



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

Τεχνικές Προβλέψεων

Αυτοπαλινδρομικά Μοντέλα Κινητού
Μέσου Όρου (ARIMA)

<http://www.fsu.gr> - lesson@fsu.gr

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 και 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*) ωστόσο έχουν την ιδιομορφία ότι εκφράζουν κάθε τιμή της χρονοσειράς ως γραμμικό άθροισμα παλαιότερων

Λογική ARIMA vs. LRL vs. ES

LRL: Η τιμή του μεγέθους Y τη χρονική στιγμή $t+1$ θα είναι ανάλογη του X την ίδια χρονική στιγμή

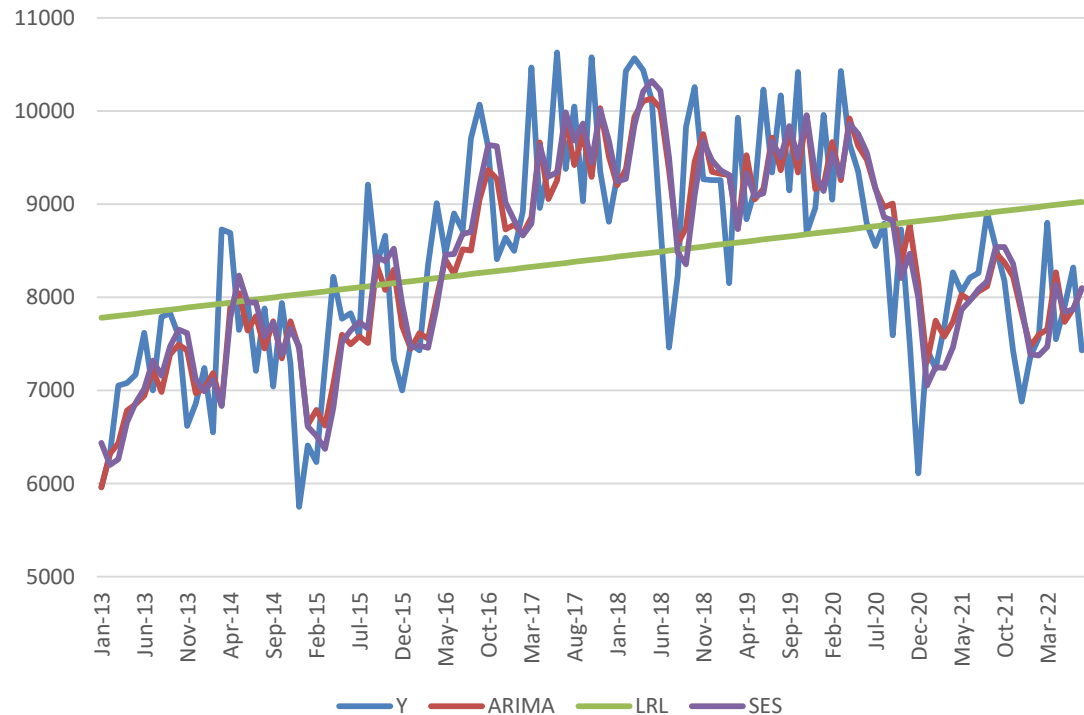
$$\widehat{Y}_{t+1} = 7768 + 11(t + 1)$$

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

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

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

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



Στοχαστικοί Παράγοντες ARMA

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

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

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

Απαιτήσεις – Περιορισμοί (1/2)

Προκειμένου ένα μοντέλο ARMA να χρησιμοποιηθεί αποδοτικά θα πρέπει η χρονοσειρά να είναι:

- **Διακριτή:** Ισαπέχουσες παρατηρήσεις
- **Στάσιμη:** Σταθερή μέση τιμή, διακύμανση και συνάρτηση αυτοσυσχέτισης στο δείγμα
 $y_{t1}, y_{t2} \dots y_{tn}$ ταυτίζεται στατιστικά με $y_{t1+\tau}, y_{t2+\tau} \dots y_{tn+\tau}$
- Να παραχθούν **βραχυπρόθεσμες** προβλέψεις

Απαιτήσεις – Περιορισμοί (2/2)

- **Διακριτότητα**

Η πρόβλεψη εξαρτάται από p παρελθοντικές παρατηρήσεις οι οποίες και οφείλουν να είναι γνωστές και να βρίσκονται σε αντιστοιχία – Συμπλήρωση τιμών

- **Στασιμότητα**

Τα μοντέλα ARMA δεν μπορούν να μοντελοποιήσουν τάση και εξηγήσουν παροδικές διακυμάνσεις – Μετασχηματισμοί και διαφορίση

- **Βραχυπρόθεσμες προβλέψεις**

Οι μακροπρόθεσμες προβλέψεις θα βασίζονται στις βραχυπρόθεσμες και συνεπώς το αναμενόμενο σφάλμα αυξάνει αισθητά

Επιλογή μοντέλου $ARIMA(p,d,q)$

Η πορεία επιλογής ενός ικανοποιητικού μοντέλου ARIMA περιλαμβάνει τρία στάδια:

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

Επεξεργασία χρονοσειράς

Μετασχηματισμοί: Περιορισμός διακύμανσης και απαλοιφή ακραίων τιμών χρονοσειράς

Πότε; Όταν έχω έντονη διακύμανση στη χρονοσειρά ή το σφάλμα πρόβλεψης που προκύπτει είναι αυξημένο παρά τη σωστή επιλογή μοντέλου

Διαφόριση: Περιορισμός των διακυμάνσεων επιπέδου (τάσης και εποχιακότητας)

Πότε; Όταν έχω εμφανή τάση ή εποχιακότητα στη χρονοσειρά.

Τα μοντέλα ARIMA αντιμετωπίζουν αποτελεσματικά από μόνα τους την τάση 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.

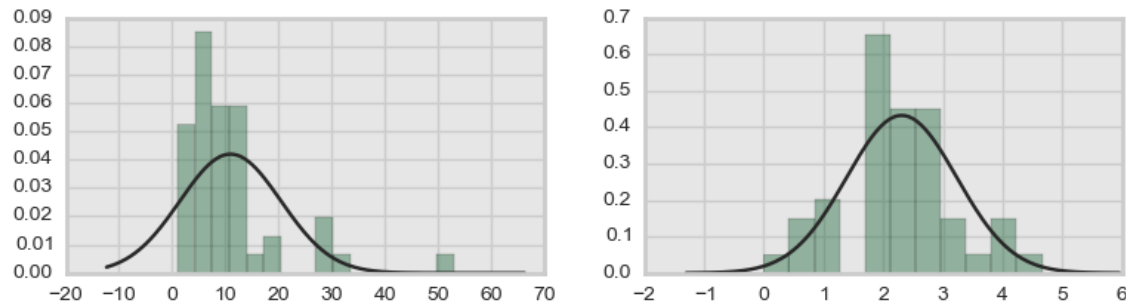
Μετασχηματισμοί (1/2)

