Εισαγωγή στις χρονοσειρές

ΣΑΣΤΑ47-17, Άνοιξη 2024

1ο μάθημα

Κωνσταντίνος Χριστόπουλος, PhD

Διδακτικοί στόχοι

- Βασικά χαρακτηριστικά και τύποι χρονοσειρών
- Διαγνωστικά εργαλεία για την συμπεριφορά των χρονοσειρών

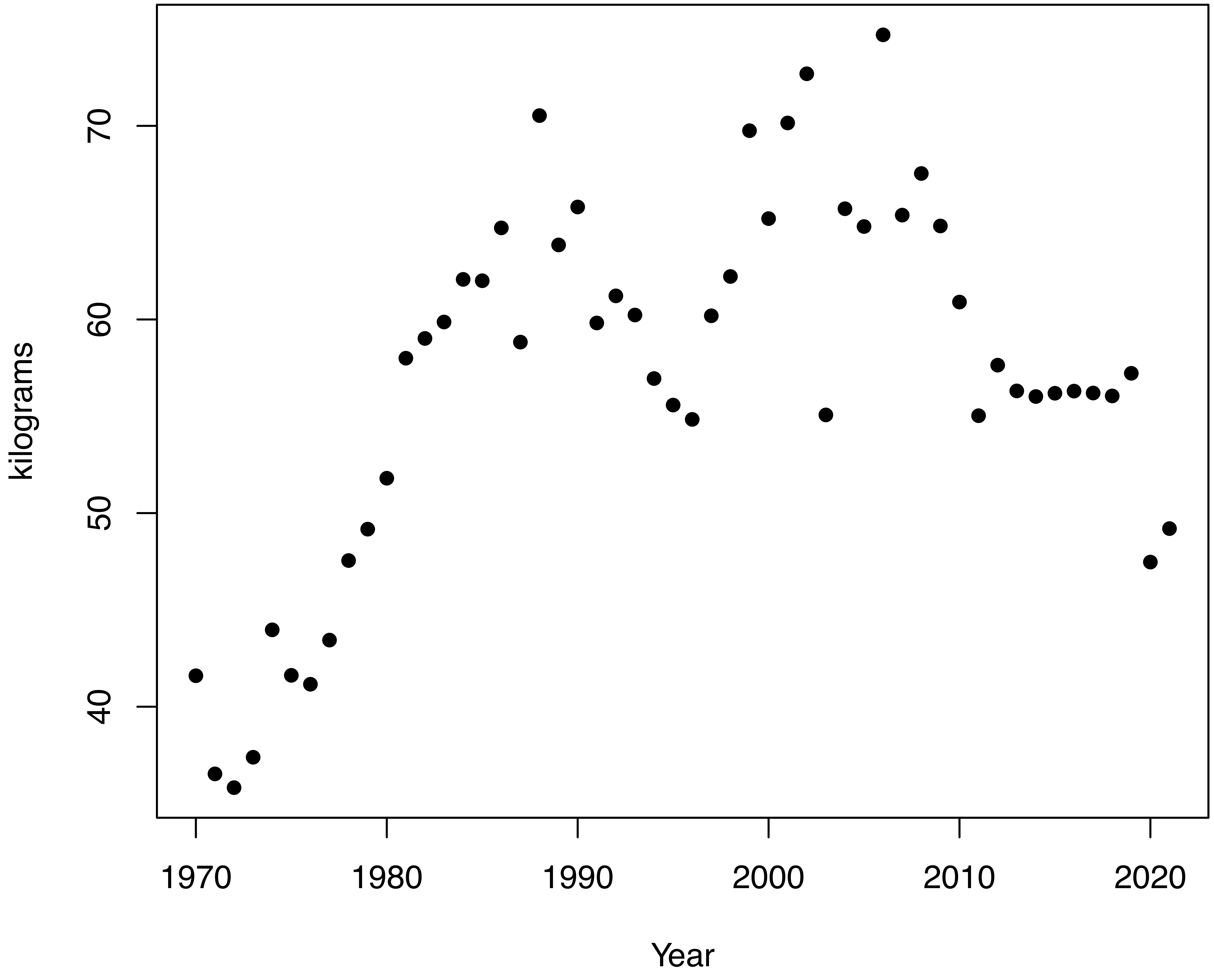
Reading material

- Αναγνώστου Α, (2023). Κλασσικά και Σύγχρονα Υποδείγματα Χρονολογικών Σειρών. Κεφάλαια 2,3
- Shumway, R. H. & Stoffer, D. S. (2010). Time series analysis and its applications (3rd ed.) κεφάλαιο 1

Τι είναι οι χρονοσειρές?

