Econometrics

Week 2

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Fall 2022

Recommended Reading

For today

- Basic regression analysis with time series data
- Chapter 10 (pp. 311 – 342)

For next week

- Further issues in using OLS with time series data
- Chapter 11 (pp. 347 – 370)

Today's Lecture

- Differences between time series data and cross-sectional data
- Gauss-Markov assumptions for time series regression:
  How do they differ from the cross-sectional case?
- Properties of OLS estimator in time-series context
- Examples of econometric models using time series data
- Trends and seasonality

Cross Sectional Data vs. Time Series Data

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i \qquad u_i \sim N(0, \sigma^2)$$

$$i = 1, 2, \ldots, n$$

vs.

$$y_t = \beta_0 + \beta_1 x_{t1} + u_t \qquad u_t \sim N(0, \sigma^2)$$

$$t = 1, 2, \ldots, T$$

The Nature of Time Series Data

- Time series vs. cross-sections: Temporal ordering.
- Until now, we have studied properties of OLS estimator based on the assumption that samples are random...
- ...but time series data are not random samples (why?)

Instead, we deal with Stochastic Processes

- "stochastic" from the Greek "stochos": aim, guess, or characterized by conjuncture and randomness.
- The observed data is one realization of a stochastic process.

- How does it challenge the CLM model?

Classical Linear Model Assumptions

- Assumption MLR.1 (Linear in Parameters)
- Assumption MLR.2 (Random Sampling)
- Assumption MLR.3 (No Perfect Colinearity)
- Assumption MLR.4 (Zero Conditional Mean)
- Assumption MLR.5 (Homoskedasticity)
- Assumption MLR.6 (Normality)

Which of the assumptions is/are challenged when using time-series data?

Unbiadesness of the OLS estimator

Consider a simple model: $y_t = \beta_0 + \beta_1 \cdot x_t + u_t$

OLS estimator:

$$\hat{\beta}_1^{OLS} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} = \beta_1 + \frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}$$

Expected value of the OLS estimator (which assumptions are used and where?):

$$E[\hat{\beta}_1^{OLS}|\mathbf{X}] = \beta_1 + \frac{E[\sum(x_t - \bar{x})u_t|\mathbf{X}]}{\sum(x_t - \bar{x})^2} = \beta_1 + \frac{\sum(x_t - \bar{x})E[u_t|\mathbf{X}]}{\sum(x_t - \bar{x})^2} = \beta_1$$

Note: All sums go from t=1 to t=T, i.e. sum over all observations

Unbiasedness: Assumptions

TS1: Linear in parameters

The stochastic process $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t) : t = 1, 2, \ldots, n\}$ follows the linear model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \tag{1}$$

with $\{u_t : t = 1, 2, = \ldots, T\}$ sequence of error disturbances and T number if observations (time periods).

TS1: Linear in parameters (matrix notation)

The model representing the stochastic process can be written as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \tag{2}$$

with $\mathbf{y}$ being a $T \times 1$ vector, $\mathbf{X}$ being a $T \times k + 1$ matrix, and $\mathbf{u}$ being a $T \times 1$ vector of error disturbances.

Unbiasedness: Assumptions

TS2: Zero Conditional Mean

$$E(u_t|X) = 0, \qquad t = 1, 2, \ldots, T. \tag{3}$$

$\Rightarrow$ error term in any given period is uncorrelated with explanatory variable in all time periods (also future!).

- TS2 implies that explanatory variable is **strictly exogenous**.
- This is needed b/c we don't have random samples
- A weaker assumption is: $E(u_t|x_{t1}, \ldots, x_{tk}) = E(u_t|x_t) = 0$
  - Means that x_{tj} are **contemporaneously exogenous**.
  - Less strict; does not require u_t to be uncorrelated with x_{sj} for $s \neq t$ as in TS2.

Note

In social sciences, many stochastic processes violate the strict exogeneity assumption! Next week we'll learn how to deal with this

Unbiasedness: Assumptions

TS3: No Perfect Collinearity

No independent variable is constant or a perfect linear combination of the others.

TS3: No Perfect Collinearity (matrix notation)

The matrix $\mathbf{X}$ has full rank, i.e. it's rank is $k+1$

These are essentially the same as for cross-sectional data.

Unbiasedness

Theorem 1: Unbiasedness of OLS

Under TS1, TS2, and TS3, the OLS estimators are unbiased conditional on X, and therefore unconditionally as well: $E(\hat{\beta}_j) = \beta_j$, $j = 0, 1, \ldots, k$.

Theorem 1: Unbiasedness of OLS (matrix notation)

Under TS1,TS2, and TS3, the OLS estimator $\widehat{\beta}$ is unbiased for β.

