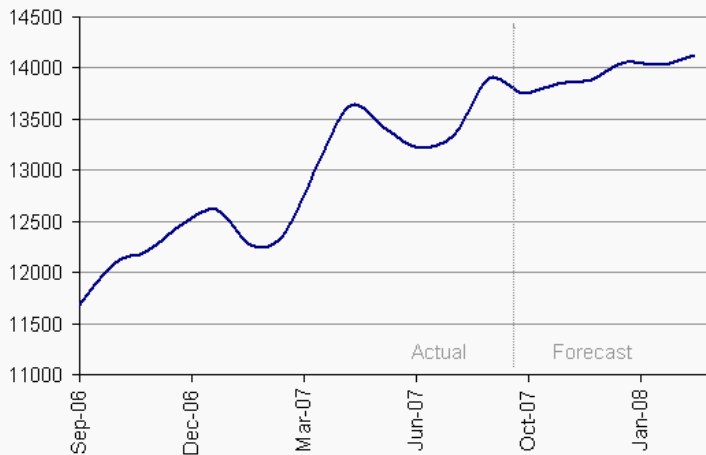


ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Μονάδα Προβλέψεων & Στρατηγικής
Forecasting & Strategy Unit

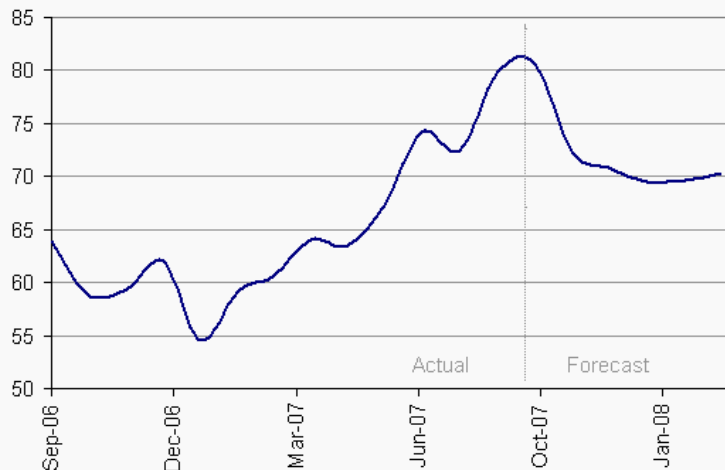


Εισαγωγή στις Μεθόδους Προβλέψεων

Πρόβλεψη



Η μεγαλύτερη **πρόκληση** στην ανάλυση χρονοσειρών είναι η **πρόβλεψη**, δηλαδή το να εκτιμήσει κανείς πώς η ακολουθία των παρατηρήσεων θα συνεχιστεί στο μέλλον.



Ζητούμενο είναι να ακολουθήσει κανείς μια **διαδικασία** που θα εξασφαλίσει ότι θα παραχθούν όσον τον δυνατόν πιο **ακριβείς** προβλέψεις, **αξιοποιώντας** στο έπακρο όλη την διαθέσιμη **ιστορική πληροφορία**.

Στατιστική Πρόβλεψη

03



Κριτική Πρόβλεψη

Είδη Προβλέψεων

Πρόβλεψη Προϋπολογισμού

04



Τελική Πρόβλεψη

Είδη Προβλέψεων

Στατιστική Πρόβλεψη

Κυριότερες Στατιστικές Μέθοδοι Πρόβλεψης

05

Naïve

Η πιο απλή στατιστική μέθοδος. Δεν παράγει ακριβείς προβλέψεις αλλά πολλές φορές χρησιμοποιείται ως benchmark για άλλες μεθόδους. Η πρόβλεψη θεωρείται πως είναι ίση με την τελευταία παρατήρηση της διαθέσιμης χρονοσειράς.



$$F_{t+1} = y_T$$

R-code: `naive(y, h)`

Στατιστική Πρόβλεψη

Κυριότερες Στατιστικές Μέθοδοι Πρόβλεψης

06

Seasonal Naïve

Πρόκειται για μία παραλλαγή της naïve και είναι χρήσιμη για δεδομένα με έντονη εποχιακότητα. Η πρόβλεψη θεωρείται πως είναι ίση με την τελευταία παρατήρηση της διαθέσιμης χρονοσειράς από την αντίστοιχη εποχιακή περίοδο



$$F_{t+1} = y_{T+h-km}$$

m = seasonal period,

$$k = \lfloor (h-1)/m \rfloor + 1,$$

R-code: `snaive(y, h)`

Κινητοί Μέσοι Όροι για πρόβλεψη

Περίοδος	Δεδομένα	ΚΜΟ(3)	ΚΜΟ(5)
1	106,5		
2	109,2		
3	117,8		
4	117,2	111,17	
5	116,9	114,73	
6	118,7	117,30	113,52
7	115,6	117,60	115,96
8	119,0	117,07	117,24
9	134,7	117,77	117,48
10	130,4	123,10	120,98
11	126,2	128,03	123,68
12		130,43	125,18

$$F_{t+1} = \frac{1}{k} \cdot \sum_{i=t-k+1}^t Y_i$$



R-code: `ma(y, h, drift=TRUE)`

Στατιστική Πρόβλεψη

Κυριότερες Στατιστικές Μέθοδοι Πρόβλεψης

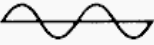
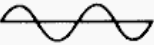
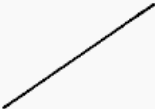
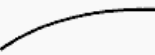
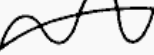
Μέθοδοι Εκθετικής Εξομάλυνσης

Αναπτύχθηκαν τις αρχές της δεκαετίας του '50. Από τότε έγιναν από τις πιο δημοφιλείς μεθόδους προβλέψεων μεταξύ των επιχειρηματιών κυρίως λόγω της ευκολίας χρήσης τους, της ελάχιστης απαίτησης σε υπολογιστικό χρόνο και την απαίτηση σχετικά λίγων παρατηρήσεων προκειμένου να παράγουν προβλέψεις. Οι μέθοδοι εξομάλυνσης είναι κατάλληλες για **βραχυπρόθεσμες** προβλέψεις ενός μεγάλου όγκου χρονοσειρών. Αποδίδουν καλύτερα σε δεδομένα που παρουσιάζουν στασιμότητα ή μικρό ρυθμό ανάπτυξης ή μείωσης ως προς το χρόνο.

Κυριότερες μέθοδοι εξομάλυνσης:

- ✓ Simple Exponential Smoothing
- ✓ Holt
- ✓ Damped
- ✓ Winter

Τύποι Μοντέλων Εξομάλυνσης

	Nonseasonal	Additive Seasonality	Multiplicative Seasonality
Constant Level			
Linear Trend			
Exponential Trend			
Damped Trend			

• Σταθερού Επιπέδου

- ✓ Για βραχυπρόθεσμες προβλέψεις
- ✓ Για χρονοσειρές που εμπεριέχουν υψηλό θόρυβο και τυχαιότητα

• Γραμμικής τάσης

- ✓ Για σταθερή αύξηση τους μεγέθους στο μέλλον

• Εκθετικής τάσης

- ✓ Για εκθετική αύξηση τους μεγέθους στο μέλλον (π.χ. στις αρχές του κύκλου ζωής ενός προϊόντος)
- ✓ Είναι υπεραισιόδοξες για μακροπρόθεσμες προβλέψεις

• Φθίνουσας τάσης

- ✓ Για μεσοπρόθεσμες προβλέψεις.

Μοντέλο Σταθερού Επιπέδου – SES

Χρόνος t	Δεδομένα X_t	Πρόβλεψη F $\hat{X}_{t-1}(1)$	Σφάλμα e_t	Επίπεδο στο τέλος της περιόδου t $S_t = S_{t-1} + h_1 e_t$
0				$S_0 = 585$
1	545	585	-40	$S_1 = 585 + 0.4(-40) = 569$
2	635	569	66	$S_2 = 569 + 0.4(6.6) = 595.4$
3	420	595,4	-175,4	$S_3 = 595.4 + 0.4(-175.4) = 525.2$
4	716	525,2	190,8	$S_4 = 525.2 + 0.4(190.8) = 601.5$
5	699	601,5	97,5	$S_5 = 601.5 + 0.4(97,5) = 640.5$
6	681	640,5	40,5	$S_6 = 640,5 + 0.4(40.5) = 656.5$
7	763	656,5	106,3	$S_7 = 656.5 + 0.4(106.3) = 699.2$
8	778	699,2	78,8	$S_8 = 699.2 + 0.4(78.8) = 730.5$
9	690	730,5	-40,5	$S_9 = 730.5 + 0.4(-40.5) = 711.4$
10	707	711,4	-7,4	$S_{10} = 711.4 + 0.4(-7.4) = 711.5$
11	716	711,5	4,5	$S_{11} = 711.5 + 0.4(4.5) = 713.3$
12		713,3		

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \alpha \cdot e_t$$

$$F_{t+1} = S_t$$

Μοντέλο Σταθερού Επιπέδου – SES (I)

Εξίσωση Σφάλματος

$$e = Y_{t-1} - F_{t-1}$$

Εξίσωση Επιπέδου & Πρόβλεψης

$$F_t = F_{t-1} + \alpha e$$

$$F_t = F_{t-1} + \alpha(Y_{t-1} - F_{t-1})$$

$$F_t = \alpha Y_{t-1} + (1-\alpha)F_{t-1}$$

Μοντέλο Σταθερού Επιπέδου – SES (II)

Εξίσωση Σφάλματος

$$e = Y_{t-1} - F_{t-1}$$

$$F_t = \alpha Y_{t-1} + (1-\alpha)F_{t-1}$$

Εξίσωση Επιπέδου
& Πρόβλεψης

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + (1-\alpha)^2 F_{t-1}$$

Ομοίως αντικαθιστώντας στην (3) το F_{t-1} , κ.ο.κ. , προκύπτει:

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \alpha(1-\alpha)^3 Y_{t-3} + \alpha(1-\alpha)^4 Y_{t-4} + \dots \\ \dots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t F_1$$

Μοντέλο Σταθερού Επιπέδου – SES (III)

Από την εξίσωση (4) παρατηρούμε

- Ότι οι συντελεστές (βάρη) των ιστορικών δεδομένων Y **μειώνονται** εκθετικά, εξού και το όνομα της μεθόδου «*εκθετική εξομάλυνση*».
- Ότι ο τελευταίος όρος είναι ο $(1-\alpha)^t F_1$. Αυτό σημαίνει ότι *η αρχική πρόβλεψη παίζει ρόλο σε όλες τις επόμενες προβλέψεις*. Στο παράδειγμα μας υπολογίζονται τα βάρη για $t = 11$, ισχύει :
 - $(1-\alpha)^t = 0.3138$ αν $\alpha = 0.1$
 - $(1-\alpha)^t = 0.0004$ αν $\alpha = 0.5$
 - $(1-\alpha)^t = 0.0000$ αν $\alpha = 0.9$

Μοντέλο Σταθερού Επιπέδου – SES (IV)

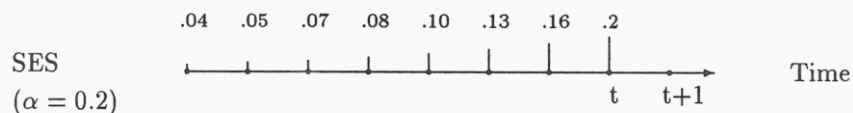
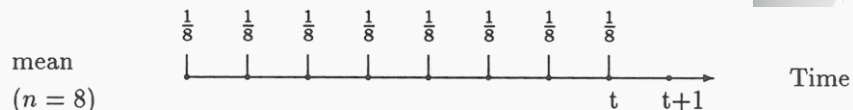
- ✓ Όσο **μικρότερη τιμή του α** επιλέξουμε τόσο **μεγαλύτερο** ρόλο παίζει η **πρώτη** τιμή της πρόβλεψης που θα επιλέξουμε $F1$.
- ✓ Όσο **περισσότερα δεδομένα** έχουμε τόσο μεγαλύτερη είναι η τιμή του t , οπότε τόσο μικρότερο είναι το βάρος του $F1$.

