

# Time Series Cheat Sheet

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

## 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:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \dots + \beta_{s+1} x_{t-s} + u_t$$

The long term cumulative effect in  $y$  when  $\Delta x$  is:

$$\beta_1 + \beta_2 + \dots + \beta_{s+1}$$

- **Dynamic models** - a temporal drift of the dependent variable is part of the independent variables (endogeneity). Conceptually:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_s y_{t-s} + u_t$$

- Combinations of the above, like the rational distributed-lag models (distributed-lag + dynamic).

## 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.**

- $y_t$  must be a linear function of the  $\beta$ 's.
- The stochastic  $\{(x_t, y_t) : t = 1, 2, \dots, T\}$  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.**

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

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

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 (**i.i.d.** so on):  $u \sim N(0, \sigma^2)$

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  $\sigma^2$** .  $E(\hat{\sigma}^2) = \sigma^2$
- 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

**Spurious regression** - is when the relation between  $y$  and  $x$  is due to factors that affect  $y$  and have correlation with  $x$ ,  $Corr(x, u) \neq 0$ . Is the **non-fulfillment of ts3**.

### 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 **spurious regression**. Given the model:

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

where:

$$y_t = \alpha_0 + \alpha_1 Trend + v_t$$

$$x_t = \gamma_0 + \gamma_1 Trend + v_t$$

Adding a trend to the model can solve the problem:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 Trend + u_t$$

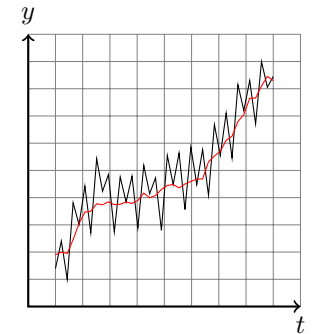
The trend can be linear or non-linear (quadratic, cubic, exponential, etc.)

Another way is making use of the **Hodrick-Prescott filter** to extract the trend (smooth) and the cyclical component.

### Seasonality

A time series with can manifest seasonality. That is, the series is subject to a seasonal variations or pattern, usually related to climatology conditions.

For example, GDP (black) is usually higher in summer and lower in winter. Seasonally adjusted series (red) for comparison.



- This problem is **spurious regression**. A seasonal adjustment can solve it.

A simple **seasonal adjustment** could be creating stationary variables and adding them to the model. For example, for quarterly series ( $QX_t$  are binary variables):

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

Another way is to seasonally adjust (sa) the variables, and then, do the regression with the adjusted variables:

$$z_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + v_t \rightarrow \hat{v}_t + E(z_t) = \hat{z}_t^{sa}$$

$$\hat{y}_t^{sa} = \beta_0 + \beta_1 \hat{x}_{1t}^{sa} + \dots + \beta_k \hat{x}_{kt}^{sa} + u_t$$

There are much better and complex methods to seasonally adjust a time series, like the **X-13ARIMA-SEATS**.

## Auto-correlation

The residual of any observation,  $u_t$ , is correlated with the residual of any other observation. The observations are not independent. Is the **non-fulfillment of ts5**.

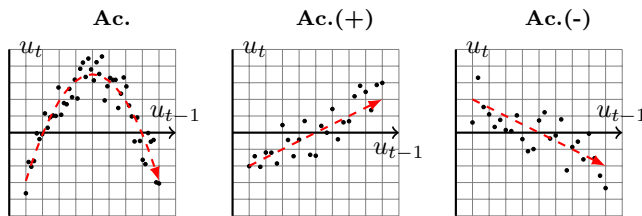
$$\text{Corr}(u_t, u_s|x) \neq 0 \text{ for any } t \neq s$$

