FORECASTING FOR BUSINESS & ECONOMICS EBC2089

Prof. dr. Alain Hecq
A3.14
a.hecq@maastrichtuniversity.nl

March 2023

Main features of a univariate time series

- Let 's take the following intuitive decomposition for a univariate time series variable y_t (income or GDP, inflation, stock prices, unemployment,...)

$$T(y_t) = f(\text{deterministic}) + g(\text{past of the series}) + \text{noise (unforecastable)}$$

- ISSUES:
  - What 's T, the transformation we should apply to the data (logs, first differences, growth rates,...)
  - What is the "deterministic" part: trend, seasonality, outliers,...?
  - What 's the forecastable past: AR, MA,...i.e. the past
  - What 's f and what 's g (linear, non linear).
- Remark: we can add regressors, i.e. explanatory variables.

Main features of a time series

- Say differently:

$$\textit{Series} \text{ or its transformation} = \textit{trend} + \textit{seasons} + \textit{cycle} + \textit{noise}$$

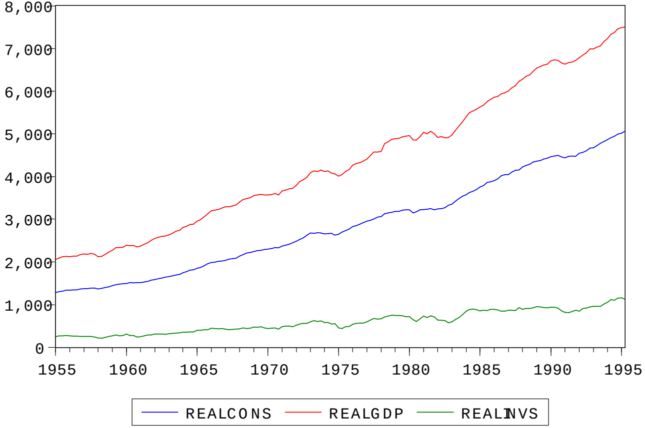
or more exactly

$$\textit{Series} \text{ or its transformation} = \textit{trend} \ \& \ \textit{seasons} \ \& \ \textit{cycle} \ \& \ \textit{noise}$$

- We must identify these components to try to forecast and to analyze the series correctly.

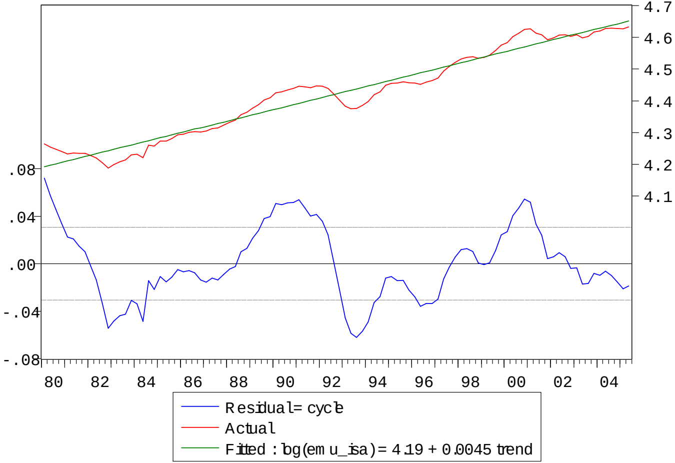
Features: trends in macro series

in levels

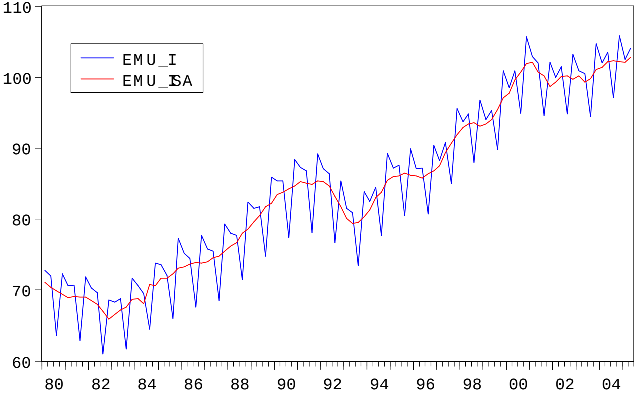


1.*pdf*

Features: cycles in EU detrended industrial production



Features: raw and SA series for EU industrial production



How to best fit/find these components?

