ΚΕΦΑΛΑΙΟ 6

ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ΧΡΟΝΟΣΕΙΡΩΝ

6.1 Εισαγωγή

Ένα μονομεταβλητό ARIMA υπόδειγμα χρησιμοποιείται για:

- Μονομεταβλητές προβλέψεις (πολύ χρήσιμες αλλά μόνο για βραχυχρόνιο ορίζοντα).
- Την κατασκευή ενός υποδείγματος εκτίμησης επιπτώσεων (Impact assessment models)*.
- Πολυμεταβλητή ανάλυση χρονικών σειρών: εύρεση συνάρτησης μεταφοράς, μελέτη αιτιότητας κατά Granger, πολυμεταβλητές προβλέψεις κ.λ.π). Σε αντίθεση με τις μονομεταβλητές προβλέψεις οι πολυμεταβλητές προβλέψεις μπορούν υπό προϋποθέσεις να είναι κατάλληλες για μεσο-μακροχρόνιο ορίζοντα καθώς στηρίζονται σε διαρθρωτικά υποδείγματα.
- Ανάλυση χρονικής σειράς (σε συνιστώσες) βασιζόμενη στο ARIMA υπόδειγμα της αρχικής σειράς (ARIMA model based decomposition).

(*) Συνώνυμοι όροι: Time series quasi-experiments, Intervention models, Box-Tiao models.

6.2 Μονομεταβλητές προβλέψεις με υποδείγματα ARIMA

- **ΒΕΛΤΙΣΤΗ ΠΡΟΒΛΕΨΗ**

Αντικειμενικός σκοπός στην πρόβλεψη είναι να εκτιμήσουμε τις μελλοντικές τιμές μιας χρονοσειράς με όσο το δυνατό μικρότερο σφάλμα. Ως βέλτιστη πρόβλεψη θεωρείται αυτή για την οποία το **μέσο τετραγωνικό σφάλμα πρόβλεψης** (mean square forecast error) ελαχιστοποιείται. Καθώς το σφάλμα πρόβλεψης είναι μία τυχαία μεταβλητή επιδιώκεται η ελαχιστοποίηση της αναμενόμενης τιμής του:

$$E\left[e_T^2(l)\right] = E\left[(Y_{T+l} - \hat{Y}_T(l))\right]^2$$

Σχετικό είναι και το θεώρημα που ακολουθεί:

- **ΘΕΩΡΗΜΑ ΒΕΛΤΙΣΤΗΣ ΠΡΟΒΛΕΨΗΣ**

Η βέλτιστη (μονομεταβλητή) πρόβλεψη για την τιμή μιας χρονοσειράς $\mathbf{Y_t}$ *l* περιόδους στο μέλλον (*l* step ahead forecast) δίνεται από την υπό συνθήκη αναμενόμενη τιμή της $\mathbf{Y_t}$ με δεδομένη όλη τη διαθέσιμη πληροφορία από το παρελθόν της σειράς μέχρι και την παρούσα χρονική περίοδο.

ΠΟΡΙΣΜΑ

Η βέλτιστη πρόβλεψη *l* περιόδων στο μέλλον δίνεται από τη σχέση:

$\hat{y}_T(l) = \sum_{j=0}^{\infty} \psi_{l+j}\hat{\varepsilon}_{T-j}$ όπου ψ_l οι γνωστές μας σταθμίσεις Ψ.

Απόδειξη

Σύμφωνα με το Θεώρημα Βέλτιστης Πρόβλεψης η $\hat{y}_T(l)$ είναι βέλτιστη (optimal) ανν:

$$\hat{y}_T(l) = E(y_{T+l} / y_T, y_{T-1},...y_1)$$

Εξάλλου από το Θεώρημα του Wold γνωρίζουμε ότι:

$$y_t = \Psi(B)\varepsilon_t = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i}$$

