

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ & ΔΙΟΙΚΗΣΗΣ
ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΣΤΑΤΙΣΤΙΚΟΣ ΠΟΙΟΤΙΚΟΣ ΕΛΕΓΧΟΣ ΓΡΑΜΜΗΣ
ΠΑΡΑΓΩΓΗΣ ΠΥΡΗΝΩΝ ΜΕΤΑΣΧΗΜΑΤΙΣΤΩΝ**

ΖΑΧΑΡΙΑΣ ΔΑΣΚΑΛΑΚΗΣ

ΧΑΝΙΑ, ΙΟΥΝΙΟΣ 1993

ΕΠΙΤΡΟΠΗ ΕΞΕΤΑΣΗΣ

Ι. ΓΡΥΣΠΟΛΑΚΗΣ
ΚΑΘΗΓΗΤΗΣ

Γ. ΣΑΡΙΔΑΚΗΣ
ΑΝ. ΚΑΘΗΓΗΤΗΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ Κ. ΜΗΛΟΛΙΔΑΚΗΣ
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ

Κατ' αρχήν ευχαριστώ τους γονείς μου για όλα όσα μου προσέφεραν αυτά τα χρόνια.

Τους αφιερώνω αυτή τη διπλωματική εργασία.

Επίσης ευχαριστώ τους κ.κ. Περρή και Μανουκιάν για τη βοήθεια που μου προσέφεραν καθ' όλη τη διάρκεια συλλογής στοιχείων στην ΕΛΒΗΜ, όπως επίσης και το διευθυντή του εργοστασίου της εταιρείας κ. Σπηλιόπουλο.

Ακόμη ευχαριστώ τον κ. Ε. Βουντουράκη, υπεύθυνο συστημάτων του εργαστηρίου Δυναμικών Συστημάτων και Προσομοίωσης, την κ. Β. Δινοπούλου, Ε.Μ.Υ., καθώς επίσης και την συμφοιτήριά μου κ. Μαίρη Αλεξάκη για την πολύτιμη βοήθειά τους κατά την συγγραφή αυτής της διπλωματικής.

Τέλος ευχαριστώ τον Επίκουρο Καθηγητή κ. Μηλολιδάκη ο οποίος επέβλεψε κάθε μέρος αυτής της διπλωματικής εργασίας και μου παρείχε κάθε βοήθεια τόσο κατά την ανάλυση όσο και κατά την συγγραφή.

ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή	
1		
2. Διαγράμματα Ελέγχου		
2.1 Εισαγωγή στα Διαγράμματα Ελέγχου	
5		
2.2 Διαγράμματα Ελέγχου μέσω μεταβλητών	
9	2.2.1 \bar{X} -Διαγράμματα	Ελέγχου
	10
2.2.2 R-Διαγράμματα Ελέγχου	
11		
2.2.3 S-Διαγράμματα Ελέγχου	
11		
2.2.4 S^2 -Διαγράμματα Ελέγχου	
12		
2.2.5 ρ -Διαγράμματα Ελέγχου	
13		
2.3 Εφαρμογή των Δ.Ε. στη διαδικασία της κοπής	13
2.4 Εφαρμογή των Δ.Ε. στη διαδικασία της πλέξης	17
3. Γραμμή παραγωγής πυρήνων μετασχηματιστών		
3.1 Περιγραφή	21
3.2 Μεταβλητές	
23		
4. Γραμμικό Μοντέλο		
4.1 Γενικά	
30		
4.2 Πολυμεταβλητή Γενική Γραμμική Υπόθεση	32
4.2.1 Επιλογή καλύτερου υποσυνόλου ανεξαρτήτων μεταβλητών...	
39		
4.2.2 Ελεγχοι καλής προσαρμογής του μοντέλου	
42		
4.2.3 Έλεγχος κανονικότητας στην κατανομή των residuals	
43		

4.2.4	Ελεγχος ομοιογένειας της διασποράς	
	43		
4.2.5	Το μοντέλο πρόβλεψης της διασποράς	
	45		
4.2.6	Διόρθωση του μοντέλου με την τεχνική των		
	Weighted	Least	Squares
	48	
4.2.7	Εύρεση των outliers και παρατηρήσεων με σημαντική		
	επίδραση	στο	μοντέλο
	49	
4.2.8	Ελεγχοι της στοχαστικής ανεξαρτησίας των ϵ_i	
	50		
4.2.9	Διόρθωση των σειριακών συσχετίσεων	
	51		
4.3	Ελεγχοι πρόβλεψης απωλειών του μοντέλου προς		
	την πρόβλεψη απωλειών της μελέτης	
	55		
4.4	Συμπεράσματα	
	59		

Παραρτήματα

Παράρτημα Α

Παράρτημα Β1

Παράρτημα Β2

Παράρτημα Β3

Παράρτημα Β4

Παράρτημα Β5

Παράρτημα Β6

Παράρτημα Β7

Παράρτημα Β8

Βιβλιογραφία

ΕΙΣΑΓΩΓΗ

Η συλλογή στοιχείων από την ΕΛΒΗΜ Α.Ε. ξεκίνησε στις αρχές του Μαΐου 1992 στο εργοστάσιο της εταιρείας που βρίσκεται στα Οινόφυτα Βοιωτίας. Διήρκεσε όλο τον μήνα. Μετά από μία διακοπή μερικών μηνών συνεχίστηκε τον Σεπτέμβριο του ίδιου έτους -πάντα στο εργοστάσιο- και τελείωσε περί το τέλος Νοεμβρίου. Η όλη διαδικασία αποδείχτηκε πιο χρονοβόρα απ' ό,τι είχε θεωρηθεί στην αρχή. Ο λόγος ήταν ότι χρειάστηκε εκ των υστέρων να ληφθούν αποφάσεις για το είδος και το πλήθος των στοιχείων που ήταν απαραίτητα για την επεξεργασία που θα επακολουθούσε.

Η Ελληνική Βιομηχανία Ηλεκτρομηχανολογικών Α.Ε. ιδρύθηκε το 1969 με την επωνυμία ΜΕΤΑΛΛΟΤΕΧΝΙΚΑ ΕΛΕΚΤΡΑ Α.Ε. με την συμμετοχή της Αμερικάνικης "WESTINGHOUSE ELECTRIC CORPORATION". Με την μορφή αυτή, η εταιρεία παραμένει μέχρι το 1977 οπότε αποχώρησε η WESTINGHOUSE. Τον Σεπτέμβριο του 1981, οι μετοχές της εταιρείας αγοράστηκαν από την ΑΓΕΤ ΗΡΑΚΛΗΣ και την ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ ΤΗΣ ΕΛΛΑΔΟΣ. Τον Ιούνιο του 1982 η εταιρεία μετονομάστηκε σε ΕΛΒΗΜ Α.Ε. (Ελληνική Βιομηχανία Ηλεκτρομηχανολογικών Α.Ε.). Τον Σεπτέμβριο του 1991 το σύνολο των μετοχών

της εταιρείας αγοράσθηκαν από τη Γαλλική MERLIN - GERIN. Η ΕΛΒΗΜ έχει μεγάλες πωλήσεις μετασχηματιστών στην Ελλάδα και στην Μέση Ανατολή.

Αντικείμενο της διπλωματικής αυτής είναι η μελέτη μέσω στατιστικών μεθόδων των παραγόντων που επιδρούν στις απώλειες του πυρήνα ενός μετασχηματιστή, και η μέσω αυτών των μεθόδων :

1. Ρύθμιση και παραμονή σε "κατάσταση ελέγχου" ορισμένων παραγωγικών διαδικασιών.

2. Δημιουργία γραμμικού μοντέλου που διέπει τις απώλειες κενού ενός μετασχηματιστή κατά την τριφασική μέτρηση.

Οι παρούσες καμπύλες που χρησιμοποιούνται από το τμήμα μελετών της ΕΛΒΗΜ για την κατασκευή των μετασχηματιστών έχουν προκύψει κυρίως από την συγκέντρωση στοιχείων που αφορούν την συμπεριφορά μετασχηματιστών ονομαστικής ισχύος 160 KVA. Όμως το εύρος των μετασχηματιστών που κατασκευάζει η εταιρεία είναι από μετασχηματιστές ονομαστικής ισχύος 50 KVA έως μετασχηματιστές ονομαστικής ισχύος 1600 KVA. Επομένως προκύπτει η ανάγκη για την -μέσω στατιστικών μεθόδων- σχεδίαση των ανάλογων καμπυλών και για τα υπόλοιπα ονομαστικά μεγέθη μετασχηματιστών. Εδώ πρέπει να σημειωθεί ότι στο εύρος των μετασχηματιστών που κατασκευάζει η ΕΛΒΗΜ υπάρχουν μετασχηματιστές με τέσσερις, έξι, και οκτώ πυρήνες. Επειδή όμως η συντριπτική πλειοψηφία τους έχει τέσσερις πυρήνες, στην μελέτη που κάναμε περιοριστήκαμε σε τέτοιους μετασχηματιστές αποκλειστικά. Στόχος μας είναι η μελέτη της συσχέτισης αυτών των παραγόντων με τις απώλειες, ώστε να εμπλουτισθεί το παρόν θεωρητικό μοντέλο της εταιρείας που πάνω του στηρίζονται οι μελέτες κατασκευής ενός μετασχηματιστή.

Αρχική προϋπόθεση για την μελέτη που επακολούθησε ήταν η συγκέντρωση στοιχείων, που αφορούσαν τις διάφορες παραγωγικές διαδικασίες για την κατασκευή των μετασχηματιστών. Το 1991 στην ΕΛΒΗΜ παρήχθησαν 3226 μετασχηματιστές όλων των ονομαστικών ισχύων. Από αυτόν τον πληθυσμό αποφασίστηκε να ληφθεί ένα δείγμα της τάξης του 10%, πάνω στο οποίο στηρίχτηκαν οι παραπέρα επεξεργασίες. Οι 320 μετασχηματιστές που ελήφθησαν επιλέχτηκαν με τη βοήθεια μιας γεννήτριας τυχαίων αριθμών και συγκεντρώθηκαν σε μια βάση δεδομένων. Για τον σκοπό αυτό χρησιμοποιήθηκε η DBASE III PLUS. Κατόπιν η βάση δεδομένων του δείγματος εισήχθη στο στατιστικό πακέτο SYSTAT, version 5.03, τ' οποίο είναι το ισχυρότερο που υπάρχει σε επίπεδο προσωπικού

υπολογιστή. Στο πακέτο αυτό έγινε η όλη στατιστική επεξεργασία για την εύρεση του γραμμικού μοντέλου.

Όπως έχει ήδη αναφερθεί ο πρώτος αντικειμενικός σκοπός αυτής της διπλωματικής είναι η ρύθμιση και παραμονή ορισμένων παραγωγικών διαδικασιών σε "κατάσταση ελέγχου". Την περίοδο 1992 θεωρήθηκε πλέον πρόσφορος ο έλεγχος δύο παραγωγικών διαδικασιών. Η πρώτη ήταν η κοπή της μαγνητικής λαμαρίνας η οποία χρησιμοποιείται σαν πρώτη ύλη για τους μετασχηματιστές, και η δεύτερη παραγωγική διαδικασία που προσπαθήσαμε να ελέγξουμε για το αν ήταν σε "κατάσταση ελέγχου", ήταν η πλέξη των πυρήνων.

