Further Issues in Using OLS with Time Series Data. Wooldridge (2013), Chapter 11

- Covariance Stationary Process
- Weakly Dependent Time Series
- Examples for weakly dependent time series
  - moving average process of order one
  - autoregressive process of order one
- Assumptions for consistency and asymptotic normality of OLS
- Autoregressive distributed Lag model
- Random Walks
- Transforming Persistent Series
- Dynamically Complete Models and the Absence of Serial Correlation

Further Issues in Using OLS with Time Series Data.

- So far we showed that under Assumptions TS.1 to TS.6 the Ordinary Least Squares estimator has exactly the same properties that in the cross sectional case.
- However, these Assumptions are *very strong*, and will not be satisfied in some models (e.g. Assumption TS.3 [$E(u_t|X) = 0$] does not hold if we have lagged dependent variables as regressors).
- It will be shown that under a different set of Assumptions the inference procedures introduced in the cross-sectional case can be used in such models in the time series context.
- In some cases we would like to have as regressors lagged dependent variables.

Example: *Partial adjustment model*: Suppose y_t^* is the desired level of inventories of a firm and y_t is the actual level and x_t is the sales. Assume that the desired level of inventories depends of sales plus a error term

$$y_t^* = \alpha + \beta x_t + v_t.$$

Because of frictions in the market the gap between the actual and desired level cannot be closed instantaneously but only with some lag. That is the inventory in time t would equal that at time $t-1$ plus an adjustment factor.

$$y_t = y_{t-1} + \lambda(y_t^* - y_{t-1}),$$
$$0 < \lambda < 1$$

In this case combining these two equations we obtain

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 x_t + u_t$$

where $\gamma_0 = \alpha\lambda$, $\gamma_1 = (1-\lambda)$,$\gamma_2 = \beta\lambda, u_t = \lambda v_t$.

Further Issues in Using OLS with Time Series Data.

Main points:

- If lagged dependent variables are included as regressors OLS is *biased* and **is** *not BLUE* (the Gauss Markov theorem does not hold).
- The justification for the use of the inference procedures relies on *Large Sample analysis* (If lagged dependent variables are included as regressors OLS is *consistent* and *Asymptotically normal*).

Stability

- We need to assume *stability*: If we allow the relationship between two variables (say y_t and x_t) to change arbitrarily each time period we cannot hope to learn much about how a change in one variable affects the other variable if we only have access to a single realization, hence we require a definition of stability (in time series this is given by the notion of *stationary processes*).

Dependence

- In the cross sectional case we had a random sample $\{X_1, X_2, ..., X_n\}$. Each random variable X_1 to X_n were independent with the same mean μ_X and finite variance σ_X^2. Under independence a Law of Large numbers (LLN) and a Central limit theorem (CLT) would be valid. That is,

$$\text{plim } \bar{X} = \mu_X, \textbf{ (Law of Large numbers)}$$

$$\sqrt{n}\frac{(\bar{X} - \mu)}{\sigma_X} \overset{a}{\sim} N(0,1), \textbf{ (Central limit theorem)}$$

- However in the time series context we do not have a random sample as the random variables X_1 to X_n are not independent.
- *Dependence* is typical with time series: what happens this period and what happened last period are related, as is what happened the period before last and so on.
- Thus with time-series data we typically have to deal with *dependence* between the observations (with cross-section we typically do not have to deal with such dependence).
- We have to introduce key concepts that address these issues.

Definition

A stochastic process $\{x_t, t = 1, ..\}$ is *covariance stationary* if

- $E(x_t)$ is constant (does not vary with t)
- $Var(x_t)$ is constant,
- for any $h \neq 0$, $Cov(x_t, x_{t+h})$ depends only on h and not on t.

- **Example:** The process $\{\varepsilon_t, t = 1, ...\}$ such as $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$ and $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0$ with $h \neq 0$ is known as a *white noise process*. It is covariance stationary.

- Stationarity is important in time series because
  - if we want to understand the relationship between two or more variables using regression analysis we need to assume some sort of *stability* of the relationship over time.
  - it also simplifies the assumptions required for a *LLN* and a *CLT* to hold.

We need to replace the concept of independence by a different concept in time series.