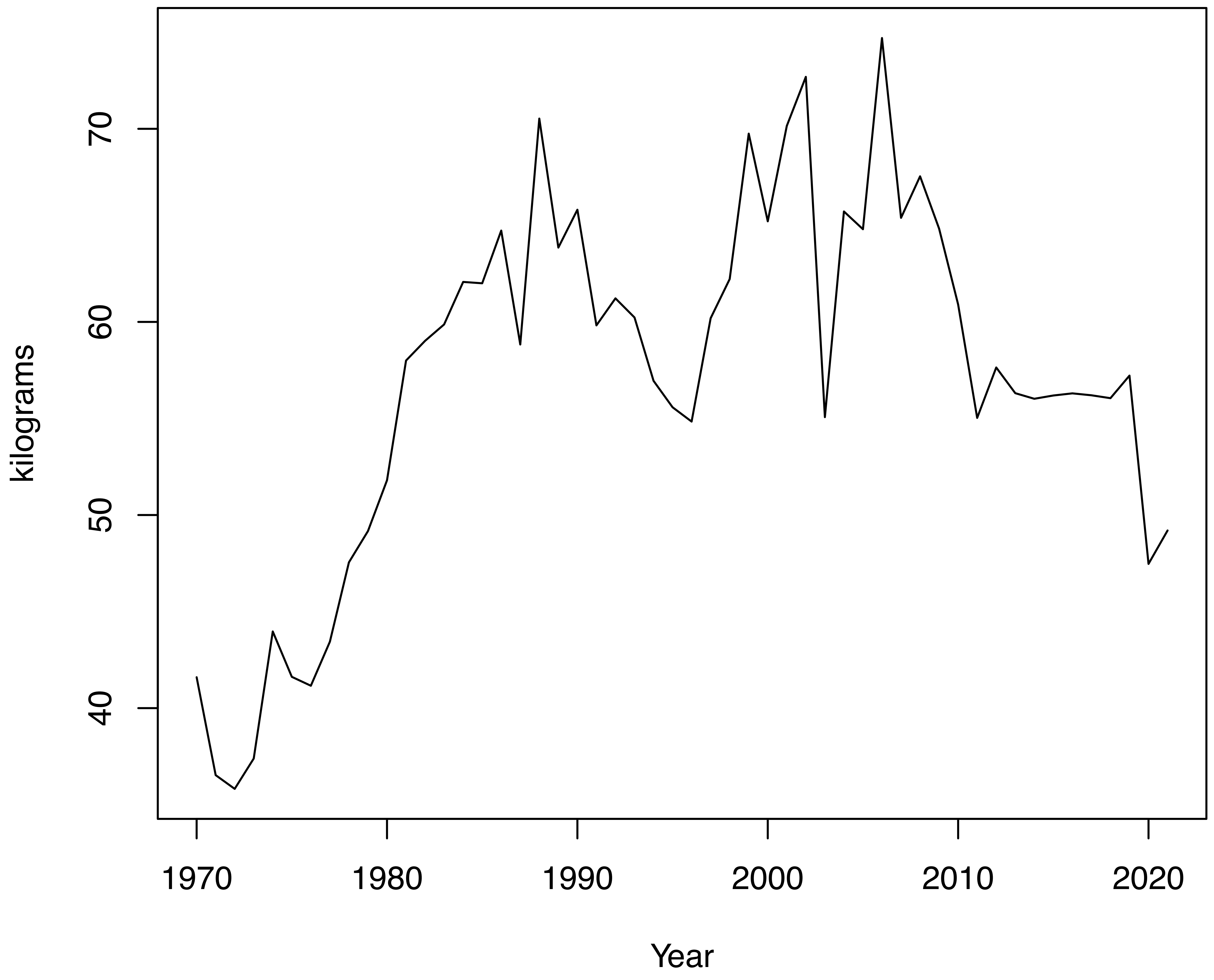
- Μια σειρά επαναλαμβανόμενων με μετρήσεων στον χρόνο ανά ίσα (ή και όχι) διαστήματα (π.χ. κάθε βδομάδα) για το ίδιο άτομο $x_1, x_2, x_3, \ldots x_t$ όπου t το σύνολο των μετρήσεων
- Ο χρόνος μπορεί να είναι είτε συνεχείς είτε διακριτός
- Οι παρατηρήσεις ΔΕΝ είναι ανεξάρτητες μεταξύ τους
- Αυτή η μη-ανεξαρτησία αφορά το βασικό αντικείμενο μελέτης στην ανάλυση χρονοσειρών
- Η σειρά των παρατηρήσεων έχει σημασία γιατί παρέχει πληροφορίες

Yearly per capita alcohol consumption in Greece



Source: FAOSTAT

Yearly per capita alcohol consumption in Greece



Source: FAOSTAT

Σκοπός της ανάλυσης χρονοσειρών

- Να καταλάβουμε την δομή της χρονοσειράς=> πως μεταβάλετε στον χρόνο, πως επηρεάζεται από τις προηγούμενες τιμές της, πως επηρεάζεται από άλλες χρονοσειρές ή παράγοντες
- Να προβλέψουμε μελλοντικές τιμές της χρονοσειρά (forecasting)

Ταξινόμηση χρονοσειρών

Με βάση τον χρόνο

- Διάστημα χρόνου $t \in [t_1, t_2]$ (συνεχής χρόνος)
- Ίσα διαστήματα (με ή χωρίς missing values) t={1, 2, 3 ,4 ,5}
- Άνισα διαστήματα t={1, 2, 5 ,6, 9}

Ταξινόμηση χρονοσειρών

Βάση τον αριθμό των καταγεγραμμένων τιμών

- Μονομεταβλητή (univariate) =>μια χρονοσείρα
- Πολυμεταβλητή (multivariate) =>πολλές χρονοσειρές

Ταξινόμηση χρονοσειρών

Βάση τον τύπο των τιμών

- Ακέραιος (integer)
- Αναλογική (rational)
- Πραγματική (real)
- Σύνθετη (complex)

Συχνότητα μετρήσεων
Measurement frequency

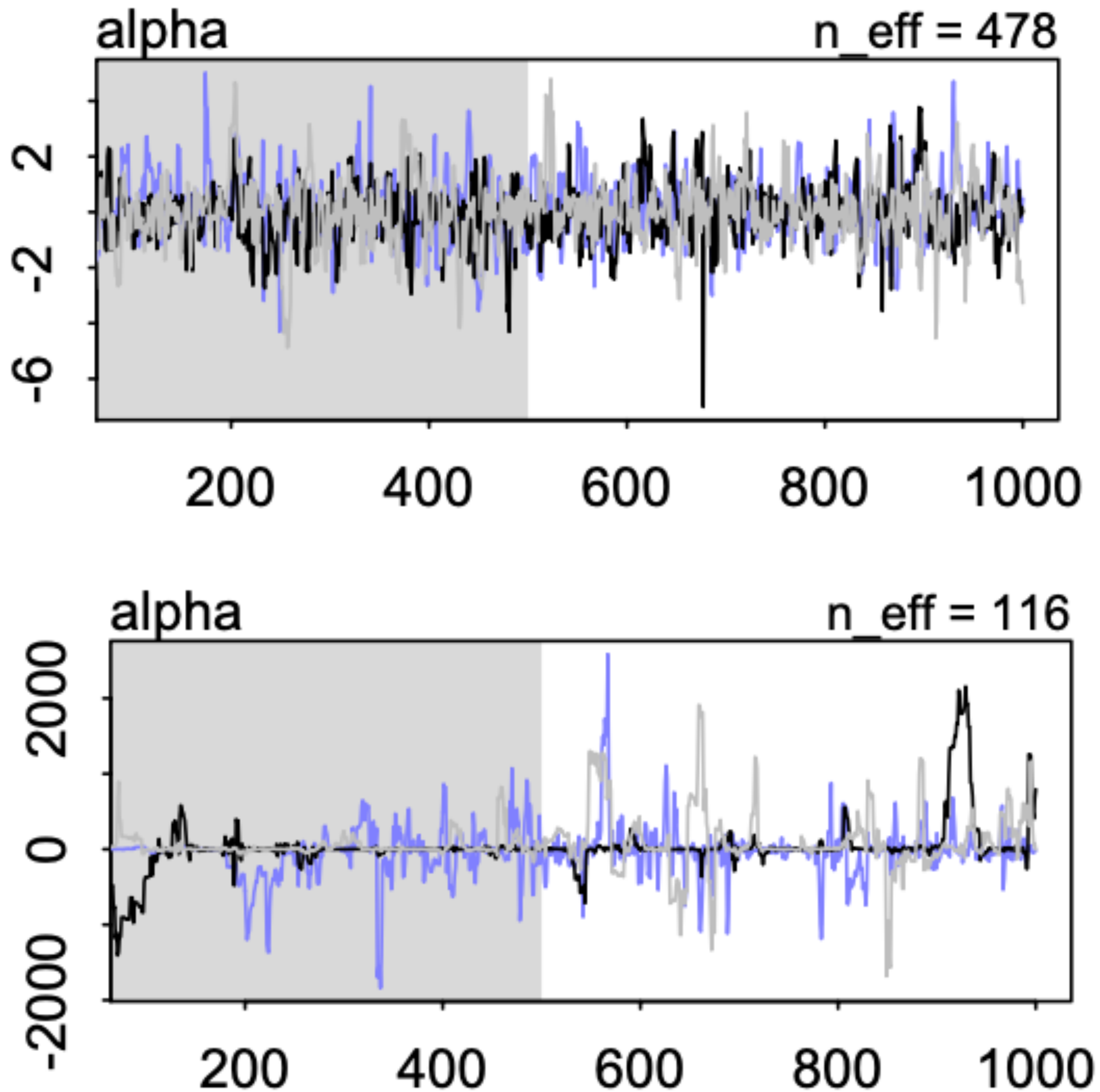
- Υψηλή=>συχνές μετρήσεις (π.χ. ημερήσια συχνότητα)
- Χαμηλή=>αραιές μετρήσεις (π.χ. ετήσια συχνότητα)
- μικρότερη ή μεγαλύτερη μεταβλητότητα?

