PS2030

Political Research and Analysis

Unit 2: Regression Models: Problems and Extensions

2. Multicollinearity, Heteroskedasticity, Autocorrelation

Spring 2025, Weeks 6-7

WW Posvar Hall 3600

Professor Steven Finkel

More Assumptions of OLS

- No perfect correlation between independent variables ("multicollinearity")
- Population error term ε has following properties:
  - The expected value of ε is 0, or $E(\varepsilon_i)=0$
  - The expected covariance of X and ε is 0, or $E(X_i\varepsilon_i)=0$
  - "Homoskedasticity" (equal error variance) $E(\sigma_i^2)=\sigma^2$ for all X_i
  - No "Autocorrelation", or $E(\varepsilon_i\varepsilon_j=0)$ for all *i* and *j*
  - Residuals are normally distributed, i.e., $\varepsilon_i \sim N$
- This session and next devoted to violations (or near-violations) of these assumptions: how to detect them, what their consequences are for OLS, how to deal with them to obtain best ("BLUE") estimates of population regression parameters

A. Multicollinearity

- Have seen previously that perfect "collinearity" exists when one X is a perfect linear combination of other Xs (e.g. three dummy variables corresponding the three categories of a nominal outcome variable – any two of the dummies are perfectly correlated with the third
- This is a violation of OLS assumption (and indeed every other estimation method). You can't get unique effects of one variable since it is perfectly related to other variable(s). The multivariate OLS formula is undefined, given 0 in the denominator:

$$\beta_1(x_1) = \left(\frac{SD_y}{SD_{x_1}}\right)\left(\frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}\right)$$

- In other instances, though, the correlation approaches 1 and causes problems. This is *not* a violation of assumptions, but it is a situation that can lead to unreliable estimates (large standard errors) and incorrect inferences if we are not careful

- Can see this with formula for standard error of the slope

$$\hat{\sigma}_{b_1} = \sqrt{\frac{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{N-k-1}}{\Sigma(X_i - \bar{X})^2(1 - r^2_{x_1x_2})}}$$

  - Or more general multivariate formula:

$$\hat{\sigma}_{b_1} = \sqrt{\frac{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{N-k-1}}{\Sigma(X_i - \bar{X})^2(1 - R^2_{1.k})}}$$

  - As can see, when variables are very highly correlated, the standard error gets bigger and bigger until it finally becomes undefined only at R=1.
  - This leads to problems in statistical inference: you may obtain an insignificant *t* value even though there is a true effect of a given variable (Type II error)

- Further Comments
  - OLS estimates in the presence of multicollinearity are still unbiased and efficient. OLS estimates of β will, on average, neither under- or overestimate the true β
  - The variance of OLS estimates (standard errors) are still the minimum possible
  - If you have specified the variables properly, there is no violation of OLS specification error assumptions
  - ε is still uncorrelated with X
  - Multicollinearity does not imply anything about heteroskedasticity or autocorrelated errors either
- Therefore, *nothing* in the multicollinearity situation that, in itself, violates OLS assumptions unless R=1

- Seriousness of the multicollinearity problem depends on the nature and purpose of the model and the analysis
  - If you are interested in, say, the effect of X_1 and you include X_2 and X_3 as control variables, then it is not important that X_2 and X_3 are highly collinear.
  - So long as X_1 is not highly related to the other variables, then no problem at all
  - If you are interest in *prediction*, not explanation, then again multicollinearity is not a problem. You still get unbiased estimates of the β and hence accurate predictions of $E(Y|X)$
- The *only* time multicollinearity is a problem is when the purpose of the analysis is to make inferences about the effects of a given X, and that X is highly correlated with other Xs.

- The problem with multicollinearity is not of OLS assumptions about the population. It is a problem of lack of information in your *sample*
  - Not enough information in sample about variation in Y that is associated with changes in each X to make accurate estimates
  - In an experiment, for example, we can randomly include individuals with different combinations of Xs (and assign them to be "treated" and "control") to minimize covariation between the Xs and with "treatment"
  - In real world samples, can't do that so left with "deficient sample" where correlation between the Xs is too high for precise estimates to be made
    - Example: all college educated people are wealthy in our sample, all non-college educated people are poor; therefore we don't have enough information on college/poor and non-college/rich people to estimate the unique effects of education and income
  - For this reason, multicollinearity is sometimes called a problem of "micronumerosity" (inadequate sample size)

- This situation leads to problems in estimating effects in your sample
  - You often get strange estimates of the slopes. Why? Not enough independent variation in each X, so estimates are sensitive to one or two points
  - Again, this is the sample problem. Our estimate of the effect of, e.g., income will be based on those very few rich non-college people and those very few college-poor people, and so, from sample to sample as we obtain information on very few of these people, the estimates will jump around
  - You get very large standard errors (and possible statistical insignificance) for the same reasons
- Statistically: multicollinearity means that the estimates of the β themselves will be correlated, the more so, the more highly correlated the variables are. Sometimes this produces estimates of the wrong sign, again due to limited information that your sample is providing.

Detecting Multicollinearity

- Inspect bivariate correlation: .7 and .8 is starting to get high. But there is no threshold level where you can say that you definitely have a problem
- Sub-sample within your sample and run regressions on both. If you see huge differences in estimates, then possible multicollinearity
- Run regression of each X on all other Xs – to extent that R^2 is high, then high multicollinearity.
  - Variance Inflation Factor (VIF) = $1/(1-R_{12}^2)$
  - "Tolerance" = $1-R^2$ of this equation
  - Can see that as R^2 goes up, VIF goes up and Tolerance goes down
  - Some suggest that when VIF hits 10 you have a multicollinearity problem
- Look for pattern of a highly significant F, but insignificant individual t test results. However, don't use this as an excuse for not getting significant results!

What to Do About Multicollinearity?

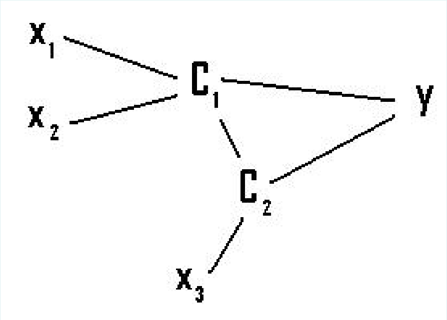
- Nothing if it is not theoretically important or if your interest is in prediction
- Get more information!!! (Solve the "micronumerosity" problem!) Increasing N will:
  - offset multicollinearity problems, since part of standard error formula is based on number of cases
  - Also, more information raises possibility that you will get more *independent* variation in each of the Xs. So increasing sample size is your best bet if you can!
- Throw out one of the offending variables? Not recommended, as you are then willfully committing specification error, and, since the offending variable is by definition highly related to the variable you leave in, you will almost certainly estimate the included variables' effects with bias!

