Week 1: Repetition
Time Series Regression Models (TSRMs)

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

Outline

Introduction

- Stochastic (random) process:
  Sequence of random variables indexed by time

- Time series is one realization of a stochastic process

TSRMs examples

- Static Models

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

- Finite Distributed Lag (FDL) Models

$$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

Order of the FDL model,
Impact multiplier $\times$ long-run multiplier,
Temporary (one-off) $\times$ permanent increase in x,
Lag distribution.

G-M assumptions for TSRMs

Definitions:

- $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$
  Set of all explanatory variables (matrix $n \times k$)
- $\boldsymbol{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$
  Set of all n observations of the j-th regressor (vector $n \times 1$)
- $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tk})$
  Set of all explanatory variables observed at time t (vector $1 \times k$)

- Usually, the first item in $\boldsymbol{x}_t$ and the first column in $\boldsymbol{X}$ correspond to the intercept.

G-M assumptions for TSRMs

- TS.1 Linearity
  The stochastic process $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t); t = 1, 2, \ldots, n\}$ follows a linear model $y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$.

- TS.2 No perfect collinearity
  There is no perfect collinearity among regressors.
  Comment: It allows collinearity among regressors.

- TS.3 Strict exogeneity
  For each t, the expected value of error conditionally on the explanatory variables at all time periods is zero:
  $E(u_t|\boldsymbol{X}) = 0, \quad t = 1, 2, \ldots, n$
  Comment: If $E(u_t|x_{t1}, x_{t2}, \ldots, x_{tk}) = E(u_t|\boldsymbol{x}_t) = 0$, then x_{tj} variables are contemporaneously exogenous

G-M assumptions for TSRMs

- TS.4 Conditional Homoskedasticity

$$\text{var}(u_t|\boldsymbol{X}) = \text{var}(u_t) = \sigma^2, \qquad t = 1, 2, \ldots, n$$

- TS.5 Serial Correlation (Autocorrelation) is not present

$$\text{cov}(u_t, u_s|\boldsymbol{X}) = 0, \qquad t \neq s$$

- TS.6 Normality
  u_t are independent of $\boldsymbol{X}$ and *i.i.d.* [1] : $u_t \sim N(0, \sigma^2), i.i.d.$

- **CLRM**: Classical linear regression model
  TS.1 - TS.6 conditions hold

[1] Independently and identically distributed

Properties of OLS estimators

- Under TS.1 - TS.3, OLS estimators are unbiased.

- Under assumptions TS.1 - TS.5,
  variance of $\hat{\beta}_j$ conditional on $\boldsymbol{X}$ is given by:

$$\mathrm{var}(\hat{\beta}_j|\boldsymbol{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \quad j = 1, \ldots, k$$

  where SST_j is the total sum of squares of $\boldsymbol{x}_j$ and R_j^2 is the coefficient of determination from auxiliary regression of x_j on other explanatory variables.

Properties of OLS estimators

- Under assumptions TS.1 - TS.5, the estimator

$$\hat{\sigma}^2 = \frac{SSR}{df} \quad ; \quad df = n - k - 1$$

  is an unbiased estimator of σ^2 (variance of u_t).

- Under assumptions TS.1 - TS.5, OLS estimators are **BLUE** conditional on $\boldsymbol{X}$

- Under assumptions TS.1 - TS.6, $\hat{\beta}_j$ are normally distributed. Under H_0, each t statistics has a t distribution and F statistic has a F distribution (small-sample and asymptotically). The usual construction of confidence intervals is also valid.

Trends and spurious regression

- Regression: y on t
  If there is a linear trend in y
- Regression: $\log y$ on t
  Exponential trend, constant rate of growth of y
- Spurious regression:
  We can find relationship between two or more trending variables even if it does not exist in reality

Detrending algorithm (based on FWL theorem):

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t$$

- Regress each variable on constant, time and save residuals

$$\ddot{y}_t, \ddot{x}_{t1}, \ddot{x}_{t2}, \qquad t = 1, 2, \ldots, T$$

- Regress $\ddot{y}_t$ on $\ddot{x}_{t1}$, $\ddot{x}_{t2}$

$$\hat{\ddot{y}}_t = \hat{\beta}_1 \ddot{x}_{t1} + \hat{\beta}_2 \ddot{x}_{t2}$$

- Coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ from this regression are the same as in the original regression

