

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



Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές
Μέθοδος Theta – Διαγωνισμοί Προβλέψεων Διάλεξη 9

Το Μοντέλο Θ (1/3)

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

- Η μέθοδος Θ (Assimakopoulos & Nikolopoulos, 2000) είναι μια μονοδιάστατη μέθοδος πρόβλεψης.
- Η μέθοδος βασίζεται στην μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς μέσα από την παράμετρο θ (Theta), η οποία εφαρμόζεται (πολλαπλασιαστικά) στις διαφορές δεύτερης τάξης των δεδομένων.
- Η καινούργια χρονοσειρά που δημιουργείται διατηρεί τη μέση τιμή και την κλίση (LRL) της αρχικής χρονοσειράς αλλά όχι και τις τοπικές της καμπυλότητες. Οι χρονοσειρές που παράγονται με αυτή την διαδικασία ονομάζονται γραμμές Θ (Theta Lines).
- Βασικό ποιοτικό χαρακτηριστικό αυτών των γραμμών είναι η καλύτερη προσέγγιση της μακροπρόθεσμης συμπεριφοράς-τάσης των δεδομένων ή η ανάδειξη-τονισμός των βραχυπρόθεσμων χαρακτηριστικών τους, ανάλογα με την τιμή της παραμέτρου θ ($<$, $>$ 1).

Το Μοντέλο Θ (2/3)

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

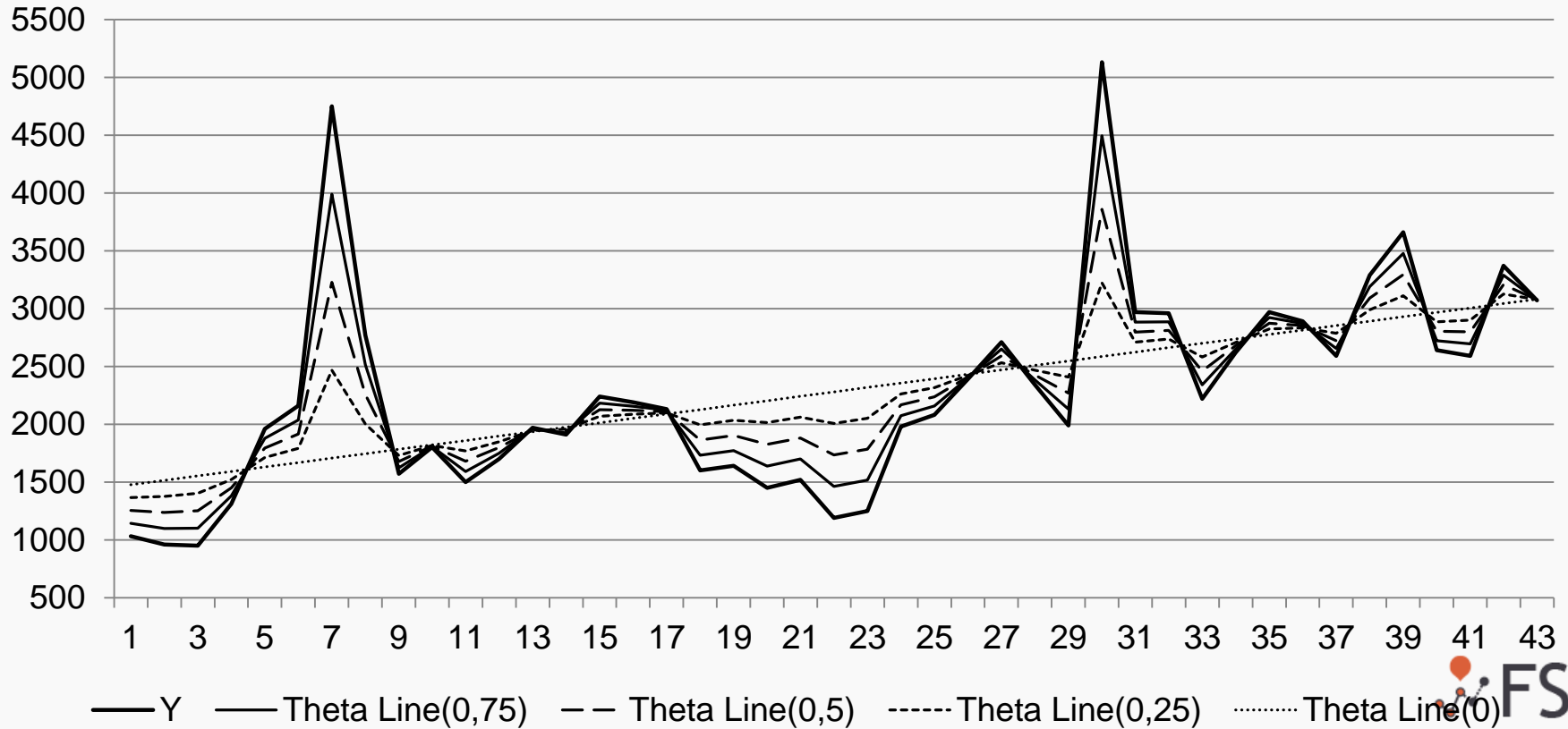
- Η προτεινόμενη μέθοδος αποσυνθέτει (διαχωρίζει) την αρχική χρονοσειρά σε δύο ή περισσότερες γραμμές Theta.
- Η κάθε γραμμή Theta προεκτείνεται στο μέλλον ξεχωριστά (με την ίδια ή και με διαφορετικές μεθόδους πρόβλεψης) και οι παραγόμενες προβλέψεις συνδυάζονται για να προκύψει η τελική πρόβλεψη.
- Ο απλός συνδυασμός δύο γραμμών Theta, 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 et al., 2000).

