Fundação Getúlio Vargas
Graduação em Economia

ECONOMETRIA I

Notas de Aula - Autocorrelação

Ilton G. Soares
(iltonsoares@fgvmail.br)

1 Autocorrelação

1.1 Introdução

Uma das hipóteses do Modelo Clássico de Regressão Linear (MCRL) estabelece que não há autocorrelação ou correlação serial entre os termos de perturbação incluídos na Função de Regressão Populacional (FRP). Apesar de ser um fenômeno típico de séries temporais, pode também ocorrer em dados do tipo *cross-section*, entretanto neste tipo de dado a disposição das informações deve apresentar alguma lógica ou interesse econômico para que possamos compreender qualquer decisão sobre a presença ou não de autocorrelação.

A maioria das séries temporais em economia apresenta autocorrelação positiva, ainda que seja possível a ocorrência de autocorrelação negativa. Os gráficos (a) e (b) da figura 01 exibem os padrões de autocorrelação positiva e negativa, respectivamente, quando a série é representada ao longo do tempo. Já os gráficos (a) e (b) da figura 02 ilustram os padrões de autocorrelação positiva e negativa (na forma autoregressiva de ordem 1) quando avaliamos o gráfico de dispersão de série em função de seu valor defasado em um período.

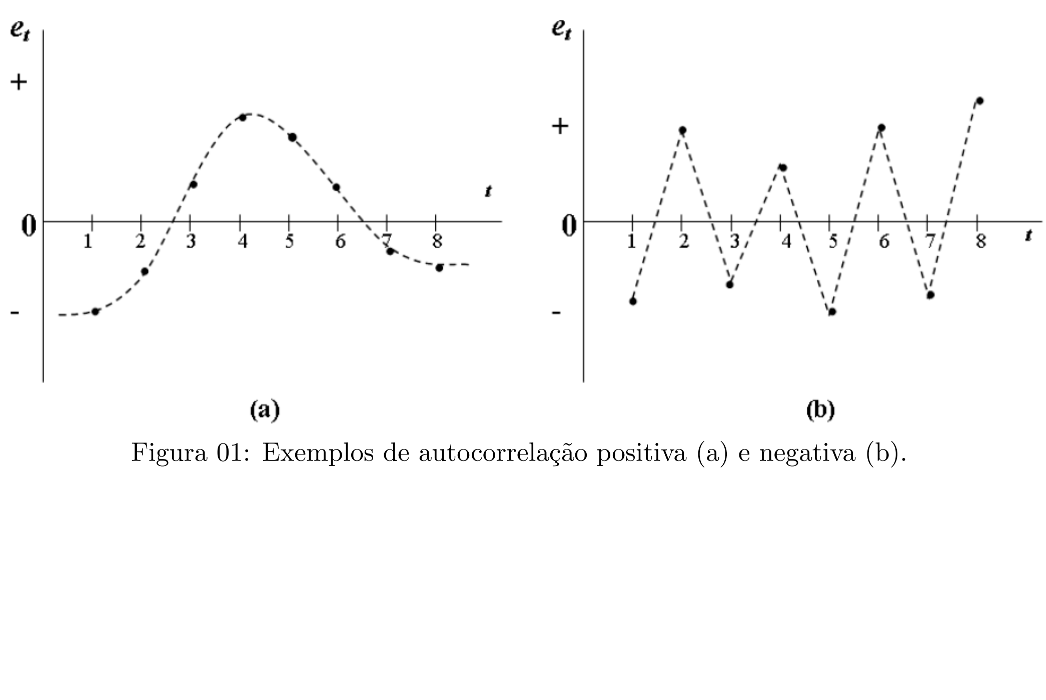


Figura 01: Exemplos de autocorrelação positiva (a) e negativa (b).

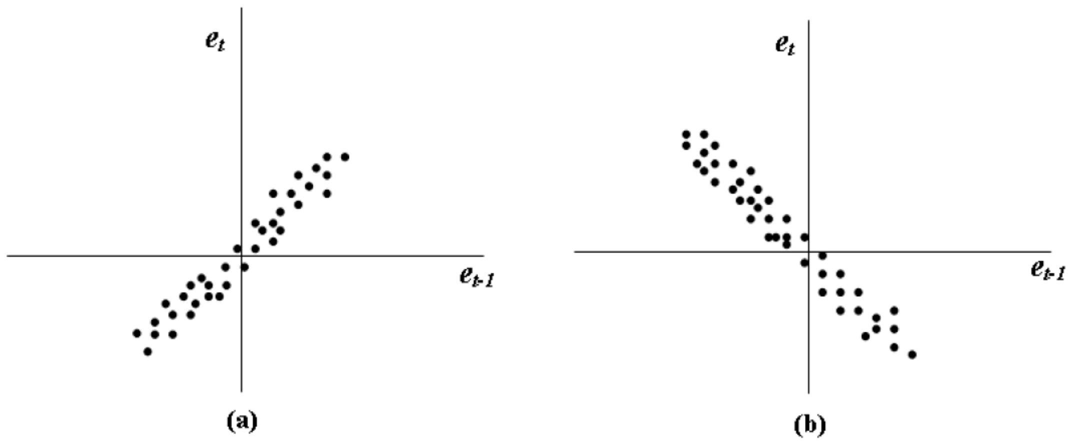


Figura 02: Exemplos de autocorrelação positiva (a) e negativa (b).

Visando tornar a nossa análise mais clara, usaremos alguns conceitos da análise de séries temporais:

Definição 1 *Um processo estocástico (ou processo de série temporal)* $\{Y_t\}_{t=0}^{\infty}$ *é uma seqüencia de variáveis aleatórias indexadas no tempo*

Definição 2 *Dizemos que o processo* $\{Y_t\}_{t=0}^{\infty}$ *com segundo momento finito* $\left[E\left(Y_t^2\right) < \infty\right]$ *é estacionário (fraco ou em covariância) se:*
i) $E\left(Y_t\right) = \mu$ *constante;*
ii) $Var\left(Y_t\right) = \sigma_y^2$ *constante;*
iii) $Cov\left(Y_t, Y_{t-h}\right) = \gamma_h$ *depende apenas de* h *e não de* t*, onde* $h > 0$.

Definição 3 *Dizemos que o processo* $\{\varepsilon_t\}_{t=0}^{\infty}$ *com segundo momento finito* $\left[E\left(\varepsilon_t^2\right) < \infty\right]$ *é um ruído branco se:*
i) $E\left(\varepsilon_t\right) = 0$ *para todo* t*;*
ii) $Var\left(\varepsilon_t\right) = \sigma_{\varepsilon}^2$ *constante;*
iii) $Cov\left(\varepsilon_t, \varepsilon_{t-h}\right) = 0$ *para todo* t *e* $h > 0$.

É fácil ver que o processo do tipo ruído branco é um caso particular de processo estacionário.

Considere o modelo linear em notação matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Assumindo as Hipóteses Clássicas do Modelo de Regressão Linear,

$$Var\left(\boldsymbol{\varepsilon}\right) = E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = \sigma^2\mathbf{I}$$

Mas note que

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) = E\begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{bmatrix} = \begin{bmatrix} E\left(\varepsilon_1^2\right) & E\left(\varepsilon_1\varepsilon_2\right) & \cdots & E\left(\varepsilon_1\varepsilon_n\right) \\ E\left(\varepsilon_2\varepsilon_1\right) & E\left(\varepsilon_2^2\right) & \cdots & E\left(\varepsilon_2\varepsilon_n\right) \\ \vdots & \vdots & \ddots & \vdots \\ E\left(\varepsilon_n\varepsilon_1\right) & E\left(\varepsilon_n\varepsilon_2\right) & \cdots & E\left(\varepsilon_n^2\right) \end{bmatrix}$$

de modo que a diagonal principal de $E(\boldsymbol{\varepsilon\varepsilon}')$ apresenta as variâncias dos erros[1] e fora da diagonal principal temos as covariâncias. Assim, se a matriz $E(\varepsilon\acute{\varepsilon})$ não for diagonal, teremos pelo menos um elemento $E(\varepsilon_i\varepsilon_j) \neq 0$, com $i \neq j$, ou seja, o modelo apresenta autocorrelação. Para contemplar a possibilidade de autocorrelação, faremos $E(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2\boldsymbol{\Omega}$, com $\boldsymbol{\Omega}$ perfeitamente geral (podendo apresentar ou não heteroscedasticidade).

