Introductory Econometrics

Serial Correlation

Monash Econometrics and Business Statistics

Semester 2, 2018

Recap

- We have studied the multiple regression model and learnt that when:

1. model is linear in parameters: $y = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
2. conditional mean of errors is zero: $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$
3. columns of $\mathbf{X}$ are linearly independent

$\Rightarrow$ then the OLS estimator $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$

Recap

- We have studied the multiple regression model and learnt that when:

1. model is linear in parameters: $y = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
2. conditional mean of errors is zero: $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$
3. columns of $\mathbf{X}$ are linearly independent

$\Rightarrow$ then the OLS estimator $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$

- if in addition,

4. sample is random and errors are homoskedastic: $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n$,

$\Rightarrow$ then $\widehat{\boldsymbol{\beta}}$ is the BLUE and $Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Recap

- We have studied the multiple regression model and learnt that when:

1. model is linear in parameters: $y = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
2. conditional mean of errors is zero: $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$
3. columns of $\mathbf{X}$ are linearly independent

$\Rightarrow$ then the OLS estimator $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$

- if in addition,

4. sample is random and errors are homoskedastic: $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$,

$\Rightarrow$ then $\widehat{\boldsymbol{\beta}}$ is the BLUE and $Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

- If, in addition to the above,

5. errors are normally distributed,

$\Rightarrow$ then conditional on $\mathbf{X}$, $\widehat{\boldsymbol{\beta}}$ is normally distributed, and we can use the usual t and F tests to make inferences based on the OLS estimator

Lecture Outline

- We define the concept on serial correlation [textbook 10-3; p.353-354]
- We discuss how the properties of the OLS estimator alter in the presence of serial correlation [textbook 12-1; p.412-414]
- We test for serial correlation in the error term: Breusch-Godfrey test [textbook 12-2; p.416-418, 421-423]
- We provide two solutions in this lecture:
  1. using the OLS estimator but correcting its standard errors: HAC standard errors [textbook 12-5; p.431-434]
  2. using a generalised least squares (GLS) estimator instead of OLS [textbook 12.3; p.423-429]
- And we study more general dynamic specifications in the next lecture.

Serial Correlation

- In the list of assumptions of the classical linear model in non-matrix form, we had "sample is randomly selected"
- This assumption is very unlikely to hold for a time series sample
- However, the list of assumptions in matrix form did not have the explicit assumption of random sampling. We only needed $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ and $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$.
- We can have $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ and $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$ with time-series (non-random) samples.
- The second one, in particular, says that in addition to homodkedasticity, we should also have that conditional on $\mathbf{X}$, the errors in two different time periods are uncorrelated, or

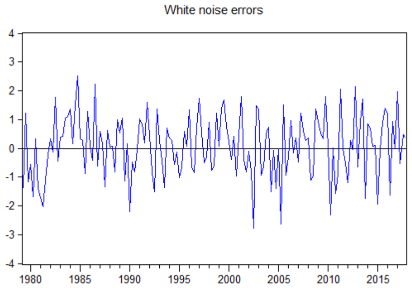
$$Corr(u_t, u_s \mid \mathbf{X}) = 0, \text{ for all } t \neq s \tag{1}$$

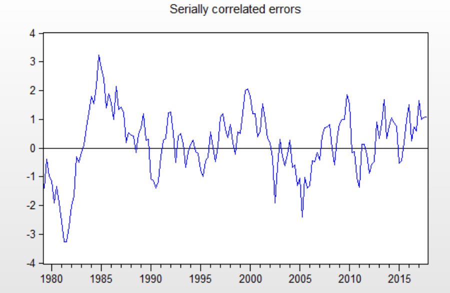
- When (1) does not hold we say that the errors of regression model, $y = \mathbf{X}\beta + \mathbf{u}$, suffer from *serial correlation*, or *autocorrelation*

Consequences of serial correlation in errors for OLS

- Serial correlation in errors does not affect the assumptions (i) $y = \mathbf{X}\beta + \mathbf{u}$, (ii) $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$, and (iii) columns of $\mathbf{X}$ are linearly independent, therefore **the OLS estimator will remain unbiased**
- However, since it violates $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$, **the OLS estimator will no longer be BLUE and** $Var(\widehat{\beta}) \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- This means that, as in the case of heteroskedasticity, the default standard errors reported by the statistical package for the OLS estimator will be incorrect
- As a result, the usual t and F statistics based on OLS standard errors no longer have t and F distributions

