Introductory Econometrics

Modelling Dynamics

Monash Econometrics and Business Statistics

Semester 2, 2018

Recap

- We looked at the concept of 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
  - Using a more flexible dynamic specification
- Serial correlation is predominantly detected in time series analysis

Lecture Outline

- Dynamic specification: why is it important and what is a well specified dynamic model?
- Temporal dependence in the dependent and/or independent variables of a regression equation
- The interpretation of parameters of a dynamic model: the impact and long-run effects of x on y
- We introduce the following concepts that are important for dynamic models:
  - Consistent and asymptotically normal estimators (more on this next week)
  - Stationary versus non-stationary time series [textbook 11-1; p.381-384]
  - Highly persistent time series with a "unit root" [textbook 11-2; p.391-395]

Learning Goals

- To appreciate that modelling dynamics is important and useful;
- To understand that in dynamic models the OLS estimator is no longer unbiased (and therefore won't be BLUE);
- To get the message that as long as variables are stationary (or trend stationary) and dynamics are well-specified (no autocorrelation left in errors), then the OLS estimator will be reasonably accurate and we can use all of our inferential tools (such as the usual t and F tests) as long as the sample size is large;
- To get the message that if a variable has a "unit root" we can make it stationary by differencing and then use our modelling skills to model its first difference.

Modelling dynamics

- With cross-sectional data, our goal is to model $E(y_i \mid x_i)$ (assume only one x for simplicity)
- With time-series data, our goal is to model $E(y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots)$
- This is because:
  1. time series variables are often persistent (autocorrelated), so we can use the lagged values of y_t to predict y_t
  2. the effect of a change in x on y may not be immediate (e.g. the effect of an interest rate change on housing construction)

Why do dynamic models make sense?

- Some factors which can create dynamics in how explanatory variables affect a target variable are:
  - Habit persistence (think about your consumption habits).
  - Institutional arrangements, such as legally binding contracts, which constrain the ability of consumers and firms to immediately adjust their in response to changes in their economic environment (think about employment contracts).
  - Administrative lags (think about permits required to build a house).
  - Gradual adjustment is prudent behaviour in the face of uncertainty (think about Monash reacting to changes in year to year variations in international student numbers).
- All of these suggest that the effect of a change in x on y may not be immediate and it may take some time for the full effect to be realised. We need a model that can measure such dynamics.
- In addition, common sense suggests that we can exploit temporal dependence and use history to construct better forecasting models for the future.

The autoregressive distributed lag model

- The simplest dynamic model of relationship between y and x is

$$E(y_t \mid x_t, y_{t-1}, x_{t-1}, \ldots) = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} \quad (1)$$

  which we can write it as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + u_t \quad (2)$$

  with $E(u_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots) = 0$
- This model is called an "autoregressive distributed lag model of order 1,1" or ARDL(1,1) for short.
- It is clear that the art of dynamic specification is to figure out how many lags of y_t and x_t we need to include so that the errors have no dependence on the past.
- It is also clear that with time series data we can have predictive models that have no x variables! These are called univariate time series models, which can be very useful for forecasting.

Univariate time series models

The autoregressive (AR) model

- The most popular univariate time series model is the AR(p) model

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + u_t, \tag{3}$$

  where u_t is a "white noise process" – a white noise process is an uncorrelated sequence of random variables with mean zero and constant finite variance, we use the short-hand $\{u_t\} \sim WN\left(0, \sigma^2\right)$.
- The number of lags, p, is called the *lag length* or *order* of the model.
- The simplest AR model is the AR(1) model

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + u_t.$$

- This model can produce:
  1. an uncorrelated sequence when $\varphi_1 = 0$
  2. a stationary (mean reverting) process when $|\varphi_1| < 1$
  3. a random walk when $\varphi_0 = 0$ and $\varphi_1 = 1$
  4. a random walk with drift when $\varphi_0 \neq 0$ and $\varphi_1 = 1$
  5. an explosive process when $\varphi_1 > 1$