Deseasonalizing algorithm (based on FWL theorem):
Example based on quarterly data

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 dummy1 + \hat{\beta}_4 dummy2 + \hat{\beta}_5 dummy3$$

- Regress variables on constant, seasonal dummies and save residuals:

$$\ddot{y}_t, \ddot{x}_{t1}, \ddot{x}_{t2}, \qquad t = 1, 2, \ldots, T$$

- Regress $\ddot{y}_t$ on $\ddot{x}_{t1}$, $\ddot{x}_{t2}$

$$\hat{\ddot{y}}_t = \hat{\beta}_1 \ddot{x}_{t1} + \hat{\beta}_2 \ddot{x}_{t2}$$

- Coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ from this regression are the same as in the original regression

Coefficient of determination when y is trending

- With trending y, coefficient of determination overshoots.

$$\overline{R}^2 = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2}$$

- $\hat{\sigma}_u^2$is unbiased estimator of error variance and is easily estimated in the case trend is among regressors.

- With trending y, $\hat{\sigma}_y^2 = \frac{SST}{(n-1)}$, where $SST = \sum_{t=1}^{n}(y_t - \overline{y})^2$ is neither unbiased nor consistent estimator.

Coefficient of determination when y is trending

A better coefficient of determination is:

- Regress $\ddot{y}_t$ on $\ddot{x}_{t1}$,$\ddot{x}_{t2}$

  Coefficient of determination from this regression, i.e.:

$$R^2 = 1 - \frac{SSR}{\sum_{t=1}^{n} \ddot{y}_t^2}$$

  is more reliable (does not overshoot) than that one from the original regression.

Stationary and weakly dependent time series

- Strict exogeneity, homoskedasticity, absence of serial correlation and error-normality assumptions are very limiting
- With large samples, weaker assumptions are sufficient
- For large samples, key assumptions are:
  Stationarity and **weak dependency**

Stationary and weakly dependent time series

- Time series is stationary if its probability distributions are stable over time.
- Covariance stationarity: first two moments and auto-covariance do not change over time.
- Weak dependency: correlation between x_t and x_{t+h} "quickly" converges to zero with h growing to infinity.

Stationary and weakly dependent time series

- For Central Limit Theorem (CLT) and Law of Large Numbers (LLN) to hold, dependency between observations must not be too strong and must sufficiently quickly decrease with growing time distance between them.
- Time series can be non-stationary and weakly dependent.

Examples of weakly dependent time series:

- Moving average process of order one: MA(1)

$$x_t = e_t + \alpha_1 e_{t-1}$$

e_t is an *iid* time series. Observations with higher time distance than 1 are uncorrelated.

- Autoregressive proces of order 1 AR(1)

$$y_t = \rho y_{t-1} + e_t \Rightarrow \mathrm{corr}(y_t, y_{t+h}) = \rho^h$$

If stability condition $|\rho| < 1$ holds, TS is weakly dependent because correlation converges to zero with growing h.

Asymptotic properties of OLS estimators

- TS.1' Linearity
  The stochastic process $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t); t = 1, 2, \ldots, n\}$ follows the linear model $y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$
  We assume both dependent and independent variables are stationary and weakly dependent.

- TS.2' No perfect collinearity
  There is no perfect collinearity among regressors.
  Comment: the same assumption as TS.2

- TS.3' Contemporaneous exogeneity
  Null conditional expected value of errors:

$$E(u_t | x_{t1}, \ldots, x_{tk}) = E(u_t | \boldsymbol{x}_t) = 0$$

Asymptotic properties of OLS estimators

- Under assumptions TS.1', TS.2' and TS.3', OLS estimators are consistent (not unbiased)

  TS.1' - TS.3' $\Rightarrow \operatorname{plim} \hat{\beta}_j = \beta_j, \quad j = 0, 1, \ldots, k$

Asymptotic properties of OLS estimators

By removing strict exogeneity (changing TS.3 to TS.3'):

- There is no restriction on how u_t is related to regressors in other time periods. Hence:
- We allow for feedback from (lagged) explained variable to "future" values of explanatory variables
- We can use lagged dependent variable as regressors.

