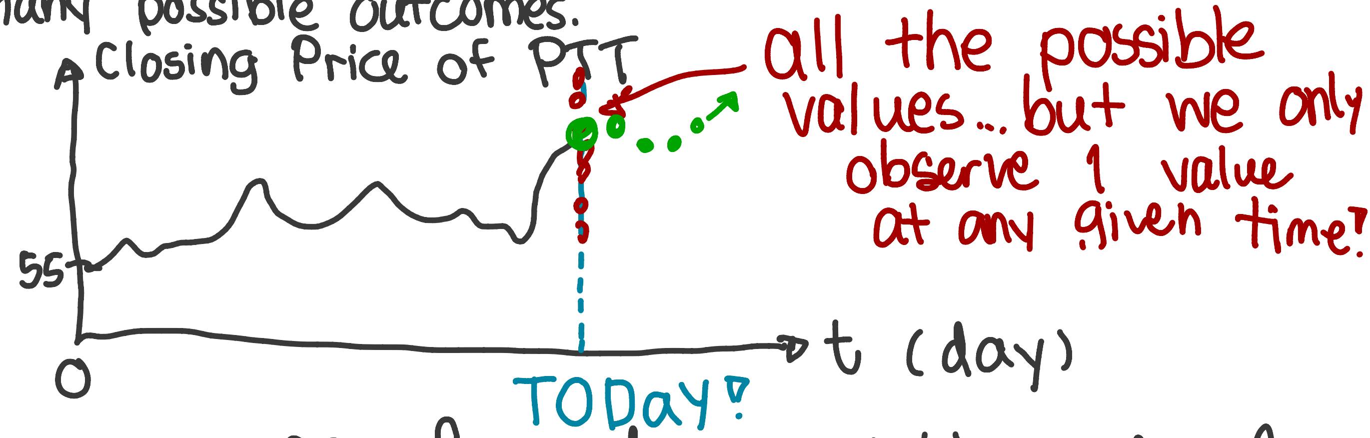
Serial Correlation and Heteroskedasticity in Time Series Regressions

1 The Nature of Time Series Data

Observations from the same object of interest (individual, country, province, stock, etc.) for MANY periods of time.

- For each time period, we only observe 1 realization of many possible outcomes.



- A sequence of random variables indexed by time is called a "stochastic process" or a "time-series" process.

TABLE 10.1

Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–2003

← object of interest

Year	Inflation	Unemployment
1948	8.1	3.8
1949	−1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
⋮	⋮	⋮
1998	1.6	4.5
1999	2.2	4.2
2000	3.4	4.0
2001	2.8	4.7
2002	1.6	5.8
2003	2.3	6.0

← Before the end of 1950, unemployment in the US is a random process.

2 Examples of Time Series Regression Models EE435

There are many time series regression models. Different models would be suitable for different types of relationship we want to estimate. Some examples of time series models are Static Model, AR (Autoregressive), ADL (Autoregressive Distributed Lag), FDL (Finite Distributed Lag), ARMA (Autoregressive Moving Average), ARCH (Autoregressive Conditional Heteroskedasticity) etc.

In this class we will talk about 2 examples 1) Static Models and 2) FDL.

↑ similar to cross-sectional

2.1 *Static Models*

Studies a contemperaneous (occurring in the same period of time) relationship of variables.

For example: How unemployment affect inflation → Phillips Curve.

e.g. inflation of Thailand in year t. → $inflation_t = \beta_0 + \beta_1 unemp_t + error$

How the price of gasohol affect demand for benzene.

$q\_benzene_t = \beta_0 + \beta_1 p\_benzene_t + \beta_2 p\_gashol_t$ (own-price; price of substitutes.)

$+$ price of other subsitutes & other factors $+ u_t$

These static regressions are no different to the cross-sectional regression. But, we have each observation being each "time(t)" rather than each "object (i)" → get long-run equilibrium relationship → not forecast of future values!

2.2 *Finite Distributed Lag Models*

- X and/or the lagged variables of X_t can affect Y_t.
- lagged variables of X_t are $X_{t-1}, X_{t-2}, X_{t-3}$, etc.

For example,

- The price of PTT stock in the past (yesterday, last week, last month) can affect the price of PTT today?
- Investment of education in the past can affect GDP (productivity) now.