- Better solution: Conduct a joint hypothesis tests with F*

$$F^* = \frac{\dfrac{ESS(R) - ESS(F)}{df(R) - df(F)}}{\dfrac{ESS(F)}{df(F)}}$$

Where (R) is the “reduced model” without including the collinear variables, and (F) is the “full model”

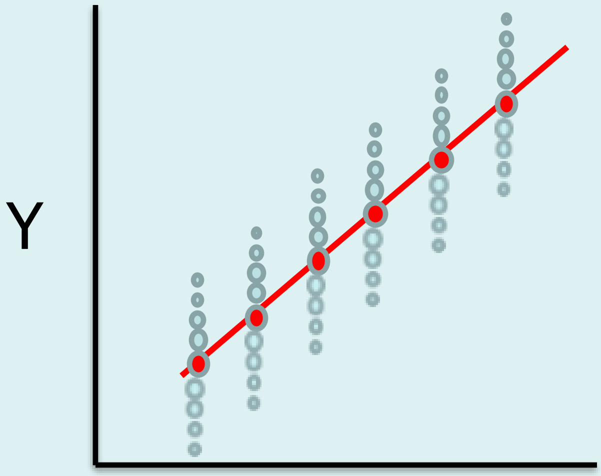
- If, and only if, the two offending variables are measures of the same underlying concept, can make a scale which combines the two things together.
  - The scale will be a better measure than each indicator taken separately anyway, so this is ideal solution. Either an additive scale or a “latent factor” scale through path analysis
  - Example: Education and Income as a “SES factor” (C_1)



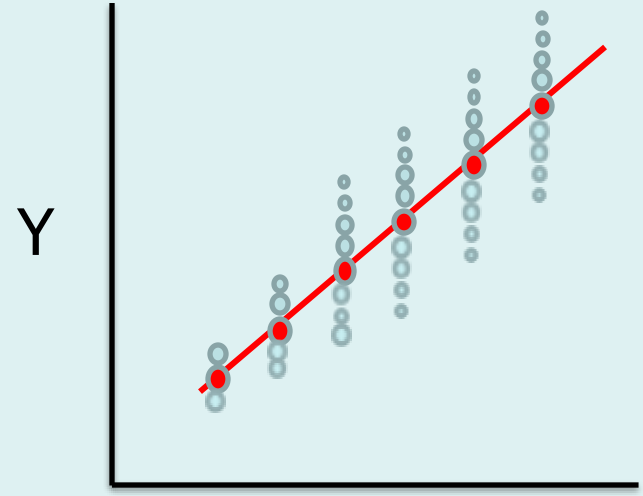
  - If the underlying concept does not encompass both indicators, however, then it is nonsense to make scale from them

B. Heteroskedasticity

When error variances are unequal for different observations



Homoskedasticity

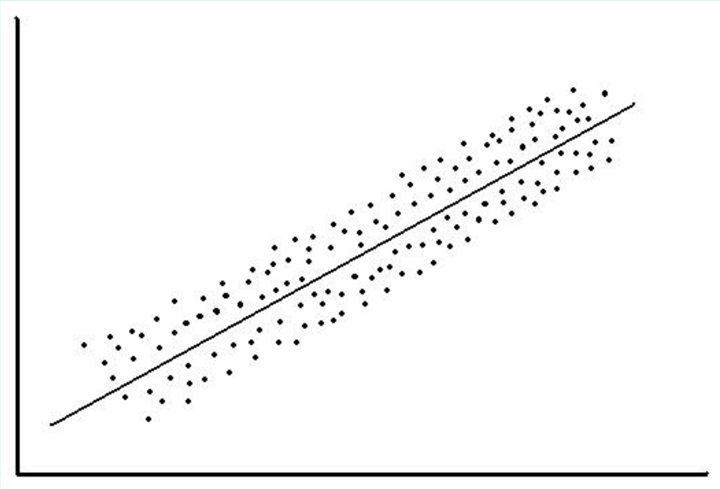


Heteroskedasticity

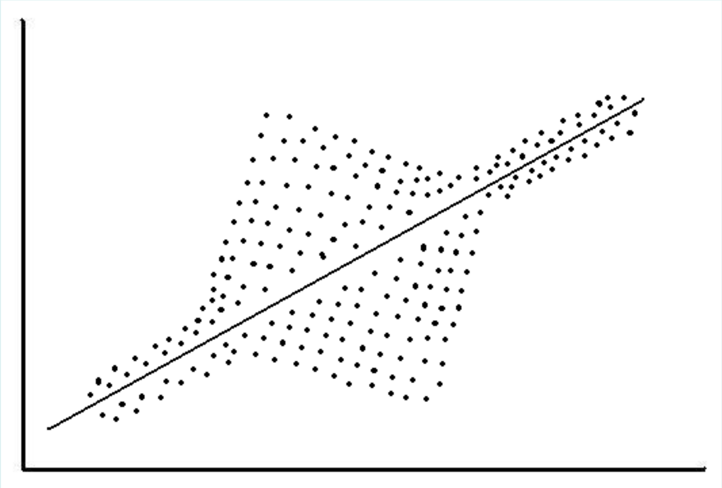
Homoskedasticity assumption of OLS: E(σ^2)=σ^2 for all X_i

