Making sense of Econometrics: Basics

Lecture 5: Autocorrelation

Hany Abdel-Latif & Anita Staneva

Egypt Scholars Economic Society

November 8, 2014

Assignment & feedback

enter classroom at

http://b.socrative.com/login/student/

room name c28efb78

Outline

1. Autocorrelation
   - meaning
   - sources
   - consequences
   - detection
   - example
2. What Next?

nature of autocorrelation

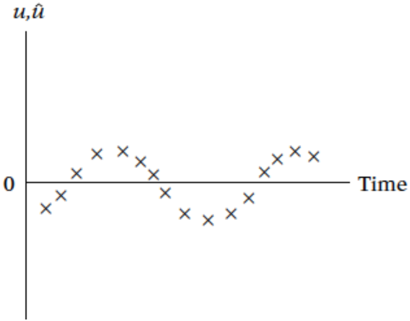
- correlation between members of series of observations ordered in time or space
- the CLRM assumes that
    - $E(u_i u_j) = 0$ for $i \neq j$
    - the disturbance term relating to any observation is not influenced by the disturbance term relating to any other observation
- example
    - there is no reason to believe that the disruption due to a labour strike affecting output in one quarter will be carried over to the next quarter
    - the effect of an increase of one family's income on its consumption expenditure is not expected to affect the consumption expenditure of another family

nature of autocorrelation

- if there is autocorrelation $E(u_i u_j) \neq 0$ for $i \neq j$
- example
  - the disruption caused by a strike this quarter may very well affect output next quarter
  - the increases in the consumption expenditure of one family may very well prompt another family to increase its consumption expenditure

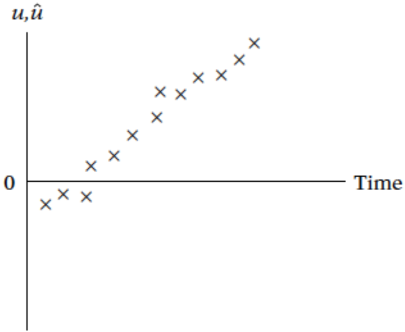
nature of autocorrelation

- a discernible pattern among the u's
  - a cyclical pattern



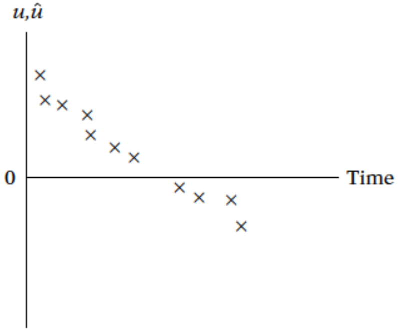
nature of autocorrelation

- a discernible pattern among the u's
  - an upward linear trend in the disturbances



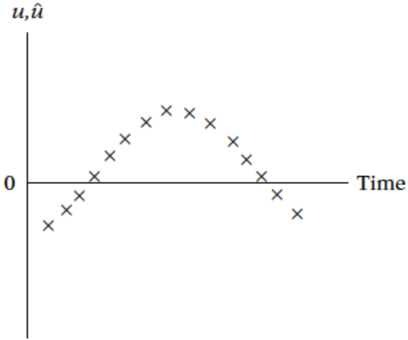
nature of autocorrelation

- a discernible pattern among the u's
  - a downward linear trend in the disturbances



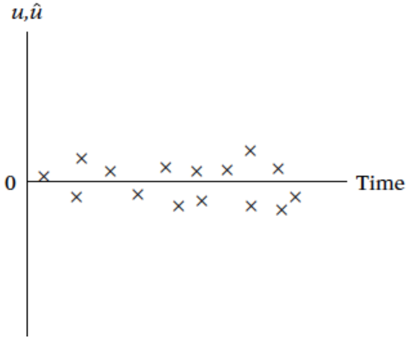
nature of autocorrelation

- a discernible pattern among the u's
  - a linear and quadratic trend terms are present in the disturbances



nature of autocorrelation

- a discernible pattern among the u's
  - no systematic pattern, supporting the non-autocorrelation assumption of the CLRM



sources of autocorrelation

1. inertia - business cycles
2. specification bias
   - excluded variables
   - incorrect functional form
3. Cobweb phenomenon
4. lags
5. manipulation of Data
6. data transformation
7. nonstationarity

أن كل عقل يفرق

consequences of autocorrelation

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

- $E(u_t, u_{t+s}) \neq 0$, for $s \neq 0$, is too general assumption
- for any practical use, we must assume the mechanism that generates u_t
- assume that the disturbance terms are generated by the following mechanism

