Block 2
Regression models based on time series
Stationarity, cointegration, ECM, Distributed lag models, PAM, AEH

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

Outline

1. Time series-based LRMs – repetition from BSc. courses
2. Stationarity, unit root tests, cointegration
   - Unit root tests
   - Cointegration, ECM
3. TS & forecasting
4. Finite and infinite distributed lag models
   - Polynomial distributed lag
   - Geometric distributed lag (Koyck)
   - Rational distributed lag (RDL)
   - Partial adjustment model (PAM)
   - Adaptive expectations hypothesis (AEH)
   - Rational expectations

Time series-based LRMs – repetition from BSc. courses

- TS is a stochastic (random) process, a sequence of observations indexed by time.
  Observed TS: one realization of a stochastic process.

- Static models

$$y_t = \beta_1 + \beta_2 x_t + u_t, \qquad t = 1, 2, \ldots, n$$

- Finite distributed Lag (FDL) model:

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

- Infinite distributed lag (IDL) model:

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

- Dynamic models: lag order
  For convenience, β subscripts follow lag order,
  Impact / lagged / long-run multiplier,
  Effect of temporary (one-off) $\times$ permanent increase in x,
  Lag distribution (function).

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 $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + u_t$.
- 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:
  $E(u_t|\boldsymbol{X}) = 0, \quad t = 1, 2, \ldots, n$

  Note: Contemporaneous exogeneity (relaxed version):
  $E(u_t|x_{t1}, x_{t2}, \ldots, x_{tk}) = E(u_t|\boldsymbol{x}_t) = 0.$

G-M assumptions for TSRMs

- TS.4 Conditional Homoscedasticity

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

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

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

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

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

Properties of OLS estimators

- Under TS.1 - TS.3, OLS estimators are unbiased.
- Under assumptions TS.1 - TS.5,

  $\text{var}(\hat{\beta}_j) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\hat{\sigma}^2 = \frac{SSR}{n-K-1}$.
- 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 (non-stationarity and cointegration topics discussed next).

Detrending and deseasonalizing

Detrending algorithm (based on FWL theorem):

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 x_{t3} + \hat{\beta}_4 t$$

- Regress each variable on constant, time and save residuals

$$\ddot{y}_t, \ddot{x}_{t2}, \ddot{x}_{t3}, \qquad t = 1, 2, \ldots, n$$

- Regress $\ddot{y}_t$ on $\ddot{x}_{t2}$, $\ddot{x}_{t3}$

$$\hat{\ddot{y}}_t = \hat{\beta}_2 \ddot{x}_{t2} + \hat{\beta}_3 \ddot{x}_{t3}$$

- Coefficients $\hat{\beta}_2$, $\hat{\beta}_3$ 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 (if 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.

- Better approach: 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.

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

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 x_{t3} + \hat{\gamma}_1 \, dummy1 + \hat{\gamma}_2 \, dummy2 + \hat{\gamma}_3 \, dummy3$$

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

$$\ddot{y}_t, \ddot{x}_{t2}, \ddot{x}_{t3}, \qquad t = 1, 2, \ldots, n$$

- Regress $\ddot{y}_t$ on $\ddot{x}_{t2}$, $\ddot{x}_{t3}$

$$\hat{\ddot{y}}_t = \hat{\beta}_2 \ddot{x}_{t2} + \hat{\beta}_3 \ddot{x}_{t3}$$

- Coefficients $\hat{\beta}_2$, $\hat{\beta}_3$ from this regression are the same as in the original regression

Stationary and weakly dependent time series

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

  Time series is strictly stationary if its marginal and all joint distributions are invariant across time.
- **Covariance stationarity:** first two moments and auto-covariance do not change over time.
  $E(y_t) = \mu, \qquad \mathrm{var}(y_t) = \sigma^2, \qquad \mathrm{cor}(y_t, y_{t+h}) = f(h)$
- **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.

Notes on CLT & LLN

CLT implies that the sum of independent random variables (or weakly dependent RVs) – if centered and standardized by its s.d. – has an asymptotic distribution $N(0, 1)$.

LLN theorem implies that the average from a random sample converges in probability to the population average; LLN holds for stationary and weakly dependent series.

Examples of weakly dependent time series:

- Moving average process of order one: `ma(1)`

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

  where e_t is an *iid* time series.

  $$\text{corr}(y_t, y_{t+1}) = \tfrac{\alpha_1}{1+\alpha_1^2},$$

  observations with higher time distance than 1 are uncorrelated.

- Stable autoregressive process of order 1: `ar(1)`

  Under stability condition $|\rho| < 1$, it can be demonstrated that (Wooldridge, Introductory econometrics, ch. 11.1):

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

  If stability condition 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

- TS.4' Contemporaneous homoscedasticity

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

- 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

Under assumptions TS.1', TS.2' and TS.3',

- OLS estimators are consistent (not unbiased)

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

Removing strict exogeneity (TS.3 $\rightarrow$ TS.3'): 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.

Theorem: Asymptotic normality of OLS: Under assumptions TS.1' – TS.5', OLS estimators are asymptotically normally distributed.

- Usual OLS standard errors, t-statistics and F-statistics are asymptotically valid.

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

Serial correlation in TSRM

- Causes and effects of autocorrelation

  DGP, dynamic (in)completeness of models

  unbiased $\hat{\beta}_j$, biased inference

- FGLS (only with strictly exogenous regressors)
- OLS + robust inference (Newey-West & other HAC s.e.)

Serial correlation in TSRM

Testing AR(1) with strictly exogenous regressors

- ρ estimation:

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

- Durbin-Watson test:

$$
d = \frac{\sum_{t=2}^{n} \left(\hat{u}_t - \hat{u}_{t-1}\right)^2}{\sum_{t=1}^{n} \hat{u}_t^2}
$$

.

  - small sample validity (conditions apply)
  - d-statistic: symmetric distribution $\langle 0, 4 \rangle$, $E(d) = 2$
  - $d \approx 2(1 - \hat{\rho})$ i.e. $\hat{\rho} \approx 1 - d/2$. Test for: $H_0 : \rho = 0$.

Serial correlation

Testing serial correlation with general regressors

- Testing AR(1):

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

$$H_0 : \rho = 0$$

We can use a heteroscedasticity-robust version of the t-test.

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

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

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

Use F-test or LM-test: $(n-q)R_a^2 \underset{H_0}{\sim} \chi^2_{(q)}$.

GLSM, matrix form description

Generalized linear regression model (GLRM) accomodates both autocorrelated and heteroscedastic residuals in the LRM.

- GLRM properties:
  - $E(\boldsymbol{u}) = \mathbf{0}$
  - $E(\boldsymbol{u}\boldsymbol{u}^T) = \sigma^2 \boldsymbol{H}$
    (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 $\boldsymbol{THT}^T = \boldsymbol{I}_n$) to get:

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

  in the transformed model.

GLSM, matrix form description

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

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

- 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 description

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 description

- Matrices $\boldsymbol{H}$ and $\boldsymbol{T}$ of the GLSM algorithm differ for heteroscedasticity and autocorrelation.
  - heteroscedasticity (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 description

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 :

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

(we assume homoscedasticity and known variance $\mathrm{var}(u_t)$)

GLSM, matrix form description

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

- That is why: $H = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{bmatrix}$

- 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).

GLSM, matrix form description

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.

Stationarity, unit root tests, cointegration

- Stationarity
- Unit root tests
- Cointegration

Weakly and strongly dependent TS

Weakly dependent time series

- Moving average process of order one `ma(1)`
  $x_t = e_t + \alpha_1 e_{t-1}$, where e_t is *i.i.d.* time series.
  Observations with higher time distance than 1 are uncorrelated. This process is stationary.

- For stable autoregressive process of order 1 `ar(1)`:
  $y_t = \rho_1 y_{t-1} + e_t \Rightarrow \text{cor}(y_t, y_{t+h}) = \rho_1^h$

  If stability condition $|\rho| < 1$ holds, the process is weakly dependent because correlation converges to zero with growing h. Also, this process is stationary for $y_0 = 0$.