Heteroskedasticity

Error variances are unequal for different observations



Homoskedasticity



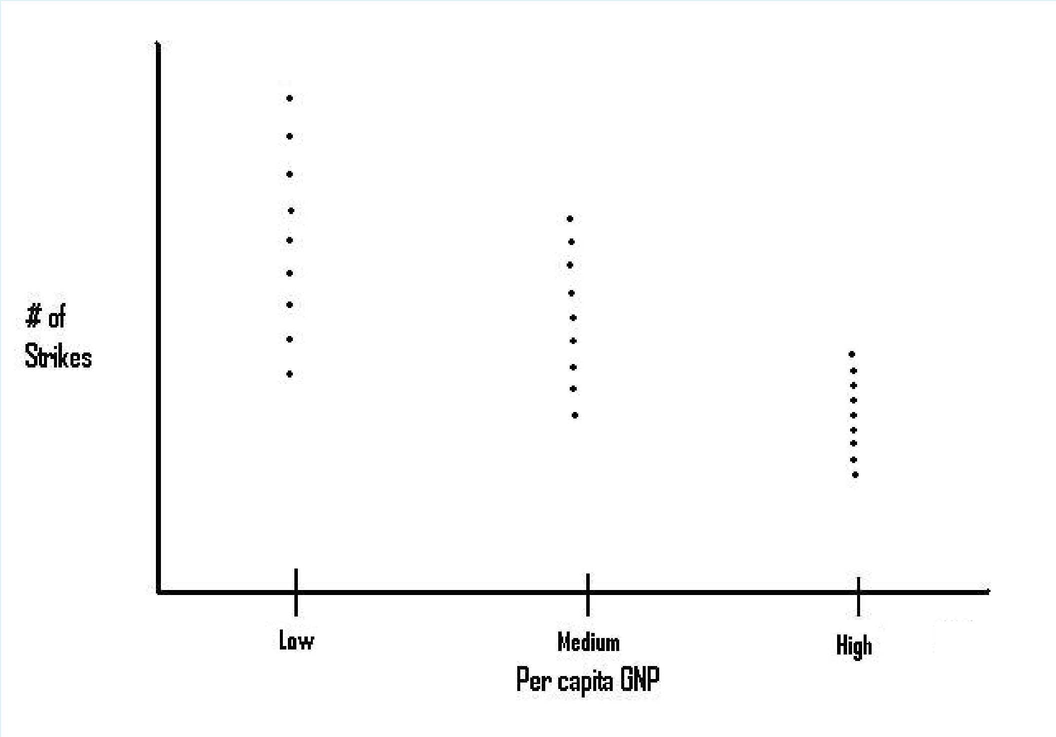
Heteroskedasticity

Summary of Problems and Solutions

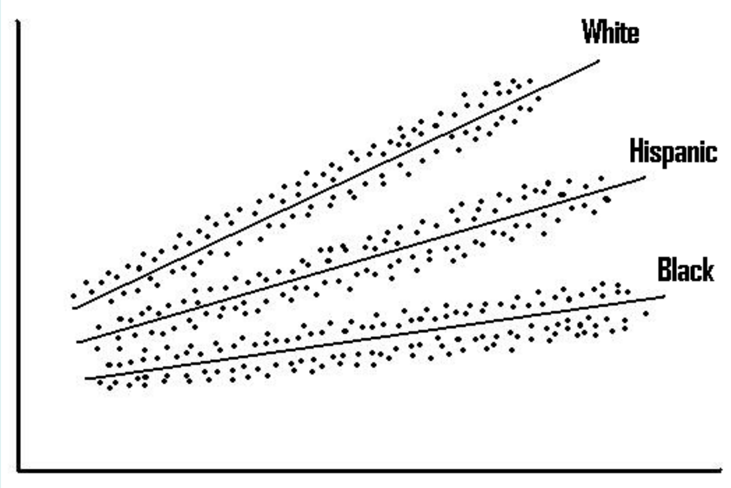
- OLS is still "unbiased" – will neither on average overestimate or underestimate the population slope value β
- OLS ignoring heteroskedasticity will yield **incorrect standard errors**, so hypothesis tests will no longer be valid
- Can correct OLS to take heteroskedasticity into account, but resultant standard errors will no longer be "BLUE", not "efficient"
- If know the **exact** form of the heteroskedasticity, can use "generalized least squares" (GLS) or "weighted least squares" solution
- If don't know exact form of the heteroskedasticity, can:
  - use "feasible generalized least squares" (FGLS) or "weighted least squares" (WLS) solution if have sense of the general form
  - use OLS with "heteroskedastic consistent", or White-Huber "robust" standard errors if agnostic completely about the form of the heteroskedasticity. **This is implemented in STATA via the "robust" option and R in various ways (e.g. coeftest, lm_robust)**
- Logic and procedures parallel the situation for autocorrelated errors

When will Heteroskedasticity be Present?

1. Measurement error in y that is systematic for different groups
   - If we have data measured on countries, and data collected is more reliably in some countries than others, will have more error in these countries with improper measures

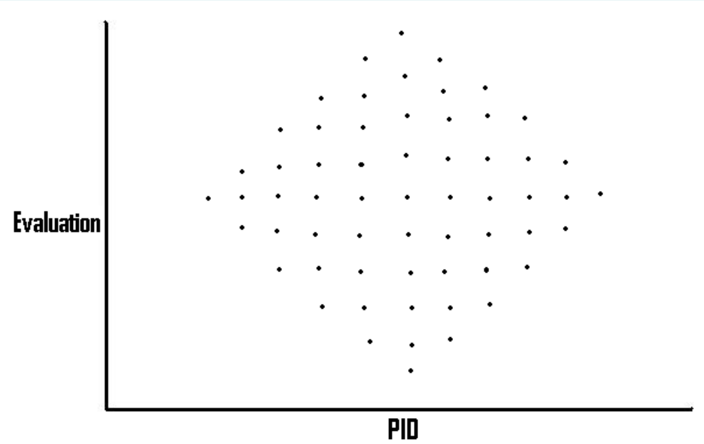


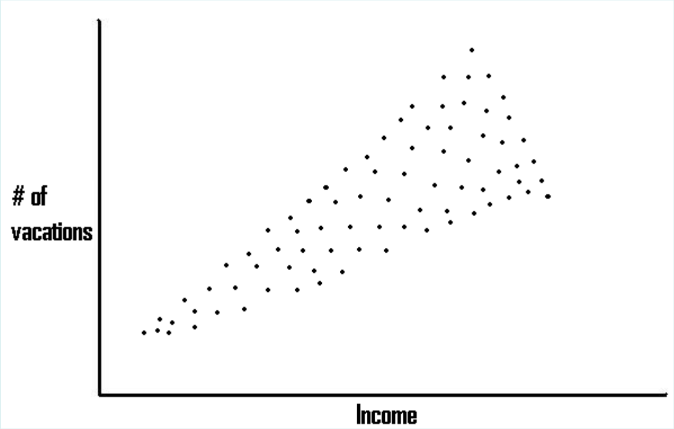
2. There may be small excluded variable effects that are presumed to be related to x, but related like an interaction dummy variable, such that there are different slopes for different values of the excluded variable.



- Without including the slope dummies, you would see heteroskedasticity.
- In this case the correction is obvious – include the variable!

3. Theoretical reasons for suspecting different error variance at different levels of X

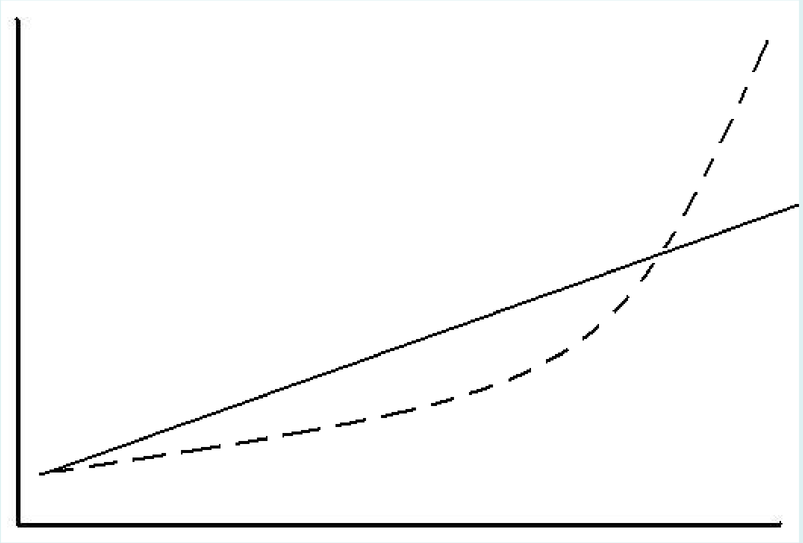




Some units display more variance in outcomes than others, depending on things such as developmental processes, group heterogeneity, and the like