Diagnosing serial correlation in errors

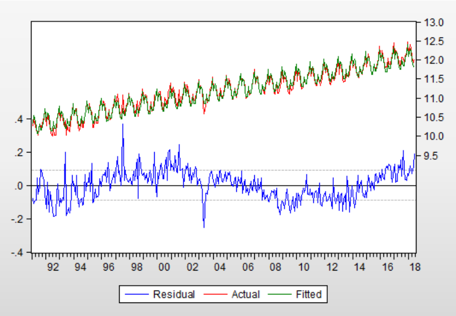




Diagnosing serial correlation in errors

Visual tools: 1. Plotting the residuals

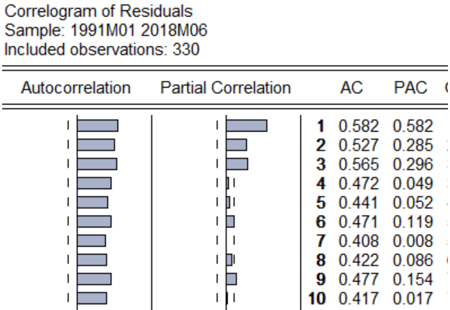
- Strong positive serial correlation can be seen from the time series plot of OLS residuals. In this case, positive residuals are likely to be followed by positive residuals and negative residuals are more likely to be followed by negative residuals
- Example: Monthly international visitor arrivals in Victoria: Plots of actual values, fitted values and residuals of the regression of $\log(VIC)$ on a constant, time trend and 11 monthly dummies:



Diagnosing serial correlation in errors

Visual tools: 2. The correlogram of residuals

- The correlogram shows the estimated autocorrelations of a time series (correlations with its own lags).
- This correlations are often shown as bar charts with some band around zero. If the bar crosses the band, then it is likely to be significant.
- Example: Monthly international visitor arrivals in Victoria: The correlogram of the residuals of the regression of log(VIC) on a constant, time trend and 11 monthly dummies:


Correlogram of Residuals
Sample: 1991M01 2018M06
Included observations: 330

Autocorrelation	Partial Correlation		AC	PAC
		1	0.582	0.582
		2	0.527	0.285
		3	0.565	0.296
		4	0.472	0.049
		5	0.441	0.052
		6	0.471	0.119
		7	0.408	0.008
		8	0.422	0.086
		9	0.477	0.154
		10	0.417	0.017

Diagnosing serial correlation in errors

Testing for serial correlation: The null and the alternative

- In order to formally test the null hypothesis that errors are not serially correlated, we have to be specific on how the errors are correlated if the null in not true.
- The most useful model of dependence is to assume that errors are *autoregressive of order q*, denoted by $AR(q)$
- This means that the model is:

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \tag{2}$$
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_q u_{t-q} + e_t \tag{3}$$

where $\{e_t, t = 1, \ldots, n\}$ are i.i.d. with mean 0 and variance σ^2

- The null and the alternative can be written as:

$H_0 : \rho_1 = \rho_2 = \cdots = \rho_q = 0$

H_1 : at least one of the q autoregressive parameters is not zero

- It is important that q is determined before looking at the data (think about why). We determine q with reference to the frequency of the data (annual 1 or 2, quarterly 4, ...).

Diagnosing serial correlation in errors

Testing for serial correlation: Breusch-Godfrey test

- If we could observe the errors, u_t, then we could estimate (3) by OLS and test this null.
- However, u_t are not observable so, instead, we ...

S1 Estimate

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

by OLS and obtain the OLS residuals, $\widehat{u}_t,\ t = 1, 2, \ldots, n$.

S2 Estimate

$$\widehat{u}_t = \alpha_0 + \alpha_1 x_{t1} + \ldots + \alpha_k x_{tk} + \rho_1 \widehat{u}_{t-1} + \ldots + \rho_q \widehat{u}_{t-q} + e_t \quad (4)$$

by OLS. Let $R^2_{\widehat{u}}$ denote the R^2 obtained from (4).

$$BG = (n-q) R^2_{\widehat{u}} \overset{asy}{\sim} \chi^2(q) \text{ under } H_0.$$

S3 We reject H_0 at the 5% significance level if

$$BG_{calc} > BG_{crit},$$

where BG_{crit} denotes the 95th percentile of a χ^2 distribution with q degrees of freedom

Diagnosing serial correlation in errors

Testing for serial correlation: Example

- The Victorian international tourist arrivals. Data is monthly, so we test the null that errors are not serially correlated against the alternative that they are $AR(12)$.

