Time Series Cheat Sheet

By Marcelo Moreno Porras - Universidad Rey Juan Carlos
The Econometrics Cheat Sheet Project

Basic concepts

Definitions

Time series - succession of observations ordered in time with a fixed frequency.

Given the format of a time series:

- **Point-in-time (stock)** - a single value is recorded for each period.
- **Aggregated (flow)** - values represent totals or averages over the period.
- **Range/interval (OHLC)** - each period records multiple summary statistics, such as min, max, open, close.

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

Components of a time series

- **Trend** - the long-term general movement of a series.
- **Seasonal variations** - periodic oscillations that are produced in a period equal to or inferior than a year, and can be easily identified across different years (usually the result of climatology).
- **Cyclical variations** - periodic oscillations that are produced in a period greater than a year (are the result of the economic cycle).
- **Residual variations** - movements that do not follow a recognizable periodic oscillation (irregular events).

Type of time series models

- **Static models** - the relation between y and x is contemporary. Conceptually:
$$y_t = \beta_0 + \beta_1 x_t + u_t$$
- **Distributed-lag models** - the relation between y and x is not contemporary. Conceptually:
$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \cdots + \beta_s x_{t-(s-1)} + u_t$$
The long-term cumulative effect in y when Δx is:
$$\beta_1 + \beta_2 + \cdots + \beta_s$$
- **Dynamic models** - lags of the dependent variable (endogeneity). Conceptually:
$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_s y_{t-s} + u_t$$
- Combinations of the above, like the rational distributed-lag models (distributed-lag + dynamic).

Assumptions and properties

OLS model assumptions under time series

Under these assumptions, the OLS estimator will present good properties. **Gauss-Markov assumptions** extended for time series:

t1. **Parameters linearity and weak dependence**.
  a. y_t must be a linear function of the β's.
  b. The stochastic $\{(x_t, y_t) : t = 1, 2, \ldots, T\}$ is stationary and weakly dependent.

t2. **No perfect collinearity**.
  - There are no independent variables that are constant: $\mathrm{Var}(x_j) \neq 0,\ \forall j = 1, \ldots, k$
  - There is no exact linear relation between independent variables.

t3. **Conditional mean zero and correlation zero**.
  a. There are no systematic errors: $\mathrm{E}(u \mid x_1, \ldots, x_k) = \mathrm{E}(u) = 0 \rightarrow$ **strong exogeneity** (a implies b).
  b. There are no relevant variables left out of the model: $\mathrm{Cov}(x_j, u) = 0,\ \forall j = 1, \ldots, k \rightarrow$ **weak exogeneity**.

t4. **Homoscedasticity**. The variability of the residuals is the same for any x: $\mathrm{Var}(u \mid x_1, \ldots, x_k) = \sigma_u^2$

t5. **No autocorrelation**. Residuals do not contain information about any other residuals:
$\mathrm{Corr}(u_t, u_s \mid x_1, \ldots, x_k) = 0,\ \forall t \neq s$

t6. **Normality**. Residuals are independent and identically distributed (**i.i.d.**): $u \sim \mathcal{N}(0, \sigma_u^2)$

t7. **Data size**. The number of observations available must be greater than $(k+1)$ parameters to estimate. (It is already satisfied under asymptotic situations)

Asymptotic properties of OLS

Under the econometric model assumptions and the Central Limit Theorem:

- Hold t1 to t3a: OLS is **unbiased**. $\mathrm{E}(\hat{\beta}_j) = \beta_j$
- Hold t1 to t3: OLS is **consistent**. $\mathrm{plim}(\hat{\beta}_j) = \beta_j$ (to t3b left out t3a, weak exogeneity, biased but consistent)
- Hold t1 to t5: **Asymptotic normality** of OLS (then, t6 is necessarily satisfied): $u \underset{a}{\sim} \mathcal{N}(0, \sigma_u^2)$
- Hold t1 to t5: **Unbiased estimate** of σ_u^2. $\mathrm{E}(\hat{\sigma}_u^2) = \sigma_u^2$
- Hold t1 to t5: OLS is BLUE (Best Linear Unbiased Estimator) or **efficient**.
- Hold t1 to t6: 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 a correlation with x, $\mathrm{Corr}(x_j, u) \neq 0$. Is the **non-fulfillment of t3**.