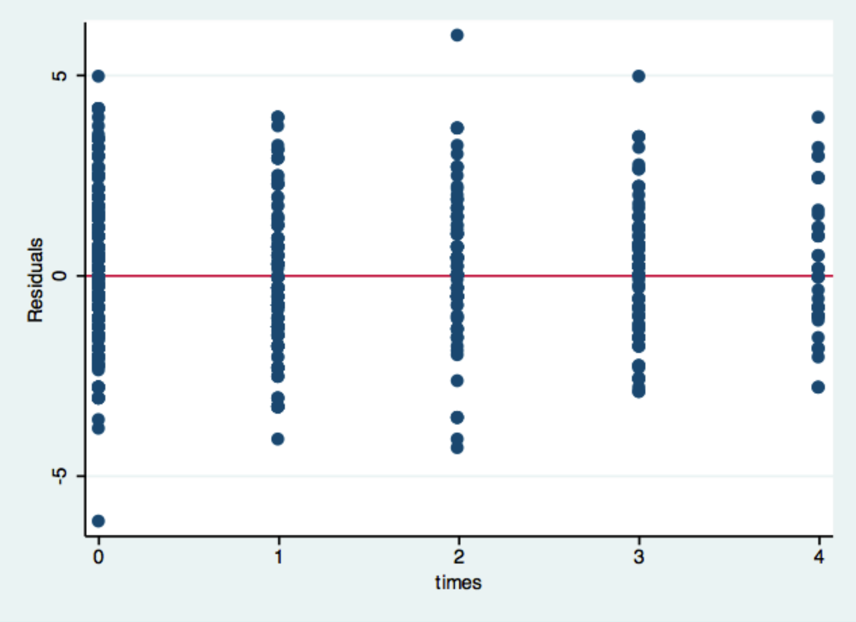
4. Functional form problems

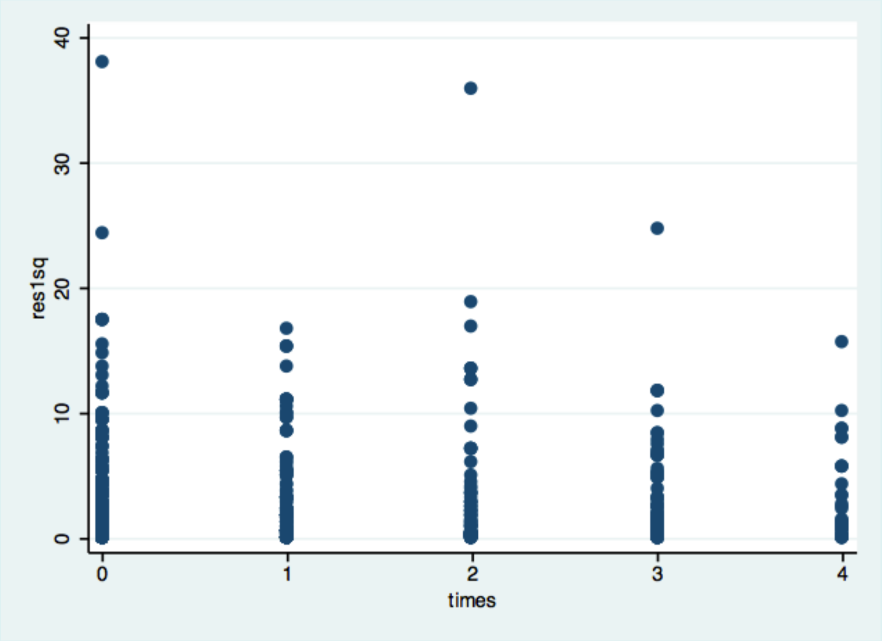
- If a linear form is being used to appropriate a non-linear relationship, then there will be ranges for x where estimates consistently over- or under-estimate the values.
- Result: Unequal variance
- Again, “easy” solution”: model the non-linearity!



Detection and Tests for Heteroskedasticity

- Visual inspection of OLS residuals or squared residuals as our best guess of the population residuals

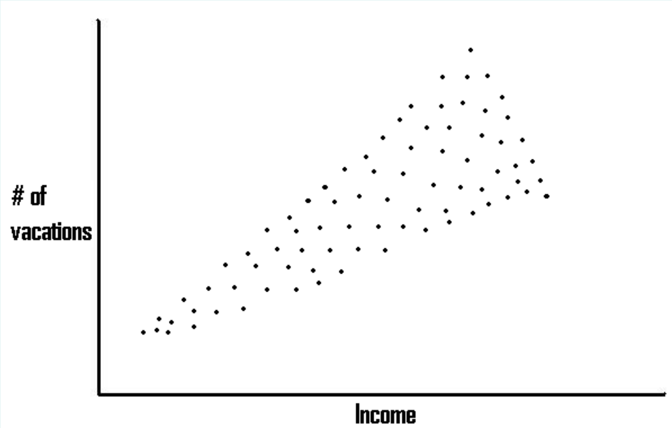




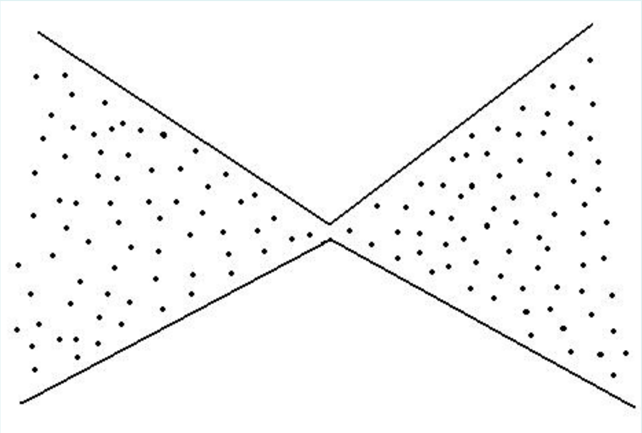
Residuals against X

Squared Residuals against X

- Goldfield-Quandt test: if believe that heteroskedasticity follows "classic" fan shaped form



not:



- Delete some number of observations *m* around the mean of x. Larger the number, the more powerful the test. Left with (N-m) observations
- Split remaining sample into two parts and run OLS on both – Get error variance for both halves and do a F test on the ratio of the variances –
- If equal, F = 1, and not significantly different. If unequal, the F → ∞, and difference of variances will be significant at the .05 level or beyond.

Breusch-Pagan and White Tests

- Breusch-Pagan (BP)
  - Take squared OLS residuals and regress on predicted Y
  - If homoskedastic, this regression should have zero R^2
  - Test is a chi-square test with 1 df
  - Implemented in STATA with "hettest" or "estat hettest" (with more options in the latter specification such as relaxing normality assumption in the residuals)
  - Example: South Africa regression of KNOW with TIMES and EDUC1 (resids plotted on slide 20)

```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of know

         chi2(1)      =    25.53
         Prob > chi2  =   0.0000
```

- White test
  - Regress squared OLS residuals against: IVs, squared IVs, cross-products (interactions) of IVs (and possibly cubed, quadratic IVs)

    $$\hat{\varepsilon}^2 = f(x_1, x_2, x_1^2, x_2^2, x_1 x_2)$$

  - The R^2 from this regression, multiplied by N, is distributed as chi-square with *k* degrees of freedom

    $$N * R^2 \sim x^2$$

  - Problems: Could also be due to specification error and not heteroskedasticity, plus low "power" of the test if many IVs Probably most popular of current tests.
  - STATA: "estat imtest, white"