Ντετερμινιστικές και στοχαστικές διαδικασίες
Deterministic and stochastic processes

- Στοχαστική διαδικασία=ένα στατιστικό φαινόμενο που εξελίσσεται στον χρόνο βάση νόμων των πιθανοτήτων (propabilistic laws)
- Οι χρονοσειρές μπορούν να θεωρηθούν ως μια πραγματοποίηση (realisation) μια τέτοιας διαδικασίας
- Ντετερμινιστικές όταν μπορεί να γίνει ακριβής πρόβλεψη
- Στοχαστικές όταν καθορίζονται μερικώς από τις προηγούμενες τιμές—δεν είναι δυνατή η ακριβής πρόβλεψη

Στασιμότητα
Stationarity

- Στάσιμη διαδικάσια=>ειδική περίπτωση στοχαστικής διαδικασίας
- Βασίζεται στην υπόθεση ότι η διαδικασία βρίσκεται σε κατάσταση στατιστικής ισορροπίας (statistical equilibrium)
- Αυστηρή στασιμότητα=> οι ιδιότητες της στοχαστικής διαδικασίας δεν επηρεάζονται από αλλαγές στην έναρξη του χρόνου
- Με άλλα λόγια, για να είναι μια (διακριτή) διαδικασία αυστηρά στάσιμη, από κοινού κατανομή πιθανοτήτων οποιουδήποτε συνόλου παρατηρήσεων δεν πρέπει να επηρεάζεται μετατοπίζοντας όλους τους χρόνους παρατήρησης προς τα εμπρός ή προς τα πίσω κατά οποιοδήποτε ακέραιο ποσό k
- Στασιμότητα 2η τάξης (weak stationarity)=σταθερός μέσο και (συν)διακύμανση για κάθε t



Source: McElreath, R. (2018). Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC

Lag operators

- Είναι κοινό να αναλύονται χρησιμοποιώντας τις προηγούμενες τιμές τους $lag(1) = L1 = x_{t-1}$

- $forward(1) = F1 = x_{t+1}$

- $difference(1) = D1 = x_t - x_{t-1}; difference(2) = D2 = D1 - (x_{t-1} - x_{t-2})$ η διαφορά βοηθάει στο να γίνει στάσιμη. Πάνω από 2 διαφορές=φάουλ

- Εποχικές διαφορές (seasonal differencing) “βγάζουν” την εποχικότητα $seasonal(1) = S1 = D1 = x_t - x_{t-1}; seasonal(2) = S2 = x_t - x_{t-2}$

Διαγνωστικά εργαλεία

- Πάντα plot the series first!
- Συνάρτηση αυτοσυσχέτισης ACF (autocorrelation function)
- Συνάρτηση μερικής αυτοσυσχέτισης PACF (partial autocorrelation function)
- Decomposition
- Εποχικότητα (seasonality)

Time series plot

Κοιτάω για

- τάση (trend)=κατά μέσο όρο οι μετρήσεις αυξάνουν (ή μειώνουν) με τον χρόνο
- Εποχικότητα (seasonality)=τακτικό/επαναλαμβανόμενο σχήμα αυξήσεων (ή μειώσεων) σχετικά με τον ημερολογιακό χρόνο (εποχές, μήνες, μέρες, κλπ)
- Outliers (τιμές μακριά από άλλα δεδομένα)
- Μακροχρόνιοι κύκλοι (άσχετα με την εποχικότητα)
- Σταθερή διακύμανση στο χρόνο
- Απότομες αλλαγές στις τιμές ή την διακύμανση

Αυτοδιακύμανση
Autocovariance

- Διακύμανση $Cov_{x,y} = E[(X - \mu_x)(Y - \mu_y)]$
- Αυτοδιακύμανση $Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k$, για κ τάξη (order)
- Θετική=οι επόμενες τιμές έχουν την τάση να έχουν το *ίδιο* πρόσημο με τις προηγούμενες
- Αρνητική=οι επόμενες τιμές έχουν την τάση να έχουν *διαφορετικό* πρόσημο με τις προηγούμενες

Μήτρα αυτοδιακύμανσης

Autocovariance matrix

- Συμμετρική (symmetric)
- Θετικά ορισμένη (positive-definite)
- γ_0 η διακύμανση του X

$$\Gamma = V(X) = \begin{bmatrix} \gamma_o & \gamma_1 & \gamma_2 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{k-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{k-3} \\ . & . & . & \cdots & . \\ \gamma_{k-1} & \gamma_{k-2} & \gamma_{k-3} & \cdots & \gamma_0 \end{bmatrix}$$