Όσον αφορά στην κοπή της μαγνητικής λαμαρίνας, οι πυρήνες κατασκευάζονται από πρώτη ύλη μαγνητικού σιδήρου, ο οποίος έρχεται στο εργοστάσιο σε ρολλούς (coils) προδιαγεγραμμένου εύρους, και από διάφορους κατασκευαστές. Το φύλλο του σιδήρου χωρίζεται σε τυποποιημένα στενότερα φύλλα σε μία μηχανή slitter, και κατόπιν τυλίγεται σε ρόλλους. Η μηχανή slitter χρησιμοποιεί κυκλικά μαχαίρια για να κόβει την μαγνητική λαμαρίνα. Όταν τελειώσει η κοπή ορισμένων coils τα μαχαίρια βγαίνουν από την μηχανή και τροχίζονται. Αυτό που μας απασχόλησε ήταν το γρέζι που άφηναν τα μαχαίρια στα φύλλα της μαγνητικής λαμαρίνας. Χρησιμοποιήσαμε διαγράμματα ελέγχου \bar{x} -R, \bar{x} -S, \bar{x} -S² για να ελέγξουμε αν αυτή η παραγωγική διαδικασία ήταν σε "κατάσταση ελέγχου".

Όσον αφορά στην πλέξη των πυρήνων, τα φύλλα του μαγνητικού σιδήρου που αποτελούν κάθε πυρήνα φτάνουν στο τμήμα πλέξης. Εκεί "πλέκονται" πάνω στο μεταλλικό πλαίσιο του μετασχηματιστή. Αυτή η δουλειά γίνεται χειρωνακτικά, από συνεργεία εργατών. Υπάρχουν περιπτώσεις κατά τις οποίες για να γίνει καλή εφαρμογή των πυρήνων στον μετασχηματιστή τα φύλλα του μαγνητικού σιδήρου κτυπιούνται με σφυρί ή ακόμα πετάγονται. Δυστυχώς, ακόμη και μετά την ανάλυση που κάναμε, δεν έγινε εφικτό να πάρουμε αποτελέσματα τα οποία θα μπορούσαν να χρησιμοποιηθούν για την παραπέρα ανάλυσή μας. Καταλήγουμε στο συμπέρασμα πως, αν υπάρχουν παράγοντες που θέτουν αυτήν τη διαδικασία εκτός ελέγχου, αυτοί θα επηρεάζουν τον αριθμό των outliers που πιθανόν θα συναντήσουμε στην ανάλυσή μας.

Το δεύτερο και κυριότερο μέρος της διπλωματικής αφορούσε στην εύρεση ενός γραμμικού μοντέλου, το οποίο θα διέπει τις απώλειες κενού ενός μετασχηματιστή κατά την τριφασική μέτρηση.

Από τη γραμμή παραγωγής πυρήνων μετασχηματιστών της ΕΛΒΗΜ προκύπτει μια ομάδα από 29 μεταβλητές, οι οποίες θα μπορούσαν να χρησιμοποιηθούν σαν ανεξάρτητες μεταβλητές στο γραμμικό μοντέλο. Είναι φανερό ότι, σαν εξαρτημένη μεταβλητή, θα χρησιμοποιηθούν οι απώλειες της τριφασικής μέτρησης για κάθε μετασχηματιστή. Μετά από συζήτηση με το τμήμα μελετών της ΕΛΒΗΜ, επιλέχθηκαν 12 μεταβλητές που θεωρήθηκαν εμπειρικά ως αυτές που επιδρούν περισσότερο στις απώλειες τις τριφασικής μέτρησης. Με αυτήν την ομάδα των 13 μεταβλητών έγινε η πρώτη προσπάθεια να δημιουργηθεί ένα γραμμικό μοντέλο. Όπως ήταν φυσικό, το αρχικό αυτό μοντέλο δεν έδωσε ικανοποιητικά αποτελέσματα. Κατόπιν χρησιμοποιήσαμε διάφορες μεθόδους για την επιλογή του "καλύτερου" υποσυνόλου ανεξαρτήτων μεταβλητών. Έγινε έλεγχος για αλληλεπιδράσεις μεταξύ των μεταβλητών που χρησιμοποιήσαμε, Stepwise Backward Elimination, εύρεση και αφαίρεση των outliers. Η παρουσία outliers στο μοντέλο κατά πάσα πιθανότητα οφείλεται σε στατιστικά ακανόνιστη συμπεριφορά των συνεργείων πλέξης των πυρήνων. Πειραματιστήκαμε με την προσθήκη και άλλων μεταβλητών στο μοντέλο, εκτός από το αρχικό σετ που πρότεινε η εταιρεία, χωρίς όμως καλύτερευση του μοντέλου. Μετά από αυτή την ανάλυση, καταλήξαμε σ' ένα μοντέλο με 8 μεταβλητές, που είχε την καλύτερη συμπεριφορά απ' όλα τα προηγούμενα που είχαμε χρησιμοποιήσει. Υστερα και από τους διάφορους ελέγχους που έγιναν -καλής προσαρμογής του μοντέλου, κανονικότητας στην κατανομή των residuals, ομοιογένειας της διασποράς-, όπως επίσης και διαφόρων διορθώσεων που έπρεπε να γίνουν -τεχνικές Weighted Least Squares, Cochran-Orcutt lag1 και lag2- καταλήξαμε στην παρουσίαση ενός μοντέλου με το τετράγωνο του συντελεστή πολλαπλής συσχέτισης, R^2 (squared multiple correlation coefficient), κοντά στο 1 ($R^2=0.982$), που δεν παραβιάζει τους απαραίτητους ελέγχους υποθέσεων, πάνω στους οποίους στηρίζεται.

Εως σήμερα δεν υπάρχει -στην ΕΛΒΗΜ- οργάνωση στον έλεγχο της ποιότητας των πυρήνων σε σχέση με την ποιότητα της προμηθευόμενης πρώτης ύλης (μαγνητικός σίδηρος). Δεν υπάρχει επίσης αρχείο ιστορικών δεδομένων ελέγχου για την εξαγωγή χρήσιμων συμπερασμάτων που θα δράσουν με καταλυτικό τρόπο στη βελτίωση της μελέτης και κατασκευής των μετασχηματιστών. Ως εκ τούτου υπάρχει δυσκολία στην αναγνώριση της σοβαρότητας της επίδρασης των παραγόντων στην ποιότητα του τελικού προϊόντος. Ένα αξιόπιστο μοντέλο πρόβλεψης των απωλειών της τριφασικής μέτρησης μπορεί να οδηγήσει στην κατάλληλη διαφοροποίηση της μελέτης, ώστε η ποιότητα του τελικού προϊόντος να κινείται μέσα στα προδιαγεγραμμένα πλαίσια με ταυτόχρονη εξοικονόμηση πρώτης ύλης.

ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

2.1 ΕΙΣΑΓΩΓΗ ΣΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Η μεταβλητότητα σε μια παραγωγική διαδικασία οφείλεται στην επίδραση δύο κλάσεων προβλημάτων [Βλέπε Μηλολιδάκης Κ. (1987)] :

α. Συστηματική. Είναι η μεταβλητότητα που οφείλεται στην επίδραση παραγόντων οι οποίοι, μπορούν ν' αναγνωριστούν εύκολα, και τα αίτιά τους να διακριθούν. Για παράδειγμα, μπορεί να υπάρχει διαφορά στα προϊόντα που παράγει η πρωινή και η νυκτερινή βάρδια.

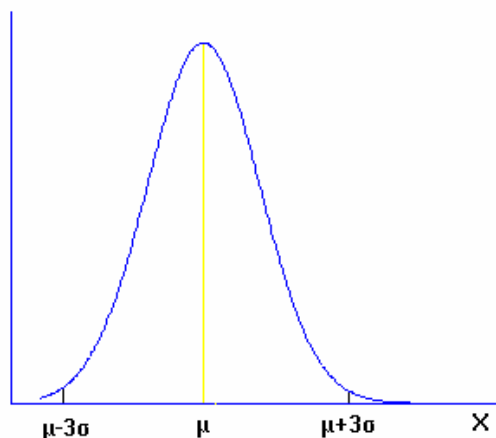
β. Τυχαία. Είναι η μεταβλητότητα που οφείλεται σε παράγοντες που δύσκολα μπορούν να εντοπιστούν, ή είναι οικονομικά ασύμφορο να μελετηθούν. Για παράδειγμα, οι παράγοντες που καθορίζουν αν ένα νόμισμα έρθει κορώνα ή γράμματα.

Όταν οι μεταβολές της παραμέτρου, που μετρά την ποιότητα του προϊόντος, δεν οφείλονται στην επίδραση συστηματικών παραγόντων, αλλά οφείλονται αποκλειστικά σε τυχαίους, θα λέμε ότι η διαδικασία βρίσκεται σε κατάσταση ελέγχου. Βέβαια, μια παραγωγική διαδικασία σπάνια μπορεί να βρεθεί σ' ένα τέτοιο απόλυτο σημείο. Μπορούμε όμως να την οδηγήσουμε, ώστε να το προσεγγίσει σε ικανοποιητικό βαθμό. Στην περίπτωση αυτή μπορούμε, με χρήση στατιστικών μεθόδων, να μελετήσουμε την μεταβλητότητα της ποιότητας, της παραγωγικής διαδικασίας. Δηλαδή είναι δυνατό, να εκτιμήσουμε την κατανομή της τυχαίας μεταβλητής, έστω X , που μετρά την ποιότητα του προϊόντος, και να προχωρήσουμε, ανά τακτά χρονικά διαστήματα, στον στατιστικό έλεγχο της υπόθεσης της τυχαιότητας (δηλαδή της υπόθεσης ότι, η παραγωγική διαδικασία, βρίσκεται σε κατάσταση ελέγχου). Με άλλα λόγια, θα λέμε ότι, μια διαδικασία βρίσκεται κάτω από έλεγχο, όταν παραμένει μέσα στα πλαίσια ορισμένων ορίων ποιότητας, που καθορίζονται από την συγκεκριμένη κατανομή πιθανότητας.

Παραγωγικές διαδικασίες, που βρίσκονται σε κατάσταση ελέγχου, είναι επιθυμητές για τους εξής λόγους:

1. Είναι δυνατή, με μαθηματικά κριτήρια, η ανάλυση και τυποποίηση της παραγωγικής διαδικασίας. Αυτό οδηγεί στη δυνατότητα ενός πολύ αποτελεσματικού ελέγχου.
2. Βοηθάει στη βελτίωση της ποιότητας των προϊόντων. Μπορούμε να εντοπίσουμε μια παραγωγική διαδικασία η οποία είναι από τη φύση της "λειψή". Δηλαδή αν και βρίσκεται σε κατάσταση ελεγχόμενης ποιότητας, παράγει ελαττωματικά προϊόντα. Αυτό οδηγεί στην ανάπτυξη και βελτίωση των τεχνολογιών.
3. Βοηθάει στην ομοιομορφία των προϊόντων. Με αυτές τις μεθόδους, διευκολύνεται ο καθορισμός ορθολογικών κριτηρίων για τις προδιαγραφές που πρέπει να τηρεί το προϊόν.

να εκτιμήσουμε τον μέσο (μ) και τη διασπορά (σ^2) της κανονικής κατανομής. Η συνάρτηση πυκνότητας πιθανότητας θα έχει την παρακάτω μορφή (Σχήμα 1).



Σχήμα 1

Γνωρίζουμε ότι, η μάζα πιθανότητας που περικλείεται στο διάστημα ($\mu-3\sigma, \mu+3\sigma$) θα είναι 0.9974. Οπότε η πιθανότητα να δούμε μια τιμή της τυχαίας μεταβλητής X μεγαλύτερη της $\mu+3\sigma$ είναι 0.0013, δηλαδή περίπου μία στις χίλιες, και ανάλογα για τιμές της X μικρότερες του $\mu-3\sigma$.

