Анализ стационарности временных рядов

А.Г. Трофимов
к.т.н., доцент, НИЯУ МИФИ

lab@neuroinfo.ru
http://datalearning.ru

Курс "Статистическая обработка временных рядов"

Сентябрь 2018

Difference-Stationary Process

Detrending and deseasonalizing can make a non-stationary (in the mean) time series stationary

But there is another type of non-stationarity in mean that cannot be removed by detrending and deseasonalizing with according to decomposition model. It is a unit root nonstationarity

Definition

A difference-stationary process (DSP, or unit root process) is a stochastic process whose first difference is stationary:

$$Y_t = c + Y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a stationary process, c is a constant (drift)

Usually it is assumed that $\{\varepsilon_t\}$ is a zero-mean stationary process (not necessary white noise)

Trend-Stationary Process

Definition

A process stationary around trend (or trend-stationary process, TSP) is a stochastic process represented as:

$$Y_t = \mu(t) + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a stationary process, $\mu(t)$ is a deterministic function (trend)

Usually it is assumed that $\{\varepsilon_t\}$ is zero-mean stationary process (not necessary white noise)

If $\mu(t) = ct$, where c is a constant, then the trend-stationary process $\{Y_t\}$ has a linear trend

Mean and Variance of TSP and DSP

TSP with linear trend:

$$Y_t = ct + \varepsilon_t, \quad \{\varepsilon_t\} \text{ is zero-mean stationary}$$

$$\mathrm{M}[Y_t] = ct, \qquad \mathrm{D}[Y_t] = \sigma^2$$

Trend-stationary process has linearly growing expectation (deterministic linear trend) and constant variance

DSP:

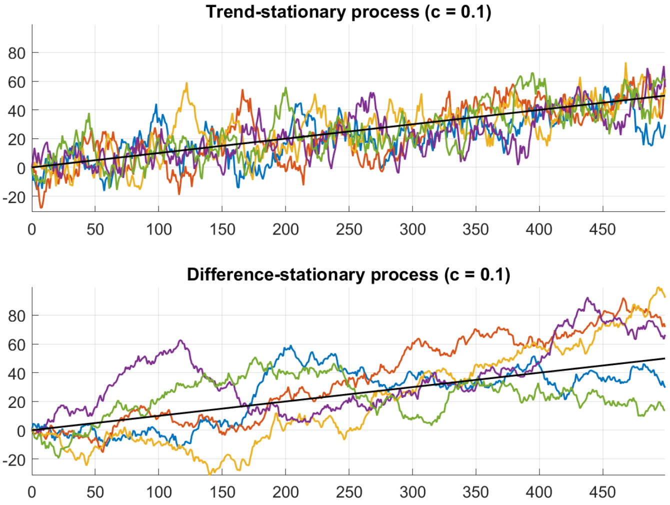
$$Y_t = c + Y_{t-1} + \varepsilon_t, \quad Y_0 = 0, \quad \{\varepsilon_t\} \text{ is zero-mean stationary}$$

$$\mathrm{M}[Y_t] = c + \mathrm{M}[Y_{t-1}] \quad \Rightarrow \quad \mathrm{M}[Y_t] = ct$$

$$\mathrm{D}[Y_t] = \mathrm{D}[Y_{t-1}] + \sigma^2 \quad \Rightarrow \quad \mathrm{D}[Y_t] = \sigma^2 t$$

Difference-stationary process with non-zero drift c has linearly growing expectation (so called stochastic trend) and linearly growing variance

TSP vs DSP. Illustration



Detrending by Subtraction

TSP:

$$Y_t = \mu(t) + \varepsilon_t$$

Detrending by subtraction:

$$\tilde{Y}_t = Y_t - \mathrm{M}[Y_t] = \mu(t) + \varepsilon_t - \mu(t) = \varepsilon_t$$

The detrended process $\{\tilde{Y}_t\}$ is a stationary process

DSP:

$$Y_t = c + Y_{t-1} + \varepsilon_t$$

Detrending by subtraction:

$$\tilde{Y}_t = Y_t - \mathrm{M}[Y_t] = c + Y_{t-1} + \varepsilon_t - ct = c + \tilde{Y}_{t-1} + c(t-1) + \varepsilon_t - ct = \tilde{Y}_{t-1} + \varepsilon_t$$

$$\mathrm{D}[\tilde{Y}_t] = \mathrm{D}[Y_t - ct] = \mathrm{D}[Y_t] = \sigma^2 t$$