Concept of stationarity and mean reversion

- History is only practically useful if it has recurring patterns
- We already know how to model periodic patterns like seasonality
- Mean reversion (that a time series returns to its mean often) is a recurring pattern of stationary time series.
- A time series $\{y_t\colon\ t = 1, 2, \ldots\}$ is said to be *weakly stationary* (or *covariance stationary*) if:
  1. $E(y_t) = \mu$ (a finite constant) for all $t \Rightarrow$ mean is time invariant
  2. $Var(y_t) = E[(y_t - \mu)^2] = \gamma_0$ (a finite constant) for all $t \Rightarrow$ variance is time invariant
  3. $Cov(y_t, y_{t+j}) = E[(y_t - \mu)(y_{t+j} - \mu)] = \gamma_j$ for all t and $j \Rightarrow$ the covariance between any two observations depends only on the time interval separating them and not on time itself.
     - Eg. $Cov(y_1, y_{10}) = Cov(y_{10}, y_{19}) = Cov(y_{101}, y_{110})$
     - It follows that
       $$Corr(y_t, y_{t+j}) = \rho_j = \frac{\gamma_j}{\gamma_0}$$
       i.e. autocorrelations also only depend on j and not t.

The stationary AR(1) process

- If y_t is generated by an AR(1) model

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + u_t, \text{ with } |\varphi_1| < 1 \text{ and } \{u_t\} \sim WN(0, \sigma^2)$$

then we have:

P1 Mean of y

$$E(y_t) = \frac{\varphi_0}{1 - \varphi_1} \text{ for all } t$$

P2 Variance of y

$$Var(y_t) = \frac{\sigma^2}{1 - \varphi_1^2} \text{ for all } t$$

P3 Autocovariances and autocorrelations of y

$$Cov(y_t, y_{t-j}) = \gamma_j = \varphi_1^j Var(y_t), \text{ for all } t \text{ and } j$$

$$Corr(y_t, y_{t-j}) = \rho_j = \frac{\gamma_j}{\gamma_0} = \varphi_1^j, \text{ for all } t \text{ and } j.$$

Estimating dynamic models

- Would OLS estimator of parameters of equation (2) be unbiased?
- No. Because $E(\mathbf{u} \mid \mathbf{X})$ in this model is not zero. To see this, let's write it in matrix form:

$$\mathbf{y} = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_t \\ y_{t+1} \\ \vdots \\ y_n \end{pmatrix}, \ \mathbf{u} = \begin{pmatrix} u_2 \\ u_3 \\ \vdots \\ u_t \\ u_{t+1} \\ \vdots \\ u_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & y_1 & x_2 & x_1 \\ 1 & y_2 & x_3 & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{t-1} & x_t & x_{t-1} \\ 1 & y_t & x_{t+1} & x_t \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{n-1} & x_n & x_{n-1} \end{pmatrix}$$

- As we can see y_t is in the $\mathbf{X}$ matrix, so $E(u_t \mid \mathbf{X}) \neq 0$.
- Since $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ is necessary for the OLS estimator to be unbiased, we can conclude that the OLS estimator will not be an unbiased estimator of parameters of a dynamic model.

Estimating dynamic models

The OLS estimator is OK in large samples

- However, as long as y and x are stationary (at least around a trend) and the model is "dynamically complete", i.e. $E(u_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots) = 0$, the OLS estimator will be consistent and asymptotically normal.
- This means that as long as the sample is large, OLS will be reliable and we can use the usual tools such as t and F test for hypothesis testing.
- We study consistency and asymptotic normality in more detail in the next lecture.
- Therefore, when estimating a model given data, we have to make sure that variables are stationary (at least around a deterministic trend) and the residuals show no sign of correlation with the past.

Example: Reserve Bank's reaction function

- How does the Reserve Bank of Australia (RBA) set the cash rate? It makes sense that the Reserve Bank reacts to data on inflation and growth rate. A regression of the cash rate on inflation and growth rate yields:

Dependent Variable: CRATE
Method: Least Squares
Sample (adjusted): 1990Q4 2016Q2
Included observations: 103 after adjustments

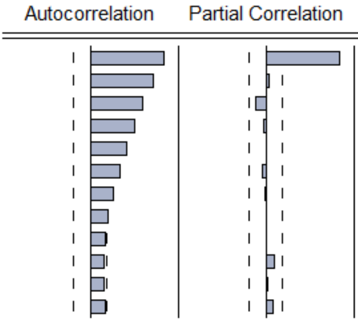
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.140237	0.411475	12.49221	0.0000
GDPGR	-0.102513	0.082255	-1.246288	0.2156
INFL	0.167738	0.089400	1.876276	0.0635