Como $E(\varepsilon_i\varepsilon_j) = E(\varepsilon_j\varepsilon_i)$ para todo $i \neq j$, concluímos que a matriz $E(\boldsymbol{\varepsilon\varepsilon}')$ é simétrica e, por definição, positiva definida (por se tratar de uma matriz de variância e covariância). Uma vez que $E(\boldsymbol{\varepsilon\varepsilon}')$ é simétrica, não temos que estimar seus n^2 termos. No caso de autocorrelação e heteroscedasticidade precisamos estimar $n(n+1)/2$ termos. Em caso de autocorrelação e homoscedasticidade, precisamos estimar $[n(n-1)+2]/2$ termos. Note que, como temos apenas n observações, não podemos estimar $n(n+1)/2$ ou $[n(n-1)+2]/2$ parâmetros, de modo que precisamos fazer alguma suposição sobre a natureza autocorrelação.

1.2 Alguns Padrões de Autocorrelação

Dentre as incontáveis formas de autocorrelação, podemos destacar os seguintes padrões[2]:

1. Processo auto-regressivo de 1^a ordem, ou simplesmente AR(1). Consiste no processo mais comum em séries econômicas:
$$\varepsilon_t = \rho_1\varepsilon_{t-1} + u_t$$

2. Processo auto-regressivo de 2^a ordem, ou simplesmente AR(2):
$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-1} + u_t$$

3. Processo auto-regressivo de ordem p, ou simplesmente AR(p):
$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \ldots + \rho_p\varepsilon_{t-p} + u_t$$

4. Processo de média móvel de 1^a ordem, ou simplesmente MA(1)[3]:
$$\varepsilon_t = u_t - \theta_1 u_{t-1}$$

5. Processo de média móvel de 2^a ordem, ou simplesmente MA(2):
$$\varepsilon_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

6. Processo de média móvel de ordem q, ou simplesmente MA(q):
$$\varepsilon_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \ldots - \theta_q u_{t-q}$$

7. Processo auto-regressivo de 1^a ordem e de média móvel de 1^a ordem, ou simplesmente ARMA(1,1)
$$\varepsilon_t = \rho_1\varepsilon_{t-1} + u_t - \theta_1 u_{t-1}$$

8. Processo auto-regressivo de ordem p e de média móvel de ordem q, ou simplesmente ARMA(p,q)
$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \ldots + \rho_p\varepsilon_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \ldots - \theta_q u_{t-q}$$

Dentre os padrões de autocorrelação e correlação serial possíveis, o mais comum em séries temporais econômicas é o processo autoregressivo de primeira ordem, AR(1). No apêndice 03 tratamos das principais características do referido processo.

[1] Sob a hipótese de homoscedasticidade todos os elementos da diagonal principal dessa matriz são iguais
[2] Em todos os tipos apresentados, assumimos que $\{u_t\}_{t=1}^{\infty}$ é um ruído branco.
[3] MA tem origem do termo em inglês "*moving average*".

1.3 Quais as Conseqüências?

Sabemos que sob as Hipóteses Clássicas do Modelo de Regressão Linear,

$$Var\left(\varepsilon\right)=E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\sigma^{2}\mathbf{I}$$

e

$$Var\left(\widehat{\boldsymbol{\beta}}\right)=\sigma^{2}\left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

Contudo, nos modelos com autocorrelação, $E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\neq\sigma^{2}\mathbf{I}$, pois os elementos fora da diagonal principal da matriz de variância e covariância dos erros não serão todos nulos. Nesse caso, faremos $E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)=\sigma^{2}\boldsymbol{\Omega}$. Note primeiramente que na presença de autocorrelação os estimadores de mínimos quadrados continuam não viesados, pois na devivação desse resultado não precisamos supor ausência de autocorrelação. Porém, não observamos mais a eficiência dos referidos estimadores dentro da classe dos estimadores lineares e não viesados, isto é, não observamos mais a validade do Teorema de Gauss-Markov. A variância de $\widehat{\beta}$ na presença de autocorrelação em termos matriciais é derivada a seguir:

$$\begin{aligned}
Var\left(\widehat{\boldsymbol{\beta}}\right) &= E\left\{\left[\widehat{\boldsymbol{\beta}}-E\left(\widehat{\boldsymbol{\beta}}\right)\right]\left[\widehat{\boldsymbol{\beta}}-E\left(\widehat{\boldsymbol{\beta}}\right)\right]'\right\}\\
&= E\left\{\left[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]\left[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]'\right\}\\
&= E\left\{\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\right]\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\right]'\right\}\\
&= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]\\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left(\varepsilon\varepsilon'\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^{2}\boldsymbol{\Omega}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\\
&= \sigma^{2}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}
\end{aligned}$$

Além disso, as variâncias amostrais são viesadas, sendo o viés geralmente negativo. Isso faz com que tanto o R^2 quanto as estatísticas t e F tendam a ser exageradas. Adicionalmente, veremos que as propriedades assintóticas devem ser derivadas caso a caso. Por exemplo, no modelo

$$Y_t=\alpha+\beta X_{t-1}+\varepsilon_t$$

com erros ε_t se comportando como

$$\varepsilon_t=\rho\varepsilon_{t-1}+u_t$$

com $|\rho|<1$ e u_t ruído branco, $\widehat{\beta}$ é inconsistente se, por exemplo, $X_t=Y_{t-1}$ e consistente se $Cov\left(X_t,\varepsilon_t\right)=0$.

1.4 Como Testar?

1.4.1 Teste Durbin-Watson

Dado o modelo de regressão linear

$$Y_t=\beta_0+\beta_1X_{t1}+\beta_2X_{t2}+\ldots+\beta_kX_{tk}+\varepsilon_t\ , \tag{1}$$

a estatística de teste

$$d=\frac{\sum\limits_{t=2}^{n}\left(\widehat{\varepsilon}_t-\widehat{\varepsilon}_{t-1}\right)^2}{\sum\limits_{t=1}^{n}\widehat{\varepsilon}_t^2}$$

é comumente usada para o propósito de testar a presença de autocorrelação, onde $\widehat{\varepsilon}_t$ corresponde aos resíduos de mínimos quadrados do modelo (1). Contudo, antes de utilizá-la, é preciso saber sob quais hipóteses esse teste é válido:

H1) O processo gerador dos erros é um AR(1) do tipo $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ onde $|\rho| < 1$ e u_t é um ruído branco;
H2) Os coeficientes estimados do modelo a ser testado são consistentes[4];
H3) O modelo de regressão contém intercepto;
H4) A matriz $\mathbf{X}$ de variáveis explicativas é composta de colunas não estocásticas, ou fixas em amostragem repetida.

Considerando uma amostra suficientement grande, $\sum_{t=1}^{n}\widehat{\varepsilon}_t^2 \approx \sum_{t=2}^{n}\widehat{\varepsilon}_{t-1}^2$ e $\sum_{t=1}^{n}\widehat{\varepsilon}_t^2 \approx \sum_{t=2}^{n}\widehat{\varepsilon}_t^2$, de modo que

$$
\begin{aligned}
d &= \frac{\sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2} = \frac{\sum_{t=2}^{n}\left(\widehat{\varepsilon}_t^2 + \widehat{\varepsilon}_{t-1}^2 - 2\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}\right)}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2} \\
&= \frac{\sum_{t=2}^{n}\widehat{\varepsilon}_t^2}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2} + \frac{\sum_{t=2}^{n}\widehat{\varepsilon}_{t-1}^2}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2} - 2\frac{\sum_{t=2}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2} \approx 1 + 1 - 2\widehat{\rho}
\end{aligned}
$$

onde $\widehat{\rho}$ corresponde ao estimador de mínimos quadrados de ρ do modelo

$$\widehat{\varepsilon}_t = \rho\widehat{\varepsilon}_{t-1} + u_t$$

Note que

$$\widehat{\rho} = \frac{\sum_{t=2}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}}{\sum_{t=2}^{n}\widehat{\varepsilon}_{t-1}^2} \approx \frac{\sum_{t=2}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}}{\sum_{t=1}^{n}\widehat{\varepsilon}_t^2}.$$

Desse modo, temos que

$$d \approx 2\left(1 - \widehat{\rho}\right)$$

Dessa forma,

$\widehat{\rho} = 0 \Longrightarrow d \approx 2$ (ausência de autocorrelação)
$\widehat{\rho} = 1 \Longrightarrow d \approx 0$ (autocorrelação positiva)
$\widehat{\rho} = -1 \Longrightarrow d \approx 4$ (autocorrelação negativa)

Um **problema** associado ao teste de DW é a presença de regiões inconclusivas, como mostra a figura 03. Caso o valor calculado de d se localize entre d_l e d_s ou entre $4 - d_s$ e $4 - d_l$, nada podemos concluir.

