Time Series Data

Typical features: serial correlation / nonindependence of observations

Static models: $y_t = \beta_0 + \beta_1 x_t + u_t$.

Finite distributed lag models: $y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$

$$\frac{\Delta y_t}{\Delta x_{t-s}} = \beta_s$$ temporary shock / transitory shock.

: one time shock $\Rightarrow$ variable change β_s.

$$\frac{\Delta y_t}{\Delta x_{t-q}} + \cdots + \frac{\Delta y_t}{\Delta x_t} = \beta_1 + \cdots + \beta_q$$: permanent shock.

- **Finite sample properties of OLS under classical assumptions**
- *Assumption TS.1 (Linear in parameters)*

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

The time series involved obey a linear relationship. The stochastic processes y_t, x_{t1},…, x_{tk} are observed, the error process u_t is unobserved. The definition of the explanatory variables is general, e.g. they may be lags or functions of other explanatory variables.

- *Assumption TS.2 (No perfect collinearity)*
  "In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others."

- Notation

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ \boxed{x_{t1} \quad x_{t2} \quad \cdots \quad x_{tk}} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

This matrix collects all the information on the complete time paths of all the explanatory variables

The values of all the explanatory variables in period number t

- *Assumption TS.3 (Zero conditional mean)*

$$E(u_t|\mathrm{X}) = 0, t = 1, 2, \ldots, n$$

The mean value of the unobserved factors is uncorrelated to the values of the explanatory variables in all periods

Exogeneity: $E(u_t|\mathbf{x}_t) = 0$ ← The mean of the error term is uncorrelated to the explanatory variables of the *same period*

Strict exogeneity: $E(u_t|\mathbf{X}) = 0$ ← The mean of the error term is uncorrelated to the values of the explanatory variables of *all periods*

if ~~u_t is related~~ to x_{t-s}, we need to add x_{t-s} into regressors.

- **Theorem 10.1 (Unbiasedness of OLS)**

$$TS.1 - TS.3 \quad \Rightarrow \quad E(\widehat{\beta}_j) = \beta_j, \quad j = 0, 1, \ldots, k$$

- ***Assumption TS.4 (Homoskedasticity)***

$$Var(u_t|\mathrm{X}) = Var(u_t) = \sigma^2, t = 1, 2, \ldots, n$$ ← The volatility of the errors must not be related to the explanatory variables in any of the periods

- ***Assumption TS.5 (No serial correlation)***

$$Corr(u_t, u_s|\mathbf{X}) = 0, \ t \neq s$$ ← Conditional on the explanatory variables, the unobserved factors must not be correlated over time

- Discussion of assumption TS.5
  - Why was such an assumption not made in the cross-sectional case?
  - The assumption may easily be violated if, conditional on knowing the values of the indep. variables, omitted factors are correlated over time.
  - The assumption may also serve as substitute for the random sampling assumption if sampling a cross-section is not done completely randomly.
  - In this case, given the values of the explanatory variables, errors have to be uncorrelated across cross-sectional units (e.g. states).

- **Theorem 10.2 (OLS sampling variances)**

Under assumptions TS.1 – TS.5:

$$Var(\widehat{\beta}_j|\mathbf{X}) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad j = 1, \ldots, k$$

The conditioning on the values of the explanatory variables is not easy to understand. It effectively means that, in a finite sample, one ignores the sampling variability coming from the randomness of the regressors. This kind of sampling variability will normally not be large (because of the sums).

- **Theorem 10.3 (Unbiased estimation of the error variance)**

$$TS.1 - TS.5 \quad \Rightarrow \quad E(\widehat{\sigma}^2) = \sigma^2$$

- **Theorem 10.4 (Gauss-Markov Theorem)**
  - Under assumptions TS.1 – TS.5, the OLS estimators have the minimal variance of all linear unbiased estimators of the regression coefficients.
  - This holds conditional as well as unconditional on the regressors.

- *Assumption TS.6 (Normality)*

$$u_t \sim \text{Normal}(0, \sigma^2) \quad \text{independently of } \mathbf{X}$$ ← This assumption implies TS.3 – TS.5

- **Theorem 10.5 (Normal sampling distributions)**
  - Under assumptions TS.1 – TS.6, the OLS estimators have the usual normal distribution (conditional on X). The usual F and t-tests are valid.

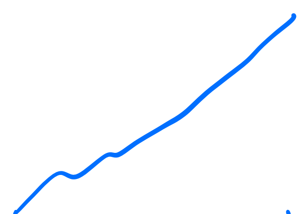
Time series with trends.

- **Modelling a linear time trend**

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta y_t) = E(y_t - y_{t-1}) = \alpha_1$$

$\Delta y_t / \Delta t = \alpha_1$ ← Abstracting from random deviations, the dependent variable increases by a constant amount per time unit

$E(y_t) = \alpha_0 + \alpha_1 t$ ← Alternatively, the expected value of the dependent variable is a linear function of time



- **Modelling an exponential time trend**

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta \log(y_t)) = \alpha_1$$

$(\Delta y_t / y_t) / \Delta t = \alpha_1$ ← Abstracting from random deviations, the dependent variable increases by a constant percentage per time unit