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

Τεχνικές Προβλέψεων

Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου (ARIMA)

Δύο λόγια για τα μοντέλα

- Ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων (***A**uto **R**egressive **I**ntegrated **M**oving **A**verage*)
- Ανήκουν στα στοχαστικά μοντέλα πρόβλεψης
- Μελετήθηκαν από τους *Box & Jenkins* (1971) και συχνά συναντώνται στη βιβλιογραφία με την αντίστοιχη ονομασία
- Προσεγγίζουν τη λογική των κλασικών μοντέλων παλινδρόμησης (π.χ. *LRL*) και εκθετικής εξομάλυνσης (π.χ. SES) με την έννοια ότι συσχετίζουν τις μελλοντικές τιμές τις χρονοσειρές με παρελθοντικές της ή/και σφάλματα που εντοπίστηκαν. Η ιδιομορφία τους έγκειται στο ότι η γραμμική συσχέτιση γίνεται χωρίς την άμεση χρήση εξομάλυνσης ή την αξιοποίηση ερμηνευτικών μεταβλητών.

Λογική ARIMA vs. LRL vs. ES

LRL: Η τιμή του μεγέθους *Y* τη χρονική στιγμή *t+1* θα είναι ανάλογη του χρόνου την αντίστοιχη χρονική περίοδο

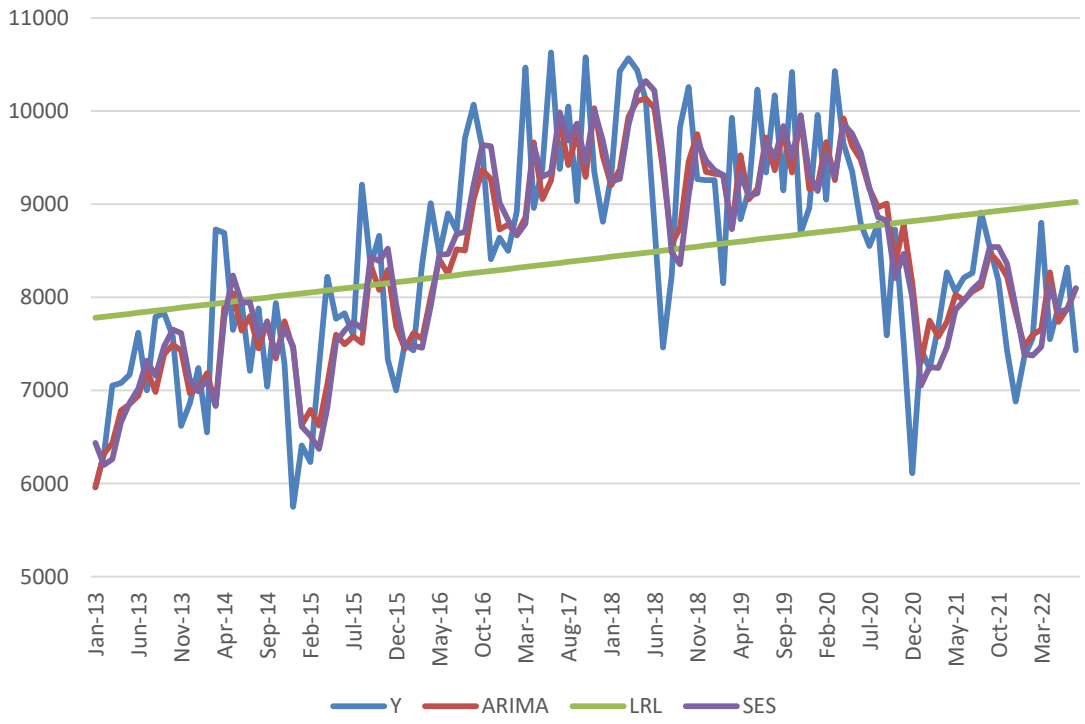
$$\widehat{Y_{t+1}} = 7768 + 11(t+1)$$

Exponential Smoothing: Η τιμή του μεγέθους *Y* τη χρονική στιγμή *t+1* θα είναι ίδια με της περιόδου *t*, αναθεωρώντας τη βάσει του σφάλματος που εντοπίστηκε νωρίτερα

$$\widehat{Y_{t+1}} = \hat{Y}_t + 0.48e_t$$

ARIMA: Η τιμή του μεγέθους *Y* τη χρονική στιγμή *t+1* θα είναι ανάλογη των προηγουμένων της τιμών, αναθεωρώντας τη βάσει του σφάλματος που εντοπίστηκε νωρίτερα

$$\widehat{Y_{t+1}} = 1.31Y_{t-1} - 0.31Y_{t-2} + 0.8e_t$$



FORECASTING & STRATEGY UNIT

Στοχαστικοί Παράγοντες ARIMA (1/3)

- Ο τυχαίος παράγοντας (*σφάλμα πρόβλεψης* e_t)
- Προηγούμενες παρατηρήσεις (Y_{t-1}, Y_{t-2}...)
- Άλλοι παράγοντες (π.χ. σταθερές και εξωτερικές μεταβλητές - ARIMAX)

Στόχος: Η εύρεση του βέλτιστου **γραμμικού συνδυασμού** των παραπάνω παραγόντων για την προέκταση της χρονοσειράς

FORECASTING & STRATEGY UNIT

Στοχαστικοί Παράγοντες ARIMA (2/3)

Μη εποχιακά μοντέλα ARIMA(p,d,q)

$$\left(1 - \varphi_1 B - \cdots - \varphi_p B^p\right)(1 - B)^n y_t = c + \left(\theta_1 B + \cdots + \theta_q B^q\right) e_t$$

Παράγοντες AR(p) | **Διαφόριση(d)** | **Σταθερά** | **Παράγοντες MA(q)**

FORECASTING & STRATEGY UNIT

Στοχαστικοί Παράγοντες ARIMA (3/3)

Εποχιακά μοντέλα ARIMA(p,d,q)(P,D,Q)

ARIMA	(p, d, q)	$(P, D, Q)_m$
	Μη εποχιακοί παράγοντες	Εποχιακοί παράγοντες

$$(1 - \varphi_1 B - \cdots - \varphi_p B^p)(1 - B)^n (1 - B^m)^N y_t = c + (\theta_1 B + \cdots + \theta_q B^q) e_t$$

Παράδειγμα ARIMA(1,1,1)(1,1,1), m=4

$$(1 - \varphi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)\, y_t = c + (\theta_1 B)(\Theta_1 B^4) e_t$$

- Μη εποχιακό AR(1)
- Εποχιακό AR(1)
- Μη εποχιακή Διαφόριση
- Εποχιακή Διαφόριση
- Μη εποχιακό MA(1)
- Εποχιακό MA(1)

FORECASTING & STRATEGY UNIT

Απαιτήσεις – Περιορισμοί (1/2)

Προκειμένου ένα μοντέλο ARMA να χρησιμοποιηθεί αποδοτικά θα πρέπει:

- Η χρονοσειρά να είναι **Διακριτή:** *Ισαπέχουσες παρατηρήσεις*
- Η χρονοσειρά να είναι **Στάσιμη:** *Σταθερή μέση τιμή, διακύμανση και συνάρτηση αυτοσυσχέτισης στο δείγμα*

  Οι τιμές y_{t1}, y_{t2} … y_{tn} ταυτίζονται στατιστικά με τις $y_{t1+\tau}$, $y_{t2+\tau}$ … $y_{tn+\tau}$
- Να παράγονται **βραχυπρόθεσμες** προβλέψεις

FORECASTING & STRATEGY UNIT

Απαιτήσεις – Περιορισμοί (2/2)

- **Διακριτότητα**

Η πρόβλεψη εξαρτάται από *p* παρελθοντικές παρατηρήσεις οι οποίες και οφείλουν για αυτό το λόγο να είναι γνωστές και ισαπέχουσες χρονικά – *Συμπλήρωση κενών τιμών*

- **Στασιμότητα**

Τα μοντέλα ARMA δεν μπορούν να μοντελοποιήσουν από μόνα τους ακραίες τιμές, τάση και εποχιακότητα – *Μετασχηματισμοί και διαφόριση*

- **Βραχυπρόθεσμες προβλέψεις**

Οι μακροπρόθεσμες προβλέψεις βασίζονται αποκλειστικά στις βραχυπρόθεσμες προβλέψεις και όχι στα πραγματικά δεδομένα. Συνεπώς το παραγόμενο σφάλμα αυξάνει σημαντικά καθώς μεγαλώνει ο ορίζοντας πρόβλεψης.

FORECASTING & STRATEGY UNIT

Επιλογή μοντέλου ARIMA

Η πορεία επιλογής ενός προβλεπτικά άρτιου μοντέλου ARIMA περιλαμβάνει τρία στάδια:

- **Αναγνώριση**: Εύρεση πιθανών αντιπροσωπευτικών μοντέλων (*στατιστική ανάλυση ή παρατήρηση διαγραμμάτων αυτοσυσχέτισης*)
- **Εκτίμηση**: Υπολογισμός των παραμέτρων *p, d και q* (*προσδοκώμενη πιθανοφάνεια, information criteria, μέθοδος ελαχίστων τετραγώνων κ.α.*)
- **Διαγνωστικός έλεγχος**: Έλεγχος στατιστικής σημαντικότητας παραμέτρων (*t-test, έλεγχος αυτοσυσχέτισης σφαλμάτων κ.α.*)

FORECASTING & STRATEGY UNIT

Επεξεργασία χρονοσειράς

Μετασχηματισμοί: *Περιορισμός τυχαίων διακυμάνσεων και απαλοιφή ακραίων τιμών χρονοσειράς*

Πότε; Όταν έχω έντονες μη συστηματικές διακυμάνσεις στη χρονοσειρά ή το σφάλμα πρόβλεψης που προκύπτει είναι αυξημένο παρά τη φαινομενικά σωστή επιλογή μοντέλου

Διαφόριση: *Περιορισμός των συστηματικών διακυμάνσεων επιπέδου (τάσης και εποχιακότητας)*

Πότε; Όταν έχω εμφανή τάση ή εποχιακότητα στη χρονοσειρά.

