O problema da autocorrelação nos termos de erro

1. Natureza da autocorrelação

- O modelo clássico de regressão linear supõe que inexiste correlação entre os termos de erro do modelo, isto é,
  $\operatorname{cov}(\mu_j, \mu_h) = 0$, para $j \neq h$ **(H.5)**

- Ou seja, o termo de erro associado a qualquer uma das observações não é influenciado pelo termo de erro de qualquer outra observação.

- Se for verificado que $\operatorname{cov}(\mu_j, \mu_h) \neq 0$, a autocorrelação estará presente.

Autocorrelação em dados de corte transversal

Dados em corte transversal frequentemente são gerados por meio de uma amostra aleatória de várias unidades econômicas, como famílias ou firmas. A aleatoriedade da amostra implica que os termos de erro para diferentes observações (famílias ou firmas) sejam não-correlacionados.

Autocorrelação em dados de séries temporais

Já em dados de séries temporais, em que as observações seguem um ordenamento natural ao longo do tempo, sempre existe a possibilidade de que erros sucessivos estejam correlacionados uns com os outros.

2. Consequências teóricas e práticas

- O estimador de MQO seguirá sendo **não tendencioso** e **consistente**, porém não mais apresentará variância mínima na classe de estimadores lineares não tendenciosos. Ou seja, não será eficiente (em termos relativos) e, portanto, não será MELNT.
- Sendo $\hat{V}(b)$ desnecessariamente grande, obteremos razões $t = \dfrac{b - \beta_{Hipotético}}{\sqrt{\hat{V}(b)}}$ menores que o adequado. Assim, estaremos propensos a declarar que um coeficiente é estatisticamente não significativo, embora na realidade possa não ser.
- Além disso, os IC serão desnecessariamente grandes, tendo em vista que o IC para β assume o formato $b - t_{\text{crítico}}\sqrt{\hat{V}(b)} < \beta < b + t_{\text{crítico}}\sqrt{\hat{V}(b)}$.

3. Detecção da autocorrelação

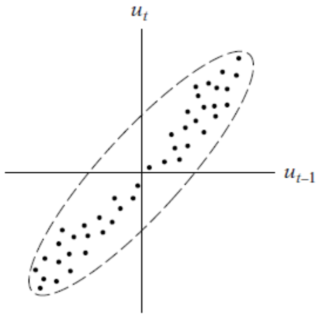
3.1 Métodos Gráficos

- Consiste no exame visual dos resíduos (e_t) da regressão. Esse método, contudo, não é infalível, tendo em vista que os e_t são proxies do verdadeiro termo de erro do modelo, μ_t.
- Uma primeira maneira de analisar os resíduos é plotando-os contra o tempo, a fim de verificar se exibem um padrão sistemático (assemelhando-se a uma função de 1º grau, a uma parábola etc.) ou se são aleatórios.
- Como alternativa, pode-se realizar a plotagem sequencial no tempo dos resíduos padronizados, que são os resíduos divididos pelo erro padrão da regressão: $\frac{e_t}{\sqrt{s^2}}$, em que s^2 é o quadrado médio residual.

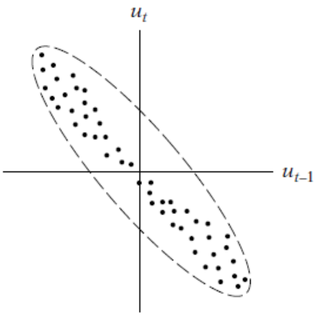
- Finalmente, pode-se plotar e_t contra e_{t-1}.

Caso os μ_t sigam um processo AR(1): $\mu_t = \rho\mu_{t-1} + \varepsilon_t$, será de esperar um resultado como (a) em caso de correlação positiva, ou como (b), se a autocorrelação for negativa.

(a)



(b)



3.2 Teste de Durbin-Watson (1951)

Finalidade:

- Testar se existe autocorrelação de 1ª ordem nos termos de erro, isto é, se eles são gerados pelo processo AR(1): $\mu_t = \rho\mu_{t-1} + \varepsilon_t$.

- Portanto, o teste não pode ser usado para detectar processos AR de ordem mais elevada.