Asymptotic properties of OLS estimators

- TS.4' Contemporaneous homoskedasticity

$$\mathrm{var}(u_t|\boldsymbol{x}_t) = \mathrm{var}(u_t) = \sigma^2$$

- TS.5' No serial correlation
  (autocorrelation in residuals is not present)

$$\mathrm{corr}(u_t, u_s|\boldsymbol{x}_t, \boldsymbol{x}_s = 0), t \neq s$$

Asymptotic properties of OLS estimators

- Theorem: Asymptotic normality of OLS estimators

  Under assumptions TS.1' – TS.5', OLS estimators are asymptotically normally distributed. Usual OLS standard errors, t-statistics and F-statistics are asymptotically valid.

  $$\hat{\boldsymbol{\beta}} \rightarrow N(\boldsymbol{\beta}, \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \text{ as } n \rightarrow \infty$$

Additional TS topics

- Using trend-stationary time series in regression analysis.
- Using highly persistent (strongly dependent) time series in regression analysis
  - Random walk, random walk with drift, unit root process
  - Order of integration, testing of unit root
  - Cointegration
- Dynamically Complete Models

Serial correlation

- Causes and effects
- Robust inference
- Testing AR(1) serial correlation with strictly exogenous regressors

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t \\
u_t &= \rho u_{t-1} + e_t \\
\hat{u}_t &= \rho \hat{u}_{t-1} + error \\
H_0 &: \rho = 0
\end{aligned}
$$

- Durbin-Watson test

Serial correlation

- Testing AR(1) serial correlation with general regressors

$$\hat{u}_t = \alpha_0 + \alpha_1 x_{t1} + \cdots + \alpha_k x_{tk} + \rho \hat{u}_{t-1} + error$$

$$H_0 : \rho = 0$$

(We can use a heteroskedasticity-robust version of the test)

- Breusch-Godfrey test for AR(q) serial correlation

$$\hat{u}_t = \alpha_0 + \alpha_1 x_{t1} + \cdots + \alpha_k x_{tk} + \rho_1 \hat{u}_{t-1} + \cdots + \rho_q \hat{u}_{t-q} + error$$

$$H_0 : \rho_1 = \cdots = \rho_q = 0$$

GLSM – matrix form

- In the generalized linear regression model (GLRM):
  - $E(\boldsymbol{u}) = \mathbf{0}$
  - $E(\boldsymbol{u}\boldsymbol{u}^T) = \sigma^2 \boldsymbol{H}$
    (i.e. not $\sigma^2 \boldsymbol{I}_n$) - covariance matrix of disturbances

- Principle of GLSM:
  Transformation of LRM using a transformation matrix $\boldsymbol{T}$ (such that $\boldsymbol{T}\boldsymbol{H}\boldsymbol{T}^T = \boldsymbol{I}_n$) to get:

$$\mathrm{var}(\boldsymbol{u}) = \sigma^2 \boldsymbol{I}_n$$

  in the transformed model.

GLSM – matrix form

CLRM: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$; estimator: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

GLRM: $\boldsymbol{T}\boldsymbol{y} = \boldsymbol{T}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{T}\boldsymbol{u}$

- For covariance matrix of the transformed model, we get:

$$E(\boldsymbol{T}\boldsymbol{u}(\boldsymbol{T}\boldsymbol{u})^T) = E(\boldsymbol{T}\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{T}^T) = \boldsymbol{T}E(\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{T}^T = \sigma^2\boldsymbol{T}\boldsymbol{H}\boldsymbol{T}^T$$

We need $\boldsymbol{T}\boldsymbol{H}\boldsymbol{T}^T = \boldsymbol{I}$; from this relation we derive the transformation matrix $\boldsymbol{T}$. Transformation gives us model that fulfills previously broken G-M assumption.