Το Μοντέλο Θ (3/3)

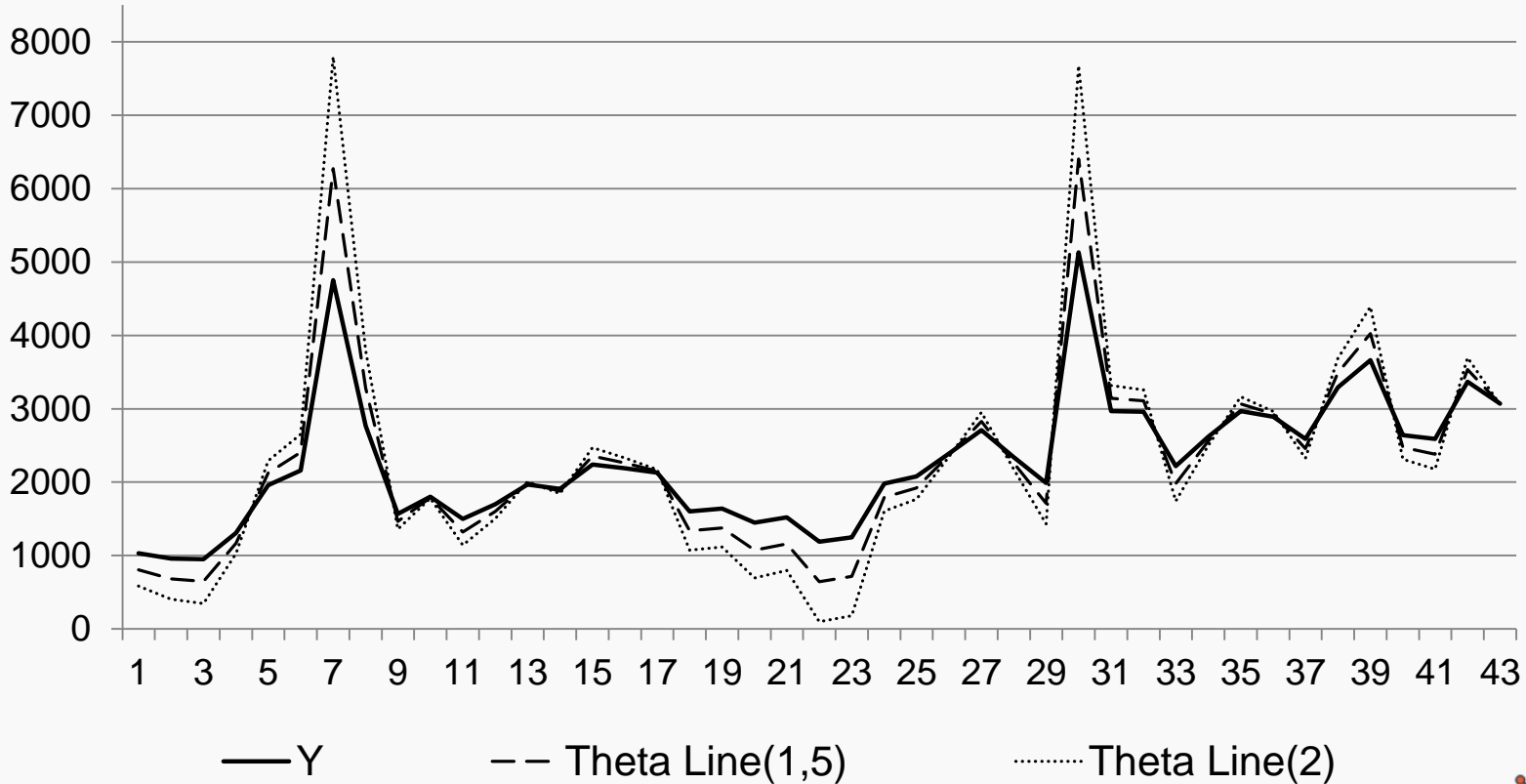
Η Μέθοδος Θ : Μια μέθοδος παραγωγής προβλέψεων βασισμένη σε μια διαφορετική προσέγγιση αποσύνθεσης

- 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.
- 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.

