(1) Nature of autocorrelation

G. 416

5) Specification bias: incorrect functional form

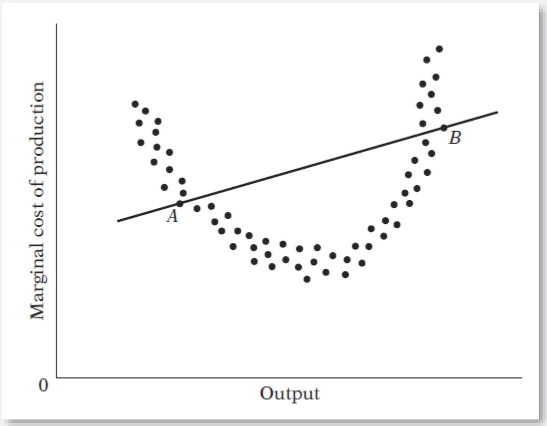
For example, a curved marginal cost which the ‘true’ model is

- $MC_i = \beta_1 + \beta_2 Q_i + \beta_3 Q_i^2 + u_i$

but if we try to fit our data with a linear model instead

- $MC_i = \beta_1 + \beta_2 Q_i + u_i$

Marginal cost model



Between point A and B, linear model underestimates marginal cost while before point A and beyond point B, the model overestimates marginal cost.

If we correlate u_i, u_j we can see clearly that there is a pattern following the curve, hence autocorrelation is present in the linear model.

However, this problem does not surface when we fit the data with polynomial model.

From this point on, we will focus on a time-series model, varying Y_t and X_t through time, not across groups of observation in the same period, the model becomes

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

(2) Effect on estimation

If the error terms are correlated, assumed linearly

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

This equation is called **first-order autoregressive scheme** or shortly **AR(1)**. If the error term t also correlates with two period back, the model is called **AR(2)**, and so on.

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

Normally, OLS with no autocorrelation will yield the variance of an estimator as

- $Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$

If u_t follows AR(1), the variance becomes

- $Var(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}[1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \ldots + 2\rho^{n-1} \frac{\sum x_t x_n}{\sum x_t^2}]$

We cannot say for sure whether $Var(\hat{\beta}_2)$ is more or less than $Var(\hat{\beta}_2)_{AR(1)}$. Though $\hat{\beta}_2$ is still linear and unbiased, variance is not minimum, or not being efficient. See this proof of GLS on page 422.

Regressing disregarding autocorrelation

- $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$ is likely to underestimate the true σ^2.
- Overestimation of R^2.
- If σ^2 is not underestimated, $Var(\hat{\beta}_2)$ may still underestimate $Var(\hat{\beta}_2)_{AR(1)}$, and the latter is still inefficient compared to $Var(\hat{\beta}_2)_{GLS}$.
- Both t and F tests are no longer valid.

(3) Detecting autocorrelation

G. 430

1) Graphical method

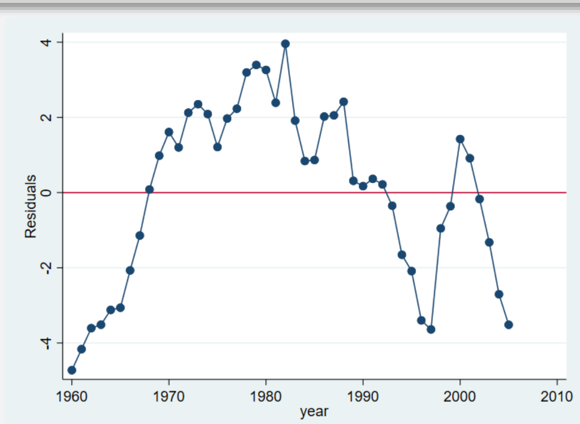
The first one, once again, is not informal but visually telling. If we plot the residuals with time period, we might see interconnection between time period.

Example: Data of real wage (Y_t) and output per hour (X_t) in the US are taken from the book page 428. Each observation is taken yearly from 1960 to 2005, We replicate the regression exactly

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

Now see the result of the estimation on the right-hand side. After that, we can predict $\hat{u}_t$ and plot them over time.

Detecting autocorrelation using a plot of residuals over time



Disregarding autocorrelation

```
. reg rwage output

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(1, 44)        =   1830.24
       Model |  10406.1452         1  10406.1452   Prob > F        =    0.0000
    Residual |  250.169395        44  5.68566807   R-squared       =    0.9765
-------------+----------------------------------   Adj R-squared   =    0.9760
       Total |  10656.3145        45   236.80699   Root MSE        =    2.3845

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .6704057   .0156705    42.78   0.000     .6388238    .7019876
       _cons |    32.7419   1.394021    23.49   0.000     29.93244    35.55137
------------------------------------------------------------------------------
```

If autocorrelation is not present, residuals should be randomly distributed around 0 and has no obvious interconnection with the previous period.

