Basic Regression Analysis with Time Series Data

Chapter 10

Wooldridge: Introductory Econometrics:
A Modern Approach, 5e

with some adjustments by me

Analyzing Time Series: Basic Regression Analysis

- **Effect of a Transitory Shock in a FDL**
  - A *one time shock* to the explanatory variable in a past period, will change the dependent variable *temporarily* by the amount indicated by the coefficient of the corresponding lag.
  - The *immediate* change in *y* due to a one-unit increase in *z* at time *t*, is usually called impact propensity or impact multiplier.
- **Effect of Permanent Shock in a FDL**
  - For a *permanent shock* to the explanatory variable in a past period, the effect on the dependent variable will be the cumulated effect of all relevant lags.
  - The sum of the coefficients on current and lagged *z* is the *long-run change* in *y* given a permanent one-unit increase in z, and is called the long-run propensity (LRP) or long-run multiplier.

- Summarizing:
  - Lagged explanatory variables (FDL),
  - Lagged dependent variables (ARDL),
  - Lagged error-terms

  or combinations of those

Multiple Regression Model: Standard assumptions

Standard assumptions for the multiple regression model

- **Assumption MLR.1 (Linear in parameters)**
- **Assumption MLR.2 (Random sampling)**
- **Assumption MLR.3 (No perfect collinearity)**
- **Assumption MLR.4 (Zero conditional mean)**
- **Assumption MLR.5 (Homoscedasticity)**
- **Assumption MLR.6 (Normality)**

Theorem 3.4 (Gauß-Markov Theorem)

- Under assumptions MLR.1 - MLR.5, the OLS estimators are the best linear unbiased estimators (BLUEs) of the regression coefficients

Analyzing Time Series: Basic Regression Analysis

10.3 **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, $x_{t1}, \ldots, x_{tk}$ 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,

while $\{u_t ; t=1, 2, \ldots, n\}$ is the sequence of errors or disturbance terms, where n is the number of observations (time periods)

Analyzing Time Series: Basic Regression Analysis

- **<u>Finite sample properties of OLS under classical assumptions</u>**
- **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.“

➔ *TS.1 and TS.2 are essentially the same as MLR.1 and MLR.3*

Analyzing Time Series: Basic Regression Analysis

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

$$E(u_t|\mathbf{X}) = 0$$

- **Notation**

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

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

The values of all explanatory variables in period number *t* :
$\mathbf{x}_t = (\ x_{t1},\ x_{t2},\ \ldots,\ x_{tn}\ )$

Analyzing Time Series: Basic Regression Analysis

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

$$E(u_t|\mathbf{X}) = 0$$

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

  - Conditional on **X**, hence, the model should be specified correctly.

Analyzing Time Series: Basic Regression Analysis

- **Discussion of assumption TS.3**

Exogeneity: $E(u_t|\mathbf{x}_t) = 0$

The mean of the error term is unrelated to the explanatory variables of the same period (contemporaneous exogeneity)
→ This would mimic the assumption in cross-section data

Strict exogeneity: $E(u_t|\mathbf{X}) = 0$

The mean of the error term is unrelated to the values of the explanatory variables of all periods

Analyzing Time Series: Basic Regression Analysis

- **Discussion of assumption TS.3** *(cont.)*

➔ If u_t is independent of **X** and E(u_t) = 0, the assumption TS.3 automatically holds.

➔ Note that in a time series context, random sampling (MLR.2) is almost never appropriate

- individuals vs. time periods

➔ No restrictions are put on correlations across time neither in the independent variables nor in u_t

- An average, u_t must be unrelated to explan. var. in all time periods

Analyzing Time Series: Basic Regression Analysis

- **Strict exogeneity is stronger than contemporaneous exogeneity**

  - In the cross-sectional case, we did not explicitly state how the error term for person i, u_i, is related to the explanatory variables for other people.

    → This was unnecessary because with random sampling (Assumption MLR.2), u_i is automatically independent of the explanatory variables for observations other than i

Analyzing Time Series: Basic Regression Analysis

- **Strict exogeneity is stronger than contemporaneous exogeneity**

  - TS.3 rules out <u>feedback</u> from the dependent variable on future values of the explanatory variables

    → Changes in the error term today cannot cause future changes in *z*

    → This is often questionable, especially if explanatory variables „adjust" to past changes in the dependent variable

Analyzing Time Series: Basic Regression Analysis

- **Strict exogeneity is stronger than contemporaneous exogeneity**
  - Example: In a simple static model to explain a city's murder rate in terms of police officers per capita:

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t.$$

  - If $Cov(polpc_{t-j}, mrdrte_t) \neq 0$, then a FDL model must be estimated
  - If $Cov(polpc_{t+1}, u_t) \neq 0$ then TS.3 fails even if $Cov(u_t, polpc_{t-j}) = 0$ for $j = 0, 1, 2, \ldots$
  - For example, if police force ($polpc_{t+1}$) is adjusted based on previous murder rates ($mrdrte_{t-j}$), it is highly likely that TS.3 fails.

Analyzing Time Series: Basic Regression Analysis

- **Strict exogeneity is stronger than contemporaneous exogeneity**
  - Example: In an agricultural production function, the rainfall is strictly exogenous, while the labor input is not. It is chosen by the farmer, and the farmer may adjust the amount of labor based on last year's yield.

➔ If the error term is related to past values of the explanatory variables, one should include these values as contemporaneous regressors, *i.e.*, use the FDL model instead of a static model.

Analyzing Time Series: Basic Regression Analysis

→ Note that also *omitted variables* and *measurement errors* in some of the regressors are also obvious candidates for non-compliance of TS.3, as they are in cross-section data, and cause biased estimates.

