{\rho}^{(1)}$ de (23); *(ii)* derivar o estimador FGLS $\widehat{\beta}^{(1)}$ de (40) com recurso a $\widehat{\rho}^{(1)}$; *(iii)* obter os resíduos $u_t^{(2)}$ através do modelo original (16) com $\widehat{\beta}^{(1)}$ e depois estimar $\widehat{\rho}^{(2)}$ em (23) com $u_t^{(2)}$; *(iv)* derivar $\widehat{\beta}^{(2)}$ de (40) com $\widehat{\rho}^{(2)}$; *(v)* repetir este processo até que $\left|\widehat{\rho}^{(m)} - \widehat{\rho}^{(m-1)}\right| < 0.001$, por exemplo. Com base em $\widehat{\rho}^{(m)}$, o estimador FGLS é $\widehat{\beta}^{(m)}$.

Para erros $AR(p)$ estacionários (raízes do polinómio fora do círculo unitário),

$$u_t = \rho_1 u_{t-1} + ... + \rho_p u_{t-p} + \varepsilon_t, \tag{42}$$

o estimador (F)GLS é obtido de uma forma semelhante. A transformação adequada é subtrair

$$\rho y_{t-1} + ... + \rho_p y_{t-p}, t = p+1, ..., T \tag{43}$$

a ambos os membros da equação (16). A análise do GLS e FGLS é muito semelhante à apresentada para o caso de $p = 1$.

0.6 Estimador (F)GLS e Estimação HAC de $V(\widehat{\beta}_{OLS})$

Nesta secção apresentamos, muito sucintamente, a derivação dos estimadores BLUE GLS e FGLS e a estimação HAC robusta de $V(\widehat{\beta}_{OLS})$ quando os erros do MRLM são heterocedasticos e autocorrelacionados de uma forma funcional geral, $E(\mathrm{u}\,\mathrm{u}'|X) = \sigma^2\Omega_{T\times T}$, onde $\Omega \equiv \Omega(X)$ é uma matriz simétrica e positiva definida. Em termos de notação, note-se que pré-multiplicamos Ω por σ^2 para obter $E(\mathrm{u}\,\mathrm{u}'|X)$. Ver os apontamentos de heterocedasticidade (tome-se $\sigma^2 = 1$ para ter uniformidade na notação!). O objectivo é demonstrar que o modelo transformado terá erros que satisfazem as hipóteses clássicas do MRLM, $E(\mathrm{u}\,\mathrm{u}'|X) = \sigma^2 I_{T\times T}$.

O estimador OLS tem como matriz de variâncias-covariâncias

$$\Sigma_{k\times k} = V\left(\widehat{\beta}|X\right) = \sigma^2\left(X^{'}X\right)^{-1}X^{'}\Omega X\left(X^{'}X\right)^{-1}. \tag{44}$$

Na estimação HAC de $\Sigma, \widehat{\Sigma}$, precisamos de um estimador consistente para a matriz de variâncias-covariâncias de $u, \sigma^2\Omega$, de Newey and West (1987). Por isso,

$$\widehat{\Sigma} = \left(X^{'}X\right)^{-1}X^{'}E(\widehat{\mathrm{u}\,\mathrm{u}^{'}}|X)X\left(X^{'}X\right)^{-1}, \tag{45}$$

onde

$$E(\widehat{\mathrm{u}\,\mathrm{u}^{'}}|X) = \widehat{S}_0 + \sum_{l=1}^{m_T}\kappa_{l,m_T}\left(\widehat{S}_l + \widehat{S}_l^{'}\right); \widehat{S}_l = \frac{1}{T}\sum_{t=l+1}^{T}\widehat{u}_t\widehat{u}_{t-l}^{'}, l = 0, 1, ..., T-1. \tag{46}$$

Para finalizar, deduzimos as expressões dos estimadores GLS and FGLS. Suponha que a matriz $\Omega_{T\times T}$ é conhecida. Como Ω é SPD (simétrica e positiva definida), Ω^{-1} é também SPD e, por isso, admite a decomposição $\Omega^{-1} = P^{'}P$, onde P é invertível e $P\Omega P^{'} = I_{T\times T}$. Pode-se provar que a diagonalização de Ω, SPD, é $\Omega = C\Lambda C^{'} = C\Lambda^{1/2}\Lambda^{1/2}C^{'}$, onde Λ é uma matriz diagonal com os valores próprios de Ω, todos eles positivos. A transformação do modelo original é

$$y = X\beta + u \Leftrightarrow Py = PX\beta + Pu \Leftrightarrow y^* = X^*\beta + u^*, \tag{47}$$

onde

$$\begin{aligned} E(u^*|X) &= E(Pu|X) = PE(u|X) = 0, \\ E(u^*|X^*) &= E\left[E(u^*|X)|X^*\right] = E\left[0|X^*\right] = 0, \\ E(u^*u^{*'}|X) &= PE(uu^{'}|X)P^{'} = \sigma^2P\Omega P^{'} = \sigma^2 I, \\ E(u^*u^{*'}|X^*) &= E\left[E(u^*u^{*'}|X)|X^*\right] = E\left[\sigma^2 I|X^*\right] = \sigma^2 I. \end{aligned} \tag{48}$$

O estimador GLS é

$$\widehat{\beta} = \left(X^{*'}X^*\right)^{-1}X^{*'}y^* = \left(X^{'}\Omega^{-1}X\right)^{-1}X^{'}\Omega^{-1}y, \tag{49}$$

em que

$$V\left(\widehat{\beta}|X\right) = \sigma^2\left(X^{'}\Omega^{-1}X\right)^{-1}; \widehat{\sigma}^2 = \frac{\widehat{u}^{*\prime}\widehat{u}^*}{T-k} = \frac{\widehat{u}^{'}\Omega^{-1}\widehat{u}}{T-k}. \tag{50}$$

Quando $\sigma^2\Omega = E(\mathrm{u}\,\mathrm{u}^{'}|X)$ não é conhecida, obtemos o estimador FGLS e a sua variância à custa do estimador consistente $E(\widehat{\mathrm{u}\,\mathrm{u}^{'}}|X)$, definida anteriormente. Nestas condições, o FGLS é assimptóticamente eficiente (mesma distribuição assimptótica que o GLS).

Aplicações

(...)

Exercícios

1. O processo $y_t = \beta_1 + \beta_2 t$ é estacionário? Se não é, qual a transformação que o torna estacionário (estacionário à volta de uma tendência)? Existem condições para y_t ter memória que se dissipa?
2. Obtenha Ω e $V\left(\widehat{\beta}_2|\mathbf{x}\right)$ para o caso de u_t ser um processo MA(1).
3. ...