Weakly and strongly dependent TS

Strongly dependent time series:
Random walk:

$$
\begin{aligned}
y_t &= y_{t-1} + e_t \\
y_t &= y_{t-2} + e_{t-1} + e_t \\
y_t &= y_{t-3} + e_{t-2} + e_{t-1} + e_t \\
&\dots \\
y_t &= y_0 + e_1 + \cdots + e_{t-1} + e_t
\end{aligned}
$$

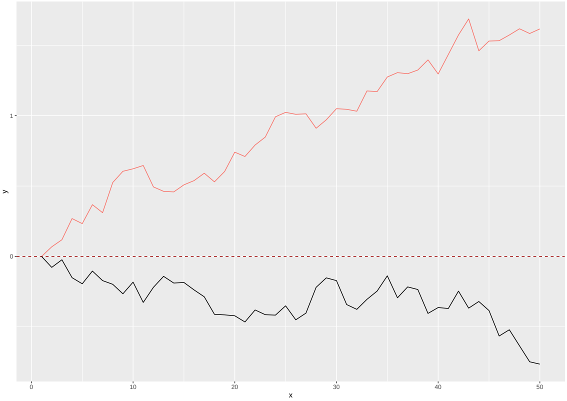
Shocks have permanent effects, the series is not covariance stationary and is strongly dependent.

$$
\begin{aligned}
E(y_t) &= E(y_0) \\
\mathrm{var}(y_t) &= \sigma_e^2 t \\
\mathrm{cor}(y_t, y_{t+h}) &= \sqrt{t/(t+h)}
\end{aligned}
$$

Correlation decreases very slowly and speed depends on t.

Strongly dependent TS

- Two realizations of a random walk



Strongly dependent TS

- Random walk with a drift

$$y_t = \alpha_0 + y_{t-1} + e_t \Rightarrow y_t = \alpha_0 t + e_t + e_{t-1} + \cdots + e_1 + y_0$$

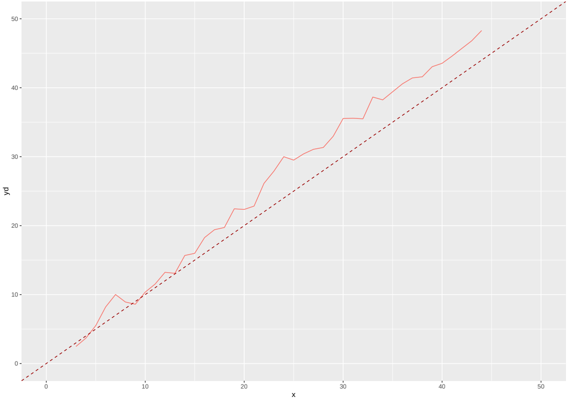
A linear trend with random walk around the trend.
It is neither covariance stationary nor weakly dependent.

$$\begin{aligned} E(y_t) &= \alpha_0 t + E(y_0) \\ \mathrm{var}(y_t) &= \sigma_e^2 t \\ \mathrm{cor}(y_t, y_{t+h}) &= \sqrt{t/(t+h)} \end{aligned}$$

Correlation decreases very slowly and decline speed depends on t.

Weakly and strongly dependent TS

- Realization of random walk with a drift



- Different realizations of trending TS (weakly dependent around the trend) may produce similar time series.

Weakly and strongly dependent TS

$$y_t = \boxed{1 \cdot y_{t-1}} + \boxed{u_t} = y_{t-1} + u_t$$

- Unit root process: $y_t = y_{t-1} + u_t$;
  where: u_t is a weakly dependent series.
- Random walk is a special case of the unit root process
  where: $u_t \sim Distr(0, \sigma_u^2), iid$

We need to distinguish strongly and weakly dependent TS:

- Economic reasons:
  In strongly dependent series, shocks or policy changes have long or permanent effects; in weakly dependent series, their effect is only temporary.
- Statistical reasons:
  Analysis with strongly dependent series must be handled in specific ways.

Integrated series

Terminology - Order of integration

- Weakly dependent TS are integrated of order zero: $I(0)$.
- If we have to difference a TS once to get a weakly dependent TS, then it is integrated of order 1: $I(1)$.
- Example of a $I(1)$ process:

$$\begin{aligned} y_t &= y_{t-1} + e_t \qquad \Rightarrow \Delta y_t = y_t - y_{t-1} = e_t \\ \log y_t &= \log y_{t-1} + e_t \Rightarrow \Delta \log y_t = e_t \end{aligned}$$

- A time series is integrated of order d: $I(d)$, if it becomes a weakly dependent TS after being differenced d times.

Unit roots tests

Unit root tests help to decide if a time series is $I(0)$ or not

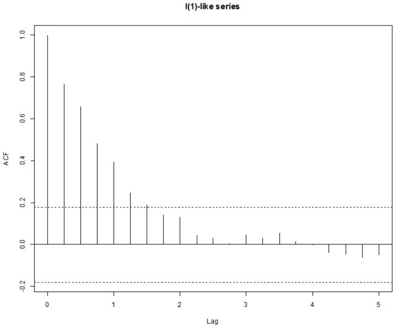
- Use either some informal procedure or a unit root test
- Informal procedures
  - Analyze autocorrelation of the first order

$$\hat{\rho}_1 = \widehat{\mathrm{corr}}(y_t, y_{t-1})$$

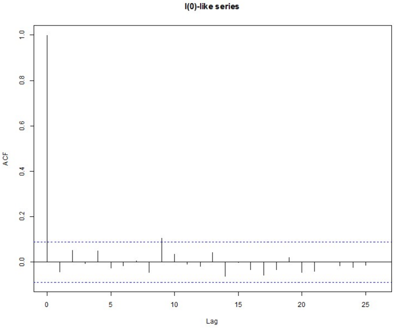
  - If $\hat{\rho}_1$ approaches 1, it indicates that the series can have unit root. Alternatively, it could have a deterministic trend.
- We can analyze sample autocorrelations using a correlogram

Unit root tests

Correlogram: $\rho_h = \frac{cov(y_t, y_{t-h})}{\sigma_{y_t} \cdot \sigma_{y_{t-h}}}$



$I(1)$-like series



$I(0)$-like series

Dickey-Fuller (DF) test – motivation

Unit root test in an `ar(1)` process:

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

$H_0 : \rho = 1, \quad H_1 : \rho < 1$

- Under H_0, y_t has a unit root.
    - For $\rho = 1 \ \wedge \ \alpha = 0 \rightarrow y_t$ is a random walk.
    - For $\rho = 1 \ \wedge \ \alpha \neq 0 \rightarrow y_t$ is a randomw walk with a drift and $E(y_t)$ is a linear function of t .
- Under H_1, y_t is a weakly dependent `ar(1)` process.

Dickey-Fuller (DF) test – motivation

Unit root test in an `ar(1)` process:

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

$$H_0 : \rho = 1, \quad H_1 : \rho < 1$$

For DF tests, $H_1 : \rho < 1$ is a common simplification to the full space of alternatives to $H_0 : \rho = 1$.

- For $|\rho| < 1$, y_t is weakly dependent (as $\mathit{plim}\, \rho^h = 0$) However, if unit root is likely to be present, the probability of $\rho < 0$ is negligible.
- We usually ignore the possibility of $\rho > 1$, as it would lead to explosive behavior in y_t.
  $\ldots |\rho| > 1$ would allow for explosive oscillations in y_t.

Dickey Fuller (DF) test

- Basic equation for unit root test in an `ar(1)` process:
  $y_t = \alpha + \rho y_{t-1} + e_t$
- For DF test, we apply a suitable transformation to y_t:
  we subtract y_{t-1} from both sides of the equation:

  $\Delta y_t = \alpha + (\rho - 1) y_{t-1} + e_t;$ apply substitution: $\theta = (\rho - 1)$

  i.e.