- The theorem is similar to the one we used in cross-sectional data, but we have omitted random sampling assumption (thanks to assumption TS2).
- Violation of the (strict) exogeneity assumption brings bias as in the cross-sectional data.

Variance of the OLS estimator

Consider a simple model: $y_t = \beta_0 + \beta_1 \cdot x_t + u_t$

OLS estimator:

$$\hat{\beta}_1^{OLS} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} = \beta_1 + \frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}$$

Variance of the OLS estimator (which assumptions are used and where?):

$$Var[\hat{\beta}_1^{OLS}|\mathbf{X}] = Var[\frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}|\mathbf{X}] = \frac{Var[\sum(x_t - \bar{x})u_t|\mathbf{X}]}{(\sum(x_t - \bar{x})^2)^2} =$$

$$= \frac{\sum(x_t - \bar{x})^2 Var[u_t|\mathbf{X}]}{(\sum(x_t - \bar{x})^2)^2} = \frac{\sigma^2}{\sum(x_t - \bar{x})^2} = \frac{\sigma^2}{SST_x}$$

Note: All sums go from t=1 to t=T, i.e. sum over all observations

Variance of the OLS Estimator: Assumptions

TS4: Homoskedasticity

Conditional on X, the variance of u_t is the same for all t:
$Var(u_t|X) = Var(u_t) = \sigma^2$, $t = 1, 2, \ldots, n$.

- It means that the error variance should be independent of all x's and is constant over time.
- When not satisfied, we say that the data are **heteroskedastic**.

Variance of the OLS Estimator: Assumptions

TS5: No Serial Correlation

Conditional on X, the errors in any two different time periods are uncorrelated:

$$Corr(u_t, u_s | X) = 0 \qquad \text{for all } t \neq s.$$

- It means that the cross-correlations in error (disturbance) structure are independent of all x's, constant over time and equal to zero.
- When not satisfied, we say that the errors (disturbances) are correlated across time and suffer from **serial correlation** or **autocorrelation**.

Variance of the OLS Estimator: Assumptions

TS4: Homoskedasticity and No Serial Correlation (matrix)

Conditional on X, one can write these two assumptions as:

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_T$$

$$Var(\mathbf{u}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2T} \\ \ldots & \ldots & & \ldots \\ \sigma_{T1} & \sigma_{T2} & \ldots & \sigma_{TT} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \ldots & \ldots & & \ldots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix}$$

Variance of the OLS Estimator

Theorem 2: OLS Sampling Variances

Under the time series Gauss-Markov Assumptions (TS1-TS5), the variance of $\hat{\beta}_j$, conditional on X is:

$$Var(\hat{\beta}_j|X) = \sigma^2/[SST_j(1 - R_j^2)], \qquad j = 1, \ldots, k,$$

where SST_j is the total sum of squares of x_{tj} and R_j^2 is the R-squared of a regression of x_{tj} on all other explanatory variables.

Theorem 3: Unbiased Estimation of σ^2

Under TS1-TS5, the estimator $\hat{\sigma}^2 = SSR/df$ is an unbiased estimator of σ^2, where $df = T - k - 1$.

Theorem 4: Gauss-Markov Theorem

Under TS1-TS5, the OLS estimators are the best linear unbiased estimator (BLUE) conditional on X.

Inference Under the Classical Linear Model (CLM) Assumptions

- Under assumptions TS1-TS5, OLS in time-series context has the same desirable finite sample properties as in the cross-sectional data case
- If we add assumption on normality of errors, inference is also the same

TS6: Normality

The errors u_t are independent of X and are independently and identically distributed as $N(0, \sigma^2)$.

TS6 implies TS3-TS5, but is stronger $\rightarrow$ independence and normality.

Inference Under the Classical Linear Model (CLM) Assumptions

Theorem 5: Normal Sampling Distribution

Under TS1-TS6, the CLM assumptions for time series, the OLS estimators are normally distributed, conditional on X. Further, under the null hypothesis, each t statistic has a t distribution, and each F statistic has an F distribution.

Inference Under the Classical Linear Model (CLM) Assumptions

- $\Rightarrow$ everything we have learned about estimation and inference for cross-sectional regressions applies directly to time series regressions.

BUT this also applies to the problems

Inference is only as good as the underlying assumptions !

!!! REMEMBER !!!

- CLM assumptions for time series data are much more restrictive than those for the cross-sectional data.
- In particular strict exogeneity and no serial correlation.
- $\rightarrow$ unrealistic in social sciences for many data sets.
- We will learn how to overcome this during next lectures

Examples of Time Series econometric models: a static model

A contemporaneous (in Czech: “souběžný”) relation between y and z can be captured by a **static model**:

$$y_t = \beta_0 + \beta_1 z_t + u_t, \qquad t = 1, 2, \ldots, n.$$

When to use?