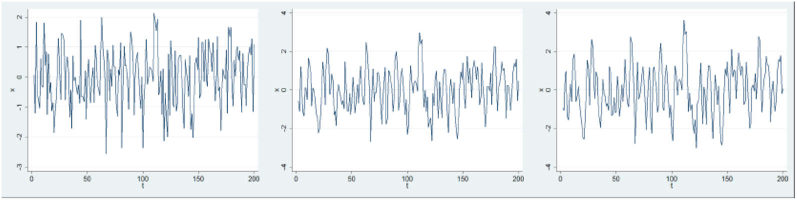
- A *stationary time series* is *weakly dependent* if x_t and x_{t+h} are "almost uncorrelated" as h increases.
- If for a covariance stationary process $Corr(x_t, x_{t+h}) \rightarrow 0$ as $h \rightarrow \infty$, we'll say this covariance stationary process is *weakly dependent*.
- We need weak dependence for LLN's and CLT's to hold.
- We won't give a rigorous technical definition of weak dependent process.

- A stochastic process is a *moving average process of order one*, MA(1), if

$$x_t = e_t + \theta_1 e_{t-1}, t = 1, 2, \ldots$$

with e_t being a white noise process with variance σ_e^2.
- This is a stationary, weakly dependent sequence as variables 1 period apart are correlated, but 2 periods apart they are not.
- Notice that
  - $E(x_t) = 0$.
  - $Var(x_t) = \sigma_e^2(1 + \theta_1^2)$.
  - $Corr(x_t, x_{t-1}) = \theta_1/(1 + \theta_1^2)$.
  - $Corr(x_t, x_{t-h}) = 0, h \geq 2$

Figure: MA(1) with $\theta_1 = 0.1,\ \theta_1 = 0.6,\ \theta_1 = 0.9$



- An *autoregressive process of order one* , AR(1), can be characterized as one where

$$x_t = \rho_1 x_{t-1} + e_t,$$

  $t = 1, 2, \ldots$with e_t being a white noise process.
- A necessary condition for an AR(1) process to be stationary and weakly dependent is that $-1 < \rho_1 < 1$: an $AR(1)$ process satisfying this condition is called stable.

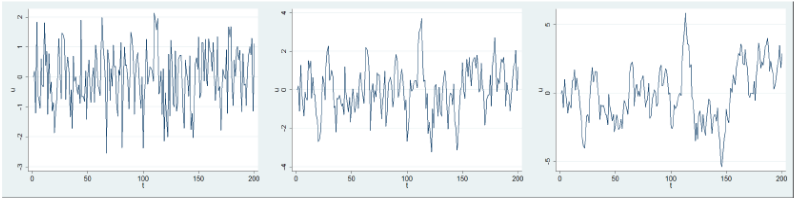
It is possible to show that if $|\rho_1| < 1$:

- $E(x_t) = 0$.
- $Var(x_t) = \frac{\sigma_e^2}{1-\rho_1^2}$
- $Corr(x_t, x_{t+h}) = Cov(x_t, x_{t+h}) / Var(x_t) = \rho_1^h$ which becomes small as h increases.

Further Issues in Using OLS with Time Series Data.

An AR(1) Process

Figure: AR(1) with $\rho_1 = 0.1,\ \rho_1 = 0.6,\ \rho_1 = 0.9$



$$y_t = \alpha + \beta t + u_t,$$

where $E(u_t) = 0$.

- A *trending series* cannot be stationary, since the mean is changing over time $E(y_t) = \alpha + \beta t$.
- A trending series is weakly dependent if u_t is weakly dependent.
- If a u_t is weakly dependent and stationary, we will call y_t a *trend-stationary process*.
- As long as a trend is included, all is well.

The following Assumptions are required to show that the OLS estimator is consistent.

Assumption (TS.1′ - linearity in parameters)

The stochastic process $\{(y_t, x_{t1}, x_{t2}, ..., x_{tk}); t = 1, 2, ..., n\}$ *is stationary and weakly dependent and follows the linear model:*

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

- Note that some regressors may be lagged values of other regressors or lagged values of y.

Assumption (TS.2′ - no perfect collinearity)

No regressor independent variable is a constant nor a perfect linear combination of the other regressors.

Write $\mathbf{x}_t = (x_{t1,...,}x_{tk})$.

