

خودهمبستگی

Autocorrelation

دانشجو مربوطه: فانم تجری

عدم خودهمبستگی بین جملات خطا یکی از فروض اساسی برای روش OLS است. همبستگی خطاها با یکدیگر را اصطلاحاً «خودهمبستگی» می‌گویند.

$$E(U_i U_j) \neq 0 \quad i \neq j$$

به طور کلی خودهمبستگی بیانگر رابطه بین e_t با وقفه‌های آن است.

مفهوم وقفه

X_{t-1}	X_t	زمان
-	1	1380
1	2	1381
2	4	1382
4	-2	1383
-2	-3	1384
-3	0	1385
0	5	1386

پیامدهای خودهمبستگی

اگر معادله $Y_t = \alpha + \beta X_t + u_t$ دارای خودهمبستگی به صورت $u_t = \rho u_{t-1} + \varepsilon_t$ باشد ولی آن را نادیده بگیریم:

۱- تخمین زنده های OLS، همچنان بدون تورش هستند.

۲- تخمین زنده های OLS، سازگارند.

۳- تخمین زنده های OLS کارا نخواهد بود (دیگر $\frac{\sigma^2}{\sum x_t^2} = \text{var}(\hat{\beta})$ نیست):

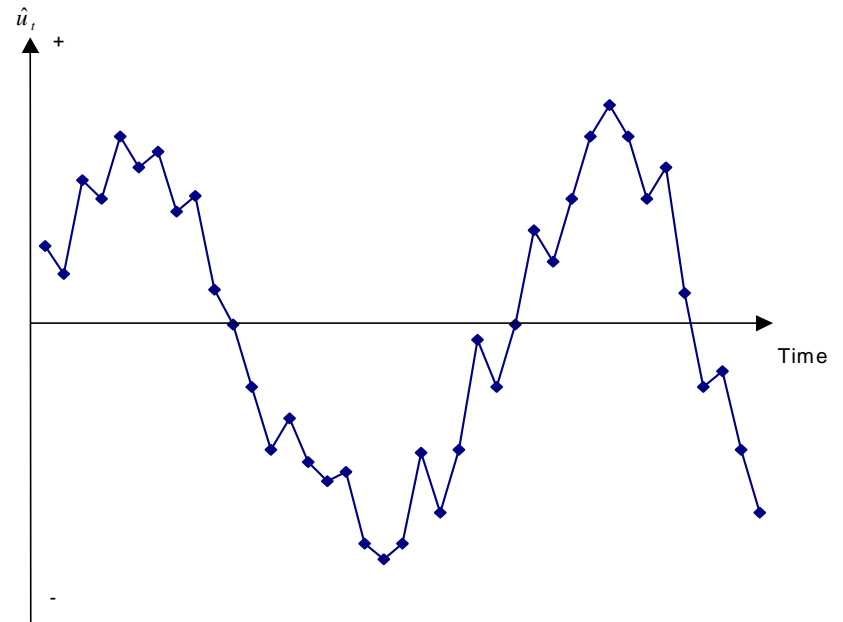
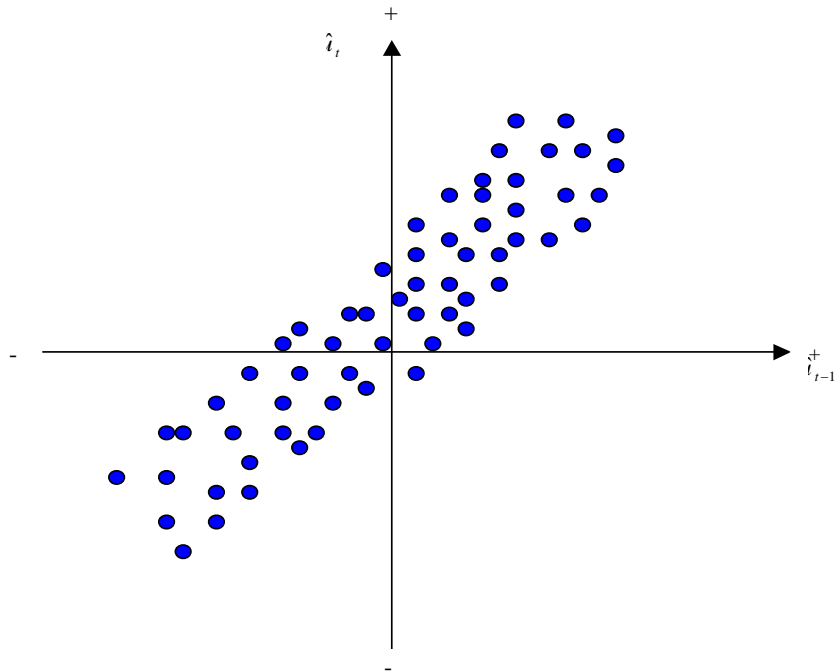
$$\begin{aligned}\text{var}(\hat{\beta}) &= E(\hat{\beta} - \beta)^\top = E\left(\sum_t w_t u_t\right)^\top, \quad w_t = \frac{x_t}{\sum x_t^\top} \\ &= E\left(\sum_t w_t^\top u_t^\top\right) + E\left(\sum_{t>s} \sum_t w_t w_s^\top u_t u_s^\top\right) \\ &= \sum_t w_t^\top E(u_t^\top) + \sum_{t>s} \sum_t w_t w_s^\top E(u_t u_s^\top)\end{aligned}$$