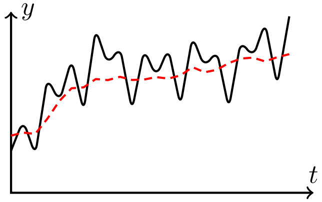
Trends

Two time series can have the same (or contrary) trend, which should lead 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:
$$y_t = \alpha_0 + \alpha_1 \mathrm{Trend} + v_t$$
$$x_t = \gamma_0 + \gamma_1 \mathrm{Trend} + v_t$$
Adding a trend to the model can solve the problem:
$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \mathrm{Trend} + u_t$$
The trend can be linear or non-linear (quadratic, cubic, exponential, etc.)

Another way is to make use of the **Hodrick-Prescott filter** to extract the trend and the cyclical component.

Seasonality

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



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

- Regressing time series that present seasonality can lead to **spurious results**.

A simple **seasonal adjustment** could be creating stationary binary variables and adding them to the model. For example, for quarterly series (Qq_t are binary variables):
$$y_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + \beta_4 x_{1t} + \cdots + \beta_k x_{kt} + u_t$$
Another way is to seasonally adjust (sa) the variables, and then, do the regression with the adjusted variables:
$$z_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + v_t \rightarrow \hat{v}_t + \mathrm{E}(z_t) = \hat{z}_t^{sa}$$
$$\hat{y}_t^{sa} = \beta_0 + \beta_1 \hat{x}_{1t}^{sa} + \cdots + \beta_k \hat{x}_{kt}^{sa} + u_t$$
There are much better and complex methods to seasonally adjust a time series, like the **X-13ARIMA-SEATS**.

Autocorrelation

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 **t5**.

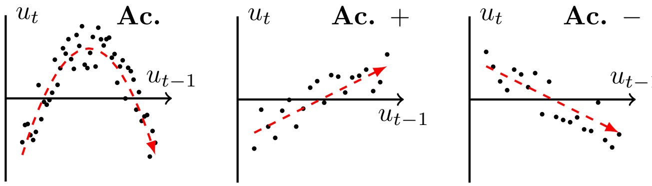
$$\text{Corr}(u_t, u_s \mid x_1, \dots, x_k) = \text{Corr}(u_t, u_s) \neq 0, \; \forall t \neq s$$

Consequences

- OLS estimators are still unbiased.
- OLS estimators are still 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 testing is not reliable.

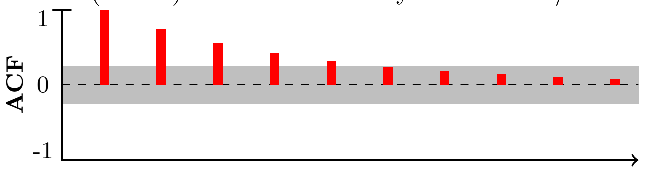
Detection

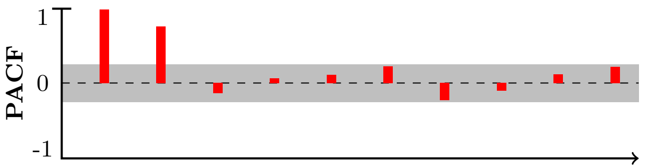
Scatter plots - look for scatter patterns on u_{t-1} vs. u_t.



Correlogram - autocorrelation function (ACF) and partial ACF (PACF).

- Y axis: correlation.
- X axis: lag number.
- Grey area: $\pm 1.96/T^{0.5}$





- **MA**(q) **process**. ACF: only the first q coefficients are significant, the remaining 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 remaining are abruptly canceled.
- **ARMA**(p,q) **process**. ACF and PACF: the coefficients are not abruptly canceled and present a fast decay.

If the ACF coefficients do not decay rapidly, there is a clear indicator of a lack of stationarity in mean.

Formal tests - Generally, H_0: No autocorrelation.

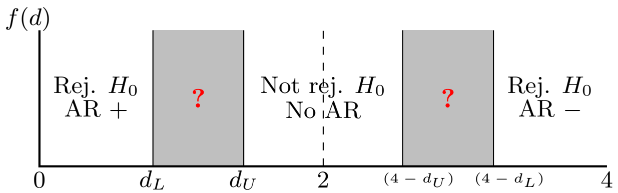
Supposing that u_t follows an AR(1) process:

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

where ε_t is white noise.

