Time Series Cheat Sheet

By Marcelo Moreno - King Juan Carlos University
As part of the Econometrics Cheat Sheet Project

THIS IS A WORK IN PROGRESS

NOT INTENDEND FOR GENEAL PURPOSE

Basic concepts

Definitions

Time series - is a succession of quantitative observations of a phenomena ordered in time.

There are some variations of time series:

- **Panel data** - consist of a time series for each observation of a cross section.
- **Pooled cross sections** - combines cross sections from different time periods.

Stochastic process - sequence of random variables that are indexed in time.

Components of a time series

- **Trend** - is the long-term general movement of a series.
- **Seasonal variations** - are periodic oscillations that are produced in a period equal or inferior to a year, and can be easily identified on different years (usually are the result of climatology reasons).
- **Cyclical variations** - are periodic oscillations that are produced in a period greater than a year (are the result of the economic cycle).
- **Residual variations** - are movements that do not follow a recognizable periodic oscillation (are the result of eventual non-permanent phenomena that can affect the studied variable in a given moment).

Type of time series models

- **Static models** - the relation between y and x's is contemporary. Conceptually:
$$y_t = \beta_0 + \beta_1 x_t + u_t$$
- **Distributed-lag models** - the relation between y and x's is not contemporary. Conceptually:
$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + ... + \beta_{q+1} x_{t-q} + u_t$$
The long term cumulative effect in y when Δx is:
$$\beta_1 + \beta_2 + ... + \beta_{q+1}$$

Assumptions and properties

OLS model assumptions under time series

Under this assumptions, the estimators of the OLS parameters will present good properties. **Gauss-Markov assumptions extended applied to time series**:

ts1. **Parameters linearity and weak dependence**.
   a. y_t must be a linear function of the β's.
   b. The stochastic $\{(x_t, y_t) : t = 1, 2, ..., T\}$ is stationary and weakly dependent.

ts2. **No perfect collinearity**.
   - There are no independent variables that are constant: $Var(x_j) \neq 0$
   - There is not an exact linear relation between independent variables.

ts3. **Conditional mean zero and correlation zero**.
   a. There are no systematic errors: $E(u_t|x_{1t}, ..., x_{kt}) = E(u_t) = 0 \rightarrow$ **strong exogeneity** (a implies b).
   b. There are no relevant variables left out of the model: $Cov(x_{jt}, u_t) = 0$ for any $j = 1, ..., k \rightarrow$ **weak exogeneity**.

ts4. **Homoscedasticity**. The variability of the residuals is the same for any t: $Var(u_t|x_{1t}, ..., x_{kt}) = \sigma^2$

ts5. **No auto-correlation**. The residuals do not contain information about other residuals: $Corr(u_t, u_s|x) = 0$ for any given $t \neq s$.

ts6. **Normality**. The residuals are independent and identically distributed: $u \sim N(0, \sigma^2)$

ts7. **Data size**. The number of observations available must be greater than $(k+1)$ parameters to estimate. (IS already satisfied under asymptotic situations)

Asymptotic properties of OLS

Under the econometric model assumptions and the Central Limit Theorem:

- Hold (1) to (3a): OLS is **unbiased**. $E(\hat{\beta}_j) = \beta_j$
- Hold (1) to (3): OLS is **consistent**. $plim(\hat{\beta}_j) = \beta_j$ (to (3b) left out (3a), weak exogeneity, biased but consistent)
- Hold (1) to (5): **asymptotic normality** of OLS (then, (6) is necessarily satisfied): $u \sim_a N(0, \sigma^2)$.
- Hold (1) to (5): **unbiased estimate of** σ^2. $E(\hat{\sigma}^2) = \sigma^2$
- Hold (1) to (5): OLS is BLUE (Best Linear Unbiased Estimator) or **efficient**.
- Hold (1) to (6): hypothesis testing and confidence intervals can be done reliably.

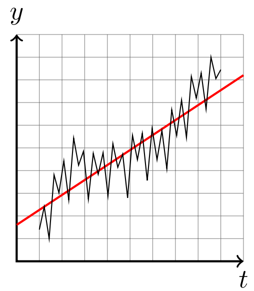
Trends and seasonality

Trends

Two time series can have the same (or contrary) trend, that should lend to a high level of correlation. This, can provoke a false appearance of causality, the problem is known as **spurious regression**. For example, given the model:
$$y_t = \beta_0 + \beta_1 x_t + u_t$$
where:
$$y_t = \alpha_0 + \alpha_1 Trend + e_t$$
$$x_t = \gamma_0 + \gamma_1 Trend + v_t$$
Adding a trend to the model can solve the problem:
$$y_t = \beta_0 + \beta_1 x_t + \beta_2 Trend + u_t$$
The trend can be linear or non-linear (quadratic, cubic, etc.)

Seasonallity

A time series with a frequency higher than a year (CORREGIR), like, quarterly, monthly, weekly, ..., can manifest seasonallity. That is, the series are subject to a seasonal pattern. For ex: the GDP is usually higher in summer and



lower in winter.

Solution: seasonal adjustment.

Solution 1: stational binary variables. For example, quarterly series:

$y_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + \beta_4 x_{1t} + ... + \beta_k x_{kt} + u_t$

Or $y_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + u_t$, and then $newy_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_k x_{kt} + u_t$

Stationarity (estacionariedad)

Strong exogeneity is violated in static models and lag models (modelos de rezagos distribuidos).