[4] Veremos na próxima seção que isso implica que não podemos incluir como variáveis explicativas termos defasados da variável dependente.

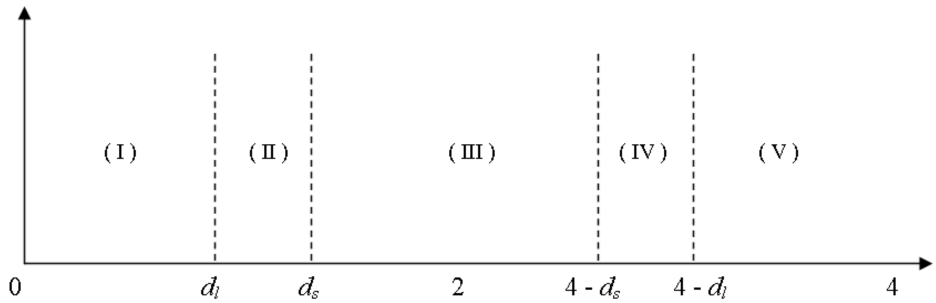


Figura 03: Regiões do teste de Durbin-Watson

onde:	(I)	H1: Existe autocorrelação positiva de primeira ordem;
	(II)	Zona inconclusiva;
	(III)	H0: Ausência de autocorrelação de primeira ordem;
	(IV)	Zona inconclusiva;
	(V)	H1: Existe autocorrelação negativa de primeira ordem;
	$d_l =$	Limite inferior;
	$d_s =$	Limite superior.

Exemplo 4 *Considere os dados da tabela 3.11 (ver no anexo destas notas). A função de produção a ser estimada é:*

$$\ln X = \alpha + \beta_1 \ln L_1 + \beta_2 \ln K_1 + v_t.$$

Os resultados da estimação são apresentados na figura a seguir:

Dependent Variable: LOG(X)
Method: Least Squares
Date: 05/16/07 Time: 15:29
Sample: 1929 1967
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.937714	0.236999	-16.61488	0.0000
LOG(L1)	1.450786	0.083228	17.43137	0.0000
LOG(K1)	0.383808	0.048018	7.993035	0.0000

R-squared	0.994627	Mean dependent var	5.687449
Adjusted R-squared	0.994329	S.D. dependent var	0.460959
S.E. of regression	0.034714	Akaike info criterion	-3.809542
Sum squared resid	0.043382	Schwarz criterion	-3.681576
Log likelihood	77.28607	F-statistic	3332.181
Durbin-Watson stat	0.858080	Prob(F-statistic)	0.000000

Consultando a tabela DW para o nível de significância de 5%, 2 variáveis explicativas e $n = 39$*, temos* $d_l = 1.38$*. Como a estatística d calculada é menor do que 1.38* ($d = 0.85 < 1.38 = d_l$)*, rejeitamos a hipótese* $\rho = 0$ *e verificamos evidências de autocorrelação de primeira ordem positiva.*

1.4.2 Teste h de Durbin

Vimos que dentre as hipóteses do teste de Durbin-Watson há uma que requer a consistência de $\widehat{\beta}$. Mostraremos agora que no caso simples de autocorrelação de primeira ordem, a inclusão de uma variável dependente defasada como explicativa torna o estimador de mínimos quadrados $\widehat{\beta}$ inconsistente. Considere o modelo

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$$

com

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

onde $|\beta| < 1$, $|\rho| < 1$ e u_t satisfaz as hipóteses clássicas do termo de erro (u_t é um ruído branco).

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum\left(Y_{t-1}-\overline{Y}\right)\left(Y_t-\overline{Y}\right)}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} = \frac{\sum\left(Y_{t-1}-\overline{Y}\right)Y_t}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} = \frac{\sum\left(Y_{t-1}-\overline{Y}\right)\left(\alpha+\beta Y_{t-1}+\varepsilon_t\right)}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} \\
&= \alpha\frac{\sum\left(Y_{t-1}-\overline{Y}\right)}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} + \beta\frac{\sum\left(Y_{t-1}-\overline{Y}\right)Y_{t-1}}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} + \frac{\sum\left(Y_{t-1}-\overline{Y}\right)\varepsilon_t}{\sum\left(Y_{t-1}-\overline{Y}\right)^2} \\
&= \beta + \frac{\sum\left(Y_{t-1}-\overline{Y}\right)\varepsilon_t}{\sum\left(Y_{t-1}-\overline{Y}\right)^2}
\end{aligned}$$

assim,

$$\begin{aligned}
\text{plim}\widehat{\beta} &= \text{plim}\beta + \frac{\text{plim}\left(\frac{1}{n}\sum\left(Y_{t-1}-\overline{Y}\right)\varepsilon_t\right)}{\text{plim}\left(\frac{1}{n}\sum\left(Y_{t-1}-\overline{Y}\right)^2\right)} \\
&= \beta + \frac{Cov\left(Y_{t-1},\varepsilon_t\right)}{Var\left(Y_{t-1}\right)}.
\end{aligned}$$

Mas note que

$$\begin{aligned}
Cov\left(Y_{t-1},\varepsilon_t\right) &= Cov\left(Y_{t-1},\rho\varepsilon_{t-1}+u_t\right) \\
&= \rho Cov\left(Y_{t-1},\varepsilon_{t-1}\right) + Cov\left(Y_{t-1},u_t\right) = \rho Cov\left(Y_{t-1},\varepsilon_{t-1}\right) \\
&= \rho Cov\left(Y_t,\varepsilon_t\right) = \rho Cov\left(\alpha+\beta Y_{t-1}+\varepsilon_t,\varepsilon_t\right) \\
&= \rho\beta Cov\left(Y_{t-1},\varepsilon_t\right) + \rho Var\left(\varepsilon_t\right)
\end{aligned}$$

logo,

$$Cov\left(Y_{t-1},\varepsilon_t\right) = \frac{\rho Var\left(\varepsilon_t\right)}{(1-\rho\beta)} = \frac{\rho\sigma_u^2}{(1-\rho\beta)\left(1-\rho^2\right)}.$$

pois $Var\left(\varepsilon_t\right) = \frac{\sigma_u^2}{(1-\rho^2)}$, onde se utilizou a estacionaridade de ε garantida pela hipótese de que $|\rho| < 1$. A condição $|\beta| < 1$ indica que Y é estacionário em covariância, logo $Var\left(Y_{t-1}\right) = Var\left(Y_t\right)$. Assim, para completar a dedução do plim$\widehat{\beta}$ resta apenas encontrar $Var\left(Y_t\right)$.

$$\begin{aligned}
Var\left(Y_t\right) &= Var\left(\alpha+\beta Y_{t-1}+\varepsilon_t\right) \\
&= \beta^2 Var\left(Y_{t-1}\right) + Var\left(\varepsilon_t\right) + 2\beta Cov\left(Y_{t-1},\varepsilon_t\right) \\
\left(1-\beta^2\right)Var\left(Y_t\right) &= Var\left(\varepsilon_t\right) + 2\beta Cov\left(Y_{t-1},\varepsilon_t\right) \\
&= \frac{\sigma_u^2}{(1-\rho^2)} + 2\beta\frac{\rho\sigma_u^2}{(1-\rho\beta)\left(1-\rho^2\right)}
\end{aligned}$$

de modo que

$$\begin{aligned}
Var\left(Y_t\right) &= \frac{1}{\left(1-\beta^2\right)}\left[\frac{\sigma_u^2}{(1-\rho^2)} + 2\beta\frac{\rho\sigma_u^2}{(1-\rho\beta)\left(1-\rho^2\right)}\right] \\
&= \frac{(1-\rho\beta)\sigma_u^2 + 2\beta\rho\sigma_u^2}{\left(1-\beta^2\right)(1-\rho\beta)\left(1-\rho^2\right)} \\
&= \frac{(1+\beta\rho)\sigma_u^2}{\left(1-\beta^2\right)(1-\rho\beta)\left(1-\rho^2\right)}.
\end{aligned}$$