The process $\{\tilde{Y}_t\}$ is zero-mean process but it is still unit root process with linearly growing variance

Detrending by Differencing

TSP:

$$Y_t = \mu(t) + \varepsilon_t$$

Differenced process:

$$\Delta Y_t = Y_t - Y_{t-1} = \Delta\mu(t) + \varepsilon_t - \varepsilon_{t-1}$$

The differenced process $\{\Delta Y_t\}$ is MA(1) process with trend $\Delta\mu(t)$

If $\mu(t) = ct$ and $\{\varepsilon_t\}$ is a white noise, then $\Delta Y_t = c + \varepsilon_t - \varepsilon_{t-1}$ is MA(1) process, $c(0) = 2\sigma^2$, $c(1) = -\sigma^2$, $\rho(1) = -0.5$

DSP:

$$Y_t = c + Y_{t-1} + \varepsilon_t$$

Differenced process:

$$\Delta Y_t = Y_t - Y_{t-1} = c + \varepsilon_t$$

The differenced process $\{\Delta Y_t\}$ is a stationary process

TSP and DSP Detrending

- Trend-stationary processes must be detrended by subtraction the trend $\mu(t)$. In practice, the trend $\mu(t)$ is unknown and should be estimated using some regression model. The detrended process:

$$\tilde{Y}_t = Y_t - \hat{\mu}(t)$$

- Difference-stationary processes must be detrended by differencing. The detrended process:

$$\Delta Y_t = Y_t - Y_{t-1}$$

- Difference-stationary processes can have a deterministic trend:

$$Y_t = \mu(t) + Y_{t-1} + \varepsilon_t$$

  It is unit root process with deterministic trend. It must be detrended by subtraction the trend $\mu(t)$ and then differenced to remove non-stationarity

Subtraction vs Differencing

DSP with deterministic trend:

$$Y_t = \mu(t) + Y_{t-1} + \varepsilon_t$$

Trend subtraction:

$$\tilde{Y}_t = Y_t - \mathrm{M}[Y_t] \Rightarrow \tilde{Y}_t = \tilde{Y}_{t-1} + \varepsilon_t \Rightarrow \text{ zero-mean DSP}$$

Differencing:

$$\Delta Y_t = Y_t - Y_{t-1} = \mu(t) + \varepsilon_t \Rightarrow \text{ TSP}$$

	Subtraction	**Differencing**
TSP	Leads to stationary process	Leads to MA(1) process with trend $\Delta\mu(t)$
DSP	Leads to zero-mean DSP process	Leads to stationary process
DSP with trend	Leads to zero-mean DSP process	Leads to TSP

Stochastic Seasonality

Models of time series decomposition:

- Additive decomposition: $Y_t = \mu(t) + s_t + I_t$
- Multiplicative decomposition: $Y_t = \mu(t) s_t I_t$
- Log-additive decomposition: $\log Y_t = \mu(t) + s_t + I_t$

The seasonal component $\{s_t\}$ can be seasonal stochastic process (for example, ARMA-process). It means that the value at each moment in cycle (with period Δ) depends on the values at corresponding moments in previous cycles and random innovations

AR(1) seasonal process:

$$s_t = \phi_{s1} s_{t-\Delta} + \varepsilon_t$$

If the seasonal stochastic process $\{s_t\}$ has a unit root, then the process $\{y_t\}$ has a unit root stochastic seasonality that can be handled by seasonal differencing

Persistence in Time Series

Persistence is defined as continuance of an effect after the cause is removed

Persistence of time series is related to its memory properties

A time series is persistent if the effect of infinitesimally small shock will be influencing the time series for a very long time. Thus the longer the time of influence the longer is the memory and the persistence