Αυτοσυσχέτιση
Autocorrelation

- Συσχέτιση $\rho_{x,y} = \dfrac{Cov(x,y)}{\sigma_x \sigma_y}$
- Αυτοσυσχέτιση για κ τάξη (order) $\rho_k = \dfrac{Cov(X_t, X_{t+k})}{\sigma_t \sigma_{t+k}} = \dfrac{\gamma_k}{\gamma_0}$
- $-1 < \rho_k < 1$, συνήθως θετική
- Durbin-Watson test (υπό πολλές προϋποθέσεις)
- Breusch-Godfrey test (υπό λιγότερες προϋποθέσεις)
- Ljung-Box, Box-Pierce test

Μήτρα αυτοσυσχέτισης

Autocorrelation matrix

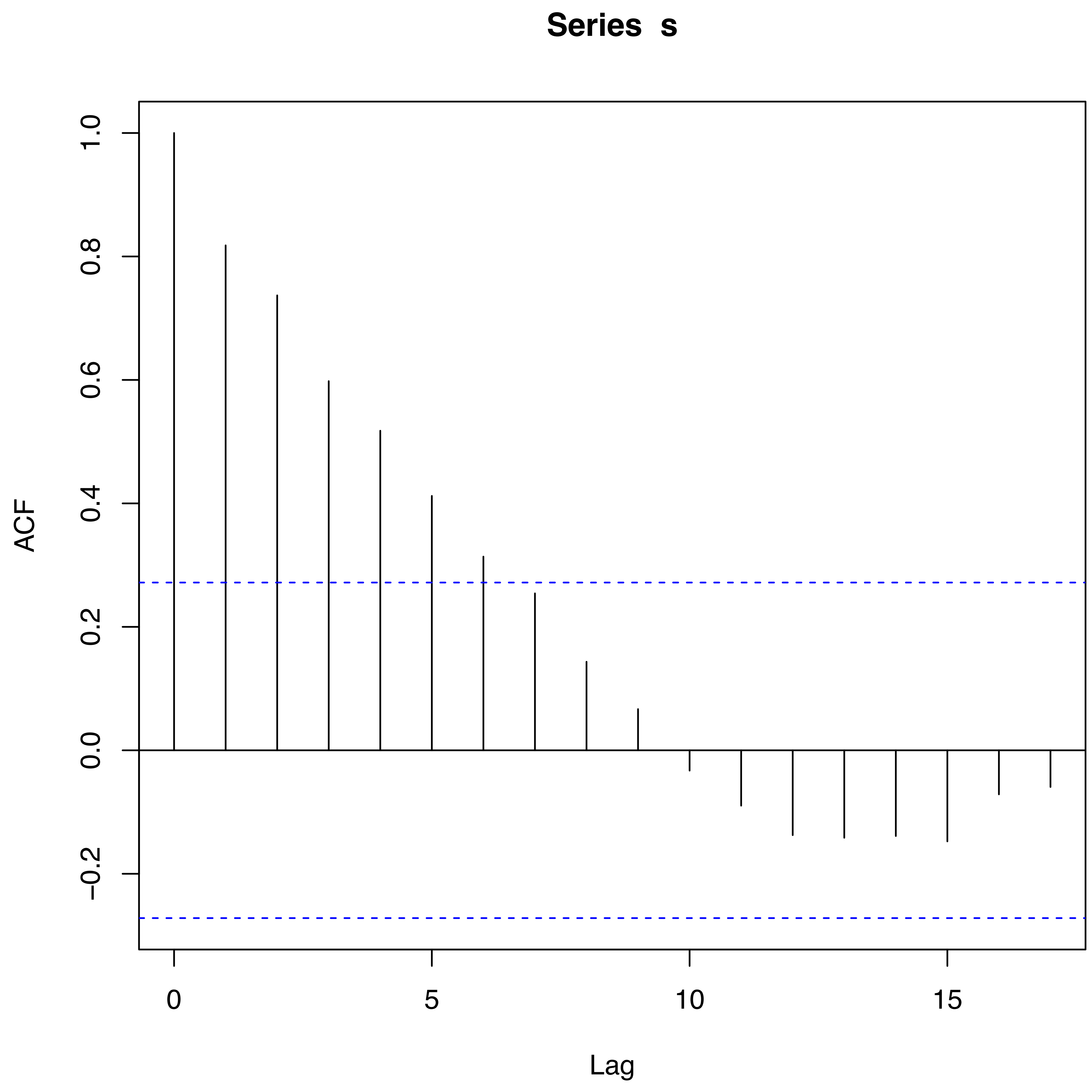
- Συμμετρική (symmetric)
- Θετικά ορισμένη (positive-definite)
- $\rho_0 = 1$
- Ανεξάρτητη από κλίμακες μέτρησης (dimensionless)
- $R = \dfrac{1}{\gamma_0}\Gamma$

$$R = \begin{bmatrix} \rho_o & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & \rho_0 & \cdots & \rho_{k-3} \\ . & . & . & \cdots & . \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_0 \end{bmatrix}$$

Συνάρτηση αυτοσυσχέτισης

Autocorrelation function (ACF)

- Συσχέτιση με προηγούμενες τιμές
- Σε στάσιμες χρονοσειρές θα “σβήνει” πολύ γρήγορα (σε λίγα lags)
- Σε μη-στάσιμες η (αυτο)συσχέτιση επιμένει