- **AR(1) t test** (exogenous regressors):
  $$t = \frac{\hat{\rho}_1}{\text{se}(\hat{\rho}_1)} \sim t_{T-k-1, \alpha/2}$$
  H_1: Autocorrelation of order one, AR(1).
- **Durbin-Watson statistic** (exogenous regressors and residual normality):
  $$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \approx 2 \cdot (1 - \hat{\rho}_1)$$
  Where $0 \leq d \leq 4$
  H_1: Autocorrelation of order one, AR(1).

$d =$	0	2	4
$\rho \approx$	1	0	-1



- **Durbin's h** (endogenous regressors):
  $$h = \hat{\rho} \cdot \sqrt{\frac{T}{1 - T \cdot v}}$$
  where v is the estimated variance of the coefficient associated with the endogenous variable.
  H_1: Autocorrelation of order one, AR(1).
- **Breusch-Godfrey test** (endogenous regressors): it can detect MA(q) and AR(p) processes (ε_t is w. noise):
  - MA(q): $u_t = \varepsilon_t - m_1 u_{t-1} - \cdots - m_q u_{t-q}$
  - AR(p): $u_t = \rho_1 u_{t-1} + \cdots + \rho_p u_{t-p} + \varepsilon_t$

  Under H_0: No autocorrelation:
  $$T \cdot R^2_{\hat{u}_t} \underset{a}{\sim} \chi^2_q \qquad \text{or} \qquad T \cdot R^2_{\hat{u}_t} \underset{a}{\sim} \chi^2_p$$
  H_1: Autocorrelation of order q (or p).
- **Ljung-Box Q test**:
  H_1: Autocorrelation up to lag h.

Correction

- Use OLS with a variance-covariance matrix estimator that is **robust to heteroscedasticity and autocorrelation** (HAC), for example, the one proposed by **Newey-West**.
- Use **Generalized Least Squares** (GLS). 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**, use a **quasi-differentiated model**:
    $$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + u_t - \rho u_{t-1}$$
    $$y_t^* = \beta_0^* + \beta_1' x_t^* + \varepsilon_t$$
    where $\beta_1' = \beta_1$; and estimate it by OLS.
  - If ρ is **not known**, estimate it by -for example- the **Cochrane-Orcutt iterative method** (Prais-Winsten's 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:
       $$y_t - \hat{\rho} y_{t-1} = \beta_0(1-\hat{\rho}) + \beta_1(x_t - \hat{\rho} x_{t-1}) + u_t - \hat{\rho} u_{t-1}$$
       $$y_t^* = \beta_0^* + \beta_1' x_t^* + \varepsilon_t$$
       where $\beta_1' = \beta_1$; and estimate it by OLS.
    4. Obtain $\hat{u}_t^* = y_t - (\hat{\beta}_0^* + \hat{\beta}_1' x_t) \neq y_t - (\hat{\beta}_0^* + \hat{\beta}_1' x_t^*)$.
    5. Repeat from step 2. The algorithm ends when the estimated parameters vary very little between iterations.
- If not solved, look for **high dependence** in the series.

Exponential smoothing

Given $\{y_t\}$, the smoothed series $\{f_t\}$:

$$f_t = \alpha y_t + (1-\alpha) f_{t-1}$$

where $0 < \alpha < 1$ is the smoothing factor and $f_0 = y_0$.

Predictions

Two types of predictions:

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

Theil's U statistic - compares the forecast results with the ones of forecasting with minimal historical data.

$$U = \sqrt{\frac{\sum_{t=1}^{T-1} \left(\frac{\hat{y}_{t+1} - y_{t+1}}{y_t} \right)^2}{\sum_{t=1}^{T-1} \left(\frac{y_{t+1} - y_t}{y_t} \right)^2}}$$

- < 1: The forecast is better than guessing.
- $= 1$: The forecast is about as good as guessing.
- > 1: The forecast is worse than guessing.

Stationarity

Stationarity allows to correctly identify relations –that stay unchanged with time– between variables.

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

Weak dependence

Weak dependence replaces the random sampling assumption for time series.

- An stationary process $\{x_t\}$ is **weakly dependent** when x_t and x_{t+h} are almost independent as h increases without a limit.
- A covariance stationary process is **weakly dependent** if the correlation between x_t and x_{t+h} tends to 0 fast enough when $h \to \infty$ (they are not asymptotically correlated).