$GDP\_percap_t = \alpha_0 + \delta_0 edu\_inv_t + \delta_1 edu\_inv_{t-1} + \ldots + \delta_{10} educ\_inv_{t-10} + u_t$

The longest lag is "10". This is called FDL of order 10

In general, FDL of order 2. How many lags would be correct?

X variables and their Lags.

$$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + u_t$$

It depends on many factors + your intuition. So, case-by-case.

$\delta_0 = \frac{dy_t}{dx_t}$ is the change in Y due to 1 unit increase in X during the same period. Called the "impact propensity" or "impact multiplier".

$\delta_1 = \frac{dy_t}{dx_{t-1}}$ is the change in Y due to a temporary change in X 1 period prior.

No need to use MLR3 - Random sample
already a ramdom process

Time Series.

3 Properties of OLS under classical assumptions

parameter can't be non-linear.

Assumption TS1. Linear in Parameter – Y is linear in X. → $Y_t = \beta_0 + \beta_1 X_{1t} + \ldots + U_t$

Assumption TS2. No Perfect Collinearity

Assumption TS3. Zero Conditional Mean

↳ $E(U_t | X_t) = 0$ → no omitted variable bias.

***** Under Assumptions TS1 to TS3, $\hat{\beta}_{OLS}$ would be unbiased *****

Assumption TS4. Homoskedasticity
Assumption TS5. No Serial Correlation

} for OLS estimators to be efficient.

$Corr(U_t, U_{s \neq t} | X) = 0 \; ; \; \forall \, t \neq s$

* Conditional on X, the error at time t should not be correlated with the error at time s (at other periods.)

***** Under Assumptions TS1 to TS5, $\hat{\beta}_{OLS}$ would be BLUE (best linear unbiased estimators)*****

The variance of $\hat{\beta}_{OLS}$ (under ass.TS1-TS5)

(Serial correlation doesn't affect the biasness.)

$$Var(\hat{\beta}_j | X) = \frac{\sigma^2}{TSS_j(1-R_j^2)}, \quad j=1,\dots,k$$

↑

if homoskedasticity + no serial correlation, we can use this calculation.

TSS_j = total sum of squares of X_{tj} = $\sum_{t=1}^{n}(X_{tj} - \bar{X}_j)^2$

R_j^2 = the R^2 from the regression of X_j on all other X (all other explanatory variables)

4 Properties of OLS with Serially Correlated Errors

- $Corr[u_t, u_s | X] \neq 0 \quad \forall t \neq s$ → serial correlation

  Why? Inertia, momentum (over valuation of stock price at time t may lead to over-valuation of stock price at time t+1, etc.)

- $Corr[u_i, u_h | X] \neq 0 \quad \forall i \neq h$ ~> auto-correlation

  cross-sectional data. Wrong functional form, non-random sampling, systematic measurement errors.

5 Unbiasedness and Consistency

- As long as TS.1 – TS.3 are satisfied, then $\hat{\beta}_{OLS}$ will be unbiased.

Strictly exogenous → requires u_t to be uncorrelated with X_t and X_s ($\forall t \neq s$) in ALL Periods!

V. difficult to achieve.

→ If strict exogeneity can't be achieved, then we compromise with weak exogeneity

$E(u_t | X_t) = 0$ ~> u_t uncorrelated with X_t (only X in period t.)

☆ If we have only weak exogeneity, our $\hat{\beta}_{OLS}$ will be consistent (but not unbiased)

^ need large observations

6 Efficiency and Inference

$y_{t-1} = \beta_0 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + u_{t-1}$

With serial correlation, $\hat{\beta}_{OLS}$ would not be BLUE ($var(\hat{\beta}_{OLS})$ would not be minimized).