Μερική αυτοσυσχέτιση

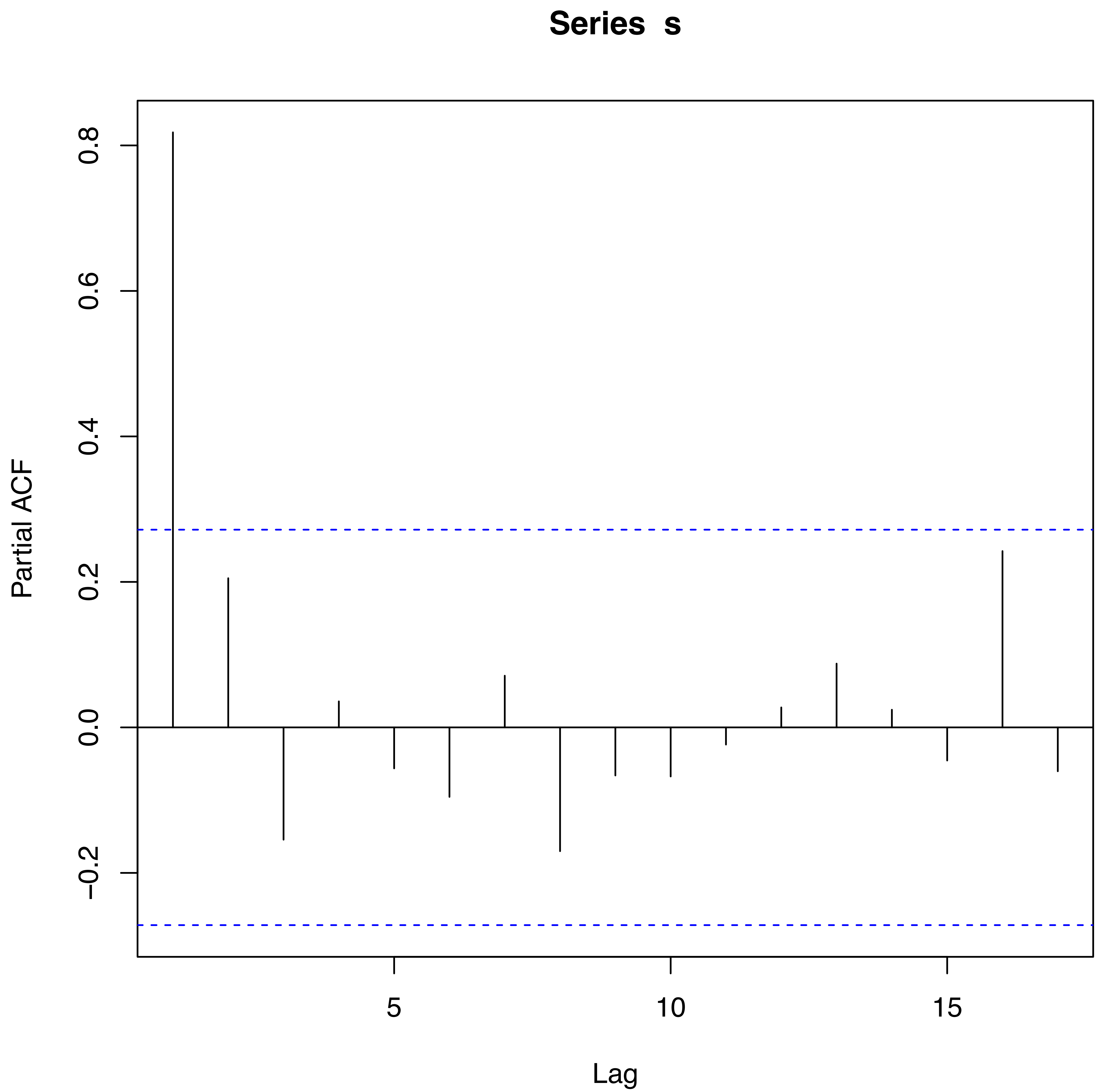
Partial autocorrelation

- Μερική συσχέτιση
- Μερική αυτοσυσχέτιση τάξης k p_k
- Αυτοσυσχέτιση μεταξύ των τιμών X_t, X_{t+k} κρατώντας σταθερές τις ενδιάμεσες τιμές $X_{t+1}, \ldots, X_{t+k-1}$

Συνάρτηση μερικής αυτοσυχέτισης

Partial autocorrelation function (PACF)

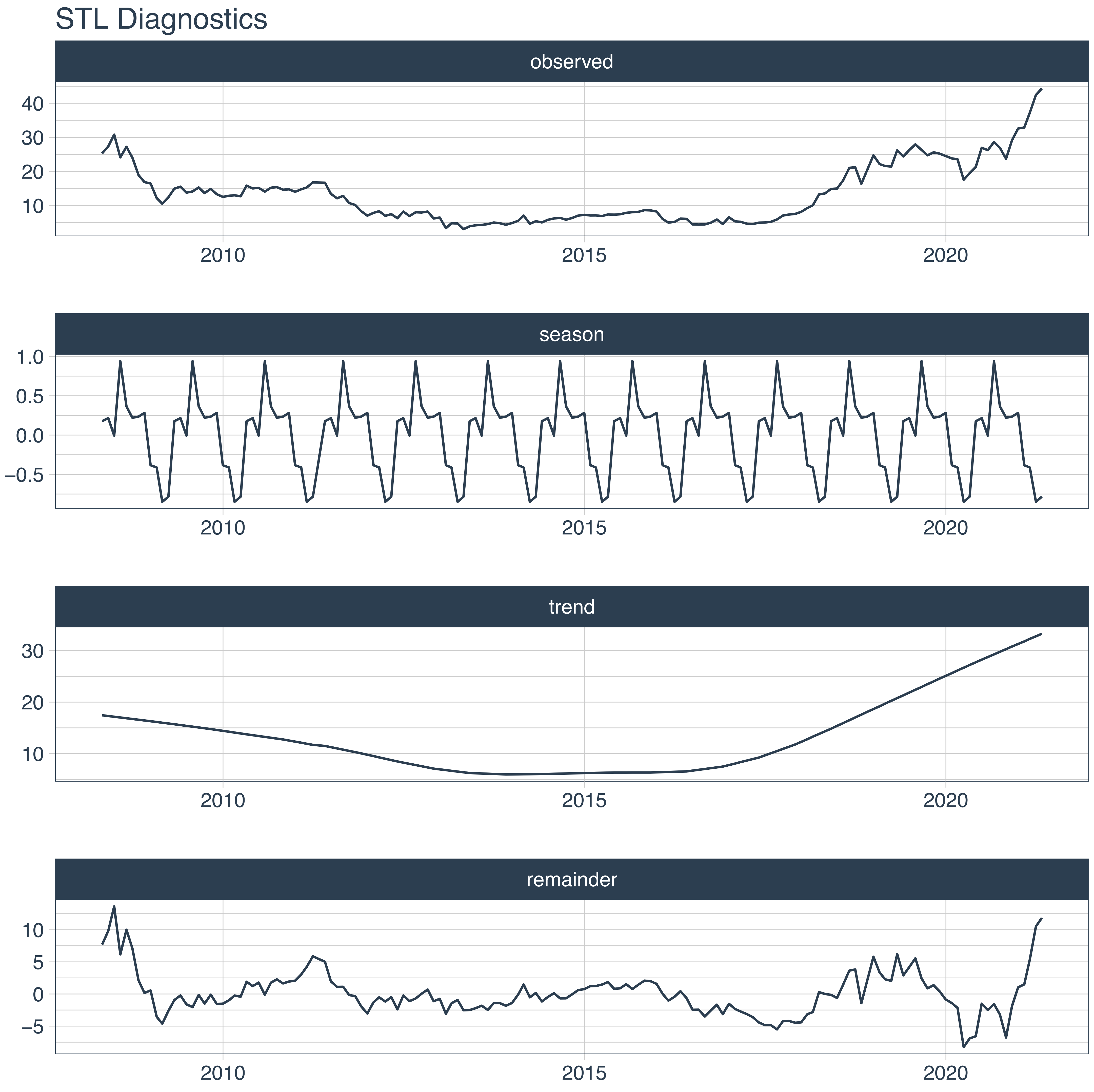
- Για lag=1 είναι το ίδιο με την συνάρτηση αυτοσυσχέτισης (ACF)



