



National and Kapodistrian
UNIVERSITY OF ATHENS

Τεχνικές Ανάλυσης και Πρόβλεψης Τηλεπικοινωνιακών Αγορών

Διάσπαση Χρονοσειρών

Ανάλυση Χρονοσειρών

- › Οι χρονοσειρές μελετώνται κάτω από την υπόθεση, ότι με βάση τη συμπεριφορά στο παρελθόν μπορεί να προβλεφθεί η μελλοντική πορεία της διακύμανσης της παρατηρούμενης μεταβλητής.
- › Οι τιμές των χρονολογικών σειρών που παρατηρούμε είναι το αποτέλεσμα της ταυτόχρονης επίδρασης τεσσάρων διαφορετικών συνιστωσών:
 - Τάση
 - Κυκλικότητα
 - Εποχικότητα
 - Τυχαίες κυμάνσεις

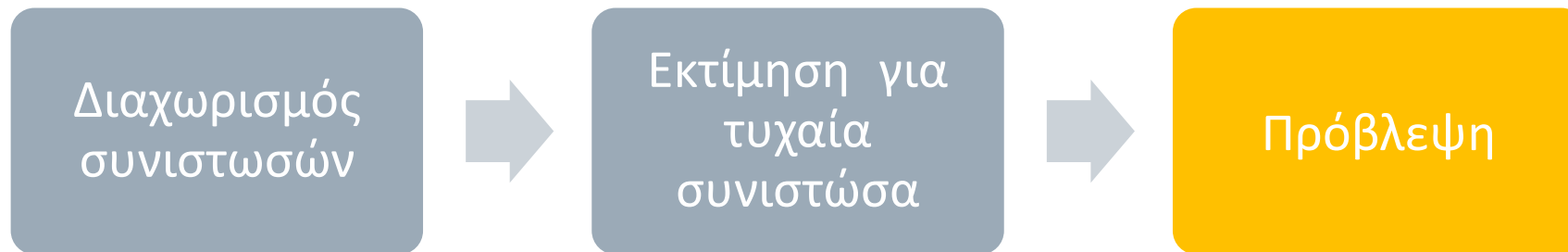
Διάσπαση Χρονοσειρών

- › Η διάσπαση των χρονοσειρών (time series decomposition) στηρίζεται στην υπόθεση, ότι οι τιμές μιας χρονοσειράς σχηματίζονται από τα στοιχεία, που τη συνθέτουν.
- › Η χρονοσειρά διασπάται στα τέσσερα προαναφερόμενα συνθετικά συστατικά της και προσδιορίζεται η επιρροή, η φύση και η έκταση, που έχει καθένα από αυτά τα συστατικά στη διαμόρφωση των τιμών της μεταβλητής.

$$\begin{aligned} \text{Δεδομένα} &= \text{πρότυπο (υπόδειγμα)} + \text{σφάλμα} \\ &= f(\text{συστηματικές συνιστώσες}) + \text{σφάλμα} \end{aligned}$$

Μοντέλα Διάσπασης Χρονοσειρών

- › Υπόθεση: Οι τρεις συστηματικές συνιστώσες και η τυχαία συνιστώσα είναι αποτέλεσμα διαφορετικών αιτίων



Μοντέλα Διάσπασης Χρονοσειρών

Εάν:

- › Y_t η πραγματική παρατήρηση σε χρόνο t ,
- › T_t η τάση της σειράς σε χρόνο t
- › S_t ο εποχικός παράγοντας σε χρόνο t ,
- › C_t η κυκλική συνιστώσα,
- › I_t ο μη – κανονικός παράγοντας

Μοντέλα Διάσπασης Χρονοσειρών

Πολλαπλασιαστικό Μοντέλο

$$Y_t = T_t \cdot S_t \cdot C_t \cdot I_t$$

Προσθετικό Μοντέλο

$$Y_t = T_t + S_t + C_t + I_t$$

- › Το προσθετικό μοντέλο υποθέτει ότι οι συντελεστές μεταξύ τους είναι ανεξάρτητοι.

Προσαρμογή και Ανάλυση της Τάσης

Αξία ανάλυσης της τάσης:

- › Προβολή τάσης στο μέλλον (πρόβλεψη)
- › Εξάλειψη τάσης – μελέτη διακυμάνσεων λοιπών συνιστωσών.
- › Σύνηθες πρώτο βήμα: Γραφική αναπαράσταση χρονοσειράς

Μοντέλα Τάσης

Η τάση υπολογίζεται:

- › Ως μεταβολές και προσέγγιση μέσω ευθύγραμμων τμημάτων, ή
- › Σαν μία μέση καμπύλη σε επίπεδο τιμών με τη μέθοδο των ελαχίστων τετραγώνων

Μοντέλα Τάσης

Τα μοντέλα τάσης βασίζονται σε δύο παράγοντες:

- › Στην επιλογή της κατάλληλης καμπύλης.
- › Στην παραδοχή ότι η καμπύλη, που συνδέει με κάποια σχετική ακρίβεια τις παρελθούσες τιμές, είναι ενδεικτική και για το μέλλον.

Υπάρχουν κάποια αντικειμενικά κριτήρια, μέσω των οποίων μπορεί να επιλεχθεί η κατάλληλη καμπύλη. Αυτά είναι:

- › Το “Akaike Information Criterion (AIC)”
- › Το “Bayesian Information Criterion (BIC)”

Μοντέλα Τάσης

- › Γραμμικό Μοντέλο
 - › Παραβολικό Μοντέλο
 - › Εκθετικό Μοντέλο
 - › Μοντέλο Gompertz
 - › Μοντέλο Pearl Reed Logistic
-
- › Η εφαρμογή των παραπάνω μοντέλων βασίζεται στη μέθοδο των ελαχίστων τετραγώνων.