- To identify those components of y_t, we'll try to "discover" how the data are actually generated,
- This is the so called **data generating process** of the series (**DGP**)
- To do so our tools are
  - Test statistics (e.g., t-test on significance of variables), information criteria, $\bar{R}^2$.
  - **Graphs** (series, residuals, ACF, etc)
  - Misspecification tests (e.g., normality, breaks, autocorrelation)

Model checking and misspecifications

- Check whether the residuals do not contain obvious information we can exploit
    - Test for no autocorrelation, normality, no heteroscedasticity.....
- We already did normality, homoskedasticity and linearity

Mispecification 4: autocorrelation

- Only for time series, this is why we did not see it before in cross section.
- An hypothesis underlying the "classical linear regression model" is violated.

$$\begin{aligned} y_t &= \alpha + \beta_1 x_{1t} + ... + \beta_k x_{kt} + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t \end{aligned}$$

- Typically (always the case) in time series

Autocorrelation

- What are the consequences for our OLS estimator?
  - Unbiased and consistent if X is non stochastic (although biased and non consistent if y_{t-i} on the right hand side)
  - Inefficient $\Rightarrow$ bad inference and so don't use t tests (in usual output)
- How to detect the presence of autocorrelation ?
  - look at graphs: (i) residuals against time allows to discriminate between positive and negative autocorrelation, (ii) residuals versus past value residuals.
  - traditional tests (DW).
  - LM tests for autocorrelation, BP tests.

Autocorrelation

- What is the cure?
  - Traditional answer: GLS (e.g. Cochrane-Orcutt, but please don't abuse)
  - Add new variables with dynamics in y and X, namely improve the model
  - Robust standard errors (HAC and not HCSE).
  - Use row data because $y^{SA} = \Psi(L)y$

Autocorrelation

- The Durbin Watson test

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

where e_t are the residuals from the OLS regression $y = Xb + e$

Durbin-Watson test

$$
\begin{aligned}
DW &= \frac{\sum_{t=2}^{T} e_t^2}{\sum_{t=1}^{T} e_t^2} + \frac{\sum_{t=2}^{T} e_{t-1}^2}{\sum_{t=1}^{T} e_t^2} - 2\frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} \\
&= 1 - \frac{e_1^2}{\sum_{t=1}^{T} e_t^2} + 1 - \frac{e_T^2}{\sum_{t=1}^{T} e_t^2} - 2r \\
&= 2(1-r) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2}
\end{aligned}
$$

Durbin-Watson test

If T is large

$$DW \approx 2(1 - r)$$

Interpretation $(r = \hat{\rho})$:

- if $r = 0$, no autocorrelation and $DW = 2$
- if $r = 1$, strong positive autocorrelation and $DW = 0$
- if $r = -1$, strong negative autocorrelation and $DW = 4$

Durbin-Watson test

- DW test (look at critical values in the tables for)

$$H_0 : \rho = 0 \quad \text{versus} \quad H_0 : \rho > 0$$

or

$$H_0 : \rho = 0 \quad \text{versus} \quad H_0 : \rho < 0$$

Simple test but:

Durbin-Watson test: drawbacks

- Only to detect autocorrelation of order 1
- Bound test, so inconclusive between dL and dU
- DW converges to 2 when lagged dependent variable present in the model

Other tests: use them instead: Breusch-Godfrey (LM test)

- `View -> residuals diagnostic -> LM test`
-

$$\begin{aligned} H_0 &: \quad \text{no autocorrelation} \\ H_1 &: \quad \varepsilon_t \text{ is a dynamic process} \end{aligned}$$