Assumption (TS.3′ - zero conditional mean)

$$E(u_t|\mathbf{x}_t) = 0, t = 1, 2, \ldots, n$$

Theorem

Under assumptions TS.1′ through TS.3′ the OLS estimator is consistent:

$$\begin{aligned} plim\ \hat{\beta}_j &= \beta_j, \\ j &= 0, 1, ...k \end{aligned}$$

Remarks:

- Thus, can have correlation between u_{t-1} and x_t (OK if $y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t$, where y_{t-1} naturally depends on u_{t-1}). Can have feedback from y (or u) to future values of the regressors.
- Weaker assumptions than those for unbiasedness.
- Main difference: Assumption TS.3 does not hold with lagged dependent variables as regressors ($E(u_t|X) \neq 0$) hence OLS is biased. However, in this case Assumption TS.3' holds and therefore it is consistent.

Why do lagged dependent variables violate strict exogeneity?
Consider the model

$$y_t = \alpha_1 y_{t-1} + u_t \tag{1}$$

This is the simplest possible regression model with a lagged dependent variable

- *Contemporaneous exogeneity*: $E(u_t|y_{t-1}) = 0$
- *Strict exogeneity*: $E(u_t|y_0, y_1, y_2, ..., y_{n-1}) = 0$.
- Strict exogeneity would imply that $cov(y_t, u_t) = 0$ for all $t = 1, \ldots, n-1$
- But this is incompatible with model (1) as this leads to a contradiction:
  - Notice that equation (1) implies that for all $t = 1, \ldots, n-1$

    $$cov(y_t, u_t) = \alpha_1 cov(y_{t-1}, u_t) + var(u_t)$$

  - On the other hand, strict exogeneity implies also that $cov(y_{t-1}, u_t) = 0$, it follows that $cov(y_t, u_t) = var(u_t) > 0$, hence contradiction!
- The solution to this problem is to drop the assumption of strict exogeneity and assume *contemporaneous exogeneity*.

Let us consider:

Assumption (TS.4′ - contemporaneous homoskedasticity)

For each $t = 1, 2, ..., n$:

$$Var(u_t|\mathbf{x}_t) = Var(u_t) = \sigma^2$$

Assumption (TS.5′ - No Serial Correlation)

For each $t; s = 1, 2, ..., n$ *such that* $t \neq s$:

$$Corr(u_t, u_s|\mathbf{x}_t, \mathbf{x}_s) = 0$$

Theorem

With assumptions TS.1′ through TS.5′, we have asymptotic normality of the OLS estimators. The usual standard errors, t statistics, F statistics and LM statistics are valid asymptotically (that is if the sample size is large).

Now we have the tools to estimate a dynamic model with lagged dependent variables as regressors.
The *Autoregressive distributed Lag model* is given by

$$y_t = \alpha + \sum_{i=0}^{q} \beta_{i+1} x_{t-i} + \sum_{i=1}^{p} \gamma_1 y_{t-i} + u_t$$

Let us consider a simple case

$$y_t = \alpha + \gamma_1 y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + u_t,$$
$$|\gamma_1| < 1$$

We are going to study the change in y_t, y_{t+1}, y_{t+2} as x changes *temporarily* or *permanently* in period t :
The change in y_t as x changes *temporarily* in period t :

$$\frac{\partial y_t}{\partial x_t} = \beta_1$$

The change in y_t as x changes *permanently* in period t :

$$\frac{\partial y_t}{\partial x_t} = \beta_1$$

Now in period $t+1$:

$$y_{t+1} = \alpha + \gamma_1 y_t + \beta_1 x_{t+1} + \beta_2 x_t + u_t$$

The change in y_{t+1} as x changes *temporarily* in period t :

$$\begin{aligned} \frac{\partial y_{t+1}}{\partial x_t} &= \gamma_1 \frac{\partial y_t}{\partial x_t} + \beta_2 \\ &= \gamma_1 \beta_1 + \beta_2 \end{aligned}$$

The change in y_{t+1} as x changes *permanently* in period t :

$$\frac{\partial y_{t+1}}{\partial x_t} + \frac{\partial y_{t+1}}{\partial x_{t+1}} = \gamma_1 \beta_1 + \beta_2 + \beta_1$$