  $\Delta y_t = \alpha + \theta y_{t-1} + e_t;$ now:

  $$\begin{aligned} H_0 &: \rho = 1 \Leftrightarrow \boxed{H_0 : \theta = 0} \\ H_1 &: \rho < 1 \Leftrightarrow \boxed{H_1 : \theta < 0} \end{aligned}$$

- We use a t-ratio for testing $H_0 : \theta = 0$. However:
  Under H_0, t-ratios don't have a t-distribution, but follow a DF-distribution. (-negative- critical values of the DF distribution are much farther from zero)
- Critical values for the DF distribution are available from statistical tables and implemented in most relevant SW packages.

DF test & ADF test

Unit root time series can manifest various levels of complexity. Hence, DF test is usually performed using the following three specifications:

$\Delta y_t = \boxed{\theta y_{t-1}} + e_t$	random walk
$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + e_t$	random walk with a drift
$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + \delta t + e_t$	random walk with a drift and trend

DF test is the same ($H_0 : \theta = 0$) for all specifications /critical values diffrer/

Augmented Dickey-Fuller (ADF) test is a common generalization of DF test (example: Augmentation of the DF test for the 2^{nd} specification)

$$\Delta y_t = \alpha + \boxed{\theta y_{t-1}} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$$

- When estimating θ, we control for possible $ar(p)$ behavior in Δy_t.
- ADF test has the same null hypothesis as a DF test $\rightarrow H_0 : \theta = 0$.

Unit root tests in R: package {urca}

Description of the options for the `ur.df()` function:

1. `type "none"`
   $\Delta y_t = \theta y_{t-1} + e_t$
   `tau1`: we test for $H_0 : \theta = 0$ (unit root)

2. `type "drift"`
   $\Delta y_t = \alpha + \theta y_{t-1} + e_t$
   `tau2`: $H_0 : \theta = 0$ (unit root)
   `phi1`: $H_0 : \theta = \alpha = 0$ (unit root and no drift)

3. `type "trend"`
   $\Delta y_t = \alpha + \theta y_{t-1} + \delta t + e_t$
   `tau3`: $H_0 : \theta = 0$ (unit root)
   `phi2`: $H_0 : \theta = \alpha = \delta = 0$ (unit root, no drift, no trend)
   `phi3`: $H_0 : \theta = \delta = 0$ (unit root and no trend)

Multiple other unit root tests exist:
(KPSS, tests for seasonal data, break in the DGP, etc.).

Unit root tests

- ADF test for TS with trend

$$\Delta y_t = \alpha + \theta y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$$

  Under the alternative hypothesis of no unit root, the process is trend-stationary.

- The critical values in the ADF distribution with time trend are even more negative as compared to random walk and random walk with a drift.

- When using DF/ADF specification 1 or 2 (R-W, R-W with drift) to test for unit root in a clearly trending TS, the test would not have sufficient power (we would not reject H_0 for trending weakly dependent TS).

Unit roots and trend-stationary series

- $\Delta y_t = \alpha + \theta y_{t-1} + \delta t + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + e_t$
- Terminology:
  - Stochastic trend: $\theta = 0$
    Also called **difference-stationary process**: y_t can be turned into $I(0)$ series by differencing. Terminology emphasizes stationarity after differencing y_t instead of weak dependence in differenced TS.
  - Deterministic trend: $\delta \neq 0, \quad \theta < 0$
    Also called **trend-stationary process**: has a linear trend, not a unit root. y_t is weakly dependent - $I(0)$ - around its trend. We can use such series in LRMs, if trend is also used as regressor.
- DF/ADF tests are not precise tools. Distinguishing between stochastic and deterministic trend is not easy (sample size!).

Stationarity in ar(p) processes

Lag operators:

$$\begin{aligned} Lx_t &= x_{t-1} \\ L(Lx_t) = L^2 x_t &= x_{t-2} \\ &\cdots \\ L^p x_t &= x_{t-p} \end{aligned}$$

Using lag operators,

$$AR(p) : x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t$$

can be rewritten as:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) x_t = \alpha + u_t$$

Stationarity in ar(p) processes

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)x_t = \alpha + u_t \quad (1)$$

Stochastic process (1) will only be stationary if the roots of corresponding equation (2) are all greater than unity in absolute value

$$1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p = 0 \quad (2)$$

Illustration 1 – AR(1) process:

$$x_t = \alpha + \phi x_{t-1} + u_t \quad (3)$$

$$(1 - \phi L)x_t = \alpha + u_t$$

$$1 - \phi L = 0$$

$$L = 1/\phi$$

$$\text{For (3) to be stationary}, |L| > 1 \leftrightarrow -1 < \phi < 1$$

Stationarity in ar(p) processes

Illustration 2 – AR(1) process:

$$x_t = 2 + 3.9x_{t-1} + 0.6x_{t-2} - 0.8x_{t-3} + u_t$$

To evaluate stationarity of x_t, we use

$$1 - 3.9L - 0.6L^2 + 0.8L^3 = 0,$$

which can be factorized:

$$(1 - 0.4L)(1 + 0.5L)(1 - 4L) = 0$$

$$1^{st} root : L = 2.5$$
$$2^{nd} root : L = -2$$
$$3^{rd} root : L = 0.25 \Rightarrow \text{ TS is non-stationary}$$

Handling trend-stationary time series

- Trend-stationary TS fulfill TS.1' assumption. We can use them in regressions if we have time trend among regressors.

- Strongly dependent time series do not fulfill TS.1' assumption. We cannot use them in regressions directly.
- Sometimes, we can transform such series into weakly dependent time series.
  - Sometimes, taking logarithms helps.
  - Differencing is popular, but it has drawbacks.

Handling strongly dependent time series

Example

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \qquad y_t, x_t \sim I(1) \qquad (4)$$
$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1} \qquad \varepsilon_t \sim i.i.d. \qquad (5)$$
$$\Delta y_t = \beta_1 \Delta x_t + v_t \qquad v_t = \varepsilon_t - \varepsilon_{t-1} \qquad (6)$$

1. Coefficient β_1 does not change between (4) and (6). However, equations (4) and (6) are different.
   $\Rightarrow$ β_1 is the change in y_t for a unit change in x_t, but is it also the change in the growth of y for a unit change in the growth of x.
2. Three problems
   1. v_t is no longer i.i.d.
   2. We loose information linked with the levels of variables, short term relation are stressed.
   3. Estimates often generate bad long-term predictions: $\Delta \hat{y}_t = \hat{\beta}_1 \Delta x_t; \ldots$ what if $\beta_0 \neq 0$?

Handling strongly dependent time series

Some properties of integrated processes

1. The sum of stationary and non-stationary series must be non-stationary.
2. Consider a process $y_t = \alpha + \beta x_t$:
   - If x_t is stationary then y_t will be stationary.
   - If x_t is non-stationary then y_t will be non-stationary.
3. If two time series are integrated of different orders, then any linear combination of the series will be integrated at the higher of the two orders of integration.
4. Sometimes it turns out a linear combination of two $I(d)$ series is integrated of order less then d.

Spurious regression or cointegration

- **Spurious regression** Regressing one $I(1)$-series on another $I(1)$-series may lead to extremely high t-statistics even if the series are completely independent. Similarly, the R^2 of such regressions tend to be very high.
  Regression analysis involving time series that have a unit root may generate completely misleading inferences.

- **Cointegration** Fortunately, regressions with $I(1)$-variables are not always spurious: If there is a stable relationship between time series that, individually, display unit root behavior, these time series are called "cointegrated".

Spurious regression or cointegration

General definition of cointegration

Two $I(1)$-time series y_t, x_t are said to be cointegrated if there exists a stable relationship between them, where:

$$y_t = \alpha + \beta x_t + e_t, \quad e_t \sim I(0)$$

Cointegration (CI) test if CI parameters are known

For residuals of the known CI relationship:

$$e_t := y_t - \alpha - \beta x_t,$$

test whether the residuals have a unit root (DF/ADF and other unit root tests may be applied "directly").
If the unit root H_0 is rejected, y_t, x_t are cointegrated.