Estimate by OLS

$$y_t = X_t'\beta + \varepsilon_t$$

Take the residuals e_t and compute the auxiliary regression

$$e_t = X_t'\gamma + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \ldots + \alpha_p e_{t-p} + \eta_t$$

F-test on $\alpha_1 = \alpha_2 = ... = \alpha_p = 0$

Box-Pierce

`View -> residuals diagnostic -> Q test`

$$Q = T\sum_{j=1}^{p} r_j^2, \text{ with } r_j = \frac{\sum\limits_{t=j+1}^{T} e_t e_{t-j}}{\sum\limits_{t=1}^{T} e_t^2}$$

$Q \sim \chi^2(p)$ under the null hypothesis

Ljung and Box (better then Box Pierce in small sample)

$$Q = T(T+2)\sum_{j=1}^{p}\frac{r_j^2}{T-j} \sim \chi^2(p)$$

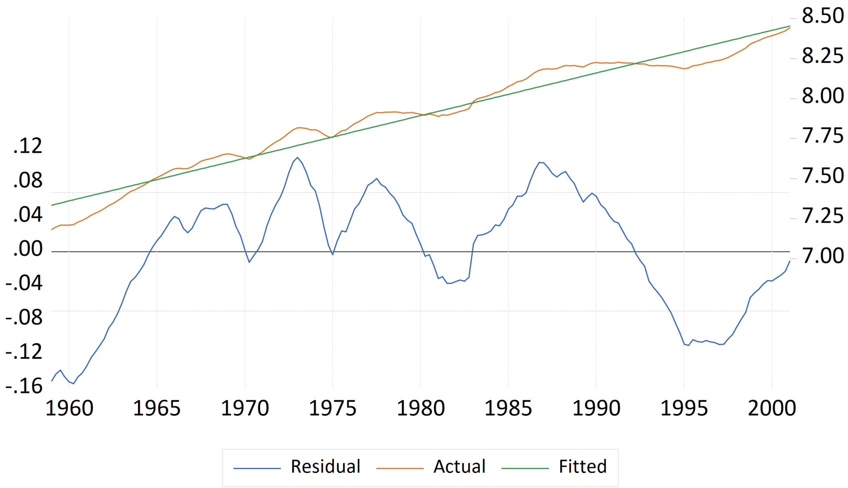
The latter three tests can be used in the presence of lagged dependent variables and to test more then 1 autoregressive coefficient.

Example: real money

Dependent Variable: LOG(M2/PRICE)
Method: Least Squares
Date: 02/19/21 Time: 08:58
Sample: 1959Q1 2001Q1
Included observations: 169

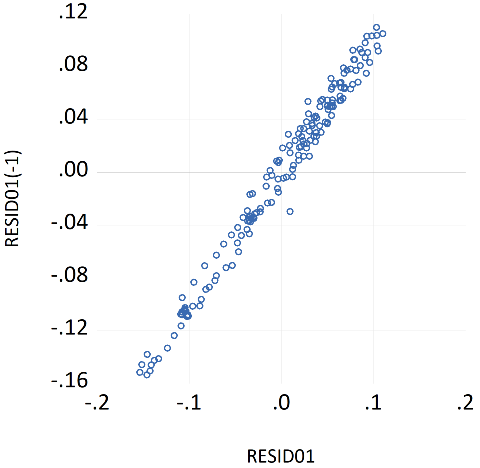
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.334988	0.010589	692.6902	0.0000
@TREND	0.006666	0.000109	61.15003	0.0000

R-squared	0.957249	Mean dependent var	7.894929
Adjusted R-squared	0.956993	S.D. dependent var	0.333371
S.E. of regression	0.069135	Akaike info criterion	-2.493746
Sum squared resid	0.798202	Schwarz criterion	-2.456706
Log likelihood	212.7216	Hannan-Quinn criter.	-2.478715
F-statistic	3739.327	Durbin-Watson stat	0.018608
Prob(F-statistic)	0.000000		