The results also suggest autocorrelation due to its highly significance of t and F, thus R^2.

(3) Detecting autocorrelation

G. 434

2) Durbin-Watson *d* Test

Durbin-Watson ***d*** statistics is defined as

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

implies sum of squared differences to the RSS.

For Durbin-Watson test, we assume

- Regression model includes intercept term.
- Regressors X are stochastic or fixed in repeated sampling.
- u_t are generated by AR(1) and normally distributed.
- The model does not include lagged variable(s) of the dependent variable, such as

$Y_t = \beta_1 + \beta_2 X_t + \gamma Y_{t-1} + u_t$

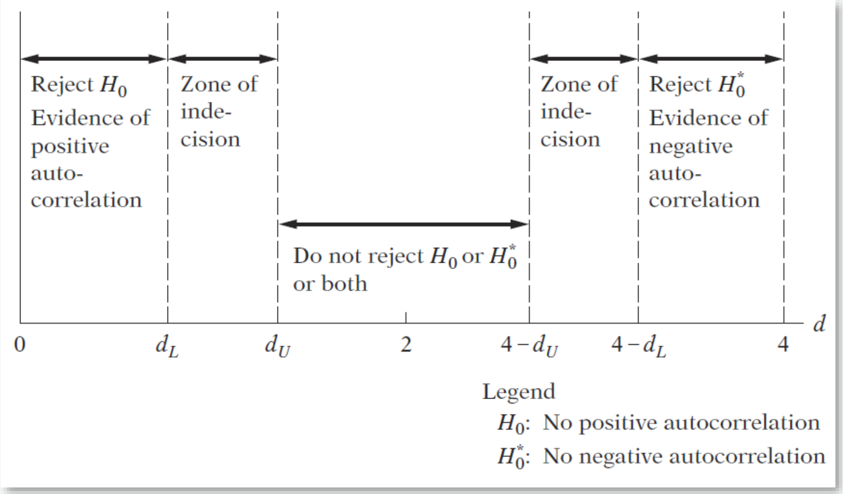
- No missing observations in the data.

We can also define 'approximate' version of the d stat as

- $d \approx 2(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2})$ now let's define
- $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$ then
- $d \approx 2(1 - \hat{\rho})$

Since $-1 \leq \hat{\rho} \leq 1$, therefore $0 \leq d \leq 4$

Testing autocorrelation with d test



Example and steps for the real wage and output model are as follows.

(3) Detecting autocorrelation

G. 436

Real wage and output model

```
reg rwage output

    Source |       SS           df       MS      Number of obs   =        46
-----------+----------------------------------   F(1, 44)        =   1830.24
     Model |  10406.1452         1  10406.1452   Prob > F        =    0.0000
  Residual |  250.169395        44  5.68566807   R-squared       =    0.9765
-----------+----------------------------------   Adj R-squared   =    0.9760
     Total |  10656.3145        45   236.80699   Root MSE        =    2.3845

------------------------------------------------------------------------------
     rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
    output |   .6704057   .0156705    42.78   0.000     .6388238    .7019876
     _cons |    32.7419   1.394021    23.49   0.000     29.93244    35.55137
------------------------------------------------------------------------------

estat dwatson

Durbin-Watson d-statistic(  2,    46) =  .1738879
```

Step 1: State the hypothesis

Since we are testing for general autocorrelation, without any speculation of positive or negative correlation, we are going to set hypothesis for both.

- H_0: No autocorrelation ($d_U < d < 4 - d_U$)
- H_a: Positive autocorrelation ($0 < d < d_L$) or

negative autocorrelation ($4 - d_L < d < 4$)

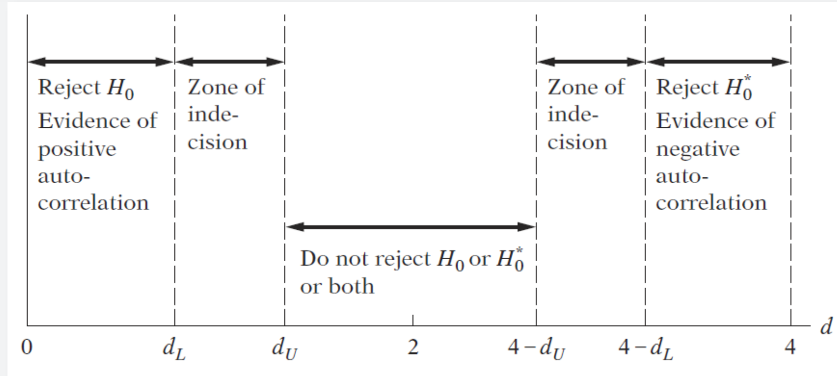
Do not forget that we can also reach the inconclusive answer for this test.