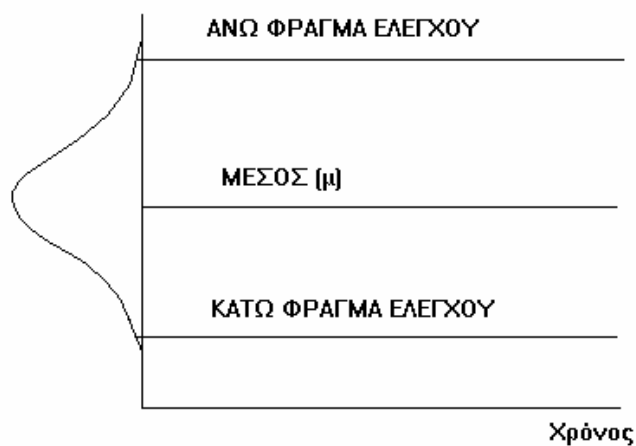
Ας ονομάσουμε τα :

$\mu-3\sigma$ ως Κάτω Φράγμα Ελέγχου (ΚΦΕ) και

$\mu+3\sigma$ ως Ανω Φράγμα Ελέγχου (ΑΦΕ)

Ας υποθέσουμε, τώρα, ότι αρχίζουμε να παίρνουμε δείγματα σε τακτά χρονικά διαστήματα. Τότε, τη διαδικασία αυτή δειγματοληψίας, μπορούμε να την παραστήσουμε με το διάγραμμα του Σχήματος 2. Στον οριζόντιο άξονα μετράμε το χρόνο και στον κατακόρυφο μετράμε τις διάφορες τιμές της μεταβλητής X . Αρα, η συνολική πιθανότητα να φύγει κάποια παρατήρηση έξω από τα όρια σ' ένα από τα δύο φράγματα είναι 0.0026. Αν λοιπόν, μία παρατήρησή μας είναι πέρα από τα όρια, είτε βρισκόμαστε στο -εξαιρετικά σπάνιο- γεγονός να είναι η διαδικασία παραγωγής σε κατάσταση ελέγχου, αλλά εμείς να εντοπίσαμε κάποιο ακρότατο, είτε η υπόθεση ότι βρίσκεται η διαδικασία κάτω από έλεγχο είναι εσφαλμένη κι

επομένως πρέπει ν' αναζητήσουμε το αίτιο του σφάλματος. Κατά την περίπτωση που αποφασίσουμε ότι ισχύει η δεύτερη περίπτωση, η πιθανότητα να κάνουμε λάθος είναι 0.0026.



Σχήμα 2

2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ ΜΕΣΩ ΜΕΤΑΒΛΗΤΩΝ

Ο έλεγχος μέσω μεταβλητών, όπως έχει ήδη αναφερθεί, αφορά διαδικασίες ελέγχου, όπου η μεταβλητή που μετράμε κινείται σε ένα συνεχές πεδίο τιμών (και δεν παίρνει, για παράδειγμα, τιμές 0 ή 1 όπως συμβαίνει στον έλεγχο μέσω χαρακτηριστικών). Προφανώς ο έλεγχος μέσω μεταβλητών είναι πιο περίπλοκος από τον έλεγχο μέσω χαρακτηριστικών. Συχνά όμως είναι πιο αποδοτικός, επειδή μπορεί να κάνει χρήση του συνόλου της πληροφορίας που περιέχεται στο δείγμα.

Η ιδέα στην οποία στηρίζεται είναι απλή. Όταν η παραγωγική διαδικασία βρίσκεται σε κατάσταση ελέγχου, η X θα είναι μια τυχαία μεταβλητή, που οι παράμετροι της κατανομής της μπορούν να εκτιμηθούν. Αρα, σκοπός του Διαγράμματος Ελέγχου (Δ.Ε.) είναι να κάνει την κατάλληλη εκτίμηση, καθώς επίσης και να καθορίσει τα όρια εκείνα, έξω από τα οποία η παραγωγική διαδικασία βρίσκεται εκτός ελέγχου. Τα Δ.Ε. που κυρίως χρησιμοποιήσαμε στον έλεγχο μέσω

μεταβλητών είναι τα \bar{x} -ΔΕ, τα R-ΔΕ, τα P-ΔΕ και τέλος τα S^2 -ΔΕ. Τα \bar{x} -ΔΕ παρακολουθούν τις διακυμάνσεις του δειγματικού μέσου από μια παραγωγική διαδικασία. Τα R-ΔΕ παρακολουθούν τις διακυμάνσεις του πεδίου τιμών του δείγματος (δηλαδή του $R = \max(X_i) - \min(X_i)$, όπου $i=1 \dots n$). Τα S-ΔΕ και S^2 -ΔΕ παρακολουθούν τις διακυμάνσεις της δειγματικής τυπικής απόκλισης S, το πρώτο, και της δειγματικής διασποράς S^2 , το δεύτερο. Όλα τα παραπάνω διαγράμματα προϋποθέτουν ότι η X είναι κανονικά κατανομημένη (σε κατάσταση ελέγχου της παραγωγικής διαδικασίας). Συμβολίζουμε με $\bar{x} \approx \bar{\mu}$ ($\bar{\mu}, \sigma^2$) όπου μ , ο μέσος και σ^2 η διασπορά της τυχαιάς μεταβλητής X.

2.2.1 \bar{x} -ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Θα αρχίσουμε την αναλυτική παρουσίαση των Διαγραμμάτων Ελέγχου από τα \bar{x} -ΔΕ. Από τη θεωρία των πιθανοτήτων και με την προϋπόθεση ότι η διασπορά της τυχαιάς μεταβλητής X υπάρχει ($\Delta(x) < \infty$), γνωρίζουμε ότι ισχύει το Κεντρικό Οριακό Θεώρημα, για δείγμα μεγέθους n :

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \longrightarrow \bar{N}(0, 1)$$

Επομένως, για αρκετά μεγάλο μέγεθος δείγματος, η X ακολουθεί περίπου κανονική κατανομή με μέσο μ και διασπορά σ^2/n . Ακολουθώς θα εξετάσουμε δύο περιπτώσεις :

α. Ο μέσος μ και η διασπορά σ είναι γνωστά. Τότε το \bar{x} -ΔΕ θα έχει την μορφή :

$$\begin{aligned} \bar{X} &= \mu \\ \bar{X} &= \mu + 3 \frac{\sigma}{\sqrt{n}} \\ \bar{X} &= \mu - 3 \frac{\sigma}{\sqrt{n}} \end{aligned}$$

β. Ο μέσος μ και η διασπορά σ είναι άγνωστα. Στην περίπτωση αυτή τα μ και σ θα εκτιμηθούν από κάποιες "καλές εκτιμήτριες" και το \bar{x} -ΔΕ γίνεται :

$$\hat{\mu} = \hat{\mu}$$

$$\hat{\mu} + 3 \frac{\hat{\sigma}}{\sqrt{n}}$$

$$\hat{\mu} - 3 \frac{\hat{\sigma}}{\sqrt{n}}$$

2.2.2 R-ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Από τη θεωρία ξέρουμε ότι ισχύει η ακόλουθη πρόταση : *Εστω η τυχαία μεταβλητή $w=R/\sigma$. Τότε ισχύει, $E[w]=f(n)$ και $D[w]=g(n)$, δηλαδή η μέση τιμή και η διασπορά της w είναι συνάρτηση αποκλειστικά του μεγέθους του δείγματος. Η παραπάνω πρόταση είναι σημαντική, επειδή οδηγεί στην εύρεση εκτιμητριών για την σ και την $\hat{\sigma}_R$, και άρα στην εύρεση των ΑΦΕ και ΚΦΕ.*

Για το R-ΔΕ θα πάρουμε :

$$\bar{R} = \bar{R}$$

$$\bar{R} + 3 \hat{\sigma}_R$$

$$\bar{R} - 3 \hat{\sigma}_R$$

όπου : $\hat{\sigma}_R = \frac{d_3}{d_2} \bar{R}$ (Οι τιμές των d_2, d_3 δίνονται από πίνακες)

2.2.3 S-ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Όπως τα \bar{x} -ΔΕ και R-ΔΕ, είναι δυνατόν να θέλουμε να ρυθμίσουμε την παραγωγική διαδικασία μέχρις ότου να εκτελείται με την επιθυμητή ακρίβεια (δηλαδή με δεδομένη τυπική απόκλιση σ). Στην περίπτωση αυτή το S-ΔΕ χρησιμοποιείται ως μέσο ρύθμισης. Στην περίπτωση που η τυπική απόκλιση είναι γνωστή, ξέρουμε από τη θεωρία ότι έχουμε :

$$c_4 \hat{\sigma}$$

$$(c_4 + 3\sqrt{1-c_4^2})\hat{\sigma}$$

$$(c_4 - 3\sqrt{1-c_4^2})\hat{\sigma}$$

Όπου c_4 είναι σταθερά εξαρτώμενη από το μέγεθος του δείγματος, που βρίσκεται από πίνακες. Στην περίπτωση που η τυπική απόκλιση σ είναι άγνωστη, γνωρίζουμε από τη θεωρία ότι έχουμε :

$$\begin{aligned} \hat{\sigma} &= \bar{S} \\ \text{Ά} &= \bar{S} + 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \\ \text{Έ} &= \bar{S} - 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \end{aligned}$$

2.2.4 S²-ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Τα S-ΔΕ δεν είναι ιδιαίτερα ακριβή και γι' αυτό σε περίπτωση που χρειαστούμε αυξημένη ακρίβεια χρησιμοποιούμε τα S²-ΔΕ (ας μην ξεχνάμε ότι θέλουμε η μάζα πιθανότητας που περικλείεται στο διάστημα $[\mu + 3\sigma, \mu - 3\sigma]$ να είναι 0.998). Για τα S²-ΔΕ θα έχουμε:

$$\begin{aligned} \hat{\sigma} &= \hat{\sigma}^2 \\ \text{Ά} &= \frac{\hat{\sigma}^2}{n-1} \chi_{0.999}^2(n-1) \\ \text{Έ} &= \frac{\hat{\sigma}^2}{n-1} \chi_{0.001}^2(n-1) \end{aligned}$$

όπου : $\hat{\sigma}_R = \bar{S}^2 = \frac{\sum_{j=1}^k S_j^2}{k}$, κ είναι το πλήθος των δειγμάτων

2.2.5 p-ΔΙΑΓΡΑΜΜΑΤΑ ΕΛΕΓΧΟΥ

Όπως γνωρίζουμε το ποσοστό ελαττωματικών στο δείγμα, το \hat{p} , που τώρα θα παίζει τον ρόλο της τυχαίας μεταβλητής X στην κατάστρωση του ΔΕ, ακολουθεί κατά προσέγγιση την κατανομή :

$$\hat{p} \cong N\left(p, \frac{p(1-p)}{n}\right)$$

Αρα, τα όρια του ΔΕ θα παρθούν :

$$\begin{aligned}\hat{\text{Ö}}\text{Ä} &= p + 3\sqrt{\frac{p(1-p)}{n}} \\ \hat{\text{E}}\text{Ö}\text{Ä} &= p - 3\sqrt{\frac{p(1-p)}{n}}\end{aligned}$$

Επομένως, αν το μέγεθος του δείγματος μεταβάλλεται, και τα ΑΦΕ και ΚΦΕ θα μεταβάλλονται. Βέβαια, υπάρχουν τρόποι για να ξεπεράσουμε αυτό το πρόβλημα.