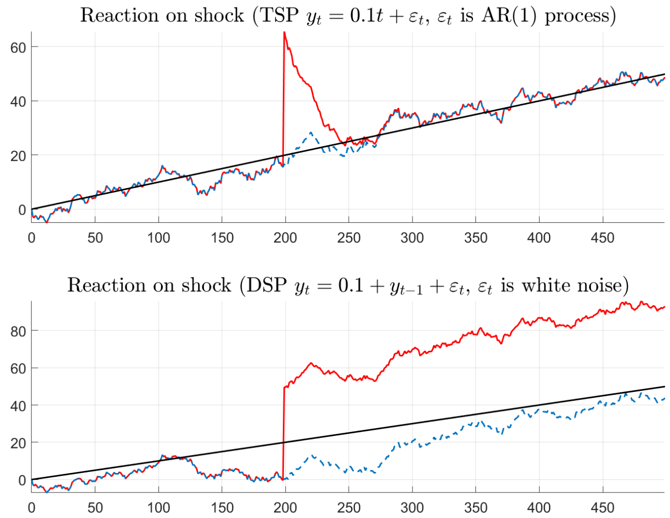
For AR(1) time series the closer ϕ_1 to 1 the more persistent it is

The measure of ARMA process persistence is related to roots of its characteristic polynomial

Random walk process is an example of extremely persistent process (information that comes from the shocks never dies out)

Persistence is not related to stationarity

Persistence of TSP and DSP. Illustration



Unit Root Tests

How to know whether the given time series has a unit root?

It can be shown that TSP and DSP are indistinguishable for finite data. In other words, there are both a TSP and a DSP models that fit a finite time series arbitrarily well

But they can be distinguished under certain assumptions. Every unit root test makes an additional assumption about the underlying (data generating) process

Unit root tests:

- Dickey-Fuller test and augmented Dickey-Fuller test
- Phillips-Perron test
- ADF-GLS test
- KPSS test
- Zivot-Andrews test
- Variance ratio test

Dickey-Fuller Test (1979)

Underlying model: $Y_t = c + bt + \phi_1 Y_{t-1} + \varepsilon_t$
where c is drift, bt is linear deterministic trend, $\{\varepsilon_t\}$ is zero-mean stationary process

Dickey-Fuller test has three versions:

- AR: AR model without trend and drift ($b = 0$, $c = 0$)

$$H_0: \; Y_t = Y_{t-1} + \varepsilon_t$$
$$H': \; Y_t = \phi_1 Y_{t-1} + \varepsilon_t, \quad \phi_1 < 1$$

- ARD: AR model with drift ($b = 0$, $c \neq 0$)

$$H_0: \; Y_t = Y_{t-1} + \varepsilon_t$$
$$H': \; Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t, \quad \phi_1 < 1$$

- TS: AR model with linear trend and drift ($b \neq 0$, $c \neq 0$)

$$H_0: \; Y_t = c + Y_{t-1} + \varepsilon_t$$
$$H': \; Y_t = c + bt + \phi_1 Y_{t-1} + \varepsilon_t, \quad \phi_1 < 1$$

Dickey-Fuller Test. Test Statistic

Underlying model: $Y_t = c + bt + \phi_1 Y_{t-1} + \varepsilon_t$

Dickey-Fuller test considers a differenced process:

$$\Delta Y_t = Y_t - Y_{t-1} = c + bt + \gamma Y_{t-1} + \varepsilon_t$$

where $\gamma = \phi_1 - 1$, as a regression of $\{\Delta Y_t\}$ on $\{Y_{t-1}\}$ and tests if γ significantly differs from 0 w.r.t. alternative $\gamma < 0$

Test statistic: $Z = \frac{\tilde{\gamma}}{\sigma[\tilde{\gamma}]}$, where $\tilde{\gamma}$ is a ordinary least square (OLS) estimation of regression coefficient γ using sample data $y_1, ..., y_T$

In classical regression tests Z has a Student's distribution, but here under null hypothesis of unit root the time series $\{Y_t\}$ is not stationary and ergodic and the usual sample moments do not converge to fixed constants

It was shown that Z tends to Dickey-Fuller distribution $(T \to \infty)$

Dickey-Fuller Test. Recommendations

- If your data shows a linear trend, choose TS version

  H_0 accepted $\Rightarrow$ difference-stationary process
  H_0 rejected and $\phi_1 = 0 \Rightarrow$ trend-stationary process
  H_0 rejected and $\phi_1 \neq 0 \Rightarrow$ AR process with deterministic trend