Τα μοντέλα ARIMA αντιμετωπίζουν αποτελεσματικά από μόνα τους την τάσης και την εποχιακότητα χωρίς να απαιτείται προετοιμασία της μέσω μεθόδων αποσύνθεσης. Οι μετασχηματισμοί είναι η μόνη επεξεργασία που μπορεί να προηγηθεί της πρόβλεψης σε αυτήν την περίπτωση.

FORECASTING & STRATEGY UNIT

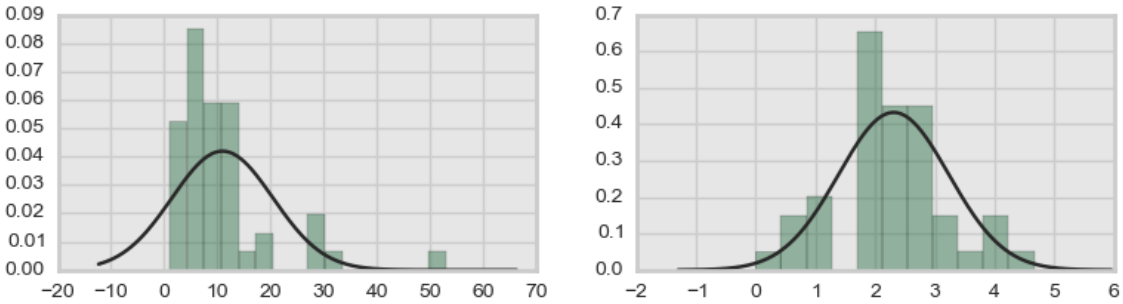
Μετασχηματισμοί (1/2)

Όταν η διασπορά των χρονοσειρών είναι αυξημένη (υψηλά επίπεδα θορύβου) η προβλεπτική ικανότητα των μοντέλων πρόβλεψης περιορίζεται καθώς δεν μπορούν να εξηγήσουν αποτελεσματικά τις τυχαίες διακυμάνσεις.

Το παραπάνω δεν αποτελεί πρόβλημα για τα μοντέλα ARMA όταν ο θόρυβος είναι λευκός. Όταν όμως αυτό δεν ισχύει (*outliers*) επηρεάζονται σημαντικά.

Λύση: Κανονικοποίηση μέσω λογαρίθμισης ή άλλων μετασχηματισμών π.χ. *Box-Cox*.

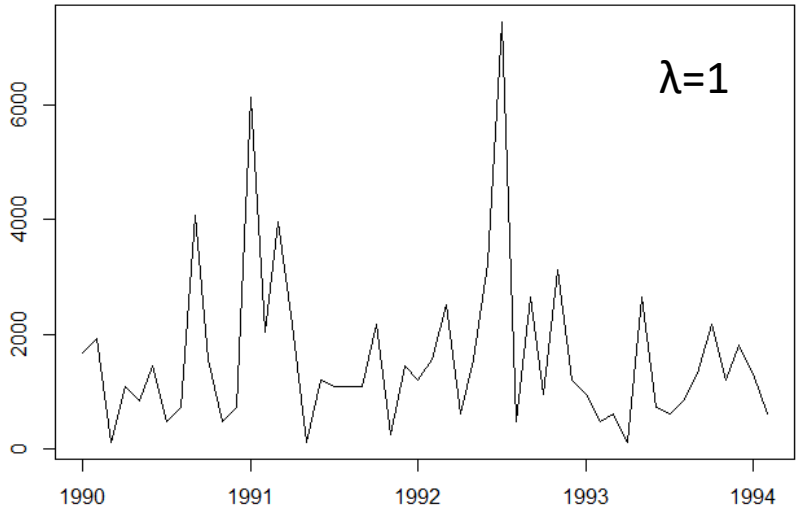
$$x_t = \frac{{y_t}^{\lambda} - 1}{\lambda}$$



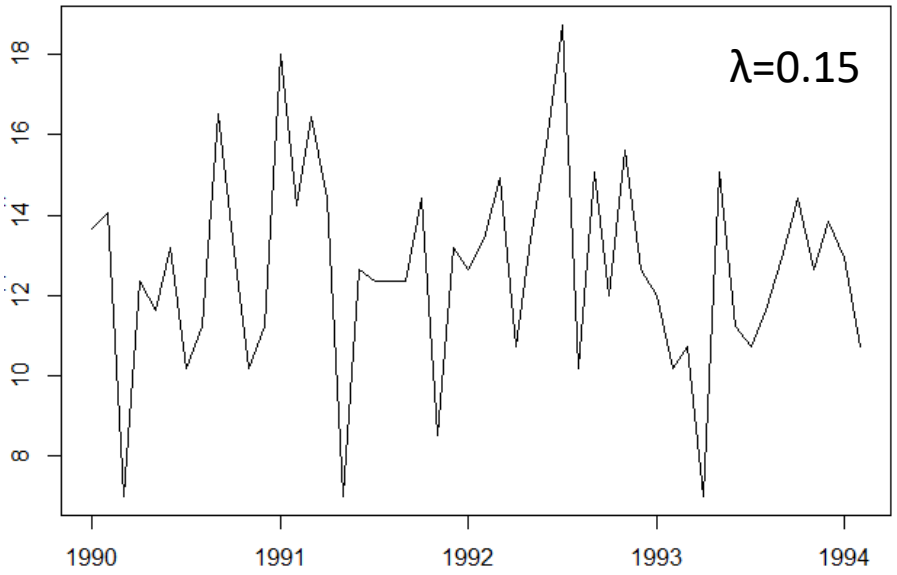
Σύγκριση γραφήματος της περιθώριας κατανομής μίας μη κανονικής χρονοσειράς πριν (αριστερά) και μετά (δεξιά) την εφαρμογή του μετασχηματισμού Box-Cox

FORECASTING & STRATEGY UNIT

Μετασχηματισμοί (2/2)



Sd/mean=0.89



Sd/mean=0.2

Διαφόριση (1/2)

Επιτυγχάνεται μέσω του παράγοντα ***I(d)*** και έχει ως στόχο την επίτευξη στασιμότητας, είτε μέσω της απαλοιφής της τάσης, είτε μέσω της απομάκρυνσης περιοδικών διακυμάνσεων (εποχιακότητας).

Δεδομένης χρονοσειράς *n* παρατηρήσεων, η διαφόριση έγκειται στη δημιουργία μίας νέας με στοιχεία της τις διαφορές των παρατηρήσεων της πρώτης. Μπορεί να είναι από *1ης* έως *n-1* τάξης. Στη πράξη χρησιμοποιείται μέχρι *2ης* τάξης.

- 1η τάξη: $y'_t=y_t-y_{t-1}$ για απαλοιφή γραμμικής τάσης
- 2η τάξη: $y''_t=y'_t-y'_{t-1}=y_t-2y_{t-1}+y_{t-2}$ για απαλοιφή μεταβαλλόμενης τάσης

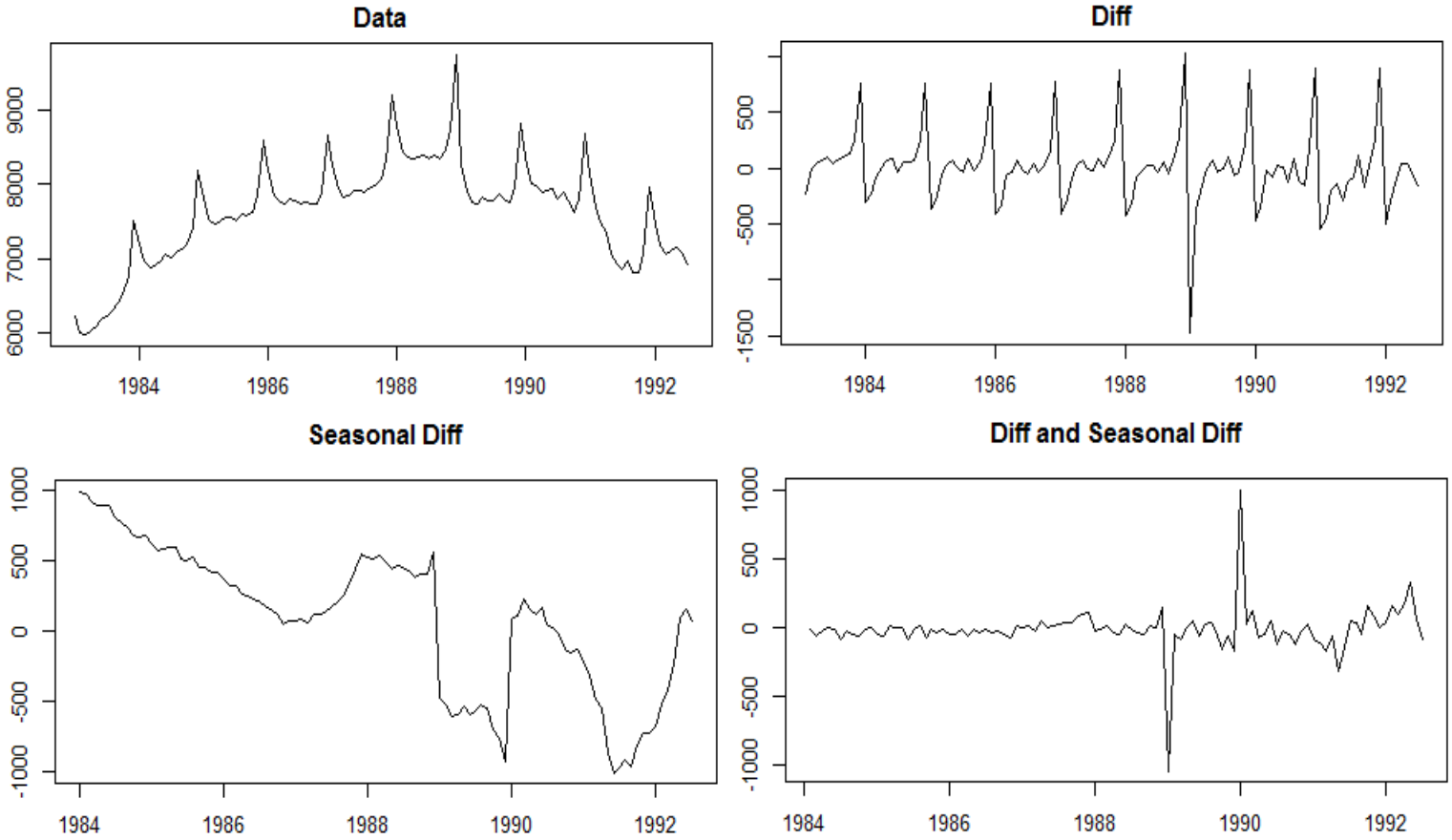
Σε περίπτωση έντονης εποχιακότητας εφαρμόζεται εποχιακή διαφόριση. Εδώ η χρονοσειρά που παράγεται είναι αποτέλεσμα της διαφόρισης μεταξύ των παρατηρήσεων της αρχικής χρονοσειράς και προηγούμενων αντίστοιχων εποχιακότητας περιόδων. Στη πράξη χρησιμοποιείται μέχρι *1ης* τάξης εποχιακή διαφόριση.