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

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

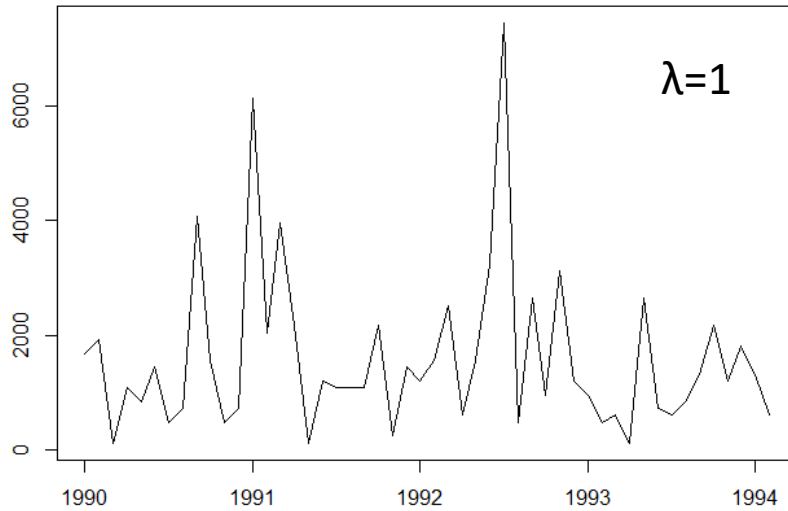
Λύση: Κανονικοποίηση μέσω λογαρίθμησης ή άλλων μετασχηματισμών π.χ. Box-Cox.

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

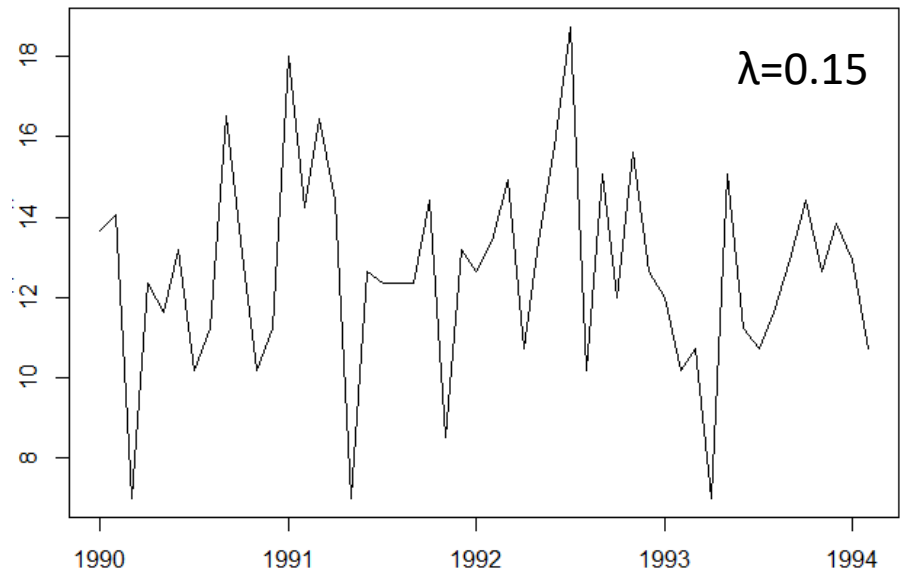


Σύγκριση γραφήματος της περιθώριας κατανομής μίας μη κανονικής χρονοσειράς πριν (αριστερά) και μετά (δεξιά) την εφαρμογή του μετασχηματισμού Box-Cox

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 συστηματικών 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 $1^{η}$ έως $n-1$ τάξης. Στη πράξη χρησιμοποιείται μέχρι $2^{η}$ τάξης.

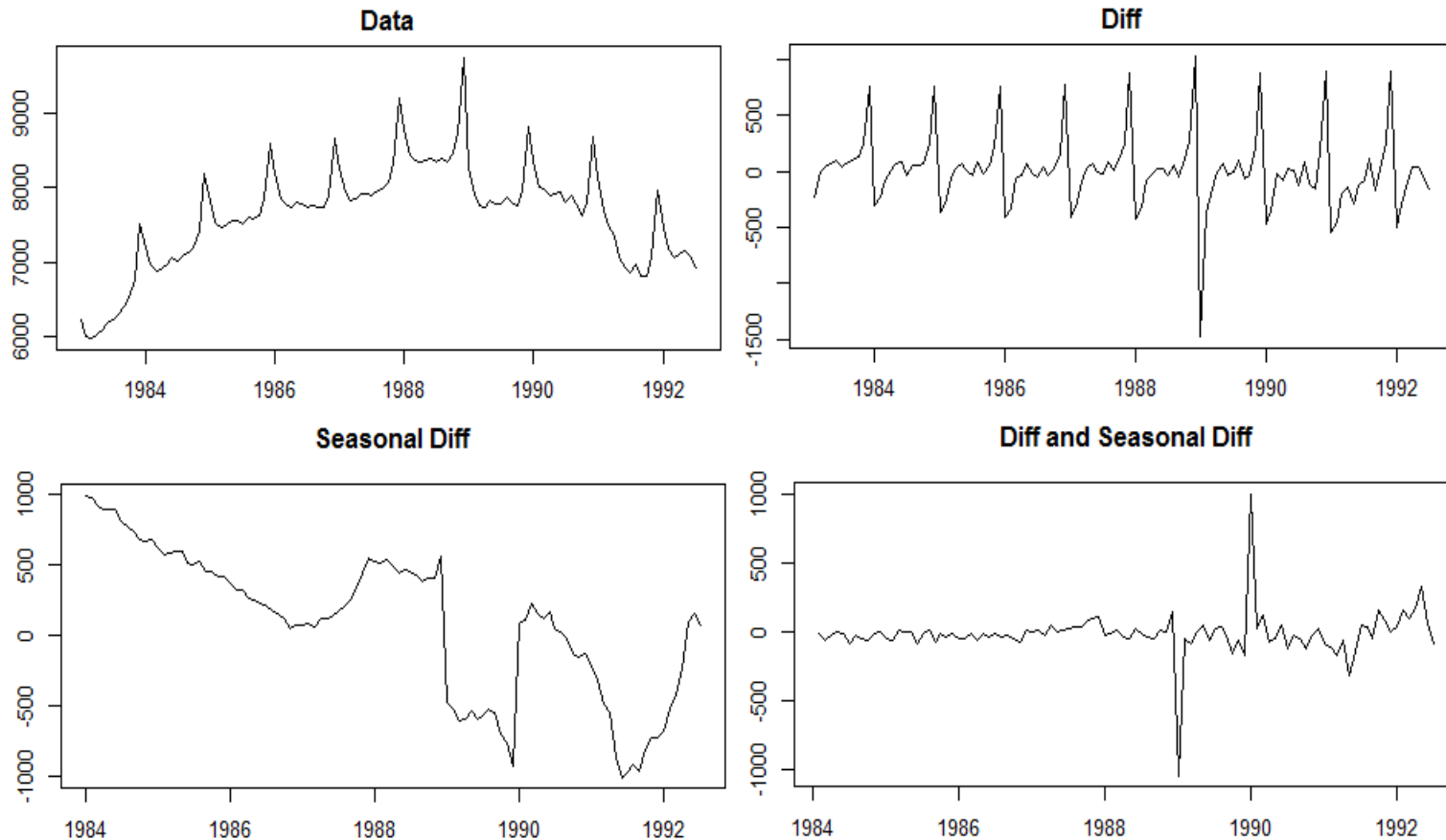
- $1^{η}$ τάξη: $y'_t = y_t - y_{t-1}$ για απαλοιφή γραμμικής τάσης
- $2^{η}$ τάξη: $y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}$ για απαλοιφή μεταβαλλόμενης τάσης

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 $1^{η}$ 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 $(1-B)^n y_t$ και την εποχιακή διαφόριση m τάξης N ως $(1-B^m)^N y_t$.