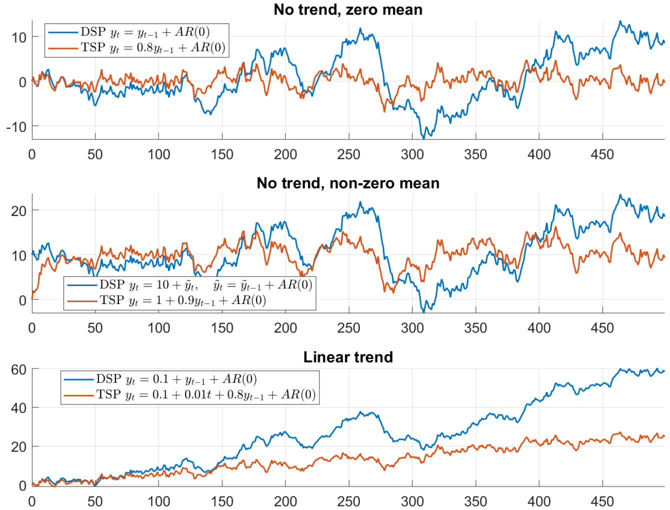
- If your data shows no trend, but seems to have a non-zero mean, choose ARD version

  H_0 accepted $\Rightarrow$ zero-mean difference-stationary process
  H_0 rejected $\Rightarrow$ AR process with constant mean $\mu = \frac{c}{1-\phi_1}$

- If your data shows no trend and seem to have a zero mean, choose AR version

  H_0 accepted $\Rightarrow$ zero-mean difference-stationary process
  H_0 rejected $\Rightarrow$ zero-mean AR process

AR, ARD and TS Processes. Illustration



Phillips–Perron Test (1988)

Phillips–Perron test (PP-test) like Dickey-Fuller test involves fitting the regression model

$$Y_t = c + bt + \phi_1 Y_{t-1} + \varepsilon_t$$

but it uses modified Dickey-Fuller statistics that have been made more robust to autocorrelation and heteroscedasticity in the innovation process $\{\varepsilon_t\}$

Phillips–Perron test has the same variants as Dickey-Fuller test: AR, ARD and TS

It was shown that PP-test performs worse in small samples than the Dickey-Fuller test*. It works well only for large samples

Davidson R., MacKinnon J. Econometric theory and methods. New York: Oxford University Press; 2004.

Augmented Dickey-Fuller Test (1984)

Augmented Dickey-Fuller (ADF) test is an extension of Dickey-Fuller test to AR(p) underlying model

ADF test has three versions:

- AR: AR model without trend and drift ($b = 0$, $c = 0$)

$$H_0: \; Y_t = Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t$$
$$H': \; Y_t = \phi_1 Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t, \quad \phi_1 < 1$$

- ARD: AR model with drift ($b = 0$, $c \neq 0$)

$$H_0: \; Y_t = Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t$$
$$H': \; Y_t = c + \phi_1 Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t, \quad \phi_1 < 1$$

- TS: AR model with linear trend and drift ($b \neq 0$, $c \neq 0$)

$$H_0: \; Y_t = c + Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t$$
$$H': \; Y_t = c + bt + \phi_1 Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t, \phi_1 < 1$$

ADF Underlying Model and Unit Root

The underlying model in ADF test

$$Y_t = \phi_1 Y_{t-1} + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \varepsilon_t$$

is equivalent to AR($p+1$) model

$$Y_t = \phi_1' Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_{p+1} Y_{t-(p+1)} + \varepsilon_t$$

with the coefficients $\phi_1', \phi_2, \ldots, \phi_{p+1}$ related to $\phi_1, \beta_1, \ldots, \beta_p$

Why $\phi_1 = 1$ leads to unit root?

$$Y_t = \phi_1 L Y_t + \beta_1 (1-L) L Y_t + \ldots + \beta_p (1-L) L^p Y_t + \varepsilon_t$$

$$(1 - \phi_1 L) Y_t = (1-L)(\beta_1 L + \ldots + \beta_p L^p) Y_t + \varepsilon_t$$

$$\phi(L) = (1 - \phi_1 L) - (1-L)(\beta_1 L + \ldots + \beta_p L^p)$$

$\phi_1 = 1 \Rightarrow$ characteristic polynomial $\phi(z)$ has unit root

Determine Appropriate Lags for ADF-Test

ADF-test requires to specify the number of lagged difference terms p in underlying model