Series de tiempo estacionarias y no estacionarias

- Stationary process - is the one in that the probability distribution are stable in time. If any collection of random variables are taken, and are shifted h periods. The joint probability distribution should be inaltered.
- Non stationary process - for example, a series with trend, at least the mean change with the time (at least). IF THE MEAN IS CONSTANT, STRONG STATIONARITY.
  IF ONLY THE COVARIANCE IS CONSTANT, $Cov(x_t, x_{t+h}$

Dummy variables and structural change in time series

Dummy (or binary) variables are used for qualitative information like sex, civil state, country, etc.

- Get the **value of 1 in a given category, and 0 on the rest**.
- Are used to analyze and modeling **structural changes** in the model parameters.

If a qualitative variable have m categories, we only have to include $(m-1)$ dummy variables.

Structural change

Structural change refers to changes in the values of the parameters of the econometric model produced by the effect of different sub-populations. Structural change can be included in the model through dummy variables.
The position of the dummy variable matters:

- **On the intercept** (β_0) - represents the mean difference between the values produced by the structural change.
- **On the parameters that determines the slope of the regression line** (β_j) - represents the effect (slope) difference between the values produced by the structural change.

The Chow's structural contrast - when we want to analyze the existence of structural changes in all the model parameters, it is common to use a particular expression of the F contrast known as the Chow's contrast, where the null hypothesis is: H_0 : No structural change

Predictions

Two types of prediction:

- Of the mean value of y for a specific value of x.
- Of an individual value of y for a specific value of x.

If the values of the variables (x) approximate to the mean values ($\overline{x}$), the confidence interval amplitude of the prediction will be shorter.

Heteroscedasticity in time series

The residuals u_i of the population regression function do not have the same variance σ^2:

$$Var(u|x) = Var(y|x) \neq \sigma^2$$

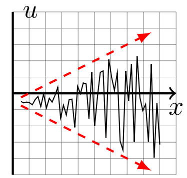
Is the **breaking of the fifth (5) econometric model assumption**.

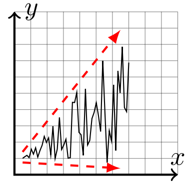
Consequences

- OLS estimators still are unbiased.
- OLS estimators still are consistent.
- OLS is **not efficient** anymore, but still a LUE (Linear Unbiased Estimator).
- **Variance estimations of the estimators are biased**: the construction of confidence intervals and the hypothesis contrast are not reliable.

Detection

- **Graphical analysis** - look for scatter patterns on x vs. u or x vs. y plots.





- **Formal tests** - White, Bartlett, Breusch-Pagan, etc. Commonly, the null hypothesis: H_0 = Homoscedasticity

DURBIN WATSON EXPLANATION AND GRAPHICAL ANALYSIS (AND MOST COMMON VALUES?)

Correction

- Use OLS with a variance-covariance matrix estimator robust to heteroscedasticity, for example, the one proposed by White.
- If the variance structure is known, make use of Weighted Least Squares (WLS) or Generalized Least Squares (GLS).
- If the variance structure is not known, make use of Feasible Weighted Least Squared (FWLS), that estimates a possible variance, divides the model variables by it and then apply OLS.
- Make assumptions about the possible variance:
  - Supposing that σ_i^2 is proportional to x_i, divide the model variables by the square root of x_i and apply OLS.
  - Supposing that σ_i^2 is proportional to x_i^2, divide the model variables by x_i and apply OLS.

- Make a new model specification, for example, logarithmic transformation.

Auto-correlation

The residual of any observation, u_t, is correlated with the residual of any other observation. The observations are not independent.

$$Corr(u_t, u_s|x) \neq 0 \text{ for any } t \neq s$$

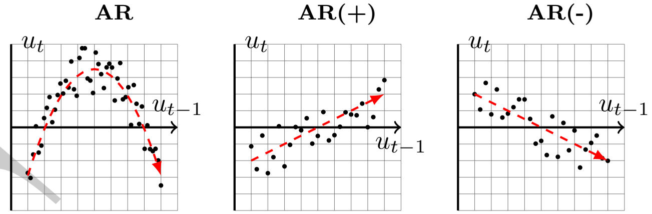
The "natural" context of this phenomena is time series. Is the **breaking of the sixth (6) econometric model assumption**.

Consequences

- OLS estimators still are unbiased.
- OLS estimators still are consistent.
- OLS is **not efficient** anymore, but still a LUE (Linear Unbiased Estimator).
- **Variance estimations of the estimators are biased**: the construction of confidence intervals and the hypothesis contrast are not reliable.

Detection

- **Graphical analysis** - look for scatter patterns on u_{t-1} vs. u_t or make use of a correlogram.



- **Formal tests** - Durbin-Watson, Breusch-Godfrey, etc. Commonly, the null hypothesis: H_0 : No auto-correlation

Correction

- Use OLS with a variance-covariance matrix estimator robust to auto-correlation, for example, the one proposed by Newey-West.
- Use Generalized Least Squares. Supposing $y_t = \beta_0 + \beta_1 x_t + u_t$, with $u_t = \rho u_{t-1} + \varepsilon_t$, where $|\rho| < 1$ and ε_t is white noise.
  - If ρ is known, create a quasi-differentiated model where u_t is white noise and estimate it by OLS.
  - If ρ is not known, estimate it by -for example- the Cochrane-Orcutt method, create a quasi-differentiated model where u_t is white noise and estimate it by OLS.

Endogeneity

Instrumental Variables Methods