- $y'_t=y_t-y_{t-m}$, όπου m η περίοδος εποχιακότητας κ.ο.κ.

Αν ορίσουμε ως B τον **τελεστή ολίσθησης** ούτως ώστε $By_t=y_{t-1}$ και $B(By_t)=B^2y_t=y_{t-2}$, τότε μπορούμε να αναπαραστήσουμε τη διαφόριση n τάξης ως $(1-B)^ny_t$ και την εποχιακή διαφόριση m τάξης N ως $(1-B^m)^Ny_t$.

FORECASTING & STRATEGY UNIT

Διαφόριση (2/2)



Εδώ η αρχική χρονοσειρά (*Data*) έχει εμφανή μηνιαία εποχιακότητα. Αν εφαρμοστεί διαφόριση 1ης τάξης (*Diff*) η εποχιακότητα παραμένει χωρίς να παρέχεται στασιμότητα, ενώ αν εφαρμοστεί εποχιακή διαφόριση (*Seasonal Diff*) η τάση διατηρείται. Έτσι, για την απόκτηση επαρκούς στασιμότητας συνιστάτε η εφαρμογή και των δύο τύπων διαφόρισης.

FORECASTING & STRATEGY UNIT

Αυτοσυσχέτιση *ACF* και μερική αυτοσυσχέτιση *PACF* (1/2)

- ACF: Ένδειξη αν η τιμή της χρονοσειράς σε μία χρονική περίοδο εξαρτάται από την τιμή της παρατήρησης k περιόδων πίσω.

$$p_k = \frac{\sum_{t=k+1}^{n}(Y_t - \mu)(Y_{t-k} - \mu)}{\sum_{t=1}^{n}(\mathrm{Y_t} - \mu)^2}$$

- PACF: Ένδειξη αν η τιμή της χρονοσειράς σε μία χρονική περίοδο εξαρτάται από την τιμή της παρατήρησης k περιόδων πίσω, μη λαμβάνοντας υπόψη την επίδραση που μπορούν ενδεχομένως να επιφέρουν οι ενδιάμεσες παρατηρήσεις.

$$\varphi_{11} = \rho_1 \ , \varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} r_j} \ \gamma\iota\alpha \ k = 3 \ldots$$

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-j} \ \gamma\iota\alpha \ k = 2 \ldots \ j = 1,2, \ldots k-1$$

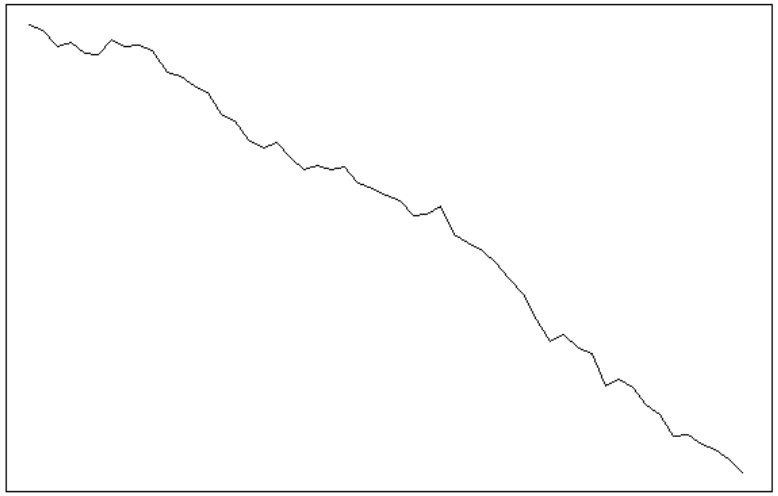
, όπου μ η μέση τιμή των παρατηρήσεων και ρ_k η εκτιμούμενη τιμή αυτοσυσχέτισης.

FORECASTING & STRATEGY UNIT

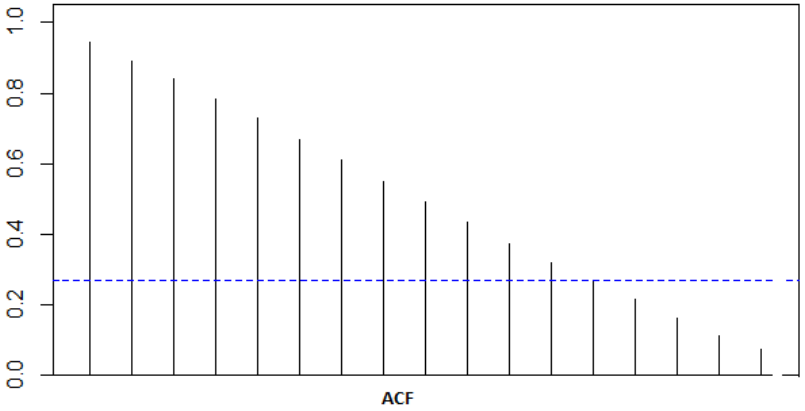
Αυτοσυσχέτιση *ACF* και μερική αυτοσυσχέτιση *PACF* (2/2)

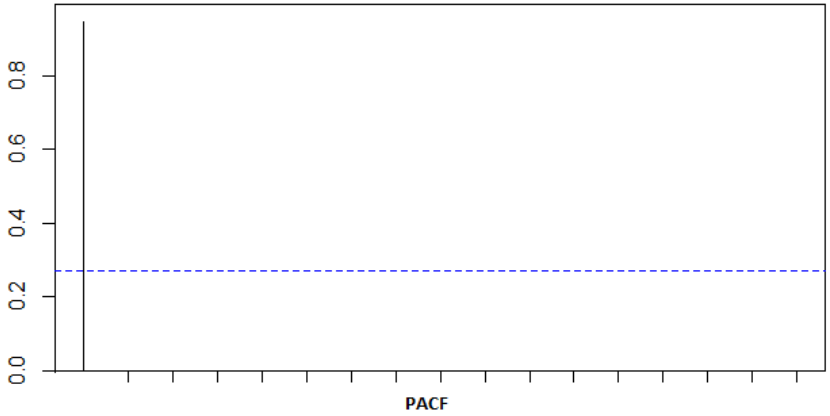
- Η συσχέτιση και η μερική αυτοσυσχέτιση αποτελούν ένδειξη για τις σχέσεις με τις οποίες συνδέονται οι παρατηρήσεις
- Συνήθως ο έλεγχος των συσχετίσεων πραγματοποιείται για υστερήσεις μικρότερες ή ίσες του 3 καθώς σε αντίθετη περίπτωση οδηγούμαστε σε ιδιαίτερα πολύπλοκα μοντέλα χωρίς όφελος σε προβλεπτική ακρίβεια
- Ο έλεγχος θα πρέπει να γίνεται αφού έχει εξασφαλιστεί η στασιμότητα της χρονοσειράς αλλιώς θα οδηγηθούμε πιθανώς σε εσφαλμένα συμπεράσματα
- Οι συσχετίσεις αποτελούν βέβαια από μόνες τους σημαντικό εργαλείο για τον έλεγχο της στασιμότητας μίας χρονοσειράς. Ιδανικά βέβαια αυτή ελέγχεται μέσα από στατιστικά τεστ (π.χ. KPSS test)

Διαγράμματα ACF – PACF (1/3)



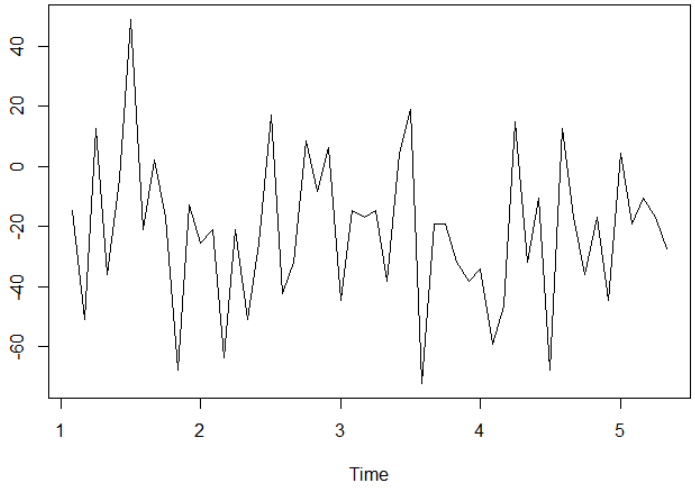
Μεγάλη συσχέτιση μεταξύ διαδοχικών παρατηρήσεων λόγω τάσης (ACF). Η επόμενη τιμή εξαρτάται αποκλειστικά από την προηγούμενη (PACF).