R-squared	0.057144	Mean dependent var	5.237379
Adjusted R-squared	0.038287	S.D. dependent var	1.974848
S.E. of regression	1.936673	Akaike info criterion	4.188514
Sum squared resid	375.0702	Schwarz criterion	4.265253
Log likelihood	-212.7085	Hannan-Quinn criter.	4.219596
F-statistic	3.030379	Durbin-Watson stat	0.129477
Prob(F-statistic)	0.052755		

Example: continued

- Neither inflation nor growth rate seems to be significant. However, the residual correlogram shows:

Sample: 1990Q4 2016Q2
Included observations: 103



	AC	PAC
1	0.863	0.863
2	0.754	0.034
3	0.630	-0.111
4	0.521	-0.026
5	0.430	0.015
6	0.345	-0.038
7	0.275	-0.001
8	0.219	0.008
9	0.173	0.000
10	0.162	0.099
11	0.157	0.030
12	0.178	0.088

which suggests that the t-statistics and p-values from the static model were not correct

Example: continued

- This also makes sense because the RBA only changes the cash rate very smoothly. So, we add one lag of every variable to the right hand side and we obtain:

Dependent Variable: CRATE
Method: Least Squares
Sample (adjusted): 1991Q1 2016Q2
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.337620	0.151658	-2.226194	0.0283
CRATE(-1)	0.915817	0.018736	48.88083	0.0000
GDPGR	0.055108	0.015714	3.506963	0.0007
GDPGR(-1)	0.041922	0.015123	2.772032	0.0067
INFL	0.090080	0.017387	5.180861	0.0000
INFL(-1)	0.062562	0.017297	3.616874	0.0005

R-squared	0.964622	Mean dependent var	5.161242
Adjusted R-squared	0.962780	S.D. dependent var	1.826377
S.E. of regression	0.352354	Akaike info criterion	0.808664
Sum squared resid	11.91875	Schwarz criterion	0.963075

which shows parameter estimates that make better sense.

Example: continued

- However, the residuals still show sign of autocorrelation:

Sample: 1991Q1 2016Q2
Included observations: 102

Autocorrelation	Partial Correlation		AC	PAC
		1	0.453	0.453
		2	0.063	-0.180
		3	-0.012	0.047
		4	-0.089	-0.119
		5	-0.062	0.041
		6	-0.085	-0.105
		7	-0.112	-0.038
		8	-0.154	-0.129
		9	-0.092	0.040
		10	-0.064	-0.088
		11	-0.113	-0.086
		12	-0.052	0.010

with value of the Breusch-Godfrey test statistic for testing the null of no serial correlation in errors against the alternative of first order serial correlation being 26.1, clearly rejecting the null.

Example: continued

- Adding the second lag takes care of serial correlation in errors. Dropping the insignificant lags leaves us with

Dependent Variable: CRATE
Method: Least Squares
Sample (adjusted): 1991Q2 2016Q2
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.027022	0.109588	0.246581	0.8058
CRATE(-1)	1.523767	0.076624	19.88626	0.0000
CRATE(-2)	-0.585687	0.071408	-8.201939	0.0000
GDPGR	0.048512	0.013588	3.570111	0.0006
INFL	0.042447	0.016436	2.582600	0.0113

R-squared	0.971664	Mean dependent var	5.093531
Adjusted R-squared	0.970483	S.D. dependent var	1.701967
S.E. of regression	0.292405	Akaike info criterion	0.426886
Sum squared resid	8.208084	Schwarz criterion	0.556347

with the value of $BG_{calc} = 1.2$ for the null of no serial correlation in errors against first order serial correlation, which is well below the 5% critical value of χ^2_1.

Interpretation of parameters in dynamic models

- In a regression with AR errors such as

$$y_t = \beta_0 + \beta_1 x_t + u_t$$
$$u_t = \rho u_{t-1} + e_t, \quad t = 2, 3, \ldots, n,$$

  the interpretation of β_1 is the same as in the static model: β_1 is the expected change in y as x increases by 1 unit.
- The autoregressive dynamics of u_t is all about part of y that is not explained by x.
- However, in the autoregressive distributed lag model

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 y_{t-1} + u_t, \quad t = 2, \ldots, n,$$

  β_1 is the contemporaneous change in y as x changes by 1 unit. But, that is not the end of the story.
- At time $t+1$, this change is x will still affect y_{t+1} in two ways: once directly because x_t still appears in the equation for y_{t+1}, and again because y_t also appears in the equation for y_{t+1} and y_t was already influenced by the change in x_t.