### Consequences

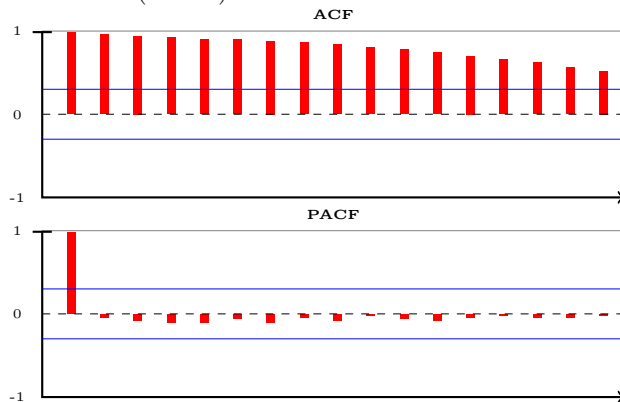
- OLS estimators still are unbiased.
- OLS estimators still are consistent.
- OLS is **not efficient** anymore, but still a LUE (Linear Unbiased Estimator).
- **Variance estimations of the estimators are biased:** the construction of confidence intervals and the hypothesis contrast are not reliable.

### Detection

- **Scatter plots** - look for scatter patterns on  $u_{t-1}$  vs.  $u_t$ .



- **Correlogram** - composed – Y axis: correlation [-1,1]. of the auto-correlation – X axis: lag number. function (ACF) and the – Blue lines:  $\pm 1.96/T^{0.5}$  partial ACF (PACF).



Conclusions differ between auto-correlation processes.

- **MA(q) process.** **ACF:** only the first  $q$  coefficients are significant. The rest are abruptly canceled. **PACF:** attenuated exponential fast decay or sine waves.
- **AR(p) process.** **ACF:** attenuated exponential fast decay or sine waves. **PACF:** only the first  $p$  coefficients are significant. The rest are abruptly canceled.
- **ARMA(p,q) process.** **ACF:** the coefficients are not abruptly canceled and presents a fast decay. **PACF:** the coefficients are not abruptly canceled and presents a fast decay.

If the ACF coefficients do not decay rapidly, there is a clear indicator of lack of stationarity in mean, which would lead to take first differences in the original series.

- **Formal tests** -  $H_0$ : No auto-correlation.

Supposing that  $u_t$  follows an AR(1) process:

$$u_t = \rho_1 u_{t-1} + \varepsilon_t$$

where  $\varepsilon_t$  is white noise.

- **AR(1) t test** (exogenous regressors):

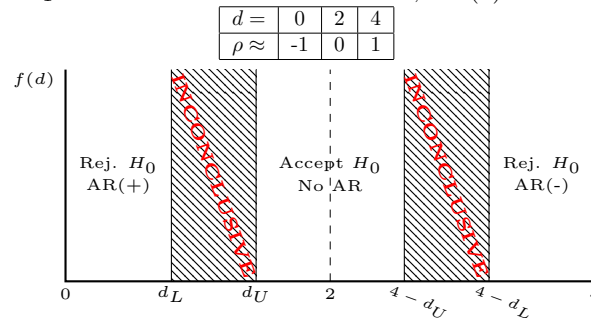
$$t = \frac{\hat{\rho}_1}{se(\hat{\rho}_1)} \sim t_{T-k-1, \alpha/2}$$

- \*  $H_1$ : Auto-correlation of order one, AR(1).

- **Durbin-Watson statistic** (exogenous regressors and residual normality):

$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2} \approx 2(1 - \hat{\rho}_1), 0 \leq d \leq 4$$

- \*  $H_1$ : Auto-correlation of order one, AR(1).



- **Durbin's h** (endogenous regressors):

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T \times v}}$$

where  $v$  is the estimated variance of the coefficient associated to the endogenous variable.

- \*  $H_1$ : Auto-correlation of order one, AR(1).