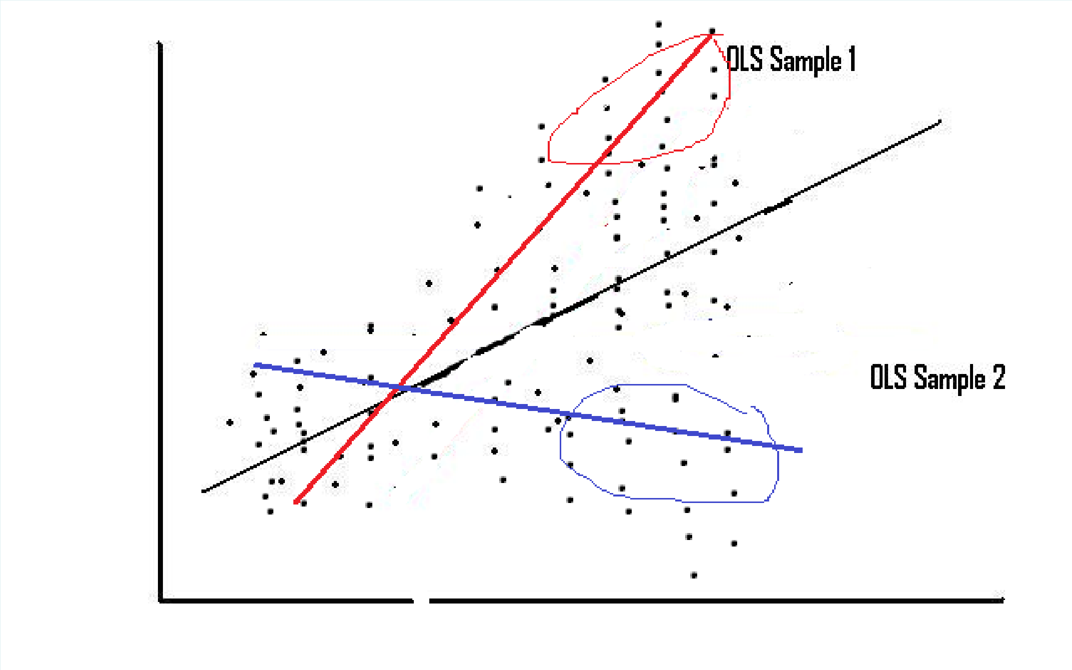
```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(5)      =     25.42
         Prob > chi2  =    0.0001
```

Consequences of Heteroskedasticity

- OLS β remain unbiased – why?
  - Estimate of "b" depends on sampling at large values of X
  - If get lots of high values by chance, line will tilt upward; if get lots of low values by chance, line will tilt downward
  - But on average will neither over- or under-estimate β



- But: variance of the slope (the standard error of β) will be **incorrect** using OLS formula
- Therefore all statistical tests will be incorrect!!!
- If OLS assumptions satisfied in population, the variance of the slope is : $$\frac{\sigma^2}{\Sigma(X_i - \bar{X})^2}$$
- But given heteroskedasticity in the population, the "true" variance of the slope is: $$\frac{\Sigma(X_i - \bar{X})^2 \sigma_i^{\,2}}{(\Sigma(X_i - \bar{X})^2)^2}$$
- So using "regular OLS" to estimate the standard error (i.e., the square root of the variance) will produce incorrect estimates

- Can also see that estimates of β vary quite a bit from sample to sample, despite being unbiased
- This means that, even if we were to use the correct variance (standard error) formula in the previous slide, the OLS procedure ("estimator") itself may not be "BLUE" in the sense of efficiency or minimum variance
- We may be able to find estimators that have less variance from sample to sample, smaller standard errors, greater chances for rejecting the null hypothesis, etc.
- That is:

$$\frac{\Sigma(X_i - \bar{X})^2 \hat{\sigma}^2_{i-OLS}}{(\Sigma(X_i - \bar{X})^2)^2}$$

is bigger than it "needs" to be, we might be able to do better

- Summary for Heteroskedasticity
  - OLS not biased
  - Estimate of error variance via standard OLS calculation is wrong, leading to incorrect estimates of the standard error of the slope
  - Correcting this problem is possible, but may still result in OLS β no longer being as efficient as other estimators

Corrections for Heteroskedasticity

- If you know or can ascertain that heteroskedasticity is present due to non-linearities or some other model misspecification, then by all means correct the problem by changing your model!
- If the heteroskedasticity is simply due to the scale of the independent variable (e.g., as dollars increase, the variance in outcomes increases as well), you can try log transformations of the variables as a possible solution
- Problems, though, in turning the model into a fully logged model if not theoretically plausible

Corrections for Heteroskedasticity, Continued

- If you know the **exact** form of the heteroskedasticity, i.e., you know the error variance at every level of X, then you can estimate parameters via a **Generalized Least Squares (GLS)** procedure that weights the model by the error variance for each X_i

$$\frac{Y_i}{\sigma_i} = \beta_0(\frac{1}{\sigma_i}) + \beta_1(\frac{X_i}{\sigma_i}) + (\frac{\varepsilon_i}{\sigma_i})$$

- Since the σ^2 are known for each *i*, the variance of the GLS error term is constant (and equals 1) for all levels of X
- Intuition: We give less weight to cases that have larger σ_i . This yields less variance in the GLS estimate of β than standard OLS procedure

Feasible Generalized Least Squares

- Usually we don't know σ_i^2, however. If we are willing to make some assumptions about the general form of the heteroskedasticity, we can estimate a "Feasible GLS" (FGLS) through the same kind of weighting procedure to yield more efficient estimates of the standard errors.
- This is sometimes just referred to as "**Weighted Least Squares**"
- Assume, for example, that the heteroskedasticity follows the classic fan-shape pattern. We might characterize the pattern as:

$$\mathrm{E}(\varepsilon_i^2) = \sigma_i^2 = kx_i^2$$

- Where the size of the error variance is proportional (by some constant value *k)* to the square of X

- Weighted Least Squares – divide the entire equation through by X_i

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \text{ where } \sigma_i^2 = kX_i^2$$

- Then:

$$\frac{Y_i}{X_i} = \beta_0 \left(\frac{1}{X_i}\right) + \beta_1 \left(\frac{X_i}{X_i}\right) + \left(\frac{\varepsilon_i}{X_i}\right)$$

- This equation's error term is now homoskedastic, equal to the constant term *k*:

$$\mathrm{var}\left(\frac{\varepsilon_i}{X_i}\right) = \frac{\sigma_i^2}{X_i^2} = \frac{kX_i^2}{X_i^2} = k$$

- So we could estimate this "weighted" equation with OLS, since it now satisfies the OLS assumption of homoskedasticity!

- Weighted Least Squares: minimizes the ***weighted*** sum of squared residuals
  - OLS minimizes $\Sigma(y_i - a - bx_i)^2$
  - WLS minimizes $\Sigma w_i(y_i - a - bx_i)^2$
  - where $w_i = \dfrac{1}{x_1^2}$
  - The weight is the inverse of the (proportional) error variance
    - So larger values of X get less weight than smaller values, since we know that the error variance is higher at those values, and that given points are likely to be farther away from the population regression function
    - Smaller values of X get greater weight, since we know that error variance is low at those values and that given points are likely to be closer to the PRF
    - This improves efficiency of the WLS estimator over OLS, which weights all observations equally.
    - Multiple regression: Divide all variables through by the "offending" variable, or follow more exact procedure outlined in Woolridge, *Introductory Econometrics* on pp. 283-284

White's Heteroskedastic-Consistent Standard Errors

- If don't know the form of the heteroskedasticity, then WLS won't help (and can even hurt if you model the form incorrectly)
- Alternative procedure: use OLS and then correct the standard errors
- Recall slide 25: With heteroskedasticity, the true variance of the slope is:

$$\frac{\Sigma(X_i - \bar{X})^2 \sigma_i^2}{(\Sigma(X_i - \bar{X})^2)^2} \quad \text{not} \quad \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2}$$