Step 2: Run the OLS regression and obtain the residuals.

Step 3: Calculate the *d* statistics. Two 'attributes' that are important for finding d_L and d_U are number of coefficients and number of observation.

- $d_L =$ and $4 - d_L =$
- $d_U =$ and $4 - d_U =$

Testing autocorrelation with d test



Step 4: Conclude the test.

(3) Detecting autocorrelation

G. 439

3) General test of autocorrelation: The Breusch-Godfrey test (BG)

Assume a model of

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

Now if the error term u_t follows the pth-order autoregressive AR(P) as follows

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

Main concept is to test these coefficients simultaneously by following the steps.

Step 1: State the hypothesis

- H_0: No autocorrelation ($\rho_1 = \rho_2 = \cdots = \rho_p = 0$)
- H_a: Autocorrelation (ρ are not simultaneously zero)

Step 2: Run the OLS regression and obtain the residuals.

Step 3: Estimate this equation, also including regressor(s) into this one.

- $\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t$

Then obtain the R^2 from this model

BG test

```
. reg rwage output

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(1, 44)        =   1830.24
       Model |  10406.1452         1  10406.1452   Prob > F        =    0.0000
    Residual |  250.169395        44  5.68566807   R-squared       =    0.9765
-------------+----------------------------------   Adj R-squared   =    0.9760
       Total |  10656.3145        45   236.80699   Root MSE        =    2.3845

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .6704057   .0156705    42.78   0.000     .6388238    .7019876
       _cons |    32.7419   1.394021    23.49   0.000     29.93244    35.55137
------------------------------------------------------------------------------

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |         34.618               1                   0.0000
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Step 4: Calculate LM-statistics, if the sample is large, BG has shown that

- $(n - p)R^2 \sim \chi^2_p$

Step 5: Look for the critical value in chi-square table and reject the null hypothesis if the LM exceeds the critical value.

Note that, in STATA, the default lag is 1 but there is an option to include more lags.

(4) Remedial measures

G. 443

First of all, make sure that the model is correctly specified (pure autocorrelation). Most of the time we do not know the relationship between u_t and u_{t-1}. In other words, we do not know the value of ρ.

1) First-difference method

The first difference equation takes the form of

- $Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$ or
- $\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t$ where $\varepsilon_t = \Delta u_t$

The rule of thumb is that we can use this equation to estimate when $d < R^2$.

Note that this model has no intercept term. If included, the intercept is interpreted as **time trend**.

As we can see on the result on the right-hand side, when we use the first-difference model, we cannot reject the null hypothesis of BG test any longer because ε_t is a stationary white-noise error term.

However, note that we lose 1 observation due to using difference term.

First-difference method

```
. reg d1.rwage d1.output

      Source |       SS           df       MS      Number of obs   =        45
-------------+----------------------------------   F(1, 43)        =     23.42
       Model |  23.0747588         1  23.0747588   Prob > F        =    0.0000
    Residual |  42.3572406        43  .985052108   R-squared       =    0.3527
-------------+----------------------------------   Adj R-squared   =    0.3376
       Total |  65.4319995        44   1.4870909   Root MSE        =     .9925

------------------------------------------------------------------------------
     D.rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |
         D1. |   .5497398   .1135843     4.84   0.000     .3206753    .7788043
             |
       _cons |   .2596131   .2643692     0.98   0.332    -.2735383    .7927645
------------------------------------------------------------------------------

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          2.243               1                   0.1342
---------------------------------------------------------------------------
                        H0: no serial correlation

. reg d1.rwage d1.output, nocons

      Source |       SS           df       MS      Number of obs   =        45
-------------+----------------------------------   F(1, 44)        =    102.14
       Model |  100.532825         1  100.532825   Prob > F        =    0.0000
    Residual |  43.3071681        44  .984253821   R-squared       =    0.6989
-------------+----------------------------------   Adj R-squared   =    0.6921
       Total |  143.839993        45   3.1964443   Root MSE        =     .9921