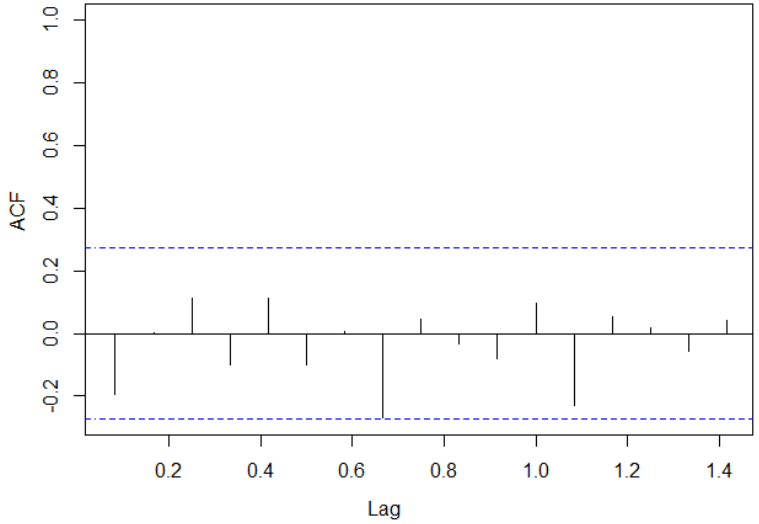


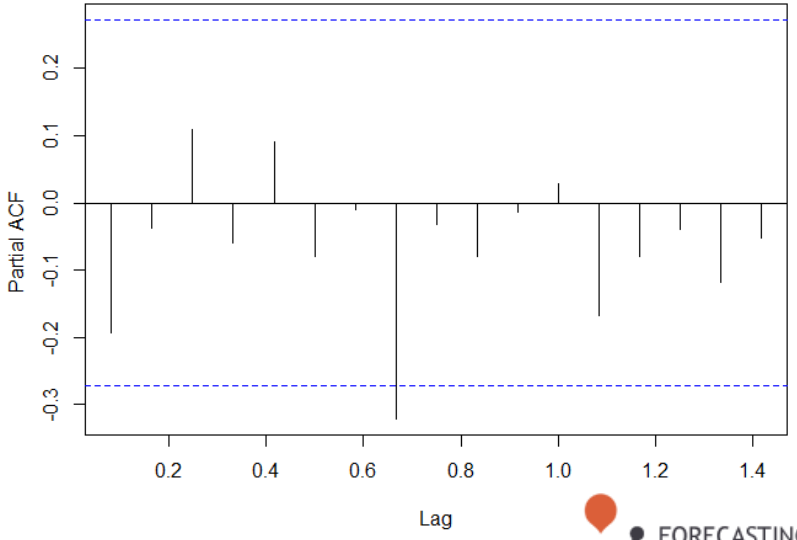
FORECASTING & STRATEGY UNIT

Διαγράμματα ACF – PACF (2/3)



Έχοντας διαφορίσει την προηγούμενη χρονοσειρά η συσχέτιση μεταξύ των παρατηρήσεων έχει εξαλειφθεί πλήρως.

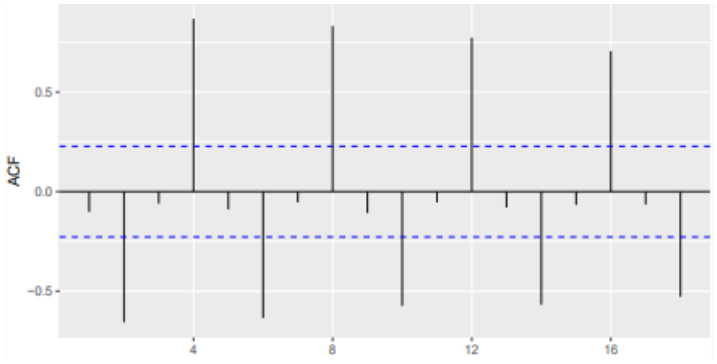


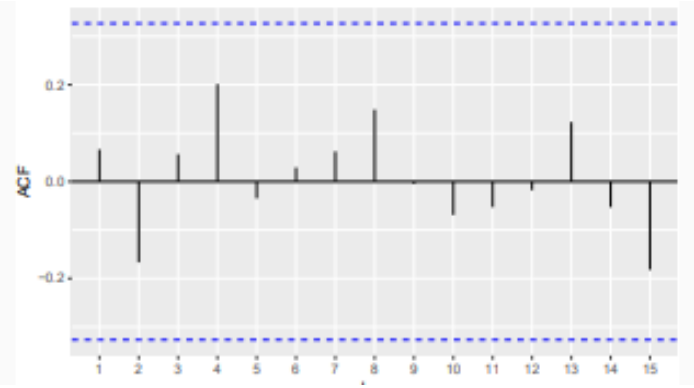


FORECASTING & STRATEGY UNIT

Διαγράμματα ACF – PACF (3/3)

- Για διάστημα εμπιστοσύνης 95%, οι αυτοσυσχετίσεις ενός λευκού θορύβου *N(0,1/n)* αναμένεται να λαμβάνουν τιμές στο διάστημα $[-1.96\sqrt{n}, 1.96\sqrt{n}]$.
- Αν η χρονοσειρά που μελετάται εμφανίζει παρόμοια συμπεριφορά, τότε δεν είναι στατιστικά διάφορη του λευκού θορύβου και συνεπώς είναι στάσιμη. Σε αντίθετη περίπτωση, είναι στατιστικά διάφορη και απαιτούνται αντίστοιχες ενέργειες.





FORECASTING & STRATEGY UNIT

Διαφόριση μέσω χρήσης τεστ υποθέσεων

- Τα τεστ κάνουν την αρχική υπόθεση (H0) ότι η χρονοσειρά είναι (ή δεν είναι) στάσιμη
- Εφαρμόζουν μία σειρά από ελέγχους και κριτήρια
- Ανάλογα με τα αποτελέσματα (H1) απορρίπτουν ή δέχονται την αρχική υπόθεση προτείνοντας αντίστοιχες ενέργειες

Augmented Dickey-Fuller (ADF): Υποθέτει ότι η χρονοσειρά δεν είναι στάσιμη. Υψηλές πιθανότητες (p.value>5%) σηματοδοτούν την έλλειψη στασιμότητας και μικρές την ύπαρξή της, ή καλύτερα τη μη ύπαρξη ενδείξεων για να υποστηριχθεί το αντίθετο.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS): Υποθέτει ότι η χρονοσειρά είναι στάσιμη.

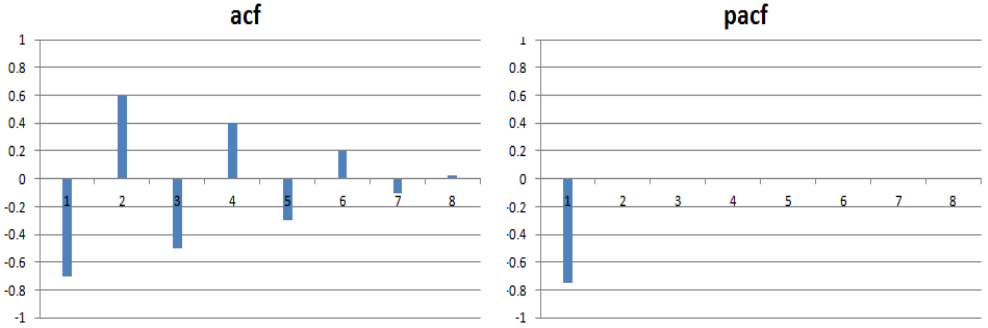
Canova-Hansen: Υποθέτει ότι η χρονοσειρά είναι στάσιμη (για έλεγχο εποχιακότητας).

FORECASTING & STRATEGY UNIT

Αναγνώριση καταλληλότητας μοντέλων ARMA (1/3)

Για ένα στάσιμο μοντέλο AR(p):

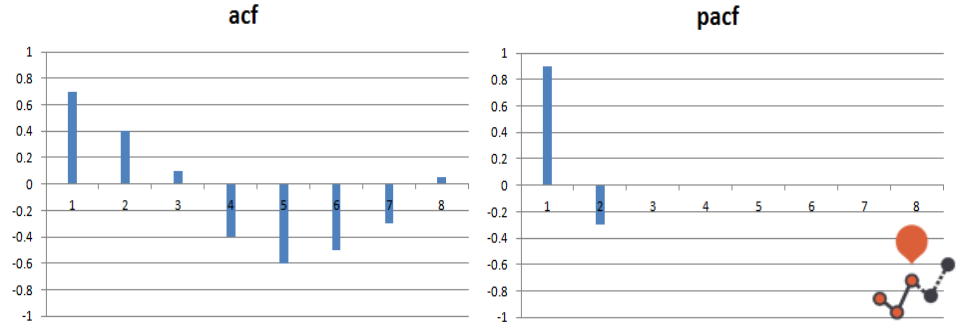
- Οι τιμές των συντελεστών ACF φθίνουν προς το μηδέν ακολουθώντας εκθετική η ημιτονοειδή πορεία
- Οι τιμές των συντελεστών PACF μηδενίζονται απότομα μετά από p περιόδους υστέρησης



$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p}$$

AR(1)

AR(2)

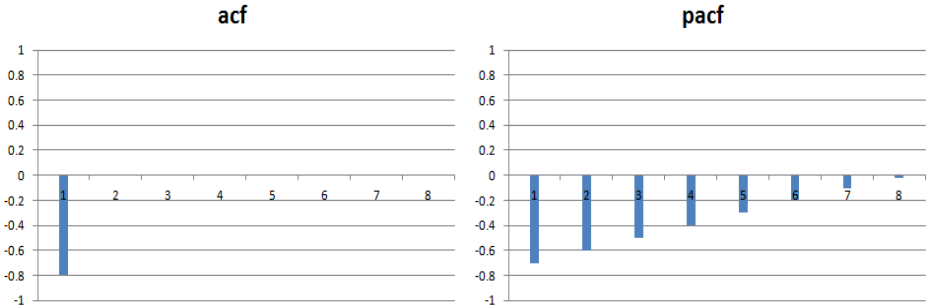


FORECASTING & STRATEGY UNIT

Αναγνώριση καταλληλότητας μοντέλων ARMA (2/3)

Για ένα στάσιμο μοντέλο MA(q):