Interpretation of parameters in dynamic models

Immediate and long-run effects

- So what is the long-run impact of a one unit increase in x at time t on y? It turns out that the long-run impact will be:

$$\text{long-run effect of a one unit increase in } x \text{ at time } t = \frac{\beta_1 + \beta_2}{1 - \beta_3}$$

  or in general the long-run effect of a unit increase in x at time t

$$= \frac{\text{sum of the coefficients } x_t \text{ and its lags}}{1 - \text{sum of the coefficients of lags of } y_t}$$

- You are required to use this formula, but not prove it.
- So the estimated model estimates the immediate effect of a unit increase in x at time t on y given by the coefficient of x_t (this is sometimes referred to as the “impact multiplier”) and the long-run effect of a unit increase in x at time t on y given by the formula above (this is sometimes referred to as the “long-run multiplier”).

Example: continued

- Our estimated reaction function of the RBA was

$$\widehat{CRATE}_t = 0.027 + 1.524 CRATE_{t-1} - 0.586 CRATE_{t-2} \\ + 0.049 GDPGR_t + 0.042 INFL_t$$

- According to this estimated equation:
  - the immediate impact of a 1 percentage point increase in the annualised GDP growth on the cash rate is ...
  - the immediate impact of a 1 percentage point increase in the annualised inflation rate on the cash rate is ...
  - the long-run impact of a 1 percentage point increase in the annualised GDP growth on the cash rate is ...
  - the long-run impact of a 1 percentage point increase in the annualised inflation rate on the cash rate is ...

The restrictive dynamics of regression with AR errors

- What is the difference between a regression with AR errors and a general dynamic model? Recall that

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

  which is a restricted ARDL model (what is the restriction?)
- In this model a one unit increase in x changes y
  - by β_1 units immediately, and
  - by $\frac{\beta_1 - \rho\beta_1}{1-\rho} = \frac{\beta_1(1-\rho)}{1-\rho} = \beta_1$ in the long-run!
- This shows that the regression with AR errors imposes that all impact of x on y is realised immediately. All the dynamics in these models are in the error (part of y that is not explained by x)

Unit roots

- Consider the ARDL(1,1) model again:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 y_{t-1} + u_t, \quad t = 2, \ldots, n,$$

  where the long-run effect of a one unit increase in x on y is

$$\frac{\beta_1 + \beta_2}{1 - \beta_3}$$

- Imagine what would happen if β_3 was equal to 1. The long-run effect will be infinite!
- In this case, y is said to have a "unit root", and is non-stationary.
- This means that we cannot use our usual tools to estimate parameters and test hypothesis about them
- Before looking at what to do in such cases, let's study the simplest unit root process: "the random walk"

The Random Walk process

- Recall the AR(1) process discussed earlier (setting $\varphi_0 = 0$):

$$y_t = \varphi_1 y_{t-1} + u_t. \tag{4}$$

- The stationarity condition for the AR(1) specification was given by

$$|\varphi_1| < 1.$$

- The "random walk" process is the above with $\varphi_1 = 1$, i.e.

$$y_t = y_{t-1} + u_t, \tag{5}$$

  where $u_t \sim WN\left(0, \sigma^2\right)$.

- The sum of the coefficients of lagged dependent variables in (5) is 1, in which case we say that the random walk has a "unit root".

Properties of the Random Walk

- A random walk process is non-stationary (i.e. it is not covariance stationary) because:

$$\begin{aligned} y_t =& y_{t-1} + u_t \\ =& y_{t-2} + u_{t-1} + u_t \\ =& y_0 + u_1 + u_2 + \ldots + u_{t-1} + u_t \end{aligned}$$

- This shows that the effect of the initial condition y_0 never fades away (random walk has "long memory"). Let's assume $y_0 = 0$ for simplicity, then
  - the mean of the random walk is:

$$E\left(y_t\right) = E\left(u_1 + u_2 + \ldots + u_{t-1} + u_t\right) = 0,$$

  - and its variance is:

$$\begin{aligned} Var\left(y_t\right) &= Var\left(u_1 + u_2 + \ldots + u_{t-1} + u_t\right) \\ &= \sigma^2 + \sigma^2 + \ldots + \sigma^2 = t\sigma^2 \end{aligned}$$