در نتیجه داریم:

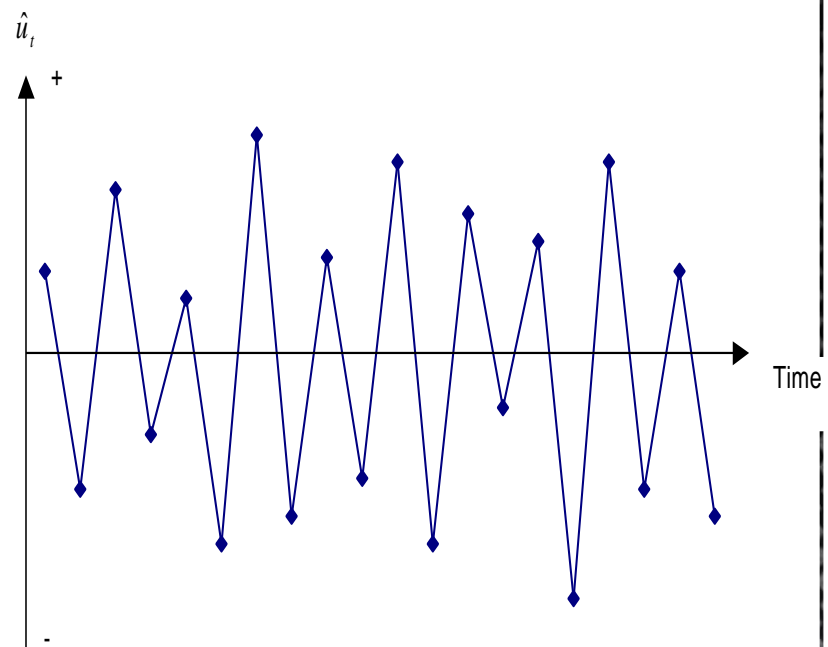
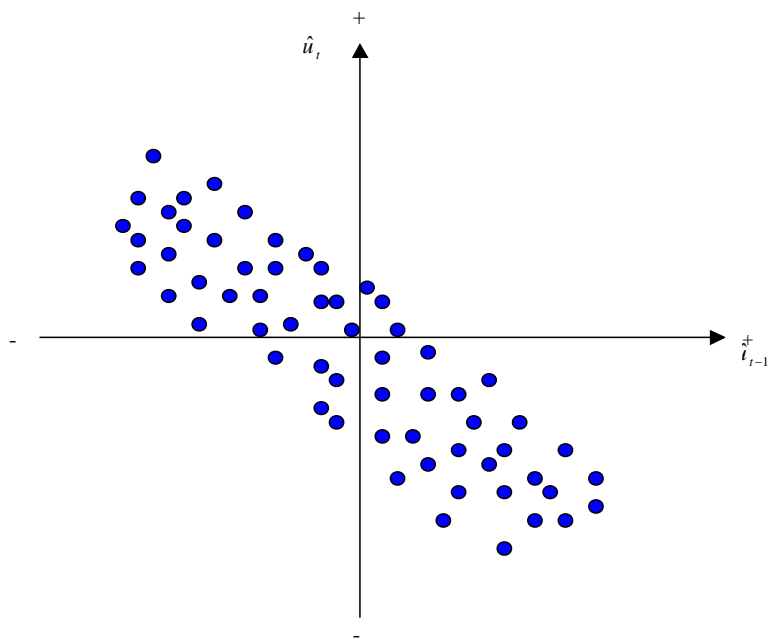
$$\text{var}(\hat{\beta}) = \frac{\sigma^\top}{\sum x_t^\top} + \sum_{t>s} \sum_t w_t w_s^\top \rho^{t-s} \sigma^\top$$