- White (1980) suggests estimating the correct variance with the OLS residuals

$$\frac{\Sigma(X_i - \bar{X})^2 \hat{\sigma}_i^2}{(\Sigma(X_i - \bar{X})^2)^2}$$

- In large samples, this estimate converges to the true value of the variance/standard error (hence the **"heteroskedastic-*consistent*"** phrase to describe it)
- In multiple regression context, the relevant formula for the standard error of variable J is:

$$\frac{\Sigma(\hat{r}_{ij.}^{2}\hat{\sigma}_{i}^{2})}{\Sigma(r_{ij.}^{2})^{2}}$$

where $\hat{r}_{ij.}$ is the residual for case *i* in a regression of variable J against all other independent variables

- These are also called **"robust standard errors"** and now done as a matter of course in contemporary political science
- But if no heteroskedasticity, they are less desirable than OLS, especially in small samples. And they are less efficient than WLS-FGLS **if** the form of the heteroskedasticity is known

```
. regress know times educ1

      Source |       SS       df       MS              Number of obs =     940
-------------+------------------------------           F(  2,   937) =  239.41
       Model |  1199.10829     2  599.554145           Prob > F      =  0.0000
    Residual |  2346.48745   937  2.50425555           R-squared     =  0.3382
-------------+------------------------------           Adj R-squared =  0.3368
       Total |  3545.59574   939  3.77592731           Root MSE      =  1.5825

------------------------------------------------------------------------------
        know |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       times |   .2470176   .0441621     5.59   0.000     .1603496    .3336856
       educ1 |   .7735603    .038075    20.32   0.000     .6988381    .8482824
       _cons |   .7447649   .1235979     6.03   0.000     .5022042    .9873256
------------------------------------------------------------------------------

. regress know times educ1, robust

Linear regression                                      Number of obs =     940
                                                       F(  2,   937) =  212.55
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3382
                                                       Root MSE      =  1.5825

------------------------------------------------------------------------------
             |               Robust
        know |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       times |   .2470176    .045377     5.44   0.000     .1579652      .33607
       educ1 |   .7735603   .0409446    18.89   0.000     .6932064    .8539141
       _cons |   .7447649   .1191949     6.25   0.000     .5108451    .9786847
------------------------------------------------------------------------------
```

C. Autocorrelation

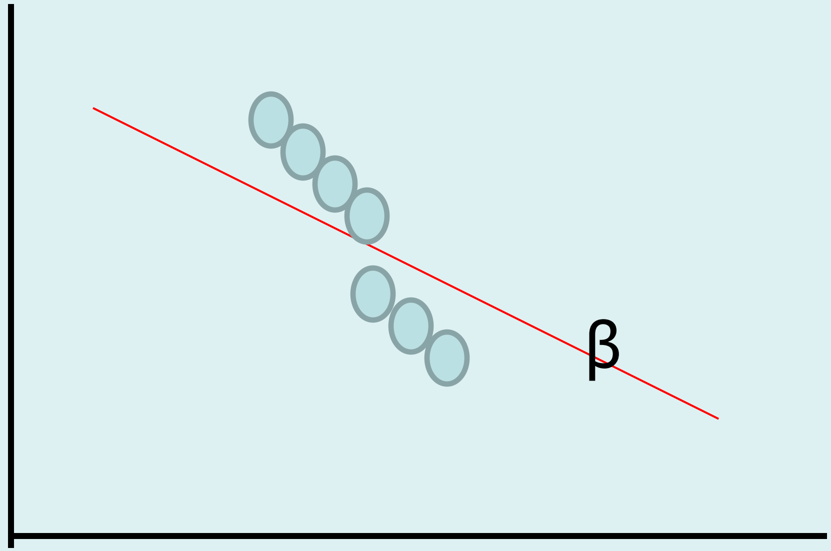
- OLS population error term assumption: residual for one observation is assumed to be unrelated to the population error term of all other observations, i.e., $E(\varepsilon_i\varepsilon_j=0)$ for all *i* and *j*
- This may not be true because, in the population:
  - Units that are ***spatially*** near one another may have error terms that are related as well
  - Units that are ***temporally*** near one another may have error terms that are related as well

– The former problem can occur in cross-sectional analyses of macro-level data or micro-level survey data, or other kinds of data

– The latter problem is associated with *longitudinal analysis*, either:
  - *Time-series analysis*, which involves the analysis of data sets with many time points but very few (and often only one) unit
  - *Panel analysis*, data sets with relatively few time points and many units

– We will focus here on the time-series case; panel case later

Example: Police Budget and Crime Rates in One U.S. City



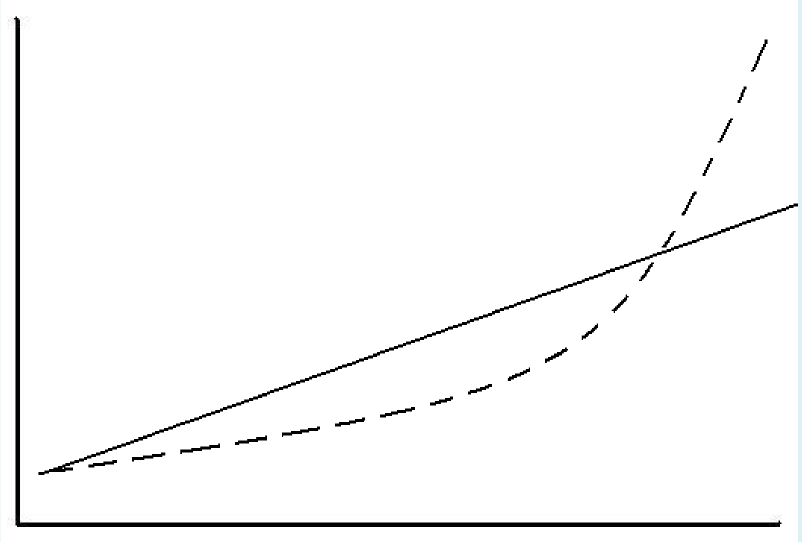
A series of positive residuals, followed by negative, then possibly shift again

Sources of Autocorrelated Errors

- Omitted variables that are relatively stable over time (best solution: include them in the model!)
- Omitted exogenous shocks that linger for several time periods
  - In the crime rate example: weather-related disturbances, idiosyncratic patterns of gang warfare, the extent of drug trafficking in the city, temporary spikes in drug prices, etc., all of which may affect crime rates over time, independently from both the main theoretical measured variables
- Correlated measurement errors
- "Inertia" and other effects of the lagged DV
  - If the true model should have lagged Y as a predictor and you omit it from consideration, the error term will be autocorrelated as a result
  - Some time series models will include the lagged DV in part as a "control" for autocorrelation, which indicates the close relationship of these concepts (and the difficulty of disentangling them at times)

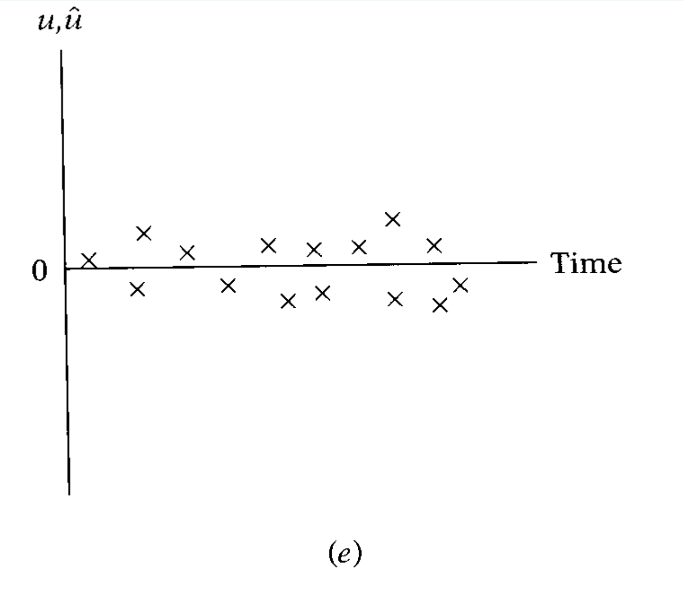
Sources of Autocorrelated Errors (Continued)

- Functional form misspecification
  - If a linear form is being used to appropriate non-linear relationship, then there will be ranges for x where estimates consistently over or under estimate the values. In time-series data, this will show up as autocorrelated disturbances – a series of negative residuals followed by positive ones, for example
  - "Easy" solution, just like with heteroskedasticity: model the non-linearity!