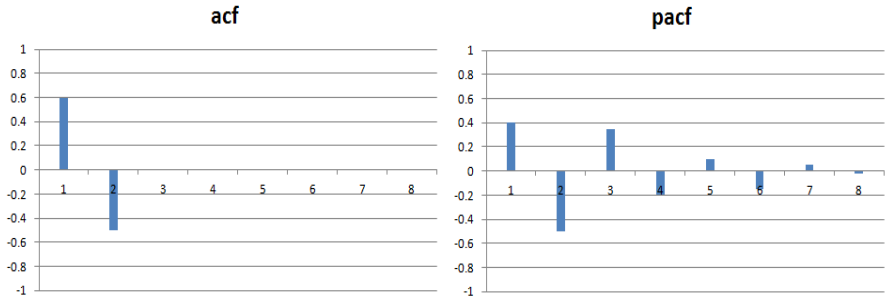
- Οι τιμές των συντελεστών ACF μηδενίζονται απότομα μετά από q περιόδους υστέρησης
- Οι τιμές των συντελεστών PACF φθίνουν προς το μηδέν ακολουθώντας εκθετική η ημιτονοειδή πορεία



$$y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

MA(1)

MA(2)

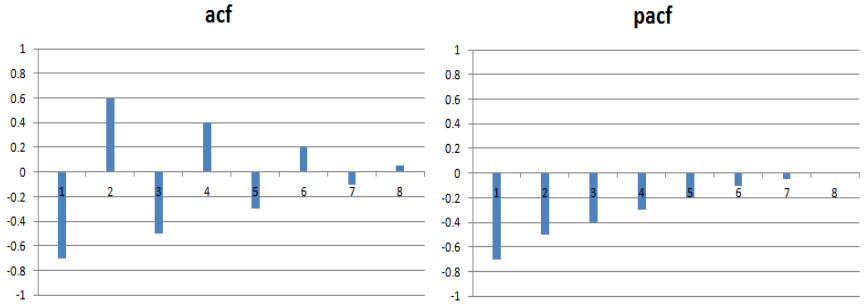


FORECASTING & STRATEGY UNIT

Αναγνώριση καταλληλότητας μοντέλων ARMA (3/3)

Για ένα στάσιμο μοντέλο ARMA(p,q):

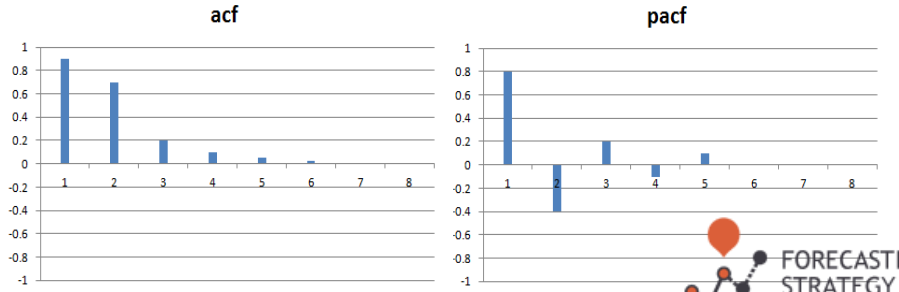
- Οι τιμές των συντελεστών ACF φθίνουν προς το μηδέν μετά από q-p περιόδους υστέρησης
- Οι τιμές των συντελεστών PACF φθίνουν προς το μηδέν μετά από p-q περιόδους υστέρησης



$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

ARMA(1,1)

ARMA(1,1)



FORECASTING & STRATEGY UNIT

Κριτήρια επιλογής μοντέλων (1/2)

Παρέχουν πληροφορία σχετικά με την ποιότητα προσαρμογής του μοντέλου αξιολογώντας την πιθανότητα ταύτισης των παραγόμενων προβλέψεων με τις πραγματικές και συσχετίζοντάς την με την πολυπλοκότητά του.

- **Προσδοκώμενη πιθανοφάνεια (Likelihood):**

$$-2logL = n\log(2\pi) + n\log(\sigma^2) + \frac{\sum_{t=1}^{n} {e_t}^2}{\sigma^2}$$

- **Akaike's Information Criterion (AIC):**

$$AIC = -2logL + 2(p + q + P + Q + k + 1)$$

$$AIC_c = AIC + \frac{2(p + q + P + Q + k + 1)(p + q + P + Q + k + 2)}{n - p - q - P - Q - k - 2}$$

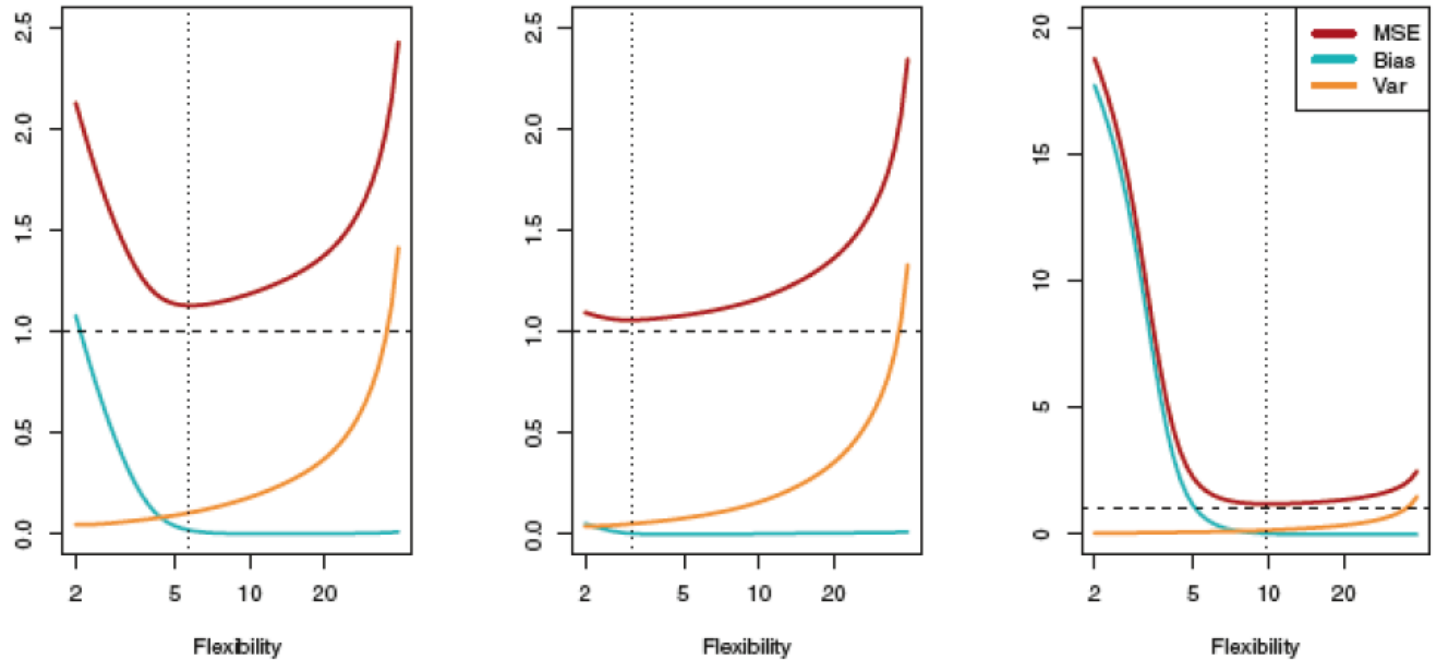
- **Bayesian Information Criterion (BIC):**

$$BIC = AIC + (log(n) - 2)(p + q + P + Q + k - 1)$$

,όπου n ο αριθμός των παρατηρήσεων και p, q, P, Q οι συντελεστές του μοντέλου ARMA. Ο συντελεστή k ισούται με 1 ή 0 ανάλογα με το αν το μοντέλο έχει ή όχι σταθερά.

FORECASTING & STRATEGY UNIT

Κριτήρια επιλογής μοντέλων (2/2)



Bias-variance trade-off & Over-fitting

- Όσο αυξάνεται η πολυπλοκότητά (*p*, *q*, *P*, *Q*, *c*) τόσο μειώνεται η προκατάληψη των προβλέψεων (*bias*).
- Η αύξηση των μεταβλητών (*flexibility-complexity*) αυξάνει τη διακύμανσης των παραγόμενων σφαλμάτων και συνεπώς μειώνει την ακρίβεια πρόβλεψης (*variance*).

FORECASTING & STRATEGY UNIT

Στατιστικός διαγνωστικός έλεγχος μοντέλων

Μπορούμε ωστόσο να εξακριβώσουμε αν οι συσχετίσεις ΄των υπολειπόμενων σφαλμάτων είναι στατιστικά σημαντική, αν δηλαδή τα σφάλματα ακολουθούν ένα συγκεκριμένο μοτίβο. Σε μία τέτοια περίπτωση το μοντέλο κρίνεται ακατάλληλο αφού οδηγεί σταθερά σε μεροληπτικές και ανακριβείς προβλέψεις.

$$t_{r_k} = \frac{r_k(e)}{S(r_{k(e)})}$$

$$S(r_k(e)) = n^{-1/2}(1 + 2\sum_{j=1}^{k-1} {r_{j(e)}}^2)^{1/2}$$

Για διάστημα εμπιστοσύνης 95%, ο δείκτης *t* πρέπει να είναι μικρότερος του 1.96 ώστε μία συσχέτιση να μην είναι σημαντική. Στην πράξη, για υστέρηση 1,2 και 3 πρέπει να είναι μικρότερη του 1.25 και για μεγαλύτερη υστέρηση μικρότερη του 1.6.