Assim, temos finalmente que

$$\begin{aligned}
\text{plim}\widehat{\beta} &= \beta + \frac{Cov\left(Y_{t-1}, \varepsilon_t\right)}{Var\left(Y_{t-1}\right)} \\
&= \beta + \frac{\frac{\rho\sigma_u^2}{(1-\rho\beta)(1-\rho^2)}}{\frac{(1+\beta\rho)\sigma_u^2}{(1-\beta^2)(1-\rho\beta)(1-\rho^2)}} \\
&= \beta + \frac{\rho\left(1-\beta^2\right)}{(1+\beta\rho)}.
\end{aligned}$$

Com isso mostramos que MQO é inconsistente[5] a menos que $\rho = 0$ (Note que se $\rho = 0$, então $\varepsilon_t = u_t$ que é um ruído branco).

Neste caso, temos uma versão modificada do teste DW (corrigindo o problema da inconsistência), chamado teste h de Durbin, cuja estatística de teste é:

$$h = \widehat{\rho}\sqrt{\frac{n}{1 - n\widehat{V}\left(\widehat{\beta}\right)}} \xrightarrow{d} N(0,1)$$

onde $\widehat{\rho}$ é a correlação serial de primeira ordem estimada dos resíduos MQO, $\widehat{V}\left(\widehat{\beta}\right)$ é a variância estimada da estimativa de MQO de β e n é o tamanho da amostra.

Exemplo 5 *A estimação de uma equação de demanda por alimentos com 50 observações gerou os seguintes resultados:*

$$\ln q_t = const - \underset{(0.05)}{0.31}\ln P_t + \underset{(0.20)}{0.45}\ln y_t + \underset{(0.14)}{0.65}\ln q_{t-1}$$

$R^2 = 0.90$ *e* $d = 1.8$. *Como a variável* $\ln q_{t-1}$ *foi incluída entre as explicativas, sabemos que o teste DW não é válido. Para construir a estatística do teste h de Durbin, note que* $\widehat{\beta} = 0.65$, $\widehat{V}\left(\widehat{\beta}\right) = (0.14)^2 = 0.0196$ *e* $\widehat{\rho} = 0.1$ *(pois* $d = 1.8 \approx 2\left(1 - \widehat{\rho}\right)$*). Desse modo, a estatística h de Durbin é*

$$h = \widehat{\rho}\sqrt{\frac{n}{1 - n\widehat{V}\left(\widehat{\beta}\right)}} = 0.1\sqrt{\frac{50}{1 - 50 \times 0.0196}} = 5$$

Consultando a tabela da distribuição normal padrão, temos que o valor tabelado para o nível de significância de 1% é 2.3. Assim, mesmo sendo a estatística d bem próxima de 2, rejeitamos a hipótese de $\rho = 0$.

1.4.3 Simulação de Erros autoregressivos de quarta ordem e teste de Wallis

Vimos que os dois testes apresentados anteriormente tratam apenas do caso de autocorrelação onde os erros seguem um processo AR(1). Contudo, existe uma grande quantidade de casos de interesse em que essa suposição não é razoável. Por exemplo, se estivermos tratando do caso de séries trimestrais, pode ser mais adequado supor que o padrão de autocorrelação é do tipo AR(4) do que AR(1). Isso está ligado com a suposição de que observações associadas ao mesmo trimestre em anos distintos tendem a apresentar maior corelação do que observações de trimestres distintos de um mesmo ano. Para verificar o que acontece com a relação entre ε_t e ε_{t-1}, ε_{t-2}, ε_{t-3} e ε_{t-4} quando os erros obedecem a relação

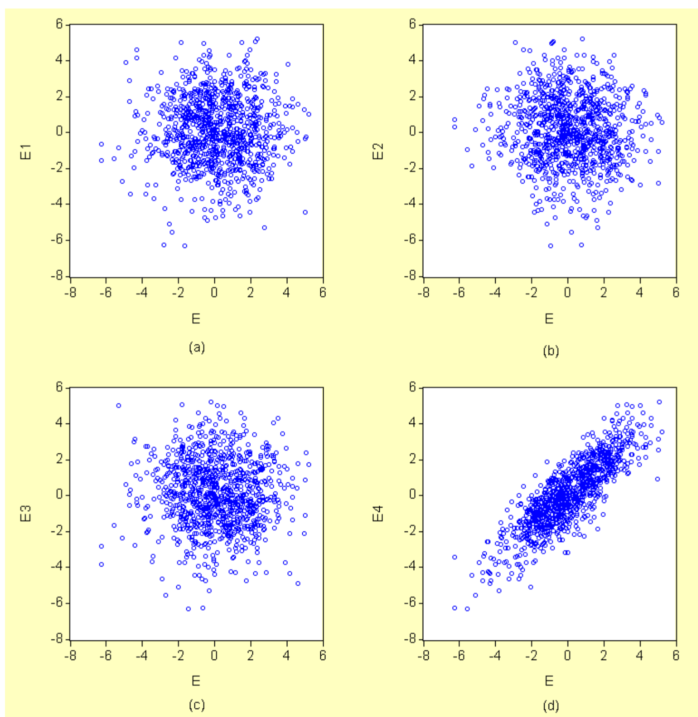
[5] Note que se os erros ε_t fossem do tipo ruído branco, $\widehat{\beta}$ seria consistente.

$$\varepsilon_t = 0.85\varepsilon_{t-4} + u_t \tag{2}$$

onde u_t é um ruído branco com distribuição normal (ou ruído branco gaussiano), construimos o seguinte experimento no EViews:

```
create u 1 1000
genr e=0
genr u=@nrnd
smpl 5 @last
genr e=0.85*e(-4)+u
genr e1=e(-1)
genr e2=e(-2)
genr e3=e(-3)
genr e4=e(-4)
```

A figura a seguir apresenta os gráficos de dispersão entre ε_t e ε_{t-1}, ε_{t-2}, ε_{t-3}, ε_{t-4}. Como é possível notar, não é observado padrão algum de autocorrelação até a defasagem de ordem 3. No gráfico de dispersão de ε_t contra ε_{t-4} percebemos claramente a presença de uma relação linear positiva entre as séries.



O propósito desse experimento é alertar para o fato de que a rejeição da hipótese de autocorrelação de até uma certa ordem p não implica que não há autocorrelação, uma vez que não podemos excluir a possibilidade de existência de autocorrelação de ordem superior a p. Assim, quando decidimos pela não rejeição da hipótese nula de que $\rho = 0$ no teste DW, estamos apenas com evidências estatísticas contrárias à presença de autocorrelação de primeira ordem, o que não significa que não podemos ter autocorrelação de ordem maior do que 1. Isto posto, como podemos testar a presença de autocorrelação de acordo com o padrão (2)? Uma possibilidade é utilizar o teste sugerido por Wallis (*Econometrica*, v.40, 1972, pp. 617-636). A estatística sugerida é semelhante à utilizada no teste DW:

$$d_4 = \frac{\sum_{t=5}^{n} (\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-4})^2}{\sum_{t=1}^{n} \widehat{\varepsilon}_t^2}.$$

Assim como no teste DW, temos aqui também a presença de regiões inconclusivas. Os valores tabelados podem ser encontrados em Wallis (*op. cit.*).

1.5 Teste de Breusch-Godfrey

Diferentemente do teste de Durbin-Watson para erros do tipo AR(1), o teste BG contempla a possibilidade de erros do tipo ARMA(p,q), e é aplicável caso haja ou não termos defasados do lado direito da equação.