------------------------------------------------------------------------------
     D.rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |
         D1. |   .6421773   .0635411    10.11   0.000     .5141187    .7702359
------------------------------------------------------------------------------
```

(4) Remedial measures

2) Estimating ρ

The reason why we want to know the value of ρ is that we can use this value to transform variables, then use the transformed ones in the GLS estimation.

Assume that the error term follows AR(1) scheme,

- $u_t = \rho u_{t-1} + \varepsilon_t$ where $1 < \rho < 1$

we transform $t-1$ equation by multiply ρ

- $\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1}$

Then, subtract the t with the equation above to remove the coexistence of the element that are the same between time

- $(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$ or
- $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$

There are several methods that we can estimate this value of ρ.

2.1) ρ based on Durbin-Watson d statistics.

From the Durbin-Watson test, we can derive that

- $\hat{\rho} \approx 1 - \frac{d}{2}$

2.2) ρ estimated from the residuals

We can also estimate another model postestimation,

- $\hat{u}_t = \rho\hat{u}_{t-1} + \nu_t$

2.3) Cochrane-Orcutt iterative procedure

Iterative method estimates ρ by starting at some value, mostly 0, then successively approximate multiple times until the value of ρ is stable. Then, ρ can be put into the transformation.

2.4) Prais-Winsten transformation

Using the same concept of iterative ρ, Prais-Winsten fixed losing 1 observation from the first-difference method because of no antecedent by adding

- $Y_1\sqrt{1-\rho^2}$ and $X_1\sqrt{1-\rho^2}$

(4) Remedial measures

Disregarding autocorrelation

```
reg rwage output

    Source |       SS           df       MS      Number of obs   =        46
-----------+----------------------------------   F(1, 44)        =   1830.24
     Model |  10406.1452         1  10406.1452   Prob > F        =    0.0000
  Residual |  250.169395        44  5.68566807   R-squared       =    0.9765
-----------+----------------------------------   Adj R-squared   =    0.9760
     Total |  10656.3145        45   236.80699   Root MSE        =    2.3845

------------------------------------------------------------------------------
     rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
    output |   .6704057   .0156705    42.78   0.000     .6388238    .7019876
     _cons |    32.7419   1.394021    23.49   0.000     29.93244    35.55137
------------------------------------------------------------------------------

estat dwatson

Durbin-Watson d-statistic(  2,    46) =  .1738879
```

Note that Prais-Winsten transformation will retain original number of observation, which might be very important especially for a small sample dataset.

Cochrane-Orcutt and Prais-Winsten method

```
. prais rwage output, corc

Cochrane-Orcutt AR(1) regression -- iterated estimates

      Source |       SS           df       MS      Number of obs   =        45
-------------+----------------------------------   F(1, 43)        =    193.55
       Model |  160.769164         1  160.769164   Prob > F        =    0.0000
    Residual |  35.7178906        43  .83064862    R-squared       =    0.8182
-------------+----------------------------------   Adj R-squared   =    0.8140
       Total |  196.487054        44  4.46561487   Root MSE        =     .9114

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .5722474   .0411331    13.91   0.000     .4892947    .6552002
       _cons |   42.97793   4.315771     9.96   0.000     34.27435    51.68152
-------------+----------------------------------------------------------------
         rho |   .8809751
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.173888
Durbin-Watson statistic (transformed) 1.631290

. prais rwage output

Prais-Winsten AR(1) regression -- iterated estimates

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(1, 44)        =    316.44
       Model |  311.554182         1  311.554182   Prob > F        =    0.0000
    Residual |  43.3203419        44  .984553225   R-squared       =    0.8779
-------------+----------------------------------   Adj R-squared   =    0.8752
       Total |  354.874524        45  7.88610054   Root MSE        =    .99225

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .6628417   .0386835    17.14   0.000     .5848803    .7408032
       _cons |   32.04334   3.718262     8.62   0.000     24.54968    39.53701
-------------+----------------------------------------------------------------
         rho |   .9193386
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.173888
Durbin-Watson statistic (transformed) 1.512079
```

(4) Remedial measures

G. 447

3) Newey-West method

This method does not deal with autocorrelation directly, instead, it is very much like White's robust standard error.

The corrected standard errors are known as **HAC standard error.** (heteroscedasticity and autocorrelation).

Newey-West approach is strictly speaking valid in large samples.

Comparing normal model to the Newey-west method

```
. reg rwage output

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(1, 44)        =   1830.24
       Model |  10406.1452         1  10406.1452   Prob > F        =    0.0000
    Residual |  250.169395        44  5.68566807   R-squared       =    0.9765
-------------+----------------------------------   Adj R-squared   =    0.9760
       Total |  10656.3145        45   236.80699   Root MSE        =    2.3845

------------------------------------------------------------------------------
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .6704057   .0156705    42.78   0.000     .6388238    .7019876
       _cons |    32.7419   1.394021    23.49   0.000     29.93244    35.55137
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  2,    46) =  .1738879


. newey rwage output, lag(1)

Regression with Newey-West standard errors      Number of obs     =        46
maximum lag: 1                                  F(  1,        44) =    777.86
                                                Prob > F          =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
       rwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      output |   .6704057   .0240373    27.89   0.000     .6219616    .7188498
       _cons |    32.7419   2.269985    14.42   0.000     28.16705    37.31676
------------------------------------------------------------------------------
```