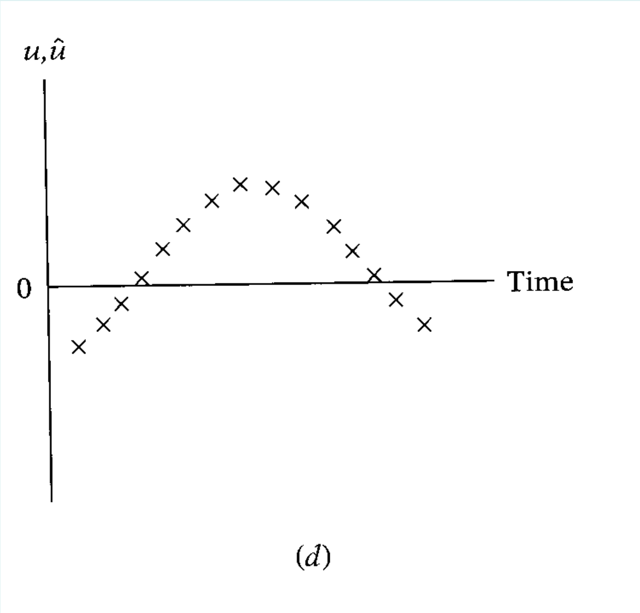


Detecting Autocorrelated Errors

- Examine OLS residual plots or standardized residual plots against time

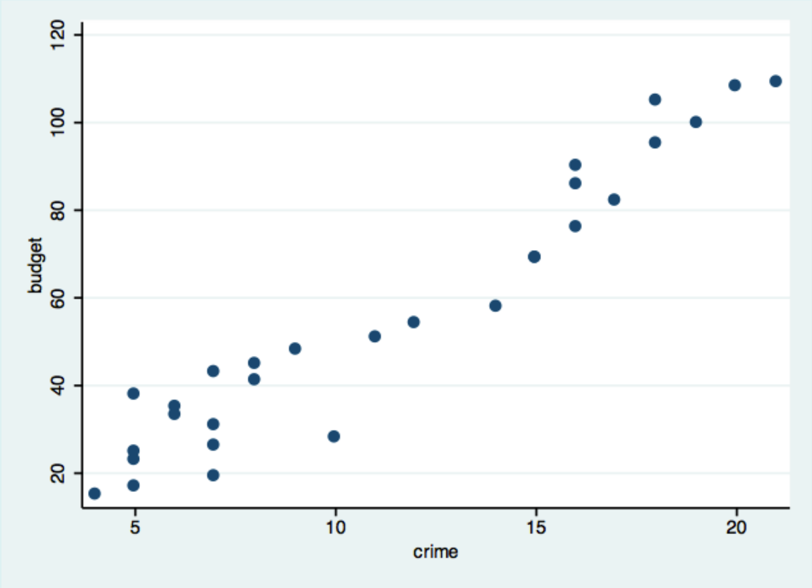


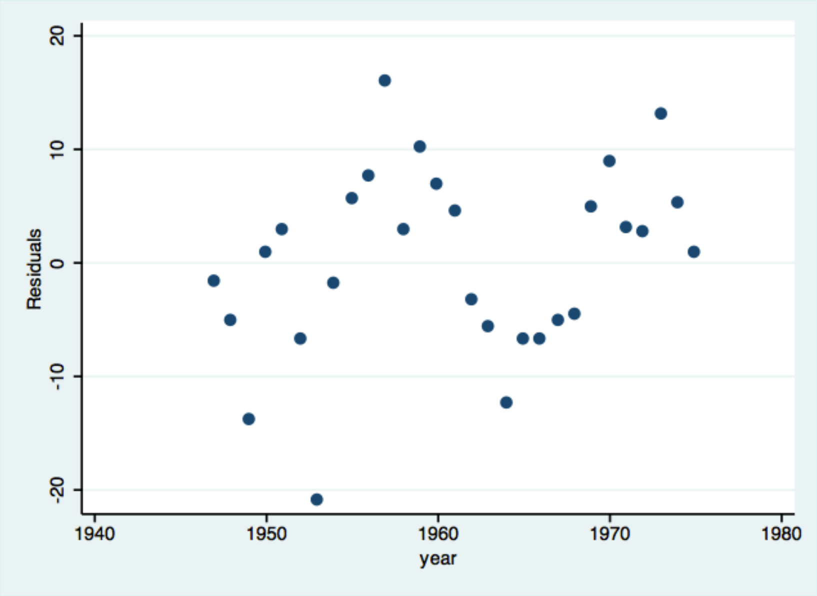
VS.



- The "Runs Test": Do we have a pattern of successive positive and then negative residuals that would be more than would be expected by chance?

Stata Example: 29 years, Police Budget as a Function of the Crime Rate





Durbin-Watson *d* Test

- If we assume that the residuals follow a "first order autoregressive pattern" (AR(1)), where

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

and v_t is "well-behaved" (normally distributed, non-autocorrelated, purely random residual), then autocorrelation in the ε is commonly tested via the Durbin-Watson *d* statistic

- Note: many other kinds of autocorrelation aside from first-order
  - Second and third, etc. order effects
  - "moving average" autocorrelation, where the error is equal to some average value (say, *v*) along with multiple lags of *v* so that the "average" is updated each time period
- Note also that the *d* test cannot be used in models with the lagged dependent variable.