Διαφόριση (2/2)



Εδώ η αρχική χρονοσειρά (*Data*) έχει εμφανή μηνιαία εποχιακότητα. Αν εφαρμοστεί διαφόριση 1^{ης} 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.

Αυτοσυσχέτιση (ACF) και μερική αυτοσυσχέτιση (PACF)

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

$$p_k = \frac{\sum_{t=k+1}^T (y_t - \mu)(y_{t-k} - \mu)}{\sum_{t=1}^T (y_t - \mu)^2}$$

- **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.

$$\varphi_{11} = \rho_1, \varphi_{kk} = p_k - \sum_{j=1}^{k-1} \frac{\varphi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} \rho_j} \text{ για } k = 2, 3 \dots \text{ και}$$

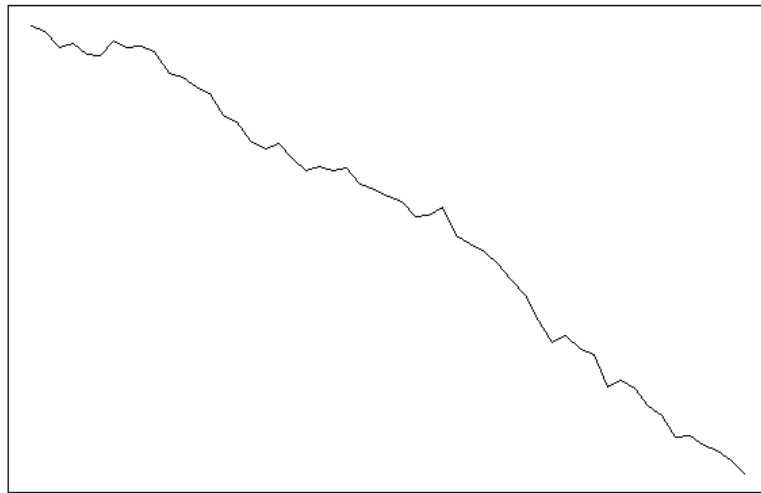
$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk} \varphi_{k-1,k-j} \text{ για } k = 3, 4, \dots, j = 1, 2, \dots$$

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

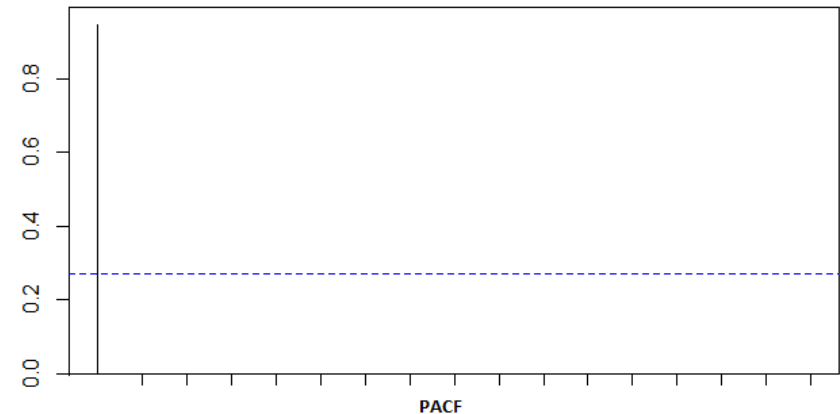
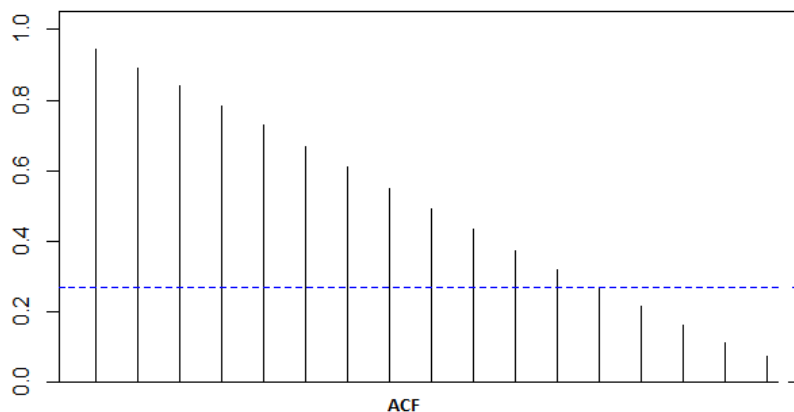
Αυτοσυσχέτιση (ACF) και μερική αυτοσυσχέτιση ($PACF$)

- 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)

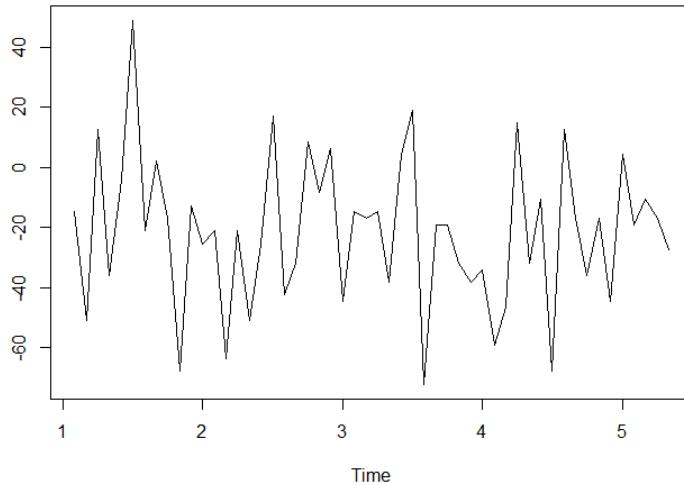
Διαγράμματα ACF – PACF (1/2)



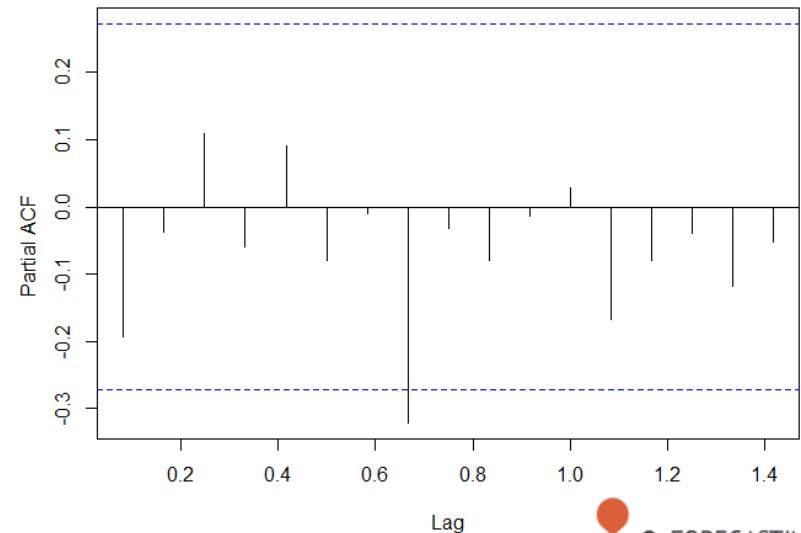
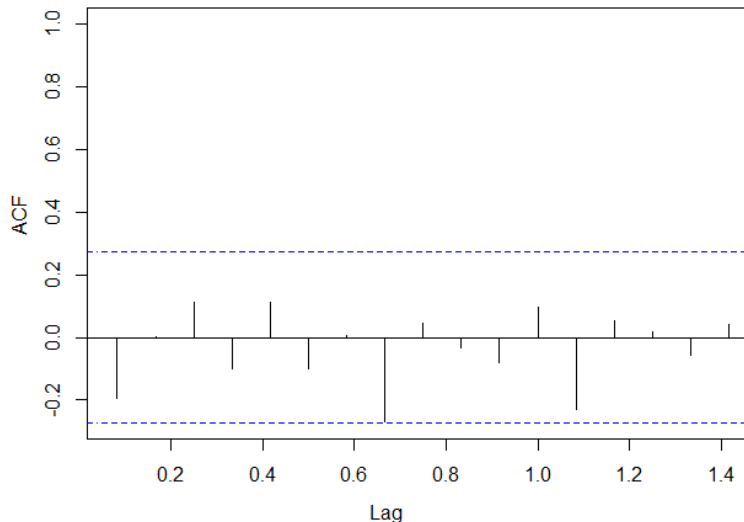
Μεγάλη συσχέτιση μεταξύ διαδοχικών παρατηρήσεων λόγω τάσης (ACF). Η επόμενη τιμή εξαρτάται αποκλειστικά από την προηγούμενη (PACF).