Consider main model: $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$

regression of errors in period t & t-1

$$u_t = \rho u_{t!1} + e_t \; ; \; t = 1, 2, ..., n \quad \text{and} \quad !\rho! < 1$$

where u_t is from a regression model

another possible main regression

$$y_t = \beta_0 + \beta_1 x_t + u_t.$$

- now assume
  1) $E(e_t) = 0$
  2) $Var(e_t) = \sigma_e^2$
  3) $Cov(e_t, e_{t-1}) = 0$
  4) $Cov(u_{t-1}, e_t) = 0$ or $E(e_t \mid u_{t-1}) = 0$
  5) $E(u_t) = E(u_{t-1})$
  6) $E(u_t^2) = E(u_{t-1}^2)$

After intensive calculation

$Var(u_t \mid X)$

$$Var(\hat{\beta}_1) = \frac{1}{SST_x^2}\left[\sum_{t=1}^{n} x_t^2 \sigma^2 + 2\sum_{t=1}^{n-1}\sum_{j=1}^{n-t} x_t x_{t+j} \rho \sigma^2\right]$$

(with serial corr.)

$$= \frac{\sigma^2}{SST_x^2} + \frac{2\rho\sigma^2}{SST_x} \sum_{t=1}^{n-1}\sum_{j=1}^{n-t} x_t x_{t+j}$$

This extra term is added when we have Serially correlation.

Similar to heteroske, we need to use the correct calculation for $\widehat{Var\, \hat{\beta}_{OLS}}$ or Std. Err.

7 Testing for Serial Correlation

(*) note that there can be many serial corr. patterns, we need to specify the pattern before doing the testing.

Given the model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$

7.1 A "t-test" for AR(1) serial correlation with strictly exogeneous regressors

The most common type of serial correlation or autocorrelation is the AR(1) type:

$$u_t = \rho u_{t-1} + e_t \; ; \; t = 1, 2, 3, \ldots n$$

- The shock carries forward 1 period with the coefficient ρ ; $|\rho| < 1$ otherwise, we will have an explosive process.
- To test for AR(1) auto/serial correlation:
  - $H_0: \rho = 0$ → no AR(1) serial correlation.
  - $H_a: \rho \neq 0$ → serial correlation.

To perform the test:

1. Estimate $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$
2. Obtain $\hat{u}_t$,$\hat{u}_{t\,!\,1}$; " $t = 1, 2, ..., n$ ← In practice, we use $\hat{u}_t$ to approximate u_t and $\hat{u}_{t-1}$ to approximate u_{t-1}.
3. Estimate $\hat{u}_t = \rho\, \hat{u}_{t\,!\,1} + error$
4. Perform the t # *test* for

$H_0: \rho = 0$ → no AR(1) serial correlation.

$H_a: \rho \neq 0$

7.2 The Durbin-Watson Test (DW test)

- Allows us to test for positive and negative serial correlation.

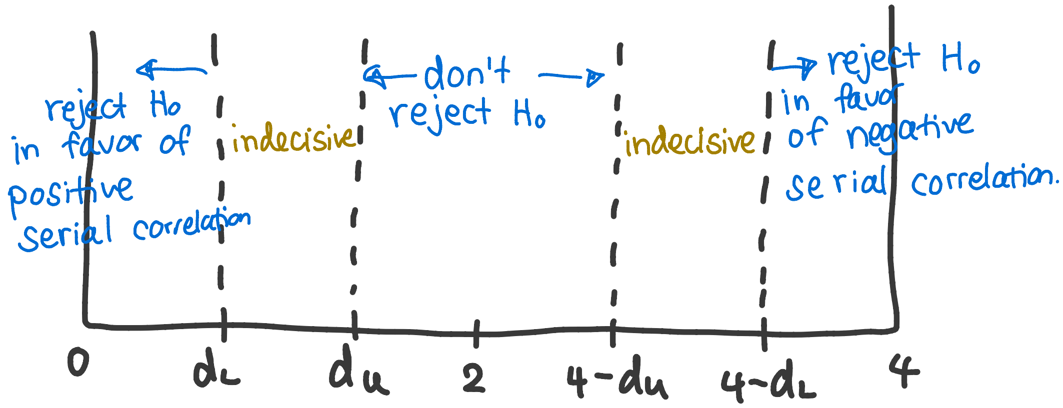
DW-statistic (DW) = $\dfrac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$

This implies

$DW \approx 2(1-\hat{\rho})$ from $u_t = \rho u_{t-1} + e_t$

no serial correlation	$\hat{\rho} = 0$	$	$DW = 2$
positive serial corr.	$\hat{\rho} > 0$	$	$DW < 2$
negative serial corr.	$\hat{\rho} < 0$	$	$DW > 2$

Since $-1 \leq \rho \leq 1$, $0 \leq DW \leq 4$!



no serial correlation

H_o : [redacted]