Sob a hipótese nula de ausência de autocorrelação até a defasagem p, com p inteiro e pré-especificado, o teste baseia-se na série de resíduos da regressão estimada:

$$y_t = \widehat{\beta}_1 + \widehat{\beta}_2 x_{2t} + \widehat{\beta}_3 x_{3t} + \ldots + \widehat{\beta}_k x_{kt} + e_t \tag{3}$$

onde e_t corresponde aos resíduos da regressão de mínimos quadrados. Suponha que o termo ε_t (erro) seja gerado pelo seguinte processo AR(p):

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots + \rho_p \varepsilon_{t-p} + u_t$$

onde u_t é um termo de ruído branco. Queremos testar

$$H_0 : \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

o que indica que $\varepsilon_t = u_t$, isto é, não temos problema de autocorrelação. A estatística de teste se baseia na seguinte regressão auxiliar:

$$e_t = \widehat{\alpha}_1 + \widehat{\alpha}_2 x_{2t} + \widehat{\alpha}_3 x_{3t} + \ldots + \widehat{\alpha}_k x_{kt} + \gamma_1 e_{t-1} + \gamma_2 e_{t-2} + \ldots + \gamma_p e_{t-p} + v_t. \tag{4}$$

Note que esta é simplesmente a regressão dos resíduos da regressão estimada (3) em relação aos regressores e aos termos de resíduos defasados[6]. Note que, para calcularmos esta regressão, teremos apenas $(n-p)$ observações devido aos termos autoregressivos. Para não perder as p observações mencionadas, um artifício comumente usado (inclusive pelo EViews) é preencher esses valores com zero O R^2 da regressão (4) é usado para calcular a estatística BG:

$$BG = TR^2 \sim \chi^2_{pgl}$$

Caso TR^2 supere o valor χ^2_{pgl} crítico para o nível de significância escolhido, rejeita-se a hipótese nula. Nesta situação, o teste indica que pelo menos um dos ρ_i's é significativamente diferente de zero, com $i = 1, 2, ..., p$.

Exemplo 6 *Considere os dados da tabela apresentada no apêndice 02 que trata das séries de vendas e estoques no período 1950-1991. Estime o modelo*

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

em que Y = estoques e X = vendas, ambos medidos em bilhões de dólares e teste a presença de erros do tipo AR(3) usando o teste BG. O resultado da estimação é apresentado a seguir:

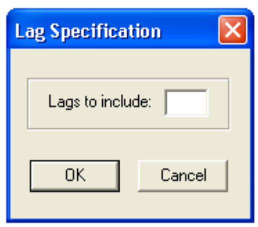
[6] Se p=1, significando um teste de autocorrelação de primeira ordem, então o teste BG é conhecido como teste n de Durbin.

Dependent Variable: ESTOQUE
Method: Least Squares
Date: 05/16/07 Time: 23:29
Sample: 1950 1991
Included observations: 42

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.654842	1.806451	0.916073	0.3651
VENDA	1.554365	0.006984	222.5642	0.0000

R-squared	0.999193	Mean dependent var	311.7254
Adjusted R-squared	0.999173	S.D. dependent var	259.1398
S.E. of regression	7.452385	Akaike info criterion	6.901393
Sum squared resid	2221.522	Schwarz criterion	6.984139
Log likelihood	-142.9293	F-statistic	49534.84
Durbin-Watson stat	1.374038	Prob(F-statistic)	0.000000

Para conduzir o teste BG no EViews, selecionamos ***View / Residual Tests / Serial Correlation LM Test...*** *na barra de ferramentas da equação estimada. Surgirá uma janela como a que segue:*



Nesta janela deve-se especificar o número de termos autoregressivos (p) que devem ser incluídos na regressão auxiliar (4). No caso em questão, p=3. O relatório padrão do teste é apresentado na figura a seguir:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.132684	Probability	0.036994
Obs*R-squared	8.507215	Probability	0.036614

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/16/07 Time: 23:32
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.062874	1.679814	-0.037429	0.9703
VENDA	0.000398	0.006520	0.060992	0.9517
RESID(-1)	0.314079	0.155054	2.025614	0.0501
RESID(-2)	0.177068	0.162221	1.091525	0.2821
RESID(-3)	-0.337657	0.156948	-2.151404	0.0380

R-squared	0.202553	Mean dependent var	3.44E-14
Adjusted R-squared	0.116342	S.D. dependent var	7.360942
S.E. of regression	6.919511	Akaike info criterion	6.817911
Sum squared resid	1771.547	Schwarz criterion	7.024776
Log likelihood	-138.1761	F-statistic	2.349513
Durbin-Watson stat	2.120679	Prob(F-statistic)	0.072099

*A estatística F exibida é um teste de significância conjunta dos termos de resíduos defasados. Guiando-nos pelo valor-p associado à estatística F, concluímos pela rejeição da hipótese nula de que os coeficientes dos termos de resíduos defasados são conjuntamente iguais a zero, considerando um nível de significância de 5%. Abaixo da estatística F é apresentada a estatística BG (Obs*R-squared). Como o valor-p associado à estatística BG é 0.036614, concluímos pela rejeição da hipótese nula*

$$H_0 : \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

para o nível de significância de 5%. Desse modo, pelo menos um dos ρ_i*'s é significativamente diferente de zero.*

*A observação "**Presample missing value lagged residuals set to zero**"indica que na regressão auxiliar de **RESID** em função de **VENDA**, **RESID(-1)**, **RESID(-2)** e **RESID(-3)**, as séries **RESID(-1)**, **RESID(-2)** e **RESID(-3)** entram com o valor zero nos campos não preenchidos com informação numérica (no caso a primeira observação de **RESID(-1)**, as duas primeiras observações de **RESID(-2)** e as três primeiras observações de **RESID(-3)**). Desta forma não perdemos observações na regressão auxiliar.*

1.6 Estimação com Erros Auto-Regressivos

Considere o modelo

$$y_t = \alpha + \beta x_t + u_t \tag{5}$$

com

$$u_t = \rho u_{t-1} + e_t \tag{6}$$

onde $t = 1, 2, ..., T$, $|\rho| < 1$ e $\{e_t\}_{t=1}^{\infty}$ é ruído branco. Como $u_t = y_t - \alpha - \beta x_t$, então $\rho u_{t-1} = \rho \left(y_{t-1} - \alpha - \beta x_{t-1}\right)$. Assim,

$$\begin{aligned} y_t &= \alpha + \beta x_t + u_t \\ &= \alpha + \beta x_t + \rho u_{t-1} + e_t \\ &= \alpha + \beta x_t + \rho \left(y_{t-1} - \alpha - \beta x_{t-1}\right) + e_t \end{aligned}$$

logo

$$y_t - \rho y_{t-1} = \alpha \left(1 - \rho\right) + \beta \left(x_t - \rho x_{t-1}\right) + e_t. \tag{7}$$

Pela hipótese de ruído branco de $\{e_t\}_{t=1}^{\infty}$, temos que o modelo transformado (7) pode ser estimado por MQO. A equação (7) énormalmente chamada de transformação *quase-diferença* de (5). A transformação que fazemos é simplesmente tomar

$$\begin{aligned} y_t^* &= y_t - \rho y_{t-1} \\ x_t^* &= x_t - \rho x_{t-1} \end{aligned}$$

para todo $t = 1, 2, ..., T$ e calcular a regressão de y_t^* em x_t^* com ou sem o termo constante, dependendo se a equação original tem ou não um termo constante. Note que ao fazer isso perdemos a primeira observação e é exatamente isso que diferencia este procedimento do procedimento de Mínimos Quadrados Generalizados (MQG). A diferença entre os dois procedimentos é que o MQG inclui a primeira observação, sendo esta definida por:

$$\begin{aligned} y_1^* &= \sqrt{1 - \rho^2} y_1 \\ x_1^* &= \sqrt{1 - \rho^2} x_1. \end{aligned}$$

Como na prática ρ não é conhecido, usamos uma das duas formas a seguir para estimá-lo:

1. Métodos iterativos: Mostraremos o procedimento de Cochrane-Orcutt.

2. Métodos grid-search: Mostraremos o método de Hildreth e Lu.

1.6.1 Método Iterativo de Cochrane-Orcutt

Vimos anteriormente que se pode obter uma estimativa de ρ a partir da estatística d de Durbin-Watson. Uma alternativa a esse método é o conhecido procedimento iterativo de Cochrane-Orctut. Considere o modelo

$$y_t = \beta_1 + \beta_2 x_t + u_t \tag{8}$$

com

$$u_t = \rho u_{t-1} + e_t \tag{9}$$

onde $t = 1, 2, ..., T$, $|\rho| < 1$ e $\{e_t\}_{t=1}^{\infty}$ é ruído branco. O procedimento de C-O consiste em:

1. Estimar o modelo (8) por MQO e obter os resíduos $\widehat{u}_t$.

2. A partir da série de resíduos obtida na etapa anterior, calcular a seguinte regressão

$$\widehat{u}_t = \widehat{\rho}\widehat{u}_{t-1} + w_t \tag{10}$$

3. Com o $\widehat{\rho}$ obtido na regressão (10), calcular a regressão de diferença generalizada

$$y_t - \widehat{\rho} y_{t-1} = \beta_1^* + \beta_2 \left(x_t - \widehat{\rho} x_{t-1}\right) + w_t \tag{11}$$

   o que pode ser feito sem problemas, já que $\widehat{\rho}$ é conhecido da etapa anterior. Além disso, note que $\beta_1^* = \beta_1 \left(1 - \widehat{\rho}\right)$. A regressão (11) nada mais é do que

$$y_t^* = \beta_1^* + \beta_2 x_t^* + w_t \tag{12}$$

   com $y_t^* = y_t - \widehat{\rho} y_{t-1}$ e $x_t^* = x_t - \widehat{\rho} x_{t-1}$.

4. Utilizando $\widehat{\beta}_1^*$ e $\widehat{\beta}_2$ da regressão (12), geramos uma nova série de resíduos $\widehat{u}_t^{**}$ por

$$\widehat{u}_t^{**} = y_t - \widehat{\beta}_1^* - \widehat{\beta}_2 x_t$$

   Note que $\widehat{u}_t^{**}$ pode ser gerado sem problemas uma vez que todos os termos do lado direito da igualdade são de nosso conhecimento.

5. Após obter $\widehat{u}_t^{**}$, calculamos a regressão

$$\widehat{u}_t^{**} = \widehat{\widehat{\rho}}\widehat{u}_{t-1}^{**} + v_t$$

   onde $\widehat{\widehat{\rho}}$ é o valor estimado de ρ na segunda rodada.

6. Esse processo deve continuar até que algum critério de convergência definido *a priori* tenha sido alcançado (por exemplo, se a variação entre o ρ estimado em duas rodadas vizinhas for menor do que 0.001).

Exemplo 7 *A partir dos dados apresentados na tabela 3.11 (ver apêndice 01), estime a seguinte função de produção:*

$$\ln X = \alpha + \beta_1 \ln L_1 + \beta_2 \ln K_1 + u$$

A figura a seguir apresenta os resultados da estimação (LS LOG(X) C LOG(L1) LOG(K1)):

Dependent Variable: LOG(X)
Method: Least Squares
Date: 05/17/07 Time: 00:48
Sample: 1929 1967
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.937714	0.236999	-16.61488	0.0000
LOG(L1)	1.450786	0.083228	17.43137	0.0000
LOG(K1)	0.383808	0.048018	7.993035	0.0000

R-squared	0.994627	Mean dependent var	5.687449
Adjusted R-squared	0.994329	S.D. dependent var	0.460959
S.E. of regression	0.034714	Akaike info criterion	-3.809542
Sum squared resid	0.043382	Schwarz criterion	-3.681576
Log likelihood	77.28607	F-statistic	3332.181
Durbin-Watson stat	0.858080	Prob(F-statistic)	0.000000

Agora, estime o mesmo modelo assumindo que os erros são AR(1)

$$u_t = \rho u_{t-1} + e_t.$$

Para isso, basta executar o comando LS LOG(X) C LOG(L1) LOG(K1) AR(1) na janela de comandos. O resultado da estimação aparece a seguir:

Dependent Variable: LOG(X)
Method: Least Squares
Date: 05/17/07 Time: 00:52
Sample (adjusted): 1930 1967
Included observations: 38 after adjustments
Convergence achieved after 15 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.473312	0.655351	-3.774028	0.0006
LOG(L1)	1.031249	0.175206	5.885927	0.0000
LOG(K1)	0.548258	0.105711	5.186413	0.0000
AR(1)	0.846255	0.117187	7.221424	0.0000

R-squared	0.996698	Mean dependent var	5.699067
Adjusted R-squared	0.996407	S.D. dependent var	0.461324
S.E. of regression	0.027653	Akaike info criterion	-4.238863
Sum squared resid	0.025999	Schwarz criterion	-4.066486
Log likelihood	84.53840	F-statistic	3421.142
Durbin-Watson stat	1.171439	Prob(F-statistic)	0.000000

Vemos no output do EViews[7] *que o procedimento encontrou convergência após 15 iterações. Além disso, é notável a diferença entre os coeficientes estimados no modelo que leva em conta a autocorrelação e o que não leva, ou seja, a correção da autocorrelação resultou em uma mudança considerável nas estimativas de* β_1 *e* β_2.

1.6.2 Método de Hildreth e Lu

O método consiste em calcular y_t^* e x_t^* para diferentes valores de ρ no conjunto

$$\{-1, \ -0.9, \ -0.8, \ ..., \ 0, \ ..., \ 0.8, \ 0.9, \ 1\}.$$

[7] O EViews estima modelos com erros AR utilizando técnicas de regressão não lineares.

Em seguida, estimamos a regressão de y_t^* em x_t^* para todos os valores de ρ e computamos a SQR. Escolhemos aquele ρ associado à menor SQR. Em seguida, refinamos nosso conjunto de $\acute{\rho}s$ agora em torno do ρ escolhido na primeira etapa. Se, por exemplo o ρ mínimo for 0.6, repetimos o método de busca para valores de ρ no conjunto

$$\{0.5,\ 0.51,\ 0.52,\ ...,\ 6,\ ...,\ 6.8,\ 6.9,\ 7\}.$$

Ao escolher o ρ associado à menor SQR nessa segunda etapa podemos interromper o processo ou continuar refinando nossa busca. Maddala (2003) observa que para amostras grandes, o método de Hildreth e Lu e o método de Máxima Verossimilhança geram resultados muito próximos.

1.7 Matriz de Variância e Covariância de Newey-West

Newey e West (1987) propuseram um estimador geral da matriz de variância e covariância de $\widehat{\boldsymbol{\beta}}$ que é consistente na presença tanto de heteroscedasticidade quato de autocorrelação com padrão desconhecido. O estimador de Newey-West é dado por:

$$\widehat{\Sigma}_{NW} = \frac{T}{T-k}\left(\mathbf{XX}\right)^{-1}\widehat{\mathbf{\Omega}}\left(\mathbf{XX}\right)^{-1}$$

onde

$$\widehat{\mathbf{\Omega}} = \frac{T}{T-k}\left\{\sum_{t=1}^{T} u_t^2 x_t x_t' + \sum_{v=1}^{q}\left(\left(1-\frac{v}{q+1}\right)\sum_{t=v+1}^{T}\left(x_t u_t u_{t-v} x_{t-v}' + x_{t-v} u_{t-v} u_t x_t'\right)\right)\right\}$$

e q, a defasagem truncada (*truncation lag*), é um parâmetro representando o número de autocorrelações usadas na avaliação da dinâmica dos resíduos MQO u_t. Newey e West sugerem que q seja obtido por:

$$q = floor\left(4\left(T/100\right)^{2/9}\right)$$

onde $floor(x)$ refere-se à função "maior inteiro menor ou igual a x". Para estimar no EViews um modelo com matriz de variância e covariância dos coeficientes estimados igual à proposta por Newey-West, você deve selecionar na janela **Equation Estimation** a opção **Options** e marcar **Heteroskedasticity Consistent Covariance** e **Newey-West**.