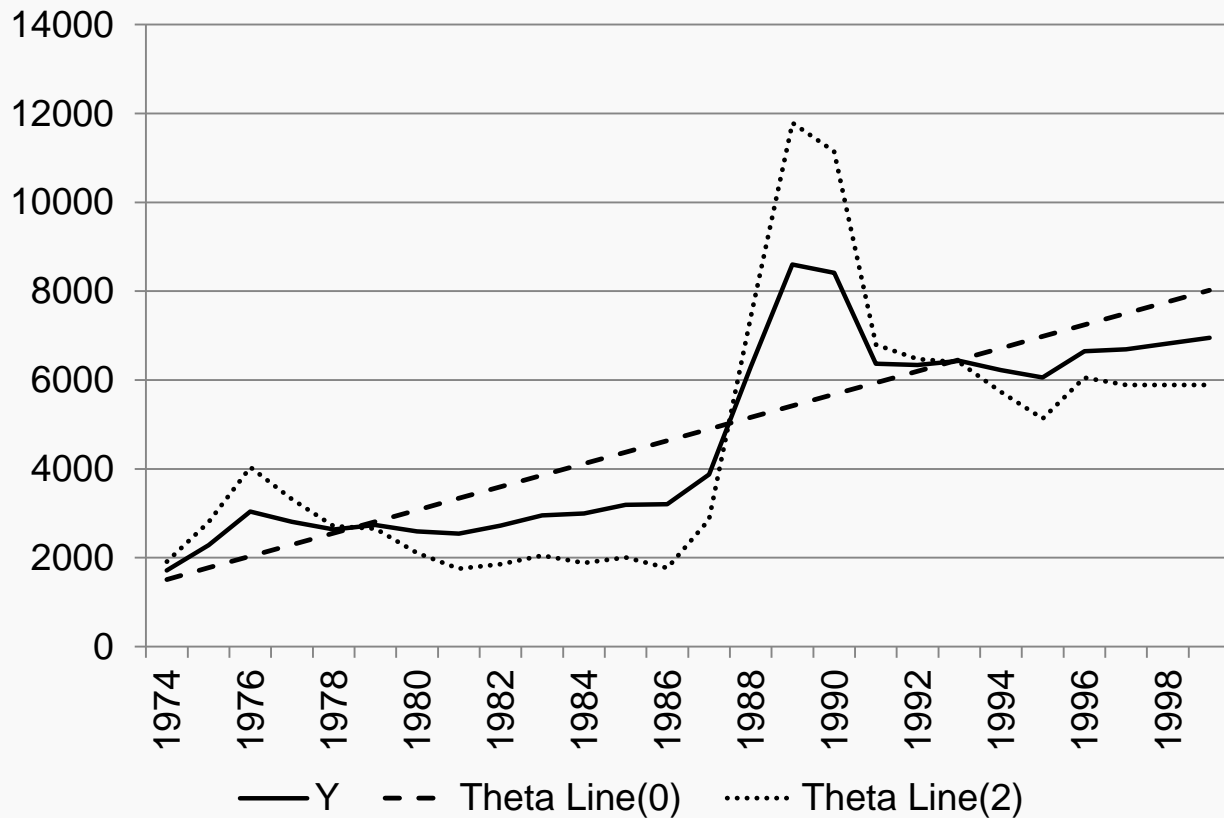
Γραμμές Θ



Γραμμές Θ



Το Κλασικό Μοντέλο Θ



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

Το κλασσικό Μοντέλο θ στην πράξη

- **Βήμα 0. Τεστ Εποχιακότητας**

Η κάθε χρονοσειρά ελέγχεται για εποχιακή συμπεριφορά με κριτήριο την τιμή του συντελεστή αυτοσυσχέτισης με καθυστέρηση ένα έτος (π.χ. για μηνιαία δεδομένα 12) συγκρινόμενη με την τιμή 1.645 (τιμή της t-κατανομής για πιθανότητα 0.1)

- **Βήμα 1. Αποεποχικοποίηση**

Η χρονοσειρά αποεποχικοποιείται με την κλασσική μέθοδο αποσύνθεσης

- **Βήμα 2. Αποσύνθεση**

Κάθε χρονοσειρά αποσυντίθεται σε δύο γραμμές θ , για $\theta=0$ και $\theta=2$.