If p is too small then the remaining serial correlation in the residuals will bias the test

If p is too large then the power of the test will suffer (false acceptances of null hypothesis occur)

Recommendation:
Perform ADF-test for $p = p_{\max}$, where $p_{\max} = \left[12\left(\frac{T}{100}\right)^{1/4}\right]$*. If the last regression coefficient β_p of the fitted ADF model is insignificant (e.g., $p > 0.1$), then reduce p by one and repeat test**

*Schwert G. Tests for unit roots: A Monte Carlo investigation. Journal of Business and Economic Statistics. 2002. Vol. 20(1), pp.5-17.

**Ng S., Perron P. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. Journal of the American Statistical Association. 1995. Vol. 90(429), pp.268-81.

KPSS Test (1992)

In KSPP-test (Kwiatkowski, Phillips, Schmidt and Shin) the null hypothesis is that a time series is trend stationary against the alternative that it is a non-stationary unit root process

Underlying model:

$$Y_t = ct + X_t + \varepsilon_t$$

$$X_t = X_{t-1} + \nu_t$$

where $\{\varepsilon_t\}$ is a stationary process, $\{\nu_t\}$ is an IID process with zero mean and variance σ^2

Null and alternative hypotheses:

$$H_0: \ \sigma^2 = 0 \qquad H': \ \sigma^2 > 0$$

Under the null hypothesis the process $\{X_t\}$ becomes a constant intercept and, thus, the process $\{Y_t\}$ becomes TSP. The alternative $\sigma^2 > 0$ implies that $\{Y_t\}$ has a unit root

Versions of KPSS Test

KPSS-test has two versions:

- Without trend ($c = 0$)

$$Y_t = X_t + \varepsilon_t$$

$$X_t = X_{t-1} + \nu_t$$

Choose this version if your time series shows no trend

- With linear trend ($c \neq 0$)

$$Y_t = ct + X_t + \varepsilon_t$$

$$X_t = X_{t-1} + \nu_t$$

Choose this version if your time series shows a linear trend

KPSS Test. Test Statistic

KPSS test considers a linear regression of Y on time t (linear trend) and regression residuals $e_1, ..., e_T$

KPSS test statistic:

$$Z = \frac{1}{T^2 \tilde{\sigma}_e^2} \sum_{t=1}^{T} S_t^2$$

where $\tilde{\sigma}_e^2$ is the estimate of residual variance and $\{S_t\}$ is the partial sum process of the residuals:

$$\tilde{\sigma}_e^2 = \frac{1}{T} \sum_{t=1}^{T} e_t^2, \quad S_t = \sum_{i=1}^{t} e_i$$

Test statistics Z follow nonstandard distributions under the null hypothesis, even asymptotically. Its critical values was obtained numerically using Monte Carlo simulations and was tabulated*

*Kwiatkowski D., Phillips P., Schmidt P., Shin. Y. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics. Vol. 54, 1992, pp. 159–178.

Variance Ratio Test (1988)

Variance ratio test is not usually used as unit root test, it is used for testing whether an innovation process is a pure random walk versus having some predictability

Underlying model:

$$Y_t = c + Y_{t-1} + \varepsilon_t$$

where c is a drift, $\{\varepsilon_t\}$ is innovation process with zero mean

Null and alternative hypotheses:

$$H_0 : \ \{\varepsilon_t\} \text{ is uncorrelated process}$$
$$H' : \ \{\varepsilon_t\} \text{ is correlated}$$

The null hypothesis can be rejected due to correlation of $\{\varepsilon_t\}$, heteroskedasticity of $\{\varepsilon_t\}$, or if $\{Y_t\}$ has no unit root

Variance Ratio Test

The variance ratio test is based on the fact that the variance of a random walk increases linearly with time:

$$\mathrm{D}[Y_t] = ct$$

It means that variance $\mathrm{D}[Y_t - Y_{t-q}] = q\mathrm{D}[Y_t - Y_{t-1}]$ for all t