GLSM – matrix form

GLSM algorithm:

1. Estimate the model using OLS.
2. Test assumption $E(\boldsymbol{u}\boldsymbol{u}^T) = \sigma^2 \boldsymbol{I}_n$, if broken:
3. Find/set appropriate transformation matrix $\boldsymbol{T}$.
4. Multiply variables of the model by $\boldsymbol{T}$, to get transformed variables.
5. Estimate the model with transformed variables (use OLS for estimation).
6. If necessary, reverse-transform the model estimated to get parameter estimates for the original specification of the LRM.

GLSM – matrix form

- Matrices $\boldsymbol{H}$ and $\boldsymbol{T}$ of the GLSM algorithm differ for heteroskedasticity and autocorrelation.
  - Heteroskedasticity (CS data, matrix form, quick recap):

$$\sigma^2 \boldsymbol{H} = \sigma^2 \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & h_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

GLS: $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \hat{\boldsymbol{H}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \hat{\boldsymbol{H}}^{-1} \boldsymbol{y}$
where WLS or FGLS method is used to estimate/set $\hat{\boldsymbol{H}}$.

GLSM – matrix form

Autocorrelation (matrix form):

- Let's assume autocorrelation of the AR(1) form
  $u_t = \rho u_{t-1} + \varepsilon_t$;
  where ρ is coefficient of autocorrelation $\rho \in (-1; 1)$

If we know the autocorrelation coefficient ρ, matrix $E(\boldsymbol{u}\boldsymbol{u}^T)$ is :

$$E(\boldsymbol{u}\boldsymbol{u}^T) = \sigma_u^2 \boldsymbol{H} = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{bmatrix}$$

(we assume homoskedasticity and known variance $\operatorname{var}(u_t)$)

GLSM – matrix form

- Explanation of matrix $E(\boldsymbol{u}\boldsymbol{u}^T)$: start with $u_t = \rho u_{t-1} + \varepsilon_t$
  $u_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \ldots$ where ε are $i.i.d.$
  Hence, $\sigma_u^2 \equiv \text{var}(u_t) = \sigma_\varepsilon^2 + \rho^2\sigma_\varepsilon^2 + \rho^4\sigma_\varepsilon^2 + \rho^6\sigma_\varepsilon^2 + \cdots = \frac{\sigma_\varepsilon^2}{1-\rho^2}$
  (Note: $\text{cov}(u_t, u_{t+s}) = \rho^s\sigma_u^2$ is independent of t if $|\rho| < 1$.)

- That is why:

$$H = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \ldots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \ldots & \rho^{n-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & \rho & 1 \end{bmatrix}$$

- Usually, we do not know the matrix $\boldsymbol{H}$, it must be estimated. In such case, we call the method Feasible General Least Squares (FGLS).

Prais - Winsten method (transformation)

- Transformation by $\boldsymbol{T}$: transformed vector $\boldsymbol{y}$ and matrix $\boldsymbol{X}$

$$\boldsymbol{T} = \sqrt{\frac{1}{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & -\rho & 1 & 0 \\ 0 & 0 & -\rho & 1 \end{bmatrix}, \quad \boldsymbol{y}^* = \begin{bmatrix} y_1\sqrt{1-\rho^2} \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \cdots \end{bmatrix},$$

$$\boldsymbol{X}^* = \begin{bmatrix} \sqrt{1-\rho^2} & x_{11}\sqrt{1-\rho^2} & x_{21}\sqrt{1-\rho^2} \\ 1-\rho & x_{12} - \rho x_{11} & x_{22} - \rho x_{21} \\ 1-\rho & x_{13} - \rho x_{12} & x_{23} - \rho x_{22} \\ \cdots & \cdots & \cdots \end{bmatrix}$$

- P-W transformation: quasi-differencing of all variables of the LRM and an approximation for the first period. In the transformation, we skip the fraction in front of $\boldsymbol{T}$-matrix. As a constant, it does not influence the regression result.

Cochrane-Orcutt method (transformation)

- P-W method without the approximation of the first observation.

For both the P-W and C-O methods, we usually use iterations to make the estimates of the autoregressive coefficient and regression parameters of the model more accurate.

Additional TS topics

- Heteroskedasticity in time series
  - Autoregressive Conditional Heteroskedasticity (ARCH) models included