Π.χ. για $t = 12$ και $\alpha=0.1$ το βάρος ισούται με 0.2824

για $t = 24$ και $\alpha=0.1$ το βάρος ισούται με 0.0798

Μοντέλο Σταθερού Επιπέδου (V)

Weight assigned to:	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
Y_t	0.2	0.4	0.6	0.8
Y_{t-1}	0.16	0.24	0.24	0.16
Y_{t-2}	0.128	0.144	0.096	0.032
Y_{t-3}	0.1024	0.0864	0.0384	0.0064
Y_{t-4}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$



Μοντέλο Σταθερού Επιπέδου – SES (VI)

Εύρεση Αρχικού Επιπέδου

- ✓ Σαν αρχική πρόβλεψη (επίπεδο) συνήθως χρησιμοποιούμε:
 - ✓ Μέσος όρος των παρατηρήσεων
 - ✓ Μέσος όρος των τεσσάρων ή πέντε πρώτων παρατηρήσεων
 - ✓ Πρώτη παρατήρηση
 - ✓ Σταθερό επίπεδο από μοντέλο γραμμικής παλινδρόμησης

Μοντέλο Σταθερού Επιπέδου – SES (VII)

Εύρεση Βέλτιστου Συντελεστή Εξομάλυνσης (1/2)

- ✓ Η βέλτιστη τιμή του α καθορίζεται από την ελαχιστοποίηση ενός δείκτη σφάλματος (MSE, MAPE, ή άλλων)
- ✓ Το βέλτιστο α μπορεί να διαφέρει ανάλογα με τον δείκτη σφάλματος που επιλέγουμε να ελαχιστοποιήσουμε
- ✓ Το α λαμβάνει τιμές στο διάστημα $[0,1]$
- ✓ Χρειάζεται να υπολογίσουμε τα σφάλματα για κάθε τιμή του α , για κάθε τιμή του in-sample δείγματος

Μοντέλο Σταθερού Επιπέδου – SES (VIII)

Εύρεση Βέλτιστου Συντελεστή Εξομάλυνσης (2/2)

- ✓ Ένας τρόπος για τη βελτιστοποίηση του α είναι ο υπολογισμός του MSE για κάποιο διακριτό πλήθος εναλλακτικών τιμών (π.χ. 0.1, 0.2, ..., 0.9) και η επιλογή εκείνου που δίνει το μικρότερο σφάλμα MSE.
- ✓ Εναλλακτικά μπορεί κανείς αν χρησιμοποιήσει μη γραμμικούς αλγορίθμους βελτιστοποίησης.

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (I)

t	Y
0	
1	200
2	135
3	195
4	197,5
5	310
6	175
7	155
8	130
9	220
10	277,5
11	235
12	???

➤ Έστω μηνιαία χρονοσειρά Y_t

➤ Ζητείται η πρόβλεψη για τον ερχόμενο

Δεκέμβρη

➤ Αρχικοποιείται το επίπεδο ως ο Μ.Ο. των

δύο πρώτων ιστορικών παρατηρήσεων

➤ Επιλέγεται παράμετρος εξομάλυνσης $\alpha=0.2$

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (II)

t	Y	F	e	$S = F + a * e$	$S = a * Y + (1-a) * F$	St
0						167,5
1	200	167,5	32,5	$167,5 + 0.2 * 32,5$	$0.2 * 200 + 0.8 * 167,5$	174,0
2	135	174,0				
3	195					
4	197,5					
5	310					
6	175					
7	155					
8	130					
9	220					
10	277,5					
11	235					
12	???					

Αρχικό
Επίπεδο

Παράμετρος
εξομάλυνσης

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (III)

t	Y	F	e	$S = F + a * e$	$S = a * Y + (1-a) * F$	S
0						167,5
1	200	167,5	32,5	$167,5 + 0.2 * 32,5$	$0.2 * 200 + 0.8 * 167,5$	174,0
2	135	174,0	-39,0	$174 + 0.2 * -39$	$0.2 * 135 + 0.8 * 174$	166,2
3	195	166,2				
4	197,5					
5	310					
6	175					
7	155					
8	130					
9	220					
10	277,5					
11	235					
12	???					

Πρόβλεψη

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (IV)

t	Y	F	e	$S = F + a * e$	$S = a * Y + (1-a) * F$	S
0						167,5
1	200	167,5	32,5	$167,5 + 0.2 * 32,5$	$0.2 * 200 + 0.8 * 167,5$	174,0
2	135	174,0	-39,0	$174 + 0.2 * -39$	$0.2 * 135 + 0.8 * 174$	166,2
3	195	166,2	28,8	$166,2 + 0.2 * 28,8$	$0.2 * 195 + 0.8 * 166,2$	172,0
4	197,5	172,0	25,5	$172 + 0.2 * 25,5$	$0.2 * 197,5 + 0.8 * 172$	177,1
5	310	177,1	132,9	$177,1 + 0.2 * 132,9$	$0.2 * 310 + 0.8 * 177,1$	203,7
6	175	203,7	-28,7	$203,7 + 0.2 * -28,7$	$0.2 * 175 + 0.8 * 203,7$	197,9
7	155	197,9	-42,9	$197,9 + 0.2 * -42,9$	$0.2 * 155 + 0.8 * 197,9$	189,3
8	130	189,3	-59,3	$189,3 + 0.2 * -59,3$	$0.2 * 130 + 0.8 * 189,3$	177,5
9	220	177,5	42,5	$177,5 + 0.2 * 42,5$	$0.2 * 220 + 0.8 * 177,5$	186,0
10	277,5	186,0	91,5	$186 + 0.2 * 91,5$	$0.2 * 277,5 + 0.8 * 186$	204,3
11	235	204,3	30,7	$204,3 + 0.2 * 30,7$	$0.2 * 235 + 0.8 * 204,3$	210,4
12	???	210,4				

Τελική
Πρόβλεψη

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (V)

$\alpha = 0.5$

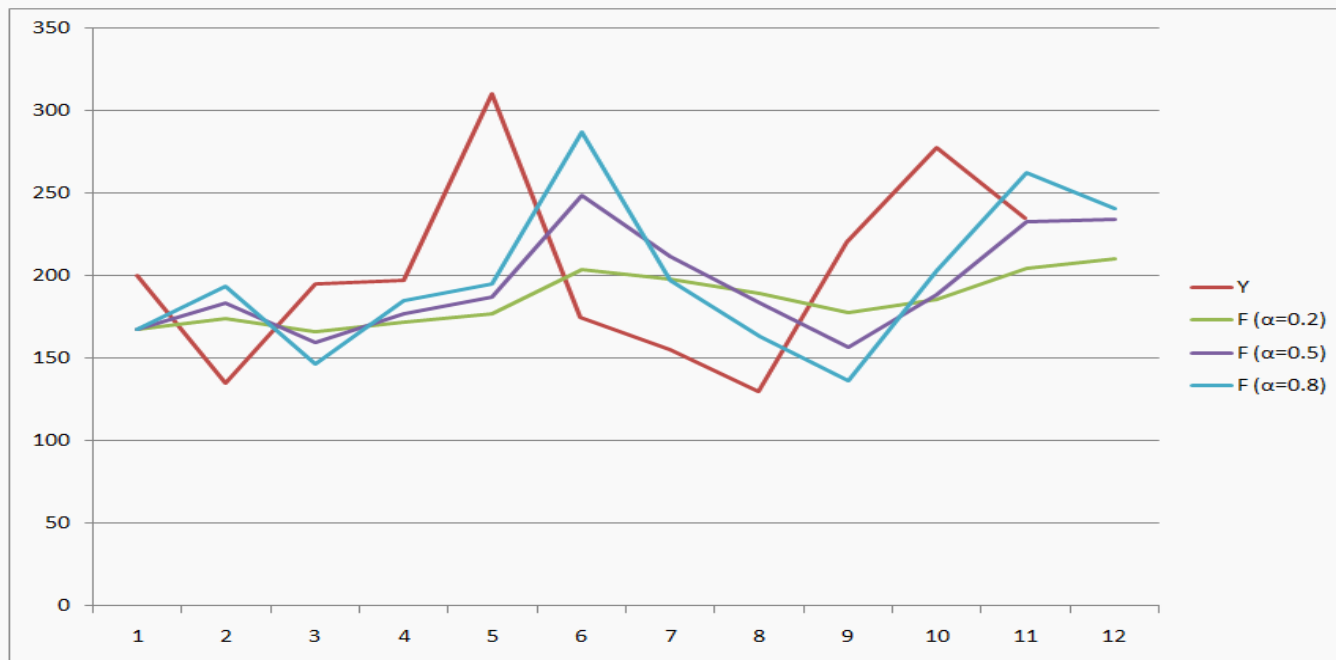
t	Y	F	e	S
0				167,5
1	200	167,5	32,5	183,8
2	135	183,8	-48,8	159,4
3	195	159,4	35,6	177,2
4	197,5	177,2	20,3	187,3
5	310	187,3	122,7	248,7
6	175	248,7	-73,7	211,8
7	155	211,8	-56,8	183,4
8	130	183,4	-53,4	156,7
9	220	156,7	63,3	188,4
10	277,5	188,4	89,1	232,9
11	235	232,9	2,1	234,0
12	???	234,0		

$\alpha = 0.8$

t	Y	F	e	S
0				167,5
1	200	167,5	32,5	193,5
2	135	193,5	-58,5	146,7
3	195	146,7	48,3	185,3
4	197,5	185,3	12,2	195,1
5	310	195,1	114,9	287,0
6	175	287,0	-112,0	197,4
7	155	197,4	-42,4	163,5
8	130	163,5	-33,5	136,7
9	220	136,7	83,3	203,3
10	277,5	203,3	74,2	262,7
11	235	262,7	-27,7	240,5
12	???	240,5		

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (VI)



Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (VII)



Η μεγαλύτερη τιμή του $\alpha = (0.8)$ εξομαλύνει πολύ λιγότερο τα δεδομένα ενώ η μικρότερη $\alpha = (0.2)$ αποδίδει σαφώς καλύτερη εξομάλυνση.



Αν $\alpha = 1$, η εκθετική εξομάλυνση θα ταυτιζόταν με την Naive, ενώ αν $\alpha = 0$ η πρόβλεψή θα ήταν σταθερή για όλες τις περιόδους και ίση με το αρχικό επίπεδο.