2.3 ΕΦΑΡΜΟΓΗ ΤΩΝ Δ.Ε. ΣΤΗ ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΚΟΠΗΣ

Προχωράμε, τώρα στην εφαρμογή των ΔΕ στην παραγωγική διαδικασία κοπής της μαγνητικής λαμαρίνας. Η τυχαία μεταβλητή X που μετράμε, είναι το γρέζι της λαμαρίνας, δηλαδή η διαφορά του πάχους μεταξύ του σημείου κοπής και ενός τυχαίου σημείου στην μέση της λαμαρίνας μετρημένο σε mm. Ισχύουν οι προϋποθέσεις κανονικότητας των \bar{X} & R ΔΕ, δηλαδή κάτω από την υπόθεση H_0 (κατάσταση ελέγχου), η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με μέσο μ και διασπορά σ^2 , κι επίσης παρατηρήσαμε ότι οι μετρήσεις ήταν αξιόπιστα οργανωμένες.

Οι μετρήσεις έγιναν σε ορισμένες ημερομηνίες και σε τυχαία επιλεγμένα coils, ως εξής : Από την ίδια πάντα πλευρά της λαμαρίνας (δεξιά), πάρθηκε από μία μέτρηση σε τυχαίο σημείο στην αρχή, στην μέση και στο τέλος του coil, τόσο της λαμαρίνας δίπλα στο σημείο κοπής όσο και της λαμαρίνας στο σημείο τομής (η διαφορά τους είναι το γρέζι). Καταγράφηκαν η ημερομηνία, η προέλευση της λαμαρίνας και το είδος, το νούμερο του coil και οι μετρήσεις (σε mm). Τα στοιχεία αφορούσαν συνολικά 220 μετρήσεις, από τις 11 Μαΐου 1992 έως τις 7 Δεκεμβρίου 1992. Επίσης, καταγράφηκε η σειρά κοπής των coils από την αρχή της ημέρας.

Τα μαχαίρια της μηχανής είναι κυκλικά κι έχουν δύο πλευρές. Επίσης τροχίζονται καθημερινά. Η διαδικασία που τηρείται είναι η εξής : Μετά την κοπή και του 3^{ου} coil από την αρχή της ημέρας, βγαίνουν τα μαχαίρια από την μηχανή και τοποθετούνται ξανά από την άλλη πλευρά τους. Μετά την κοπή και του 6^{ου} coil, τα μαχαίρια ξαναβγαίνουν, τροχίζονται και τοποθετούνται στην μηχανή από την πρώτη

πλευρά. Επομένως, η αντιστοιχία μεταξύ των coils είναι 1-4, 2-5 και 3-6. Υπήρχαν μόνο δύο μετρήσεις για τα coils 7 και 8 οπότε αγνοήθηκαν.

Επομένως, οι εξωτερικοί παράγοντες της παραγωγικής διαδικασίας διαφέρουν ανάλογα με την σειρά κοπής των coils κι έτσι τα διαθέσιμα στοιχεία ομαδοποιήθηκαν σε έξη ομάδες. Οι ομάδες αυτές είναι :

ΟΜΑΔΑ	ΠΛΗΘΟΣ ΜΕΤΡΗΣΕΩΝ	ΔΙΑΡΚΕΙΑ ΚΑΤΑΓΡΑΦΗΣ
1	46	11/5 - 7/12
2	49	11/5 - 7/12
3	34	8/5 - 30/11
4	33	12/5 - 25/11
5	30	12/5 - 27/11
6	27	14/5 - 19/11

Για την κάθε ομάδα, η ανάλυση έγινε σύμφωνα με τα \bar{x} & R Διαγράμματα Ελέγχου. Τα ιδιαίτερα συμπεράσματα επιβεβαιώθηκαν με \bar{x} & S ΔΕ και με \bar{x} & S² ΔΕ που είναι πιο ακριβή κι ευαίσθητα από τα \bar{x} & R.

Θα περιγράψουμε την μέθοδο της ανάλυσης, χρησιμοποιώντας σαν παράδειγμα την ομάδα 1 [βλέπε Παράρτημα Α]. Ομαδοποιήθηκαν τα πρώτα 16 στοιχεία, σύμφωνα με τον συνολικό αριθμό παρατηρήσεων. Εκτός ορίων βρέθηκαν η 9η, 10η, και 14η παρατήρηση που αντιστοιχούσαν στις 4 Ιουνίου, 8 Ιουνίου και 16 Ιουνίου. Θεωρούμε ότι τις ημέρες αυτές ή και ανάμεσά τους συνέβησαν γεγονότα που αναίρεσαν την τυχαιότητα κι έθεσαν την παραγωγική διαδικασία εκτός ελέγχου. Ομως, οι αιτίες που προκάλεσαν την αναίρεση της τυχαιότητας δεν μονιμοποιήθηκαν. Αυτό το επιβεβαιώνει και η στατιστική θεωρία των διαδρομών. Στην συνέχεια προστέθηκαν τα επόμενα 15 στοιχεία. Παρατηρήσαμε ότι σταδιακά εμφανίστηκε μια μετατόπιση της παραγωγικής διαδικασίας προς το χειρότερο (αύξηση του γρεζιού), που μας ανάγκασε να ενιαιοποιήσουμε τις δύο ομάδες σε μία που περιείχε 31 συνολικά παρατηρήσεις. Το νέο Διάγραμμα Ελέγχου έκανε δεκτή την μη αλλαγή των δεδομένων της παραγωγικής διαδικασίας. Εξαίρεση αποτέλεσε η 23η παρατήρηση, που αντιστοιχούσε στις 16 Ιουλίου, η οποία ήταν σαφώς outlier

όσον αφορά τόσο στον έλεγχο της τυπικής απόκλισης σ όσο και στον έλεγχο του μέσου μ . Έτσι, η 23η παρατήρηση αφαιρέθηκε από την ομάδα των 31 παρατηρήσεων, υπολογίστηκαν ξανά οι μέσοι μ και τα ΑΦΕ και ΚΦΕ του Διαγράμματος Ελέγχου. Επιπλέον, προσθέσαμε και την ομάδα των υπόλοιπων 15 μετρήσεων. Τότε εμφανίστηκε μια σοβαρή χειροτέρευση της ποιότητας κοπής της μαγνητικής λαμαρίνας, που δεν αναιρέθηκε ακόμα και όταν ενιαιοποιήσαμε και τις 46 παρατηρήσεις. Η τυχαιότητα έπαψε να υπάρχει, κι εμφανίστηκε νέος μόνιμος εξωτερικός παράγοντας που δεν υπήρχε στην αρχή των μετρήσεων. Μέσω της αφαίρεσης μετρήσεων από την αρχή του ΔΕ και επαναπροσδιορισμού των ορίων, βρέθηκε σαν κρίσιμη ημερομηνία η 7 Σεπτεμβρίου. Την άποψη αυτή υποστήριξε επίσης, η στατιστικά σημαντική μακρινή διαδρομή σημείων κάτω από τον μέσο μ στο R-ΔΕ των 46 παρατηρήσεων (στις πρώτες 16-17 παρατηρήσεις), καθώς και η αντίστοιχα μακριά διαδρομή στο S^2 -ΔΕ. Η παρατήρηση αυτή σημαίνει ότι, εκτός από την αύξηση του μέσου γρεζιού, χειροτέρεψε και η ακρίβεια στην ρύθμιση των μαχαιριών της μηχανής, κάτι που το τρόχισμα τους δεν μπορούσε να διορθώσει.

Αντίστοιχη ανάλυση έγινε και στις άλλες 5 ομάδες στοιχείων. Συνοπτικά έχουμε :

α. Ομάδα 2. Βρέθηκαν 6 outliers που αντιστοιχούσαν στις 11 Μαΐου, 17 Ιουνίου, 1 Ιουλίου, 2 Ιουλίου, 7 Ιουλίου και 22 Σεπτεμβρίου. Οι outliers που αντιστοιχούσαν στις 1 και 2 Ιουλίου, οφείλονταν στον ίδιο λόγο. Από την ανάλυση που έγινε -μέσω των ΔΕ- βρέθηκε σαν κρίσιμη ημερομηνία, εμφάνισης εξωτερικού μόνιμου παράγοντα λόγω του οποίου έπαψε να υπάρχει τυχαιότητα, αυτή της 4ης Σεπτεμβρίου. Παρατηρήσαμε, τέλος, ότι το πρόβλημα που υπήρχε ήταν όμοιο με αυτό της ομάδας 1.

β. Ομάδα 3. Βρέθηκαν 3 outliers που αντιστοιχούσαν στις 17 Ιουνίου, 1 Ιουλίου και 2 Ιουλίου. Από την ανάλυση των ΔΕ βρέθηκε σαν κρίσιμη ημερομηνία η 7 Σεπτεμβρίου. Επίσης, το πρόβλημα αυτής της ομάδας είναι όμοιο με αυτό της ομάδας 1. Σε αυτό το σημείο πρέπει να παρατηρήσουμε ότι, ορισμένοι outliers επαναλαμβάνονται και στις τρεις πρώτες ομάδες. Αυτό το γεγονός επιβεβαίωσε την άποψη ότι σε ορισμένες ημερομηνίες (16-17 Ιουνίου, 1-2 Ιουλίου), το κόψιμο της λαμαρίνας ήταν προβληματικό.

γ. Ομάδες 4,5,6. Εκτός από ορισμένους outliers, και οι τρεις ομάδες δεν εμφανίζουν τάση μονιμοποίησης εξωτερικών παραγόντων που να οδηγούν σε χειροτέρευση της ποιότητας κοπής. Πιο συγκεκριμένα, η ομάδα 4 δεν έχει outliers, η

ομάδα 5 παρουσιάζει outliers στις 18 και 23 Ιουνίου, και τέλος, η ομάδα 6 παρουσιάζει outliers στις 9 Ιουνίου.

Τα συμπεράσματα που προκύπτουν από την εφαρμογή των ΔΕ στη διαδικασία κοπής της μαγνητικής λαμαρίνας είναι ότι, τα ΔΕ που έγιναν στις ομάδες 1, 2 και 3 δείχνουν παραγωγική διαδικασία εκτός καταστάσεως ελέγχου. Δηλαδή, κατά τη διάρκεια καταγραφής των δεδομένων παρουσιάστηκε κάποιος σταθερός εξωτερικός παράγοντας που δεν υπήρχε αρχικά. Το αποτέλεσμα της εμφάνισης αυτού του παράγοντα, ήταν η χειροτέρευση της ποιότητας κοπής (αύξηση γρεζιού). Η παρουσία του είναι στατιστικά εξακριβωμένη. Παρουσιάστηκε τη δεύτερη εβδομάδα του Σεπτεμβρίου (7-14 Σεπτεμβρίου) και μονιμοποιείται από αρχές Οκτωβρίου. Εγινε στατιστικός έλεγχος, αν ο παράγοντας αυτός οφείλεται σε διαφοροποίηση του προμηθευτή της μαγνητικής λαμαρίνας (USINOR ή ARMCO), δηλαδή, ελέγχθηκε αν τα σημεία που ήταν εκτός των ορίων προέρχονταν από ένα συγκεκριμένο προμηθευτή και απορρίφθηκε η θεωρία αυτή. Επίσης, απορρίφθηκε η θεωρία ότι επρόκειτο για παροδικό φαινόμενο που εμφανίστηκε μετά τις διακοπές του καλοκαιριού, διότι είχαμε σημεία εκτός των ορίων και πριν τις διακοπές, του καλοκαιριού. Αν κάτι συνέβει μέσα στις διακοπές αυτό έχει πλέον μονιμοποιηθεί. Τέλος, έγινε έλεγχος αν ο παράγοντας αυτός συνδέεται με τις Δευτέρες, οπότε εμφανίζεται φόρτος εργασίας, δηλαδή ελέγξαμε αν οι outliers αντιστοιχούσαν σε Δευτέρες αλλά και η θεωρία αυτή απορρίφθηκε από τον στατιστικό έλεγχο. Το πιο ενδιαφέρον είναι ότι, τα Διαγράμματα Ελέγχου που έγιναν για τις ομάδες 4, 5 και 6 δείχνουν την παραγωγική διαδικασία σε κατάσταση ελέγχου. Επομένως οδηγηθήκαμε στο συμπέρασμα ότι, κάτι είχε συμβεί στην μία πλευρά των μαχαριών που χρησιμοποιούνται για την κοπή των coils.