- **Breusch-Godfrey test** (endogenous regressors): it can detect MA(q) and AR(p) processes ( $\varepsilon_t$  is w. noise):

- \* MA(q):  $u_t = \varepsilon_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$

- \* AR(p):  $u_t = \rho_1 u_{t-1} + \dots + \rho_p u_{t-p} + \varepsilon_t$

Under  $H_0$ : No auto-correlation:

$$T \times R_{\hat{u}_t}^2 \sim_a \chi_q^2 \quad \text{or} \quad T \times R_{\hat{u}_t}^2 \sim_a \chi_p^2$$

- \*  $H_1$ : Auto-correlation of order  $q$  (or  $p$ ).

- **Ljung-Box Q test:**

- \*  $H_1$ : There is 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  $\varepsilon_t$  is white noise.

- If  $\rho$  is known, create a quasi-differentiated model:

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + u_t - \rho u_{t-1}$$

$$y_t^* = \beta_0^* + \beta_1 x_t^* + u_t^*$$

where  $u_t^*$  is white noise, and estimate it by OLS.

- If  $\rho$  is not known, estimate it by -for example- the **Cochrane-Orcutt method** (Prais-Winsten method is also good):

1. Obtain  $\hat{u}_t$  from the original model.

2. Estimate  $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$  and obtain  $\hat{\rho}$ .

3. Create a quasi-differentiated model:

$$y_t - \hat{\rho} y_{t-1} = \beta_0(1 - \hat{\rho}) + \beta_1(x_t - \hat{\rho} x_{t-1}) + u_t - \hat{\rho} u_{t-1}$$

$$y_t^* = \beta_0^* + \beta_1 x_t^* + u_t^*$$

where  $u_t^*$  is white noise, and estimate it by OLS.

4. Obtain  $\hat{u}_t^*$  and repeat from step 2.

5. The method finish when the estimated parameters vary very little between iterations.

- If not solved, look for high dependence in the series.

## Stationarity and weak dependence

Stationarity means stability of the joints distributions of a process as time progresses. It allows to correctly identify the relations –that stay unchange with time– between variables.

### Stationary and non-stationary processes

- **Stationary process** (strong stationarity) - is the one in that the probability distributions are stable in time: if any collection of random variables is taken, and then, shifted  $h$  periods, the joint probability distribution should stay unchanged.

- **Non-stationary process** - is, for example, a series with trend, where at least the mean changes with time.
- **Covariance stationary process** - is a weaker form of stationarity:
  - $E(x_t)$  is constant.
  - $Var(x_t)$  is constant.
  - For any  $t, h \geq 1$ , the  $Cov(x_t, x_{t+h})$  depends only of  $h$ , not of  $t$ .

## Weak dependence time series

It is important because it replaces the random sampling assumption, giving for granted the validity of the Central Limit Theorem (requires stationarity and a form of weak dependence). Weakly dependent processes are **also known as integrated of order zero, I(0)**.

- **Weak dependence** - restricts how close the relationship between  $x_t$  and  $x_{t+h}$  can be as the time distance between the series increases ( $h$ ).

An **stationary time process**  $\{x_t : t = 1, 2, \dots, T\}$  is weakly dependent when  $x_t$  and  $x_{t+h}$  are almost independent as  $h$  increases without a limit.

A **covariance stationary time process** is weakly dependent if the correlation between  $x_t$  and  $x_{t+h}$  tends to 0 fast enough when  $h \rightarrow \infty$  (they are not asymptotically correlated).

Some examples of stationary and weakly dependent time series are:

- **Moving average** -  $\{x_t\}$  is a moving average of order one  $MA(q = 1)$ :

$$x_t = e_t + \theta_1 e_{t-1}$$

where  $\{e_t : t = 0, 1, \dots, T\}$  is an *i.i.d.* sequence with zero mean and  $\sigma_e^2$  variance.

- **Auto-regressive process** -  $\{x_t\}$  is an auto-regressive process of order one  $AR(p = 1)$ :

$$x_t = \rho_1 x_{t-1} + e_t$$

where  $\{e_t : t = 0, 1, \dots, T\}$  is an *i.i.d.* sequence with zero mean and  $\sigma_e^2$  variance.