Time series decomposition

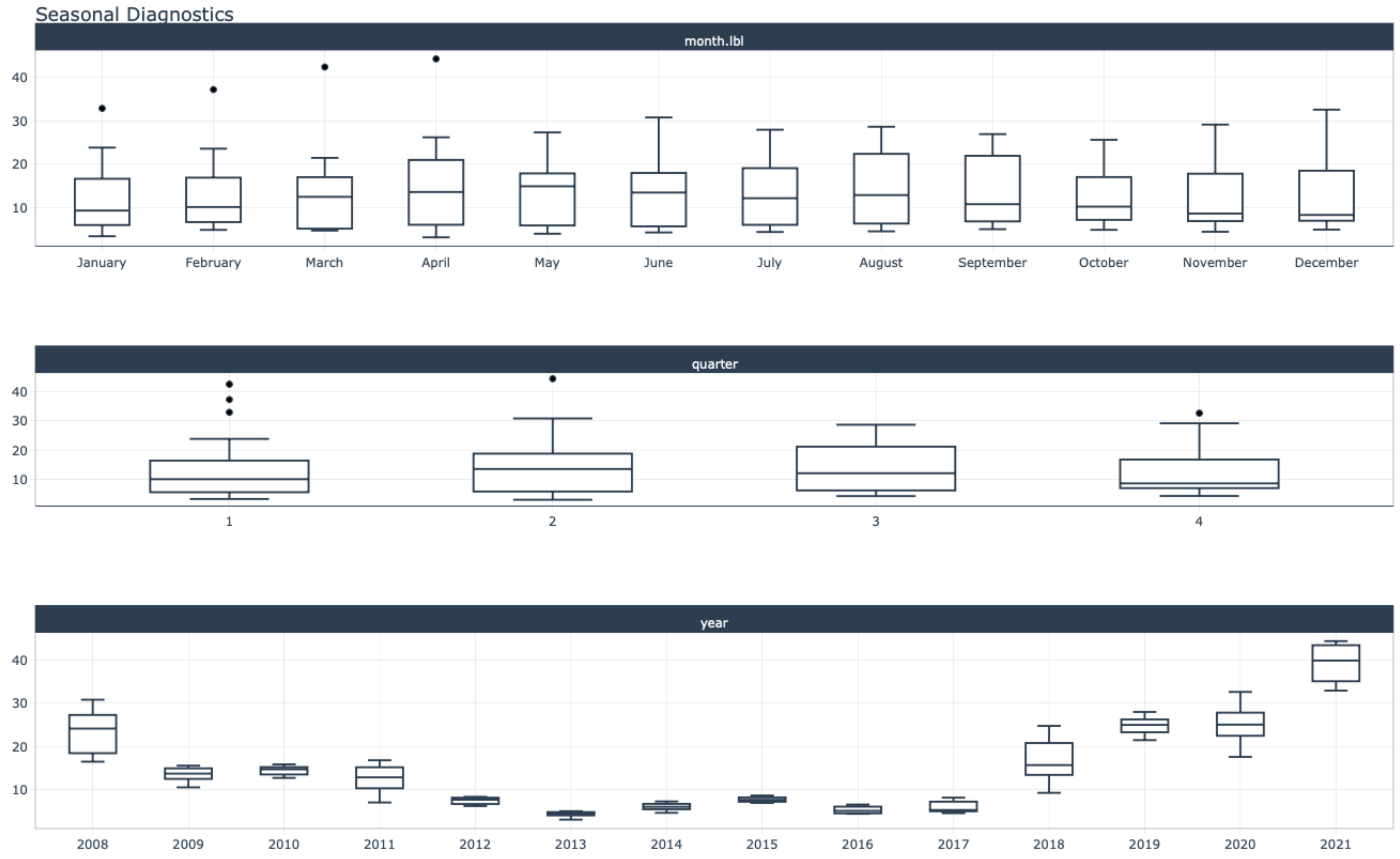
Σε 3 κύρια components

- Τάση (trend)
- Εποχικότητα (seasonality)
- Κατάλοιπα (residuals/noise)
- Αθροιστική $x_t = t + s + r$
- Πολλαπλασιαστική $x_t = t \times s \times r$