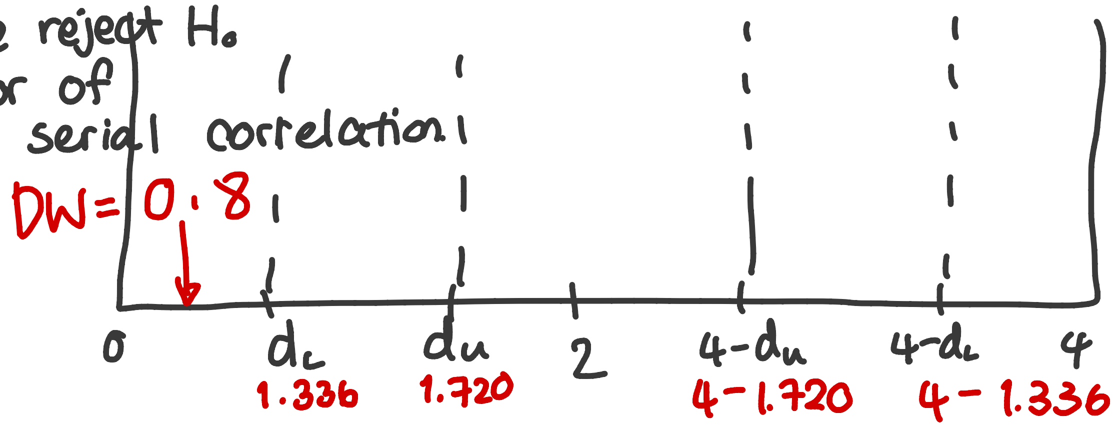
H_a : [redacted] serial correlation

To perform the test:

1. Estimate $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$
2. Obtain $\hat{u}_t$,$\hat{u}_{t!1}$; " $t = 1, 2, ..., n$
3. Calculate DW from eq.(2)
4. Find the critical d_L and d_u values (say, at the 5% level of significance) for the given sample size and # of regressors. (calculated case-by-case)
5. Follow the decision rule in the picture.

Example:
Suppose the calculated value of $DW = 0.80$, $n = 45$, $k = 4$.
From this, we get $d_L =$ ___1.336___ and $d_u =$ ___1.720___

★ So, we reject H_0 in favor of positive serial correlation



8 Correcting for serial correlation

8.1 Passive way

Use the type of standard error that is robust to the serial correlation, autocorrelation problem

- Since TS 1 – TS 3 are not violated, $\hat{\beta}_{OLS}$ unbiased.
  ↳ We just have to use the correct calculation of $\widehat{Var}(\hat{\beta}_j)$ (std. err.)

STATA Command: newey y x_1 x_2 x_3, lag(q)

q: # of lags in the AR(#) process.

8.2 Active way – use weighted-least squares or feasible weighted least squares (FGLS)

↳ the process is similar to the active remedy for heteroskedasticity.

Multicolinearity

1 The Nature of Multicolinearity

- X variables in the model are highly correlated.

⇒ though not perfectly correlated

Perfectly correlated ($|\rho|=1$)

observation	x_1	x_2	$3x_1 \# x_2$
1	6	18	0
2	12	36	0
3	7	21	0
4	-5	-15	0

Highly correlated ($|\rho| > 0.8$)

observation	x_1	x_2	$3x_1 \# x_2$
1	6	16	-2
2	12	45	9
3	7	18	-3
4	-5	-12	3

Causes

1) Data collection method, not enough sample variation.
2) Constraint on the model or on the sampled population.
3) Model specification (usually happens when there are too many polynomial terms, e.g. X^2, X^3, X^4).
4) put too many X variables.

2 Consequences of Multicolinearity

2.1 *The OLS estimator will still be BLUE.*

If MLR 1-5 are satisfied.

2.2 BUT! *The variances and covariances* (of $\hat{\beta}_{OLS}$) *will be very large. This makes precise estimation difficult.*

Recall $Var(\hat{\beta}_j) = \dfrac{\sigma^2}{SST_j(1-R_j^2)}$

If X variables are highly correlated, R_j^2 will be very large! This makes $Var(\hat{\beta}_j)$ becomes large, too.