Now in period $t + 2$:

$$y_{t+2} = \alpha + \gamma_1 y_{t+1} + \beta_1 x_{t+2} + \beta_2 x_{t+1} + u_t$$

The change in y_{t+2} as x changes *temporarily* in period t :

$$\begin{aligned}\frac{\partial y_{t+2}}{\partial x_t} &= \gamma_1 \frac{\partial y_{t+1}}{\partial x_t} \\ &= \gamma_1(\gamma_1\beta_1 + \beta_2)\end{aligned}$$

The change in y_{t+2} as x changes *permanently* in period t :

$$\frac{\partial y_{t+2}}{\partial x_{t+2}} + \frac{\partial y_{t+2}}{\partial x_{t+1}} + \frac{\partial y_{t+2}}{\partial x_t} = \beta_1 + \gamma_1\beta_1 + \beta_2 + \gamma_1(\gamma_1\beta_1 + \beta_2).$$

Long run multiplier

Suppose that the economy were in a *steady state* in which all of the variables were constant over time. Hence $x_t = x_{t-1} = x$, $y_t = y$ and in the steady state $u_t = 0$. The long run relation is given by

$$y = \alpha + \gamma_1 y + \beta_1 x + \beta_2 x$$

Hence the *long-run relationship* is given by

$$y = \frac{\alpha}{1-\gamma_1} + \frac{\beta_1 + \beta_2}{1-\gamma_1} x$$

the *long run multiplier (propensity)* is given by $\frac{\beta_1+\beta_2}{1-\gamma_1}$.

- Inference on $\frac{\beta_1+\beta_2}{1-\gamma_1}$.
- Suppose the null hypothesis is $H_0 : \frac{\beta_1+\beta_2}{1-\gamma_1} = a$, where a is a constant.
- Notice that this is equivalent to $H_0 : \beta_1 + \beta_2 + a\gamma_1 = a$
- Use the *t-statistic*

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1 - a}{se(\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1)}$$

where $se(\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1)$ is the standard error of $\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1$.

- Notice that

$$
\begin{aligned}
Var(\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1) &= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + a^2 Var(\hat{\gamma}_1) \\
&\quad + 2Cov(\hat{\beta}_1, \hat{\beta}_2) + 2a \times Cov(\hat{\beta}_1, \hat{\gamma}_1) \\
&\quad + 2a \times Cov(\hat{\beta}_2, \hat{\gamma}_1).
\end{aligned}
$$

- Hence

$$
\begin{aligned}
se(\hat{\beta}_1 + \hat{\beta}_2 + a\hat{\gamma}_1)^2 &= se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 + a^2 se(\hat{\gamma}_1)^2 \\
&\quad + 2 \times s(\hat{\beta}_1, \hat{\beta}_2) + 2a \times s(\hat{\beta}_1, \hat{\gamma}_1) \\
&\quad + 2a \times s(\hat{\beta}_2, \hat{\gamma}_1),
\end{aligned}
$$

where $se(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$, $se(\hat{\gamma}_1)$ is the standard error of $\hat{\gamma}_1$ and $s(.,.)$ is an estimator of $Cov(.,.)$.

Example: Consumption (*lc*) is expected to depend on income (*ly*) and inflation (*ir*). The latter variable is a proxy for wealth effects.

$$lc_t = \beta_0 + \beta_1 ly_t + \beta_2 ir_t + \beta_3 ly_{t-1} + \beta_4 ir_{t-1} + \beta_5 lc_{t-1} + u_t,$$

We estimate the dynamic model by OLS and obtained the following results:

lct	Coef.	Std. Err.	t
ly_t	0.7619	0.0468	16.2700
ir_t	-0.0724	0.2605	-0.2800
lc_{t-1}	0.8353	0.1169	7.1400
ly_{t-1}	-0.5975	0.1300	-4.6000
ir_{t-1}	0.3000	0.2372	1.2600
intercept	-0.0708	0.1510	-0.4700

R-squared= 0.9979
Number of observations=32
Residual sum of squares= 0.003757558

Compute the change in lc_t, lc_{t+1} as *ly* changes permanently in period t.