- *r_k, ο συντελεστής ACF για υστέρηση k*
- *e, το σφάλμα πρόβλεψης του μοντέλου*
- *n, ο αριθμός των παρατηρήσεων στο δείγμα*

Confidence level	z
99%	2.576
95%	1.96
90%	1.645
80%	1.282
70%	1.036

FORECASTING & STRATEGY UNIT

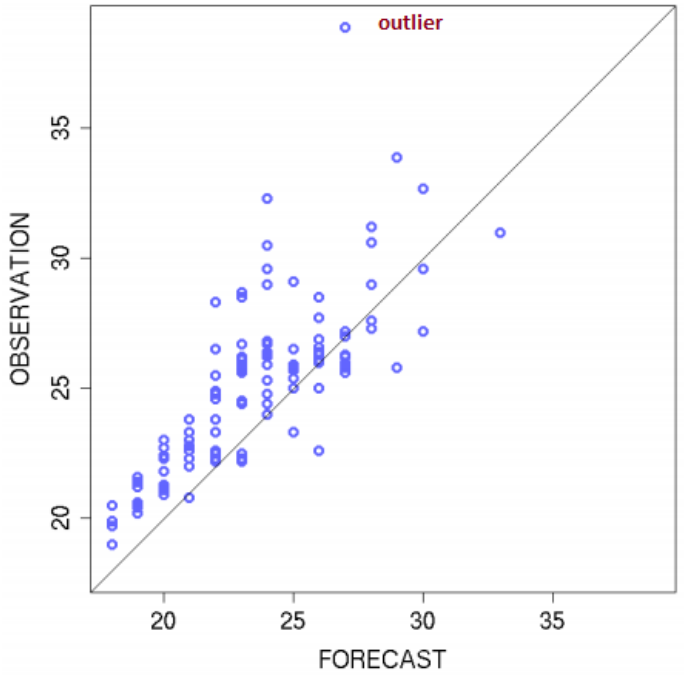
Εποπτικός διαγνωστικός έλεγχος μοντέλων (1/3)

Scatter plot

Παρουσιάζει τις πραγματικές τιμές των δεδομένων έναντι των προβλεπόμενων.

Τέλεια πρόβλεψη: Όλα τα σημεία βρίσκονται πάνω στη διαγώνιο 45ο

Πληροφορεί σχετικά με: γραμμική συσχέτιση, outliers, προκατάληψη

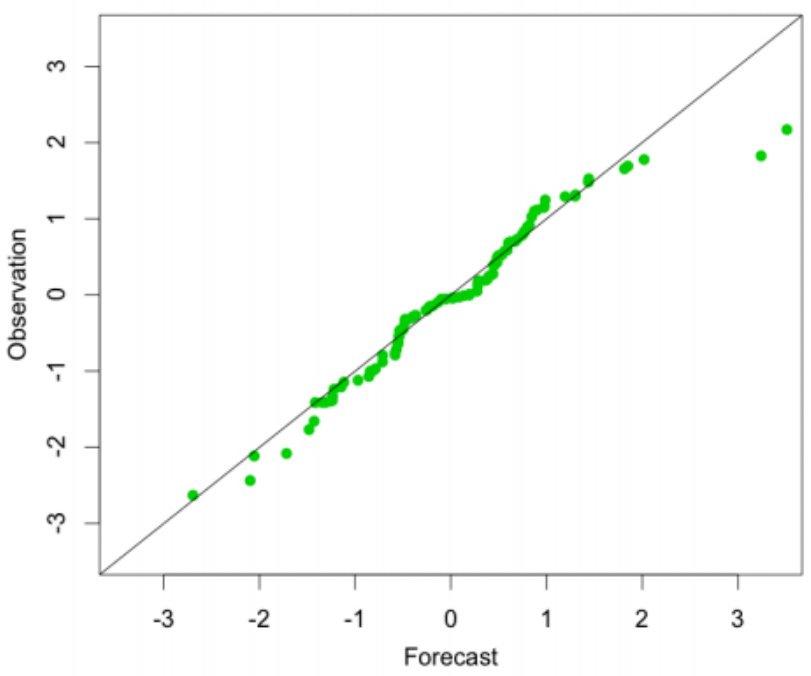


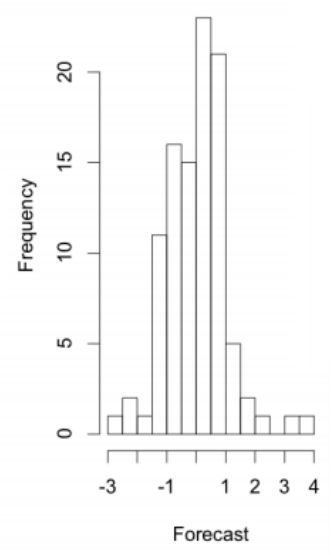
FORECASTING & STRATEGY UNIT

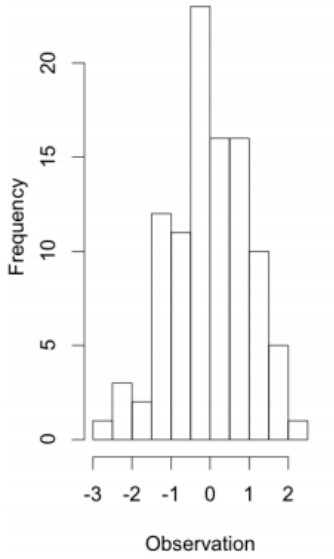
Εποπτικός διαγνωστικός έλεγχος μοντέλων (2/3)

Quantile-Quantile plot

Παρουσιάζει τις πραγματικές τιμές των δεδομένων έναντι των προβλεπόμενων, βάση της πιθανότητας εμφάνισής τους.







FORECASTING & STRATEGY UNIT

Εποπτικός διαγνωστικός έλεγχος μοντέλων (3/3)

Box-Plot

Απεικονίζει την κατανομή των σφαλμάτων πληροφορώντας σχετικά με *outliers, προκατάληψη και κανονικότητα*.

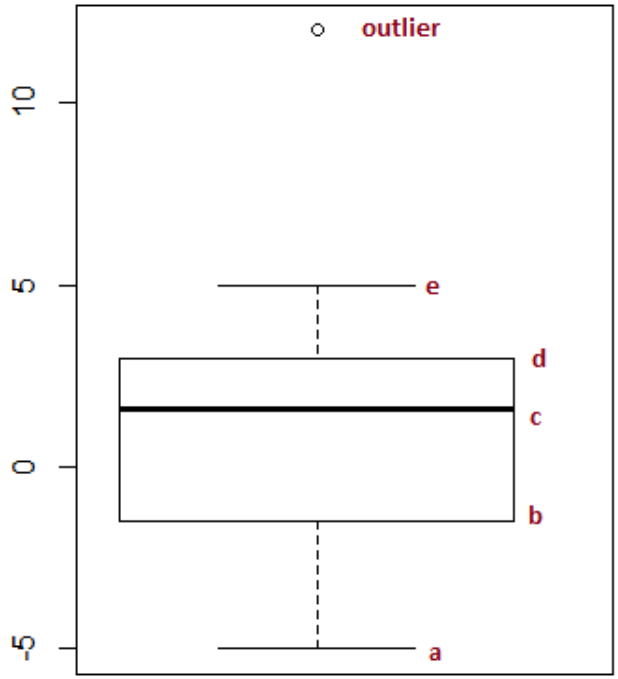
a: min

b: 25% των παρατηρήσεων

c: median

d: 75% των παρατηρήσεων

e: max



Για να έχω πραγματικά αμερόληπτες προβλέψεις πρέπει οι αποστάσεις a-b, b-c, c-d και d-e να είναι ίσες και ο ενδιάμεσος να ισούται με μηδέν.

Στο παράδειγμα φανερώνεται η απαισιοδοξία του μοντέλου καθώς c>0, δηλαδή το μεγαλύτερο μέρος των σφαλμάτων συγκεντρώνεται σε θετικότερες τιμές.

FORECASTING & STRATEGY UNIT

Μοντέλα Αυτοπαλινδρόμησης
Autoregressive models- AR(p)

Θεωρούν γραμμικές σχέσεις ανάμεσα στην παρατήρηση της χρονοσειράς που εξετάζεται και στις προηγούμενες τιμές αυτής.

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p}$$

ή αν $\overline{y_t} = y_t - \mu$

$$\left(1 - \boldsymbol{\varphi_1 B} - \boldsymbol{\varphi_2 B^2} - \cdots - \boldsymbol{\varphi_p B^p}\right)\overline{\boldsymbol{y}}_{\boldsymbol{t}} = 0$$

Η νέα μεταβλητή έχει τις ίδιες ιδιότητες με την αρχική *yt* αλλά με μηδενική μέση τιμή. Εξισώνοντας έχουμε:

$$c = \mu\left(1 - \varphi_1 - \varphi_2 - \cdots \varphi_p\right)$$

Ένα μοντέλο **AR(1)**

- Ταυτίζεται με τον λευκό θόρυβο όταν $\varphi_1=0$
- Ταυτίζεται με τον τυχαίο περίπατο όταν $\varphi_1=1$ και $c=0$
- Ταυτίζεται με τον τυχαίο περίπατο με τάση όταν $\varphi_1=1$ και c διάφορο του μηδενός
- Ταλαντώνεται μεταξύ αρνητικών και θετικών τιμών όταν $\varphi_1<0$

FORECASTING & STRATEGY UNIT

Μοντέλα Κινητού Μέσου Όρου Moving Average MA(q)

Θεωρούν γραμμικές σχέσεις ανάμεσα στην παρατήρηση της χρονοσειράς που εξετάζεται και στα σφάλματα που εμφάνισε το μοντέλο σε προηγούμενες περιόδους

$$y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

ή

$$\overline{\boldsymbol{y}}_{\boldsymbol{t}} = \left(\boldsymbol{\theta}_1 \boldsymbol{B} + \boldsymbol{\theta}_2 \boldsymbol{B}^2 - \cdots + \boldsymbol{\theta}_{\boldsymbol{q}} \boldsymbol{B}^{\boldsymbol{q}}\right) \boldsymbol{e}_{\boldsymbol{t}}$$

Εξισώνοντας έχουμε ότι:

$$c = \mu$$

FORECASTING & STRATEGY UNIT

Μοντέλα ARIMA (p,d,q)

Αποτελούν γραμμικό συνδυασμό των παραπάνω μοντέλων (AR – MA και Διαφόρισης)

$$\left(1 - \boldsymbol{\varphi_1 B} - \cdots - \boldsymbol{\varphi_p B^p}\right)(1 - \boldsymbol{B})^{\boldsymbol{n}}(1 - \boldsymbol{B^m})^{\boldsymbol{N}}\boldsymbol{y_t} \;= \boldsymbol{c} + \left(\boldsymbol{\theta_1 B} + \cdots + \boldsymbol{\theta_q B^q}\right)\boldsymbol{e_t}$$