Επομένως:

$$\hat{y}_T(l) = E[\psi_0\varepsilon_{T+l} + \psi_1\varepsilon_{T+l-1} + \ldots + \psi_{l-1}\varepsilon_{T+1} + \sum_{j=0}^{\infty}\psi_{l+j}\varepsilon_{T-j} / y_T, y_{T-1},\ldots y_1] =$$

$$= \sum_{j=0}^{\infty}\psi_{l+j}\hat{\varepsilon}_{T-j}$$

καθώς για χρονικές περιόδους μεταγενέστερες του χρόνου T ισχύει:

$$E\lfloor\varepsilon_{T+j}\rfloor = 0 \ \forall \ \mathrm{j} > 0$$

- **ΣΦΑΛΜΑ ΠΡΟΒΛΕΨΗΣ**

Για την εύρεση του σφάλματος πρόβλεψης *l* περιόδων στο μέλλον με *l* > 1 το πρώτο βήμα είναι να εκφράσουμε το ARIMA υπόδειγμα που περιγράφει τη χρονοσειρά ως μια MA(∞) διαδικασία δηλ.:

$$Y_t = \Phi(B)^{-1}(1-B)^{-d}\,\Theta(B)\varepsilon_t = \Psi(B)\varepsilon_t = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i} \qquad (1)$$

Τότε το σφάλμα πρόβλεψης *l* περιόδων στο μέλλον θα δίνεται από σχέση:

$$e_T(l) = y_{T+l} - \hat{y}_T(l) = \psi_0\varepsilon_{T+l} + \psi_1\varepsilon_{T+l-1} + \ldots + \psi_{l-1}\varepsilon_{T+1} + \sum_{j=0}^{\infty}\psi_{l+j}\varepsilon_{T-j} - \sum_{j=0}^{\infty}\psi_{l+j}\hat{\varepsilon}_{T-j} =$$

$$= \psi_0\varepsilon_{T+l} + \psi_1\varepsilon_{T+l-1} + \ldots + \psi_{l-1}\varepsilon_{T+1} \qquad (2)$$

Όπου οι σταθμίσεις *ψ* προσδιορίζονται από τη σχέση:

$\psi(B)=\varphi^{-1}(B)(1-B)^{-d}\,\theta(B)$

Υποθέτουμε ότι οι παράμετροι του ARIMA υποδείγματος $\varphi_1,..., \varphi_p$ και $\theta_1,..., \theta_q$ είναι γνωστοί και για το λόγο αυτό και οι σταθμίσεις ψ_0, ψ_1,...

είναι επίσης γνωστές. Σε αυτή την περίπτωση από τις (1), (2) προκύπτει ότι η διακύμανση του σφάλματος πρόβλεψης θα δίνεται από τη σχέση:

$$E(e_T^2(l)) = (\psi_0^2 + \psi_1^2 + \ldots + \psi_{l-1}^2)\sigma_\varepsilon^2 \qquad (3)$$

όπου σ_ε^2 η διακύμανση των ε_t.

Συνεπώς, η αλγεβρική μορφή για τη διακύμανση του σφάλματος πρόβλεψης εξαρτάται από το συγκεκριμένο υπόδειγμα ARIMA που έχει υιοθετηθεί.

- **ΠΑΡΑΤΗΡΗΣΕΙΣ**

1) Γνωρίζουμε από τον ορισμό του ψ(Β) ότι $\psi_0=1$. Για αυτό για οποιαδήποτε κατηγορία ARIMA, ξέρουμε ότι το σφάλμα πρόβλεψης μιας περιόδου είναι: $e_T(l)=\varepsilon_{T+1}$

και αυτό έχει διακύμανση σ_ε^2 .

Οπότε, η διακύμανση του σφάλματος πρόβλεψης (σ.π.) μιας περιόδου είναι η διακύμανση του λευκού θορύβου.

2) Η εκτίμηση του σφάλματος πρόβλεψης (σχέση (3)) βασίστηκε στην υπόθεση ότι γνωρίζαμε τις τιμές των παραμέτρων $\varphi_1,\ldots,\varphi_p$ και $\theta_1,\ldots,\theta_q$ με βεβαιότητα. Όμως αυτές οι παράμετροι εκτιμώνται μέσω μιας μη γραμμικής παλινδρόμησης ελάχιστων τετραγώνων και οι εκτιμήσεις είναι τυχαίες μεταβλητές με μέσους και διακυμάνσεις. Συνεπώς, η πραγματική διακύμανση του σ.π. θα είναι μεγαλύτερη από την διακύμανση που δίνεται παραπάνω. Για να καθορίσουμε ακριβώς το πόσο μεγαλύτερη είναι η πραγματική διακύμανση, πρέπει να γνωρίζουμε τις διακυμάνσεις των εκτιμώμενων παραμέτρων στο υπόδειγμα ARIMA. Επειδή, οι παράμετροι εκτιμήθηκαν μη γραμμικά το καλύτερο που μπορούμε να κάνουμε είναι να υπολογίσουμε τα τυπικά σφάλματα

βασιζόμενοι στην τελευταία επανάληψη της μη γραμμικής διαδικασίας εκτίμησης.