Το γενικό συμπέρασμα που προέκυψε από την συνολική ανάλυση της διαδικασίας κοπής είναι ότι, η παραγωγική διαδικασία είναι εύκολο να τεθεί σε κατάσταση ελέγχου. Επομένως, είμαστε σε θέση να θεωρήσουμε ότι η διαδικασία κοπής της μαγνητικής λαμαρίνας σε μια τέτοια περίπτωση δε θα επηρεάζει το μοντέλο πρόβλεψης των τριφασικών απωλειών που πρόκειται να κατασκευάσουμε. Βέβαια, το μοντέλο αυτό στην παρούσα εργασία στηρίζεται σε στοιχεία συλλεγμένα πριν να γίνει η προσπάθεια ελέγχου της κοπής. Παρ' όλα αυτά, εκτιμούμε σαν οριακή την επίδραση της διαδικασίας αυτής στο μοντέλο, αφού είναι "σχεδόν" σε κατάσταση ελέγχου.

2.4 ΕΦΑΡΜΟΓΗ ΤΩΝ Δ.Ε. ΣΤΗ ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΠΛΕΞΗΣ

Όπως ξέρουμε, κατά τη διαδικασία της πλέξης, οι πυρήνες τοποθετούνται στο μεταλλικό πλαίσιο του μετασχηματιστή από συνεργεία εργατών. Ζητήθηκε λοιπόν από το κάθε συνεργείο να συμπληρώνει στοιχεία σε ένα έντυπο, που αφορούσαν την πλέξη κάθε μετασχηματιστή. Οι μετρήσεις που είχαμε, ήταν από 1/7/1992 έως και 9/11/1992 και αφορούσαν 1537 μετασχηματιστές όλων των ονομαστικών ισχύων. Τα στοιχεία κάθε χαρτιού δεδομένων περιέχουν τα εξής : ημερομηνία, πάγκος, αριθμός μετασχηματιστή, εντολή, ώρα ολοκλήρωσης καθώς και 5 διαβαθμίσεις δυσκολίας στο πλέξιμο των πυρήνων [εύκολη προσαρμογή, δύσκολη προσαρμογή, πιάστηκε ο πυρήνας, χτυπήθηκε με σφυρί, πετάχτηκαν φύλλα μαγνητικής λαμαρίνας]. Από τα στοιχεία ήταν προφανές ότι δεν έγινε κατανοητό το σύστημα διαβάθμισης της δυσκολίας στο πλέξιμο και ως εκ τούτου δημιουργήθηκαν αμφιβολίες για την ακρίβειά τους.

Λόγω του προβλήματος στην καταγραφή των δεδομένων, ήταν δύσκολη η επιστημονική επεξεργασία τους. Προσπαθήσαμε να θεωρήσουμε τη διαδικασία της πλέξης σαν παραγωγική διαδικασία που παράγει "καλά" ή "ελαττωματικά" αντικείμενα κι επομένως θα ήταν δυνατό να τεθεί σε κατάσταση ελέγχου, μέσω p-Διαγραμμάτων Ελέγχου. Θεωρήσαμε μεταβλητό μέγεθος δείγματος με μονάδα δειγματοληψίας την ημέρα. Τέλος, θεωρήσαμε σαν "ελαττωματικό" (για τους σκοπούς της ανάλυσης), έναν μετασχηματιστή, από τους πυρήνες του οποίου αφαιρέθηκαν φύλλα μαγνητικής λαμαρίνας.

Επειδή εμφανίστηκε απροσδόκητα υψηλός αριθμός "ελαττωματικών" (από τους 1537 μετασχηματιστές που καταγράφηκαν, πετάχτηκε υλικό από τους 817, δηλαδή ποσοστό 53.15%), διατυπώσαμε την υπόθεση ότι το ποσοστό των ελαττωματικών εξαρτάται κυρίως από την μελέτη και συνακόλουθα από τον αριθμό της εντολής. Επομένως, θεωρήσαμε κάθε μελέτη σαν παραγωγική διαδικασία κι επιδιώξαμε να ελέγξουμε αν αυτή βρίσκεται σε κατάσταση ελέγχου. Την υπόθεσή μας αυτή θελήσαμε να ελέγξουμε, μέσω ελέγχου ποσοστών.

Η προσπάθεια χρήσης ενός p-ΔΕ ανά εντολή απέτυχε, για τα δεδομένα στοιχείά*. Η χρήση ελέγχου ποσοστών επιβεβαίωσε την στενή αλληλεξάρτηση αριθμού εντολής και ποσοστού μετασχηματιστών που κατατάσσονται στους "ελαττωματικούς". Έτσι έχουμε τα εξής ποσοστά ελαττωματικών p :

*
$$P_j = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i} \cdot \frac{1}{p_{i+1}}$$

$$P_j = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i} \cdot \frac{1}{p_{i+1}} = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i \cdot p_{i+1}}$$

$$P_j = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i \cdot p_{i+1}} = \frac{1}{n} \sum_{i=1}^n \frac{1}{p_i \cdot p_{i+1}}$$

ΕΝΤΟΛΗ	p	ΠΛΗΘΟΣ ΣΤΟΙΧΕΙΩΝ
1036	.20	92
1038	.38	78
1039	.44	107
1041	.35	42
1049	.10	41
1050	.91	56
1051	.39	31
1250	.80	57
1517	1	21

Αν δεχτούμε τη βάσιμα αμφισβητούμενη υπόθεση ότι τα παραπάνω δεδομένα αντιστοιχούν σε δειγματοληψία που έγινε με ορθό τρόπο, τότε για τη διαφορά του ποσοστού ελαττωματικών έχουμε :

ΕΝΤΟΛΗ 1 ΔΙΑΦΟΡΑ	ΕΝΤΟΛΗ 2	z-value	ΕΠΙΠΕΔΟ ΕΛΕΓΧΟΥ α	
1036	1038	2.59	0.01	ΝΑΙ
1038	1038	0.82	0.01	ΟΧΙ
1039	1041	1.00	0.01	ΟΧΙ
1041	1049	2.72	0.01	ΝΑΙ
1041	1051	0.35	0.01	ΟΧΙ

1050	1250	1.36	0.01	OXI
1036	1049	1.42	0.01	OXI
1250	1517	2.21	0.01	OXI

Δηλαδή, ουσιαστικά έχουμε τρεις ομάδες εντολών.

α. 1036, 1049

β. 1038, 1039, 1041, 1051

γ. 1050, 1250, 1517

Μέσα στην κάθε ομάδα δεν υπάρχει διαφοροποίηση, ενώ υπάρχει διαφοροποίηση ανάμεσα στις ομάδες.

Το συμπέρασμα που προέκυψε από την παραπάνω ανάλυση ήταν ότι -με βάση τα δεδομένα που είχαμε- δεν ήταν δυνατό, με στατιστικές μεθόδους, να καταλήξουμε αν η διαδικασία πλέξης των πυρήνων ήταν, ή όχι, σε κατάσταση ελέγχου. Αυτό για το οποίο είμασταν βέβαιοι ήταν ότι, αν η διαδικασία δεν ήταν σε κατάσταση ελέγχου, τότε σαν άμεση συνέπεια θα είχαμε την εμφάνιση υπερβολικού αριθμού outliers κατά την μελέτη εύρεσης του μοντέλου.

ΓΡΑΜΜΗ ΠΑΡΑΓΩΓΗΣ ΠΥΡΗΝΩΝ ΜΕΤΑΣΧΗΜΑΤΙΣΤΩΝ

3.1 ΠΕΡΙΓΡΑΦΗ

Η εταιρεία δέχεται μία παραγγελία για μετασχηματιστές συγκεκριμένης ονομαστικής ισχύος. Όπως έχει ήδη αναφερθεί, εμείς ασχοληθήκαμε με μετασχηματιστές που είχαν τέσσερις πυρήνες. Οι μετασχηματιστές αυτοί έχουν δύο μικρούς και δύο μεγάλους πυρήνες. Η σειρά με την οποία μπαίνουν οι πυρήνες στον μετασχηματιστή είναι -βλέποντας τον μετασχηματιστή από αριστερά προς τα δεξιά- μικρός, μεγάλος, μεγάλος, μικρός. Η εταιρεία έχει κωδικοποιήσει τους πυρήνες με αριθμούς. Έτσι η παραπάνω σειρά των πυρήνων περιγράφεται σαν 11, 12, 13, 14. Ανάλογα με τη ζητούμενη ονομαστική ισχύ της παραγγελίας, καθώς και άλλες προδιαγραφές του καταναλωτή (κυρίως ΔΕΗ), που αφορούν τις απώλειες, την τάση βραχυκυκλώσεως κ.λ.π., το τμήμα μελετών της εταιρείας εκπονεί -στηριζόμενο βασικά στις καμπύλες που διαθέτει, οι οποίες συνδέουν τις τριφασικές απώλειες με την ένταση του μαγνητικού πεδίου (δηλαδή ουσιαστικά την ονομαστική ισχύ)- μία μελέτη, η οποία παρέχει τα απαιτούμενα στοιχεία για τους πυρήνες του κάθε μετασχηματιστή. Δηλαδή από την μελέτη δίνονται :

α. Οι γεωμετρικές διαστάσεις των πυρήνων (ύψος, μήκος, πλάτος, πάχος κ.ά.)

β. Το βάρος του μικρού και του μεγάλου πυρήνα.

γ. Η ένταση του μαγνητικού πεδίου (μετρημένη σε Gauss), η οποία αντιστοιχεί στην ονομαστική ισχύ του μετασχηματιστή.

Επίσης, στην μελέτη περιέχονται συμπληρωματικά ηλεκτρικά χαρακτηριστικά των πυρήνων, όπως επίσης και η ονομαστική ισχύς (σε KVA), ο αριθμός της εντολής, και τέλος ο αριθμός των μετασχηματιστών που αφορούν την παραγγελία. Η εφαρμογή της μελέτης αρχίζει στο τμήμα των πυρήνων. Οι πυρήνες κατασκευάζονται από μαγνητικό σίδηρο. Ο μαγνητικός σίδηρος φτάνει στο εργοστάσιο υπό μορφή ρολλών, κυρίως από τρεις διαφορετικούς κατασκευαστές (ARMCO, USINOR, THYSSEN). Το φύλλο του σιδήρου χωρίζεται σε στενότερα φύλλα σε μία μηχανή slitter, και κατόπιν τυλίγεται σε ρόλλους, των οποίων οι διαστάσεις εξαρτώνται από την μελέτη, κι έτσι δημιουργούνται οι μικροί και οι μεγάλοι πυρήνες. Υστερα οι πυρήνες περνάνε τη διαδικασία της ανόπτησης, κατά την οποία τοποθετούνται σε φούρνο, ούτως ώστε να επανακτήσει -ο σίδηρος- τις μαγνητικές ιδιότητες, οι οποίες καταστράφηκαν κατά τη διαδικασία παραγωγής των πυρήνων. Τέλος, ο υπεύθυνος του τμήματος διενεργεί μία μέτρηση για τις απώλειες κάθε πυρήνα που ονομάζεται μονοφασική. Τ' αποτελέσματα αυτών των μετρήσεων καταγράφονται σε έντυπο, το οποίο συνοδεύει τους πυρήνες. Στην συνέχεια, οι πυρήνες πηγαίνουν στο τμήμα πλέξης. Εκεί λύνονται, ώστε να τοποθετηθούν τα πηνία, και στην συνέχεια ξαναπλέκονται και τοποθετούνται στο πλαίσιο του μετασχηματιστή. Η εργασία αυτή γίνεται από συνεργεία εργατών, χειρωνακτικά. Σε αρκετές περιπτώσεις, είναι δύσκολη η προσαρμογή όλων των φύλλων του σιδήρου στο χώρο που έχει προβλεφθεί, λόγω της επίδρασης διαφόρων παραγόντων. Αυτό έχει σαν αποτέλεσμα να κτυπιούνται με σφυρί τα -τελευταία κυρίως- φύλλα του σιδήρου, ή ακόμα και να πετιούνται. Κατόπιν, ο ολοκληρωμένος μετασχηματιστής πηγαίνει στο τμήμα ποιοτικού ελέγχου. Στο τμήμα αυτό γίνεται η τελική μέτρηση του μετασχηματιστή, η οποία ονομάζεται τριφασική. Η μέτρηση αυτή γίνεται σε ηλεκτρικό πεδίο καθορισμένης έντασης και το μετρούμενο μέγεθος είναι οι απώλειες του μετασχηματιστή σε Watt. Στόχος είναι να ελεγχθεί αν οι απώλειες παραμένουν στα προδιαγεγραμμένα όρια του πελάτη (εγγυημένες απώλειες). Το τμήμα ποιοτικού ελέγχου φτιάχνει ένα φάκελο για την παραγγελία, ο οποίος αναγράφει:

α. Τον αριθμό της εντολής των μετασχηματιστών

β. Τον συνολικό αριθμό των μετασχηματιστών

- γ. Την ονομαστική ισχύ
- δ. Τον αριθμό της μελέτης, βάσει της οποίας κατασκευάστηκαν οι μετασχηματιστές
- ε. Την ένταση του ηλεκτρικού πεδίου στο οποίο έγινε η τριφασική μέτρηση
- στ. Τις απώλειες κάθε μετασχηματιστή
- ζ. Τις εγγυημένες απώλειες
- η. Τη διαφορά μεταξύ πραγματικών απωλειών και εγγυημένων
- θ. Την ημερομηνία που έγινε ο ποιοτικός έλεγχος

3.2 ΜΕΤΑΒΛΗΤΕΣ

Μετά από την περιγραφή της γραμμής παραγωγής των πυρήνων, θα πρέπει να δούμε ποιές μεταβλητές προκύπτουν από αυτή τη γραμμή, και ποιές από αυτές μπορούμε να χρησιμοποιήσουμε σαν ανεξάρτητες μεταβλητές στο μοντέλο μας.

Όπως έχουμε ήδη αναφέρει, η ΕΛΒΗΜ προμηθεύεται το μαγνητικό σίδηρο από τρεις κύριους προμηθευτές. Την Αμερικάνικη ARMCO, τη Γαλλική USINOR, και τη Γερμανική THYSSEN. Συχνά, αλλά όχι πάντα, ο προμηθευτής δίνει τις μέσες απώλειες για κάθε coil, μετρημένες σε w/kg. Επειδή στο έντυπο που αναφέρονται οι μονοφασικές μετρήσεις, αναγράφεται και ο αριθμός του ρόλλου (coil) από τ' οποίο κατασκευάστηκε κάθε πυρήνας, προκύπτουν οι εξής δύο ομάδες μεταβλητών :

NOCOIL11, που είναι ο αριθμός του coil από τ' οποίο κατασκευάστηκε ο πυρήνας 11

NOCOIL12, που είναι ο αριθμός του coil από τ' οποίο κατασκευάστηκε ο πυρήνας 12

NOCOIL13, που είναι ο αριθμός του coil από τ' οποίο κατασκευάστηκε ο πυρήνας 13

NOCOIL14, που είναι ο αριθμός του coil από τ' οποίο κατασκευάστηκε ο πυρήνας 14

και

IRONLOSS(1), που είναι οι απώλειες της μαγνητικής λαμαρίνας από την οποία κατασκευάστηκε ο πυρήνας 11

IRONLOSS(2), που είναι οι απώλειες της μαγνητικής λαμαρίνας από την οποία κατασκευάστηκε ο πυρήνας 12

IRONLOSS(3), που είναι οι απώλειες της μαγνητικής λαμαρίνας από την οποία κατασκευάστηκε ο πυρήνας 13

IRONLOSS(4), που είναι οι απώλειες της μαγνητικής λαμαρίνας από την οποία κατασκευάστηκε ο πυρήνας 14

Επίσης μπορούμε να χρησιμοποιήσουμε και τους προμηθευτές σαν ανεξάρτητη μεταβλητή του μοντέλου. Ονομάσαμε την μεταβλητή αυτή SUPPLIER και της δώσαμε τις τιμές :

- 1 αν ο προμηθευτής ήταν η USINOR
- 2 αν ο προμηθευτής ήταν η ARMCO
- 3 αν ο προμηθευτής ήταν η THYSSEN

Η μεταβλητή αυτή είναι κατηγορική (ποιοτική). Αυτό σημαίνει ότι, με τις τιμές που παίρνει, χωρίζει το δειγματικό χώρο σε κατηγορίες (επίπεδα). Βέβαια η ανάλυση γίνεται πάνω στο σύνολο του δείγματος και όχι στο δείγμα κάθε επιπέδου.

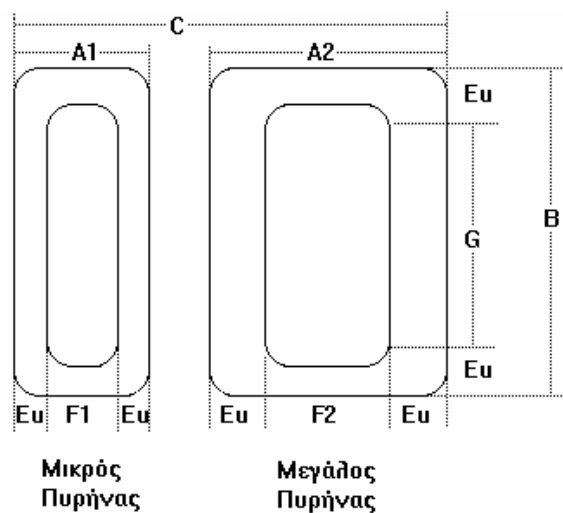
Μια άλλη ομάδα ανεξάρτητων μεταβλητών μπορεί να προέλθει από την μελέτη των μετασχηματιστών. Οι μεταβλητές αυτές είναι τα γεωμετρικά χαρακτηριστικά του κάθε πυρήνα, τα βάρη τους, και η ένταση του μαγνητικού πεδίου. Σε αυτό το σημείο αντιμετωπίσαμε το εξής πρόβλημα. Οι τιμές των διαφόρων χαρακτηριστικών του κάθε πυρήνα, που προβλέπονται από την μελέτη, δεν επαληθεύονται. Δηλαδή, μετά την κατασκευή των πυρήνων δεν παίρνονται μετρήσεις των γεωμετρικών χαρακτηριστικών και του βάρους, που οπωσδήποτε παρουσιάζουν αποκλίσεις από πυρήνα σε πυρήνα. Αρα, εμείς είμασταν αναγκασμένοι να χρησιμοποιήσουμε τα στοιχεία της μελέτης χωρίς να ξέρουμε κατά πόσον αυτά ανταποκρίνονται στην πραγματικότητα. Πρέπει επίσης να σημειώσουμε ότι όσον αφορά στα γεωμετρικά στοιχεία των πυρήνων οι τιμές των μεταβλητών είναι οι ίδιες για τους μικρούς και τους μεγάλους πυρήνες. Έτσι, από τα γεωμετρικά στοιχεία των πυρήνων, έχουμε τις μεταβλητές (Σχήμα 3) :

Eu, που είναι το πάχος του πυρήνα

F1, που είναι το μήκος της μικρής πλευράς του πυρήνα

D1, που είναι το πλάτος του πυρήνα

G, που είναι το μήκος της μεγάλης πλευράς του πυρήνα



Σχήμα 3

Το βάρος του μικρού και του μεγάλου πυρήνα χρησιμοποιήθηκαν ως :

BAROS11, που είναι το βάρος του μικρού πυρήνα (πρώτος από αριστερά)

BAROS12, που είναι το βάρος του μεγάλου πυρήνα (δεύτερος από αριστερά)

Δε χρειάστηκε να χρησιμοποιήσουμε τα βάρη των άλλων δύο πυρήνων διότι το βάρος του πυρήνα 11 είναι ίσο με το βάρος του πυρήνα 14, και αντίστοιχα το βάρος του πυρήνα 12 είναι ίσο με το βάρος του πυρήνα 13.

Η ένταση του μαγνητικού πεδίου στο οποίο γίνεται η μονοφασική μέτρηση χρησιμοποιήθηκε σαν INTENSIT (από την σύντμηση της λέξεως INTENSITY που σημαίνει ένταση).

Μία άλλη ομάδα μετρήσεων, που μπορούν να εισαχθούν σαν ανεξάρτητες μεταβλητές στο μοντέλο, είναι οι μονοφασικές μετρήσεις που γίνονται για κάθε πυρήνα. Στην ΕΛΒΗΜ, κατά τις μονοφασικές μετρήσεις λαμβάνεται μία μέτρηση που αντιπροσωπεύει τις απώλειες του πυρήνα. Όμως, δεν καταγράφεται αυτή η τιμή, αλλά υπάρχουν έξι διαβαθμίσεις απωλειών και ο κάθε πυρήνας, ανάλογα με τις απώλειές του κατατάσσεται σε μία από αυτές τις κατηγορίες. Οι κατηγορίες αυτές είναι : A, B+, B, C+, C, D όπου, ο A είναι ο πυρήνας με τις μικρότερες απώλειες και ο D είναι ο πυρήνας με τις περισσότερες απώλειες. Οι απώλειες των μονοφασικών μετρήσεων χρησιμοποιήθηκαν ως εξής :

Η μεταβλητή LOSSES11 δίνει τις απώλειες του πυρήνα 11 με τιμές :

- 1 αν ο πυρήνας είναι A
- 2 αν ο πυρήνας είναι B+
- 3 αν ο πυρήνας είναι B
- 4 αν ο πυρήνας είναι C+
- 5 αν ο πυρήνας είναι C
- 6 αν ο πυρήνας είναι D

Ανάλογα ορίζουμε τις LOSSES12, LOSSES13, και LOSSES14. Παρατηρούμε ότι εξ αιτίας του τρόπου που καταγράφονται τ' αποτελέσματα των μονοφασικών μετρήσεων στην εταιρεία, είμαστε αναγκασμένοι να εισάγουμε τις ανεξάρτητες μεταβλητές LOSSES11, LOSSES12, LOSSES13, LOSSES14 σαν κατηγορικές.

Από τη διαδικασία της πλέξης των πυρήνων δεν μπορούμε να πάρουμε - διότι δεν υπάρχουν- κάποιες μετρήσεις, οι οποίες να χρησιμοποιηθούν σαν μεταβλητές στο μοντέλο.