Γραμμικό Μοντέλο

- › Χρησιμοποιείται πολύ συχνά ως κριτήριο για τον προσδιορισμό του μοντέλου με την καλύτερη προσαρμογή.
- › Ο αναλυτής χαράζει απλώς μια γραμμή διαμέσου της σειράς των παρατηρήσεων, προκειμένου να αποκαλύψει το σχήμα και την κατεύθυνση της τάσης της χρονοσειράς.

Γραμμικό Μοντέλο

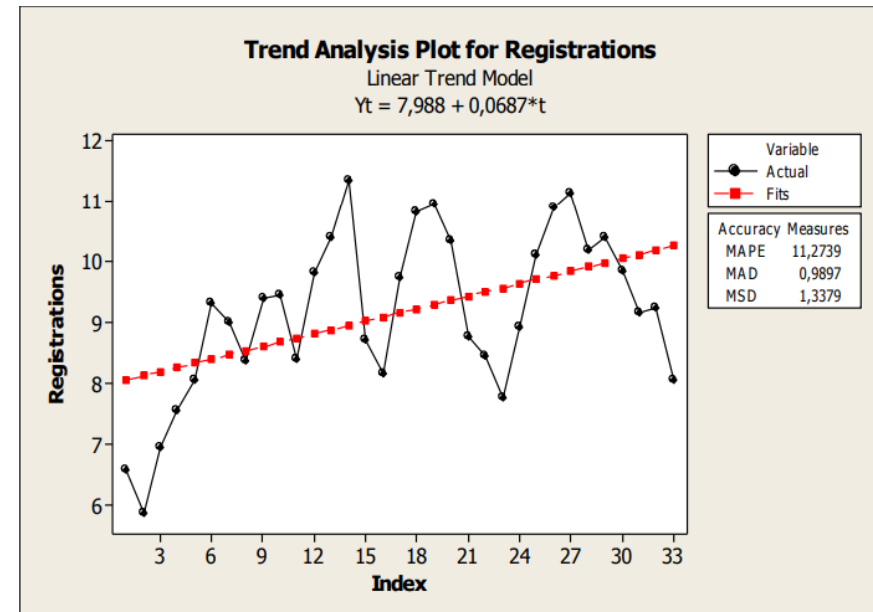
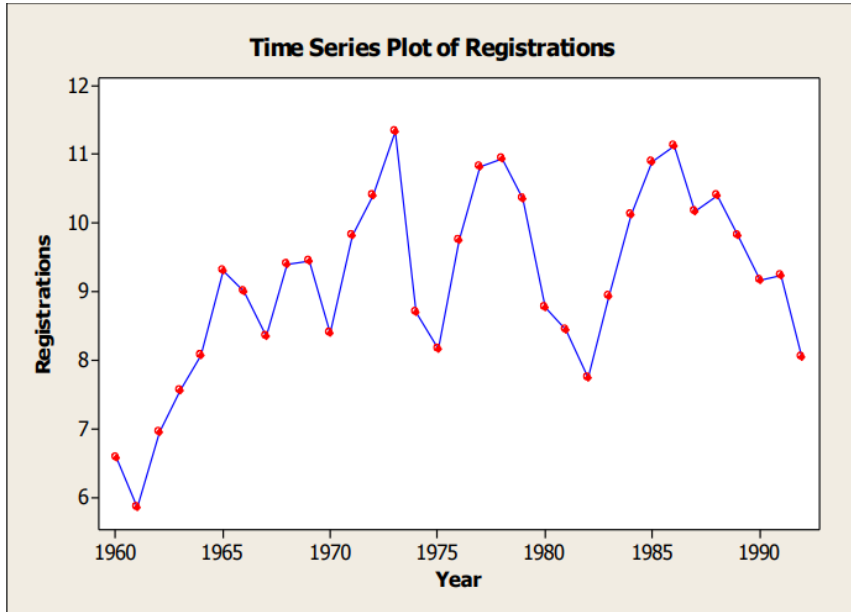
$$\hat{Y}_t = a + bt \text{ με}$$

$$a = \frac{\sum y \sum t^2 - \sum t \sum ty}{n \sum t^2 - (\sum t)^2} \text{ και}$$

$$b = \frac{n \sum ty - \sum t \sum y}{n \sum t^2 - (\sum t)^2}$$

- › \hat{Y}_t είναι η προβλεφθείσα τιμή τάσης για τη μεταβλητή Y για μια επιλεγείσα χρονική περίοδο t .
- › a είναι η τιμή της τάσης, όταν $t=0$
- › b είναι η κλίση της ευθείας (η μεταβολή στο Y που σχετίζεται με μια μοναδιαία μεταβολή στη t).