Διαγράμματα ACF – PACF (2/2)



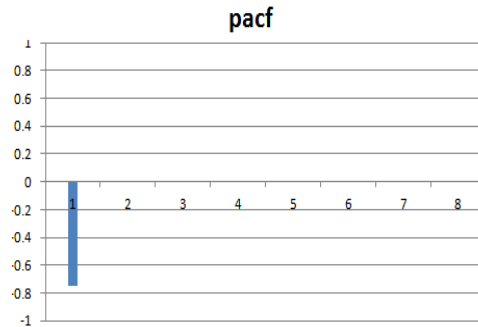
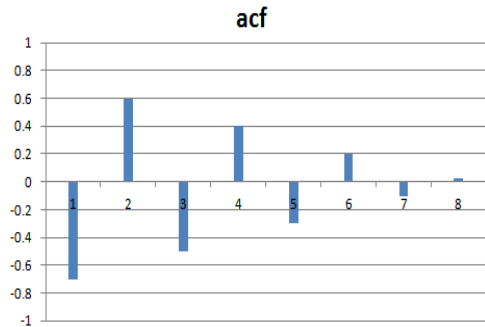
Έχοντας διαφορίσει την προηγούμενη χρονοσειρά η συσχέτιση μεταξύ των παρατηρήσεων έχει εξαλειφθεί πλήρως. Η επιλογή του μοντέλου ARMA θα πρέπει να γίνει βάσει αυτών των διαγραμμάτων



Αναγνώριση καταλληλότητας μοντέλων ARMA (1/3)

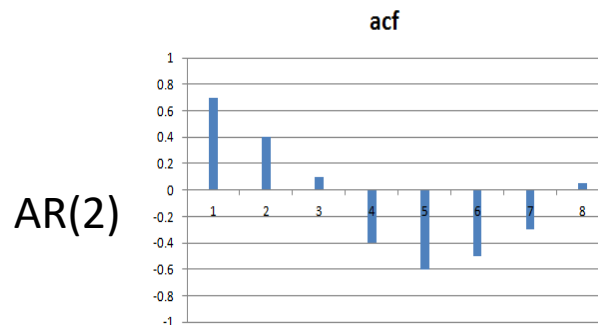
Για ένα στάσιμο μοντέλο AR(p):

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

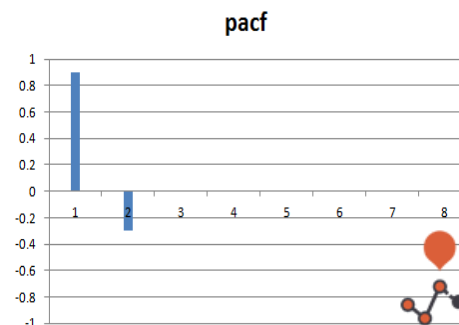


$$y_t = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p) + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t$$

AR(1)



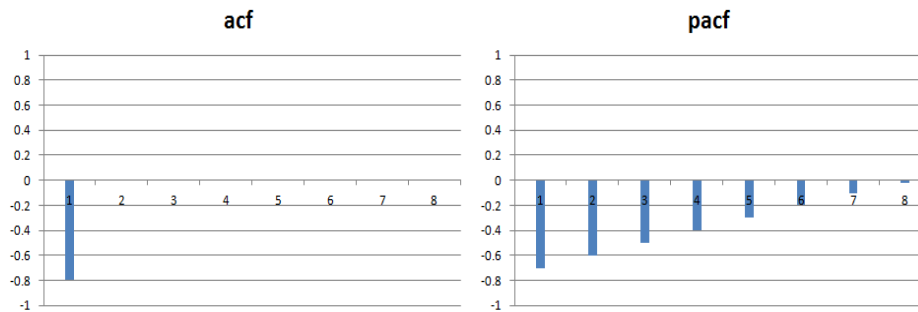
AR(2)



Αναγνώριση καταλληλότητας μοντέλων ARMA (2/3)

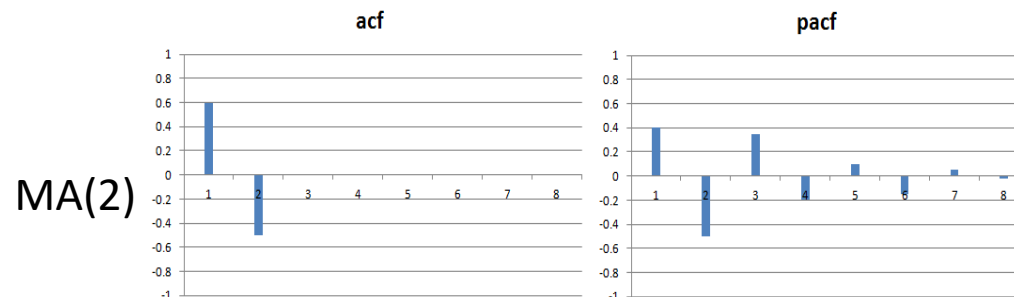
Για ένα στάσιμο μοντέλο MA(q):

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



$$y_t = \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t$$

MA(1)

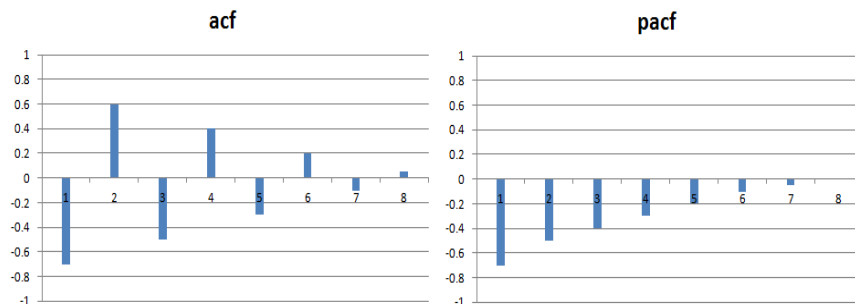


MA(2)