- **Βήμα 3. Πρόβλεψη**

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

- **Βήμα 4. Συνδυασμός**

Οι προηγούμενες προβλέψεις συνδυάζονται με ίσα βάρη

- **Βήμα 5. Εποχικοποίηση**

Οι τελικές προβλέψεις εποχικοποιούνται

Μοντέλο Θ

Hyndman & Billah (2003). Unmasking the Theta method. *International Journal of Forecasting*, 19, 287-290:

“SES with drift?”

Υπολογίζοντας τις γραμμές Theta

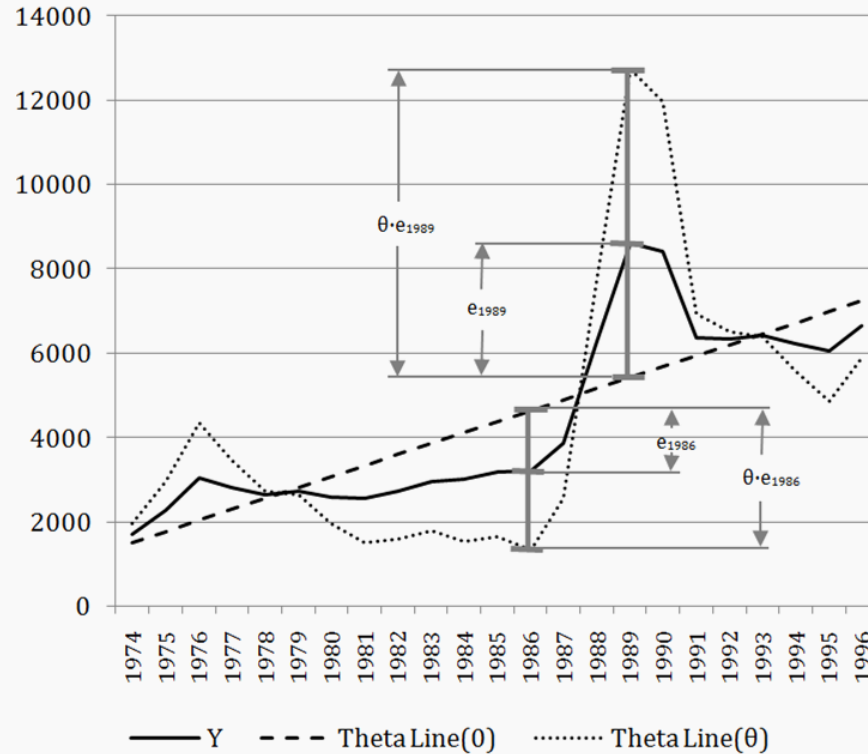
Για το κλασικό μοντέλο Theta (παράμετροι 0 και 2) οι γραμμές Theta υπολογίζονται ως εξής:

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

Γενικότερα ισχύει:

- $\text{Theta Line}(\theta) = \theta \times \text{Data} + (1-\theta) \times \text{LRL}$

Υπολογίζοντας τις γραμμές Theta



Παράδειγμα Theta (1/5)

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

	min	max	average (w/o min & max)	SI
	105,58	101,15	99,54	101,79
	95,12	96,90	97,36	99,32
97,14	98,65	97,71	101,65	
99,54	103,69	103,98	99,95	
	99,54	105,58	101,43	101,79
	95,12	99,32	97,13	97,48
	97,14	101,65	98,18	98,54
	99,54	103,98	101,82	102,19

sum	398,559
Σ.K.	0,9964

Παράδειγμα Theta (2/5)

X		Numerator			Denominator	ThetaLine(0)	
Period	Y Data	X-Mean(X)=A	Y-Mean(Y)=B	A*B	(X-Mean(X))^2	b=slope	LRL
1	4036,68	-9,5	-233,67	2219,87	90,25	444.623	3846.42
2	3974,13	-8,5	-296,22	2517,88	72,25	3801.8	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



Παράδειγμα Theta (3/5)