Spurious regression or cointegration

- **Testing for CI if the parameters are unknown**
  If the potential relationship is unknown, it can be estimated by OLS. After that, we test whether the regression residuals have a unit root. If the unit root is rejected, this means that y_t, x_t are cointegrated. Due to the pre-estimation of parameters, critical values are different than in the case of known parameters. (Software handles this automatically.)
- **The CI relationship may include a time trend**
  If the two series have differential time trends (drifts in this case), the deviation between them may still be $I(0)$ but with a linear time trend. In this case one should include a time trend in the CI-regression. Also, we have to use different critical values when testing residuals. (Software handles this automatically.)

Cointegration tests based on regression residuals

Engle-Granger test estimates a p-lag ADF equation:

$$\Delta \hat{u}_t = \theta\, \hat{u}_{t-1} + \sum_{j=1}^{p} \Delta \hat{u}_{t-j} + e_t$$

- Esentially, this is an ADF test on $\hat{u}_t$ $[\theta = (\rho - 1)]$
- Specific critical values apply (farther from 0 than t or DF).

Phillips-Ouliaris test estimates a DF equation:

$$\Delta \hat{u}_t = \theta\, \hat{u}_{t-1} + e_t$$

- The t-ratio is based on robust standard errors, different estimators exist for the robust standard errors.

In both cases (EG and PO), H_0 of unit root in $\hat{u}$ i.e. "no-cointegration" is tested.

Error correction model (ECM)

- It can be shown that when variables are cointegrated, i.e. when there exists a long-term relationship among them, their short-term dynamics are related as in a so-called error correction model (ECM).

Autoregressive distributed lag models

- Autoregressive distributed lag model with one regressor

$$\text{ADL}(p,q)\colon\ y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=0}^{q} \gamma_j x_{t-j} + u_t, \quad u_t \sim iid(0,\sigma^2)$$

- There are many useful modifications/simplifications to the ADL(p,q) process. For example:

$$\text{ADL}(1,1)\colon\ y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t\,. \qquad (7)$$

Additional ADL$(1,1)$ restriction: $\beta_1 = 1$ and $\gamma_1 = -\gamma_0$

gives a model in 1^{st} diffs.: $\Delta y_t = \beta_0 + \gamma_0 \Delta x_t + u_t\,.$

Error correction model (ECM)

For ADL(1,1) model (7), suppose there is an equilibrium value x° and in the absence of shocks, $x_t \to x^\circ$ as $t \to \infty$. Then, assuming absence of u_t errors, y_t converges to steady state: y°.

Hence, the ADL(1,1) model (7) can be re-written as:

$$y^\circ = \beta_0 + \beta_1 y^\circ + (\gamma_0 + \gamma_1) x^\circ$$

Solving this for y° as a function of x°, we get

$$y^\circ = \frac{\beta_0}{1-\beta_1} + \frac{\gamma_0+\gamma_1}{1-\beta_1} x^\circ = \frac{\beta_0}{1-\beta_1} + \lambda x^\circ$$

where $\lambda \equiv \frac{\gamma_0+\gamma_1}{1-\beta_1}$ and $|\beta_1| < 1$ is assumed.

Error correction model (ECM)

$$y^\circ = \frac{\beta_0}{1-\beta_1} + \lambda x^\circ$$

$$\lambda \equiv \frac{\gamma_0 + \gamma_1}{1-\beta_1}$$

- λ is the long-run derivative of y° with respect to x°.
- λ is an elasticity if both y° and x° are in logs.
- $\hat{\lambda}$ can be computed directly from the estimated parameters of the ADL(1,1) model (7).

Error correction model (ECM)

The ADL(1,1) equation (7) - repeated here for convenience:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t,$$

can be equivalently rewritten as follows:

$$\Delta y_t = \beta_0 + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t. \tag{8}$$

Again, $\lambda \equiv \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$ and $|\beta_1| < 1$ is assumed.

Equation (8) is an error-correction model (ECM).

Error correction model (ECM)

ECM: $\Delta y_t = \beta_0 + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t.$

- $(y_{t-1} - \lambda x_{t-1})$ measures the extent to which the long run equilibrium between y_t and x_t is not satisfied (at $t-1$).
- Consequently, $(\beta_1 - 1)$ can be interpreted as the proportion of the disequilibrium $(y_{t-1} - \lambda x_{t-1})$ that is reflected in the movement of y_t, i.e. in Δy_t.
- $(\beta_1 - 1)(y_{t-1} - \lambda x_{t-1})$ is the **error-correction term**.
- Many ADL(p, q) specifications can be re-written as ECMs.
- **ECMs can be used with non stationary TS**.
- ECMs $(\beta_1 - 1)$ is essentially the same as θ from Partial adjustment model (discussed next).

Error correction model (ECM)

Some more complicated ECMs:

1) We can use higher order lags, e.g. ADL(2,2):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \gamma_0 x_t + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + u_t,$$

to establish ECMs. It is again possible to rearrange and re-parametrize ADL(2,2) to get an ECM. More than one re-parameterization is possible.

2) More than two variables can enter into an equilibrium relationship.

ECM: non-stationary & cointegrated series

Superconsistency: $y_t = \beta_0 + \beta_1 x_t + u_t$

1. Provided x_t and y_t are cointegrated, the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ will be consistent.
2. $\hat{\beta}_j$ converge in probability to their true values β_j more quickly in the cointegrated non-stationary case than in the stationary case (asymptotic efficiency).

Consequences:

For simple static regression between two cointegrated variables: $y_t, x_t \sim C(1,1)$, super-consistency applies (with deterministic regressors such as intercept and trend added upon relevance). Dynamic misspecifications do not necessarily have serious consequences. This is a large sample property - in small samples, OLS estimators are biased.
(Specific statistical inference applies to cointegrating vectors.)

ECM: non-stationary & cointegrated series

Granger representation theorem:[1]
If two TS x_t and y_t are cointegrated, the short-term disequilibrium relationship between them can be expressed in the ECM form

$$\Delta y_t = lagged(\Delta y, \Delta x) - \delta u_{t-1} + \varepsilon_t \tag{9}$$

where $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1}$ is the disequilibrium error and δ is a short-run adjustment parameter.
Note: as u is on the scale of y, δ can be interpreted in percentages.
Example: $\delta = 0.8 \rightarrow 80\%$ of the disequilibrium error gets corrected between $t-1$ and t (on average).

Two implications:

1. The general-to-specific model search can focus on ECMs
2. Engle-Granger two-stage procedure

[1] Engle and Granger (1987)

ECM: non-stationary & cointegrated series

Engle-Granger two-stage procedure:

We short-cut the search of an ECM from a general model:

1^{st} stage: Estimation of the cointegrating (static) regression and saving residuals

$$\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$$

2^{nd} stage: Use residuals $\hat{u}_{t-1}$ in (9) instead of u_{t-1} and estimate by OLS

Estimators are consistent and asymptotically efficient, but biased in small samples.

Assumptions: y_t and x_t are non-stationary and cointegrated.

ECM: non-stationary & cointegrated series

Possibility of more cointegrating vectors:

Long-run relationship: $y_t = \beta_0 + \beta_1 x_t + \beta_2 w_t + \beta_3 z_t + u_t$, with all observed variables are $I(1)$

If this long-run relationship exists, then the disequilibrium error

$$u_t = [y_t - \beta_0 - \beta_1 x_t - \beta_2 w_t - \beta_3 z_t] \quad \sim I(0). \tag{10}$$

If a linear combination of variables such as (10) is stationary,

then the coefficients in this relationship form a cointegrating vector, e.g. $(1, -\beta_1, -\beta_2, -\beta_3)$. In the multivariate case, there

may be more than one linearly independent stationary combinations linking the cointegrated variables (topic discussed separately).

Cointegration: the existence of at least one cointegrating vector.

Cointegration among more than two variables

Testing and estimation
Cointegration can be tested using the EG and/or PO tests

Only one cointegrating vector exists
Estimation can proceed by the Engle-Granger two-stage method for ECMs.

Two or more cointegrating vectors
Engle-Granger two-stage method is not applicable. Johansen (1988) suggests a maximum likelihood approach.