Αναγνώριση καταλληλότητας μοντέλων ARMA (3/3)

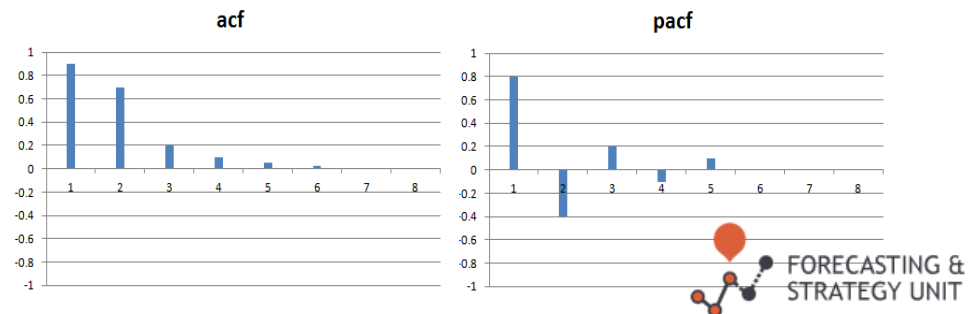
Για ένα στάσιμο μοντέλο ARIMA(p,q):

- ❑ 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



ARMA(1,1)

ARMA(1,1)



Κριτήρια επιλογής μοντέλων

Παρέχουν πληροφορία σχετικά με την ποιότητα ενός στατιστικού μοντέλου για ένα συγκεκριμένο σύνολο δεδομένων αξιολογώντας την πιθανότητα ταύτισης των παραγόμενων από αυτό τιμών με το σύνολο και την πολυπλοκότητά του.

❑ Προσδοκώμενη πιθανοφάνεια (Likelihood):

$$-2\log L = n [\log(2\pi) + 1 + \log\left(\frac{RSS}{n}\right)]$$

❑ Akaike's Information Criterion (AIC):

$$AIC = -2\log(L) + 2(p + q + k + 1)$$

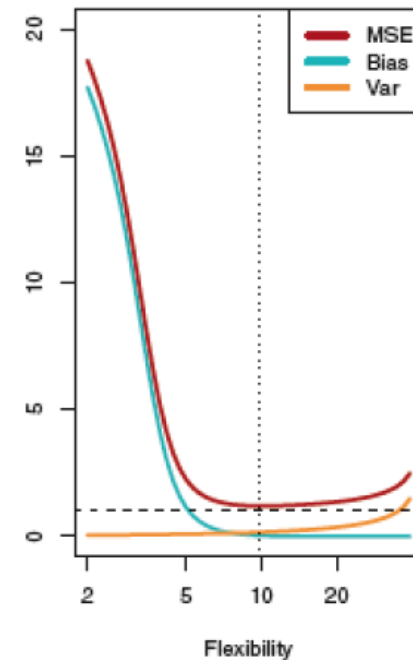
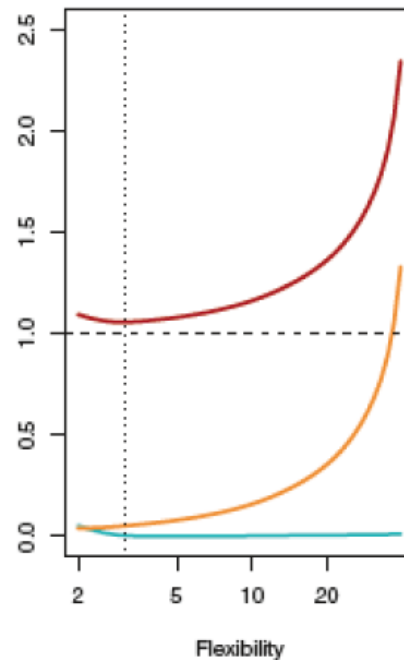
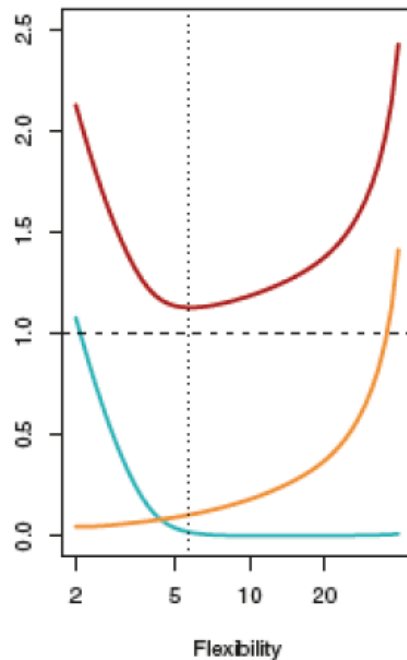
$$AIC_c = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{n - p - q - k - 2}$$

❑ Bayesian Information Criterion (BIC):

$$BIC = AIC + \log(T) (p + q + k + 1)$$

,όπου n ο αριθμός των παρατηρήσεων, RSS το άθροισμα τετραγώνων των σφαλμάτων πρόβλεψης και p , d , q οι συντελεστές του μοντέλου ARMA. Ο συντελεστής k ισούται με 1 ή 0 ανάλογα με το αν το μοντέλο έχει ή όχι σταθερά.

Κριτήρια επιλογής μοντέλων



Έλεγχος στατιστικής σημαντικότητας μοντέλου: t-test τυπικού σφάλματός $S(r_k(e))$.

Τα σφάλματα ενός μοντέλου ARIMA δεν είναι ποτέ τελείως ασυσχέτιστα, όσο καλά και αν αυτό περιγράφει τα δεδομένα. Μπορούμε ωστόσο να εξακριβώσουμε αν η ύπαρξη αυτής της συσχέτισης υποδηλώνει κάτι, αν δηλαδή τα σφάλματα ακολουθούν ένα μοτίβο. Σε μία τέτοια περίπτωση το μοντέλο κρίνεται ακατάλληλο αφού οδηγεί σταθερά σε μεροληπτικές και ανακριβείς προβλέψεις.

Λύση: Στατιστικός έλεγχος σημαντικότητας συσχέτισης σφαλμάτων

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

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

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

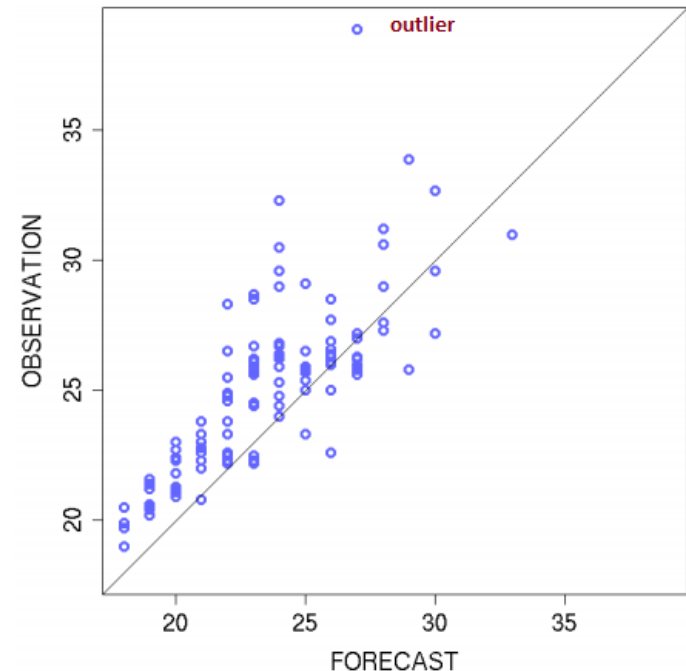
Εποπτικός έλεγχος καταλληλότητας μοντέλου ARIMA (1/3)

Scatter plot

Παρουσιάζει τις πραγματικές τιμές των δεδομένων έναντι των προβλεπόμενων.