S1 Regress $\log(VIC)$ on a constant, time trend and 11 dummies for 11 months, save residuals $\widehat{u}_t$

S2 Regress $\widehat{u}_t$ on a constant, time trend, 11 dummies for 11 months, and $\widehat{u}_{t-1}$ to $\widehat{u}_{t-12}$. The R^2 of this regression is 0.504. Note also that the number of observations in this regression is $330 - 12 = 318$

S3 We know that $(n-q)R^2_{\hat{u}} \sim \chi^2_{12}$ under H_0.

$$BG_{calc} = 318 \times 0.504 = 160.27$$

$$BG_{crit} = 21.03 \text{ from the } \chi^2_{12} \text{ table}$$

$$BG_{calc} > BG_{crit} \Rightarrow \text{ we reject } H_0$$

- As the visual tools showed there is significant evidence of serial correlation in the errors of the $\log(VIC)$ model.

Diagnosing serial correlation in errors

Breusch-Godfrey test: notes

- Note that the reason we use $n - q$ to compute the BG test statistic is that we lose q observations when we form q lags. Stop to make sure you understand why.
- Different software packages may compute the BG test statistic slightly differently.
- For example, EViews replaces all missing values in lags with zero. So, both the number of observations and the R^2 of the auxiliary regression would be different from the way we have computed it.
- However, since the BG test should only be used when n is large, these variations do not normally change the conclusion of the test.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	24.98731	Prob. F(12,305)	0.0000
Obs*R-squared	163.5945	Prob. Chi-Square(12)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Sample: 1991M01 2018M06
Included observations: 330
Presample missing value lagged residuals set to zero.

Implications of serial correlation

- Similar to heteroskedastic errors, serially correlated errors imply that the OLS estimator is no longer BLUE, and OLS standard errors are incorrect.
- However, serially correlated errors have an important implication: they imply that we can improve the prediction of a model by learning from past errors.
- Consider a two variable model with AR(1) errors:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + u_t \\ u_t &= \rho u_{t-1} + e_t \\ \Rightarrow y_t &= \beta_0 + \beta_1 x_t + \rho u_{t-1} + e_t \end{aligned}$$

which implies that we can get a better prediction for y_t by adding a fraction ρ of the previous error u_{t-1} to $\beta_0 + \beta_1 x_t$. More on this later.

Remedies for serial correlation in the errors of a linear regression model

If we find evidence of autocorrelation in the error term in our linear regression model, we have three options:

1. Use the OLS estimator, even though it is not BLUE, but use standard errors that are asymptotically valid even if the errors are autocorrelated. Use these standard errors for inference.
2. Use a feasible generalised least squares (GLS) estimator. This estimator is not BLUE either, but it is more efficient than the OLS estimator in large samples.
3. Use a more general dynamic model specification.

Solution 1: HAC standard errors

- First we focus on option 1 for dealing with autocorrelated errors.
- One can opt to continue using the OLS estimator, but to adjust the usual OLS standard errors such that these and the test statistics based upon them are valid in large samples (a.k.a. *asymptotically*) even when the errors are autocorrelated.
- This follows a similar logic to the adjustment of OLS standard errors of the coefficients of the main regression, when heteroskedasticity is detected (see Week 8 lecture notes).

Solution 1: HAC standard errors

- Recall that in the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

  the OLS estimator is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and we have

$$Var(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \left[\mathbf{X}' Var(\mathbf{u} \mid \mathbf{X})\mathbf{X}\right] (\mathbf{X}'\mathbf{X})^{-1}.$$

- When errors are homoskedastic and not serially correlated, we have $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$, which implies $Var(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- In lecture 8 we presented the solution White provided on how to estimate $Var(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})$ when errors are heteroskedastic.
- Newey and West provided a solution for this problem in the presence of serially correlated errors. Their estimator is called the Newey-West HAC (heteroskedasticity and autocorrelation consistent) estimator, and is available in all good software packages.

The logic behind robust standard errors

Only for those who are curious. Not examinable.