TS & forecasting

- Chow tests
- Forecasts from TS-based models

Chow tests

For any LRM: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$

- Say, the sample (time series) for a period $t = 1, 2, \ldots, T$ may be conveniently divided into two groups: $T_1 + T_2 = T$.
  [consider two periods: fixed vs. floating F/X rates]
  [pre-EU accession vs. post-EU accession period]
  [applies to CS data as well; e.g Male/Female]

- Now, the LRM's vectors and matrices may be partitioned as follows:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix}$$

where $\boldsymbol{y}_1' = (y_1, \ldots, y_{T_1})$, $\boldsymbol{y}_2' = (y_{T_1+1}, \ldots, y_T)$, etc.
i.e. $\{\boldsymbol{y}_1, \boldsymbol{X}_1\} \in T_1$, $\{\boldsymbol{y}_2, \boldsymbol{X}_2\} \in T_2$.

Chow tests

For any LRM: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, Chow test can be based on an auxiliary regression (unrestricted model for the F test):

- $$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2 \end{bmatrix} \gamma + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix}$$

  where $\mathbf{0}$ is a zero-matrix of the same dimensions as $\boldsymbol{X}_1$, i.e. $(T_1 \times k)$.

Also, we can see that:

- $T_1: \quad \hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$
- $T_2: \quad \hat{\boldsymbol{y}} = \boldsymbol{X}(\hat{\boldsymbol{\beta}} + \hat{\gamma})$

Note: Power of the test depends on proper T_1 vs. T_2 cutoff. Chow test may be generalized for 3+ time periods (groups).

Chow tests

For our unrestricted model:

- $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\gamma} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix}$

We can formulate the null of no structural change in model dynamics between the two time periods (groups) as follows:

- $H_0: \quad \boldsymbol{\gamma} = \mathbf{0}$, i.e.: $\gamma_1 = \gamma_2 = \gamma_3 = \cdots = \gamma_k = 0$
- $H_1: \quad \neg H_0$

This can be tested using an F-test (or its HC version):

- $F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \times \frac{n-2K}{K} \underset{H_0}{\sim} F[K, (n-2K)]$

Note: Here, β_1 and γ_1 both relate to the intercept.

Chow test - CS-based example

A simple Chow test example for CS data:
(to assess whether parameters are equal for M/F students.)

- Original model (Chow test restricted model):
  ... based on the well known Wooldridge dataset.

  $$cumgpa = \beta_1 + \beta_2 sat + \beta_3 hsperc + \beta_4 tothrs + u$$

- Auxiliary model (Chow test unrestricted model):

$$\begin{aligned} cumgpa = \beta_1 & \qquad + \gamma_1 female \\ + \beta_2 sat & \qquad + \gamma_2 (female \times sat) \\ + \beta_3 hsperc & + \gamma_3 (female \times hsperc) \\ + \beta_4 tothrs & + \gamma_4 (female \times tothrs) + u \end{aligned}$$

Chow test - CS-based example (contd.)

- Null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$

  If all interactions effects are zero, we have the same regression function for both groups.

- Estimate of the unrestricted model

$$\begin{aligned}\widehat{cumgpa} = \underset{(.21)}{1.48} - \underset{(.411)}{.353}\,female + \underset{(.0002)}{.0011}\,sat + \underset{(.00039)}{.0075}\,(female \times sat) \\ - \underset{(.0014)}{.0085}\,hsperc - \underset{(.00316)}{.00055}(female \times hsperc) \\ + \underset{(.0009)}{.0023}\,tothrs - \underset{(.00163)}{.00012}(female \times tothrs)\end{aligned}$$

... t-tests cannot be used to evaluate the joint H_0.

Chow test - CS-based example (contd.)

- F-statistic:

$$F = \frac{(SSR_r - SSR_{ur})/K}{SSR_{ur}/(n-2K)} = \frac{(85.515 - 78.355)/4}{78.355/(366-8)} \approx 8.18$$

  ... using p-value, we reject the null hypothesis

- **Important:** Chow tests (all types) assume constant error variance across groups.

Chow 1: stability test for TS

Here, the F-statistic for the Chow test is calculated in an alternative way (Chow 1):

- For a suitable (potential) "breakpoint", we divide our sample $\{t = 1, 2, \ldots, T\}$ in two groups:
  "T_1" with $\{t = 1, 2, \ldots, T_1\}$ and
  "T_2" with $\{t = T_1+1,\ T_1+2, \ldots, T\}$
  ... note that the choice of T_1 is arbitrary
  ... (breakpoint-searching algorithms can be used)
- Run separate regressions for both T_1, T_2 groups;
  the SSR_{ur} is given by the sum of the SSRs of the two separately estimated regression models.
  ... sufficient observations in T_1 and T_2 are required (d.f.)
- Run the original (restricted) regression model on the whole sample T and store SSR_r.

Chow 1: stability test for TS

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \cdot \frac{T-2K}{K} \underset{H_0}{\sim} F(\,K\,,\,T-2K\,)$$

where
$SSR_{ur} = SSR_{T_1} + SSR_{T_2}$
$SSR_r \;\; = SSR_T$
k is the number of parameters (including intercept) in LRM

H_0: stable structure of coefficients - no statistically significant differences between T_1 and T_2.
H_1: $\neg H_0$ (assume structural change in parameters over time)

Note: Chow 1 can be generalized for G time periods ($G-1$ "breakpoints").
... In such case, $SSR_{ur} = \sum_{g=1}^{G} SSR_g$, d.f. $= T - GK$
... and we assume $T_g > K$ for all time groups.
... (only usable for small G-values, problematic setup of breakpoints)

Chow 2: prediction test for TS

Sometimes, we do not have enough observations to estimate the LRM separately for T_1 and T_2 as in the Chow 1 test.

In such case, we can use Chow 2: test of prediction unsuitability (slightly different F-statistics).

- The whole period is again divided into two subsets: $T = T_1 + T_2$.
- T_1 is the “base” period (sample size)
- T_2 is the number of “additional” observations, it usually corresponds to an ex-post prediction period

Chow 2: prediction test for TS

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \cdot \frac{T_1 - K}{T_2} \underset{H_0}{\sim} F(\, T_2 \,,\, T_1 - K \,)$$

where
$SSR_{ur} = SSR_{T_1}$ (from LRM estimated for "base" period)
$SSR_r \; = SSR_T$ (from LRM estimated for the whole period)
k is the number of parameters (including intercept) in LRM

H_0: additional (T_2) observations come from the same DGP as in T_1.
H_1: $\neg H_0$ (assume significant differences between samples)
...If H_0 is rejected, we would expect large differences
...between predictions and actual observations of y_t.

If enough T_1 and T_2 observations are available, Chow 1 is preferred (compared to Chow 2) as it has more "power".

TS & forecasting

- **One-step-ahead forecast** f_t
  Forecast error $e_{t+1} = y_{t+1} - f_t$
  Information set: I_t
  Loss function: e_{t+1}^2 or $|e_{t+1}|$
  In forecasting, we minimize $E(e_{t+1}^2 | I_t) = E[(y_{t+1} - f_t)^2 | I_t]$
  Solution: $E(y_{t+1} | I_t)$

- **Multiple-step-ahead forecast** $f_{t,h}$
  Solution: $E(y_{t+h} | I_t)$

For some processes, $E(y_{t+1}|I_t)$ is easy to obtain:

1. **Martingale process (MP):**
   If $E(y_{t+1}|y_t, y_{t-1}, \ldots, y_0) = y_t, \forall t \geq 0$ then $\{y_t\}$ is MP
   $f_t = y_t$
   If a process $\{y_t\}$ is a martingale then $\{\Delta y_t\}$ is martingale difference sequence (MDS)

$$E(\Delta y_{t+1}|y_t, y_{t-1}, \ldots, y_0) = 0$$

2. **Process with exponential smoothing:**

$$E(y_{t+1}|I_t) = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \cdots + \alpha(1-\alpha)^t y_0; \;\; 0 < \alpha < 1.$$

   Set $f_0 = y_0$, then for $t \geq 1 : \; f_t = \alpha y_t + (1-\alpha) f_{t-1}$

3. **Regression models**
   - Static model: $y_t = \beta_0 + \beta_1 x_t + u_t$
     $E(y_{t+1}|I_t) = \beta_0 + \beta_1 x_{t+1} \rightarrow$ Conditional forecasting
     I_t contains $x_{t+1}, y_t, x_t, \ldots, y_1, x_1$
     Here, knowledge of x_{t+1} is assumed (forecast condition).