Analyzing Time Series: Basic Regression Analysis

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

  - Under assumptions TS.1-TS.3: $E\left[\hat{\beta}_j \middle| \mathbf{X}\right] = \beta_j, j = 0, 1, \cdots, k,$

    for any values of β_j ,

  - and therefore also,

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

Analyzing Time Series: Basic Regression Analysis

- **Theorem 10.1 (Unbiasedness of OLS)**
  - The proof is similar to that of Theorem 3.1. As in Chapter 3, we condition on the regressors **X**. Conditional unbiasedness implies unconditional unbiasedness:

$$E\left[\hat{\beta}_j\right] = E\left[E\left[\hat{\beta}_j \middle| \mathbf{X}\right]\right] = E\left[\beta_j\right] = \beta_j$$

  - *Law of Iterated Expectations* : For two r.v.'s, $E[E[Y|X]] = E[Y]$.
  - Intuition: $E[Y|X] = \mu(X)$ averages over Y for each group of individuals with a specific X value, and then $E[E[Y|X]] = E[\mu(X)]$ averages over X by the probability of each X group, which results in the unconditional mean of Y .

Analyzing Time Series: Basic Regression Analysis

➔ We‵ve now established conditions for unbiasedness of the OLS estimator in time series models, but not yet for the Gauss-Markov Theorem (BLUE).

➔ Two more assumptions are needed for that

Analyzing Time Series: Basic Regression Analysis

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

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

The volatility of the errors must not be related to the explanatory variables in any of the periods

Conditional on **X**, the variance of u_t is the same for all t.

  - A sufficient condition is that the volatility of the error is independent of the explanatory variables and that it is constant over time

➔ *Pretty similar to MLR.5 for cross-section data, but perhaps trickier …*

Analyzing Time Series: Basic Regression Analysis

- **Discussion of assumption TS.4**
  - Unobservables that affect the dependent variable must have constant variance over time
  - In the time series context, homoscedasticity may also be easily violated, *e.g*. if the volatility of the dependent variable depends on regime changes
    - For example, it could be that the variability in interest rates (dep.var.) depends on the level of inflation or relative size of the deficit (indeo.vars.).

➔ *Chapter 12 will deal with heteroskedasticity issues*

Analyzing Time Series: Basic Regression Analysis

- **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

- Key idea: errors from different time periods are uncorrelated
- *Why was such an assumption not made in the cross-sectional case?*

Analyzing Time Series: Basic Regression Analysis

- **Discussion of assumption TS.5**
  - Note that it does not assume anything about the temporal correlation in the independent variables
    - For example, when explaining interest rates (dep.var.), the level of inflation (indep.var.) is almost certainly correlated across time, but that has nothing to do with whether TS.5 holds.
  - The assumption may easily be violated if, conditional on knowing the values of the indep. variables, omitted factors are correlated over time

Analyzing Time Series: Basic Regression Analysis

- **Discussion of assumption TS.5**
  - In cross-section data ...
  - ... this 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
    - For example, when analyzing cities or states

➔ *Also in panel data, serial correlation is an issue: Chapter 13 and 14*

Analyzing Time Series: Basic Regression Analysis

- **<u>Theorem 10.2 (OLS sampling variances)</u>**

Under assumptions TS.1 – TS.5:

The same formula as in the cross-sectional case

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

Total sample variation in explanatory variable x_j:

$$\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

R-squared from a regression of explanatory variable x_j on all other independent variables (including a constant)

Multiple Regression Analysis: Estimation

- ***From Chapter 3, still valid:* Components of OLS Variances:**
- **1) The error variance (σ^2)**
  - A high error variance increases the sampling variance because there is more „noise" in the equation
  - A large error variance necessarily makes estimates imprecise
  - The error variance does not decrease with sample size
- **2) The total sample variation in the explanatory variable (*SST*)**
  - More sample variation leads to more precise estimates
  - Total sample variation automatically increases with the sample size, increasing the sample size is thus a way to get more precise estimates

Analyzing Time Series: Basic Regression Analysis

- **<u>Theorem 10.3 (Unbiased estimation of the error variance)</u>**

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

where σ^2 =SSR/df is an unbiased estimator of σ^2, where df=n-k-1 :

$$\widehat{\sigma}^2 = \left(\sum_{i=1}^{n} \widehat{u}_i^2 \right) / \, [n - k - 1]$$

Analyzing Time Series: Basic Regression Analysis

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

➔ *Note: we're still talking about finite sample properties*

Analyzing Time Series: Basic Regression Analysis

- **Assumption TS.6 (Normality)**

This assumption implies TS.3 – TS.5

$$u_t \sim N(0, \sigma^2) \quad \text{independently of} \quad \mathbf{X}$$

- **<u>Theorem 10.5 (Normal sampling distributions)</u>**
  - Under assumptions TS.1 – TS.6, the OLS estimators have the usual normal distribution (conditional on $\mathbf{X}$). The usual F- and t-tests are valid.

Analyzing Time Series: Basic Regression Analysis

- **<u>Implications ...</u>**
  - When assumptions TS.1 – TS.6 hold, everything we have learned about estimation and inference for cross-sectional regressions applies directly to time series regressions.
  - However, the classical linear model (CLM) assumptions for time series data are more restrictive than those for cross-sectional data, in particular:
    - the strict exogeneity assumption, and
    - the no serial correlation assumption
  - Moreover, finite sample properties. As with cross-section data, large sample properties are less restrictive (*even in time series ...*)

    ➔ *Chapter 11*