Όσον αφορά τις τριφασικές μετρήσεις, που παίρνονται από το τμήμα ποιοτικού ελέγχου, μπορούν να χρησιμοποιηθούν οι μεταβλητές :

NODESIGN, που είναι ο αριθμός της μελέτης βάσει της οποίας κατασκευάστηκαν οι μετασχηματιστές

ORDER, που είναι ο αριθμός της εντολής κάθε παρτίδας μετασχηματιστών

NOTRANS, που είναι ο αριθμός του μετασχηματιστή μέσα στην παρτίδα

ONISXYS, που είναι η ονομαστική ισχύ των μετασχηματιστών της παρτίδας

VOLTAGE, που είναι η ένταση του ηλεκτρικού πεδίου στο οποίο γίνεται η μέτρηση

LOSSESF, που είναι οι απώλειες των μετασχηματιστών κατά την τριφασική μέτρηση. Η μεταβλητή αυτή βέβαια είναι η εξαρτημένη μεταβλητή του μοντέλου

LOSSESC, που είναι οι εγγυημένες απώλειες που ζητάει ο πελάτης

DIFFER, που είναι η διαφορά ανάμεσα στις πραγματικές και τις εγγυημένες απώλειες

DATE, που είναι η ημερομηνία κατά την οποία έγινε ο ποιοτικός έλεγχος

PIECES, που είναι ο αριθμός των μετασχηματιστών που περιέχονται σε κάθε παρτίδα

Συγκεντρωτικά λοιπόν, η αρχική ομάδα των μεταβλητών που μπορούν να χρησιμοποιηθούν στο μοντέλο σαν ανεξάρτητες είναι :

1. SUPPLIER PIECES	11. ORDER	21.
2. NOTRANS	12. NODESIGN	22. D1
3. G BAROS11	13. F1	23.
4. BAROS12 LOSSES12	14. LOSSES11	24.
5. LOSSES13 NOCOIL11	15. LOSSES14	25.
6. NOCOIL12 NOCOIL14	16. NOCOIL13	26.
7. IRONLOSS(1) IRONLOSS(3)	17. IRONLOSS(2)	27.

8. IRONLOSS(4)	18. INTENSIT	28. DATE
9. LOSSESC VOLTAGE	19. DIFFER	29.
10. ONISXYS	20. Eu	

Παρατηρούμε ότι :

A. Από την παραπάνω ομάδα "δυνατών" μεταβλητών, ορισμένες δεν μπορούν να χρησιμοποιηθούν στην κατοπινή μελέτη για την εύρεση του γραμμικού μοντέλου. Αυτές είναι :

α. PIECES. Διότι ο αριθμός των μετασχηματιστών που περιέχονται σε κάθε παρτίδα, από αυτές που κατασκευάστηκαν το 1991, δεν μας δίνει καμμία πληροφορία στην προσπάθεια πρόβλεψης των τριφασικών απωλειών. Το ίδιο ακριβώς ισχύει και για τις μεταβλητές NOTRANS, NODESIGN, NOCOIL11, NOCOIL12, NOCOIL13, NOCOIL14, DATE, DIFFER, LOSSESC.

β. D1. Διότι όλες οι μετρήσεις που έχουμε από αυτήν την μεταβλητή είναι ένας σταθερός αριθμός, οπότε ούτε αυτή μας δίνει πληροφορία για την πρόβλεψη των τριφασικών απωλειών. Το ίδιο ακριβώς ισχύει και για την μεταβλητή VOLTAGE.

γ. IRONLOSS(1). Διότι από τους 320 μετασχηματιστές που είχε το δείγμα μας, μόνο σε 62 από αυτούς αναφέρονταν οι απώλειες τις μαγνητικής λαμαρίνας. Το δείγμα αυτό των 62 μετασχηματιστών ήταν πάρα πολύ μικρό για να χρησιμοποιηθεί στην εξαγωγή στατιστικά χρήσιμων συμπερασμάτων. Το ίδιο ακριβώς ισχύει και για τις μεταβλητές IRONLOSS(2), IRONLOSS(3), IRONLOSS(4).

B. Υπάρχουν ακόμα κάποιες μεταβλητές για τις οποίες, ενώ υπάρχουν μετρήσεις από την ELBHM, δεν τέθηκε καθόλου θέμα χρησιμοποίησής τους. Ενδεικτικά αναφέρουμε ότι, στο φάκελλο που υπάρχει για κάθε παρτίδα μετασχηματιστών, υπάρχουν και τριφασικές μετρήσεις γι' απώλειες χαλκού. Ομως, αυτές οι μετρήσεις αναφέρονται στις απώλειες των πηνίων του κάθε μετασχηματιστή. Αρα, και αυτές οι μετρήσεις είναι πέρα από το αντικείμενο της μελέτης μας.

Μετά από την παραπάνω συζήτηση καταλήγουμε σε μία τελική ομάδα μεταβλητών, τις οποίες θα χρησιμοποιήσουμε σαν ανεξάρτητες για την εύρεση του γραμμικού μοντέλου. Αυτή η ομάδα είναι :

1. SUPPLIER

5. ORDER

9. G

2. BAROS11

6. BAROS11

10. LOSSES13

3. LOSSES11

7. LOSSES12

11. ONISXYS

4. LOSSES14

8. INTENSIT

12 . Eu

ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ

4.1 ΓΕΝΙΚΑ

Το Γενικό Γραμμικό Μοντέλο ορίζεται ως εξής [Βλέπε Μηλολιδάκης Κ. (1988)]: Ας ονομάσουμε Y_i $i = 1, \dots, n$ τις τιμές της εξαρτημένης μεταβλητής στο δείγμα μας (και κατ' αναλογία, ας θεωρήσουμε το διάνυσμα Y των n αυτών μεταβλητών). Το Γραμμικό μοντέλο υποθέτει ότι $Y = X\beta + \epsilon$ όπου :

- i. Ο πίνακας X διαστάσεως $n \times p$ ($n > p$) πρέπει να βρεθεί από εμάς, ονομάζεται πίνακας σχεδιασμού και είναι συνάρτηση των τιμών των ανεξάρτητων μεταβλητών στο δείγμα.
- ii. Το $p \times 1$ διάνυσμα β εκφράζει p παραμέτρους, που θα εκτιμηθούν με στατιστικές μεθόδους, και
- iii. Το $n \times 1$ διάνυσμα ϵ εκφράζει το "θόρυβο" ανάμεσα στο θεωρητικά υποτιθέμενο μοντέλο $X\beta$ και στην παρατήρηση Y . Το ϵ θεωρείται τυχαία μεταβλητή.

Ο στόχος μας είναι να κατασκευάσουμε κάποιο μοντέλο πρόβλεψης των τριφασικών απωλειών ενός μετασχηματιστή (εξαρτημένη μεταβλητή ή αλλιώς *response variable*), σύμφωνα με τις τιμές που θα παίρνουν ορισμένες κατάλληλα επιλεγμένες μεταβλητές (ανεξάρτητες μεταβλητές ή αλλιώς *explanatory variables*), που θεωρούμε ότι καθορίζουν την συμπεριφορά της εξαρτημένης μεταβλητής. Θεωρούμε ότι $E(\epsilon) = 0$ και $\text{cov}(\epsilon) = \Sigma$. Το μοντέλο λέγεται γραμμικό λόγω της γραμμικότητας ως προς τις άγνωστες παραμέτρους β . Η σχέση των Y_i με τα X_i μπορεί να είναι εξαιρετικά μη γραμμική.

Κρίσιμα ζητήματα στην κατάστρωση του μοντέλου είναι :

- α. Η κατανομή του ϵ
- β. Η δομή του πίνακα συνδιακύμανσης Σ (ο οποίος ως γνωστόν είναι συμμετρικός θετικά ορισμένος - *symmetric positive definite*).
- γ. Η τάξη (*rank*) και η δομή του πίνακα X

Κατ' αρχήν ξεκινάμε υποθέτοντας ότι, $\text{rank}(\mathbf{X})=p$ (τ' οποίο δεν αποτελεί πρόβλημα για τις σύγχρονες βιβλιοθήκες λογισμικού) και ότι η κατανομή του $\boldsymbol{\varepsilon}$ είναι η κανονική με μέσο $\mu=0$ και πίνακα συνδιακύμανσης $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}$ όπου, η σ^2 είναι άγνωστη παράμετρος και \mathbf{I} είναι ο μοναδιαίος πίνακας. Εδώ είναι δυνατόν να υπάρξουν αποτελέσματα και με άλλες υποθέσεις όσον αφορά στην κατανομή του $\boldsymbol{\varepsilon}$ (αν και η σχετική επεξεργασία είναι σαφώς δυσκολότερη).

Ο τρόπος που δουλεύουμε είναι ο εξής : Παρατηρούμε ζεύγη (\mathbf{X}_i, Y_i) όπου, $i=1, \dots, n$ και \mathbf{X}_i είναι το διάνυσμα των μετρήσεων όπου κάθε συντεταγμένη παριστάνει και μία ανεξάρτητη μεταβλητή (explanatory variable) και Y_i είναι η αντίστοιχη μέτρηση της εξαρτημένης μεταβλητής (response variable), που στην μελέτη μας είναι οι απώλειες της τριφασικής μέτρησης. Αν υποθέσουμε ότι υπάρχει κάποιο γραμμικό μοντέλο που "διέπει" την σχέση των (\mathbf{X}_i, Y_i) (δηλαδή κάποιος, γνωστός, πίνακας \mathbf{X}) τότε, μπορούμε να απαντήσουμε στα εξής ερωτήματα :

- α. Να εκτιμήσουμε το άγνωστο διάνυσμα $\boldsymbol{\beta}$.
- β. Να δώσουμε διαστήματα πρόβλεψης (prediction intervals) για το που θα πέσει μια επόμενη μέτρηση, για παράδειγμα η Y_{n+1} , ανάλογα με την τιμή του \mathbf{X}_{n+1} .
- γ. Να προσδιορίσουμε, εκ των προτέρων, την τιμή του Y_{n+1} και να βρούμε την τιμή του \mathbf{X}_{n+1} από αυτήν (calibration problem : αυτό βασικά μας ενδιέφερε διότι απαντά στο ερώτημα "εύρεσης της καμπύλης" που τέθηκε).
- δ. Να ελέγξουμε αν δύο διαφορετικά μοντέλα είναι "ίδια".
- ε. Να οργανώσουμε πειράματα για την επίδραση των διάφορων ανεξάρτητων μεταβλητών στην εξαρτημένη μεταβλητή, την αλληλεπίδραση μεταξύ τους κ.ά.
- στ. Να ελέγξουμε αν κάποια καινούργια ανεξάρτητη μεταβλητή "προσθέτει" ουσιώδη γνώση.

Τα προβλήματα που είναι δυνατόν να προκύψουν όταν ξεκινάμε με δεδομένα (data) της μορφής (\mathbf{X}_i, Y_i) όπου $i=1, \dots, n$ είναι τα εξής :

- α. Ποιά θα είναι η πιο σωστή γενική μορφή γραμμικού μοντέλου που μπορούμε να εφαρμόσουμε στα στοιχεία. Εδώ δεν υπάρχει συστηματική θεωρία και κανένας απαντά στο ερώτημα αυτό με διάφορες τεχνικές πάνω στον υπολογιστή, ξεκινώντας από "απλά" μοντέλα -δηλαδή κρατώντας σταθερές ορισμένες μεταβλητές,

μετρώντας την καλή προσαρμογή τους στα στοιχεία, κ.λ.π. [Βλέπε ειδικά τα κεφάλαια 3, 6, 7, 8 του Draper & Smith, καθώς και τα κεφάλαια 8, 12 του Graybill].

β. Η ορθότητα των υποθέσεων για τη δομή του λάθους ϵ , και την κατανομή του (για παράδειγμα αν ισχύει ότι $\Sigma = \sigma^2 \mathbf{I}$). Εδώ υπάρχουν διάφοροι στατιστικοί έλεγχοι υποθέσεων που μπορούν να γίνουν. Γενικά πάντως, το πρόβλημα αυτό είναι δύσκολο.

γ. Η ενδεχόμενη ύπαρξη, στα (\mathbf{X}_i, Y_i) "κακών" παρατηρήσεων (ανώμαλες, outliers) που χρειάζεται να αφαιρεθούν, διότι διαφορετικά θα μας οδηγήσουν σε εσφαλμένα συμπεράσματα.