- Consider

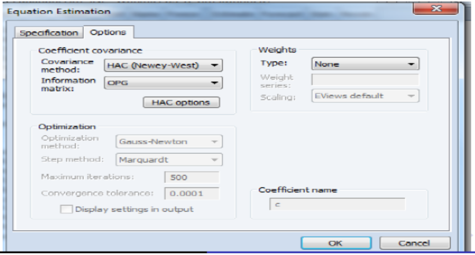
$$Var(\widehat{\beta} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\left[\mathbf{X}'Var(\mathbf{u} \mid \mathbf{X})\mathbf{X}\right](\mathbf{X}'\mathbf{X})^{-1}.$$

- With HTSK and autocorrelated errors $Var(\mathbf{u} \mid \mathbf{X})$ is an n by n matrix with lots of unknown elements that get larger with sample size.
- White's genius was to notice that to estimate $Var(\widehat{\beta} \mid \mathbf{X})$, we only need to know $[\mathbf{X}'Var(\mathbf{u} \mid \mathbf{X})\mathbf{X}]$, which is always a $k+1$ by $k+1$ matrix regardless of the sample size. In a large sample, there is enough information to estimate this matrix.
- All robust estimators of $Var(\widehat{\beta} \mid \mathbf{X})$ use this logic.

HAC standard errors

Example

- Consider Victoria's international tourist arrivals again.
- Newey-West HAC estimate of variance can be chosen in EViews estimation window under 'Options'



- The OLS results with usual standard errors and the OLS results with HAC standard errors for the first few parameters are given in the next slide.

Empirical example continued

Dependent Variable: LOG(VIC)
Method: Least Squares
Sample: 1991M01 2018M06
Included observations: 330

Variable	Coefficient	Std. Error	t-Statistic
C	10.66171	0.019773	539.2010
T	0.005401	5.30E-05	101.8685
@MONTH=1	-0.317922	0.024743	-12.84874
@MONTH=2	-0.135544	0.024743	-5.478088

Dependent Variable: LOG(VIC)
Method: Least Squares
Sample: 1991M01 2018M06
Included observations: 330
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic
C	10.66171	0.019836	537.5007
T	0.005401	0.000104	51.79185
@MONTH=1	-0.317922	0.013028	-24.40314
@MONTH=2	-0.135544	0.015520	-8.733471

Solution 2: Estimate by FGLS

- Can we transform the model to a model that does not suffer from serial correlation?
- Let's consider the simplest model with AR(1) errors:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + u_t \\
u_t &= \rho u_{t-1} + e_t, \ \{e_t, t = 1, \dots, n, \ i.i.d.(0, \sigma^2)\}
\end{aligned}
$$

- Substitute for u_t in the first equation from the second:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + \rho u_{t-1} + e_t \\
&= \beta_0 + \beta_1 x_t + \rho(y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + e_t \\
&= \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + \rho y_{t-1} + e_t
\end{aligned}
$$

- Take ρy_{t-1} to the left hand side to obtain:

$$
(y_t - \rho y_{t-1}) = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + e_t
$$

- This equation now has the same parameters β_0 and β_1, and its errors do not suffer from serial correlation!

Solution 2: Estimate by FGLS

- So, if ρ was known, an OLS regression of $(y_t - \rho y_{t-1})$ on $(1 - \rho)$ and $(x_t - \rho x_{t-1})$ could give us BLUE for β_0 and β_1.
- This is called a generalised least squares (GLS) estimator, but alas it is not feasible because we do not know ρ :-(
- We need to estimate ρ at the same time as estimating β_0 and β_1
- Different algorithms were suggested for this purpose, notably the Cochrane-Orcutt algorithm (Donald Cochrane was the foundation Dean of the Faculty of Business and Economics at Monash)
- With the advances in computing, there is no need to study various algorithms in detail any more.

Feasible GLS (FGLS)

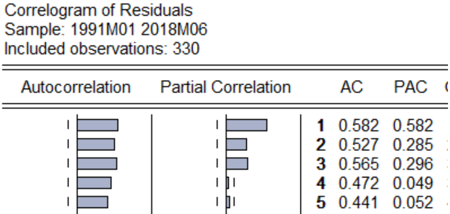
- Using some optimisation algorithm, computers find $\hat{\rho}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimise the sum of squared residuals in the model

$$y_t = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + \rho y_{t-1} + e_t \tag{5}$$

- These are called the feasible generalised least squares (FGLS) estimators.
- Since model (5) is non-linear in parameters, the FGLS estimator is not unbiased.
- However, the FGLS estimator is *consistent*, meaning that if sample is large, the probability of it being far from the truth is small (more on consistency later in this unit)
- Also in large samples FGLS estimator of β_0 and β_1 have smaller variance (are *asymptotically more efficient*) than the OLS estimator of these parameters.
- Finally, the hypothesis tests which are based on the FGLS standard errors are valid when sample is large (i.e. *asymptotically*).