     $E(y_{t+1}|I_t) = \beta_0 + \beta_1 E(x_{t+1}|I_t) \rightarrow$ Unconditional forecasting
     I_t contains $y_t, x_t, \ldots, y_1, x_1$
     Here, x_{t+1} needs to be estimated before y_{t+1}

   - Dynamic models depending on lagged variables only:
     $y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 x_{t-1} + u_t$
     $E(u_t|I_{t-1}) = 0$
     $E(y_{t+1}|I_t) = \delta_0 + \alpha_1 y_t + \gamma_1 x_t$
     Also, we can use more lags, drop or add regressors ...

One-Step-Ahead Forecasting with

$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 x_{t-1} + u_t$:

point forecast: $\hat{f}_t = \hat{\delta}_0 + \hat{\alpha}_1 y_t + \hat{\gamma}_1 x_t$

forecast error: $\hat{e}_{t+1} = y_{t+1} - \hat{f}_t$

s.e. of forecast: $\text{s.e.}(\hat{e}_{t+1}) = \{[\text{s.e.}(\hat{f}_t)]^2 + \hat{\sigma}^2\}^{1/2}$

forecast interval: essentially the same as prediction interval

approximate 95% forecast interval is: $\hat{f}_t \pm 1.96 \times \text{s.e.}(\hat{e}_{t+1})$

Example: File PHILLIPS

Forecasting US unemployment rate

$$\widehat{unem}_t = \underset{(.577)}{1.572} + \underset{(.097)}{.732}\, unem_{t-1}$$

$$n = 48, \overline{R}^2 = .544$$

$$\widehat{unem}_t = \underset{(.490)}{1.304} + \underset{(.084)}{.647}\, unem_{t-1} + \underset{(.041)}{.184}\, inf_{t-1}$$

$$n = 48, \overline{R}^2 = .677$$

Note that these regressions are not meant as causal equations. The hope is that the linear regressions approximate well the conditional expectation.

Evaluating forecast quality

- We can measure how forecasted values fit to actual observations (in-sample criteria, e.g. R^2)
- It is better, however, to evaluate the forecasting performance when forecasting out-of-sample values (out-of-sample criteria). For this purpose, use first n observations for estimation, and the remaining m observations to calculate the forecast errors $\hat{e}_{n+h}$
- Forecast evaluation measures:

  Mean Absolute Error $MAE = m^{-1} \sum_{h=1}^{m} |\hat{e}_{n+h}|$,

  Root Mean Squared Error $RMSE = (m^{-1} \sum_{h=1}^{m} \hat{e}^2_{n+h})^{1/2}$

  k-Fold Cross-Validation (kFCV) approach

- **Additional comments**
  - Multiple-step-ahead forecasts are possible, but necessarily less precise.
  - Forecasts may make use of deterministic trends, but the error made by extrapolating time trends too far into the future may be large.
  - Similarly, seasonal patterns may be incorporated into forecasts.
  - It is possible to calculate confidence intervals for the point multiple-step-ahead forecasts.
  - Forecasting $I(1)$ time series can be based on adding predicted changes (which are $I(0)$) to base levels.
  - Forecast intervals for $I(0)$ series converge to the unconditional variance, whereas for integrated series, they are unbounded.

Finite and infinite distributed lag models

- Finite and infinite distributed lag models
- Polynomial distributed lag
- Geometric distributed lag (Koyck transformation)
- Rationad distributed lag
- Partial adjustment model (PAM)
- Adaptive expectations hypothesis (AEH)
- Rational expectations

Finite and infinite distributed lag models

- Finite distributed Lag (FDL) model:

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

- Infinite distributed lag (IDL) model:

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

Dynamically complete (FDL) models

- Model is dynamically complete if we have a “sufficient” number of lags of regressors, so that no more additional lags would help with explanation of variance in the dependent variable.
- In dynamically incomplete models, we usually detect autocorrelation in the error term of the LRM.

Finite and infinite distributed lag models

Infinite distributed lag (IDL) models

- Lagged regressors extend back to infinity
- We cannot estimate IDL models without the use of simplifying restrictions on parameters,
  i.e. restrictions on lag distribution
- IDL models are useful under the assumption of lagged coefficients converging to zero as lag increases
- Order of the IDL model (∞),
  impact multiplier vs. long-run multiplier,
  temporary vs. permanent change in x,
  ... all analogical to FDL models

FDL: Polynomial distributed lag (Almon)

Used in Finite distributed lag models
...example below also extends to higher order polynomials

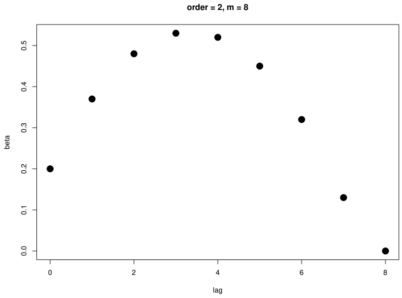
$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_m x_{t-m} + u_t \tag{11}$$

$$y_t = \alpha + (\sum_{i=0}^{m} \beta_i x_{t-i}) + u_t \ , \quad (i \text{ is the lag operator here})$$

Simplifying assumption:

$$\beta_i = k_0 + k_1 i + k_2 i^2 \tag{12}$$

$$\beta_0 = k_0$$
$$\beta_1 = k_0 + k_1 + k_2$$
$$\beta_2 = k_0 + k_1 \cdot 2 + k_2 \cdot 4$$
$$\ldots$$
$$\beta_m = k_0 + k_1 m + k_2 m^2$$



Polynomial distributed lag

- Almon-type transformation of (11) for $m = 8$ and $k = 2$:

$$y_t = \alpha + k_0 x_t + (k_0 + k_1 + k_2)x_{t-1} + (k_0 + 2k_1 + 4k_2)x_{t-2} + \\ + \cdots + (k_0 + 8k_1 + 64k_2)x_{t-8} + u_t \tag{13}$$

$$y_t = \alpha + k_0(x_t + x_{t-1} + \cdots + x_{t-8}) + \\ + k_1(x_{t-1} + 2x_{t-2} + \cdots + 8x_{t-8}) + \\ + k_2(x_{t-1} + 4x_{t-2} + \cdots + 64x_{t-8}) =$$

$$y_t = \alpha + k_0 \underbrace{\sum_{i=0}^{8} x_{t-i}}_{W_{0t}} + k_1 \underbrace{\sum_{i=1}^{8} i\, x_{t-i}}_{W_{1t}} + k_2 \underbrace{\sum_{i=1}^{8} i^2 x_{t-i}}_{W_{2t}} + u_t \tag{14}$$

$$y_t = \alpha + k_0 W_{0t} + k_1 W_{1t} + k_2 W_{2t} + u_t \tag{15}$$

We estimate (15), then calculate β_i (lags 0 to 8) as in (12) …note the reduction in estimated parameters (10 vs 4).

Polynomial distributed lag

- Method developed by Shirley Almon in the 60ies.
- Equation (15) can be generalized: for m lags, sums go to m, for higher order polynomials, we add more W-terms.
- Advantages of this approach:
  - Saves degrees of freedom
  - Removes the problem of multicollinearity
  - Does not affect the assumptions for u, because errors do not change during transformation
- In EViews, transformation is slightly modified.
- In R, routines are available.

Geometric distributed lag (Koyck)

IDL linear regression model: $y_t = f(x_t, x_{t-1}, x_{t-2}, \dots)$:

$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta_3 x_{t-3} + \cdots + u_t$

Assumption for the geometric δ_j weights:

- $\delta_j = \gamma \rho^j, \quad 0 < \rho < 1, \quad j = 0, 1, 2, \dots$

  (j is the lag operator here)
  $\delta_0 \equiv \gamma$ for convenience of notation (see RDL)
  $|\rho| < 1$ not used here (allows sign oscilations in δ_j).