Condições para operar o teste:

1. O modelo de regressão inclui o intercepto α
2. O modelo de regressão não inclui valores defasados de Y como variáveis explanatórias
3. As variáveis explanatórias são não estocásticas
4. $\mu_t \sim \mathrm{N}\left(0\,,\sigma^2\right)$
5. Não faltam observações nos dados

Operacionalização do teste

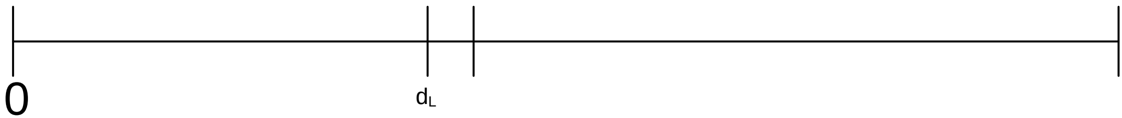
- Ajusta-se a regressão por MQO, obtendo-se os resíduos $e_t = Y_t - \hat{Y}_t$

- Calcula-se a estatística de teste: $d = \dfrac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$

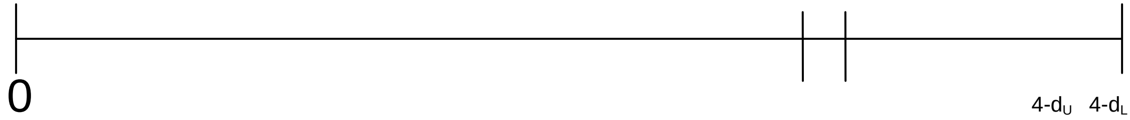
- É possível demonstrar que $d = 2(1 - r)$, em que r é o coeficiente de correlação entre e_t e e_{t-1}.

- O valor calculado d deve ser comparado com o valor crítico.

- Para testar $\begin{cases} H_0 : \rho = 0 \\ H_A : \rho > 0 \end{cases}$ d é comparado com d_L e d_U



- Para testar $\begin{cases} H_0: \rho = 0 \\ H_A: \rho < 0 \end{cases}$ d é comparado com $4 - d_U$ e $4 - d_L$



Atualmente já existem programas de computador que fornecem a probabilidade caudal associada ao valor calculado do teste de Durbin-Watson. Nesse caso basta comparar a probabilidade causal com o nível de significância adotado para decidir se o resultado é ou não significativo, evitando-se o problema do resultado inconclusivo.

Desvantagens do teste:

1. O teste não é genérico, somente podendo ser aplicado sob as hipóteses anteriormente elencadas.
2. Uma estatística d significativa pode não indicar necessariamente autocorrelação. Em vez disso, ela pode ser indicação de erro na especificação do modelo.

Exercício

A fim de estudar as variáveis que afetam o preço do cobre no mercado norte-americano, estabeleceu-se o seguinte modelo de regressão:

$$\ln(C_t) = \alpha + \beta_1 \ln I_t + \beta_2 \ln L_t + \beta_3 \ln H_t + \beta_4 \ln A_t + \mu_t \ ,$$

em que C = preço do cobre no mercado doméstico norte-americano (média dos últimos 12 meses), em centavos/libra, I = índice de produção industrial (média dos últimos 12 meses), L = preço de cobre na London Metal Exchange (média dos últimos 12 meses), em libras esterlinas, H = número de novas obras de habitação por ano, em milhares de unidades e A = preço do alumínio (média dos últimos 12 meses), em centavos/libra.

Utilizando dados anuais para o período 1951-1980, os seguintes resultados foram obtidos:

Model 1: OLS, using observations 1951-1980 (T = 30)
Dependent variable: $\ln(C_t)$

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	-1.50044	1.00302	-1.4959	0.14719	
$\ln I_t$	0.467509	0.165987	2.8165	0.00934	***
$\ln L_t$	0.279443	0.114726	2.4357	0.02233	**
$\ln H_t$	-0.00515155	0.142947	-0.0360	0.97154	
$\ln A_t$	0.441449	0.106508	4.1447	0.00034	***

Mean dependent var	3.721145	S.D. dependent var	0.447149
Sum squared resid	0.370573	S.E. of regression	0.121749
R-squared	0.936090	Adjusted R-squared	0.925864
F(4, 25)	91.54312	P-value(F)	1.49e-14
Log-likelihood	23.34039	Akaike criterion	-36.68077
Schwarz criterion	-29.67478	Hannan-Quinn	-34.43950
rho	0.520838	Durbin-Watson	0.954940

a) Relate e interprete os resultados obtidos (descreva, para cada variável explanatória do modelo, seu efeito sobre a variável dependente).
b) Realize o teste de Durbin-Watson (adote um nível de significância de 5%) e comente sobre a natureza da autocorrelação presente nos dados. Que pressupostos você teve que adotar para poder aplicar esse teste?