Γραμμικό Μοντέλο

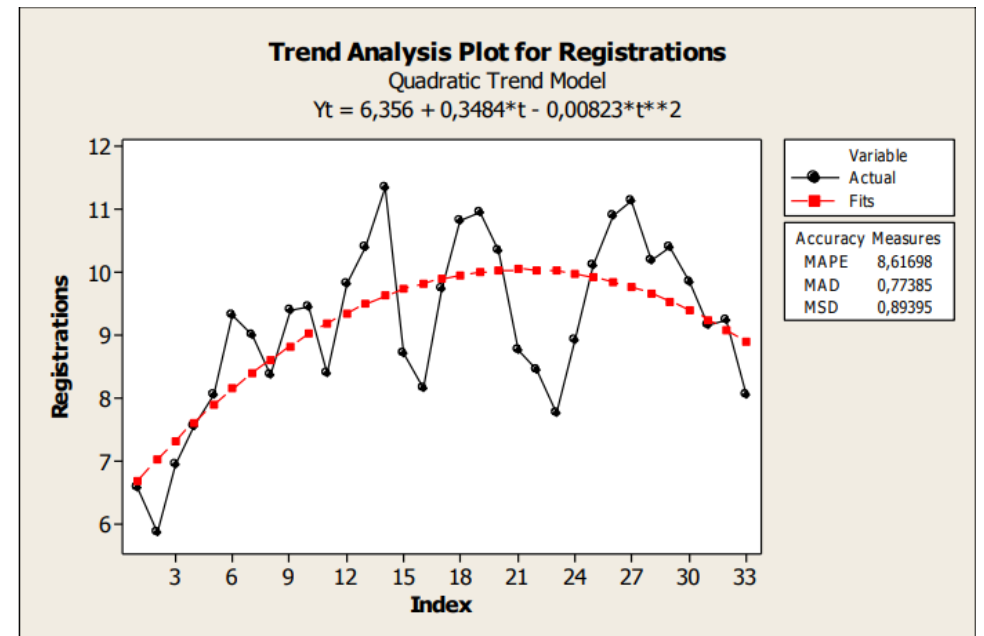


Παραβολικό Μοντέλο

› Βασίζεται στην παραβολή ελαχίστων τετραγώνων που ορίζεται από τις σχέσεις”

$$y = a + b x + c x^2, \quad \sum y = n a + b \sum x + c \sum x^2,$$

$$\sum x y = a \sum x + b \sum x^2 + c \sum x^3 \quad \sum x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4$$



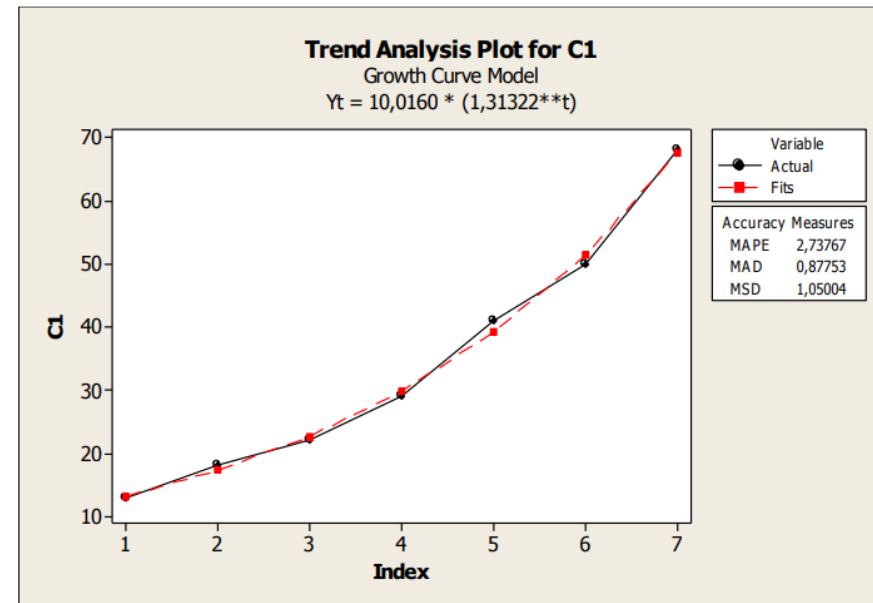
Growth Curves

- › Οι Growth Curves αναπαριστούν τάσεις που η ταχύτητα μεγέθυνσης (growth) μεταβάλλεται ανάλογα με τη φάση ωρίμασης.
- › Εκθετικό Μοντέλο
- › Μοντέλο Gompertz
- › Μοντέλο Pearl Reed Logistic

Εκθετικό Μοντέλο

› Η καμπύλη είναι της μορφής:

$$\hat{Y}_t = b_0 b_1^t$$

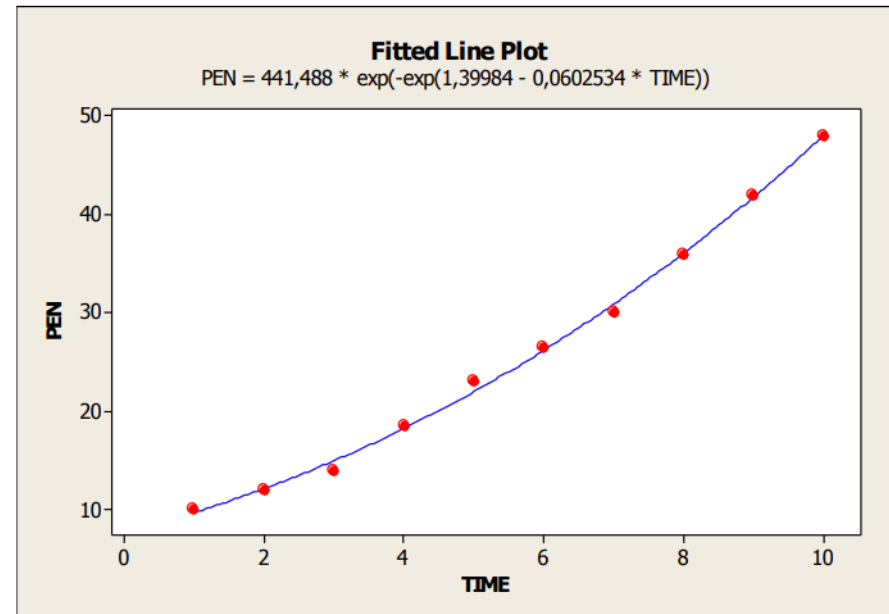


› Εάν κάναμε γραμμική προσέγγιση στα αρχικά στάδια θα είχαμε υπερεκτίμηση και στα τελικά στάδια υποεκτίμηση των αυξήσεων.