Η δυσκολία εδώ, είναι το ότι τα τυπικά σφάλματα για τη γραμμικοποίηση στη τελευταία επανάληψη δεν είναι αληθείς εκτιμήσεις των πραγματικών τυπικών σφαλμάτων για τις τιμές των παραμέτρων. Από πρακτικής απόψεως, μας δίνεται η επιλογή ή να κάνουμε χρήση τέτοιων τυπικών σφαλμάτων στον υπολογισμό της διακύμανσης του σ.π., ή μπορούμε να τα αγνοήσουμε και απλά να υπολογίσουμε την διακύμανση του σ.π. βασιζόμενοι στην εξίσωση (3).

- **ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΠΡΟΒΛΕΨΗΣ**

Πριν υπολογίσουμε ένα διάστημα εμπιστοσύνης (δ.ε.) για την πρόβλεψη, χρειαζόμαστε έναν εκτιμητή $\hat{\sigma}_{\varepsilon}^{2}$ για την διακύμανση του διαταρακτικού όρου. Αυτός ο εκτιμητής λαμβάνει υπόψη του το άθροισμα των τετραγωνικών υπολοίπων $S(\hat{\phi}_1,...,\hat{\phi}_p,\hat{\theta}_1,...,\hat{\theta}_q)$ και μπορεί να υπολογισθεί, αφού γίνουν γνωστές οι τελικές εκτιμήσεις των παραμέτρων, με τη σχέση:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{S}{T-p-q} = \frac{\sum_{i=1}^{T} \hat{\varepsilon}_{l}^{2}}{T-p-q}$$

Εδώ, το T-p-q είναι κατά τα γνωστά ο αριθμός των βαθμών ελευθερίας. Βλέπουμε από την εξίσωση αυτή και από το γεγονός ότι $\psi_{0}=1$ ότι ένα δ.ε. n τυπικών αποκλίσεων γύρω από μία πρόβλεψη l περιόδων στο μέλλον θα δινόταν από τη σχέση:

$$C_n = \hat{y}_T(l) \pm n(1+\sum_{j=1}^{l-1}\psi_j^2)^{\frac{1}{2}}\hat{\sigma}_e$$

Όπως αναμενόταν το διάστημα αυτό μεγαλώνει, καθώς μεγαλώνει το *l* αν και το ακριβές πρότυπο εξαρτάται από τις σταθμίσεις ψ_j.

Οι προβλέψεις των y_t μαζί με το 66% δ.ε. και το 99% δ.ε. φαίνονται για ένα υποθετικό ARIMA μοντέλο (d=0) στο σχήμα που επισυνάπτεται στο τέλος. Σημειώνουμε ότι οι προβλέψεις (σημειώνονται με σταυρούς) πρώτα είναι αυξανόμενες και στην συνέχεια μειώνονται στο επίπεδο του σταθερού μέσου της σειράς. Γνωρίζουμε ότι η πρόβλεψη θα προσεγγίζει το μέσο της σειράς όσο ο χρόνος προήγησης μεγαλώνει, αφού η σειρά είναι στάσιμη. Το δ.ε., φυσικά, αυξάνει καθώς ο χρόνος προήγησης αυξάνει.

6.3 Εφαρμογές

Σε όλες τις περιπτώσεις που ακολουθούν θεωρούμε ότι γνωρίζουμε με βεβαιότητα τις παραμέτρους του συγκεκριμένου υποδείγματος ARIMA.

i) ΠΡΟΒΛΕΨΗ ΜΕ ΔΙΑΔΙΚΑΣΙΑ AR(1) ΜΕ ΣΤΑΘΕΡΟ ΟΡΟ

Έστω η στάσιμη πρώτης τάξεως αυτοπαλίνδρομη διαδικασία AR(1):

$$y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t \text{ όπου } \delta \text{ σταθερά}$$

Για αυτή τη διαδικασία η πρόβλεψη της μίας περιόδου είναι

$$\hat{y}_T(1) = E(y_{T+1} \mid y_T, ..., y_1) = \phi_1 y_T + \delta$$

Ομοίως, $\hat{y}_T(2) = \phi_1 \hat{y}_T(1) + \delta = \phi_1^2 y_T + (\phi_1 + 1)\delta$