Exemplo 8 *Usando a mesma base de dados do exemplo anterior, o modelo estimado considerando a matriz de variância e covariância de Newey-West é apresentado na figura a seguir:*

Dependent Variable: LOG(X)
Method: Least Squares
Date: 05/17/07 Time: 00:46
Sample: 1929 1967
Included observations: 39
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.937714	0.345728	-11.38963	0.0000
LOG(L1)	1.450786	0.120898	12.00010	0.0000
LOG(K1)	0.383808	0.066489	5.772518	0.0000

R-squared	0.994627	Mean dependent var	5.687449
Adjusted R-squared	0.994329	S.D. dependent var	0.460959
S.E. of regression	0.034714	Akaike info criterion	-3.809542
Sum squared resid	0.043382	Schwarz criterion	-3.681576
Log likelihood	77.28607	F-statistic	3332.181
Durbin-Watson stat	0.858080	Prob(F-statistic)	0.000000

1.8 Correlação Serial Devido a Dinâmicas Mal Especificadas e Teste de Fatores Comuns (COMFAC)

Sargan (1964) argumentou que uma estatística DW significante não necessariamente implica que tenhamos um problema de autocorrelação. Para mostrar isso, considere o seguinte modelo:

$$y_t = \beta x_t + u_t \tag{13}$$

com

$$u_t = \rho u_{t-1} + e_t$$

onde e_t são independentes com variância constante σ^2. Como $u_t = y_t - \beta x_t$, então $\rho u_{t-1} = \rho\left(y_{t-1} - \beta x_{t-1}\right)$. Assim,

$$\begin{aligned} y_t &= \beta x_t + u_t \\ &= \beta x_t + \rho u_{t-1} + e_t \\ &= \beta x_t + \rho\left(y_{t-1} - \beta x_{t-1}\right) + e_t \\ &= \rho y_{t-1} + \beta x_t - \rho\beta x_{t-1} + e_t. \end{aligned} \tag{14}$$

Agora, considere o modelo dinâmico

$$y_t = \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + e_t \tag{15}$$

com $|\beta_1| < 1$. Note que as equações (14) e (15) são iguais desde que

$$\beta_1\beta_2 + \beta_3 = 0 \tag{16}$$

É fácil ver que um teste para $\rho = 0$ é um teste para $\beta_1 = 0$ (e $\beta_3 = 0$) em (15). Contudo, Sargan alerta que antes de testar $\rho = 0$, precisamos testar a validade da hipótese (16). Caso essa hipótese seja rejeitada, não teremos um modelo com autocorrelação e a correlação serial nos erros de (13) se deve a "dinâmicas mal especificadas", isto é, à omissão das variáveis y_{t-1} e x_{t-1} da equação. Para testar a restrição não linear (16) podemos usar testes baseados nos princípios de Wald, Lagrange (LM) ou da Razão de Verossimilhança (RV). Caso o teste DW indique a presença de autocorrelação, é indicado testar a restrição (16) antes de impor qualquer transformação autoregressiva nas variáveis. Sargan é mais radical, sugerindo que primeiro seja testada a restrição (16) e só então testar a presença de autocorrelação.

O teste de Wald não se mostra adequado nos casos de restrições não lineares por ser sensível ao modo como a restrição é escrita. Note que a restrição (16) pode ser escrita como

$$\beta_1 = -\frac{\beta_3}{\beta_2}$$

e o método de Wald não garante que o resultado dos testes com as duas hipóteses (iguais!) seja o mesmo. Por essa razão, prefere-se utilizar uma das outras duas alternativas (LM ou RV). No exemplo a seguir trabalharemos com o teste RV.

Exemplo 9 *No exemplo 7, a partir de uma estatística DW igual a 0.858 estimamos um modelo corrigindo quanto à presença de erros AR(1). Agora, estimaremos o mesmo modelo, porém sob a forma* (15), *cujos resultados são apresentados a seguir:*

Dependent Variable: LOG(X)
Method: Least Squares
Date: 05/17/07 Time: 12:53
Sample (adjusted): 1930 1967
Included observations: 38 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.150114	0.768017	-1.497511	0.1441
LOG(L1)	1.197453	0.216394	5.533660	0.0000
LOG(K1)	0.320345	0.205328	1.560165	0.1286
LOG(X(-1))	0.701718	0.171658	4.087874	0.0003
LOG(L1(-1))	-0.775291	0.209836	-3.694752	0.0008
LOG(K1(-1))	-0.198057	0.233729	-0.847377	0.4031

R-squared	0.996868	Mean dependent var	5.699067
Adjusted R-squared	0.996379	S.D. dependent var	0.461324
S.E. of regression	0.027760	Akaike info criterion	-4.186503
Sum squared resid	0.024660	Schwarz criterion	-3.927937
Log likelihood	85.54356	F-statistic	2037.245
Durbin-Watson stat	1.095351	Prob(F-statistic)	0.000000

Vimos que sob o pressuposto de que os erros são AR(1), a SQR obtida pelo método iterativo foi $SQR_1 = 0.025999$. *Uma vez que o nosso modelo apresenta dois coeficientes angulares, temos duas restrições da forma* (16). *No modelo dinâmico geral temos seis parâmetros* (α *e cinco* β*'s) e no modelo que contempla autocorrelação temos quatro parâmetros* (α, *dois* β*'s e* ρ). *O teste da RV baseia-se na estatística*

$$-2\ln\lambda \sim \chi^2_{rgl}$$

onde o número r *de graus de liberdade corresponde ao número de restrições (no nosso caso* $r = 2$*) e*

$$\lambda = \left(\frac{SQR_0}{SQR_1}\right)^{n/2}$$

logo, a nossa estatística de teste é

$$\begin{aligned} -2\ln\lambda &= -2\ln\left(\frac{SQR_0}{SQR_1}\right)^{38/2} = -2\ln\left(\frac{0.024660}{0.025999}\right)^{38/2} \\ &= -38\ln\left(\frac{0.024660}{0.025999}\right) = 2.0093 \end{aligned}$$

Uma vez que o valor tabelado para o nível de significância de 5% é $\chi^2_{crit} = 5.991$ *(no EViews, digitar o comando =@qchisq(0.95,2) para obter o valor tabelado em questão), não rejeitamos a hipótese nula de que o modelo AR(1) é apropriado para o problema em questão. (OBS.: Este exemplo está baseado em Maddala (2003, pp. 136-137), porém o referido livro apresenta* $SQR_0 = 0.01718$, *que não confere com os resultados encontrados nas nossas estimações. Além disso, para calcular as regressões sob teste, perdemos uma observação, logo* $n = 38$, *e não* 39 *como sugere o livro. Se considerássemos* $SQR_0 = 0.01718$ *e* $n = 39$ *(como apresentado no livro), então a nossa estatística de teste seria* 16.7, *que é significante a 1%. A conclusão nesse caso seria de que a hipótese de autocorrelação de primeira ordem é rejeitada, embora a estatística DW seja significante.)*

1.9 APÊNDICE 01: Tabela 3.11 (Maddala, 2003)

Ano	X	L1	L2	K1	K2
1929	189.8	173.3	44.151	87.80	888.90
1930	172.1	165.4	41.898	87.80	904.00
1931	159.1	158.2	36.948	84.00	900.20
1932	135.6	141.7	35.686	78.30	883.60
1933	132.0	141.6	35.533	76.60	851.40
1934	141.8	148.0	37.854	76.00	823.70
1935	153.9	154.4	39.014	77.70	805.30
1936	171.5	163.5	40.765	79.10	800.40
1937	183.0	172.0	42.484	80.00	805.50
1938	173.2	161.5	40.039	77.60	817.60
1939	188.5	168.6	41.443	81.40	809.80
1940	205.5	176.5	43.149	87.00	814.10
1941	236.0	192.4	46.576	96.20	830.30
1942	257.8	205.1	49.010	104.40	857.90
1943	277.5	210.1	49.695	110.00	851.40
1944	291.1	208.8	48.668	107.80	834.60
1945	284.5	202.1	47.136	102.10	819.30
1946	274.0	213.4	49.950	97.20	812.30
1947	279.9	223.6	52.350	105.90	851.30
1948	297.6	228.2	53.336	113.00	888.30
1949	297.7	221.3	51.469	114.90	934.60
1950	238.9	228.8	52.972	124.10	964.60
1951	351.4	239.0	55.101	134.50	1021.40
1952	360.4	241.7	55.385	139.70	1068.50
1953	378.9	245.2	56.226	147.40	1100.30
1954	375.8	237.4	54.387	148.90	1134.60
1955	406.7	245.9	55.718	158.60	1163.20
1956	416.3	251.6	56.770	167.10	1213.90
1957	422.8	251.5	56.809	171.90	1255.50
1958	418.4	245.1	55.023	173.10	1287.90
1959	445.7	254.9	56.215	182.50	1305.80
1960	457.3	259.6	56.743	189.00	1341.40
1961	466.3	258.1	56.211	194.10	1373.90
1962	495.3	264.6	57.078	202.30	1399.10
1963	515.5	268.5	57.540	205.40	1436.70
1964	544.1	275.4	58.508	215.90	1477.80
1965	579.2	285.3	60.055	225.00	1524.40
1966	615.6	297.4	62.130	236.20	1582.20
1967	631.1	305.0	63.162	247.90	1645.30

obs.: os pontos na tabela acima separam os decimais.