Μοντέλο Gompertz

- › Η καμπύλη Gompertz περιγράφεται με την ακόλουθη σχέση:

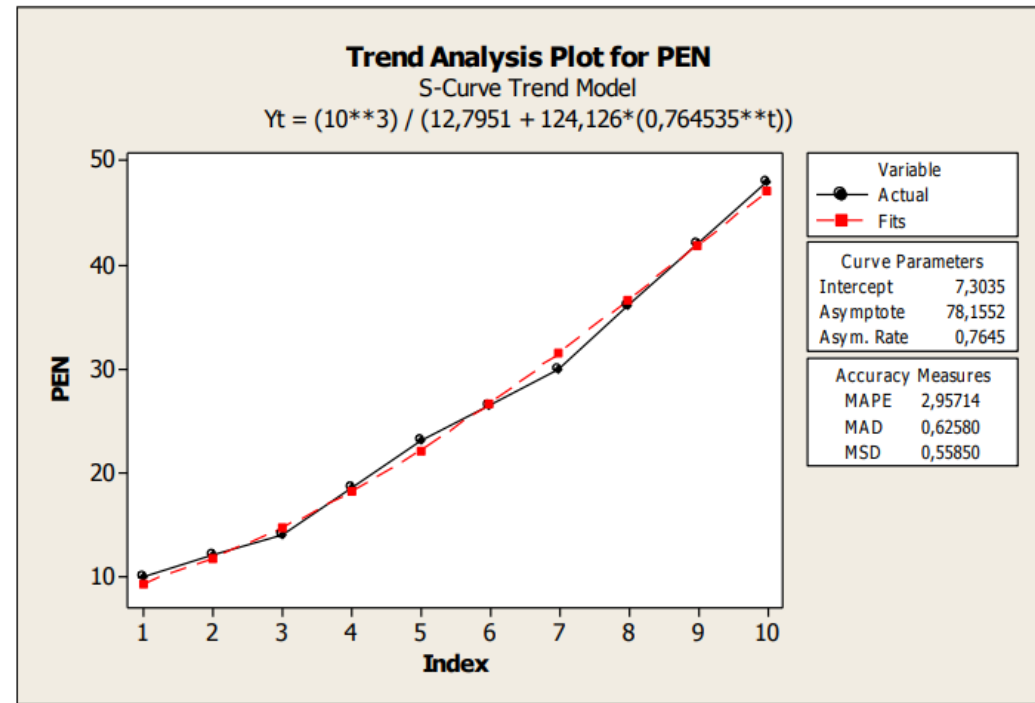
$$\hat{Y}_t = \theta_1 \exp(-\exp(\theta_2 - \theta_3 t))$$



Μοντέλο Pearl Reed Logistic

› Η καμπύλη ἸΦΤΡΒΪ ΦΦΝ ΧΘЦΨΗΨ περιγράφεται με την ακόλουθη σχέση

$$Y_t = \frac{1000}{\beta_0 + \beta_1 \beta_2^t}$$



Επιλογή Best Fit Μοντέλου

- › Ένα σπουδαίο ρόλο στην μελέτη χρονοσειρών, γενικώς, παίζουν τα κατάλοιπα (residuals).
- › Το κατάλοιπο είναι η διαφορά ανάμεσα στο πραγματικό αποτέλεσμα και την πρόβλεψη του αποτελέσματος, δηλαδή, η διαφορά των εκτιμήσεων από τις παρατηρήσεις.

Επιλογή Best Fit Μοντέλου

› Το κατάλοιπο υπολογίζεται ως εξής:

$$e_t = Y_t - \hat{Y}_t$$

όπου,

- e_t = το σφάλμα πρόβλεψης τη χρονική στιγμή t .
- Y_t = η πραγματική τιμή τη χρονική στιγμή t .
- \hat{Y}_t = η προβλεπόμενη τιμή τη χρονική στιγμή t .

Επιλογή Best Fit Μοντέλου

Για την αξιολόγηση της προβλεπτικής ικανότητας του υποδείγματος, έχουν προταθεί διάφορα κριτήρια, από τα οποία τα κυριότερα είναι:

- › Root Mean Square Error
- › Mean Absolute Deviation
- › Mean Square Error
- › Mean Absolute Percentage Error
- › Mean Percentage Error

Επιλογή Best Fit Μοντέλου

Μέση Απόλυτη Απόκλιση

Mean Absolute Deviation, MAD

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

Επιλογή Best Fit Μοντέλου

Μέσο Τετραγωνικό Σφάλμα
Mean Square Error, MSE

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

Επιλογή Best Fit Μοντέλου

Μέσο Απόλυτο Ποσοστιαίο Σφάλμα

Mean Absolute Percentage Error, MAPE

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

Επιλογή Best Fit Μοντέλου

Μέσο Ποσοστιαίο Σφάλμα

3 ФТБ ЙФЯЎФЪПЦФ ГЯӨЯ τ 3 ЙГ

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}$$

Επιλογή Best Fit Μοντέλου – Εφαρμογή

Χρόνος	Πραγματική Τιμή	Πρόβλεψη
1	58	-
2	54	58
3	60	54
4	55	60
5	62	55
6	62	62
7	65	62
8	63	65
9	70	63