  $\delta_j = \delta_{j-1}\,\rho, \quad 0 < \rho < 1.$
  alternative notation of the above

Instantaneous propensity (multiplier): $\delta_0 = \gamma\rho^0 = \gamma$

Long-run propensity (multiplier):

$\delta_0 + \delta_1 + \cdots = \gamma(1 + \rho + \rho^2 + \rho^3 + \dots) = \frac{\gamma}{1-\rho}$

Geometric distributed lag (Koyck)

Koyck transformation of the IDL model:

$$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \cdots + u_t \quad |\delta_j = \gamma\rho^j$$

$$y_t = \alpha + \gamma x_t + \gamma\rho x_{t-1} + \gamma\rho^2 x_{t-2} + \cdots + u_t \tag{16}$$

$$y_{t-1} = \alpha + \gamma x_{t-1} + \gamma\rho x_{t-2} + \gamma\rho^2 x_{t-3} + \cdots + u_{t-1} \quad | \times \rho \tag{17}$$

$$\rho y_{t-1} = \alpha\rho + \gamma\rho x_{t-1} + \gamma\rho^2 x_{t-2} + \cdots + \rho u_{t-1} \tag{18}$$

Now, we subtract (18) from (16):

$$y_t - \rho y_{t-1} = \underbrace{\alpha(1-\rho)}_{\alpha_0} + \gamma x_t + \underbrace{u_t - \rho u_{t-1}}_{v_t} \tag{19}$$

$$y_t = \alpha_0 + \gamma x_t + \rho y_{t-1} + v_t \tag{20}$$

Geometric distributed lag (Koyck)

IDL model: $y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \cdots + u_t$

Koyck transf.: $y_t = \alpha_0 + \gamma x_t + \rho y_{t-1} + v_t$

Using the Koyck transformation, we can calculate parameters of the IDL model from the estimated model after Koyck transformation:

$\hat{\delta}_0 = \hat{\gamma}$

$\hat{\delta}_j = \hat{\gamma}\,\hat{\rho}^{\,j}\ ;\ j = 1, 2, 3, \ldots$

$\hat{\alpha} = \frac{\hat{\alpha}_0}{1-\hat{\rho}}$

Problems of the Koyck transformation:

- in (20), regressor y_{t-1} is not exogenous $(v_t = u_t - \rho u_{t-1})$ …IVR discussed separately
- $v_t = u_t - \rho u_{t-1}$ is not i.i.d.
- Koyck transformation extends to models with multiple regressors only if common ρ (geometric decay) can be assumed.

Rational distributed lag (RDL)

The geometric distributed lag is a special case of rational distributed lag (RDL) model:

$$
\begin{aligned}
y_t &= \alpha_0 + \gamma x_t + \rho y_{t-1} + v_t \quad \text{(geometric distributed lag)} \\
y_t &= \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \rho y_{t-1} + v_t \quad \text{(RDL)} \qquad (21)
\end{aligned}
$$

This can be shown by successive substitution for the y_{t-1} term of RDL equation (21), where we get:

$$
\begin{aligned}
y_t = \alpha &+ \gamma_0 x_t + (\rho\gamma_0 + \gamma_1)x_{t-1} + \rho(\rho\gamma_0 + \gamma_1)x_{t-2} + \\
&+ \rho^2(\rho\gamma_0 + \gamma_1)x_{t-3} + \rho^3(\rho\gamma_0 + \gamma_1)x_{t-4} + \ldots \\
&+ \rho^{h-1}(\rho\gamma_0 + \gamma_1)x_{t-h} + \cdots + u_t \qquad (22)
\end{aligned}
$$

After estimating (21), we can calculate lag distribution for (22)

Rational distributed lag (RDL)

RDL specification:

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \rho y_{t-1} + v_t$$

can be used to calculate δ_h in the IDL model:

$$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \cdots + u_t$$

With RDLs, impact propensity $\gamma_0 \equiv \delta_0$ can differ in sign from lagged coefficients.

$\delta_h = \rho^{h-1}(\rho\gamma_0 + \gamma_1)$ corresponds to the x_{t-h} variable for $h \geq 1$.

... δ_0 may differ in sign from "lags", even if $\rho > 0$.

... for $\rho > 0$, δ_h doesn't change sign with growing $h \geq 1$.

Long-run propensity: $LRP = \frac{\gamma_0+\gamma_1}{1-\rho}$,

where $|\rho| < 1 \; \Rightarrow$ the sign of LRP follows the sign of $(\gamma_0 + \gamma_1)$.

Also, $u_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \cdots \quad MA(\infty)$

Koyck transformation and RDL model – summary

IDL model specification:

$$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta_3 x_{t-3} + \cdots + u_t$$

Koyck: $y_t = \alpha_0 + \gamma x_t + \rho y_{t-1} + v_t$

RDL: $y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \rho y_{t-1} + v_t$

Table 1: Koyck vs. RDL coefficients

	Koyck	RDL
Impact multiplier	$\delta_0 = \gamma$	$\delta_0 = \gamma_0$
Lagged multiplier (lag h)	$\delta_h = \gamma\rho^h$	$\delta_h = \rho^{h-1}(\rho\gamma_0 + \gamma_1)$
LRP	$\frac{\gamma}{1-\rho}$	$\frac{\gamma_0+\gamma_1}{1-\rho}$

Partial adjustment model

Partial adjustment model (PAM)
is based on two main assumptions:

1. LRM describes long-run behavior of y_t^*: the unobserved, expected/equilibrium/target/optimum value of y_t:

$$y_t^* = \alpha + \beta x_t + u_t \tag{23}$$

2. Between two time periods, y_t follows the process:

$$y_t - y_{t-1} = \theta(y_t^* - y_{t-1}), \quad 0 < \theta < 1 \tag{24}$$

Hence, the actual Δy_t is only a fraction of the "desirable" change from y_{t-1} to the optimum value of y_t^*.
...in the special case of $\theta = 1$, Δy_t leads to optimum.

Note: (24) can be re-written as: $y_t = \theta y_t^* + (1 - \theta) y_{t-1}$

Partial adjustment model (PAM)

Parameter estimation of PAM:

$$y_t^* = \alpha + \beta x_t + u_t \tag{23}$$

$$y_t = \theta y_t^* + (1-\theta) y_{t-1}, \qquad 0 < \theta < 1 \tag{24}$$

1. Substitute for y_t^* in (24) from (23):

$$\begin{aligned} y_t &= \alpha\theta \quad + \beta\theta x_t + \ (1-\theta) y_{t-1} \quad + \theta u_t \\ y_t &= \beta_0' \quad + \beta_1' x_t + \ \beta_2' y_{t-1} \qquad\quad + \theta u_t \end{aligned} \tag{25}$$

2. Estimate (25) and then calculate sought parameters of the PAM in (23) and (24):
   $\hat{\theta} = 1 - \hat{\beta}_2'$
   $\hat{\alpha} = \hat{\beta}_0' / \hat{\theta}$
   $\hat{\beta} = \hat{\beta}_1' / \hat{\theta}$

Note that θu_t can be independent of y_{t-1} and i.i.d.

Partial adjustment model (PAM)

Parameter interpretation of PAM:

$$\underline{\begin{aligned} y_t^* &= \alpha + \beta x_t + u_t \\ y_t &= \theta y_t^* + (1-\theta) y_{t-1}, \qquad 0 < \theta < 1 \end{aligned}}$$

$$\begin{aligned} y_t &= \alpha\theta + \ \beta\theta x_t + \ (1-\theta) y_{t-1} + \ \theta u_t \\ y_t &= \beta_0' \ + \ \beta_1' x_t + \ \beta_2' y_{t-1} + \ \theta u_t \end{aligned}$$

Parameters:

θ : $\theta = (1 - \beta_2')$ is the adjustment coefficient; higher $\hat{\theta}$ indicates higher speed of adjustment towards equilibrium.

β_1' : impact multiplier (short-run marginal propensity)

β : $\beta = \beta_1'/\theta$ is the long run multiplier.