روش های نموداری جهت تشخیص خودهمبستگی

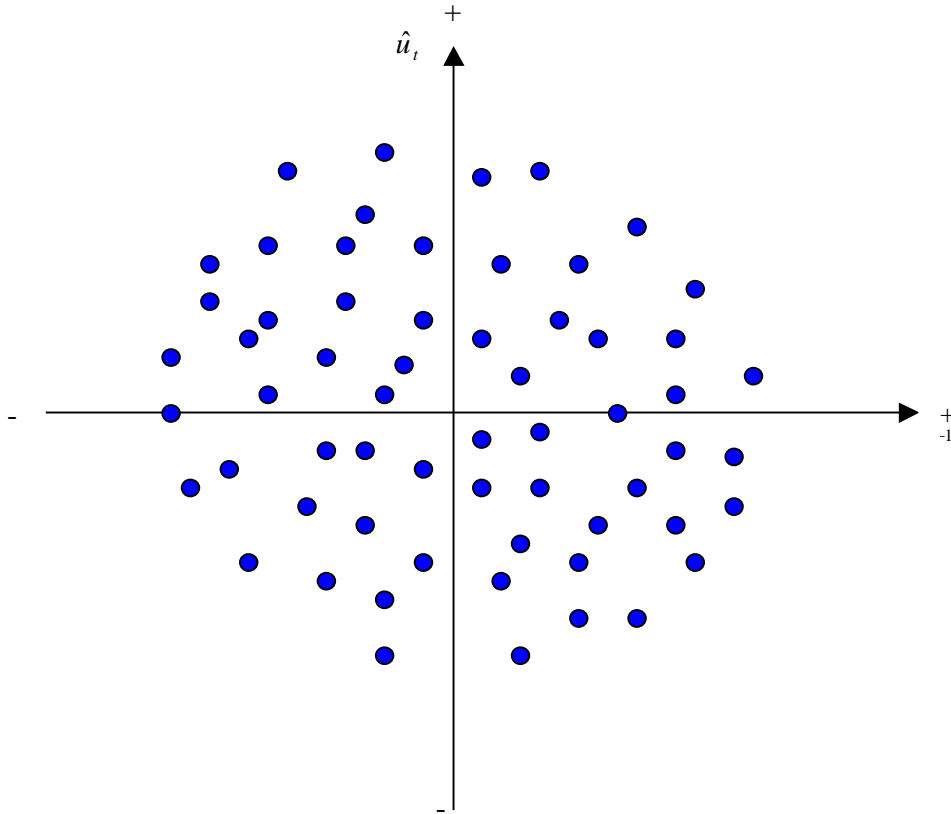
خودهمبستگی مثبت



خودهمبستگی منفی



عدم خود همبستگی



کشف خودهمبستگی:

1- آزمون دوربین-واتسون

برای تشخیص خودهمبستگی مرتبه اول به کار می رود:

$$u_t = \rho u_{t-1} + v_t \quad v_t \sim N(0, \sigma_v^2)$$

فرضیه ها در این آزمون بصورت زیر است:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

براساس e_t ها (برآورد u_t ها)، آماره DW بدین صورت است:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