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

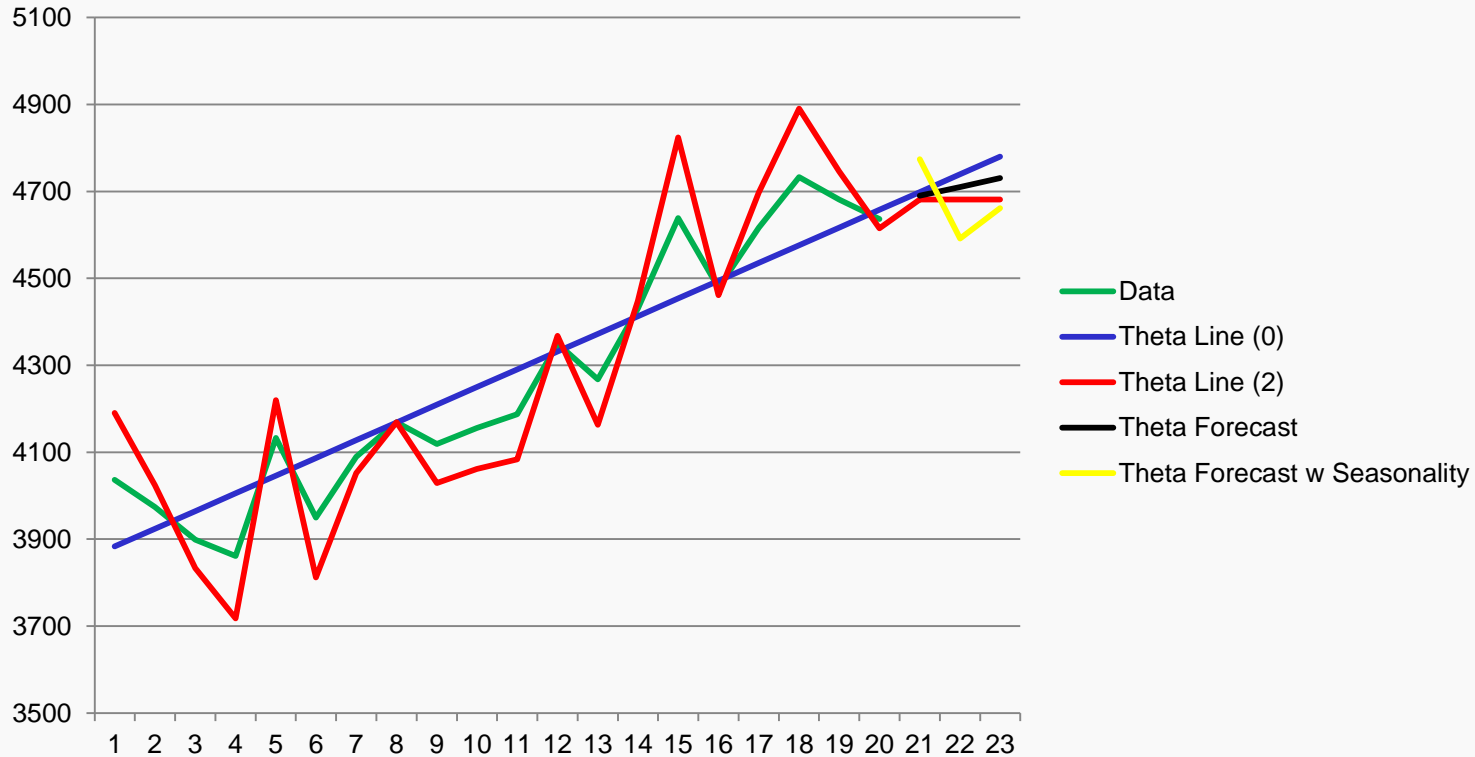
Παράδειγμα Theta (4/5)

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

- 01** Δημιουργία ερεθισμάτων στους ερευνητές για την υλοποίηση νέων μεθόδων πρόβλεψης
- 02** Σύγκριση και ταξινόμηση των μεθόδων πρόβλεψης με κριτήριο την ελαχιστοποίηση του σφάλματος
- 03** Έλεγχος της εγκυρότητας των αποτελεσμάτων προηγούμενων διαγωνισμών πρόβλεψης

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 (1,428 μηνιαίες χρονοσειρές)	sMAPE (3,003 χρονοσειρές)
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 | 61 μέθοδοι πρόβλεψης

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
<i>Total</i>	<i>25,121</i>	<i>18,798</i>	<i>19,402</i>	<i>24,534</i>	<i>8,708</i>	<i>3,437</i>	<i>100,000</i>

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.