Feasible GLS (FGLS) in practice

- To use FGLS in practice, we need to determine the autoregressive order of the errors.
- Breusch-Godfrey test regression can give us an idea about the dependence in errors (how many of the lags are significant)
- Another easy visual aid is the correlogram of the residuals.
- The correlogram provides information about autocorrelations and partial autocorrelations.
- The significant partial autocorrelations show the order of the AR process that can produce the pattern of correlation seen in the autocorrelations.
- Example: Correlogram of residuals of Victoria visitors model:



Correlogram of Residuals
Sample: 1991M01 2018M06
Included observations: 330

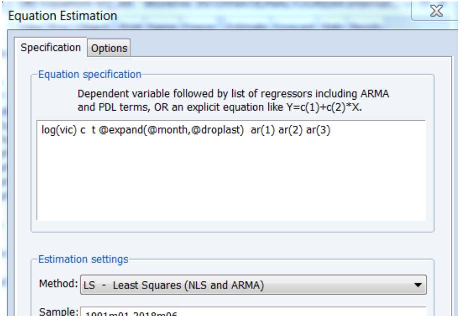
Autocorrelation	Partial Correlation		AC	PAC
		1	0.582	0.582
		2	0.527	0.285
		3	0.565	0.296
		4	0.472	0.049
		5	0.441	0.052

Feasible GLS (FGLS) - Example

- AR(3) model for errors should be adequate.
- So we amend the Victorian international tourist model with an AR(3) equation for errors, resulting in:

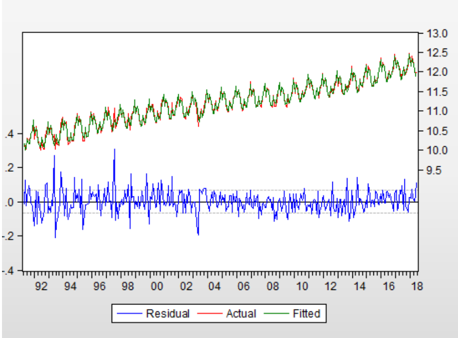
$$\log(VIC) = \beta_0 + \beta_1 t + \beta_2 JAN_t + \beta_3 FEB_t + \cdots + \beta_{12} NOV_t + u_t$$
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + e_t$$

  where e_t is homoskedastic and not serially correlated.
- Estimating this model in EViews can be done by simply adding 'ar(1) ar(2) ar(3)' to the end of the estimation command.



Feasible GLS (FGLS) - Example

- Always a good idea to check if the residuals of the amended model are now “white noise” (i.e. not autocorrelated).



Sample: 1991M04 2018M06
Included observations: 327
Q-statistic probabilities adjusted for 3 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC
		1	-0.017	-0.017
		2	-0.028	-0.029
		3	-0.071	-0.072
		4	-0.013	-0.016
		5	-0.031	-0.036

The implication of autocorrelated errors for prediction

- A very important implication of this model is for prediction (forecasting).
- The last month of the sample in June 2018. The last value for t is 330. If we want to use this model for forecasting visitors to Victoria in July 2018, we compute:

$$\widehat{\log(VIC)}_{331} = \widehat{\beta}_0 + \widehat{\beta}_1 \times 331 + \widehat{\beta}_8 \times JUL_{331} + \hat{\rho}_1 \hat{u}_{330} + \hat{\rho}_2 \hat{u}_{329} + \hat{\rho}_3 \hat{u}_{328}$$

where:

$$\hat{u}_{330} = \log(VIC)_{330} - \widehat{\beta}_0 - \widehat{\beta}_1 \times 330 - \widehat{\beta}_7 \times JUN_{330}$$
$$\hat{u}_{229} = \log(VIC)_{229} - \widehat{\beta}_0 - \widehat{\beta}_1 \times 229 - \widehat{\beta}_6 \times MAY_{229}$$
$$\hat{u}_{228} = \log(VIC)_{228} - \widehat{\beta}_0 - \widehat{\beta}_1 \times 228 - \widehat{\beta}_5 \times APR_{228}$$

Summary

- We focused on serial correlation in the error term of a linear regression model:
- How to define serial correlation
- What implications does the existence of serial correlation have on the properties of the OLS estimator
- How to detect it (Breusch-Godfrey test)
- How to correct for it:
  - HAC standard errors
  - FGLS estimation
  - Use a more general dynamic specification (next lecture)