Μοντέλο Σταθερού Επιπέδου

Παράδειγμα (VIII)

t	Y	F ($\alpha=0.2$)	E	AE	APE	SAPE
1	200	167,5	32,50	32,50	0,163	0,177
2	135	174,0	-39,00	39,00	0,289	0,252
3	195	166,2	28,80	28,80	0,148	0,159
4	197,5	172,0	25,54	25,54	0,129	0,138
5	310	177,1	132,93	132,93	0,429	0,546
6	175	203,7	-28,65	28,65	0,164	0,151
7	155	197,9	-42,92	42,92	0,277	0,243
8	130	189,3	-59,34	59,34	0,456	0,372
9	220	177,5	42,53	42,53	0,193	0,214
10	277,5	186,0	91,52	91,52	0,330	0,395
11	235	204,3	30,72	30,72	0,131	0,140
12		210,4				
		$\alpha=0.2$	19,51	<u>50,41</u>	<u>0,25</u>	<u>0,25</u>
		$\alpha=0.5$	12,08	54,39	0,27	0,27
		$\alpha=0.8$	<u>8,30</u>	58,13	0,29	0,29

Μεγαλύτερη
προκατάληψη

Καλύτερη ακρίβεια

Μοντέλο Σταθερού Επιπέδου – SES

Χρόνος t	Δεδομένα X_t	Πρόβλεψη F $\hat{X}_{t-1}(1)$	Σφάλμα e_t	Επίπεδο στο τέλος της περιόδου t $S_t = S_{t-1} + h_1 e_t$
0				$S_0 = 585$
1	545	585	-40	$S_1 = 585 + 0.4(-40) = 569$
2	635	569	66	$S_2 = 569 + 0.4(6.6) = 595.4$
3	420	595,4	-175,4	$S_3 = 595.4 + 0.4(-175.4) = 525.2$
4	716	525,2	190,8	$S_4 = 525.2 + 0.4(190.8) = 601.5$
5	699	601,5	97,5	$S_5 = 601.5 + 0.4(97.5) = 640.5$
6	681	640,5	40,5	$S_6 = 640.5 + 0.4(40.5) = 656.5$
7	763	656,5	106,3	$S_7 = 656.5 + 0.4(106.3) = 699.2$
8	778	699,2	78,8	$S_8 = 699.2 + 0.4(78.8) = 730.5$
9	690	730,5	-40,5	$S_9 = 730.5 + 0.4(-40.5) = 711.4$
10	707	711,4	-7,4	$S_{10} = 711.4 + 0.4(-7.4) = 711.5$
11	716	711,5	4,5	$S_{11} = 711.5 + 0.4(4.5) = 713.3$
12		713,3		

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \alpha \cdot e_t$$

$$F_{t+1} = S_t$$

Ο συντελεστής α πρέπει να επιλέγεται βέλτιστα ώστε να ελαχιστοποιείται *συνήθως* το μέσο τετραγωνικό σφάλμα (MSE), όπου $0 < \alpha < 1$

Μοντέλο Γραμμικής Τάσης (Holt)

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + a * e_t$$

$$T_t = T_{t-1} + a * b * e_t$$

$$F_{t+m} = S_t + mT_t$$

Οι συντελεστές a και b πρέπει να επιλέγονται βέλτιστα ώστε να ελαχιστοποιείται συνήθως το μέσο τετραγωνικό σφάλμα (MSE), όπου $0 < a < 1$, $0 < b < a$

- ✓ Χρειάζεται προσοχή στην αρχικοποίηση του μοντέλου.
- ✓ Ουσιαστικά εκτελείται μία γραμμική παλινδρόμηση, με το χρόνο ως ανεξάρτητη μεταβλητή.
- ✓ Ως αρχικό επίπεδο συνήθως ορίζεται η σταθερά A της παλινδρόμησης **$y = A + Bt$** .
- ✓ Ως αρχική τάση συνήθως ορίζεται η κλίση B της παλινδρόμησης

Μοντέλο Γραμμικής Τάσης (Holt)

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$$

$$T_t = T_{t-1} + \beta \cdot e_t$$

$$F_{t+m} = S_t + m \cdot T_t$$

$$h_1 = 0.20, h_2 = 0.10$$

Time t	Data X_t	Forecast $X_{t-1}(1)$	Error e_t	Level at End of t $S_t = S_{t-1} + T_{t-1} + h_1 e_t$	Trend at End of t $T_t = T_{t-1} + h_2 e_t$	Forecast for $t + 1$ $\hat{X}_t(1) = S_t + T_t$
0				$S_0 =$ 54.0	$T_0 =$ 2.0	$\hat{X}_0(1) = 54.0 + 2.0 = 56.0$
1	54.0	56.0	-2.0	$S_1 = 54.0 + 2.0 + 0.2(-2.0) = 55.6$	$T_1 = 2.0 + 0.1(-2.0) = 1.8$	$\hat{X}_1(1) = 55.6 + 1.8 = 57.4$
2	55.0	57.4	-2.4	$S_2 = 55.6 + 1.8 + 0.2(-2.4) = 56.9$	$T_2 = 1.8 + 0.1(-2.4) = 1.6$	$\hat{X}_2(1) = 56.9 + 1.6 = 58.5$
3	57.0	58.5	-1.5	$S_3 = 56.9 + 1.6 + 0.2(-1.5) = 58.2$	$T_3 = 1.6 + 0.1(-1.5) = 1.5$	$\hat{X}_3(1) = 58.2 + 1.5 = 59.7$
4	60.0	59.7	0.3	$S_4 = 58.2 + 1.5 + 0.2(0.3) = 59.8$	$T_4 = 1.5 + 0.1(0.3) = 1.5$	$\hat{X}_4(1) = 59.8 + 1.5 = 61.3$
5	66.0	61.3	4.7	$S_5 = 59.8 + 1.5 + 0.2(4.7) = 62.2$	$T_5 = 1.5 + 0.1(4.7) = 2.0$	$\hat{X}_5(1) = 62.2 + 2.0 = 64.2$
6	62.0	64.2	-2.2	$S_6 = 62.2 + 2.0 + 0.2(-2.2) = 63.8$	$T_6 = 2.0 + 0.1(-2.2) = 1.8$	$\hat{X}_6(1) = 63.8 + 1.8 = 65.6$
7	59.0	65.6	-6.6	$S_7 = 63.8 + 1.8 + 0.2(-6.6) = 64.3$	$T_7 = 1.8 + 0.1(-6.6) = 1.1$	$\hat{X}_7(1) = 64.3 + 1.1 = 65.4$
8	65.0	65.4	-0.4	$S_8 = 64.3 + 1.1 + 0.2(-0.4) = 65.3$	$T_8 = 1.1 + 0.1(-0.4) = 1.1$	$\hat{X}_8(1) = 65.3 + 1.1 = 66.4$
9	69.0	66.4	2.6	$S_9 = 65.3 + 1.1 + 0.2(2.6) = 66.9$	$T_9 = 1.1 + 0.1(2.6) = 1.4$	$\hat{X}_9(1) = 66.9 + 1.4 = 68.3$
10	70.0	68.3	1.7	$S_{10} = 66.9 + 1.4 + 0.2(1.7) = 68.6$	$T_{10} = 1.4 + 0.1(1.7) = 1.6$	$\hat{X}_{10}(1) = 68.6 + 1.6 = 70.2$
11	63.0	70.2	-7.2	$S_{11} = 68.6 + 1.6 + 0.2(-7.2) = 68.8$	$T_{11} = 1.6 + 0.1(-7.2) = 0.9$	$\hat{X}_{11}(1) = 68.8 + 0.9 = 69.7$
12	75.0	69.7	5.3	$S_{12} = 68.8 + 0.9 + 0.2(5.3) = 70.8$	$T_{12} = 0.9 + 0.1(5.3) = 1.4$	$\hat{X}_{12}(1) = 70.8 + 1.4 = 72.2$
13		72.2				

Μοντέλο Μη Γραμμικής Τάσης (Damped)

$$\begin{aligned}
 e_t &= Y_t - F_t \\
 S_t &= S_{t-1} + \varphi T_{t-1} + a * e_t \\
 T_t &= \varphi T_{t-1} + a * b * e_t \\
 F_{t+m} &= S_t + \sum_{i=1}^m \varphi^i T_t
 \end{aligned}$$

Οι συντελεστές a και β πρέπει να επιλέγονται βέλτιστα ώστε να ελαχιστοποιείται συνήθως το μέσο τετραγωνικό σφάλμα (MSE), όπου $0 < a < 1$, $0 < \beta < a$

- Χρειάζεται προσοχή στην αρχικοποίηση του μοντέλου
- Ουσιαστικά εκτελείται μία **φθίνουσα** γραμμική παλινδρόμηση, με το χρόνο ως ανεξάρτητη μεταβλητή
- Ως αρχικό επίπεδο ορίζεται η σταθερά A της παλινδρόμησης
- Ως αρχική τάση ορίζεται η κλίση B της παλινδρόμησης

Μοντέλο Μη Γραμμικής Τάσης (Damped)

$$\begin{aligned}
 e_t &= Y_t - F_t \\
 S_t &= S_{t-1} + \phi T_{t-1} + a * e_t \\
 T_t &= \phi T_{t-1} + a * b * e_t \\
 F_{t+m} &= S_t + \sum_{i=1}^m \phi^i T_t
 \end{aligned}$$

➤ Το μοντέλο Μη Γραμμικής Τάσης μπορεί να χρησιμοποιηθεί σαν ένα αυτόματο μοντέλο πρόβλεψης για κάθε μη εποχιακή χρονοσειρά, ανάλογα με τον damping factor που θα επιλέξουμε:

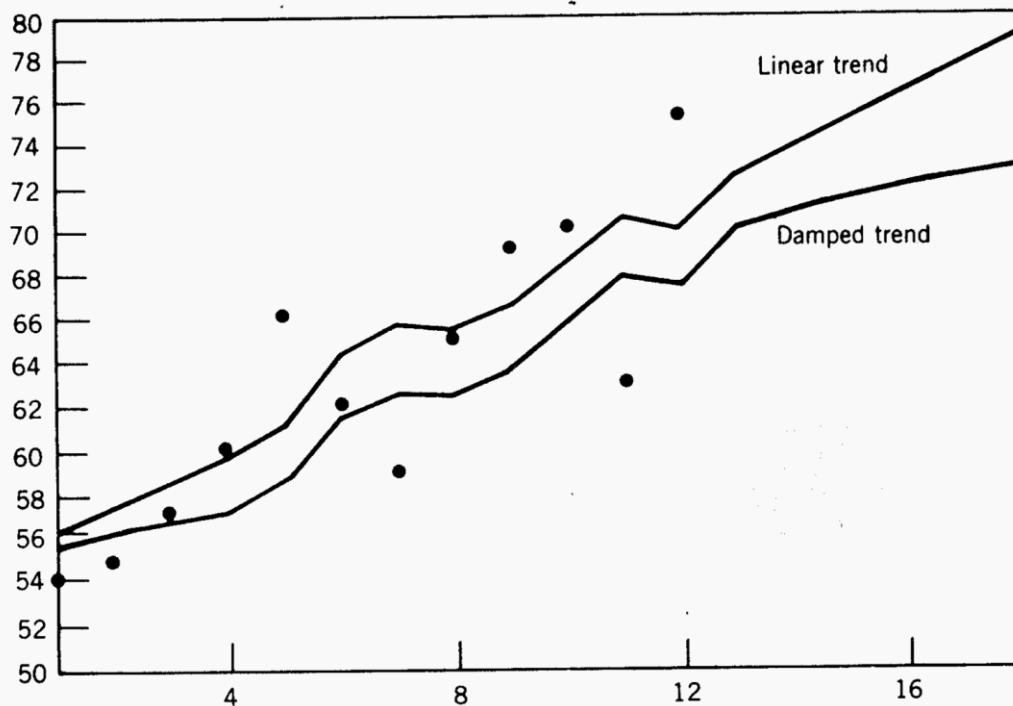
- $\phi \approx 0$, σταθερού επιπέδου
- $\phi < 1$, φθίνουσας τάσης
- $\phi \approx 1$, γραμμικής τάσης
- $\phi > 1$, εκθετικής τάσης