Επιλογή Best Fit Μοντέλου – Εφαρμογή

Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} = \sqrt{\frac{188}{8}} = 4.8$$

Mean Absolute Deviation

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| = \frac{34}{8} = 4.3$$

Mean Squared Error

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \frac{188}{8} = 23.5$$

Mean Absolute Percentage Error

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} = \frac{0.556}{8} = 0.0695 = 6.95 \%$$

Mean Percentage Error

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} = \frac{0.162}{8} = 0.0203 = 2.03 \%$$

Επιλογή Best Fit Μοντέλου – Εφαρμογή

- › Το MAD μας λέει ότι κατά μέσον όρο θα έχουμε απόκλιση 4.3 πελάτες ανά πρόβλεψη.
- › Τα MSE και MAPE συγκρίνονται με τα αντίστοιχα άλλων μεθόδων, προκειμένου να επιλέξουμε τη πιο ακριβή μέθοδο.

Ανάλυση και Διαχωρισμός της Εποχικότητας

Η γνώση των εποχικών κινήσεων επιτρέπει:

- › Τη σχεδίαση του επιχειρηματικού προγράμματος
- › Την ανάλυση των επιδόσεων του παρελθόντος
- › Τη μεταβολή του εποχικού μοντέλου

Ανάλυση και Διαχωρισμός της Εποχικότητας

Η ταυτοποίηση του εποχικού παράγοντα σε μία χρονοσειρά διαφέρει από την ανάλυση τάσης σε τουλάχιστον δύο σημεία:

- › Η τάση ορίζεται απευθείας από τα δεδομένα, αλλά ο εποχικός παράγοντας ορίζεται άμεσα μετά την εξάλειψη των άλλων παραγόντων και αφήνοντας μόνο τον εποχικό.
- › Η τάση αναπαριστάται με ένα “best fitting curve”, ενώ η εποχικότητα θα πρέπει να υπολογιστεί για κάθε παρατηρούμενο διάστημα του χρόνου (εβδομάδα, μήνας κλπ) και γράφεται συνήθως σε μορφή ενός δείκτη.
 - Δείκτες είναι ποσοστά που δείχνουν αλλαγές στο χρόνο.

Ανάλυση και Διαχωρισμός της Εποχικότητας

- › Υπάρχουν διάφορες μέθοδοι για τη μέτρηση της εποχικής διακύμανσης
- › Μέθοδος του *λόγου της πραγματικής τιμής προς τον κινητό μέσο (ratio - to - moving average)*

Κινητός Μέσος

- › Η ονομασία κινητός μέσος όρος χρησιμοποιείται για να περιγράψει την διαδικασία, κατά την οποία καθώς μια νέα παρατήρηση γίνεται διαθέσιμη, ένας νέος μέσος όρος μπορεί να υπολογιστεί, στον οποίο παραλείπεται η πιο παλιά παρατήρηση, προκειμένου να συμπεριληφθεί η πιο πρόσφατη.
- › Ο αριθμός των δεδομένων, που χρησιμοποιούνται στον υπολογισμό του μέσου όρου, παραμένει σταθερός

Κινητός Μέσος

- › Ως κινητός μέσος τάξης Ψ είναι η μέση τιμή των Ψ διαδοχικών παρατηρήσεων και υπολογίζεται από την ακόλουθη σχέση

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$$

όπου

- › $H_{\tau\sigma\ddot{u}}$ η πρόβλεψη για την επόμενη περίοδο
- › H_{τ} η πραγματική τιμή
- › Ψ ο αριθμός των παραγόντων του κινητού μέσου