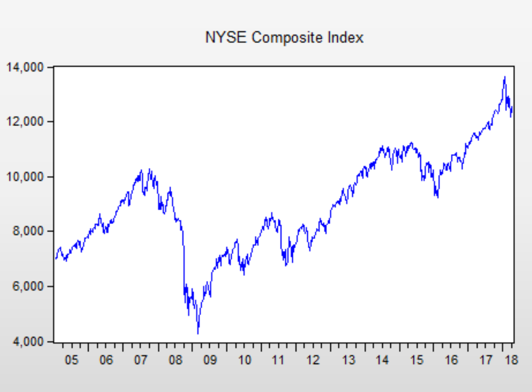
- Therefore, the mean is constant over time but the variance depends on t and increases to infinity as $t \to \infty$.
- Hence the random walk is non-stationary.

The practical importance of unit roots

- Are unit root processes a theoretical curiosity for mathematicians with no practical importance?
- Unfortunately not. They are prevalent in finance and economics.
- How can we tell if a time series has a unit root?
- A time series with a unit root is not mean reverting, so if we plot the data, and observe that the time series crosses its mean only a handful of times, then we can conclude that it has a unit root.
- However, a trend-stationary time series also does not cross its sample mean. So, if the plot shows a deterministic trend, we run a regression on a trend and then look at the residuals to see if they are mean reverting or not.
- Also, a time series with a unit root is highly persistent. This shows up in a correlogram that decays very slowly, and a first order autocorrelation that is very close to 1.
- There are formal tests for unit roots that we do not cover here (they will be covered in ETC3450 and ETC3460).

Example 1: NYSE Composite Index

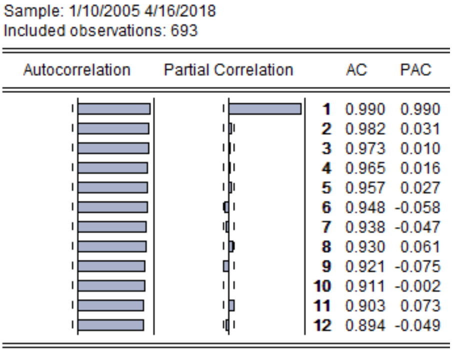
- NYSE composite index weekly data (an index made of prices at the end of the week adjusted for dividends) from 2005 to mid-2018



- Obviously this series
  - is not mean reverting, and
  - its non-stationarity is not due to a time trend

Example 1 continued

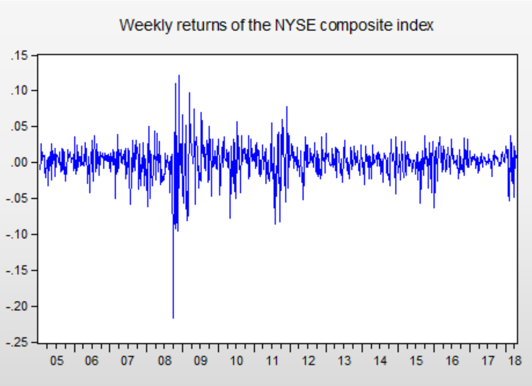
- A look at the correlogram of this series confirms that it has a unit root:



- Note that
  - the first order autocorrelation is very close to 1
  - the autocorrelations decay very slowly

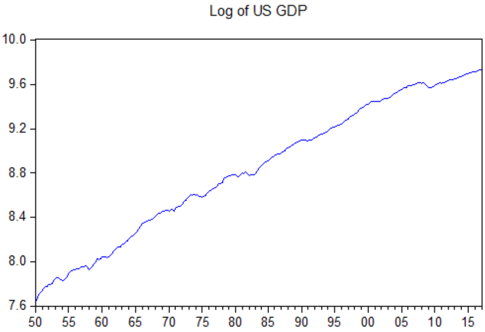
Example 1 conclusion

- Should we be flustered that we cannot use our econometric tools to model this series?
- Not really, because we are not interested in the value of the index
- We are interested in returns $\Delta \log(NYSE)$, and first differencing gets rid of the unit root!