The variance covariance matrix of the Ordinary Least squares estimator is given by

	ly_t	ir_t	lc_{t-1}	ly_{t-1}	ir_{t-1}	intercept
ly_t	0.0022					
ir_t	0.0014	0.0679				
lc_{t-1}	-0.0001	0.0191	0.0137			
ly_{t-1}	-0.0019	-0.0216	-0.0143	0.0169		
ir_{t-1}	0.0004	-0.0527	-0.0149	0.0155	0.0563	
intercept	-0.0020	0.0187	0.0134	-0.0131	-0.0165	0.0228

Test the hypothesis that the long-run multiplier of income is equal to 1.

- A *random walk* is an $AR(1)$ model where $\rho_1 = 1$, meaning the series is not weakly dependent:

$$y_t = y_{t-1} + e_t.$$

- e_t is a white noise process with variance σ_e^2.
- With a random walk, the expected value of y_t is always y_0 – it doesn't depend on t
- $Var(y_t) = \sigma_e^2 t$, so it increases with t
- A random walk is *not covariance stationary*.
- We say a random walk is *highly persistent* since $E(y_{t+h}|y_t) = y_t$ for all $h \geq 1$
- Contrast conditional expectation of random walk, $E(y_{t+h}|y_t) = y_t$, with conditional expectation of the stable $AR(1)$ process, $E(y_{t+h}|y_t) = \rho_1^h y_t$.
- For stable $AR(1)$ process ($|\rho_1| < 1$) $E(y_{t+h}|y_t)$ approaches zero (unconditional expected value) exponentially fast as $h \to \infty$.

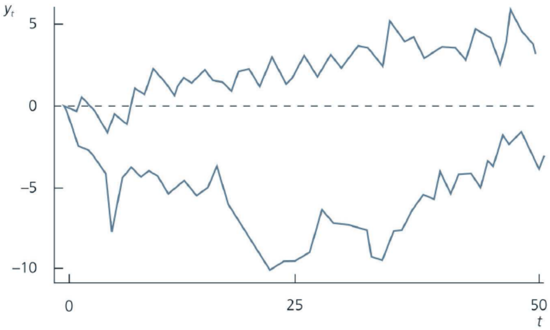
- A random walk is a special case of what's known as a *unit root process*. A unit root process is defined as

$$y_t = y_{t-1} + e_t.$$

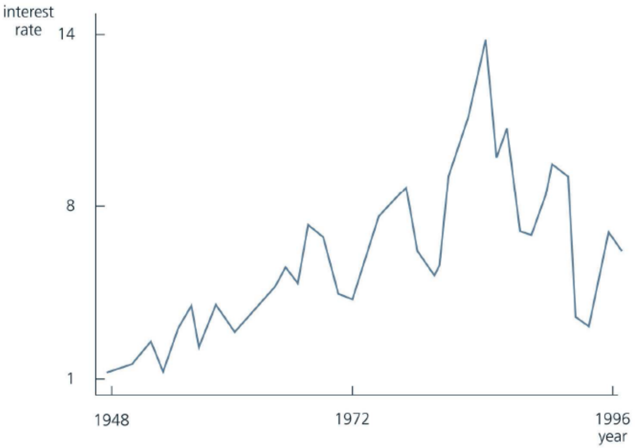
  where e_t is a weakly dependent process (like an AR(1) or MA(1) etc...).
- **Example:** GDP is (most likely) a unit root process.
- A *random walk with drift (intercept)* is an example of a highly persistent series that is trending

$$y_t = \alpha_0 + y_{t-1} + e_t$$

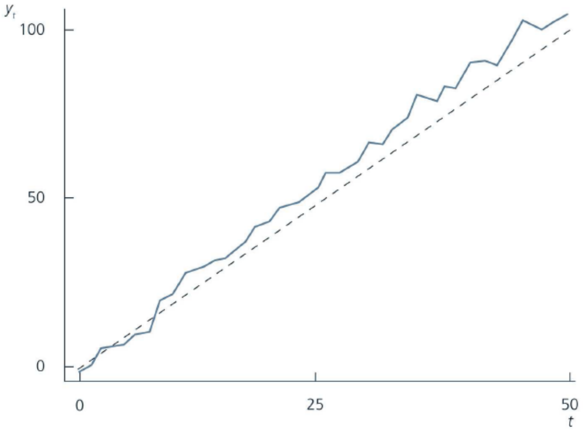
Two realizations of the random walk $y_t = y_{t-1} + e_t$, with $y_0 = 0$, $e_t \sim \text{Normal}(0,1)$, and $n = 50$.