Και η πρόβλεψη *l*-περιόδων είναι:

$$\hat{y}_T(l) = \phi_1^l y_T(1) + (\varphi_1^{l-1} + \phi_1^{l-2} + ... + \phi_1 + 1)\delta$$

Σημειώνουμε ότι στο όριο καθώς το *l* μεγαλώνει, η πρόβλεψη συγκλίνει στην τιμή:

$$\lim_{l \to \infty} \hat{y}_T(l) = \delta \sum_{j=0}^{\infty} \phi_1^j = \frac{\delta}{1-\phi_1} = \mu_y$$

Βλέπουμε επομένως, ότι η πρόβλεψη τείνει στο μέσο της σειράς καθώς το *l* μεγαλώνει. Φυσικά αυτό δεν αποτελεί έκπληξη αφού η σειρά είναι στάσιμη. Καθώς το *l* γίνεται πολύ μεγάλο δεν υπάρχει στην ουσία καμία χρήσιμη πληροφορία στις πρόσφατες τιμές της χρονοσειράς, y_T, y_{T-1} ,....κ.λ.π. που να μπορεί να χρησιμοποιηθεί για να προσαρμόσουμε την πρόβλεψη μακριά από την μέση τιμή. Επομένως, για ένα πολύ μεγάλο *l* η καλύτερη πρόβλεψη είναι ο σταθερός μέσος της σειράς.

Ας υπολογίσουμε τώρα το σ.π. για αυτή τη διαδικασία. Το σ.π. για *l* περιόδους δίνεται από τη σχέση:

$$e_T(l) = y_{T+l} - \hat{y}_T(l) = \phi_1 y_{T+l-1} + \delta + \varepsilon_{T+l} - \hat{y}_T(l)$$
$$= \phi_1^2 y_{T+l-2} + (\phi_1 + 1)\delta + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} - \hat{y}_T(l)$$

....

$$= \varphi_1^l y_T + (\varphi_1^{l-1} + \phi_1^{l-2} + ... + \phi_1 + 1)\delta + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + ... + \phi_1^{l-1} \varepsilon_{T+1} - \hat{y}_T(l)$$

Τώρα αντικαθιστώντας στην εξίσωση για $\hat{y}_T(l)$ παίρνουμε:

$$e_T(l) = \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + ... + \phi_1^{l-1} \varepsilon_{T+1}$$

το οποίο έχει διακύμανση:

$$E[e_T^2(l)] = (1 + \phi_1^2 + \phi_1^4 + ... + \phi_1^{2l-2})\sigma_\varepsilon^2$$

Να σημειωθεί ότι η διακύμανση αυτού του σφάλματος πρόβλεψης αυξάνεται μη γραμμικά καθώς μεγαλώνει το *l* (βλ. και σχήμα στο τέλος).

ii) ΠΡΟΒΛΕΨΗ ΜΕ ΜΑ(1) ΔΙΑΔΙΚΑΣΙΑ ΜΕ ΣΤΑΘΕΡΟ ΟΡΟ

Τώρα ας εξετάσουμε την απλή πρώτης τάξης διαδικασία κινητού μέσου ΜΑ(1) με σταθερό όρο:

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Η πρόβλεψη μιας περιόδου είναι

$$\hat{y}_T(1) = E(y_{T+1} \mid y_T, ..., y_1) = \delta - \theta_1 \hat{\varepsilon}_T$$

Όπου $\hat{\varepsilon}_T$ είναι τα πραγματικά υπόλοιπα από την πρόσφατη παρατήρηση.

Από την άλλη πλευρά η πρόβλεψη της l περιόδου για $l > 1$ δίνεται από τη σχέση:

$$\hat{y}_T(1) = E(y_l \mid y_T, ..., y_1) = E[\delta + \varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1}] = \delta$$

Αυτό είναι επίσης αναμενόμενο αφού η διαδικασία ΜΑ(1) έχει μνήμη <u>μιας μόνο περιόδου</u>. Έτσι, τα πρόσφατα δεδομένα είναι άχρηστα στο να υπολογίσουμε πρόβλεψη για δύο ή περισσότερες περιόδους στο μέλλον. Η καλύτερη πρόβλεψη είναι ο μέσος της σειράς.