- Basic idea of D-W *d*: Use OLS residuals "e" as our best guess of the population residuals ε
- Construct *d* as

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

- It can be seen that if one error term is highly (positively) related to another, *d* will be relatively small as the numerator will be small
- If there is no positive autocorrelation, *d* gets relatively high, as the numerator will be large

- Mathematical logic of the test: *d* can be shown to be equal to 2(1-ρ), and since rho is bounded by -1 and 1, *d* is bounded by 0 and 4

$$d = \frac{\Sigma e_t^2 + \Sigma e_{t-1}^2 - 2\Sigma e_t e_{t-1}}{\Sigma e_t^2}$$

$$d \approx \frac{2\Sigma e_t^2 - 2\Sigma e_t e_{t-1}}{\Sigma e_t^2}$$

$$d \approx 2(1 - \frac{\Sigma e_t e_{t-1}}{\Sigma e_t^2})$$

$$d \approx 2(1 - \hat{\rho})$$

because

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

So: as rho →1, *d* →0
as rho→0, *d* → 2
as rho→-1 *d* → 4

- Testing the statistical significance of d

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

- Very rarely will see *negative* autocorrelation; d can handle this but we will ignore it
- It turns out that the exact sampling distribution of d is not known for a given number of cases. But it is known that d must lie between 0 and 2 (for positive autocorrelation), and between 2 and 4 (for negative autocorrelation)
- Durbin and Watson worked out the *lower* and *upper* bounds of where d must be in order to reject or fail to reject the null hypothesis, given the number of cases in the analysis and the number of independent variables in the model
- The so-called d_L and d_U for a given N and k (including the intercept in the total of k) are given in any statistics text

- So: compare obtained *d* to the upper and lower bounds, given the df and number of independent variables
- Decision rules:
  - If *d* is greater than d_U, **do not reject H_0**.
  - If *d* is less than d_L, **reject H_0**.
  - If *d* is in between d_L and d_U, the test is **inconclusive**
- Our example:
  - N=29
  - k=2 (including the intercept)
  - d_L =1.054, d_u =1.332

```
. dwstat //THE DURBIN-WATSON TEST :  RESULT IS LOWER THAN DL CRITICAL VALUE, THEREFORE AUTOCORRELATION

Durbin-Watson d-statistic(  2,    29) =  .9568562
```

d is lower than d_L, so we reject H_0. We reject the hypothesis that there is no positive autocorrelation.

Breusch-Godfrey (BG) Test

- Limitations of D-W *d*
  - Only handle first-order autocorrelation, not higher orders nor MA processes
  - No lagged dependent variables allowed in the model
- More general test provided by Breusch-Godfrey
- Idea: Obtain OLS residuals, and regress them against the Xs, lagged Y (if you like) and as many lags (*p*) of the residuals as you think necessary (constrained by N of cases and common sense)

$$e_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 e_{t-1} + \hat{\rho}_2 e_{t-2} + \hat{\rho}_3 e_{t-3} + \ldots v_t$$

- In large samples, $(N-p)*R^2 \sim x^2$ with *p* degrees of freedom

```
. estat bgodfrey, lags(2)

Breusch-Godfrey LM test for autocorrelation
```

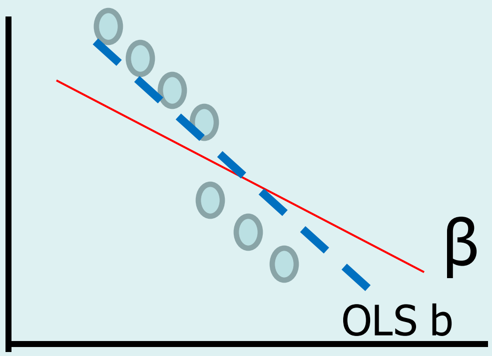
lags(p)	chi2	df	Prob > chi2
2	7.218	2	0.0271

H0: no serial correlation

Consequences of Autocorrelation

- Same consequences as with heteroskedasticity
- OLS is "unbiased"; it neither overestimates nor underestimates the population slope on average
- But OLS yields incorrect standard errors and thus incorrect hypothesis tests
- Typically, OLS *underestimates* the true standard error, and thus leads to rejecting null hypotheses that should not be rejected
- Can correct OLS standard errors, but result is *inefficient* estimates (i.e., larger than they "need" to be)
- BLUE estimator with autocorrelation is provided by Generalized Least Squares (GLS) if rho (ρ) is known; Feasible Generalized Least Squares (FGLS) if rho (ρ) is estimated from the data
- Robust standard errors with autocorrelation: "HAC" (Heteroskedastic and Autocorrelation Consistent)

OLS Disregarding Autocorrelation



Can see that autocorrelated errors lead OLS to estimate a line that is "too close" to the data points. The reason those observations are where they are is not because of X, but because of the autocorrelation in the residuals that OLS assumes don't exist

So the line in this case is biased downward; but if our first couple of observations had been ones with negative residuals, we would then estimate a line that was also "too close" to the points, but was biased upward.

So OLS, on average, gets it right, it is *unbiased*. But in a given sample it will be way off.

Correcting for Autocorrelated Disturbances: Generalized Least Squares (GLS)

- AR(1) Model: $y_t = \beta_0 + \beta x_t + \varepsilon_t$

  where

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

with v_t as pure "white noise" disturbance

- If we knew the value of rho (ρ), we could transform the model in such a way as to satisfy the OLS assumptions
- Different ways to estimate rho:
  - Estimate rho from a regression of the OLS residuals at time *t* against time *t-1*
  - Use the D-W *d* test to approximate rho
  - Iterative procedures such as "Prais-Winsten" or "Cochrane-Orcutt" methods
  - Assume it is 1, the maximum value, and use that as an approximation of "very high" autocorrelation
- Plug estimates into the AR(1) model and arrive at "Feasible GLS", usually estimated via "Prais-Winsten" regression ("prais" in STATA)

```
. regress budget crime

      Source |       SS       df       MS              Number of obs =      29
-------------+------------------------------           F(  1,    27) =  339.33
       Model |  23783.6827     1  23783.6827           Prob > F      =  0.0000
    Residual |  1892.45521    27  70.0909338           R-squared     =  0.9263
-------------+------------------------------           Adj R-squared =  0.9236
       Total |  25676.1379    28  917.004926           Root MSE      =   8.372

------------------------------------------------------------------------------
      budget |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       crime |   5.373169   .2916902    18.42   0.000      4.77467    5.971668
       _cons |  -4.759523   3.637971    -1.31   0.202    -12.22402    2.704978
------------------------------------------------------------------------------
```

OLS

```
. prais budget crime, twostep robust // THIS STOPS THE ITERATIONS AFTER TWO ROUNDS

Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.5208

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

Linear regression                                      Number of obs =      29
                                                       F(  1,    27) =  199.71
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7904
                                                       Root MSE      =  6.8958

------------------------------------------------------------------------------
             |             Semirobust
      budget |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       crime |    4.77184   .3376612    14.13   0.000     4.079016    5.464663
       _cons |   2.044059   4.770144     0.43   0.672    -7.743468    11.83159
-------------+----------------------------------------------------------------
         rho |   .5207877
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.956856
Durbin-Watson statistic (transformed) 1.615307
```

FGLS
“Prais-Winsten”

“Heteroskedasticity and Autocorrelation Consistent Standard Errors” (HAC)

- If agnostic about the form of autocorrelation, can estimate via OLS (since it is “consistent”) and correct standard errors for possible autocorrelation via “**Newey-West**”, or HAC standard errors, which also provides correction for possible heteroskedasticity as well
- Takes White robust standard error and weights it by a factor related to the cross-products of the OLS error terms to take autocorrelation into account. Which cross-products are used depends on researcher specifying how many lags of autocorrelation are likely to be relevant
- These will not be BLUE if you have AR(1) autocorrelation, though !!!

```
. newey budget crime, lag(5)

Regression with Newey-West standard errors          Number of obs  =        29
maximum lag: 5                                      F(  1,     27) =    183.91
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
      budget |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       crime |   5.373169   .3962155    13.56   0.000     4.560202    6.186136
       _cons |  -4.759523   5.108712    -0.93   0.360    -15.24173    5.722689
------------------------------------------------------------------------------
```