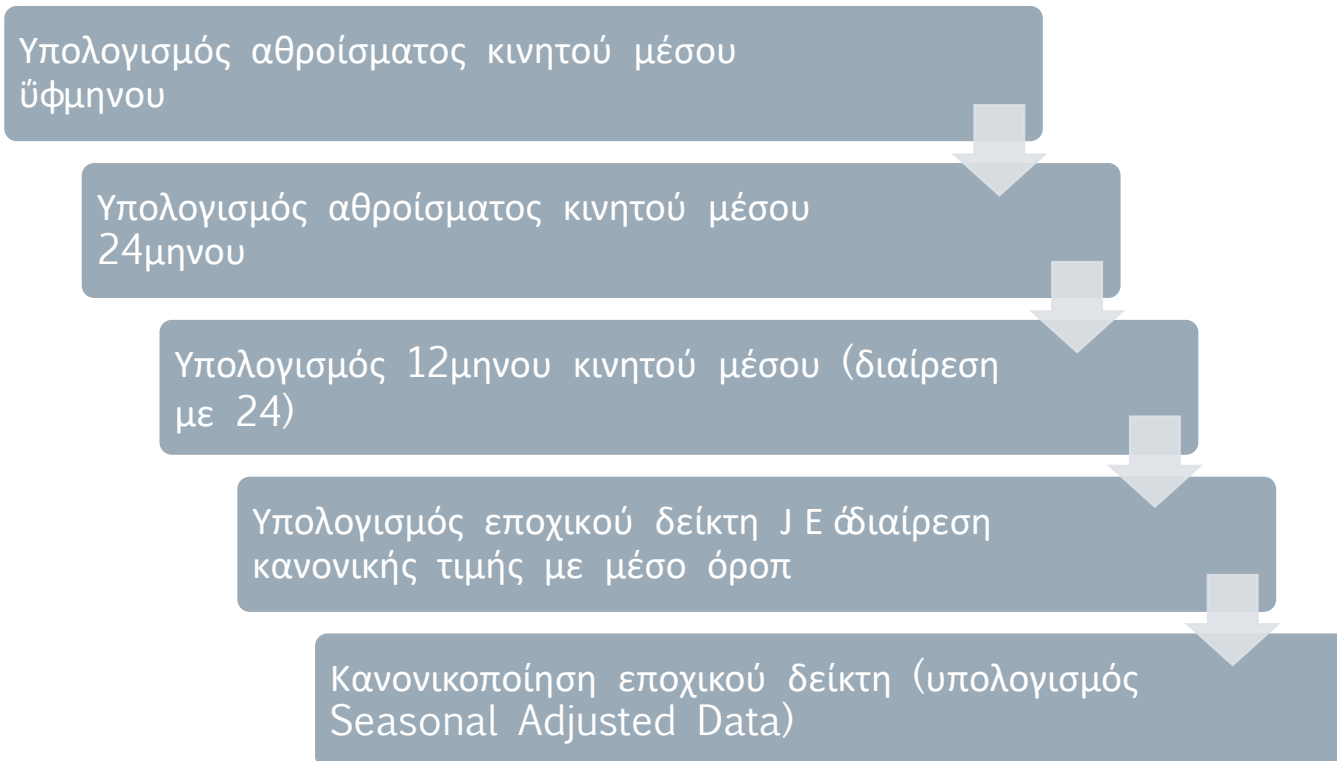
ï ΤΤΘ : ΤΘ : 3 ΘΑΥЦ , ΑΦЯГ ЦФ

- › Ο όρος ratio – to – moving – average χρησιμοποιείται επειδή παίρνουμε τη μεταβλητή Y_t της χρονοσειράς και τη διαιρούμε με τον κινητό μέσο όρο.
- › Η ποσότητα που βρίσκουμε, η S_t είναι η καλύτερη δυνατή μέτρηση της εποχικής συνιστώσας στα δεδομένα της χρονοσειράς.

$$\frac{Y_t}{MA} = \frac{S_t \times C_t \times T_t \times I_t}{C_t \times T_t \times I_t}$$

Ratio – to – Moving Average

- › Εάν υποθέσουμε ότι η χρονική περίοδος που εξετάζεται η κυκλικότητα είναι ένας (1) χρόνος:



Ratio – to – Moving Average - Εφαρμογή

Μήνας	2024	2025
Ιαν	518	613
Φεβ	404	392
Μαρ	300	273
Απρ	210	322
Μαι	196	189
Ιουν	186	257
Ιουλ	247	324
Αυγ	343	404
Σεπ	464	677
Οκτ	680	858
Νοε	711	895
Δεκ	610	664

Ratio – to – Moving Average - Εφαρμογή

Κανονικοποίηση SI

- › Για κάθε μήνα και για όλες τις διαθέσιμες τιμές όλων των ετών βρίσκουμε τον διάμεσο.
- › Προσθέτουμε τους διάμεσους του εποχικού δείκτη και βρίσκουμε έναν πολλαπλασιαστή ως ακολούθως:

$$\text{Multiplier} = \frac{12}{\sum_{i=1}^{12} (\text{Median})_i}$$

- › Πολλαπλασιάσουμε τον κάθε συντελεστή εποχικότητας με τον πολλαπλασιαστή.

Ratio – to – Moving Average - Εφαρμογή

Μήνας	2010	2011	2012	2013	2014	2015
Ιαν		1,208	1,202	1,272	1,411	1,431
Φεβ		0,7	0,559	0,938	1,089	0,903
Μαρ		0,524	0,564	0,785	0,8	0,613
Απρ		0,444	0,433	0,480	0,552	0,697
Μαι		0,424	0,365	0,488	0,503	0,396
Ιουν		0,49	0,459	0,461	0,465	0,528
Ιουλ	0,639	0,904	0,598	0,681	0,603	0,662
Αυγ	1,115	0,913	0,889	0,830	0,830	0,830
Σεπ	1,371	1,560	1,346	1,128	1,128	1,395
Οκτ	1,792	1,863	1,796	1,638	1,638	1,771
Νοε	1,884	2,012	2,012	1,695	1,695	1,846
Δεκ	1,519	1,088	1,224	1,282	1,445	

Ratio – to – Moving Average - Εφαρμογή

Μήνας	Median	Normal SI
Ιαν	1,272	1,278
Φεβ	0,903	0,907
Μαρ	0,613	0,616
Απρ	0,480	0,482
Μαι	0,424	0,426
Ιουν	0,465	0,467
Ιουλ	0,651	0,654
Αυγ	0,860	0,864
Σεπ	1,359	1,365
Οκτ	1,782	1,790
Νοε	1,857	1,865
Δεκ	1,282	1,288
SUM	11,948	12,002

Απομόνωση της Κυκλικότητας

- › Η διαδικασία απομόνωσης της κυκλικότητας είναι παρόμοια με τη διαδικασία απομόνωσης της εποχικής συνιστώσας