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

- where ρ is known by the coefficient of autocovariance
- $-1 < \rho < 1$
- ϵ_t is the stochastic disturbance satisfies the standard OLS assumptions

consequences of autocorrelation

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

- an error term with the preceding properties is called a white noise error term
- the value of the disturbance term in period t is equal to ρ times its value in the previous period plus a purely random error term
- this scheme is known as Markov first-order autoregressive, usually denoted as $AR(1)$
- it is first order because u_t and its immediate past value are involved; the maximum lag is 1

consequences of autocorrelation

- if the model was

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t$$

- this is a second-order autoregressive scheme, and so on

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

- ρ, the coefficient of autocovariance, can also be interpreted as the first-order coefficient of autocorrelation, or more accurately, the coefficient of autocorrelation at lag 1

consequences of autocorrelation

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

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

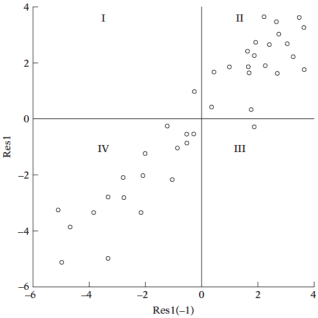
- it can be shown that
  - the usual OLS formulas to compute the variances and standard errors of β's could give seriously misleading results
  - $\hat{\beta}$ is still linear and unbiased but not BLUE
    - $\hat{\beta}$ does not have minimum variance

detection of autocorrelation

- two ways in general
  - informal (graphically)
  - formal (Durbin-Watson)
- examine the OLS residuals $\hat{u}_i$
  - since they are the ones we observe, and not the disturbances u_i
  - one hopes that they are good estimates of u_i
  - a hope that may be fulfilled if the sample size is fairly large

informal method

- there are various ways of examining the residuals
  - plot $\hat{u}_t$ against time; or
  - plot $\hat{u}_t$ against $\hat{u}_{t-1}$
- if the residual are non-random, you will obtain pictures similar to what we have shown before



formal method

Durbin-Watson test

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2}$$

assumptions underlying the d statistic

- regression includes the intercept term
- explanatory variables are nonstochastic; fixed in repeated sampling
- u_t are generated by $AR(1)$ scheme; cannot be used to detect higher-order autoregressive scheme

formal method

Durbin-Watson test

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} \hat{u}_t^2}$$

assumptions underlying the d statistic

- u_t is normally distributed
- regression does not include lagged values of the dependent variable as one of the explanatory variables
- no missing observations in the data

formal method

Durbin-Watson test

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} \hat{u}_t^2}$$

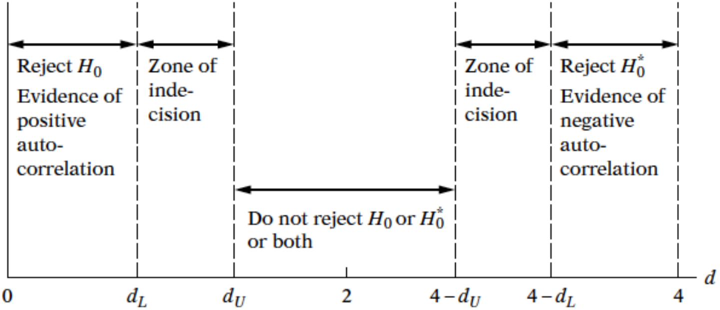
- $d \approx 2(1 - \hat{\rho})$; where $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$
- since that $-1 < \rho < 1$, then $2 \leq d \leq 4$
- these are the bounds of d; any estimated d value must lie within these limits

hypotheses

- H_0 : no positive autocorrelation
- H_0^* : no negative autocorrelation

formal method

how to make a decision?



formal method

Durbin-Watson

1. run OLS regression and obtain residuals
2. compute d, Stata gives d
3. for the given sample size and given number of explanatory variables, find out the critical d_L and d_U values
4. follow the decision role shown before

example

- suppose we calculated $d = 0.1229$ for a given regression model with $n = 40$ and one explanatory variable
- from the Durbin-Watson tables, $d_L = 1.44$ and $d_U = 1.54$ at $\alpha = 5\%$
- d lies below d_L, we can not reject the null hypothesis that there is positive serial correlations in the residuals

data

- data come from (Klein, 1950)

```
use http://www.stata-press.com/data/r13/klein
des
```

variable name	storage type	display format	value label	variable label
yr	float	%9.0g		year
consump	float	%9.0g		consumption
profits	float	%9.0g		private profits
wagepriv	float	%9.0g		private wage bill
invest	float	%9.0g		investment
capital1	float	%9.0g		lagged value of capital stock
totinc	float	%9.0g		total income/demand
wagegovt	float	%9.0g		government wage bill
govt	float	%9.0g		government spending
taxnetx	float	%9.0g		indirect bus taxes + net export
wagetot	float	%9.0g		total US wage bill
year	float	%9.0g		calendar year - 1931
profits1	float	%9.0g		last year's private profits
totinc1	float	%9.0g		last year's total income/demand

model

```
tab yr
```

- we have 22 observations from 1920 to 1941
- suppose we want to fit an OLS model of consumption on the government wage bill

$$consump_t = \beta_1 + \beta_2 wagegovt_t + u_t$$

```
tsset yr
```

```
time variable:  yr, 1920 to 1941
        delta:  1 unit
```