Κατά τα γνωστά προκύπτει ότι: $c = \mu(1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p)$, για n=N=0

και $c = 0$ σε κάθε άλλη περίπτωση

Ένας πρώτος και απλός τρόπος ελέγχου της **καταλληλότητας** του μοντέλου είναι η ικανοποίηση των παρακάτω συνθηκών:

- ✓ AR(1): $-1<\varphi_1<1$
- ✓ AR(2): $-1<\varphi_2<1$ και $\varphi_1 + \varphi_2 <1$ και $\varphi_2 - \varphi_1 <1$
- ✓ MA(1): $-1<\vartheta_1<1$
- ✓ MA(2): $-1<\vartheta_2<1$ και $\theta_1 + \vartheta_2 >-1$ και $\vartheta_1 - \vartheta_2 <1$

FORECASTING & STRATEGY UNIT

Πότε χρησιμοποιώ σταθερά;

- Αν c=0 και d=0, μακροπρόθεσμα οι προβλέψεις θα ισούνται με μηδέν
- Αν c=0 και d=1, μακροπρόθεσμα οι προβλέψεις θα ισούνται με μία μη μηδενική σταθερά
- Αν c=0 και d=2, μακροπρόθεσμα οι προβλέψεις θα ακολουθούν μία ευθεία γραμμή
- Αν c≠0 και d=0, μακροπρόθεσμα οι προβλέψεις θα ισούνται με τη μέση τιμή της χρονοσειράς
- Αν c≠0 και d=1, μακροπρόθεσμα οι προβλέψεις θα ακολουθούν μία ευθεία γραμμή
- Αν c≠0 και d=2, μακροπρόθεσμα οι προβλέψεις θα ακολουθούν μία εκθετική καμπύλη

FORECASTING & STRATEGY UNIT

Υπολογισμός συντελεστών σε μοντέλα ARIMA

ARMA(p, q)	p_1	p_2
AR(1)	φ_1	
MA(1)	$\frac{-\theta_1}{1+{\theta_1}^2}$	
AR(2)	$\frac{\varphi_1}{1-\varphi_2}$	$\frac{{\varphi_1}^2}{1-\varphi_2}+\varphi_2$
MA(2)	$\frac{-\theta_1(1-\theta_2)}{1+{\theta_1}^2+{\theta_2}^2}$	$\frac{-\theta_2}{1+{\theta_1}^2+{\theta_2}^2}$
ARMA(1,1)	$\frac{(1-\varphi_1\theta_1)(\varphi_1-\theta_1)}{1+{\theta_1}^2-2\varphi_1\theta_1}$	$\rho_1\varphi_1$

Γενικά αποδεικνύεται ότι για αμιγώς MA(q) διαδικασίες ισχύει για τη συνάρτηση αυτοσυσχέτισης:

$\rho_k = \frac{-\theta_k+\theta_{k+1}\theta_1+\cdots+\theta_{q-k}\theta_q}{1+{\theta_1}^2+{\theta_2}^2+\cdots+{\theta_q}^q}$ για k=1,2,..q και $\rho_k = 0$ για k>q.

Αντίστοιχα για αμιγώς AR(p) διαδικασίες ισχύει για τη συνάρτηση αυτοσυσχέτισης:

$\rho_k = {\rho_1}^k$ για k=1,2,..p

FORECASTING & STRATEGY UNIT

Πρόβλεψη με μοντέλα ARIMA

Όταν θέλουμε να προβλέψουμε μέσω ενός μοντέλου ARIMA την τιμή της χρονοσειράς y την περίοδο t, τότε απαιτείται γνώση των τιμών $y_{t-1}, y_{t-2} \ldots y_{t-p}$ *ή/και των τιμών* $e_{t-1}, e_{t-2} \ldots e_{t-q}$.

Για μεγάλους ορίζοντες πρόβλεψης οι παραπάνω τιμές ενδέχεται να μην είναι διαθέσιμες. Όταν συμβαίνει αυτό χρησιμοποιούμε τις αντίστοιχες τιμές πρόβλεψης του μοντέλου. Με αυτήν τη λογική υπάρχει πάντα τιμή y_{t-p} διαθέσιμη, όχι όμως και e_{t-q} αφού δε μπορούμε να γνωρίζουμε το μελλοντικό σφάλμα.

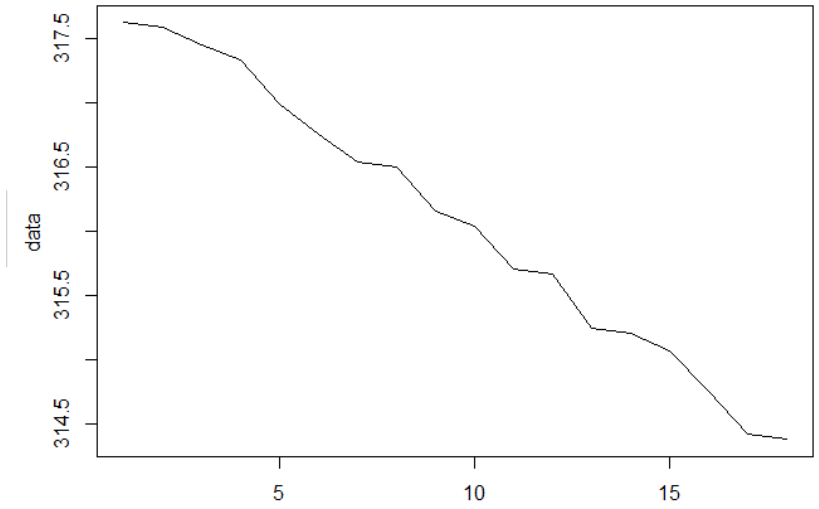
Μακροχρόνια το μοντέλο ARIMA εκφυλίζεται σε μοντέλο AR εξαρτώμενο μόνο από τις προβλέψεις που έχει κάνει το ίδιο και όχι από τις τιμές των δεδομένων. Αυτός είναι και ο λόγος που χρησιμοποιείται κυρίως για βραχυπρόθεσμες προβλέψεις.

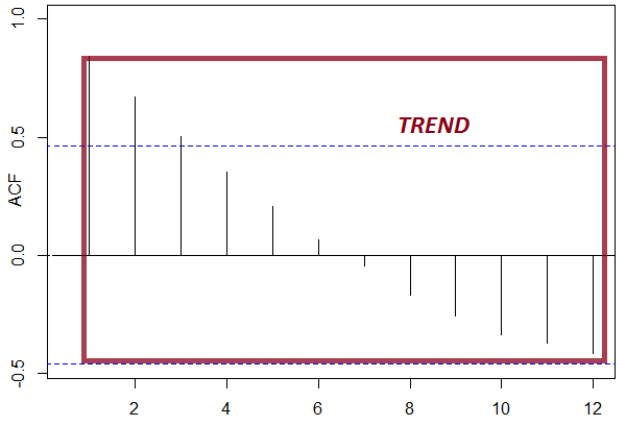
FORECASTING & STRATEGY UNIT

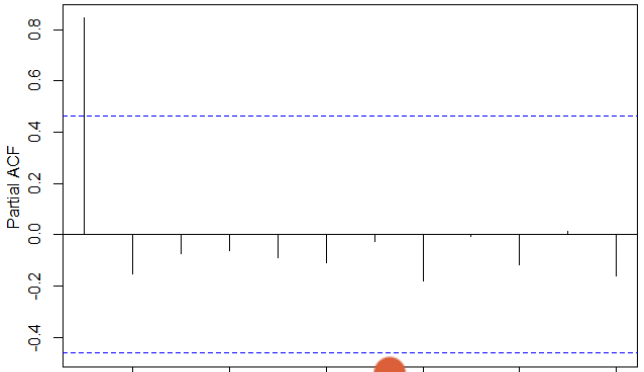
Εφαρμογή

Δίνεται η παρακάτω χρονοσειρά. Ζητείται πρόβλεψη για ορίζοντα τρία.

#	Data
1	317.62
2	317.58
3	317.44
4	317.33
5	316.99
6	316.75
7	316.53
8	316.49
9	316.16
10	316.04
11	315.70
12	315.66
13	315.24
14	315.21
15	315.07
16	314.75
17	314.41
18	314.38