8.50
8.25
8.00
7.75
7.50
7.25
7.00
.12
.08
.04
.00
-.04
-.08
-.12
-.16
1960
1965
1970
1975
1980
1985
1990
1995
2000
Residual
Actual
Fitted

`Proc -> make residual series`

then scatter plot between residuals and residuals laggged



Univariate time series with a deterministic trend

- y_t is a single series on grossdomestic product, unemployment, advertisement expenditures,...
- it is possible to estimate different trend models to capture the nonstationarity feature of the data:

$$
\begin{aligned}
(i)\ y_t &= \alpha + \beta.trend + \varepsilon_t \\
(ii)\ y_t &= \alpha + \beta.trend + \gamma.trend^2 + \varepsilon_t \\
\\
(iii)\ \ln(y_t) &= \ln(\alpha) + \beta.trend + \varepsilon_t \\
& \quad or \\
(iv)\ y_t &= \alpha \exp^{\beta.trend_t} \varepsilon_t \\
&\neq (v)\ y_t = \alpha \exp^{\beta.trend} + \varepsilon_t
\end{aligned}
$$

- Next week we test whether the trend is deteministic or stochastic (unit root tests).

- (i) for a linear trend (by OLS)
- (ii) and (iii) for a non-linear trend by OLS.
- (v) by NLS for comparing R^2 with (ii)

Structural Breaks (graphs)

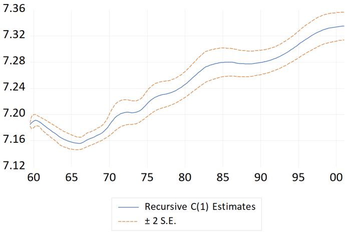
- Recursive coefficients in EViews to have a first flavor
- Consider estimating coefficients $\hat{\beta}$ on an increased sample

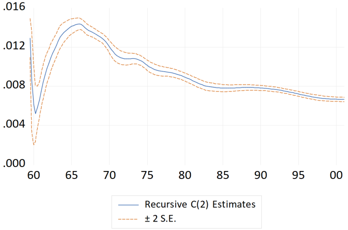
$$
\begin{aligned}
t &= \overbrace{1,2,...T_{10}}^{\text{estimation period}}, t = \overbrace{1,2,...T_{11},}^{\text{estimation period}} t = \overbrace{1,2,...T_{12},}^{\text{estimation period}} \\
.......t &= \overbrace{1,2,...T_{200},}^{\text{estimation period}}
\end{aligned}
$$

- Look whether estimated coefficients $\hat{\beta}$ "move" or are rather constant through time.

Structural Breaks (graphs)

`View -> stability diagnostic -> recursive residuals`





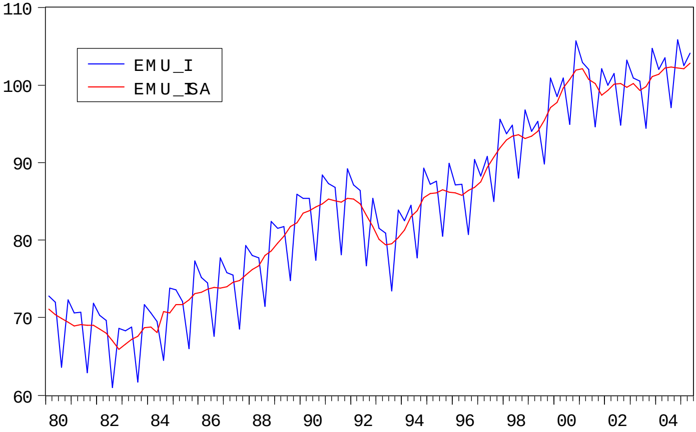
Structural Break tests

- Chow tests (F-tests)
    - 1 break/known break date
    - >1 breaks/known break dates
    - 1 break/unknown break date
    - >1 breaks/unknown break dates

Modelling deterministic seasonality

- Series are observed at a higher frequency than annual: quarterly, monthly, weekly, daily, intra-daily...
- Positive point:
  - in general more observations give more accurate estimation and forecasts.
  - Seasonality is observed. This means that movements that are repeated (reproduced) every year on the same quarter, month, week,...can be used to forecast series.
- Negative point: more work.