Adaptive expectations hypothesis (AEH)

Adaptive expectations hypothesis (AEH)
model is based on two main assumptions:

1. LRM describes behavior of y_t, as a function of x_t^*: the unobserved, expected/equilibrium/target/optimum value of x_t (permanent income, potential output, etc.):

$$y_t = \alpha + \beta x_t^* + u_t. \tag{26}$$

2. The unobserved x_t^* process is defined as:

$$\begin{gathered} x_t^* - x_{t-1}^* = \phi(x_t - x_{t-1}^*), \qquad 0 < \phi < 1 \\ \Downarrow \\ x_t^* = \phi x_t + (1-\phi)x_{t-1}^*. \end{gathered} \tag{27}$$

with $\phi = 0$ for static expectations
and $\phi = 1$ for immediate adjustment.

Note: alternative 2^{nd} hypothesis: $x_t^* = \phi x_{t-1} + (1-\phi)x_{t-1}^*$.

Adaptive expectations hypothesis (AEH)

Parameter estimation of AEH model:

$$y_t = \alpha + \beta x_t^* + u_t, \tag{26}$$

$$x_t^* = \phi x_t + (1-\phi)x_{t-1}^* \tag{27}$$

Successive substitution for x_t^* from (27) to (26): IDL process

$$y_t = \alpha + \beta\phi x_t + \beta\phi(1-\phi)x_{t-1} + \beta\phi(1-\phi)^2 x_{t-2} + \cdots + u_t \tag{28}$$

After applying Koyck transformation, we get

$$\begin{aligned} y_t &= \alpha\phi \quad + \beta\phi x_t + \ (1-\phi)y_{t-1} \quad + v_t \\ y_t &= \beta_0' \quad + \beta_1' x_t + \ \beta_2' y_{t-1} \qquad + v_t \end{aligned} \tag{29}$$

Estimate (29), then calculate parameters in (26) and (27).

$\hat{\phi} = 1 - \hat{\beta}_2'$ (ϕ is the "adaptive expectations coefficient")

$\hat{\alpha} = \hat{\beta}_0' / \hat{\phi}$

$\hat{\beta} = \hat{\beta}_1' / \hat{\phi}$ (β_1' and β are the SR and LR propensities)

Note: Problems of Koyck-transformed model estimation apply.

Koyck, PAM, AEH: regression of y_t on x_t and y_{t-1}

The same underlying regression model (statistical form) is used:

- The Koyck transformation:
  $y_t = \alpha_0 + \gamma x_t + \rho y_{t-1} + v_t$
- The Partial adjustment model (PAM):
  $y_t = \alpha\theta + \beta\theta x_t + (1-\theta)y_{t-1} + \theta u_t$
- The Model with adaptive expectations (AEM):
  $y_t = \alpha\phi + \beta\phi x_t + (1-\phi)y_{t-1} + v_t$

We can make three different interpretations from one estimated equation.

Of course, not all interpretations are always relevant, we must choose according to application and test assumptions.

Koyck, PAM, AEH: regression of y_t on x_t and y_{t-1}

Example: Model for $c_t \leftarrow f(x_t, x_{t-1}, x_{t-2}, \dots)$:
private consumption (c_t) as a function of disposable income (x_t).
Same estimated model for Koyck, PAM, AEH:

$$\begin{array}{rrrrr} & \hat{c}_t = & 1,038 + & 0.404x_t + & 0.501c_{t-1} \\ \textbf{Koyck:} & \hat{c}_t = & \hat{\alpha}_0 + & \hat{\gamma}x_t + & \hat{\rho}c_{t-1} \end{array}$$

Koyck: IDL, geometric decay in δ parameters assumed:

- $\hat{\rho} = 0.501$
- $\hat{\alpha}_0 = 1,038 = \hat{\alpha}(1-\hat{\rho}) = \hat{\alpha}(1-0.501) \Rightarrow \hat{\alpha} = \frac{1,038}{0.499} = 2,080$
- $\hat{\delta}_j = \hat{\gamma}\hat{\rho}^j = 0.404 \times 0.501^j$
- $LRP = \frac{\hat{\gamma}}{1-\hat{\rho}} = \frac{0.404}{0.499} \doteq 0.81$
- IDL: $\hat{c}_t = 2,080 + \underbrace{0.404}_{\hat{\gamma}\hat{\rho}^0} x_t + \underbrace{0.202}_{\hat{\gamma}\hat{\rho}} x_{t-1} + \underbrace{0.101}_{\hat{\gamma}\hat{\rho}^2} x_{t-2} + \dots$

Koyck, PAM, AEH: regression of y_t on x_t and y_{t-1}

Example continued:

$$
\begin{aligned}
\hat{c}_t &= \quad 1,038 + \quad 0.404 x_t + \quad 0.501 c_{t-1} \\
\textbf{PAM: } \hat{c}_t &= \quad \hat{\alpha}\hat{\theta} + \quad \hat{\beta}\hat{\theta} x_t + \quad (1-\hat{\theta}) c_{t-1}
\end{aligned}
$$

- $(1-\hat{\theta}) = 0.501 \Rightarrow \hat{\theta} = 0.499$
- $\hat{\alpha}\hat{\theta} = 1,038 \Rightarrow \hat{\alpha} = \frac{1{,}038}{0.499} = 2,080$
- $\hat{\beta}\hat{\theta} = \ 0.404 \Rightarrow \hat{\beta} = \frac{0.404}{0.499} \doteq 0.81$

- PAM: $\hat{c}_t^* = 2,080 + 0.81 x_t$

  $c_t - c_{t-1} = 0.499 \cdot (c_t^* - c_{t-1})$

- If c_t has a prominent inertia and Δc_t significantly follows changes in habits, we might use the PAM approach.

Koyck, PAM, AEH: regression of y_t on x_t and y_{t-1}

Example continued:

$$\begin{aligned} \hat{c}_t &= \quad 1,038 + \quad 0.404 x_t + \quad 0.501 c_{t-1} \\ \textbf{AEH:}\ \hat{c}_t &= \quad \hat{\alpha}\hat{\phi} + \quad \hat{\beta}\hat{\phi} x_t + (1-\hat{\phi}) c_{t-1} \end{aligned}$$

- $(1-\hat{\phi}) = 0.501 \Rightarrow \hat{\phi} = 0.499$
- $\hat{\alpha}\hat{\phi} = 1,038 \Rightarrow \hat{\alpha} = \frac{1,038}{0.499} = 2,080$
- $\hat{\beta}\hat{\phi} = \ 0.404 \Rightarrow \hat{\beta} = \frac{0.404}{0.499} \doteq 0.81$

- AEH: $\hat{c}_t = 2,080 + 0.81 x_t^*$

  $x_t^* = 0.499 x_t + 0.501 x_{t-1}^*$

- If c_t if formed as a function of expected (e.g. permanent) income, we might prefer AEH.

Rational expectations

- Rational expectations

$$\mathbf{E}_{t-1}(x_t) = a_0 + a_1 x_{t-1} + b_1 z_{1,t-1} + b_2 z_{2,t-2} + \dots$$

$\mathbf{E}_{t-1}(x_t)$: expected value of x_t at time $t-1$
$z_{k,t-j}$: exogenous variables with impact on $\mathbf{E}_{t-1}(x_t)$

We put $x_t^* = \mathbf{E}_{t-1}(x_t)$ into (26): $y_t = \alpha + \beta x_t^* + u_t$

We assume that agents:

- know all relevant information
- know how to use this information

Agents can make prediction errors (v_t), so:

$$x_t = x_t^* + v_t$$

Rational vs. adaptive expectations

Under rational expectations:

- Expected value of prediction errors $[v_t = x_t - x_t^*]$ must be zero. If they were systematically different from zero, rational agents would immediately adjust their forecasting methods accordingly.
- $[v_t = x_t - x_t^*]$ prediction error must be uncorrelated with any information available when the prediction is made $(t-1)$. If not, this would imply that the forecaster has not made use of all available information.
- These properties can be used for testing the rational expectations hypothesis in different applications (e.g. through ex-post simulated dynamic forecasts).

Some economic application that use expectations

- Philips curve (Expectations-augmented Phillips curve)
- Efficient market hypothesis (EMH)
- Consumption function - Permanent income hypothesis