Η διακύμανση του σ.π. για την ΜΑ(1) είναι σ_ε^2 για μία πρόβλεψη μιας περιόδου ενώ για την πρόβλεψη l περιόδων με l>1 δίνεται από τη σχέση:

$$E[e_T^2(l)] = E\{[y_{T+l} - \hat{y}_T(l)]^2\} = E[(\varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1})^2 = (1 + \theta_1^2)\sigma_\varepsilon^2$$

Έτσι, η διακύμανση του σφάλματος πρόβλεψης είναι η ίδια για δύο περιόδους, τρεις περιόδους, κλπ.

Το διάστημα εμπιστοσύνης για τις προβλέψεις φαίνεται στο σχήμα που επισυνάπτεται στο τέλος.

iii) ΠΡΟΒΛΕΨΗ ΜΕ ARIMA(0,1,0) ΔΙΑΔΙΚΑΣΙΑ

Ως γνωστόν η διαδικασία ARIMA(0,1,0) είναι μη-στάσιμη και στην περίπτωση που δεν υπάρχει σταθερός όρος μπορεί να γραφεί ως:

$Y_t = Y_{t-1} + \varepsilon_t$

Οι σταθμίσεις ψ στην περίπτωση αυτή θα είναι:

$\psi_0 = \psi_1 = \psi_2 = ...=1$

Επομένως οι σημειακές προβλέψεις θα είναι:

$$\hat{y}_T(1) = E(y_{T+1} \mid y_T,..., y_1) = E(y_T + \varepsilon_{T+1}) = y_T$$

…………………………………………….

$$\hat{y}_T(l) = E(y_{T+l} \mid y_T,..., y_1) = E(y_T + \varepsilon_{T+1} + ...\varepsilon_{T+l}) = y_T$$

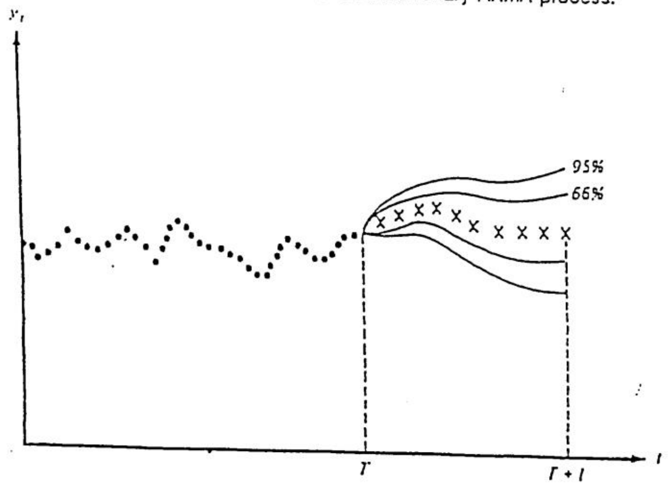
Άρα η καλύτερη πρόβλεψη είναι πάντα ίση με την τιμή της τελευταίας διαθέσιμης παρατήρησης.

Η διακύμανση του σφάλματος πρόβλεψης θα είναι:

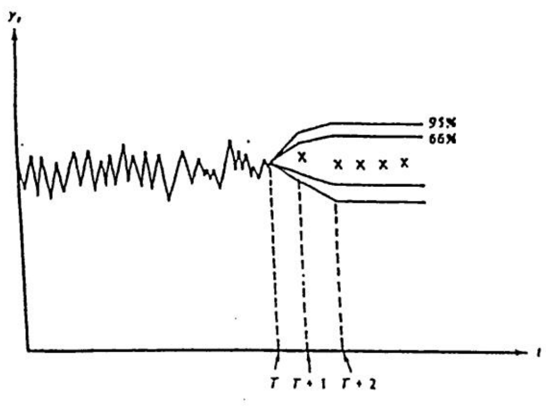
$$E[e_T^2(l)] = E\{[y_{T+l} - \hat{y}_T(l)]^2\} = E[(\varepsilon_{T+l} + \varepsilon_{T+l-1} + ... + \varepsilon_{T+1})^2 = l\sigma_\varepsilon^2$$

Άρα η διακύμανση του σφάλματος πρόβλεψης αυξάνει γραμμικά με το *l* και φυσικά, καθώς έχουμε μη στάσιμη διαδικασία, τείνει στο άπειρο καθώς το *l* τείνει στο άπειρο (βλ. και σχήμα).