- Change in z at time t has immediate effect on y: $\Delta y_t = \beta_1 \Delta z_t$, when $\Delta u_t = 0$.
- We would like to know the tradeoff between y and z.

Example: Demand Curve

- Contemporaneous tradeoff between price and consumption:
$$D_t = \beta_0 + \beta_1 Price_t + u_t$$
- This specification assumes that both parameters are time-invariant.

Examples of Time Series econometric models: Finite Distributed Lag (FDL) models

- This is a dynamic model
- We allow one or more variables to affect y with a **lag**.
- FDL of order two:
  $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t,$
- **FDL model of order** q will include q lags of z.
- Can be used to test the effect of a 1-period temporary increase in z
  - δ_0 – **"impact propensity"** – reflects immediate change in y when z changes by one unit.
- Can be used to test the effect of a permanent increase in z
  - $\delta_0 + \delta_1 + \ldots + \delta_q$ – **"long-run propensity"** (LRP)– reflects the long-run change in y after a permanent change in z.

Examples of Time Series econometric models

Note

Static model is a special case of a FDL model

when setting $\delta_0, \delta_1, \ldots, \delta_q$ to zero:

$$y_t = \alpha_0 + \delta_0 \; z_t + \underset{\substack{\downarrow \\ 0}}{\delta_1} \; z_{t-1} + \underset{\substack{\downarrow \\ 0}}{\ldots} + \underset{\substack{\downarrow \\ 0}}{\delta_q} \; z_{t-q} + u_t$$

Note

Often, we have substantial correlation across z at different lags $\Rightarrow$ multicollinearity

Example: Effect of EET (electronic sales record) introduction on prices in restaurants

Since December 2016 restaurants in the Czech Republic are obliged to report each sale through an electronic system

- Introduced to reduce tax evasion
- Lot of attention in the **media**
- Opponents worried EET would cause price increase
- Use monthly **data** on restaurant price index to verify this hypothesis

$$priceindex_t = \alpha_0 + \delta_0 EET_t + \delta_1 EET_{t-1} + \delta_2 EET_{t-2} +$$
$$+ \delta_3 EET_{t-3} + \delta_4 EET_{t-4} + u_t$$

Trends in Time Series

- Economic time series often trend
(i.e. change regularly over time).
- What happens when we regress two trending time series on each other?

▸ Example

- We might come across the so called **spurious correlation**
  - We find relationship between two or more trending variables simply because each is growing over time.
  - Think about it as of omitted variable problem:
Trending factors that affect y_t are correlated with an explanatory variable
- It is important to properly account for time trend

Dealing with trending variables

- In presence of spurious correlation, the CLM assumptions are violated
- Which assumption(s) is/are violated?
- How to eliminate this problem?
    - Add a time trend to the regression model, or
    - “Detrend” the trending variables

Adding a time trend to the model

- Simple idea: we add the missing variable to the model
- Example on the board
- Pros:
  - This eliminates the spurious correlation problem
  - It is easy
- Cons:
  - We get unnaturally high R-squared
  - It is impossible to say what portion of total variation in the dependent variable is explained by the explanatory variables (usually we don't think of time trend as of an explanatory variable)

Detrending

- This involves regressing each variable in the model on t...
- ...and re-running the original model with detrended variables
- Example on the board
- Pros:
  - This eliminates the spurious correlation problem
  - We explicitly see which variables are trending and to what extend
  - R-squared in the final regression is informative of what portion of total variation in the dependent variable is explained by the explanatory variables
- Cons:
  - The procedure is complicated
  - We have to be careful and use the same form of time trend for each variable

Trends in Time Series

We can control for trend in several simple ways.

A trending time series $\{y_t\}$ can be written as:

1. Linear trend: $y_t = \alpha_0 + \alpha_1 t + \epsilon_t$
2. Exponential trend: $log(y_t) = \alpha_0 + \alpha_1 t + \epsilon_t$
3. Quadratic trend: $y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t$
4. ...

with i.i.d. $\{\epsilon_t\}$ and $E(\epsilon_t) = 0$ and $var(\epsilon_t) = \sigma_\epsilon^2$

Example: Linear Trend

We can think of a sequence with linear trend as:
$E(y_t) = \alpha_0 + \alpha_1 t$,
while change of y_t, $E(\Delta y_t) = E(y_t - y_{t-1}) = \alpha_1$.

- For $\alpha_1 > 0$, we have upward trend.
- For $\alpha_1 < 0$, we have downward trend.

Seasonality

- Data observed at monthly or quarterly interval may exhibit *seasonality*, or periodicity.
- Example: Quarterly data on retail sales will tend to jump up in the 4th quarter.
- Seasonality can be dealt with by adding a set of seasonal dummies.
- As with trends, the series can be seasonally adjusted before running the regression.