The U.S. three-month T-bill rate, for the years 1948–1996.



A realization of the random walk with drift, $y_t = 2 + y_{t-1} + e_t$, with $y_0 = 0$, $e_t \sim$ Normal(0, 9), and $n = 50$. The dashed line is the expected value of y_t, $E(y_t) = 2t$.



- In order to use a highly persistent series and get meaningful estimates and make correct inferences, we want to transform it into a weakly dependent process.
- To have consistency and make inference, since TS.1′ fails! TS.1′ through TS.6′ are hard to get with highly persistent series!
- We refer to a weakly dependent process as being *integrated of order zero*, [$I(0)$]
- A random walk (or in general a unit-root process) is *integrated of order one*, [$I(1)$], meaning a first difference will be $I(0)$.($\Delta y_t = y_t - y_{t-1}$) are $I(0)$.
- In practice we do not know whether a series is $I(0)$ or $I(1)$.

- A simple tool for determining if the process is I(1) is to consider an AR(1) model

$$y_t = \rho_1 y_{t-1} + u_t.$$

- If the process is *I(0)*, $|\rho_1| < 1$, but it is *I(1)* if $\rho_1 = 1$.
- $\rho_1 = Corr(y_t, y_{t-1})$, thus we can estimate it from the sample correlation between y_t and y_{t-1} : $\hat{\rho}_1 = \widehat{Corr}(y_t, y_{t-1})$; it is called *first order autocorrelation* of $\{y_t\}$.
- If the sample first order autocorrelation is close to one, this suggests that the time series may be highly persistent (= contains a unit root)
- When estimating ρ_1, we should consider a trend in the series; detrend the series first, or include a trend in the regression.
- Both unit root and trend may be eliminated by differencing.
- It is possible to test the hypothesis $H_0 : \rho_1 = 1$, although we are not going to cover these tests in this module.

Example: Fertility equation:

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

- For large sample analysis, the fertility series and the series of the personal tax exemption have to be stationary and weakly dependent. This is questionable because the two series are highly persistent::

$$\hat{\rho}_{gfr} = .977, \ \hat{\rho}_{pe} = .964$$

- It is therefore better to estimate the equation in first differences. This makes sense because if the equation holds in levels, it also has to hold in first differences:

$$\Delta\widehat{gfr} = \underset{(.468)}{-\ .964} - \underset{(.027)}{.036}\ \Delta pe - \underset{(.028)}{.014}\ \Delta pe_{-1} + \underset{(.027)}{.110}\ \Delta pe_{-2}$$

$$n = 69, R^2 = .233, \bar{R}^2 = .197$$

δ_2

- Consider the general model

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_k x_{kt} + u_t,$$

where the explanatory variables $\mathbf{x}_t = (x_{1t}, ..., x_{kt})$ may or may not contain lags of y_t or x_{jt}.

- A model is said to be *dynamically complete model*. if enough lagged variables have been included as explanatory variables so that further lags do not help to explain the dependent variable:

$$E(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, ...) = E(y_t|\mathbf{x}_t).$$

- This implies that $E(u_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, ...) = 0$.
- A dynamically complete model *must* satisfy assumption about uncorrelated regression errors (TS 5′)

$$E(u_t u_s|\mathbf{x}_t, \mathbf{x}_s) = 0.$$

- One can easily test for dynamic completeness: If lags cannot be excluded, this suggests there is serial correlation

Sequential exogeneity

- A set of explanatory variables is said to be sequentially exogenous if "enough" lagged explanatory variables have been included:

$$E(u_t|\mathbf{x}_t, \mathbf{x}_{t-1}, ...) = E(u_t) = 0.$$

- Sequential exogeneity is *weaker* than strict exogeneity
- Sequential exogeneity is equivalent to dynamic completeness if the explanatory variables contain a lagged dependent variable
- Should all regression models be dynamically complete?
- Not necessarily: If sequential exogeneity holds, $E(y_t|\mathbf{x}_t, \mathbf{x}_{t-1}, ...)$ will be correctly estimated; absence of serial correlation is not crucial.