Διαγωνισμός M4

- Ο συνδυασμός μεθόδων πρόβλεψης συνεχίζει να είναι η αποτελεσματικότερη στρατηγική παραγωγής προβλέψεων υψηλής ακρίβειας.
- Η χρήση «υβριδίων» που συνδυάζουν μεθόδους Μηχανικής Μάθησης με στατιστικές, αποδίδει βέλτιστα.
- Το σύνολο των μεθόδων υποεκτιμά την αβεβαιότητα του μέλλοντος. Ωστόσο, υπάρχουν κάποιες που την εκτιμούν με υψηλή ακρίβεια.
- Πιο πολύπλοκες μέθοδοι ενδέχεται να οδηγούν σε υψηλότερη ακρίβεια.
- Η αξιοποίηση πληροφοριών που προέρχονται από πολλές χρονοσειρές (cross-learning) οδηγεί σε καλύτερα αποτελέσματα από την παραγωγή προβλέψεων μέσω μίας και μόνο χρονοσειράς.
- Οι μέθοδοι πρόβλεψης που βασίζονται αποκλειστικά σε αλγόριθμους Μηχανικής Μάθησης δεν παράγουν ακριβείς προβλέψεις.

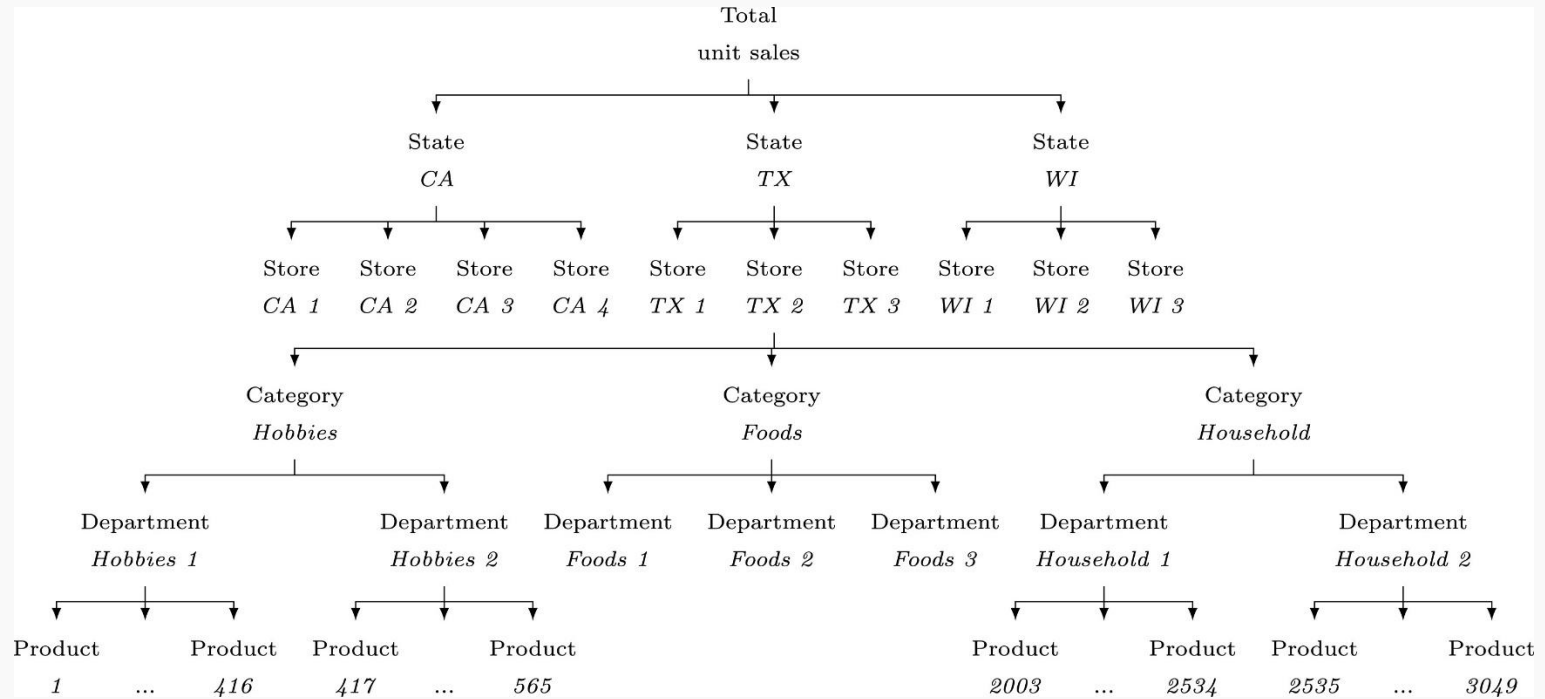
Διαγωνισμός M5