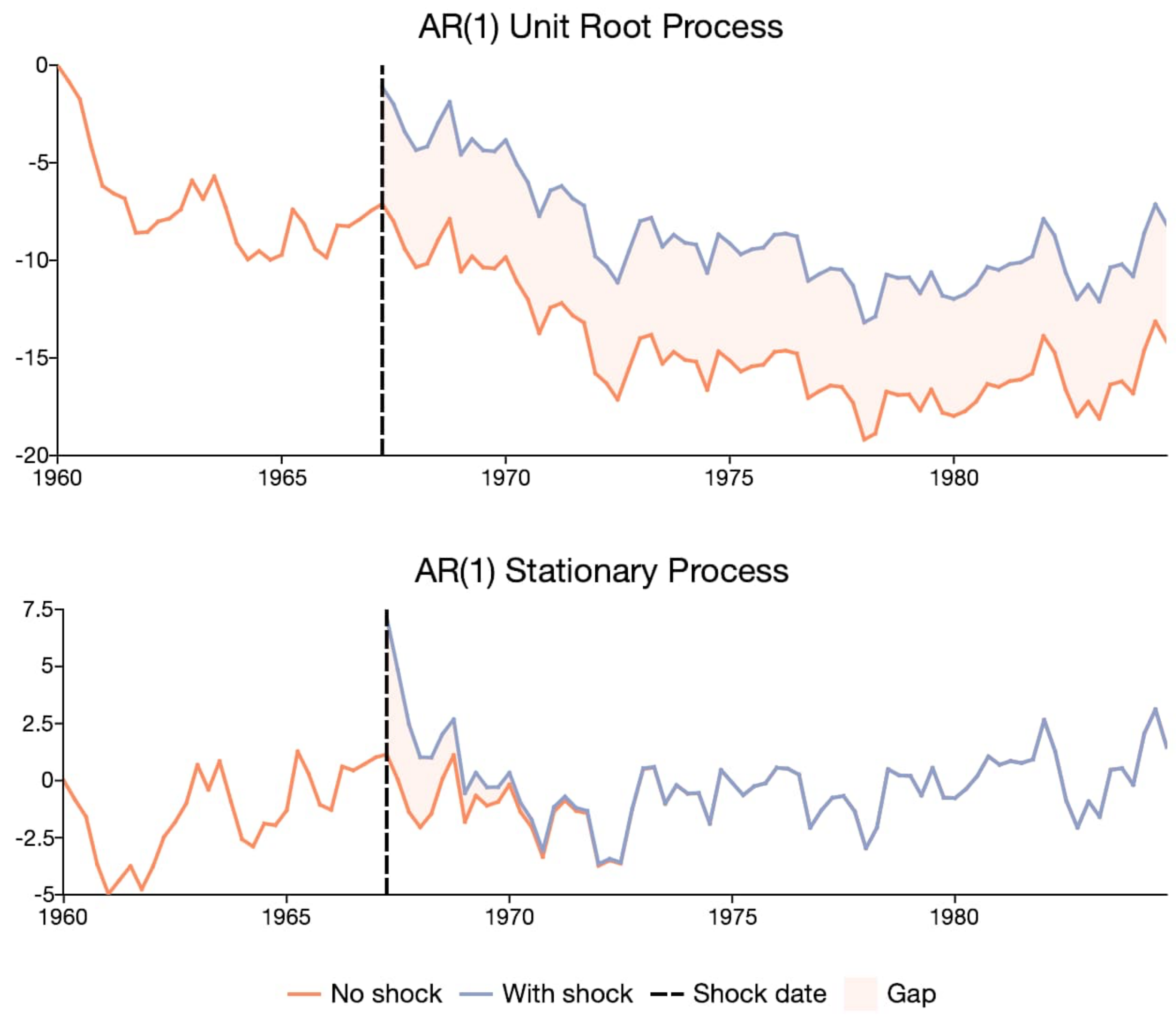
Εποχικότητα
Seasonality

- Συστηματικές μεταβολές της τιμής ανάλογα την εποχή (π.χ. μήνας, τρίμηνο, ΣΚ) προς κάποια κατεύθυνση



Μοναδιαία ρίζα
Unit root

- όταν η σειρά έχει μοναδιαία ρίζα σημαίνει ότι υπάρχει πάνω από μια τάση (trend)
- Dickey-Fuller test (Ho=έχει μοναδιαία ρίζα)
- Structural break=αναπάντεχη/απότομη αλλαγή των τιμών στον χρόνο
- Chow test (Ho=no break)



Προβλήματα των παραδοσιακών μεθόδων ανάλυσης

- Ιδιότητες του πληθυσμού από ένα δείγμα του
- Πολλές φορές αδύνατο να πάρουμε πάνω από μια μέτρηση/χρόνο (time)
- Οι παρατηρήσεις δεν είναι ανεξάρτητες (αυτοσυσχέτιση)
- Η μη ανεξαρτησία σημαίνει ότι οι πραγματικά ανεξάρτητες παρατηρήσεις είναι *λιγότερες* από το δείγμα
- Η αυτοσυσχέτιση αυξάνει την στατιστική σημαντικότητα