برای ساده نمودن DW :

$$\sum_{t=2}^n (e_t - e_{t-1})^2 = \sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}$$

$$\sum_{t=2}^n e_t^2 = e_2^2 + e_3^2 + \dots + e_n^2 \cong \sum_{t=2}^n e_{t-1}^2 = e_1^2 + e_2^2 + \dots + e_{n-1}^2$$

$$DW \cong \frac{2 \sum_{t=2}^n e_t^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2} = 2 \left(1 - \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2} \right)$$




$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

داریم:

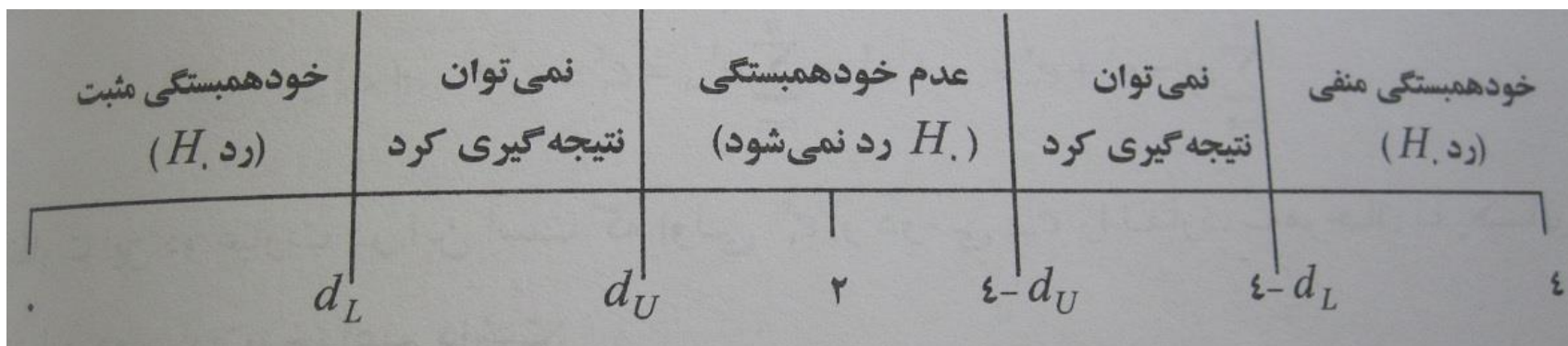
$$DW \cong 2(1 - \hat{\rho})$$

در نتیجه داریم:

از آنجا که $\hat{\rho}$ ضریب همبستگی است، باید $-1 \leq \hat{\rho} \leq 1$ باشد. $\Leftarrow 0 \leq DW \leq 4$
سه حالت امکان پذیر است:

- $\hat{\rho}=1$ و $DW=0$  وجود خودهمبستگی مثبت و کامل بین باقیمانده ها
- $\hat{\rho}=-1$ و $DW=4$  خودهمبستگی کامل و منفی
- $\hat{\rho}=0$ و $DW=2$  عدم خودهمبستگی

نواحی رد و عدم رد در DW:



شرایط اعتبار داشتن آزمون DW:

- 1- عرض از مبدا باید در رگرسیون اصلی وجود داشته باشد.
- 2- برآوردها باید غیرتصادفی باشند (فرض 4 مدل رگرسیون خطی)
- 3- در رگرسیون اصلی، نباید هیچ گونه وقفه ای از متغیر وابسته وجود داشته باشد.

اشکالات آزمون DW:

فقط برای تشخیص خود همبستگی نوع اول بکار می رود.
اگر داده مفقود داشته باشیم، این آماره قابل اعتنا نیست.
نباید متغیر وابسته تاخیری در مدل وجود داشته باشد.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \theta Y_{t-1} + u_t$$

در رابطه با مشکل آخر این معادله را در نظر بگیرید:

در این صورت برای بررسی خودهمبستگی از آماره h دوربین استفاده می کنیم:

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \text{var}(\hat{\theta})}} \quad , \quad \hat{\rho} = 1 - \frac{DW}{2}$$

که توزیع نرمال استاندارد دارد. و برای هر توزیع نرمال استاندارد داریم:

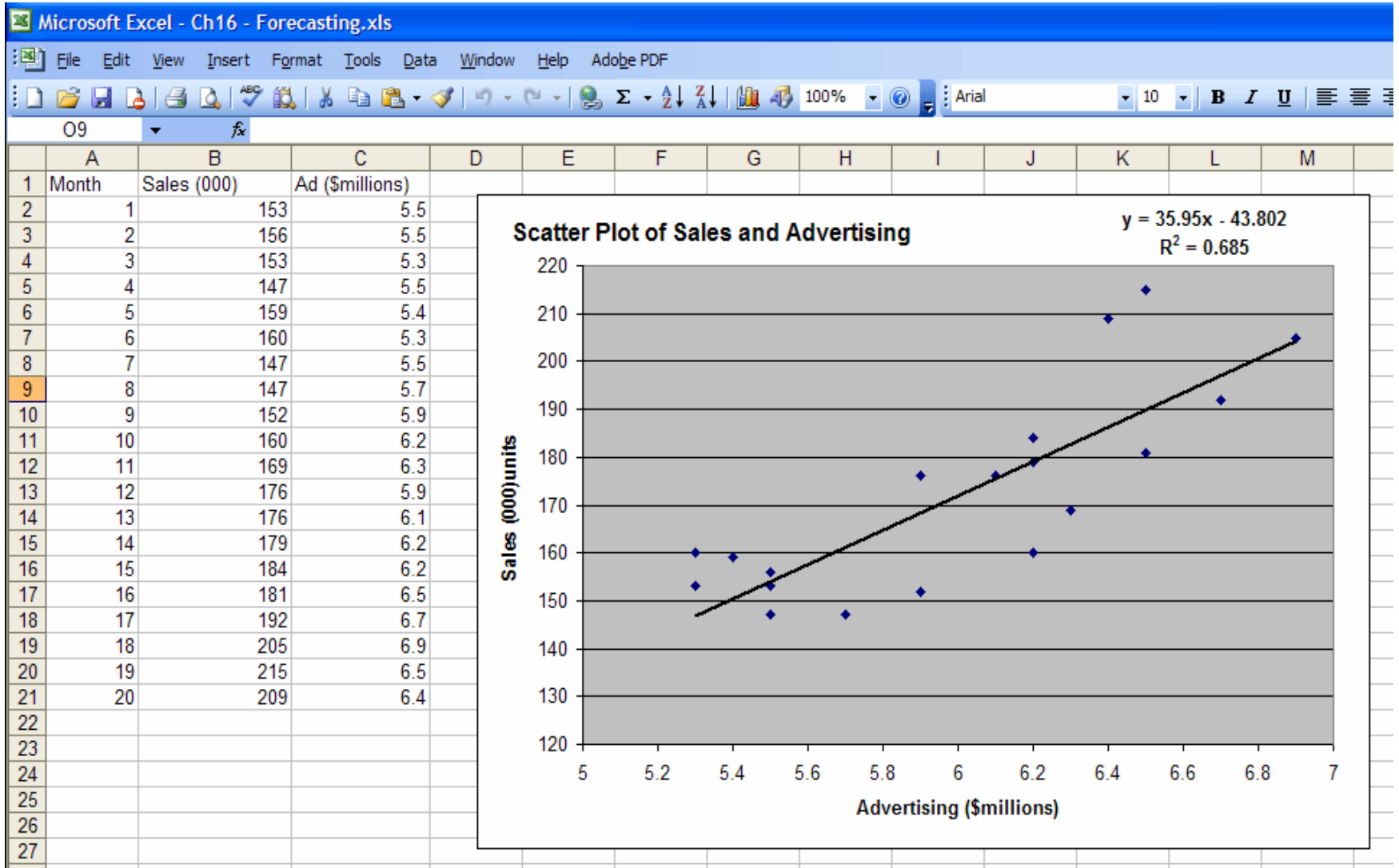
$$P(-1/96 \leq h \leq 1/96) = 0/95$$

یک مثال

آزمون خودهمبستگی برای داده های شرکت

Month	Sales (000)	Ad (\$millions)
1	153	5.5
2	156	5.5
3	153	5.3
4	147	5.5
5	159	5.4
6	160	5.3
7	147	5.5
8	147	5.7
9	152	5.9
10	160	6.2
11	169	6.3
12	176	5.9
13	176	6.1
14	179	6.2
15	184	6.2
16	181	6.5
17	192	6.7
18	205	6.9
19	215	6.5
20	209	6.4