Example 2: US real GDP

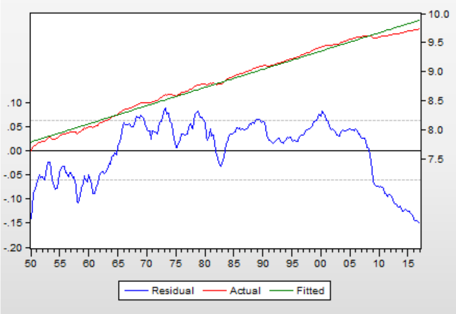
- Consider US real Gross Domestic Product (GDP) over the period Q1:1950 to Q1:2017 (269 observations) that we examined in last week's lecture.
- A line graph of the (log of) US GDP is given below:



- Persistence in this time series is evident, but it can just be a linear time trend

Example 2 continued

- However, after removing the linear trend, the residuals still show high persistence:



Example 2 continued

- And the correlogram of these residuals suggests a unit root:

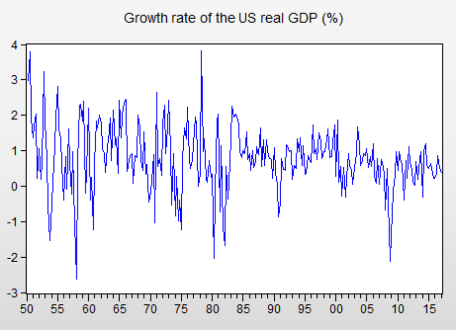
Sample: 1950Q1 2017Q1
Included observations: 269

Autocorrelation	Partial Correlation		AC	PAC
		1	0.968	0.968
		2	0.930	-0.093
		3	0.893	-0.014
		4	0.856	-0.016
		5	0.821	0.008
		6	0.790	0.042
		7	0.762	0.020
		8	0.734	-0.023
		9	0.707	0.001
		10	0.677	-0.053
		11	0.646	-0.033
		12	0.617	0.030

- So, the log of US real GDP has a unit root as well as a deterministic trend

Example 2 continued

- Should we be flustered that we cannot use our econometric tools to model log of real GDP?
- Not really, because we are interested in the growth rate of real GDP, i.e. $100\Delta \log(GDP)$, and first differencing gets rid of both deterministic trend as well as the unit root!



- We can use all tools that we have learnt to model this series.

Example 2 continued

- For example, we can find the best univariate AR model for US real GDP growth by starting with an AR(4) - since data is quarterly.
- If there is serial correlation left in the residuals, increase the lag length until residuals become white noise.
- If there isn't any serial correlation in the residuals, use t-test to indentify insignificant lags and drop them one by one, until there is no insignificant lags left, or use model selection criteria to choose between this and lower order models. When the best model is chosen, test again to make sure that no serial correlation has seeped into the residuals.
- This process will lead to the following AR model:

$$\Delta \log(GDP_t) = \underset{(0.07)}{0.49} + \underset{(0.06)}{0.36} \Delta \log(GDP_{t-1}) + \hat{u}_t$$

$$R^2 = 0.13, \;\; n = 259.$$

- There is no evidence of serial correlation in the errors of this model.

Summary of the main points

- With time series data we can build a dynamic model by adding lags of dependent and independent variables to the list of explanatory variables.
- As long as the dependent and independent variables are stationary and errors are white noise, the OLS estimator of the parameters of a dynamic model is reliable and we can use t and F tests provided the sample size is large.
- Dynamic models give us insights about the immediate impact and the long-run effect of a change in each of the independent variables on the dependent variable, all else constant.

Summary of the technical points

- A stationary time series has:
  1. constant mean over time
  2. constant variance over time
  3. the covariance between any two observations depends only on the time interval separating them and not on time itself.
- A stationary time series in which the covariance between any two observations is zero is called a "white noise".
- If any of the above conditions are violated then the time series is said to be non-stationary.
- If non-stationarity is due to a linear trend in mean, we can add a time trend to the model and proceed as usual.
- If non-stationarity is due to a unit root (perhaps as well as a linear trend), we difference the data and proceed with modelling the differenced data.
- We identify a unit root in a time series from its plot and its correlogram. The series will not be mean-reverting, will have first order autocorrelation close to 1, and its autocorrelations decay very slowly.