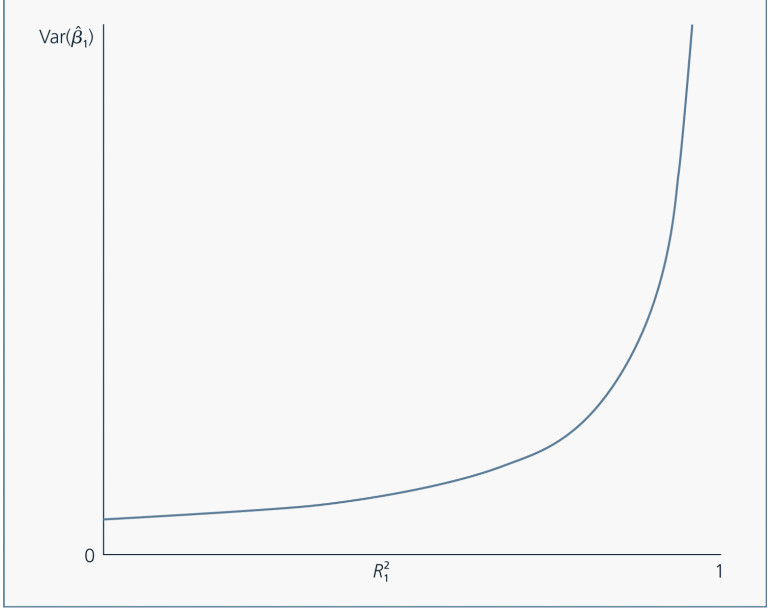
- SST_j is the SST from regressing X_j on all other X_s. For example, SST_2 is SST from the following regression:

$$X_2 = \theta_0 + \theta_1 X_1 + \theta_2 X_3 + \ldots + \theta_K X_K + \text{error.}$$

- R_j^2 is R^2 from regressing X_j on all other X_s.

FIGURE 3.1

$\text{Var}(\hat{\beta}_1)$ as a function of R_1^2.



3 Detection of multicolinearity

1. There is conflicting test between t- and F-test: if we find that the conclusion derived from the two tests are inconsistent, specifically R2 is high and F-test results in statistical overall significance; whereas, at least, one null hypothesis of some t-tests cannot be rejected, it is reasonable to suspect the multicollinearity problem.

2. Correlation of regressors is greater than 0.8: the higher the correlation, the higher the variance of estimators.

3. Variance inflation factor (VIF) is greater than 10: when the regressors face the multicollinearity problem, the value of VIF might be so high that the resulting high variance of estimators adversely affects the regression analysis.

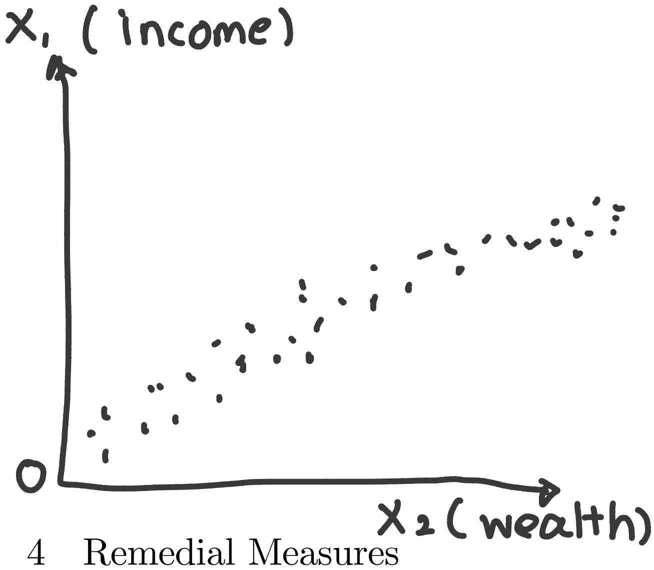
   % The VIF (variance inflation factor) to detect high multicolinearity:

$$VIF = \frac{1}{1 - R_j^2}$$

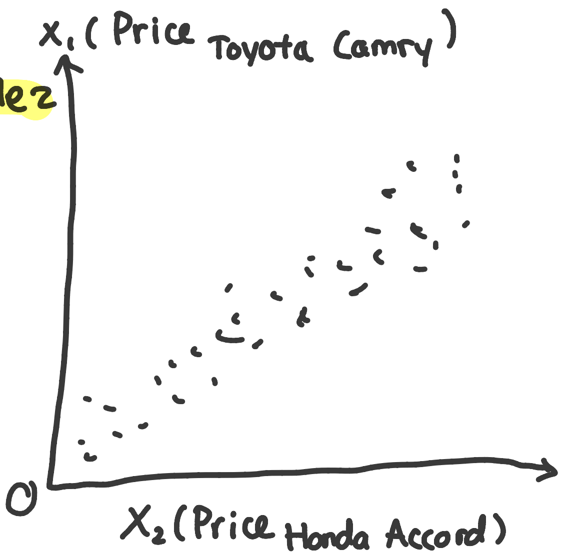
$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} \cdot VIF$$

4. Scatter plot of two regressors is relatively linear: when we plot the value of on regressor against another and we find that both of them tend to change in the same way, this fact might suggest the existence of multicollinearity.

Example 1



Example 2



4 Remedial Measures

1. Do nothing → Because we still have BLUE
   - If the multicollinearity is not too severe.
   - among the control variables, not variables of interest.

2. Apply prior relationship among explanatory variables - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. (1)

   Suppose " $\beta_2 = 0.7\beta_1$ " → β_2 = total score of a university course, β_1 = cummulative score before final exam of a university course.

   Then, (1) becomes $Y = \beta_0 + \beta_1 X_1 + 0.7\beta_1 X_2 + u = \beta_0 + \beta_1 (X_1 + 0.7X_2) + u$ → can only determine the value of $\hat{\beta}_1$, not both $\hat{\beta}_1$ & $\hat{\beta}_2$.

3. Discard some explanatory variables - the removal of the variables could mitigate the problem; but, another problem, namely specification bias problem, might occur instead. For example, suppose we want to construct the model where the production is the explained variables; and labor and capital are the explanatory ones. If there is linear relationship between labor and capital, the elimination of one variable might assuage the multicolinearity problem, but might be contrary to economic reasoning. Hence, the decision of which variables will be disposed of should be based on economic theory.

4. Collect more observations - this practice will increase _ _ SST_j _ _ _ _ _ _ _, (← Because more sample variation.) which is the component of the variances. As a result, the variances will be lower despite high correlation among explanatory variables.

   $$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} \cdot \frac{1}{(1 - R_j^2)}$$

   If this gets larger, $Var(\hat{\beta}_j)$ will decrease.

5. Transform the variables - although there is linear relationship among explanatory variables, it is not necessary that the first difference or ratio transformation of the variables will have that relationship

   For example, first difference

   $$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t \quad (1)$$
   $$Y_{t-1} = \beta_0 + \beta_1 X_{1,t-1} + \beta_2 X_{2,t-1} + u_{t-1} \quad (2)$$

   (1) − (2) gives

   $$Y_t - Y_{t-1} = \beta_0 + \beta_1 (X_{1t} - X_{1,t-1}) + \beta_2 (X_{2t} - X_{2,t-1}) + (u_t - u_{t-1})$$

Some useful points for the exam

1. Do you know that $\widehat{Var(\hat{\beta})}$ = Std. Err. !

	termgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
X_1	attend	.046594	.0036082	12.91	0.000	.0395093	.0536787
X_2	priGPA	.5329307	.0403281	13.21	0.000	.4537468	.6121146
X_3	final	.0503197	.0040339	12.47	0.000	.0423992	.0582403
X_4	frosh	.0974307	.0560211	1.74	0.082	-.0125662	.2074276
X_5	soph	.0689273	.0467006	1.48	0.140	-.0227689	.1606236
	_cons	-1.361077	.1316861	-10.34	0.000	-1.619642	-1.102513

2. Heteroskedascity ⇒ variance of the error (Var(u)) correlates with explanatory variables (x)

- To test, we can use either the BP-test or White test
- BP-test $\hat{u}^2 = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \text{error}$

  approximates Var(u) ↗

  $H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$ ⇝ Homoske.

  H_a: H_0 not true ⇝ Heteroske.

3. Serial / Auto correlation ⇒ "value" of the error terms (u_t) correlates among each other across times (across samples).

- Test for AR(1) in the error : $u_t = \rho u_{t-1} + e_t$

  $H_0: \rho = 0$ ⇝ no serial correlation

  $H_a: \rho \neq 1$ ⇝ AR(1) serial correlation