If  $|\rho_1| < 1$ , then  $\{x_t\}$  is an  $AR(1)$  stable process that is weakly dependent. It is stationary in covariance,  $Corr(x_t, x_{t-1}) = \rho_1$ .

- **ARMA process** - is a combination of the two above.  $\{x_t\}$  is an  $ARMA(p = 1, q = 1)$ :

$$x_t = e_t + \theta_1 e_{t-1} + \rho_1 x_{t-1}$$

A series with a trend cannot be stationary, but can be weakly dependent (and stationary if the series is detrended).

## Strong dependence time series

Most of the time, economics series are strong dependent (or high persistent in time). Some special cases of unit root processes,  $I(1)$ :

- **Random walk** - an  $AR(1)$  process with  $\rho_1 = 1$ .

$$y_t = y_{t-1} + e_t$$

where  $\{e_t : t = 1, 2, \dots, T\}$  is an *i.i.d.* sequence with zero mean and  $\sigma_e^2$  variance (the latter changes with time).

The process is not stationary, is persistent.

- **Random walk with a drift** - an  $AR(1)$  process with  $\rho_1 = 1$  and a constant.

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

where  $\{e_t : t = 1, 2, \dots, T\}$  is an *i.i.d.* sequence with zero mean and  $\sigma_e^2$  variance.

The process is not stationary, is persistent.

## I(1) detection

- **Augmented Dickey-Fuller (ADF) test** - where  $H_0$ : the process is unit root,  $I(1)$ .

- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test** - where  $H_0$ : the process have no unit root,  $I(0)$ .

## Transforming unit root to weak dependent

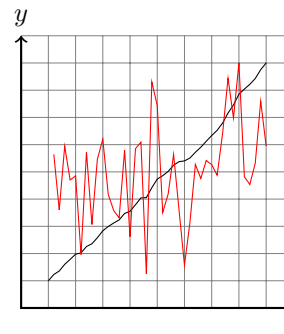
Unit root processes are **integrated of order one, I(1)**. This means that **the first difference of the process is weakly dependent** or  $I(0)$  (and usually, stationary). For example, a random walk:

$$\Delta y_t = y_t - y_{t-1} = e_t$$

where  $\{e_t\} = \{\Delta y_t\}$  is *i.i.d.*

Getting the first difference of a series also deletes its trend.

For example, a series with a trend (black), and its first difference (red).



When an  $I(1)$  series is strictly positive, it is usually converted to logarithms before taking the first difference. That is, to obtain the (approx.) percentage change of the series:

$$\Delta \log(y_t) = \log(y_t) - \log(y_{t-1}) \approx (y_t - y_{t-1})/y_{t-1}$$

## Cointegration

When **two series are I(1), but a linear combination of them is I(0)**. If the case, the regression of one series over the other is not spurious, but expresses something about the long term relation.

For example:  $\{x_t\}$  and  $\{y_t\}$  are  $I(1)$ , but  $y_t - \beta x_t = u_t$  where  $\{u_t\}$  is  $I(0)$ . ( $\beta$  get the name of cointegration parameter).

## Heterocedasticity on time series

The assumption **affected is ts4**, which leads to **OLS be not efficient**.

Some tests that work could be the Breusch-Pagan or White's, where  $H_0$ : No heterocedasticity. It is **important for the tests to work that there is no autocorrelation** (so first, it is important to test for it).

## ARCH

An auto-regressive conditional heterocedasticity (ARCH), is a model to analyze a form of dynamic heterocedasticity, where the error variance follows an  $AR(p)$  process.

Given the model:

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

where, there is  $AR(1)$  and heterocedasticity:

$$E(u_t^2 | u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$$

## GARCH

A general auto-regressive conditional heterocedasticity (GARCH), is a model similar to ARCH, but in this case, the error variance follows an  $ARMA(p, q)$  process.

## 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 ( $\bar{x}$ ), the confidence interval amplitude of the prediction will be shorter.