X = índice do produto interno bruto em dólares constantes
L_1 = índice de trabalho empregado (número de trabalhadores ajustado pelas taxas de utilização)
L_2 = população economicamente ativa
K_1 = índice de capital empregado
K_2 = estoque de capital em dólares constantes

1.10 APÊNDICE 02: Tabela exercício 12.32 (Gujarati, 2000)

Ano	Vendas	Estoques	Ano	Vendas	Estoques
1950	38.596	59.822	1971	117.023	188.991
1951	43.356	70.242	1972	131.227	203.227
1952	44.840	72.377	1973	153.881	234.406
1953	47.987	76.122	1974	178.201	287.144
1954	46.443	73.175	1975	182.412	288.992
1955	51.694	79.516	1976	204.386	318.345
1956	54.063	87.304	1977	229.786	350.706
1957	55.879	89.052	1978	260.755	400.929
1958	54.201	87.055	1979	298.328	452.636
1959	59.729	92.097	1980	328.112	510.124
1960	60.827	94.719	1981	356.909	547.169
1961	61.159	95.580	1982	348.771	575.486
1962	65.662	101.049	1983	370.501	591.858
1963	68.995	105.463	1984	411.427	651.527
1964	73.682	111.504	1985	423.940	665.837
1965	80.283	120.929	1986	431.786	664.654
1966	87.187	136.824	1987	459.107	711.745
1967	90.918	145.681	1988	496.334	767.387
1968	98.794	156.611	1989	522.344	813.018
1969	105.812	170.400	1990	540.788	835.985
1970	108.352	178.594	1991	533.838	828.184

Fonte: Economic Report of the President, 1993, Tabela B-53, p. 408.

1.11 APÊNDICE 03: Algumas derivações úteis de um processo AR(1)

Considere que o processo gerador dos erros é dado por

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

onde $|\rho| < 1$ e $\{u_t\}_{t=1}^{\infty}$ é um ruído branco. A condição $|\rho| < 1$ garante a estacionaridade[8] de $\{\varepsilon_t\}_{t=1}^{\infty}$. Com isso, podemos encontrar facilmente $E(\varepsilon_t)$, $Var(\varepsilon_t)$ e $Cov(\varepsilon_t, \varepsilon_{t-h})$.

A esperança de ε_t é dada por

$$\begin{aligned} E(\varepsilon_t) &= E(\rho\varepsilon_{t-1} + u_t) \\ &= E(\rho\varepsilon_{t-1}) + E(u_t) \\ &= \rho E(\varepsilon_{t-1}) + E(u_t) \end{aligned}$$

Assim, como o processo $\{\varepsilon_t\}_{t=1}^{\infty}$ é estacionário, $E(\varepsilon_t) = E(\varepsilon_{t-1}) = \mu$. Além disso, segue do fato de que u_t é um ruído branco que $E(u_t) = 0$ para todo t. Desse modo,

$$\begin{aligned} \mu &= \rho\mu + 0 \\ (1-\rho)\mu &= 0 \end{aligned}$$

ou seja,

$$\mu = E(\varepsilon_t) = 0.$$

A variância de ε_t também pode ser obtida de maneira simples:

$$\begin{aligned} Var(\varepsilon_t) &= Var(\rho\varepsilon_{t-1} + u_t) \\ &= Var(\rho\varepsilon_{t-1}) + Var(u_t) + 2\rho Cov(\varepsilon_{t-1}, u_t) \\ &= \rho^2 Var(\varepsilon_{t-1}) + Var(u_t) \end{aligned}$$

[8] O leitor interessado nos detalhes da demonstração de estacionaridade de um AR(1) é encorajado a consultar HAMILTON, J. D., *Time Series Analysis*, 1994, seções 2.2 e 3.4.

pois $Cov(\varepsilon_{t-1}, u_t) = 0$. Usando a estacionaridade de $\{\varepsilon_t\}_{t=1}^{\infty}$, temos que $Var(\varepsilon_t) = Var(\varepsilon_{t-1}) = \sigma^2$, logo

$$\begin{aligned} Var(\varepsilon_t) &= \rho^2 Var(\varepsilon_{t-1}) + Var(u_t) \\ \sigma^2 &= \rho^2\sigma^2 + \sigma_u^2 \\ \sigma^2 &= \frac{\sigma_u^2}{1-\rho^2}. \end{aligned}$$

Vejamos agora como obter as covariâncias do tipo $Cov(\varepsilon_t, \varepsilon_{t-h})$. Note que

$$Cov(\varepsilon_t, \varepsilon_{t-1}) = E\{[\varepsilon_t - E(\varepsilon_t)][\varepsilon_{t-1} - E(\varepsilon_{t-1})]\}$$

e que $E(\varepsilon_t) = 0$ para todo t. Assim,

$$\begin{aligned} Cov(\varepsilon_{t-1}, \varepsilon_{t-2}) &= E(\varepsilon_t\varepsilon_{t-1}) \\ &= E[(\rho\varepsilon_{t-1} + u_t)\varepsilon_{t-1}] \\ &= \rho E(\varepsilon_{t-1}^2) + E(u_t\varepsilon_{t-1}) \\ &= \rho\sigma^2 \end{aligned}$$

pois $E(u_t\varepsilon_{t-1}) = Cov(u_t, \varepsilon_{t-1}) = 0$ e $E(\varepsilon_{t-1}^2) = Var(\varepsilon_t) = \sigma^2$. De maneira inteiramente análoga,

$$\begin{aligned} Cov(\varepsilon_t, \varepsilon_{t-2}) &= E(\varepsilon_t\varepsilon_{t-2}) \\ &= E[(\rho\varepsilon_{t-1} + u_t)\varepsilon_{t-2}] \\ &= \rho E(\varepsilon_{t-1}\varepsilon_{t-2}) + E(u_t\varepsilon_{t-2}) \\ &= \rho Cov(\varepsilon_{t-1}, \varepsilon_{t-2}) \\ &= \rho^2\sigma^2 \end{aligned}$$

onde se usou o fato de que $Cov(\varepsilon_{t-1}, \varepsilon_{t-2}) = Cov(\varepsilon_{t-1}, \varepsilon_{t-2})$, pois $\{\varepsilon_t\}_{t=1}^{\infty}$ é estacionário. Procedendo desse modo de maneira sequencial, teremos

$$Cov(\varepsilon_t, \varepsilon_{t-h}) = \rho^h\sigma^2.$$

Assim, concluímos que

$$\begin{aligned} Corr(\varepsilon_t, \varepsilon_{t-h}) &= \frac{Cov(\varepsilon_t, \varepsilon_{t-h})}{\sqrt{Var(\varepsilon_t)\, Var(\varepsilon_{t-h})}} \\ &= \frac{Cov(\varepsilon_t, \varepsilon_{t-h})}{Var(\varepsilon_t)} \\ &= \frac{\rho^h\sigma^2}{\sigma^2} \\ &= \rho^h. \end{aligned}$$

logo, como $|\rho| < 1$, observamos que a correlação entre ε_t e ε_{t-h} é função decrescente de h, ou seja, da distância entre os erros.

1.12 Bibliografia

Greene, W. H. *Econometric Analysis.* Prentice Hall, 2003.
Gujarati, D. *Econometria Básica*, Makron Books, 2000.
Maddala, G. S. *Introdução à Econometria.* LTC, 2003.
Soares, I. G. e Castelar, I. *Econometria Aplicada com o Uso do EViews.* Livro Técnico, 2004.