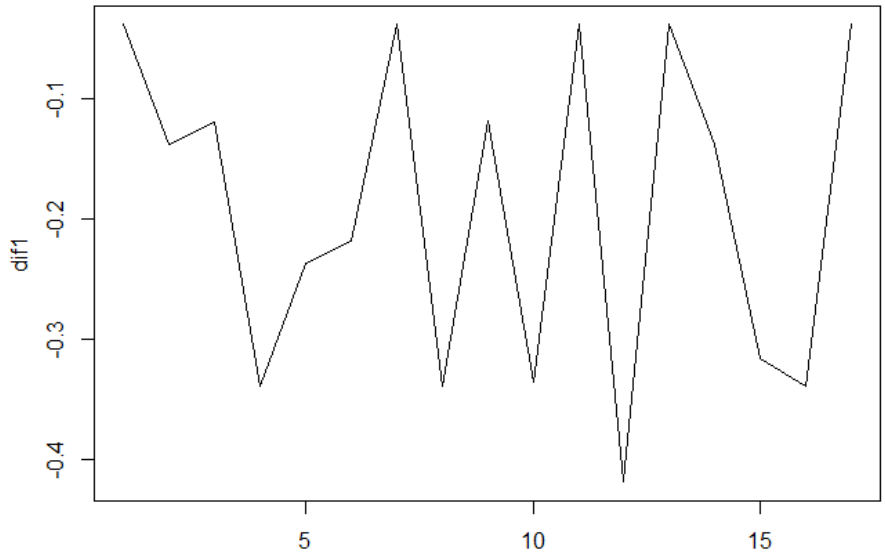


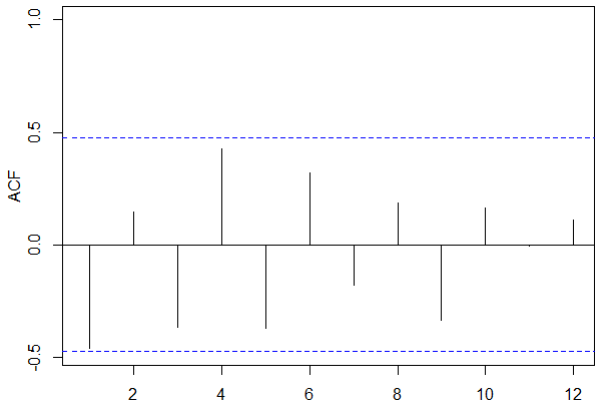


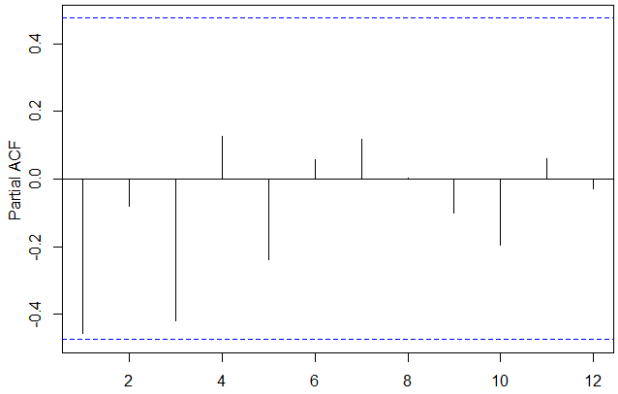
FORECASTING & STRATEGY UNIT

Η χρονοσειρά έχει εμφανή τάση. Πρέπει επομένως να προχωρήσουμε σε διαφόριση μέχρις ότου φτάσουμε σε επαρκή στασιμότητα.

#	Δεδομένα	Πρώτες διαφορές
1	317.62	
2	317.58	-0.04
3	317.44	-0.14
4	317.33	-0.11
5	316.99	-0.34
6	316.75	-0.24
7	316.53	-0.22
8	316.49	-0.04
9	316.16	-0.33
10	316.04	-0.12
11	315.7	-0.34
12	315.66	-0.04
13	315.24	-0.42
14	315.21	-0.03
15	315.07	-0.14
16	314.75	-0.32
17	314.41	-0.34
18	314.38	-0.03







Η διαφόριση 1ης τάξης επαρκεί αφού οδηγεί σε στασιμότητα (Μ.Τ γύρω από το -0.19).

FORECASTING & STRATEGY UNIT

Έχοντας εξασφαλίσει στασιμότητα επιλέγουμε μοντέλο πρόβλεψης βάση των διαγραμμάτων ACF & PACF. Στα πρώτα διαγράμματα παρατηρείται φθίνουσα πορεία των συντελεστών ACF και μεγάλη τιμή PACF για υστέρηση k=1. Επιλέγουμε λοιπόν το μοντέλο AR(1).

#	Δεδομένα	Πρώτες διαφορές	yt-μ	Αριθμητής p1	Παρονομαστής p1
1	317.62	-	-	-	-
2	317.58	-0.04	0.151	-	0.023
3	317.44	-0.14	0.051	0.008	0.003
4	317.33	-0.11	0.081	0.004	0.006
5	316.99	-0.34	-0.149	-0.012	0.022
6	316.75	-0.24	-0.049	0.007	0.002
7	316.53	-0.22	-0.029	0.001	0.001
8	316.49	-0.04	0.151	-0.004	0.023
9	316.16	-0.33	-0.139	-0.021	0.019
10	316.04	-0.12	0.071	-0.010	0.005
11	315.7	-0.34	-0.149	-0.011	0.022
12	315.66	-0.04	0.151	-0.022	0.023
13	315.24	-0.42	-0.229	-0.035	0.053
14	315.21	-0.03	0.161	-0.037	0.026
15	315.07	-0.14	0.051	0.008	0.003
16	314.75	-0.32	-0.129	-0.007	0.017
17	314.41	-0.34	-0.149	0.019	0.022
18	314.38	-0.03	0.161	-0.024	0.026
Μ.Ο.	***316.075***	***-0.191***		***-0.134***	***0.295***

Υπολογίζουμε το συντελεστή αυτοσυσχέτισης για υστέρηση 1, σύμφωνα με το γνωστό τύπο.

Αυτός προκύπτει ίσος με:
$\rho_1 = -0.134/0.295 = -0.455$

FORECASTING & STRATEGY UNIT

Συνεπώς για το μοντέλο ARIMA (1,1,0) ισχύει:

$$(1 - \varphi_1 B) y_t (1 - \mathrm{B}) = \mathrm{c}$$

$$y_t = \varphi_1 y_{t-1} + y_{t-1} - \varphi_1 y_{t-2} + \mathrm{c}$$

$$y_t = (1 + \varphi_1) y_{t-1} - \varphi_1 y_{t-2} + \mathrm{c}$$

$$y_t = 0.545 y_{t-1} + 0.455 y_{t-2} + \mathrm{c}$$

Εδώ c = μ(1-ϕ1) = -0.277

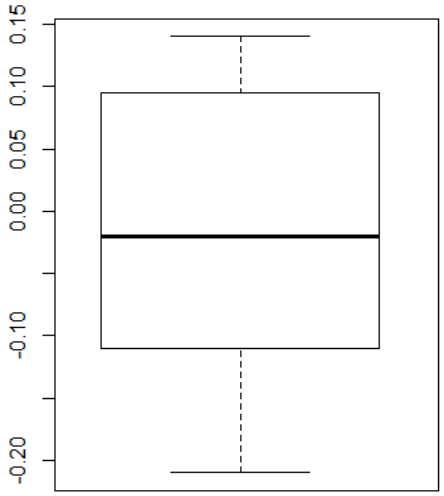
Επισημαίνεται ότι σαν *μ* χρησιμοποιείται η μέση τιμή της 1ης διαφοράς της χρονοσειράς, αφού αυτή είναι στάσιμη. Τη σταθερά τη χρησιμοποιούμε ούτως ώστε να προδώσουμε την απαραίτητη τάση.

Αν c≠0 και d=1, μακροπρόθεσμα οι προβλέψεις θα ακολουθούν μία ευθεία γραμμή
Αν c=0 και d=1, μακροπρόθεσμα οι προβλέψεις θα ισούνται με μία μη μηδενική σταθερά

FORECASTING & STRATEGY UNIT

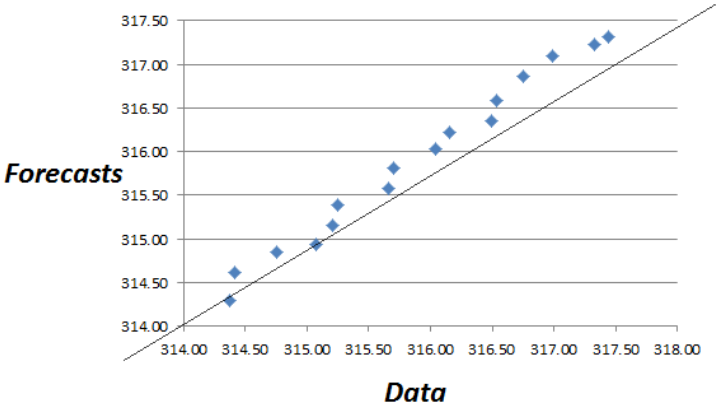
#	Δεδομένα	ARIMA(1,1,0)	et	RS
1	317.62	-	-	-
2	317.58	-	-	-
3	317.44	317.32	0.12	0.014
4	317.33	317.23	0.10	0.011
5	316.99	317.10	-0.11	0.013
6	316.75	316.87	-0.12	0.014
7	316.53	316.58	-0.05	0.003
8	316.49	316.35	0.14	0.019
9	316.16	316.23	-0.07	0.005
10	316.04	316.03	0.01	0.000
11	315.70	315.82	-0.12	0.014
12	315.66	315.58	0.08	0.007
13	315.24	315.40	-0.16	0.026
14	315.21	315.15	0.06	0.003
15	315.07	314.95	0.12	0.015
16	314.75	314.86	-0.11	0.011
17	314.41	314.62	-0.21	0.043
18	314.38	314.29	0.09	0.009
f1	-	314.12	-	-
f2	-	313.96	-	-
f3	-	313.75	-	-

- Για τις προβλέψεις F2 και F3 για τις οποίες στερούμαστε δεδομένα θεωρούμε ως yt-1, yt-2 τις αντίστοιχες προβλέψεις του μοντέλου



Το μοντέλο είναι σχετικά **αισιόδοξο**:

- Median Error = -0.02 <0
- Scatter line πάνω από 45°
- ME=-0.01



FORECASTING & STRATEGY UNIT

$$-2logL = n\log(2\pi) + n\log(\sigma^2) + \frac{\sum_{t=1}^{n} {e_t}^2}{\sigma^2}$$

$$= 16 * 1.838 - 16 * 4.368 + 0.219/0.013 = -23.63$$

- *AIC* = -2logL+2(1+0+1+1)=-17.63
- *AICc* = AIC+2(1+0+1+1)*(1+0+1+2)/(16-1-0-1-2)=-15.63
- *BIC*= AIC +(log(16)-2)(1+0+1-1)=-16.86

Αυτοσυσχέτιση σφαλμάτων

Αρκετά ασθενής για τις πρώτες υστερήσεις: Δείγμα καταλληλότητας μοντέλου

Lag (k)	rk
1	-0.196
2	-0.321
3	-0.156
5	0.283

$$t_{r_k} = \frac{r_k(e)}{S(r_{k(e)})}$$

$$S(r_k(e)) = n^{-1/2}(1 + 2\sum_{j=1}^{k-1} {r_{j(e)}}^2)^{1/2}$$

Στατιστική σημαντικότητα μοντέλου

Όλοι οι δείκτες *t* είναι μικρότεροι του 1.25 για υστέρηση 1, 2 και 3 και του 1.6 για υστέρηση 4. Συνεπώς το μοντέλο θεωρείται άρτιο.

lag	rk	sq(rk)	*Sk*	*trk*
1	-0.196	0.039	0.250	-0.786
2	-0.321	0.103	0.259	-1.238
3	-0.156	0.024	0.283	-0.550
4	0.283	0.080	0.289	0.980

FORECASTING & STRATEGY UNIT