قدم اول: ایجاد معادله رگرسیون



Microsoft Excel - Ch16 - Forecasting

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U \$ % ,

1	Month	Sales (000)	Ad (\$millions)	Predicted Sales (000)	Residuals	Lagged			
2		Y	X	\hat{Y}	$e_t = Y - \hat{Y}$	e_{t-1}	$(e_t - e_{t-1})^2$	$(e_t)^2$	
3	1	153	5.5	153.923	-0.923			0.8519	
4	2	156	5.5	153.923	2.077	-0.923	9.0000	4.3139	
5	3	153	5.3	146.733	6.267	2.077	17.5561	39.2753	
6	4	147	5.5	153.923	-6.923	6.267	173.9761	47.9279	
7	5	159	5.4	150.328	8.672	-6.923	243.2048	75.2036	
8	6	160	5.3	146.733	13.267	8.672	21.1140	176.0133	
9	7	147	5.5	153.923	-6.923	13.267	407.5361	47.9279	
10	8	147	5.7	161.113	-14.113	-6.923	51.6961	199.1768	
11	9	152	5.9	168.303	-16.303	-14.113	4.7961	265.7878	
12	10	160	6.2	179.088	-19.088	-16.303	7.7562	364.3517	
13	11	169	6.3	182.683	-13.683	-19.088	29.2140	187.2245	
14	12	176	5.9	168.303	7.697	-13.683	457.1044	59.2438	
15	13	176	6.1	175.493	0.507	7.697	51.6961	0.2570	
16	14	179	6.2	179.088	-0.088	0.507	0.3540	0.0077	
17	15	184	6.2	179.088	4.912	-0.088	25.0000	24.1277	
18	16	181	6.5	189.873	-8.873	4.912	190.0262	78.7301	
19	17	192	6.7	197.063	-5.063	-8.873	14.5161	25.6340	
20	18	205	6.9	204.253	0.747	-5.063	33.7561	0.5580	
21	19	215	6.5	189.873	25.127	0.747	594.3844	631.3661	
22	20	209	6.4	186.278	22.722	25.127	5.7840	516.2893	
23									
24							2338.5702	2744.2686	
25									
26									
27									

=43.802+35.95*C3

=(E4-F4)^2

=E4^2

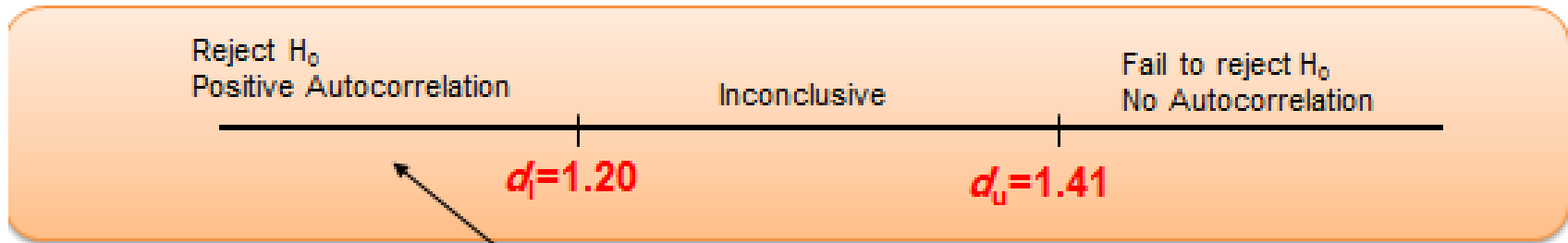
=B3-D3

=E3

$\sum(e_t - e_{t-1})^2$

$\sum(e_t)^2$

محاسبه DW از فرمول بیان شده:



$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n (e_t)^2} = \frac{2338.5829}{2744.2685} = 0.8522$$

با توجه به جدول مقادیر بحرانی دوربین-واتسون، کران بالا و پایین بدست می آید.