Υπολογισμός εποχικού δείκτη μέσω Ratio-to-Moving Average

Διαιρώντας τις πραγματικές τιμές με τον εποχικό δείκτη παραμένουν οι τιμές TCI

Υπολογισμός τάσης των τιμών TCI

Χρήση κινητών μέσων για υπολογισμό της κυκλικής συνιστώσας

Απομόνωση του I

Απομόνωση της Κυκλικότητας

- › Η απομόνωση μιας κυκλικής συνιστώσας επιτυγχάνεται με την απάλειψη της εποχικότητας από τα αρχικά δεδομένα ως εξής:

$$\frac{Y_t}{\text{εποχικός δείκτης}} = \frac{S_t \times C_t \times T_t \times I_t}{S_t} = C_t \times T_t \times I_t$$

Απομόνωση της Κυκλικότητας

- › Τα δεδομένα αυτά, τα οποία είναι απαλλαγμένα από την εποχικότητα, μπορούν να απαλλαγούν από την τάση ως εξής:

$$\frac{C_t \times T_t \times I_t}{T_t} = C_t \times I_t$$

- › Οι σειρές που προκύπτουν περιλαμβάνουν μόνο την επίδραση που σχετίζεται με την κυκλική συνιστώσα και τη μη - κανονική διακύμανση.

Απομόνωση της Κυκλικότητας

- › Η μη – κανονική διακύμανση μπορεί να εξαλειφθεί από τη χρονοσειρά μέσω υπολογισμού του κινητού μέσου όρου, που εξομαλύνει αυτές τις μη – κανονικές διακυμάνσεις από τα δεδομένα.
 - Συνήθως παίρνουμε περιττό αριθμό χρονικών περιόδων για τον κινητό μέσο.
 - Για μηνιαία δεδομένα αρκεί κινητός μέσος 5, 7, 9 ή και 11 χρονικών περιόδων, ενώ για τριμηνιαία δεδομένα αρκεί κινητός μέσος 3 περιόδων.

Απομόνωση της Κυκλικότητας

› Η συνιστώσα τυχαιότητας υπολογίζεται ως εξής:

$$I_t = \frac{C_t \times I_t}{C_t}$$

Απομόνωση της κυκλικότητας

$$Y = 232,7$$

$$T = 261,24 + 0,759t$$

$$SI = 0,77967$$

$$CI = 1,118$$

$$\hat{T}_t = 261.24 + 0.759t \quad SCI = \frac{Y}{T} = \frac{232.7}{262.0} = 0.888$$

$$TCI = \frac{232.7}{.77967} = 298.458$$

$$CI = \frac{Y}{TS} = \frac{232.7}{(262.0)(.77967)} = 1.139$$

$$I = \frac{CI}{C} = \frac{1.159}{1.118} = 1.036$$

Πρόβλεψη βάσει Διάσπασης

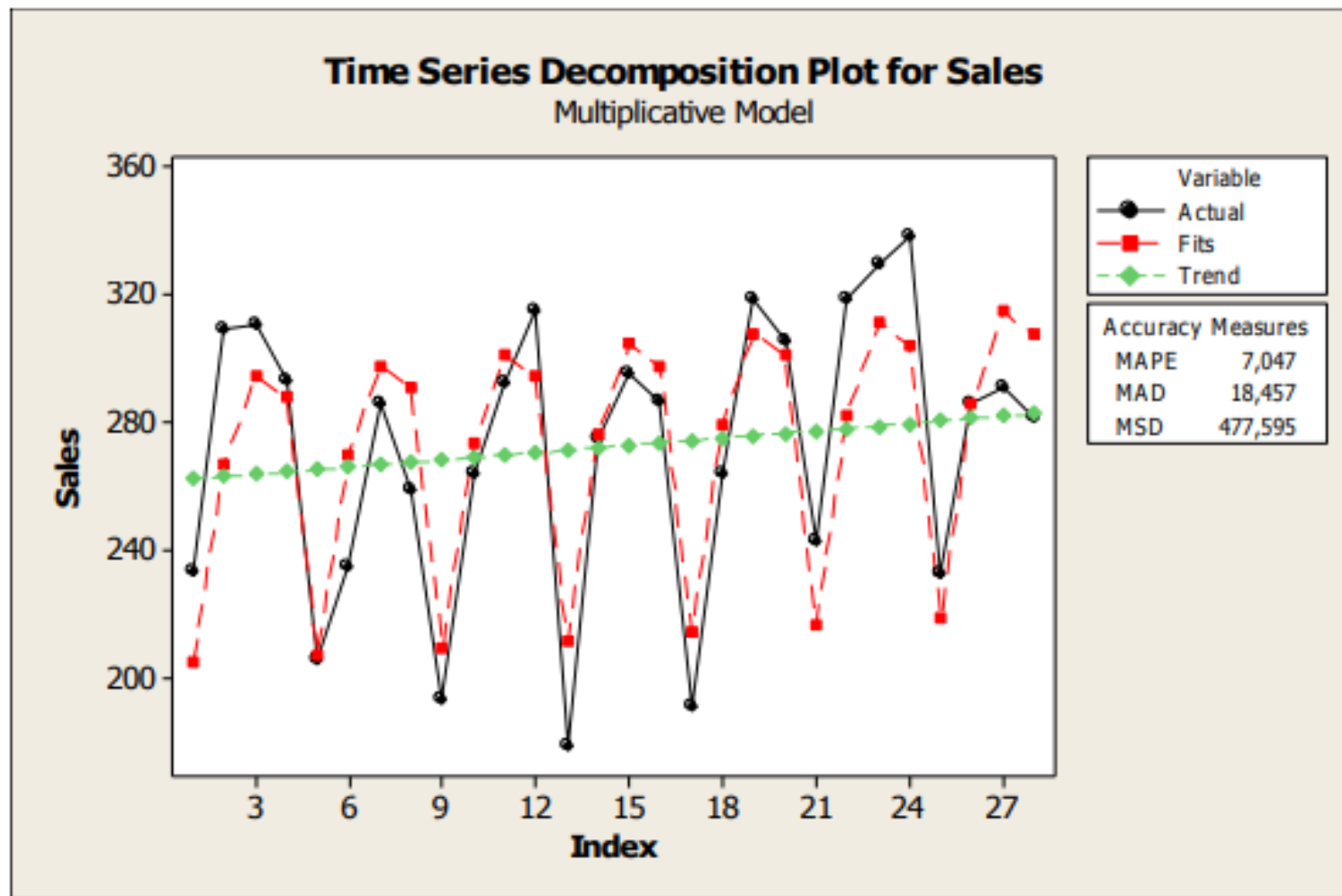
› Για να κάνουμε πρόβλεψη:

1. Βρίσκουμε τη τιμή της τάσης από την ευθεία ελαχίστων τετραγώνων
2. Παίρνουμε τη τιμή του εποχικού δείκτη S για το πρώτο τρίμηνο
3. Πολλαπλασιάζουμε μαζί με τον κυκλικό δείκτη C , που είναι περίπου 1 στη περίπτωση μας και με το δείκτη τυχαιότητας I πάλι 1, επειδή δεν ξέρουμε κάτι γι' αυτόν.

$$\text{Οπότε: } \hat{T}_{29} = 261.24 + 0.759(29) = 283.251$$

$$\hat{Y}_{29} = T_{29} \times S_{29} \times C_{29} \times I_{29} = (283.251)(.77967)(1.0)(1.0) = 220.842$$

Πρόβλεψη βάσει Διάσπασης



Επιχειρηματικοί Δείκτες

- › Κυκλικοί δείκτες χρησιμοποιούνται συνήθως για να απαντήσουμε τις ακόλουθες ερωτήσεις:
 - Εμφανίζει κυκλικότητα η προς έρευνα χρονοσειρά;
 - Εάν ναι πόσο έντονη είναι αυτή;
 - Ακολουθεί η χρονοσειρά την γενικότερη κατάσταση της οικονομίας (Επιχειρηματικοί Δείκτες);

- › Ένας τρόπος να ερευνήσουμε τα κυκλικά σχήματα στις χρονοσειρές είναι να μελετήσουμε και να συγκρίνουμε αυτά με τα σχήματα επιχειρηματικών δεικτών.

Επιχειρηματικοί Δείκτες

- › Ως επιχειρηματικός δείκτης ορίζεται μία χρονοσειρά, συσχετιζόμενη με κάποια επιχειρηματικότητα και χρησιμοποιούνται για την εκτίμηση της γενικής κατάστασης της οικονομίας.

Επιχειρηματικοί Δείκτες

Οι δείκτες αυτοί χωρίζονται σε τρεις κατηγορίες:

- › *Leading indicators*: Τους χρησιμοποιούμε για να ανιχνεύσουμε σημεία καμψής στην οικονομία και την μελλοντική κατεύθυνση της οικονομίας. Π.χ. Ο δείκτης έκδοσης των οικοδομικών αδειών και ο δείκτης μετοχών του χρηματιστηρίου.
- › *Coincident indicators*: Τους χρησιμοποιούμε για να κατανοήσουμε την τρέχουσα κατάσταση της οικονομίας. Π.χ. Ο δείκτης βιομηχανικής παραγωγής, ο δείκτης λιανικού εμπορίου κλπ.
- › *Lagging indicators*: Αυτοί οι δείκτες έχουν μία χρονοκαθυστέρηση ως προς την κατάσταση της οικονομίας. Π.χ. βασικό επιτόκιο, ποσοστό ανεργίας.

Επιχειρηματικοί Δείκτες

- › Οι Leading indicators αλλάζουν κατεύθυνση στα σημεία καμψής της γενικής οικονομίας.
- › Οι Coincident indicators αυξομειώνονται ταυτόχρονα με τις αυξήσεις, ή μειώσεις της γενικής οικονομίας.
- › Οι Lagging indicators ακολουθούν την πορεία της γενικής οικονομίας με κάποια καθυστέρηση.

Επιχειρηματικοί Δείκτες

- › Είναι δύσκολο να ταυτοποιήσουμε σημεία καμπής, γιατί δεν αναπτύσσονται όλοι οι κλάδοι της οικονομίας με τον ίδιο τρόπο, ή τον ίδιο ρυθμό.
- › Οι Leading indicators σηματοδοτούν συνήθως τις αλλαγές της οικονομίας, οπότε είναι οι σημαντικότεροι.
- › Οι Coincident και Lagging indicators μας βοηθούν να εκτιμήσουμε την απόδοση της τωρινής και παρελθούσας οικονομίας σε συγκεκριμένους κλάδους και να βοηθήσουν στην ανάπτυξη μελλοντικών πολιτικών.

CENSUS II X-13ARIMA-SEATS

- › Αποτελεί μέθοδο που βασίζεται στις επαναλήψεις βημάτων προκειμένου να αποσυνθέσει τις χρονοσειρές στα συστατικά τους.
- › Η μέθοδος περιλαμβάνει 12 βήματα.

Ερωτήσεις???