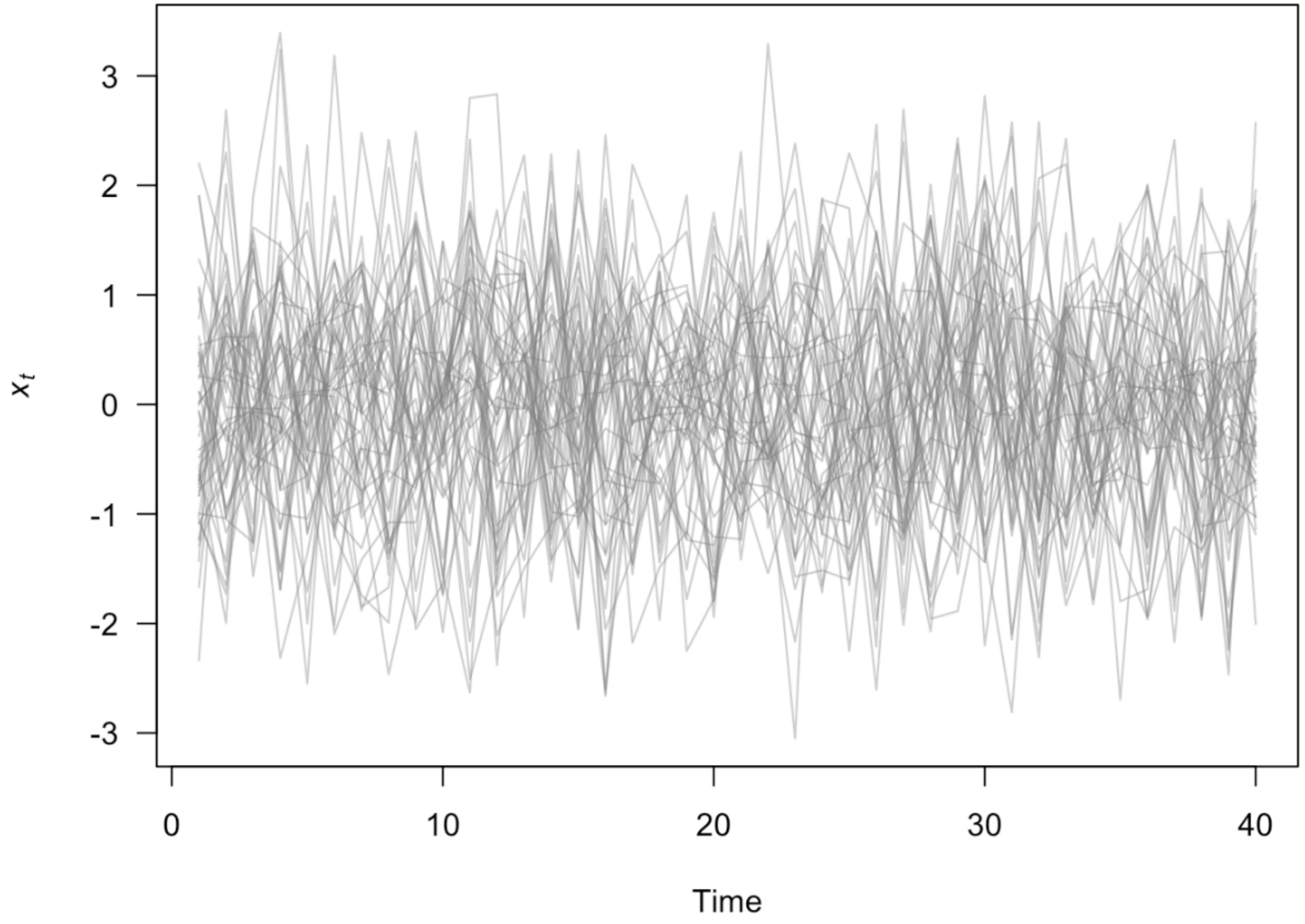
Μοντέλο χρονοσειράς

Time series model

- Το χρησιμοποιούμε για να αναλύσουμε δεδομένα από χρονοσειρές
- Ένα μοντέλο για την x_t είναι ένα υπόδειγμα (specification) των από κοινού κατανομών μιας αληλλουχίας τυχαίων μεταβλητών X_t που η x_t είναι μια πραγματοποίηση
- Οι διαφορετικές σειρές αντιπροσωπεύουν την κατανομή των πιθανών πραγματοποιήσεων (realizations) παρότι εμείς βλέπουμε μόνο μια
- Το μοντέλο βοηθάει στο να βγάλουμε συμπεράσματα από αυτή την μια παρατηρούμενη χρονοσειρά για την από κοινού κατανομή (joint distribution) από την οποία προήλθε

White noise model

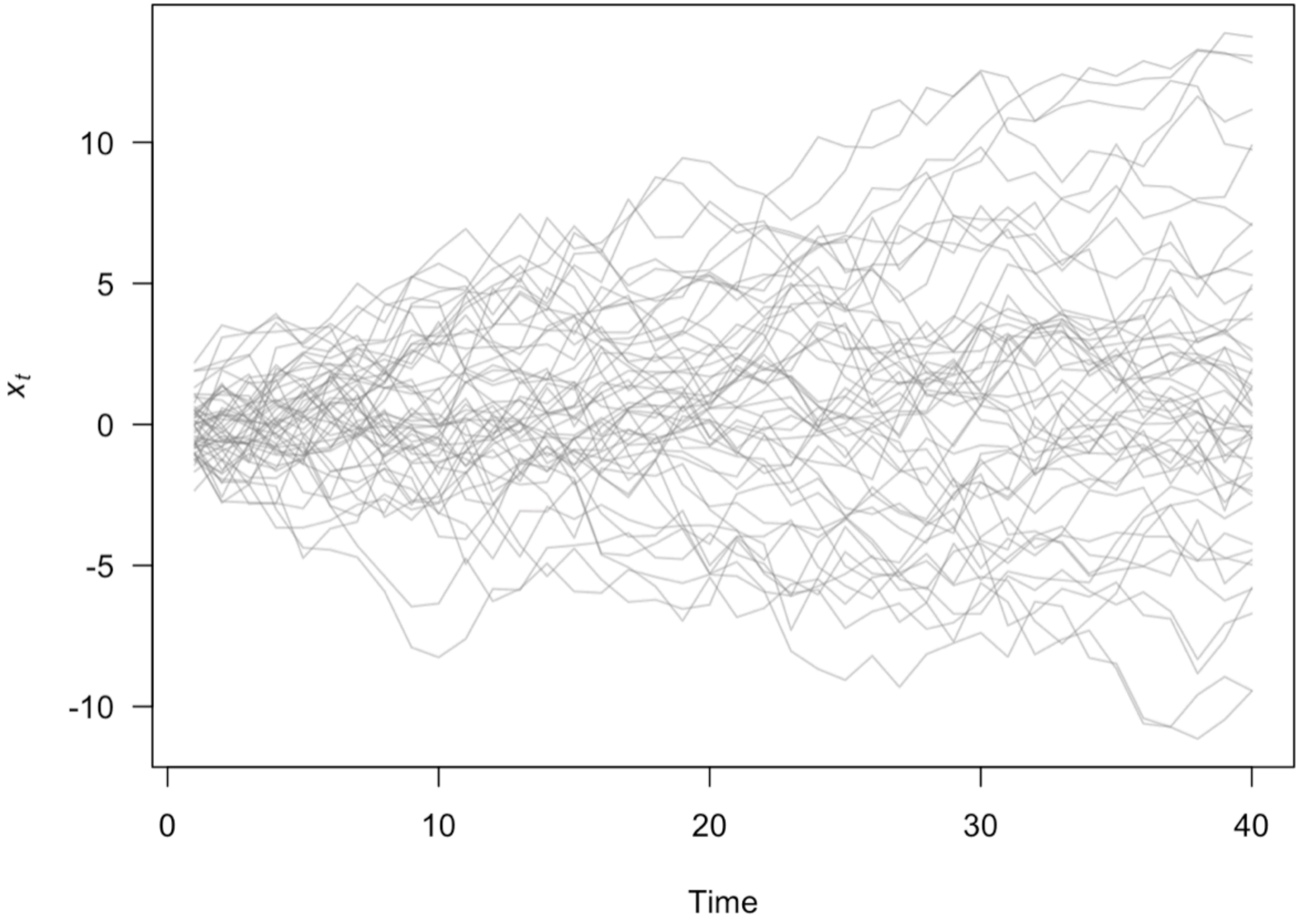
$x_t \sim N(0,1)$



Holmes, E. E., M. D. Scheuerell, and E. J. Ward. Applied time series analysis for fisheries and environmental data.

Random walk model

$$x_t = x_{t-1} + \epsilon_t, \epsilon_t \sim N(0,1)$$



Holmes, E. E., M. D. Scheuerell, and E. J. Ward. Applied time series analysis for fisheries and environmental data.

Επόμενο μάθημα

R εργαστήριο 1