Έτος διεξαγωγής 2020 | 42,840 χρονοσειρές | Περισσότερες από 6000 συμμετοχές

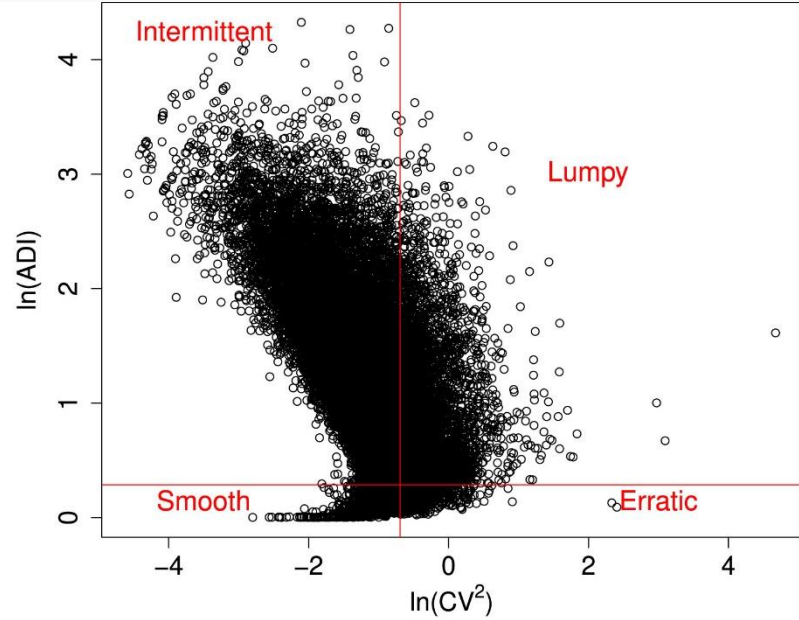
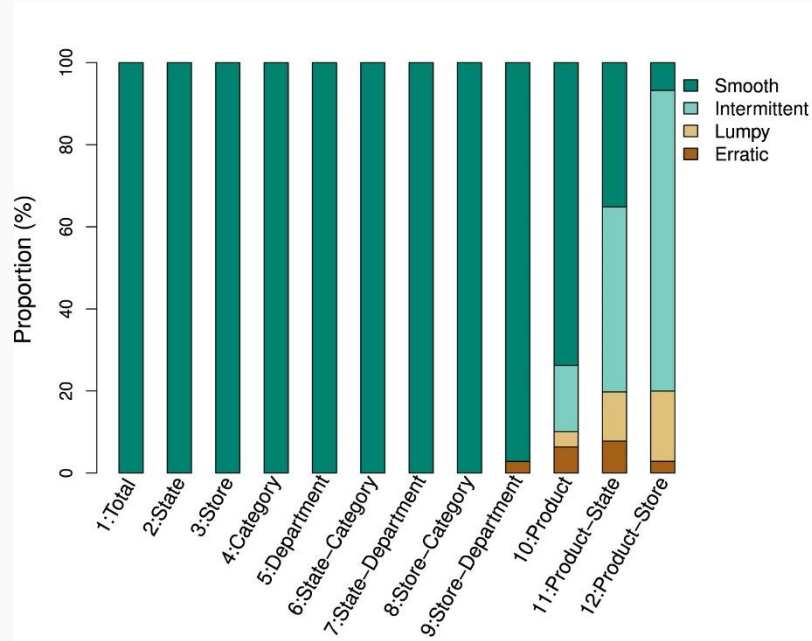
- Ιεραρχικά δεδομένα πωλήσεων που αφορούν επίπεδα προϊόντων, καταστημάτων, κατηγοριών και υποκατηγοριών προϊόντων, γεωγραφικές περιοχές κ.α.
- Εξωγενείς μεταβλητές όπως τιμή προϊόντος, προωθητικές ενέργειες, ημερομηνίες, γιορτές και αργίες.
- Αξιολόγηση τόσο σημειακών προβλέψεων όσο και πιθανοτικών προβλέψεων (50%, 67%, 95%, 99% και median)
- Έμφαση σε δεδομένα διακοπτόμενης ζήτησης
- Τα δεδομένα προέρχονται από τη Walmart και αφορούν 3,049 κωδικούς προϊόντων
- Ξεχωριστοί διαγωνισμοί για σημειακές και πιθανοτικές προβλέψεις

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation, International Journal of Forecasting, 38(4), 1325-1336