Μοντέλο Μη Γραμμικής Τάσης (Damped)

$$h_1 = 0.2, \quad h_2 = 0.1, \quad \varphi = 0.8$$

Time t	Data X_t	Forecast $X_{t-1}(1)$	Error e_t	Level at End of t $S_t = S_{t-1} + \phi T_{t-1} + h_1 e_t$	Trend at End of t $T_t = \phi T_{t-1} + h_2 e_t$	Forecast for $t + 1$ $\hat{X}_t(1) = S_t + \phi T_t$
0				$S_0 =$	$T_0 =$	$\hat{X}_0(1) = 54.0 + 0.8(2.0) = 55.6$
1	54.0	55.6	-1.6	$S_1 = 54.0 + 0.8(2.0) + 0.2(-1.6) = 55.3$	$T_1 = 0.8(2.0) + 0.1(-1.6) = 1.4$	$\hat{X}_1(1) = 55.3 + 0.8(1.4) = 56.4$
2	55.0	56.4	-1.4	$S_2 = 55.3 + 0.8(1.4) + 0.2(-1.4) = 56.1$	$T_2 = 0.8(1.4) + 0.1(-1.4) = 1.0$	$\hat{X}_2(1) = 56.1 + 0.8(1.0) = 56.9$
3	57.0	56.9	0.1	$S_3 = 56.1 + 0.8(1.0) + 0.2(0.1) = 56.9$	$T_3 = 0.8(1.0) + 0.1(0.1) = 0.8$	$\hat{X}_3(1) = 56.9 + 0.8(0.8) = 57.5$
4	60.0	57.5	2.5	$S_4 = 56.9 + 0.8(0.8) + 0.2(2.5) = 58.0$	$T_4 = 0.8(0.8) + 0.1(2.5) = 0.9$	$\hat{X}_4(1) = 58.0 + 0.8(0.9) = 58.7$
5	66.0	58.7	7.3	$S_5 = 58.0 + 0.8(0.9) + 0.2(7.3) = 60.2$	$T_5 = 0.8(0.9) + 0.1(7.3) = 1.5$	$\hat{X}_5(1) = 60.2 + 0.8(1.5) = 61.4$
6	62.0	61.4	0.6	$S_6 = 60.2 + 0.8(1.5) + 0.2(0.6) = 61.5$	$T_6 = 0.8(1.5) + 0.1(0.6) = 1.3$	$\hat{X}_6(1) = 61.5 + 0.8(1.3) = 62.5$
7	59.0	62.5	-3.5	$S_7 = 61.5 + 0.8(1.3) + 0.1(-3.5) = 61.8$	$T_7 = 0.8(1.3) + 0.1(-3.5) = 0.7$	$\hat{X}_7(1) = 61.8 + 0.8(0.7) = 62.4$
8	65.0	62.4	2.6	$S_8 = 61.8 + 0.8(0.7) + 0.2(2.6) = 62.9$	$T_8 = 0.8(0.7) + 0.1(2.6) = 0.8$	$\hat{X}_8(1) = 62.9 + 0.8(0.8) = 63.5$
9	69.0	63.5	5.5	$S_9 = 62.9 + 0.8(0.8) + 0.2(5.5) = 64.6$	$T_9 = 0.8(0.8) + 0.1(5.5) = 1.2$	$\hat{X}_9(1) = 64.6 + 0.8(1.2) = 65.6$
10	70.0	65.6	4.4	$S_{10} = 64.6 + 0.8(1.2) + 0.2(4.4) = 66.5$	$T_{10} = 0.8(1.2) + 0.1(4.4) = 1.4$	$\hat{X}_{10}(1) = 66.5 + 0.8(1.4) = 67.6$
11	63.0	67.6	-4.6	$S_{11} = 66.5 + 0.8(1.4) + 0.2(-4.6) = 66.7$	$T_{11} = 0.8(1.4) + 0.1(-4.6) = 0.7$	$\hat{X}_{11}(1) = 66.7 + 0.8(0.7) = 67.3$
12	75.0	67.3	7.7	$S_{12} = 66.7 + 0.8(0.7) + 0.2(7.7) = 68.8$	$T_{12} = 0.8(0.7) + 0.1(7.7) = 1.3$	$\hat{X}_{12}(1) = 68.8 + 0.8(1.3) = 69.8$
13		69.8				

Σύγκριση Μοντέλων Γραμμικής & Μη Γραμμικής Τάσης



Μοντέλο Σταθερού Επιπέδου με Πολ/κή Εποχιακότητα

$$h_1 = 0.1, \quad h_3 = 0.01$$

Time t	Data X_t	Forecast $\hat{X}_{t-1}(1)$	Error e_t	Deseasonalized Level at End of t $S_t = S_{t-1} + h_1 e_t / I_{t-p}$	Seasonal Index at End of t $I_t = I_{t-p} + h_3 e_t / S_t$	Forecast for $t+1$ $\hat{X}_t(1) = S_t(I_{t-p+1})$
-3					$I_{-3} =$	
-2					$I_{-2} =$	
-1					$I_{-1} =$	
0				$S_0 =$	$I_0 =$	
1	53.0	45.5	7.5	$S_1 = 74.3 + 0.1(7.5)/0.6122 = 75.5$	$I_1 = 0.6122 + 0.01(7.5)/75.5 = 0.6132$	$\hat{X}_0(1) = 74.3(0.6122) = 45.5$
2	85.0	76.2	8.8	$S_2 = 75.5 + 0.1(8.8)/1.0086 = 76.4$	$I_2 = 1.0086 + 0.01(8.8)/76.4 = 1.0098$	$\hat{X}_1(1) = 75.5(1.0086) = 76.2$
3	92.0	101.6	-9.6	$S_3 = 76.4 + 0.1(-9.6)/1.3303 = 75.7$	$I_3 = 1.3303 + 0.01(-9.6)/75.7 = 1.3290$	$\hat{X}_2(1) = 76.4(1.3303) = 101.6$
4	78.0	79.4	-1.4	$S_4 = 75.7 + 0.1(-1.4)/1.0489 = 75.5$	$I_4 = 1.0489 + 0.01(-1.4)/75.5 = 1.0487$	$\hat{X}_3(1) = 75.7(1.0489) = 79.4$
5	44.0	46.3	-2.3	$S_5 = 75.5 + 0.1(-2.3)/0.6132 = 75.2$	$I_5 = 0.6132 + 0.01(-2.3)/75.2 = 0.6129$	$\hat{X}_4(1) = 75.5(0.6132) = 46.3$
6	75.0	75.9	-0.9	$S_6 = 75.2 + 0.1(-0.9)/1.0098 = 75.1$	$I_6 = 1.0098 + 0.01(-0.9)/75.1 = 1.0096$	$\hat{X}_5(1) = 75.2(1.0098) = 75.9$
7	102.0	99.8	2.2	$S_7 = 75.1 + 0.1(2.2)/1.3290 = 75.2$	$I_7 = 1.3290 + 0.01(2.2)/75.2 = 1.3293$	$\hat{X}_6(1) = 75.1(1.3290) = 99.8$
8	60.0	78.9	-18.9	$S_8 = 75.2 + 0.1(-18.9)/1.0487 = 73.4$	$I_8 = 1.0487 + 0.01(-18.9)/73.4 = 1.0461$	$\hat{X}_7(1) = 75.2(1.0487) = 78.9$
9	55.0	45.0	10.0	$S_9 = 73.4 + 0.1(10.0)/0.6129 = 75.1$	$I_9 = 0.6129 + 0.01(10.0)/75.1 = 0.6142$	$\hat{X}_8(1) = 73.4(0.6129) = 45.0$
10	88.0	75.8	12.2	$S_{10} = 75.1 + 0.1(12.2)/1.0096 = 76.3$	$I_{10} = 1.0096 + 0.01(12.2)/76.3 = 1.0112$	$\hat{X}_9(1) = 75.1(1.0096) = 75.8$
11	108.0	101.4	6.6	$S_{11} = 76.3 + 0.1(6.6)/1.3293 = 76.8$	$I_{11} = 1.3293 + 0.01(6.6)/76.8 = 1.3302$	$\hat{X}_{10}(1) = 76.3(1.3293) = 101.4$
12	59.0	80.3	-21.3	$S_{12} = 76.8 + 0.1(-21.3)/1.0461 = 74.7$	$I_{12} = 1.0461 + 0.01(-21.3)/74.7 = 1.0433$	$\hat{X}_{11}(1) = 76.8(1.0461) = 80.3$
13		45.9				$\hat{X}_{12}(1) = 74.7(0.6142) = 45.9$

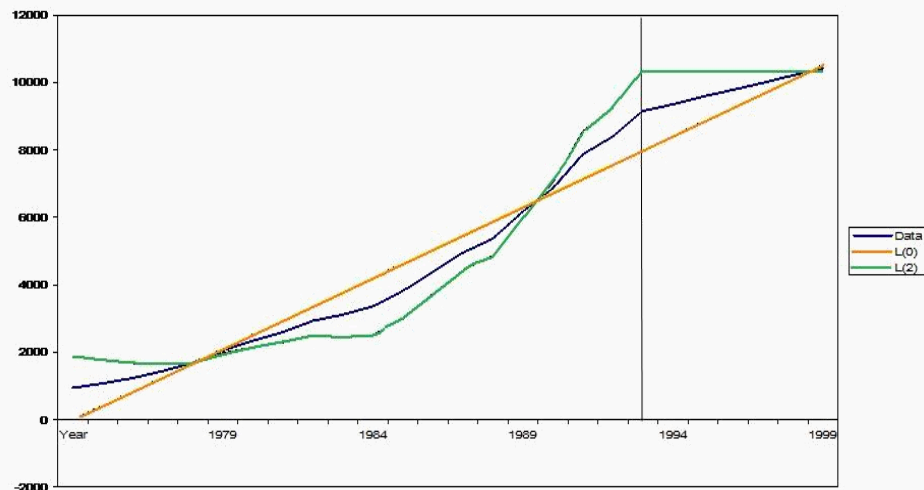
Στατιστική Πρόβλεψη

Κυριότερες Στατιστικές Μέθοδοι Πρόβλεψης

Μέθοδος Theta

Η μέθοδος πρόβλεψης **Theta** βασίζεται στην τροποποίηση των τοπικών καμπυλοτήτων της χρονοσειράς. Η αρχική χρονοσειρά αποσυντίθεται σε δύο ή περισσότερες γραμμές Theta, κάθε μία από τις οποίες προεκτείνεται ξεχωριστά. Οι ανεξάρτητες προβλέψεις συνδυάζονται για την παραγωγή των τελικών προβλέψεων.

Figure 1. M3 Competition - N0001YB001-YEARLY/MCRO



Η Μέθοδος Θ : Μια μέθοδος παραγωγής προβλέψεων βασισμένη σε μια διαφορετική προσέγγιση της αποσύνθεσης

- Η μέθοδος Θ (Assimakopoulos & Nikolopoulos, 2000) είναι μια μονοδιάστατη μέθοδος πρόβλεψης.
- Βασίζεται στην μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς μέσα από την παράμετρο θ (Theta), η οποία εφαρμόζεται στις διαφορές δεύτερης τάξης των δεδομένων.
- Η καινούργια χρονοσειρά που δημιουργείται (γραμμή Θ ή Theta Line) διατηρεί την μέση τιμή και γραμμική κλίση της αρχικής χρονοσειράς, αλλά όχι και τις τοπικές καμπυλότητες.
- Βασικό ποιοτικό χαρακτηριστικό των γραμμών Θ είναι η καλύτερη προσέγγιση της μακροπρόθεσμης συμπεριφοράς (τάσης) των δεδομένων ή η ανάδειξη των βραχυπρόθεσμων χαρακτηριστικών της (επίπεδο), ανάλογα με την τιμή της παραμέτρου θ που επιλέγεται ($<$, $>$ 1).