Thank you for your attention!

Reading for next week

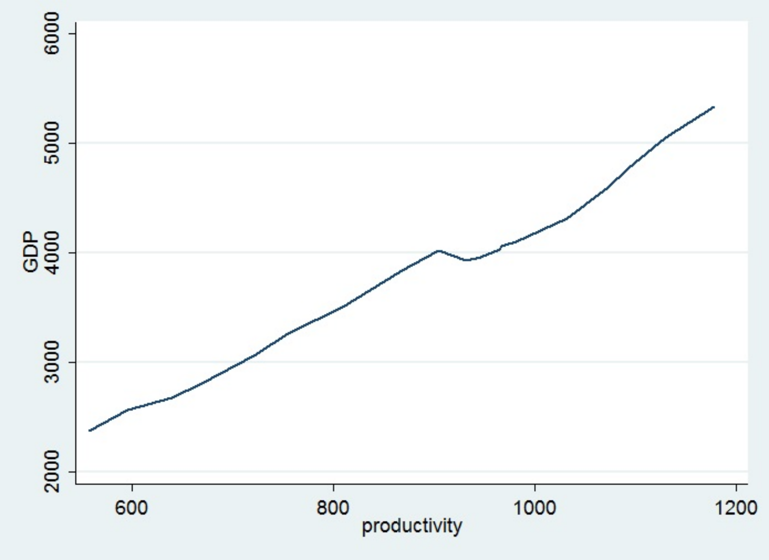
Do not forget to read Chapter 11 for the next week!

Revision for next week

Please, revise also:

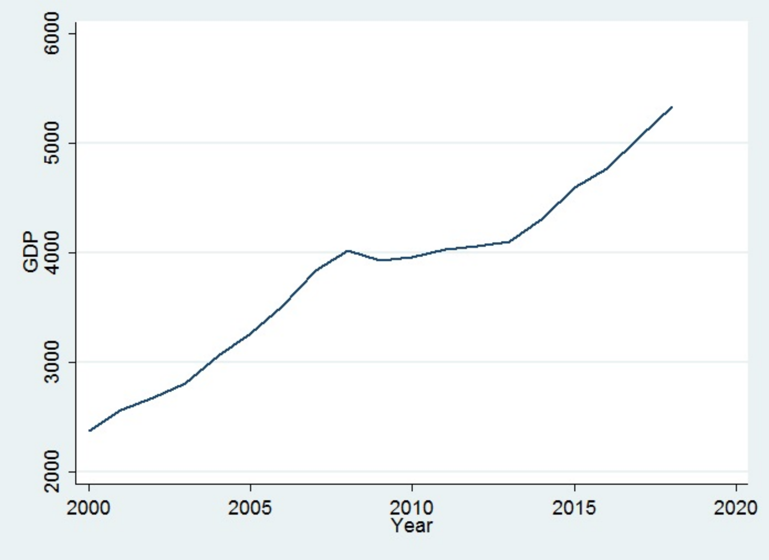
- (Statistics) expected value, variance, covariance
- (Econometrics I) asymptotic (large sample) properties of OLS

Trends in Time Series



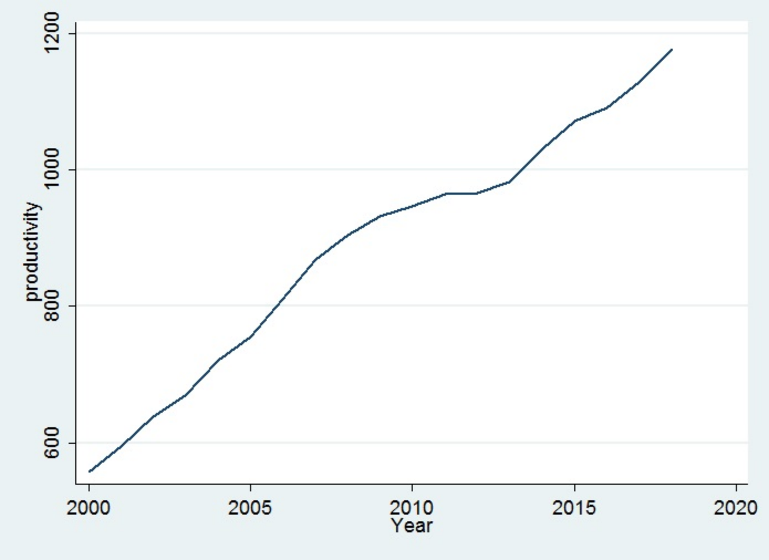
◂ Back

Trends in Time Series



Back

Trends in Time Series



◂ Back