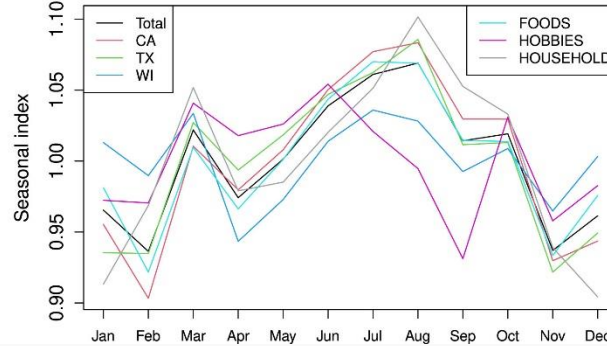
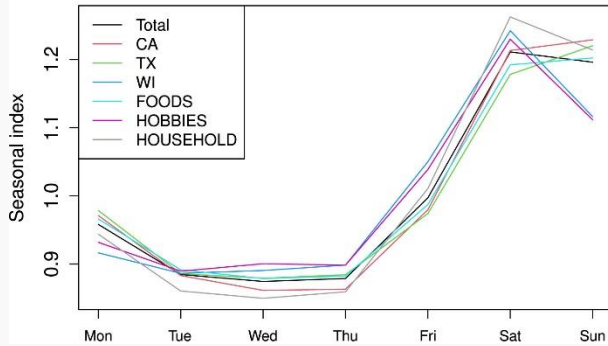
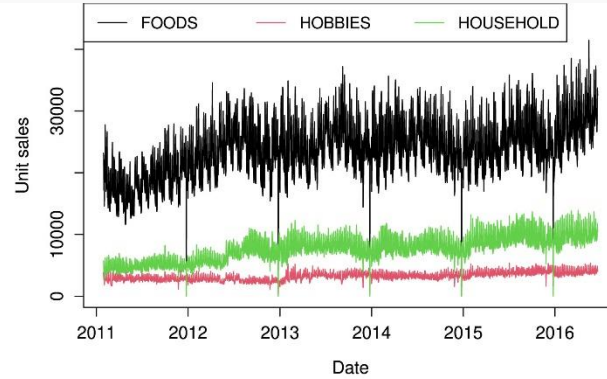
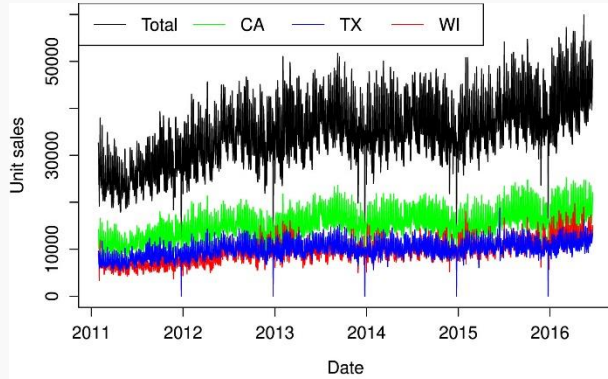
Διαγωνισμός M5



Διαγωνισμός M5



Διαγωνισμός M5



Διαγωνισμός M5

- Οι μέθοδοι μηχανικής μάθησης και κυρίως τα δέντρα αποφάσεων (π.χ. LightGBM) οδηγούν σε σημαντικά καλύτερες προβλέψεις.
- Ο συνδυασμός μεθόδων πρόβλεψης, στατιστικών αλλά κυρίως μηχανικής μάθησης, βελτιώνει περαιτέρω την ακρίβεια πρόβλεψης, ακόμα και αν αυτός πραγματοποιείται με απλές τεχνικές
- Η αξιοποίηση πληροφοριών που προέρχονται από πολλές χρονοσειρές (cross-learning) επιβεβαιώνεται ότι οδηγεί σε καλύτερα αποτελέσματα, ειδικά όταν οι χρονοσειρές ομαδοποιούνται κατάλληλα (π.χ. βάσει παρόμοιων χαρακτηριστικών)
- Η προσαρμογή των προβλέψεων μέσω κατάλληλων δεικτών διόρθωσης τάσης και επιπέδου μπορεί να βελτιώσει την ακρίβεια των προβλέψεων
- Η βελτιστοποίηση των μοντέλων μέσω τεχνικών cross-validation είναι καθοριστικής σημασίας για την κατασκευή εύρωστων προσεγγίσεων πρόβλεψης
- Η χρήση εξωγενών μεταβλητών βοηθά σημαντικά στη βελτίωση της ακρίβειας
- Οι βελτιώσεις των νικητήριων ομάδων επί των κλασικών προσεγγίσεων ήταν σημαντικές συνολικά, όχι όμως σε όλα τα επίπεδα και ποσοστημόρια πρόβλεψης



Feel free to say hi!

We are friendly and social

Ηρώων Πολυτεχνείου 9, Ζωγράφος
Αττική, 15780, Ελλάδα
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



spiliotis@fsu.gr