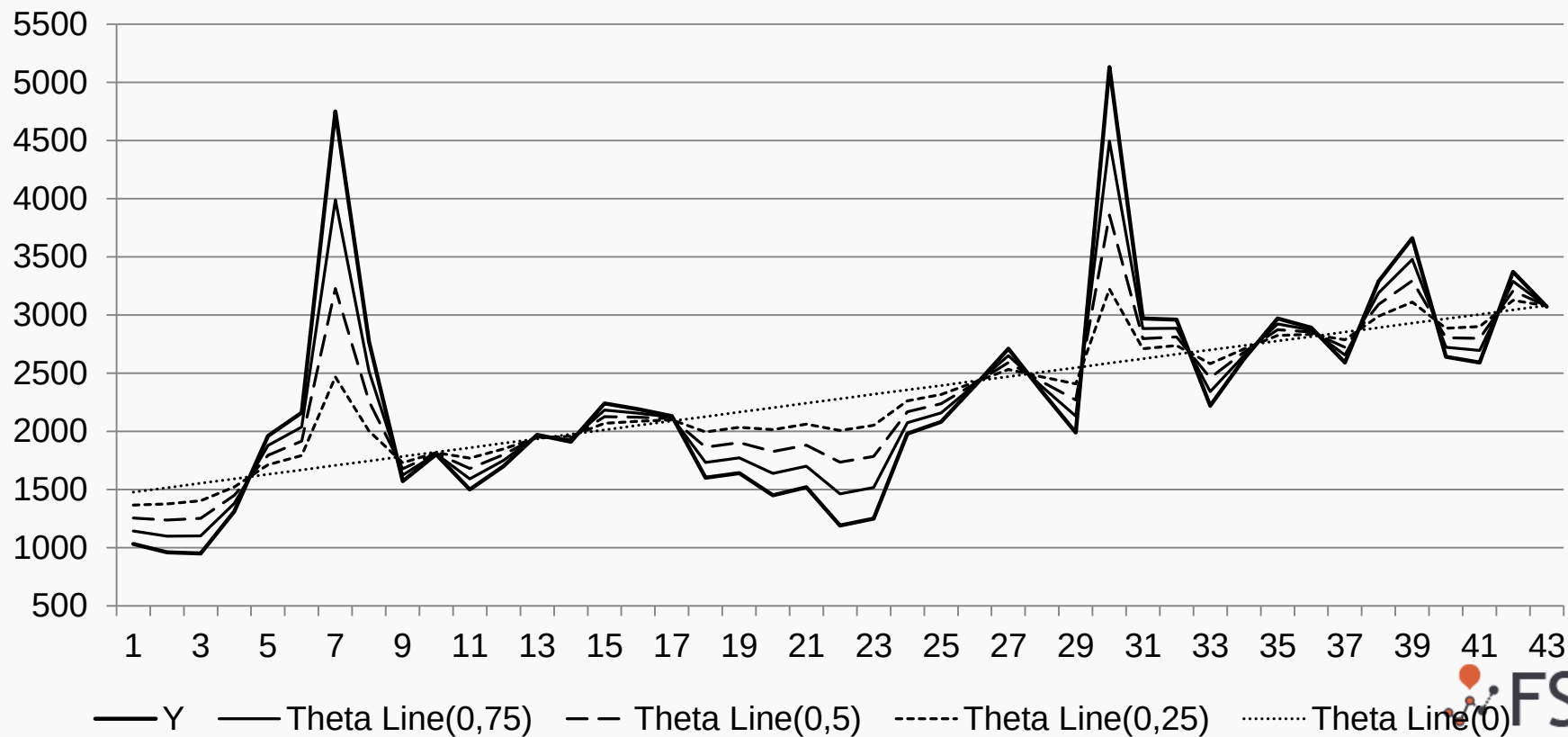
Η Μέθοδος Θ : Μια μέθοδος παραγωγής προβλέψεων βασισμένη σε μια διαφορετική προσέγγιση της αποσύνθεσης

- Η μέθοδος Θ αποσυνθέτει (διαχωρίζει) την αρχική χρονοσειρά σε δύο ή περισσότερες γραμμές Θ .
- Η κάθε γραμμή προεκτείνεται στο μέλλον ξεχωριστά (με την ίδια ή και με διαφορετικές μεθόδους πρόβλεψης) και οι παραγόμενες προβλέψεις συνδυάζονται για να προκύψει η τελική πρόβλεψη.
- Ο απλός συνδυασμός δύο γραμμών Θ , για $\theta=0$ (ευθεία γραμμή) και $\theta=2$ (διπλασιασμός των τοπικών καμπυλοτήτων) χρησιμοποιήθηκε για την παραγωγή προβλέψεων για τις 3003 χρονοσειρές του διεθνούς διαγωνισμού προβλέψεων M3 (*Makridakis & Hibon, 2000*).
- Η μέθοδος παρήγαγε πολύ καλά αποτελέσματα, ιδιαίτερα για τις μηνιαίες χρονοσειρές και τα μικροοικονομικά δεδομένα.

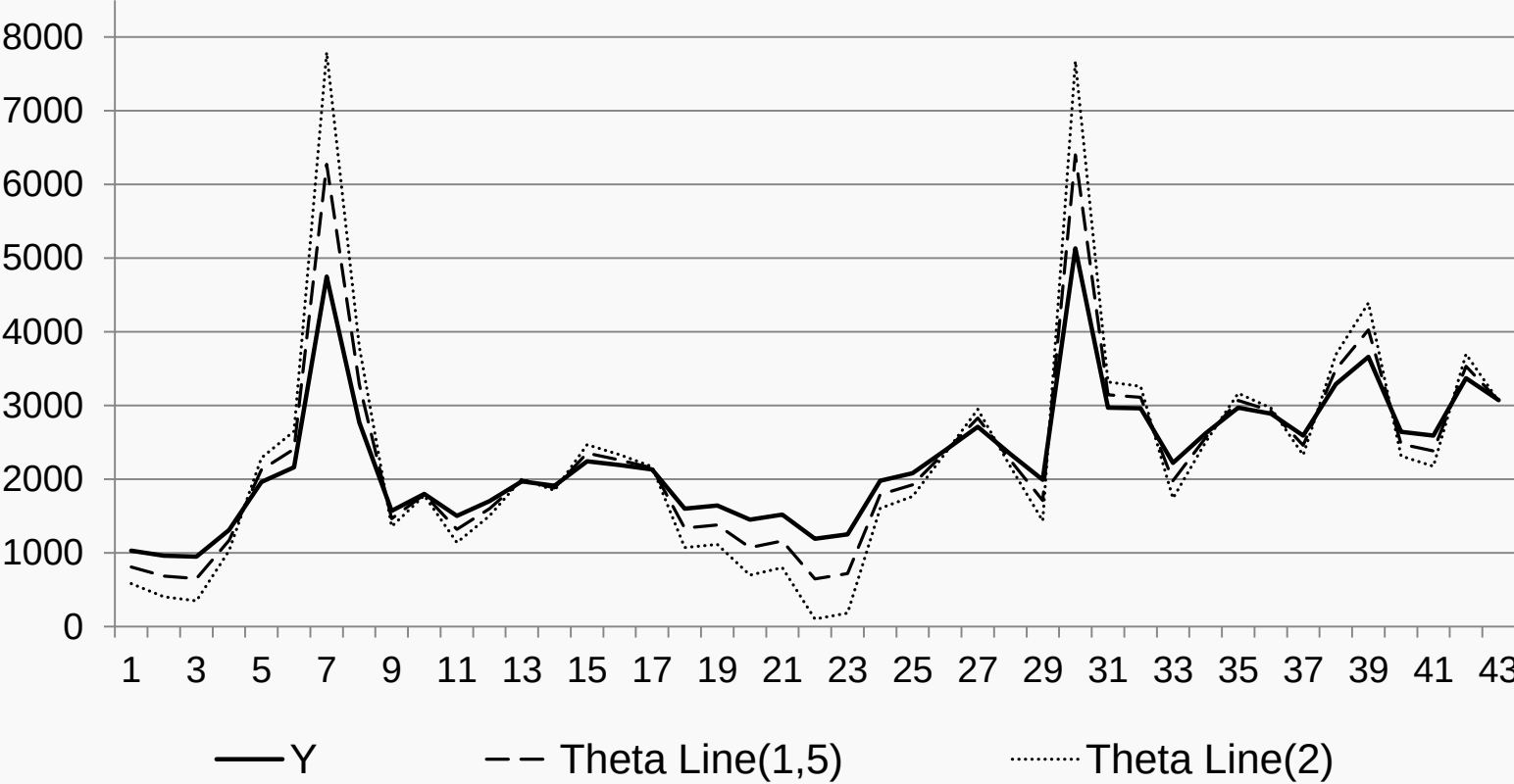
Η Μέθοδος Θ : Μια μέθοδος παραγωγής προβλέψεων βασισμένη σε μια διαφορετική προσέγγιση της αποσύνθεσης

- Η μέθοδος Theta εισήγαγε μια διαφορετική προσέγγιση της αποσύνθεσης. Ο διαχωρισμός των αποεποχικοποιημένων δεδομένων γίνεται σε συνιστώσες (γραμμές Theta) μακροπρόθεσμης και βραχυπρόθεσμης τάσης.
- Η πρόκληση για τη συγκεκριμένη μέθοδο ήταν να αυξήσει το βαθμό αξιοποίησης της χρήσιμης πληροφορίας που είναι κρυμμένη μέσα στα δεδομένα, πριν την εφαρμογή ενός μοντέλου επέκτασης των δεδομένων στο μέλλον (extrapolation model).
- Ουσιαστικά, η μέθοδος Theta λειτουργεί σαν ένας μεγεθυντικός φακός μέσα από τον οποίο οι διακυμάνσεις της χρονοσειράς μεγεθύνονται ή μικραίνουν. Ο γραμμικός συνδυασμός των προβλέψεων των συνιστωσών, γίνεται, μέσα από αυτήν την διαδικασία, πιο αποδοτικός.

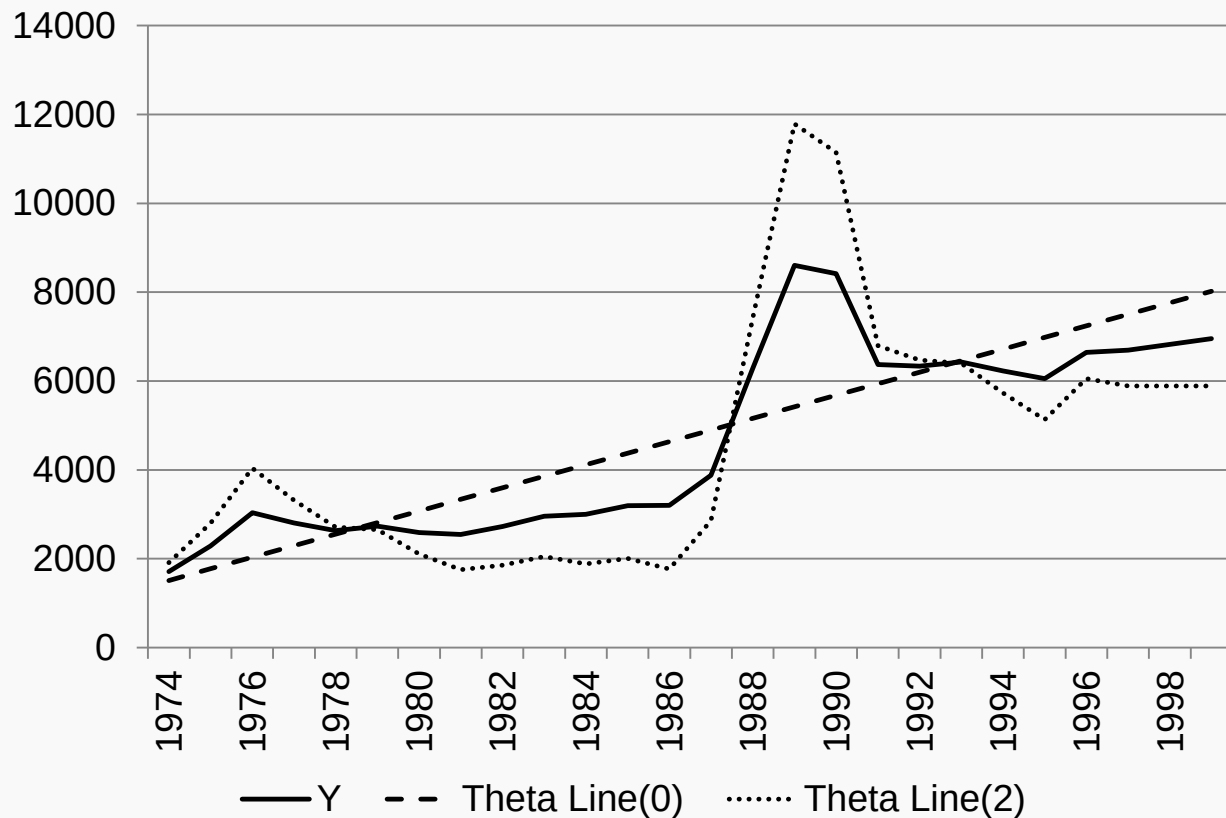
Γραμμές Θ (1/3)