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

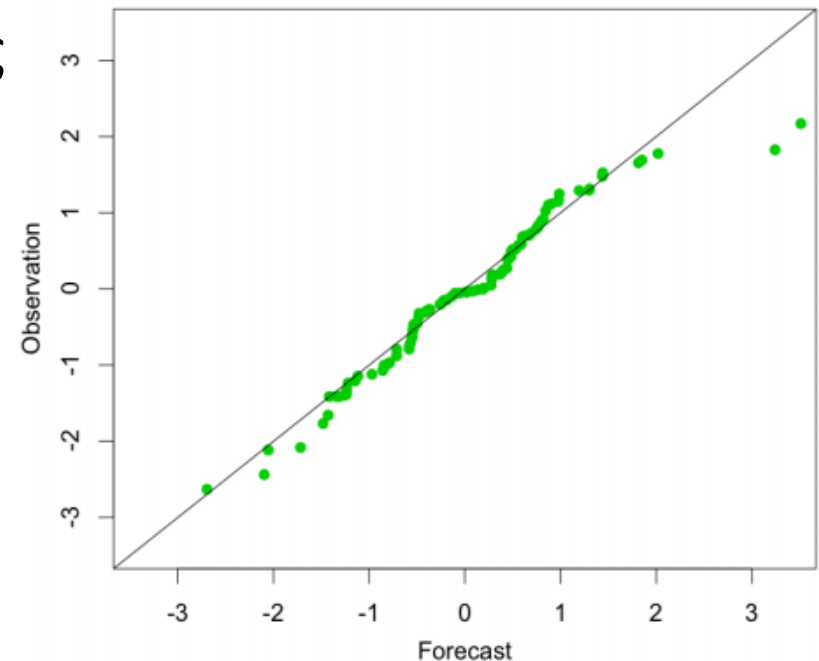
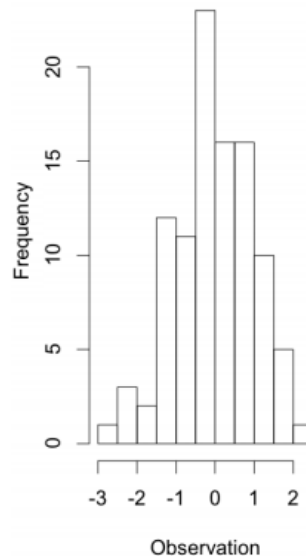
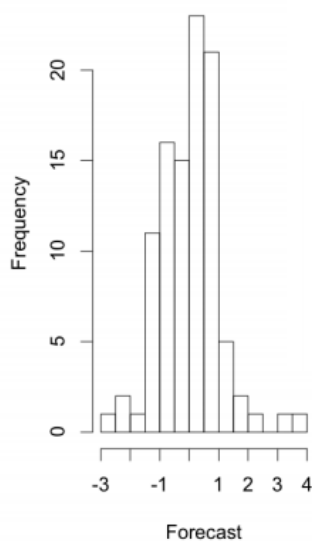
Πληροφορεί σχετικά με: γραμμική συσχέτιση, outliers, προκατάληψη



Εποπτικός έλεγχος καταλληλότητας μοντέλου ARIMA (2/3)

Quantile-Quantile plot

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



Εποπτικός έλεγχος καταλληλότητας μοντέλου ARIMA (3/3)

Box-Plot

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

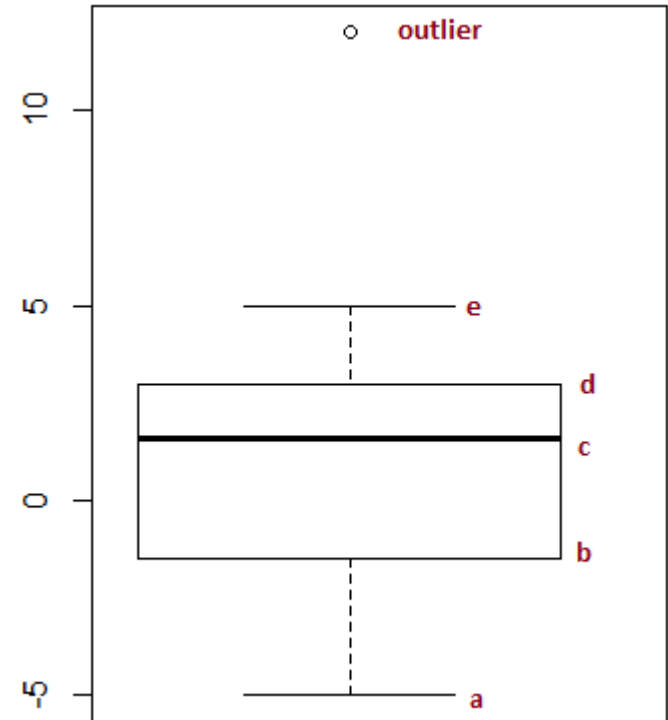
a: min

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.

Μοντέλα Αυτοπαλινδρόμησης

Autoregressive models- AR(p)

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

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

$$\text{ή αν } \bar{y}_t = y_t - \mu$$

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) \bar{y}_t = e_t$$

, η οποία μεταβλητή έχει τις ίδιες ιδιότητες με την αρχική y_t με μηδενική μέση τιμή, συνεπώς σέβεται τη στασιμότητα. Εξισώνοντας έχουμε:

$$\bar{y}_t - \varphi_1 B \bar{y}_t - \varphi_2 B^2 \bar{y}_t - \dots - \varphi_p B^p \bar{y}_t = e_t \rightarrow$$

$$\bar{y}_t - \varphi_1 \bar{y}_{t-1} - \varphi_2 \bar{y}_{t-2} - \dots - \varphi_p \bar{y}_{t-p} = e_t \rightarrow$$

$$y_t - \mu - \varphi_1 (y_{t-1} - \mu) - \varphi_2 (y_{t-2} - \mu) - \dots - \varphi_p (y_{t-p} - \mu) = e_t \rightarrow$$

$$y_t = \mu(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p) + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + e_t$$

$$c = \mu(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$$

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

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

$$\bar{y}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t$$

Εξισώνοντας έχουμε:

$$\bar{y}_t = e_t - \theta_1 B e_t - \theta_2 B^2 e_t - \dots - \theta_q B^q e_t \rightarrow$$

$$y_t = \mu - \theta_1 B e_t - \theta_2 B^2 e_t - \dots - \theta_q B^q e_{t-q} + e_t \rightarrow$$

$$y_t = \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t$$

$$c = \mu$$

Μοντέλα ARIMA (p,d,q)

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

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^n(1 - B^m)^N y_t = c + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t$$

ή

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^n(1 - B^m)^N \overline{y_t} = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t$$