Difference between raw and SA series for EU industrial production



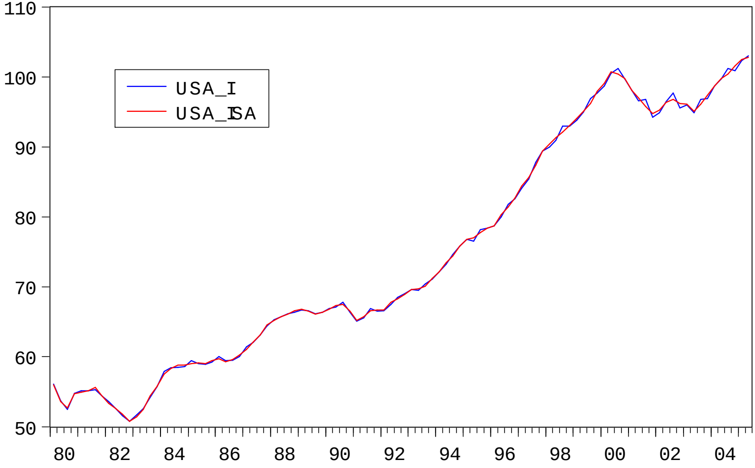
Modelling deterministic seasonality

- Often the case on industrial production indexes, prices, unemployment, stock of money,...and most of the macro-economic time series.
- Less seasonal movements in some financial data such as interest rates.
- Intra-day effects in security or bond prices.
- But stock prices have seasonal daily effect: Monday effect.

Modelling deterministic seasonality

- Observations:
  - Some macro series do not seem to display seasonal movements. It could be that there are no seasonality but it might also be that those movements have already been "deleted", filtered => seasonally adjusted (SA) series. (methods: moving averages, X12)
  - They are not raw data.
  - This could be a problem for inference.
  - This could be a problem for economic interpretation and forecasting. Example: tourism application.
  - Most of the series for the US are SA.

US industrial production: Raw and SA



Modelling deterministic seasonality

- Like many others I also favor modelling seasonality instead of removing it.
- Some seasonal movements are so strong that they visually hide (dominate) a trend (example Italy)
- This could be a problem for economic analysis, I mean about the blabla economists need to produce => could take annual growth rate of NSA

$$\frac{y_t - y_{t-4}}{y_{t-4}}$$

  if you main concern is to comment a graph...
- For modelling we must of course take this additional feature into account.
- What's the best way to modelling seasonal movements: Stochastic (seasonal differences) and or deterministic (seasonal dummies) or a part of it.

Modelling deterministic seasonality (second part of Chapter 5 in Diebold)

- In the presence of deterministic seasonality we consider $(\Delta y_t)^c$ where $(\Delta y_t)^c$ is the residual series from Δy_t on a set of deterministic quarterly dummies D_t with

$$D_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Be careful with the dummy trap!

Dummies in Eviews

- @quarter(1)=1, @quarter(2)=1,...@quarter(4)=1
- @quarter =1, @quarter =2, ...@quarter =4
- @month(1)=1, ...@month(12)=1, ...
- **@expand(@quarter, @drop(1))**

Outliers

- Outliers detection: a lot of tests and procedures
- For instance, extreme values can be detected thanks to the box plot. Very aberrant values are the observations outside the following interval:

$$\begin{aligned} Q_{25} - 3 \times IR &< \text{far outside outliers} < Q_{75} + 3 \times IR \\ IR &= Q_{75} - Q_{25}, \text{ interquatile range} \end{aligned}$$

where Q_{25} and Q_{75} are respectively the 25% and 75% quantiles, i.e. quartiles. These are calculated such as in an ascendant sorted series

$$\begin{aligned} Median &= \text{value of } \frac{T}{2} \\ Q_{25} &= \text{value of } \frac{T}{4} \\ Q_{75} &= \text{value of } \frac{3T}{4} \end{aligned}$$