Forecasts and confidence intervals for a stationary ARMA process.

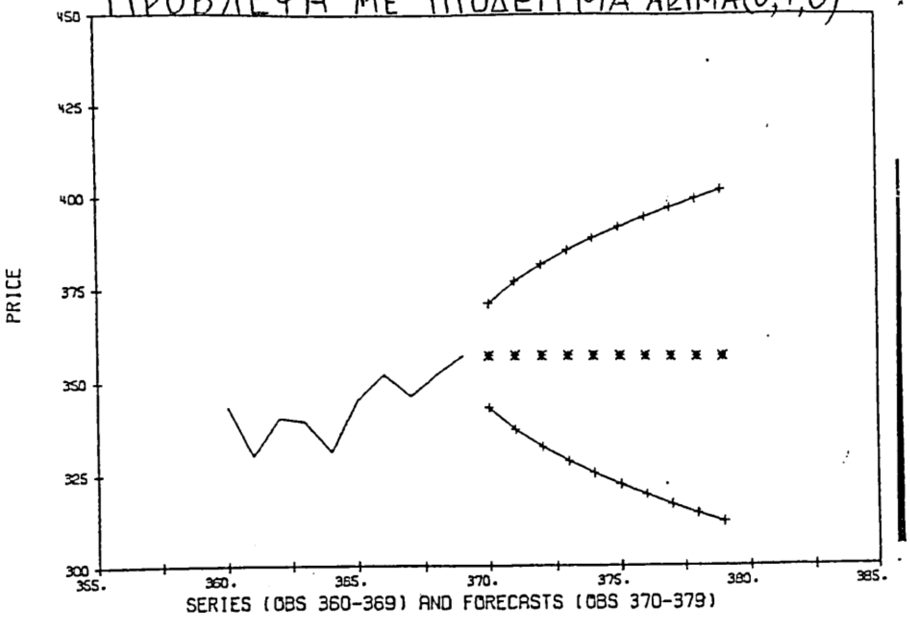


ΠΡΟΒΛΕΨΗ ΜΕ ΥΠΟΔΕΙΓΜΑ MA(1)

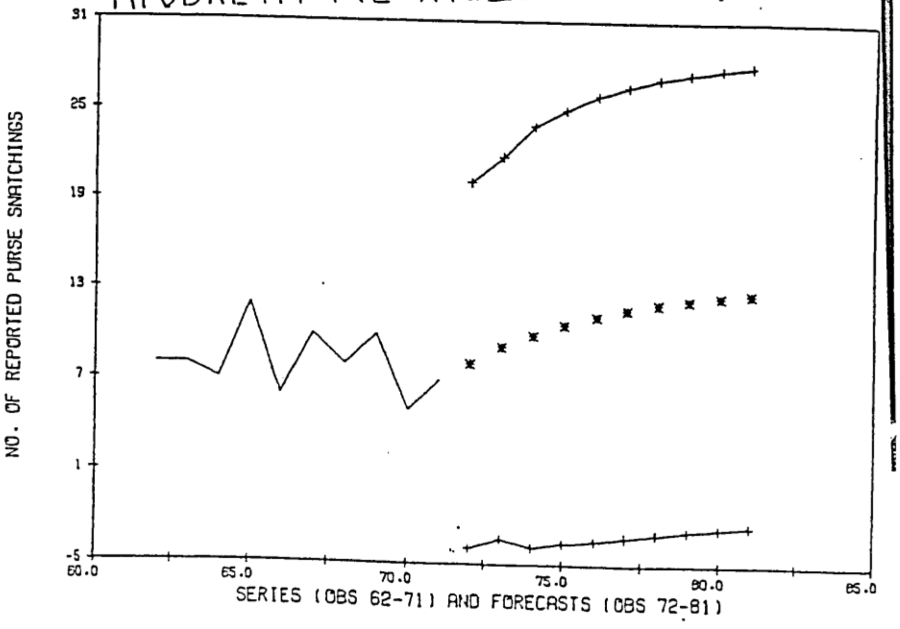


-11-

ΠΡΟΒΛΕΨΗ ΜΕ ΥΠΟΔΕΙΓΜΑ ARIMA(0,1,0)



ΠΡΟΒΛΕΨΗ ΜΕ ΥΠΟΔΕΙΓΜΑ AR



-12-