Γραμμές Θ (2/3)



Γραμμές Θ (3/3)



$$Y_t = \frac{1}{2} \cdot (Y_t^{\theta=0} + Y_t^{\theta=2})$$

Το κλασσικό μοντέλο Θ

- Βήμα 0: **Τεστ Εποχιακότητας**

Η κάθε χρονοσειρά ελέγχεται ως προς την ύπαρξη εποχιακής συμπεριφοράς με κριτήριο την τιμή του συντελεστή αυτοσυσχέτισης - υστέρηση ίση με τη συχνότητα των δεδομένων (π.χ. για μηνιαία δεδομένα $k=12$) και διάστημα εμπιστοσύνης 90% ($t_{critical}=1.645$)

- Βήμα 1: **Αποεποχικοποίηση**

Η χρονοσειρά αποεποχικοποιείται – αν το τεστ ήταν θετικό - με την κλασσική μέθοδο αποσύνθεσης

- Βήμα 2: **Αποσύνθεση**

Κάθε χρονοσειρά αποσυντίθεται σε δύο γραμμές Θ , με $\theta=0$ και $\theta=2$

- Βήμα 3: **Πρόβλεψη**

Η γραμμή $TL(0)$ προεκτείνεται με απλή γραμμική παλινδρόμηση (LRL) ενώ η γραμμή $TL(2)$ με απλή εκθετική εξομάλυνση (SES)

- Βήμα 4: **Συνδυασμός**

Οι προηγούμενες προβλέψεις συνδυάζονται με ίσα βάρη

- Βήμα 5: **Εποχικοποίηση**

Οι τελικές προβλέψεις εποχικοποιούνται – αν το τεστ ήταν θετικό

Υπολογίζοντας τις γραμμές Theta

Για το κλασσικό μοντέλο Theta ($\theta=0$ και $\theta=2$) οι γραμμές θ υπολογίζονται ως εξής:

- $\text{Theta Line}(0) = \text{LRL}$
- $\text{Theta Line}(2) = 2 \times \text{Data} - \text{LRL}$

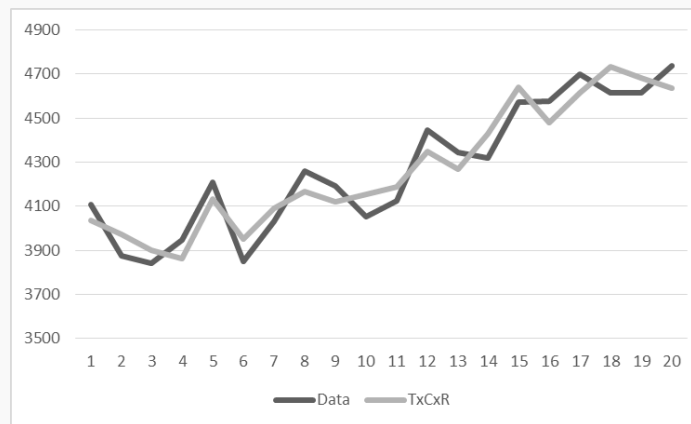
Γενικότερα ισχύει:

- **$\text{Theta Line}(\theta) = \theta \times \text{Data} + (1-\theta) \times \text{LRL}$**
- Ή ισοδύναμα: $\text{Theta Line}(\theta) = \text{LRL} + \theta \times \text{eLRL}$

Παράδειγμα Theta (1/5)

➤ Αποεποχικοποίηση εποχιακής τριμηνιαίας χρονοσειράς

Period	Data	TxC	SxRx100	I=SxR=S	TxCxR
1	4109			101,79	4036,68
2	3874			97,48	3974,13
3	3842	3955,00	97,14	98,54	3899,03
4	3946	3964,25	99,54	102,19	3861,43
5	4207	3984,75	105,58	101,79	4132,95
6	3850	4047,50	95,12	97,48	3949,51
7	4030	4085,00	98,65	98,54	4089,82
8	4260	4108,38	103,69	102,19	4168,71
9	4193	4145,50	101,15	101,79	4119,20
10	4051	4180,63	96,90	97,48	4155,70
11	4126	4222,63	97,71	98,54	4187,24
12	4445	4275,00	103,98	102,19	4349,74
13	4344	4364,13	99,54	101,79	4267,54
14	4319	4436,13	97,36	97,48	4430,63
15	4571	4496,88	101,65	98,54	4638,85
16	4576	4578,13	99,95	102,19	4477,93
17	4699	4620,25	101,70	101,79	4616,30
18	4614	4645,75	99,32	97,48	4733,25
19	4613			98,54	4681,47
20	4738			102,19	4636,46



	105,58	101,15	99,54	101,70
	95,12	96,90	97,36	99,32
97,14	98,65	97,71	101,65	
99,54	103,69	103,98	99,95	

min	max
99,54	105,58
95,12	99,32
97,14	101,65
99,54	103,98

average (w/o min & max)
101,43
97,13
98,18
101,82

SI
101,79
97,48
98,54
102,19

sum	398,559
Σ.K.	0,9964

Παράδειγμα Theta (2/5)

➤ Υπολογισμός γραμμής TL(0)

X		Y		Numerator			Denominator		ThetaLine(0)	
Period	Data	X-Mean(X)=A	Y-Mean(Y)=B	A*B	(X-Mean(X))^2		LRL			
1	4036,68	-9,5	-233,67	2219,87	90,25		3846.42			
2	3974,13	-8,5	-296,22	2517,88	72,25		3891.05			
3	3899,03	-7,5	-371,32	2784,92	56,25		3935.67			
4	3861,43	-6,5	-408,92	2657,95	42,25		3980.29			
5	4132,95	-5,5	-137,40	755,68	30,25		4024.92			
6	3949,51	-4,5	-320,84	1443,79	20,25		4069.54			
7	4089,82	-3,5	-180,53	631,86	12,25		4114.16			
8	4168,71	-2,5	-101,64	254,11	6,25		4158.79			
9	4119,20	-1,5	-151,15	226,72	2,25		4203.41			
10	4155,70	-0,5	-114,65	57,32	0,25		4248.03			
11	4187,24	0,5	-83,11	-41,55	0,25		4292.66			
12	4349,74	1,5	79,39	119,09	2,25		4337.28			
13	4267,54	2,5	-2,81	-7,02	6,25		4381.90			
14	4430,63	3,5	160,28	560,98	12,25		4426.53			
15	4638,85	4,5	368,50	1658,24	20,25		4471.15			
16	4477,93	5,5	207,58	1141,71	30,25		4515.77			
17	4616,30	6,5	345,95	2248,64	42,25		4560.40			
18	4733,25	7,5	462,90	3471,78	56,25		4605.02			
19	4681,47	8,5	411,12	3494,53	72,25		4649.64			
20	4636,46	9,5	366,11	3478,06	90,25		4694.27			
21							4738.89			
22							4783.51			
23							4828.14			
Average		10,5	4270,329	Sum	29674.52		665			

b=slope
a=constant

444.623
3801.8

Παράδειγμα Theta (3/5)

46

➤ Υπολογισμός γραμμής TL(2) και προέκταση γραμμών Θ

Period	Data	ThetaLine(0)	ThetaLine(2)
1	4036,68	3846.42	4226.94
2	3974,13	3891.05	4057.21
3	3899,03	3935.67	3862.39
4	3861,43	3980.29	3742.57
5	4132,95	4024.92	4240.98
6	3949,51	4069.54	3829.48
7	4089,82	4114.16	4065.48
8	4168,71	4158.79	4178.63
9	4119,20	4203.41	4034.99
10	4155,70	4248.03	4063.37
11	4187,24	4292.66	4081.82
12	4349,74	4337.28	4362.20
13	4267,54	4381.90	4153.18
14	4430,63	4426.53	4434.73
15	4638,85	4471.15	4806.55
16	4477,93	4515.77	4440.09
17	4616,30	4560.40	4672.20
18	4733,25	4605.02	4861.48
19	4681,47	4649.64	4713.30
20	4636,46	4694.27	4578.65
21		4738.89	
22		4783.51	
23		4828.14	

SES on ThetaLine(2) with $\alpha=0.5$
4226.94
4226.94
4142.08
4002.23
3872.40
4056.69
3943.09
4004.28
4091.46
4063.22
4063.30
4072.56
4217.38
4185.28
4310.01
4558.28
4499.18
4585.69
4723.59
4718.44
4648.55
4648.55
4648.55

Παράδειγμα Theta (4/5)

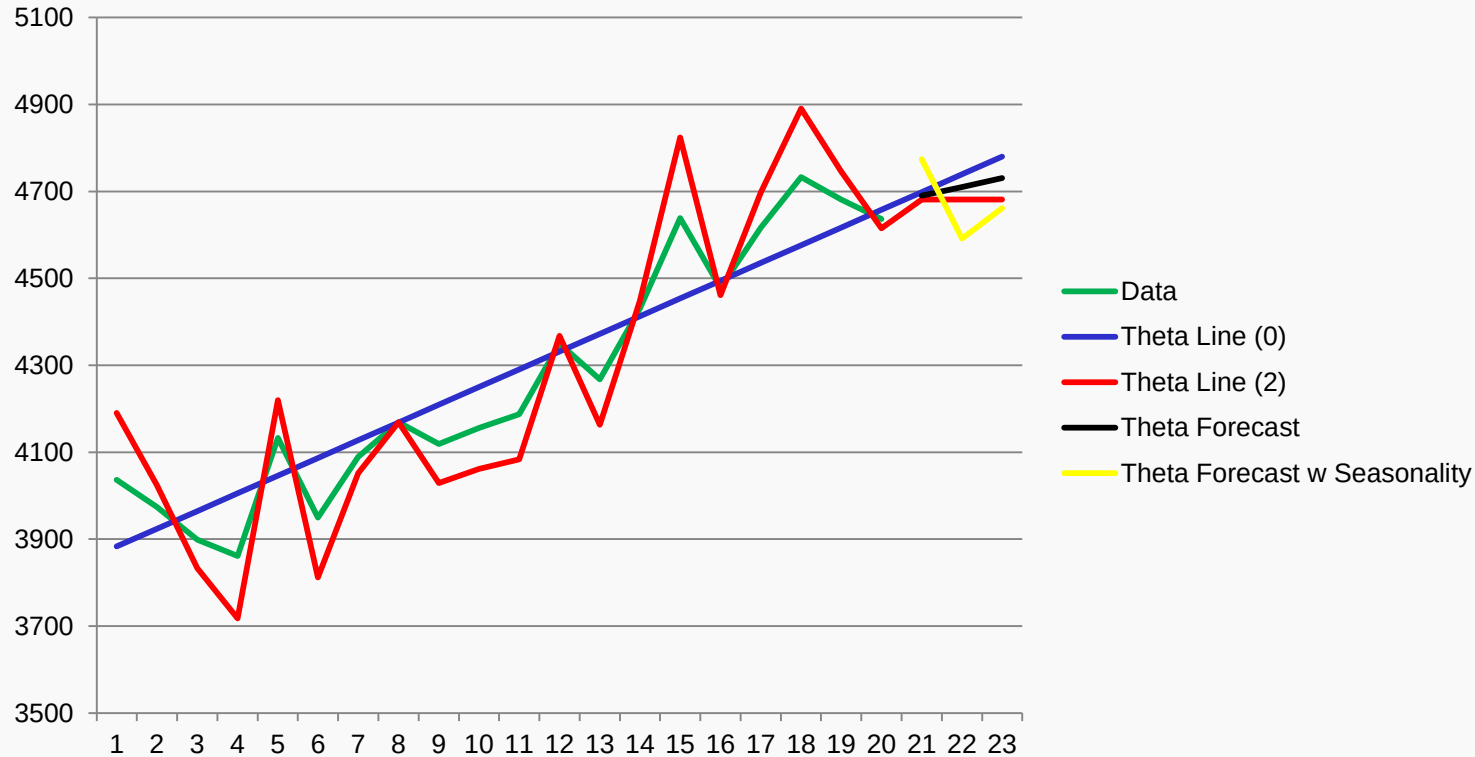
➤ Συνδυασμός προβλέψεων και εποχικοποίηση