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

και $c = 0$ σε κάθε άλλη περίπτωση

Ένας πρώτος και απλός τρόπος ελέγχου της καταλληλότητας του μοντέλου είναι η ύπαρξη στασιμότητας. Αποδεικνύεται ότι για τα πλέον χρησιμοποιούμενα μοντέλα αυτό ισχύει όταν ικανοποιούνται οι παρακάτω συνθήκες:

- ✓ 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 $\vartheta_1 - \vartheta_2 < 1$
- ✓ Για μοντέλα AR(p) με $p > 2$ πρέπει τουλάχιστον να ισχύει $\varphi_1 + \varphi_2 + \varphi_3 + \dots + \varphi_p < 1$

Πότε χρησιμοποιώ σταθερά;

- ❑ Η μη διαφόριση ($d=0$) είναι δείγμα ύπαρξης σταθερότητας στην αρχική χρονοσειρά. Ωστόσο αυτή η σταθερότητα ενδέχεται εν τέλει να είναι ανεπαρκής. Η εισαγωγή λοιπόν μίας σταθεράς c μπορεί να βοηθήσει στον καλύτερο προσδιορισμό του επιπέδου της.
- ❑ Η διαφόριση πρώτης τάξης ($d=1$) συνεπάγεται ύπαρξη σταθερής τάσης στην αρχική χρονοσειρά. Αυτή θεωρητικά έχει εξαλειφτεί μετά τη διαφόριση, ωστόσο ενδέχεται να έχει υποπέσει στην παραπάνω περίπτωση. Έτσι, η εισαγωγή μίας σταθεράς μπορεί να γίνει, έχοντας πρώτα βεβαιωθεί για την ανάγκη τόνωσης του επίπεδου της παραγόμενης χρονοσειράς. Στην πράξη αυτό αποφεύγεται.
- ❑ Η διαφόριση δεύτερης τάξης ($d=2$) συνεπάγεται χρονικά μεταβαλλόμενη τάση (τάση μέσα στην τάση) στην αρχική χρονοσειρά. Έτσι, η εισαγωγή σταθεράς σε αυτό την περίπτωση θεωρείται άστοχη επιλογή.

Υπολογισμός συντελεστών σε μοντέλα ARIMA

ARMA(p, q)	ρ_1	ρ_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:

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

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

$$\rho_k = \rho_1^k \text{ για } k=1,2,\dots,p$$

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 $y_{t-1}, y_{t-2} \dots y_{t-p}$ ή/και των τιμών $e_{t-1}, e_{t-2} \dots e_{t-q}$.

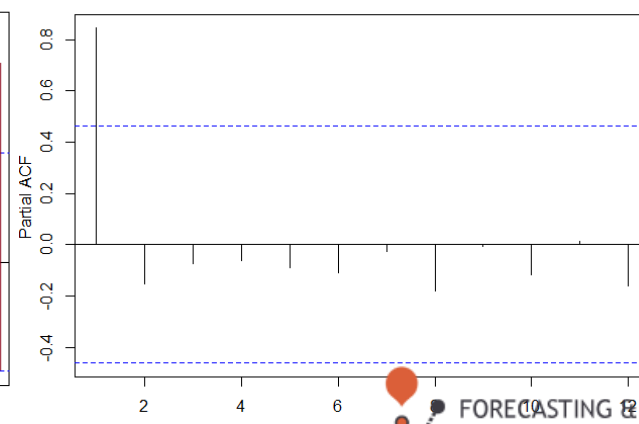
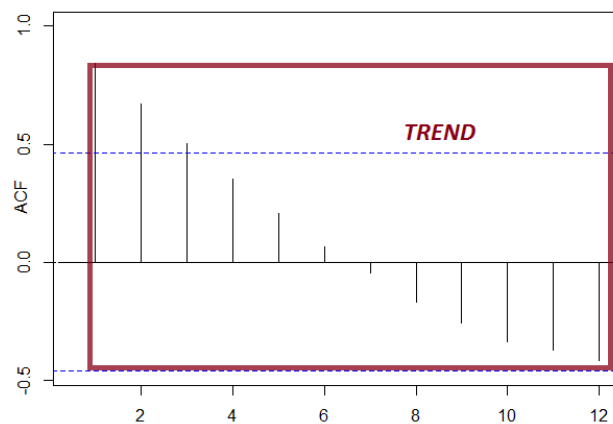
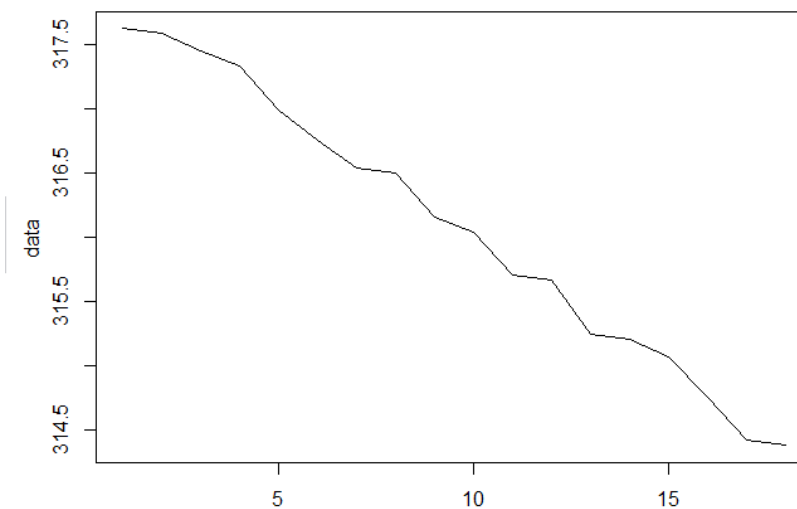
Για μεγάλους ορίζοντες πρόβλεψης οι παραπάνω τιμές ενδέχεται να μην είναι διαθέσιμες. Όταν συμβαίνει αυτό χρησιμοποιούμε τις αντίστοιχες τιμές πρόβλεψης του μοντέλου. Με αυτήν τη λογική υπάρχει πάντα τιμή y_{t-p} διαθέσιμη, όχι όμως και e_{t-q} αφού δε μπορούμε να γνωρίζουμε το μελλοντικό σφάλμα.

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.

Εφαρμογή

Δίνεται η παρακάτω χρονοσειρά. Ζητείται πρόβλεψη για ορίζοντα τρία.