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76

2- آزمون ضرایب لاگرانژ (بروش-گادفری)

یک آزمون تعمیم یافته که خودهمبستگی از مرتبه اول تا مرتبه r ام و همچنین همبستگی بین e و متغیر توضیحی، را در نظر می گیرد:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_r u_{t-r} + v_t, \quad v_t \sim N(0, \sigma_v^2)$$

فروض:

$H_0: \rho_1 = 0, \rho_2 = 0, \dots, \rho_r = 0$ عدم خودهمبستگی

$H_1: \rho_1 \neq 0$ یا $\rho_2 \neq 0$ یا ... یا $\rho_r \neq 0$ وجود خودهمبستگی

مراحل انجام آزمون:

معادله موردنظر را با روش OLS برآورد می کنیم و جملات اخلا را استخراج می نماییم:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

e_t را روی متغیرهای توضیحی و خطاهای تاخیری برآزش کرده و R^2 آن را حساب می کنیم:

$$e_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \rho_1 e_{t-1} + \dots + \rho_r e_{t-r} + v_t$$

آماره آزمون محاسبه می گردد:

$$(n-r)R^2 \sim \chi_r^2$$

اگر آماره آزمون از ارزش های بحرانی جدول بیشتر بود، فرضیه صفر رد می شود.

روش های تخفیف خودهمبستگی

روش اولین تفاضل، روش دوربین، روش تکرارپذیر کوکران-اورکات و روش حداقل مربعات
تعمیم یافته

روش تکراری کوکران-اورکات

برای تخمین ضرایب در حالت خودهمبستگی مرتبه اول، مدل زیر را در نظر بگیرید:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

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

مدل را با یک دوره تاخیر نوشته و در ρ ضرب می کنیم

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho u_{t-1}$$

حال این معادله را از معادله اولی کم می کنیم

$$Y_t - \rho Y_{t-1} = \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

$$Y_t - \rho Y_{t-1} = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + v_t$$

با جایگذاری $X_t - \rho X_{t-1} = X_t^*$ و $Y_t - \rho Y_{t-1} = Y_t^*$ و $X_t - \rho X_{t-1} = X_t^*$ داریم:

$$Y_t^* = \beta_1^* + \beta_2 X_t^* + v_t$$

این مدل خودهمبستگی ندارد، لذا می توان آن را با روش OLS تخمین زد.

مراحل روش تکراری کوکران-اورکات:

1- ابتدا معادله $Y_t = \beta_1 + \beta_2 X_t + u_t$ را با روش OLS برآورد کرده و باقیمانده ها را حساب می کنیم.

2- بر اساس باقیمانده های مرحله اول، مدل $e_t = \rho e_{t-1} + v_t$ را برآورد می کنیم و تخمین $\hat{\rho}$ بدست می آید:

$$\hat{\rho} = \frac{\sum e_{t-1} e_t}{\sum e_{t-1}^2}$$

3- با استفاده از $\hat{\rho}$ مقادیر X_t^* و Y_t^* را حساب کرده سپس با روش OLS مدل را تخمین می زنیم.

4- تکرار مراحل 2 و 3 تا جاییکه مقدار $\hat{\rho}$ تقریباً ثابت بماند.

با تشکر از خانم زهرا تجری

نکته ای که در پایان لازم است به بحث فوق اضافه شود این مورد است یک جمع بندی است که پس از تشخیص خودهمبستگی چگونه آن را برطرف نماییم!؟

روش اول استفاده از روش حداقل مربعات تعمیم یافته GLS یا FGLS است و سایر روش های ذکر شده در اسلایدها در این زمینه مثل روش تکراری و غیره

روش دوم استفاده از اجزا AR و MA بر اساس رتبه حاصل از نمودار همبستگی نگار بر روی پسماندهای حاصله

روش سوم استفاده از رگرسیون مقاوم Robust است. HAC

پایان

Econometrics.blog.ir

مدیریت وب: حسین خاندانی و دکتر امیری

Econometrics1.blog.ir

وب کارگاه: حسین خاندانی و دکتر امیری

Econometrics2.blog.ir

مدیریت وب: علیرضا خیری