Period	Data	ThetaLine(0)	ThetaLine(2)
1	4036,679	3846.42	4226.937
2	3974,128	3891.05	4057.213
3	3899,027	3935.67	3862.39
4	3861,435	3980.29	3742.567
5	4132,955	4024.92	4240.983
6	3949,508	4069.54	3829.48
7	4089,817	4114.16	4065.477
8	4168,706	4158.79	4178.633
9	4119,201	4203.41	4034.99
10	4155,703	4248.03	4063.367
11	4187,242	4292.66	4081.823
12	4349,741	4337.28	4362.2
13	4267,543	4381.90	4153.177
14	4430,63	4426.53	4434.733
15	4638,848	4471.15	4806.55
16	4477,934	4515.77	4440.087
17	4616,295	4560.40	4672.203
18	4733,255	4605.02	4861.48
19	4681,471	4649.64	4713.297
20	4636,462	4694.27	4578.653
21		4738.89	4648.55
22		4783.51	4648.55
23		4828.14	4648.55

Theta Forecast	SI	Theta Forecast w Seasonality
----------------	----	------------------------------

4693.72	101,79	4777.74
4716.03	97,48	4597.19
4738.34	98,54	4669.16

Παράδειγμα Theta (5/5)



Στόχοι Διαγωνισμών Πρόβλεψης

- Δημιουργία ερεθισμάτων στους ερευνητές για την υλοποίηση νέων και ακριβέστερων μεθόδων πρόβλεψης
- Σύγκριση και ταξινόμηση των μεθόδων πρόβλεψης με κριτήριο την ελαχιστοποίηση του σφάλματος
- Εύρεση της ακριβέστερης μεθόδου πρόβλεψης ανά τύπο χρονοσειράς (χαρακτηριστικά, είδος, συχνότητα, μήκος)
- Έλεγχος της εγκυρότητας των αποτελεσμάτων προηγούμενων διαγωνισμών πρόβλεψης

<https://robjhyndman.com/hyndsight/forecasting-competitions/>

Διαγωνισμός M1

Έτος διεξαγωγής 1982 / 1,001 χρονοσειρές / 15 μέθοδοι πρόβλεψης & 9 παραλλαγές

- Οι στατιστικά πολύπλοκες ή εξεζητημένες μέθοδοι δεν παράγουν απαραίτητα και ακριβέστερες προβλέψεις σε σχέση με τις πιο απλές.
- Η σχετική κατάταξη της απόδοσης των διαφόρων μεθόδων ποικίλει ανάλογα με το κριτήριο ακρίβειας που χρησιμοποιείται.
- Ο συνδυασμός απλών μεθόδων πρόβλεψης συνήθως οδηγεί σε καλύτερη ακρίβεια σε σχέση με αυτή των επιμέρους μεθόδων που χρησιμοποιήθηκαν.
- Η ακρίβεια πρόβλεψης εξαρτάται από την έκταση του ορίζοντα πρόβλεψης.

Διαγωνισμός M2

Έτος διεξαγωγής 1993 / 29 χρονοσειρές / 16 μέθοδοι πρόβλεψης & 3 συνδυασμοί μεθόδων

- Συνδυασμός στατιστικών προβλέψεων με κριτικές προβλέψεις βάσει εξωτερικής πληροφόρησης και εμπειρίας
- Συνεχής επικοινωνία αναλυτών με επιχειρήσεις και οργανισμούς για τη βελτίωση της προβλεπτικής ακρίβειας των μεθόδων.
- Εκπαίδευση των αναλυτών μέσω της επαναλαμβανόμενης κατάθεσης προβλέψεων σε πραγματικό χρόνο.
- Παρά τις επιπλέον πληροφορίες που δόθηκαν, οι βελτιώσεις στην ακρίβεια των προβλέψεων ήταν μικρές ή και ανύπαρκτες

Διαγωνισμός M3

Έτος διεξαγωγής 2000 / 3,003 χρονοσειρές / 24 μέθοδοι πρόβλεψης

- Επιβεβαιώθηκαν τα βασικά συμπεράσματα του διαγωνισμού M1
- Κατασκευάστηκε ένα μεγάλο σετ δεδομένων το οποίο αποτελεί ακόμα και σήμερα σημείο αναφοράς (benchmark) για τον έλεγχο της προβλεπτικής ακρίβειας νέων μεθόδων
- Καθιερώθηκε η χρήση γνωστών σετ δεδομένων και μεθόδων για την εξακρίβωση της προβλεπτικής ακρίβειας νέων μεθόδων
- Αναδείχτηκε το μοντέλο Θ ως μία νέα ακριβής μέθοδος πρόβλεψης

Διαγωνισμός M3

Συνοπτικά Αποτελέσματα

Μέθοδος Πρόβλεψης	SMAPE (σύνολο 1,428 μηνιαίων χρονοσειρών)	SMAPE (σύνολο 3,003 χρονοσειρών)
THETA	13.85	13.01
ForecastPro	13.86	13.19
ForcX	14.45	13.49
COMB S-H-D	14.48	13.52
DAMPEN	14.59	13.63
THETA _{sm}	14.66	13.88
RBF	14.77	13.75
B-J automatic	14.81	14.01
AutomatANN	14.93	14.11
SMARTFCS	15.03	14.13

Διαγωνισμός M4

Έτος διεξαγωγής 2018 | 100,000 χρονοσειρές | Αναμένεται να συμμετάσχουν μέθοδοι κάθε τύπου (Στατιστικές, Μηχανικής Μάθησης & Συνδυασμός) | Χρηματικά έπαθλα

Συχνότητα / Τύπος	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	6,538	3,716	3,903	6,519	1,088	1,236	23,000
Quarterly	6,020	4,637	5,315	5,305	1,858	865	24,000
Monthly	10,975	10,017	10,016	10,987	5,728	277	48,000
Weekly	112	6	41	164	24	12	359
Daily	1,476	422	127	1,559	10	633	4,227
Hourly	0	0	0	0	0	414	414
Total	25,121	18,798	19,402	24,534	8,708	3,437	100,000

<https://www.m4.unic.ac.cy/>

Διαγωνισμός M4

- Σύγκριση ακρίβειας διαφορετικών μεθόδων πρόβλεψης και κυρίως των στατιστικών με αυτές της μηχανικής μάθησης (Machine Learning)*
- Σύγκριση ακρίβειας μεθόδων αναλυτών (ακαδημαϊκοί, εταιρίες, φοιτητές κτλ.) με απλές μεθόδους αναφοράς.
- Αξιοποίηση πλήθους δεδομένων για την εξακρίβωση της στατιστικής σημαντικότητας των αποτελεσμάτων.
- Αξιολόγηση προκατάληψης διαφόρων τύπων μεθόδων.
- Αξιολόγηση ακρίβειας διαστημάτων εμπιστοσύνης διαφόρων προσεγγίσεων.
- Συσχέτιση της προβλεπτικής ακρίβειας των μεθόδων με την ικανότητα προσαρμογής τους και τις υπολογιστικές τους απαιτήσεις.
- Επαλήθευση των αποτελεσμάτων (replicability) μέσω της υποβολής πληροφοριών, περιγραφών και ανοιχτού κώδικα.

* Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). *Statistical and Machine Learning forecasting methods: Concerns and ways forward*. PLOS ONE 13(3): e0194889

Διαστήματα Εμπιστοσύνης

Στη Στατιστική, το διάστημα εμπιστοσύνης (*Confidence Interval – CI*) είναι ένα σύνολο τιμών εντός των οποίων αναμένεται να κινείται μία παράμετρος. Στην ουσία, αντί να εκτιμούμε την παράμετρο με μία και μόνο τιμή, αποδίδουμε ένα διάστημα τιμών εντός του οποίου αυτή αναμένεται βάσει ιστορικών στοιχείων να κυμαίνεται για μία προκαθορισμένη πιθανότητα. Συνεπώς, τα διαστήματα εμπιστοσύνης χρησιμοποιούνται για να υποδείξουν την εγκυρότητα της παραμέτρου που θέλουμε να προβλέψουμε. Η πιθανότητα της τιμής παραμέτρου να περιλαμβάνεται στο διάστημα εμπιστοσύνης καθορίζεται από το επίπεδο εμπιστοσύνης. Αυξάνοντας το επίπεδο εμπιστοσύνης, το αντίστοιχο διάστημα «πλαταίνει».

Για παράδειγμα, ένα διάστημα εμπιστοσύνης μπορεί να χρησιμοποιηθεί για να περιγράψει την εγκυρότητα των αποτελεσμάτων μιας δημοσκόπησης. Σε μια δημοσκόπηση για πρόθεση ψήφου, το αποτέλεσμα ήταν ότι το 40% των ερωτηθέντων προτίθεται να ψηφήσει το κόμμα Χ. Ένα διάστημα εμπιστοσύνης 95% θα υποδείκνυε ότι, βάσει δείγματος, η πραγματική πρόθεση ψήφου στο σύνολο του πληθυσμού για το κόμμα Χ είναι από 36% έως 44%. Συνεπώς, το αποτέλεσμα μιας δημοσκόπησης με μικρά διαστήματα εμπιστοσύνης θεωρείται πιο έγκυρο από αυτό μιας δημοσκόπησης με μεγάλα διαστήματα. Ένας από τους κύριους παράγοντες που επηρεάζουν το εκτιμώμενο εύρος είναι το μέγεθος του δείγματος των ερωτηθέντων.

Διαστήματα Εμπιστοσύνης

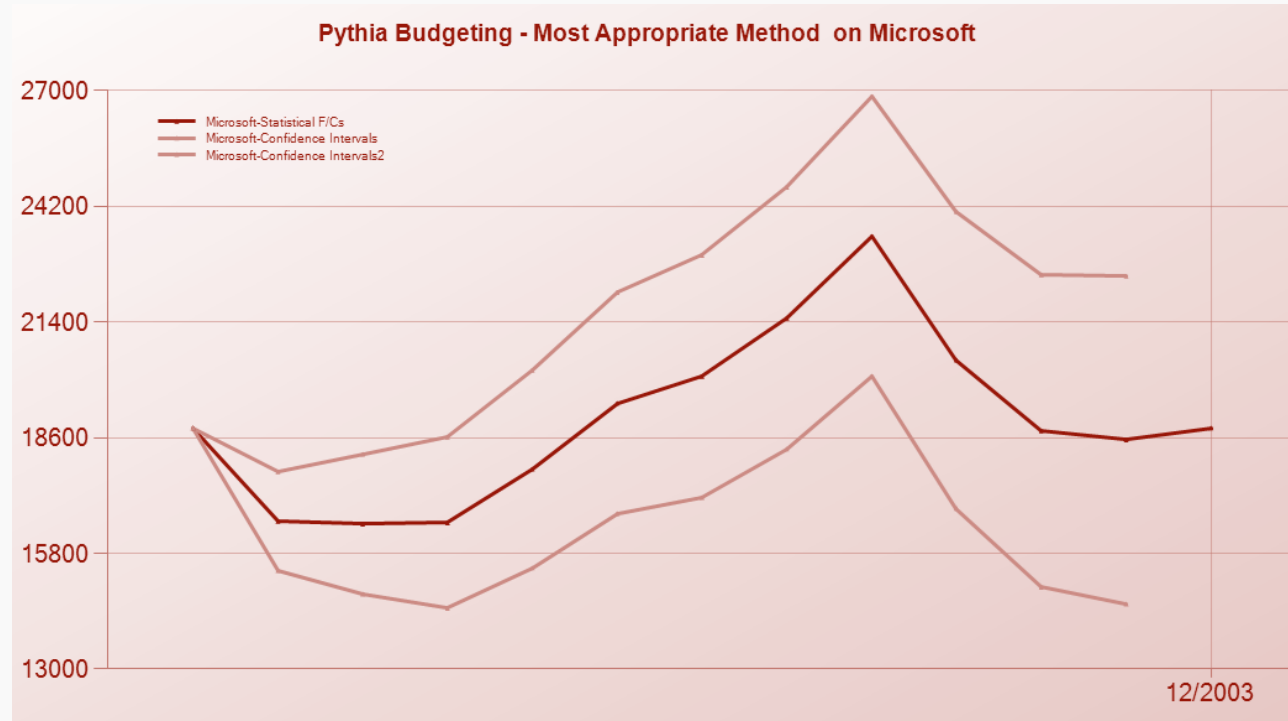
$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