estimation

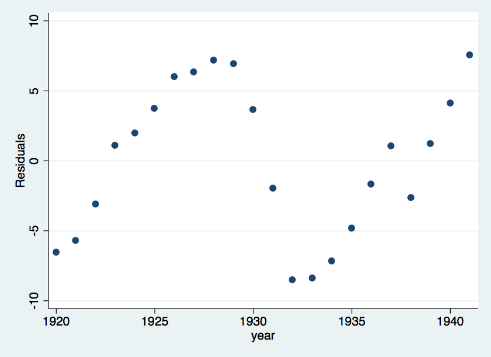
```
reg consump wagegovt
```

```
      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  1,    20) =   17.72
       Model |  532.567711     1  532.567711           Prob > F      =  0.0004
    Residual |  601.207167    20  30.0603584           R-squared     =  0.4697
-------------+------------------------------           Adj R-squared =  0.4432
       Total |  1133.77488    21  53.9892799           Root MSE      =  5.4827

------------------------------------------------------------------------------
     consump |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    wagegovt |    2.50744   .5957173     4.21   0.000     1.264796    3.750085
       _cons |   40.84699   3.192183    12.80   0.000     34.18821    47.50577
------------------------------------------------------------------------------
```

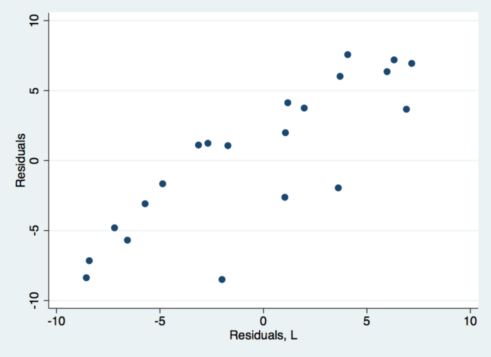
examine residuals: informal

```
predict e, residual
scatter e yr
```



examine residuals: informal

```
scatter e l.e
```



examine residuals: Durbin-Watson

```
estat dwatson
```

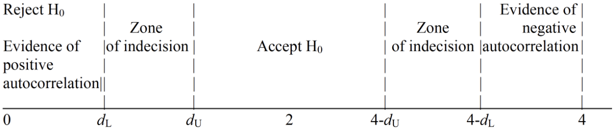
```
Durbin-Watson d-statistic(  2,    22) =  .3217998
```

- find critical values for $n = 22$ and one explanatory variable at $\alpha = 5\%$

obs	k'=1		k'=2		k'=3		k'=4		k'=5	
N	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813
34	1.993	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799

examine residuals: Durbin-Watson

- $d_L = 1.239$ and $d_U = 1.429$



- $d = 0.32$ is far from the center of its distribution $d = 2$
- reject the null of no first-order autocorrelation
- evidence of positive autocorrelation

remedies

- if we assume that *wagegovt* is strictly exogenous, we could refit the model using *arima*, and model the error term explicitly
- if not willing to assume that *wagegovt* is strictly exogenous, use Durbin's alternative test or the Breusch–Godfrey to test for first order autocorrelation

remedies

```
estat durbinalt, small
```

Durbin's alternative test for autocorrelation

lags(p)	F	df	Prob > F
1	35.035	(1, 19)	0.0000

H0: no serial correlation

```
estat bgodfrey, small
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	F	df	Prob > F
1	14.264	(1, 19)	0.0013

H0: no serial correlation

remedies

- both tests strongly reject the null of no first-order autocorrelation
- we can refit the model with two lags of *consump* included as regressors and then run the test again
- not that because the revised model includes lagged values of the dependent variable, Durbin-Watson test is not applicable

remedies

```
regress consump wagegovt L.consump L2.consump
```

Source	SS	df	MS
Model	702.660311	3	234.220104
Residual	85.1596011	16	5.32247507
Total	787.819912	19	41.4642059

Number of obs = 20
F(3, 16) = 44.01
Prob > F = 0.0000
R-squared = 0.8919
Adj R-squared = 0.8716
Root MSE = 2.307

consump	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
wagegovt	.6904282	.3295485	2.10	0.052	-.0081835	1.38904
consump						
L1.	1.420536	.197024	7.21	0.000	1.002864	1.838208
L2.	-.650888	.1933351	-3.37	0.004	-1.06074	-.241036
_cons	9.209073	5.006701	1.84	0.084	-1.404659	19.82281

remedies

```
estat durbinalt, small lags(1/2)
```

Durbin's alternative test for autocorrelation

lags(p)	F	df	Prob > F
1	0.080	(1, 15)	0.7805
2	0.260	(2, 14)	0.7750

H0: no serial correlation

```
estat bgodfrey, small lags(1/2)
```

Breusch-Godfrey LM test for autocorrelation

lags(p)	F	df	Prob > F
1	0.107	(1, 15)	0.7484
2	0.358	(2, 14)	0.7056

H0: no serial correlation

remedies

- *wagegov* and the constant term are no longer statistically different from zero at the 5% level
- including the two lags of *consump* has removed any serial correlation from the errors

Next Lecture

you should know

- assignment 2
  - due on November 22 (20:00 Cairo time)
  - must attach the cover sheet
  - to be emailed to `eses@egyptscholars.org`

next lecture

- Saturday November 15 (20:00 Cairo time)
- regression in practice
  - multicolinearity
  - specification bias

References

Klein, L. R. (1950), 'Economic fluctuations in the united states, 1921-1941'.