Variance ratio with period q:

$$V_q = \frac{\mathrm{D}[Y_t - Y_{t-q}]}{q\mathrm{D}[Y_t - Y_{t-1}]}$$

Under null hypothesis (random walk with uncorrelated innovations): $V_q \simeq 1$ (for all q)

For a mean-reverting time series (time series that revert to the mean after shock): $V_q < 1$

For a mean-averting time series: $V_q > 1$

Unit Root Tests. Statistical Decisions

Test	H_0 **accepted** $(p > \alpha)$	H_0 **rejected** $(p < \alpha)$
ADF, PP (AR)	zero-mean DSP	zero-mean AR process
ADF, PP (ARD)	zero-mean DSP	constant-mean AR process
ADF, PP (TS)	DSP	TSP
KPSS (lin. trend)	TSP	DSP with lin. trend
KPSS (no trend)	stationary	DSP
Variance ratio	DSP	TSP or DSP with correlated or heteroskedastic innovations

Any of the stationarity tests can be rejected due to incorrect underlying model assumptions

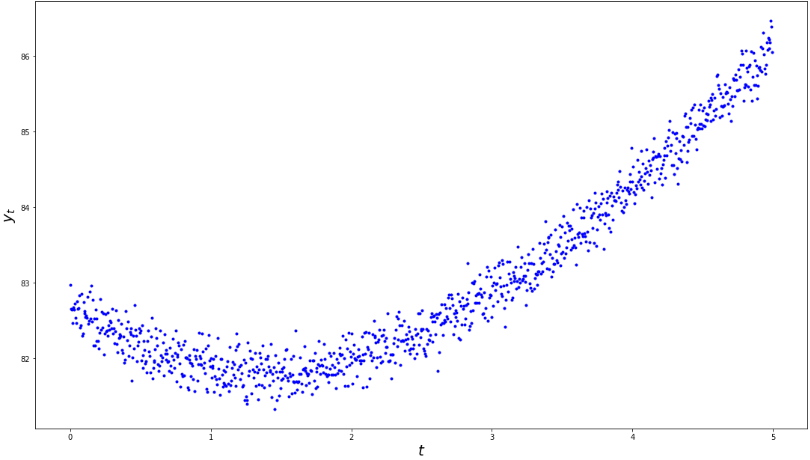
It is good idea to perform several stationarity tests to make statistical decision

Notes on Unit Root Tests

- ADF, PP and KPSS tests have very low power against roots close to 1. That is, they cannot distinguish highly persistent stationary processes from non-stationary processes
- Power of unit root tests diminish as deterministic terms are added to the underlying model. That is, tests that include drift and trend in the model have less power than tests that only include drift
- Non-stationary but memoryless time series can easily trick unit-root tests
- It is impossible to test whether a time series is non-stationary with a single path observed over a finite time interval. Thus, a null hypothesis rejection can either represent empirical evidence that the underlying model is incorrect, or that the underlying model is correct but the null hypothesis is false

Time Series Stationarity. Illustration 1

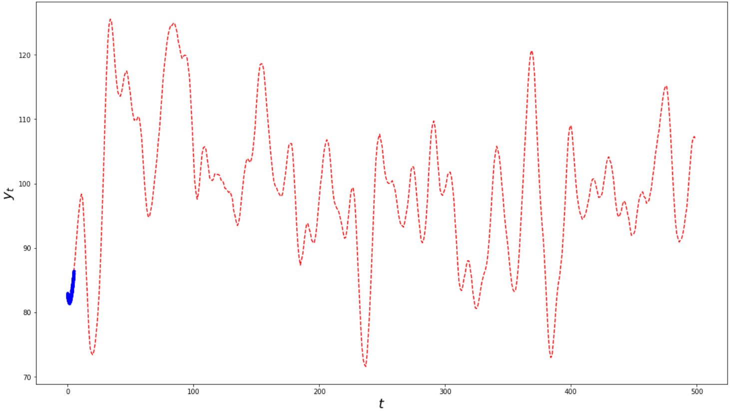
Is this time series stationary?



Any statistical test will show that it is non-stationary

Time Series Stationarity. Illustration 2

The same plot over a much longer time horizon



The process didn't change but statistical tests will show that it is stationary and has no unit roots