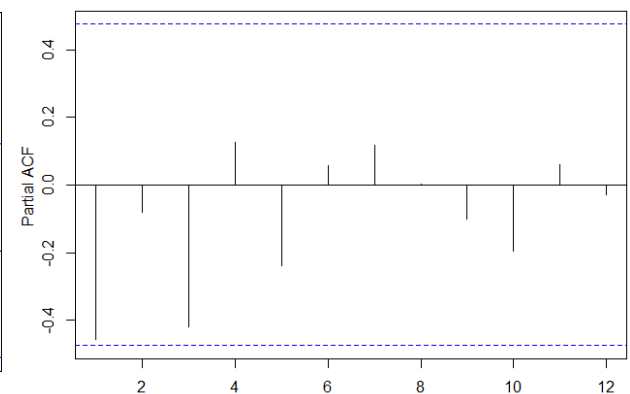
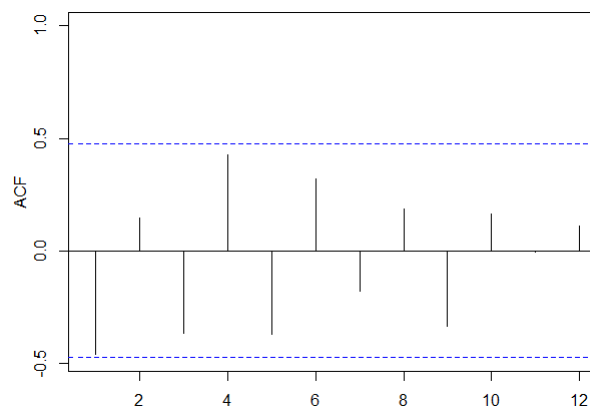
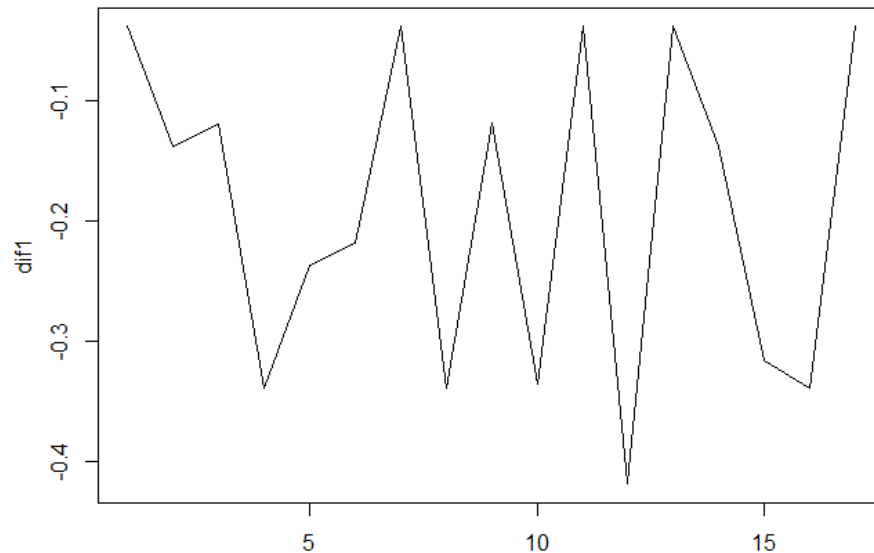
#	Data
1	317.62
2	317.58
3	317.44
4	317.33
5	316.99
6	316.75
7	316.53
8	316.49
9	316.16
10	316.04
11	315.70
12	315.66
13	315.24
14	315.21
15	315.07
16	314.75
17	314.41
18	314.38



Η χρονοσειρά έχει εμφανή τάση. Πρέπει επομένως να προχωρήσουμε σε διαφορίση μέχρι 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 διαφορίση 1^{ης} τάξης επαρκεί αφού οδηγεί σε στασιμότητα (Μ.Τ γύρω από -0.19).



Έχοντας εξασφαλίσει στασιμότητα επιλέγουμε μοντέλο πρόβλεψης βάση των διαγραμμάτων ACF & PACF. Στο πρώτο παρατηρείται ημιτονοειδής κυμάτωση ενώ στο δεύτερο μεγάλη τιμή μερικής αυτοσυσχέτισης για υστέρηση 1. Επιλέγουμε λοιπόν το μοντέλο AR(1).

#	Δεδομένα	Πρώτες διαφορές	yt-μ	Αριθμητής ρ1	Παρονομαστής ρ1
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
M.O.	316.075	-0.191		-0.134	0.295

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

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

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

$$(1 - \varphi_1 B)y_t(1 - B) = e_t + c$$

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

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

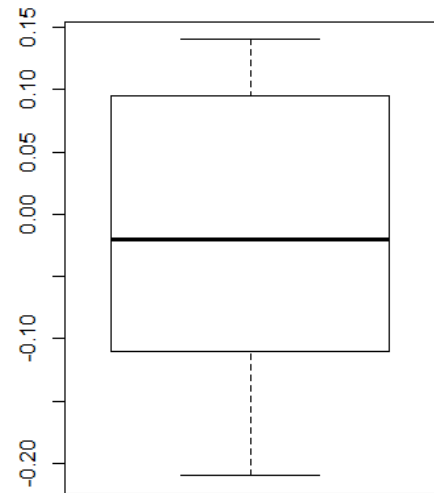
$$y_t = e_t + 0.545y_{t-1} + 0.455y_{t-2} + c$$

Εδώ $c = \mu(1 - \phi_1) = -0.277$

Επισημαίνουμε ότι σαν μέσος χρησιμοποιείται η μέση τιμή της 1^{ης} διαφοράς της χρονοσειράς, αφού αυτή είναι η στάσιμη. Τη σταθερά τη χρησιμοποιούμε ούτως ώστε να προδώσουμε στασιμότητα γύρω από το μηδέν και όχι από τον μέσο της μ , όπως ακριβώς προϋποθέτει το μοντέλο.

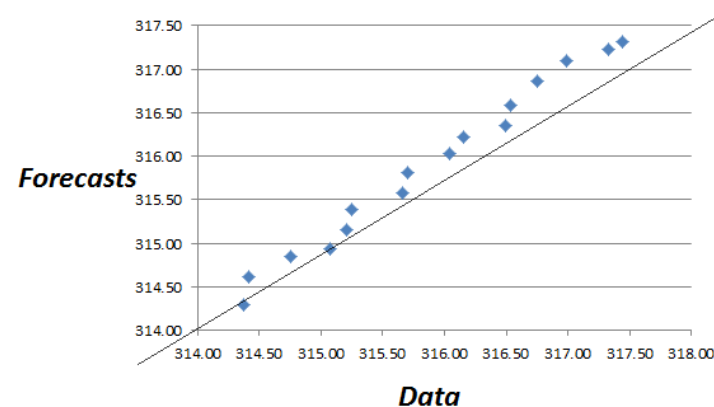
- Για τις προβλέψεις f_2 και f_3 για τις οποίες στερούμαστε δεδομένα θεωρούμε ως y_{t-1} , y_{t-2} τις αντίστοιχες προηγούμενες προβλέψεις

#	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
f_1	-	314.12	-	-
f_2	-	313.96	-	-
f_3	-	313.75	-	-



Το μοντέλο είναι σχετικά **αισιόδοξο**:

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



$$-2\log L = n \log(2\pi) + n + n \log\left(\frac{RSS}{n}\right) =$$

$$16 * 1.838 + 16 - 16 * 4.351 = -24.208, \text{ ή } L=12.104$$

Εδώ θα προτιμούσα το μοντέλο σε σχέση π.χ. με τη Naïve που έχει $L=1.3$

- $AIC = -2\log L + 2(1+0+1+1) = -18.208$
- $AICc = AIC + 2(1+0+1+1) * (1+0+1+2) / (16-1-0-1-2) = -16.208$
- $BIC = AIC + \log(16)(1+0+1+1) = -15.435$

Αυτοσυσχέτιση σφαλμάτων

Αρκετά ασθενής για τις πρώτες
υστερήσεις: Δείγμα
καταλληλότητας μοντέλου

lag	ACF error
r1	-0.196
r2	-0.321
r3	-0.156
r4	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 στατιστικά αξιόπιστο.

lag	ACF error	ACF sq.	Sk	t
r1	-0.196	0.039	0.250	-0.786
r2	-0.321	0.103	0.259	-1.238
r3	-0.156	0.024	0.283	-0.550
r4	0.283	0.080	0.289	0.980