Weakly dependent processes are known as **integrated of order zero**, I(0). Some examples:

- **Moving average** - $\{x_t\}$ is a moving average of order q, MA(q):
  $$x_t = e_t + m_1 e_{t-1} + \cdots + m_q e_{t-q}$$
  where $\{e_t : t = 0, 1, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.
- **Autoregressive process** - $\{x_t\}$ is an autoregressive process of order p, AR(p):
  $$x_t = \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + e_t$$
  where $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.
  Stability condition: if $1 - \rho_1 z - \cdots - \rho_p z^p = 0$ for $|z| > 1$ then $\{x_t\}$ is an AR(p) stable process that is weakly dependent. For AR(1), the condition is: $|\rho_1| < 1$.
- **ARMA process** - is a combination of the previous; $\{x_t\}$ is an ARMA(p, q):
  $$x_t = e_t + m_1 e_{t-1} + \cdots + m_q e_{t-q} + \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p}$$

Unit roots

A process is integrated of order d, I(d), if applying differences d times makes the process stationary.
When $d \geq 1$, the process is said to have a **unit root**. A process has a unit root when the stability condition is not met (there are roots on the unit circle).

Strong dependence

Generally, economic series are highly persistent in time. Some examples of **unit root** I(1):

- **Random walk** - an AR(1) process with $\rho_1 = 1$.
  $$y_t = y_{t-1} + e_t$$
  where $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.
- **Random walk with a drift** - an AR(1) process with $\rho_1 = 1$ and a constant.
  $$y_t = \beta_0 + y_{t-1} + e_t$$
  where $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.

Unit root tests

Test	H_0	Reject H_0
ADF	I(1)	tau $<$ Critical value
KPSS	I(0) level	mu $>$ Critical value
	I(0) trend	tau $>$ Critical value
Phillips-Perron	I(1)	Z-tau $<$ Critical value
Zivot-Andrews	I(1)	tau $<$ Critical value

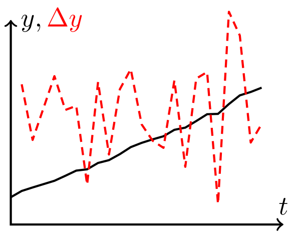
From unit root to weak dependence

Integrated of **order one**, I(1), means that **the first difference** of the process is **weakly dependent** or I(0) (and usually, stationary). Let $\{y_t\}$ be a random walk:

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

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

- The first difference of a series removes its trend.
- Logarithms of a series stabilizes its variance.



From unit root to percentage change

When an I(1) series is strictly positive, logs are often used before differencing to approximate percentage changes:

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

Ergodicity

A strictly stationary process $\{y_t\}$ is **ergodic** if time averages converge to their ensemble averages (expectations). This is often ensured by **strong mixing**, which implies asymptotic independence of distant events.

$$\frac{1}{T} \sum_{t=1}^{T} y_t \underset{a}{\rightarrow} \mathrm{E}(y_t)$$

Without it, sample moments may not reflect population moments. Estimators are inconsistent.

Cointegration

Two I(1) series are **cointegrated** if a linear combination is I(0). In that case, a regression between them is not spurious but reflects a valid **long-run** relationship. Cointegrated variables share a common stochastic trend.
For example, $\{x_t\}$ and $\{y_t\}$ are I(1), but $y_t - \beta x_t = u_t$ where $\{u_t\}$ is I(0). (β is the cointegrating parameter).

Cointegration test

1. Estimate $y_t = \alpha + \beta x_t + \varepsilon_t$ and obtain $\hat{\varepsilon}_t$.
2. Perform an ADF test on $\hat{\varepsilon}_t$ with a modified distribution.
   The result of this test is equivalent to:
   - H_0: $\beta = 0$ (no cointegration)
   - H_1: $\beta \neq 0$ (cointegration)

   if test statistic $>$ critical value, reject H_0.

Heteroscedasticity in time series

The **assumption** affected is **t4**, which leads **OLS to be not efficient**.
Use tests like Breusch-Pagan or White's, where H_0: No heteroscedasticity. For these tests to work, there should be **no autocorrelation**.

ARCH

An autoregressive conditional heteroscedasticity (ARCH) model is used to analyze a form of dynamic heteroscedasticity, 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 heteroscedasticity:

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

GARCH

A general ARCH (GARCH) model is similar to ARCH, but the error variance follows an ARMA(p, q) process.