$$F_i = F_i \pm t \cdot RMSE \cdot \sqrt{i - n}$$

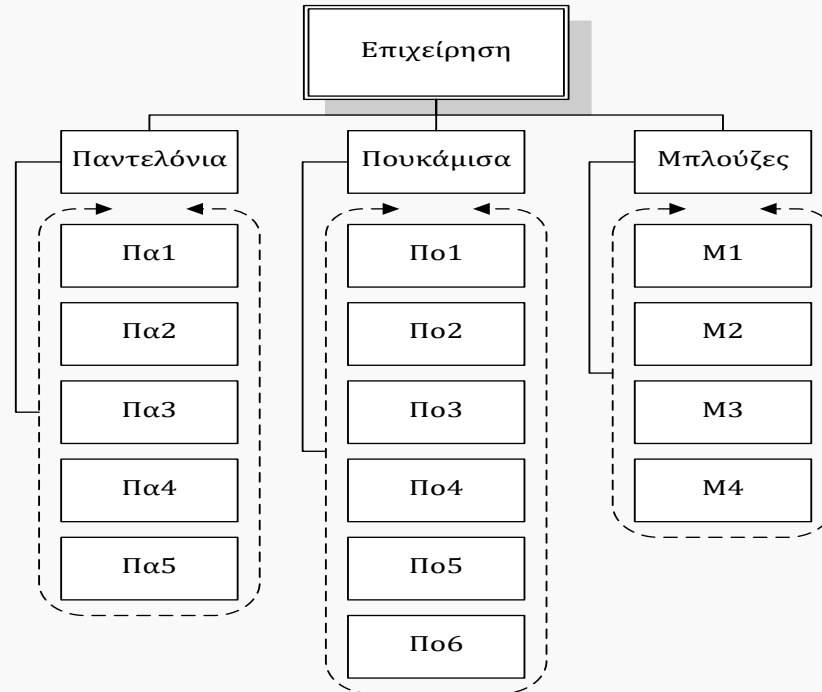
Confidence	t
99%	2.58
98%	2.33
95%	1.96
90%	1.645
80%	1.28

Όπου F είναι οι σημειακές προβλέψεις του μοντέλου, t είναι η παράμετρος εμπιστοσύνης, $RMSE$ είναι η ρίζα του μέσου τετραγωνικού σφάλματος και n είναι το σύνολο των διαθέσιμων παρατηρήσεων.

Διαστήματα Εμπιστοσύνης



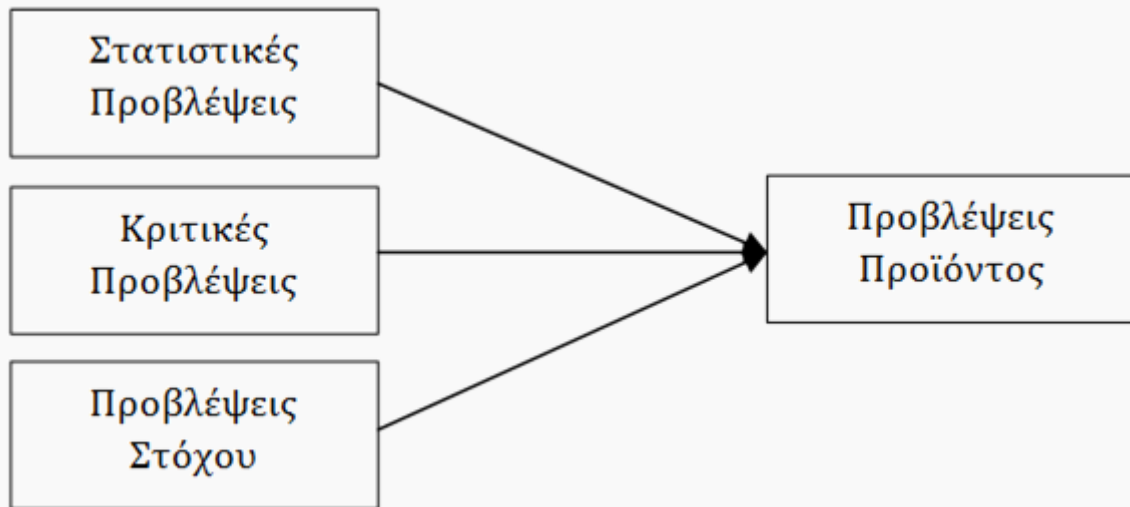
Η διαδικασία της πρόβλεψης στην επιχείρηση



Η διαδικασία της πρόβλεψης στην επιχείρηση

60

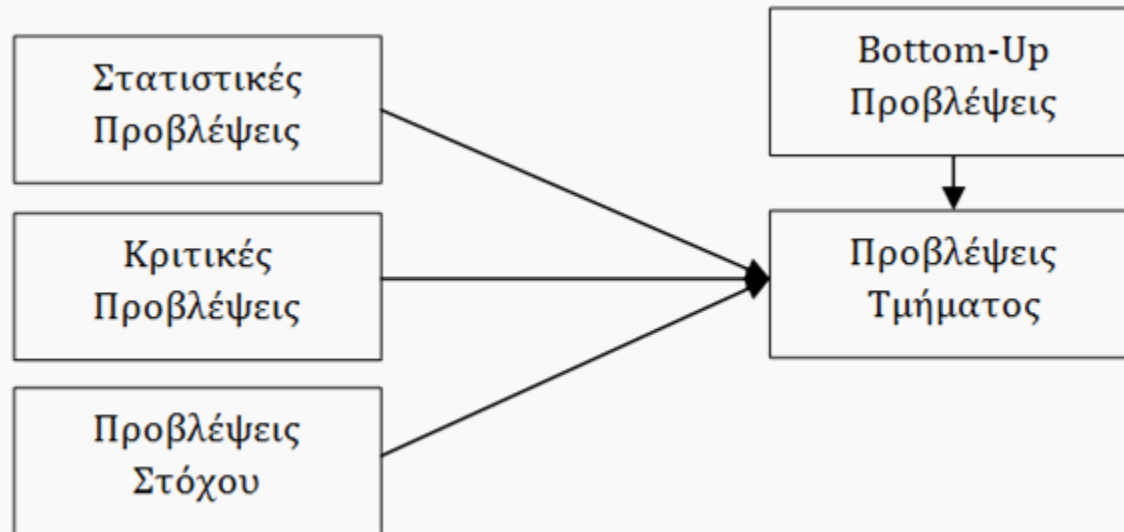
Πρόβλεψη Προϊόντος



Η διαδικασία της πρόβλεψης στην επιχείρηση

Πρόβλεψη Κατηγορίας/Τμήματος

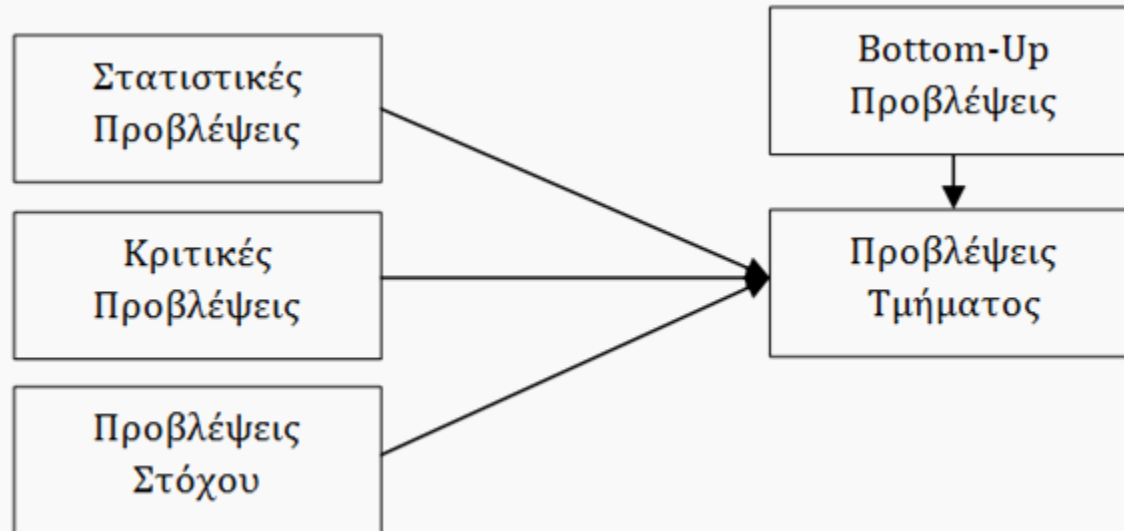
$$Y_t^{\text{department}} = \sum_{i=1}^{\text{pr}} Y_t^i \cdot p_t^i$$



Η διαδικασία της πρόβλεψης στην επιχείρηση

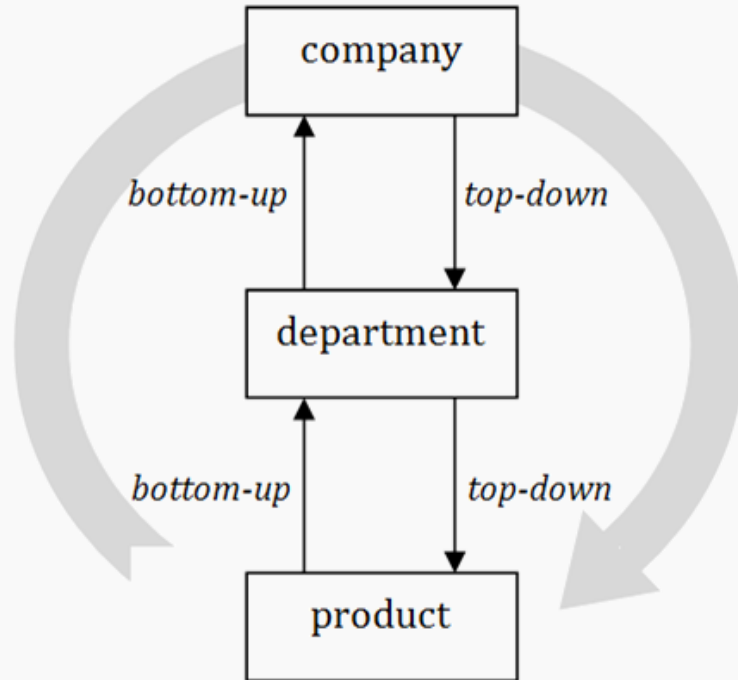
Πρόβλεψη Επιχείρησης

$$Y_t^{\text{company}} = \sum_{j=1}^d \sum_{i=1}^{\text{pr}} Y_t^{d,i} \cdot p_t^{d,i}$$



Η διαδικασία της πρόβλεψης στην επιχείρηση

63





Fell free to say hi!

We are friendly and social

Ηρώων Πολυτεχνείου 9, Ζωγράφος

Αττική, 15780, Ελλάδα

E-mail: [info\(at\)fsu.gr](mailto:info(at)fsu.gr)

Τηλέφωνο: 2107723637 Fax: 2107723740

Κτίριο της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
2ος όροφος - 2.2.1 Εργαστήριο



@FSU NTUA



Μονάδα
Προβλέψεων και
Στρατηγικής ΕΜΠ



lesson@fsu.gr