δ. Η ενδεχόμενη ύπαρξη υποσυνόλων των παρατηρήσεων, που το καθένα (υποσύνολο) να ακολουθεί ξεχωριστό γραμμικό μοντέλο. Και αν ναι, αν μπορούμε να "διακρίνουμε" τέτοια υποσύνολα (Discriminant Analysis).

ε. Αν είναι δυνατόν να απλοποιήσουμε το μοντέλο, δηλαδή να επιβεβαιώσουμε ότι ορισμένες ανεξάρτητες μεταβλητές (explanatory variables) μπορούν να απαλειφθούν, χωρίς σοβαρές επιπτώσεις στην καλή προσαρμογή του μοντέλου στα στοιχεία. Γενικά το πρόβλημα που μπορεί να αντιμετωπίσουμε είναι, πόσο ουσιώδης είναι μια ανεξάρτητη μεταβλητή για την "εξήγηση" της εξαρτημένης μεταβλητής (στην περίπτωση που θα αντιμετωπίσουμε τις τριφασικές απώλειες του μετασχηματιστή), και ακόμη πόσο πολύ αυτή επηρεάζει τις απώλειες.

στ. Ποιοί είναι τελικά οι περιορισμοί στην εφαρμογή του γραμμικού μοντέλου, και ποιά η ερμηνεία της τελικής σχέσεως που θα εξαχθεί από την ανάλυση των δεδομένων.

4.2 ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΓΕΝΙΚΗ ΓΡΑΜΜΙΚΗ ΥΠΟΘΕΣΗ

Το γραμμικό μοντέλο που χρησιμοποιούμε εδώ έχει την μορφή :

$$\mathbf{Y} = \mathbf{X} \hat{\mathbf{a}} + \hat{\mathbf{a}}$$

όπου, \mathbf{Y} είναι το διάνυσμα των 320 τιμών (10% του πληθυσμού) των απωλειών της τριφασικής μέτρησης του δείγματος. Ο πίνακας \mathbf{X} οργανώνει τις ανεξάρτητες μεταβλητές (ή αλλιώς predictor variables) του μοντέλου μας. Η επιτυχημένη επιλογή του οδηγεί σε αξιόπιστη πρόβλεψη του β κι επομένως στη δυνατότητα πρόβλεψης

των απωλειών Y , για κάποιο νέο σύνολο τιμών των ανεξάρτητων μεταβλητών. Ας ονομάσουμε \hat{a} , την εκτίμηση των αγνώστων παραμέτρων β , που γίνεται χρησιμοποιώντας τεχνικές ελαχίστων τετραγώνων. Η "πρόβλεψη" μας τότε για την τιμή των Y , ας την ονομάσουμε \hat{Y} , θα είναι η :

$$\hat{Y} = X \hat{a}$$

Το διάνυσμα $r : Y - \hat{Y} = Y - X \hat{a}$ ονομάζεται διάνυσμα των residuals και προφανώς δίνει το σφάλμα ανάμεσα στην πρόβλεψη του μοντέλου μας και την πραγματική μέτρηση. Η δύσκολη εργασία σε παρόμοια μοντέλα βρίσκεται στην ανάλυση των residuals που γίνεται με τεχνικές ελέγχου υποθέσεων. Είναι αυτή η ανάλυση που αποφασίζει το βαθμό αξιοπιστίας του μοντέλου μας.

Η βασική υπόθεση για το μοντέλο αφορά το διάνυσμα ϵ και είναι ότι το $\hat{a} \approx \hat{I} (0, \sigma^2 I)$, δηλαδή ότι οι αποκλίσεις είναι κανονικά κατανοημένες, στοχαστικά ανεξάρτητες, τυχαίες μεταβλητές με κοινή, άγνωστη, διασπορά.

Κριτήριο για την καλή προσαρμογή του μοντέλου στις μέτρησεις του δείγματός μας αποτελεί το τετράγωνο του συντελεστή πολλαπλής συσχέτισης, R^2 , (squared multiple correlation coefficient). Αν ονομάσουμε \hat{Y}_i την πρόβλεψη που δίνει το μοντέλο για τις απώλειες της τριφασικής μέτρησης του i -οστού στοιχείου του δείγματος και \bar{Y} την μέση τιμή των τριφασικών απωλειών στο δείγμα, τότε μπορεί ν' αποδειχθεί ότι :

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Επομένως, το R^2 μετρά το ποσοστό της συνολικής διακύμανσης γύρω από τον μέσο \bar{Y} , που μπορεί να εξηγήσει το μοντέλο. Ένα μοντέλο με R^2 κοντά στο 1, που δεν παραβιάζει τους απαραίτητους ελέγχους υποθέσεων πάνω στους οποίους στηρίζεται, είναι το ζητούμενο.

Η εργασία που έγινε χωρίστηκε στα εξής στάδια που συνοψίζουμε εδώ. (Αναλυτική περιγραφή μαζί με τις σχετικές βιβλιογραφικές αναφορές ακολουθεί):

α. Επιλογή των ανεξάρτητων μεταβλητών που θα εισάγουμε στο μοντέλο (δηλαδή επιλογή του "καλύτερου" υποσυνόλου predictor variables), εάν υποθέσουμε ότι, ο πίνακας X δε διαθέτει όρους δευτέρου ή/και μεγαλύτερου

βαθμού. Η επιλογή έγινε με τη βήμα-βήμα προς τα πίσω διαδικασία απαλειφής (Stepwise Backward Elimination) και με κριτήρια :

- i. Ελέγχους ανάλυσης της διασποράς για τον κάθε συντελεστή μέσω F-test.
- ii. Καθορισμό υψηλών επιθυμητών tolerances
- iii. Προσπάθεια να διατηρηθεί το R^2 σε υψηλά επίπεδα

Η διάσταση του καλύτερου μοντέλου που επιλέχθηκε, ελέγχθηκε σύμφωνα με την cross validation τεχνική του Leo Breimann. Τέλος, προχωρήσαμε στον έλεγχο της υπόθεσης βελτίωσης του μοντέλου αυτού με την εισαγωγή αλληλεπιδράσεων μεταξύ των μεταβλητών ή/και μη γραμμικών όρων, στον πίνακα σχεδιασμού X. Τέτοιες αλληλεπιδράσεις και όροι δε βρέθηκαν, κι έτσι καταλήξαμε σ' ένα πίνακα σχεδιασμού διάστασης 320x26. Η εργασία στο μέρος αυτό εξελίχθηκε ομαλά.

β. Ελεγχοι καλής προσαρμογής του μοντέλου. Χρησιμοποιήθηκαν οι εξής τεχνικές :

- i. Χρήση των διαφορετικών μετρήσεων τριφασικών απωλειών για τα ίδια διανύσματα X των ανεξάρτητων μεταβλητών (δηλαδή χρήση των repeats) για το σπάσιμο του αθροίσματος των τετραγώνων των residuals (SSR) σε αυτό που οφείλεται σε έλλειψη προσαρμογής και σε αυτό που οφείλεται σε καθαρό σφάλμα. Στην συνέχεια, έγινε έλεγχος μέσω F-test της ισότητας των διασπορών όπως προκύπτουν από τις δύο εκτιμήτριες.
- ii. Έλεγχος της συνολικής προσαρμογής του μοντέλου, μέσω ανάλυσης της διασποράς (F-test for overall regression). Τέλος, έγινε χρήση του κριτηρίου του Wetz, για την καλή προσαρμογή. Και στο μέρος αυτό η εργασία εξελίχθηκε ομαλά.

γ. Έλεγχος της κανονικότητας στην κατανομή των residuals του μοντέλου. Σαν κριτήρια χρησιμοποιήσαμε τα :

- i. Ιστόγραμμα των residuals.
- ii. Απεικόνιση της αναμενόμενης τιμής των απωλειών της τριφασικής μέτρησης σύμφωνα με το μοντέλο, συναρτήσει των residuals, σε κανονικό διάγραμμα (Normal plot).
- iii. Απαραμετρικό έλεγχο κανονικότητας Kolmogorov-Smirnov (παραλλαγή Lilliefors) των residuals.

Και στο μέρος αυτό, η εργασία εξελίχθηκε ομαλά.

δ. Έλεγχος ομοιογένειας της διασποράς (homoscedasticity test). Χρησιμοποιήθηκαν οι εξής τεχνικές :

i. Διαγράμματα των residuals συναρτήσει, αφενώς της πρόβλεψης του μοντέλου και αφ' ετέρου κάθε μιας ανεξάρτητης (predictor) μεταβλητής.

ii. Ελεγχος Bartlett, για την ομοιογένεια των διασπορών. Εδώ διαπιστώθηκε ότι η διασπορά των τριφασικών απωλειών δεν παρέμενε σταθερή, αλλά εξαρτιόταν από τις ανεξάρτητες μεταβλητές του μοντέλου που είχε κατασκευαστεί στο α (εκτός από τις τέσσερις μεταβλητές που αφορούσαν τη διαβάθμιση των μονοφασικών απωλειών) και αυξανόταν καθώς αυξανόταν η πρόβλεψη των τριφασικών απωλειών του μοντέλου. Για τον λόγο αυτό, απαιτήθηκε η διόρθωση του μοντέλου, μέσω της τεχνικής των weighted least squares. Για τον σκοπό αυτό προχωρήσαμε στην κατασκευή μοντέλου πρόβλεψης της διασποράς της εξαρτημένης μεταβλητής του αρχικού μοντέλου.

ε. Το μοντέλο προβλεψής της διασποράς. Κατασκευάσαμε, χρησιμοποιώντας από την αρχή τις τεχνικές του γραμμικού μοντέλου, που συνοπτικά έχουμε μέχρι τώρα περιγράψει, το καλύτερο δυνατό μοντέλο πρόβλεψης της διασποράς [βλέπε το αναλυτικό σχετικό κεφάλαιο παρακάτω]. Σύμφωνα με τα κριτήρια ελέγχου που χρησιμοποιούμε [βλέπε α, β, γ και δ], το μοντέλο αυτό ήταν ικανοποιητικό.

στ. Διόρθωση του αρχικού μοντέλου πρόβλεψης των τριφασικών απωλειών, μέσω της τεχνικής των weighted least squares. Μετά τον μετασχηματισμό εξαρτημένης και ανεξάρτητων μεταβλητών, επαναλάβαμε τις εργασίες α, β, γ, δ και τ' αποτελέσματα ήταν ικανοποιητικά.

ζ. Εντοπισμός μη αναμενομένων τιμών στα residuals (outliers). Για τον εντοπισμό τους χρησιμοποιήθηκε η studentized version των residuals, ώστε να ληφθεί υπ' όψιν η τυπική απόκλιση. Η ύπαρξη πολλών τέτοιων τιμών (στην περίπτωσή μας έως 3 θεωρούνται φυσιολογικές για το μέγεθος του δείγματος που χρησιμοποιούμε), εφ' όσον οι έλεγχοι α έως και δ είναι ικανοποιητικοί, υπαινίσσεται την επίδραση κάποιου μεμονωμένου παράγοντα που αύξησε υπέρμετρα την απόσταση μεταξύ πραγματικής μέτρησης τριφασικών απωλειών και πρόβλεψης. Σαν τέτοιο, για παράδειγμα, θα μπορούσαμε να θεωρήσουμε το υπέρμετρο πέταγμα λαμαρίνας κατά την πλέξη ενός μετασχηματιστή με εξαιρετικά καλή σύνθεση πυρήνων. Οι outliers μπορούν να μην λαμβάνονται υπ' όψη κατά την ανάλυση του μοντέλου (να "ξεσκαρτάρονται"), εφ' όσον όμως εντοπιστούν οι μεμονωμένοι παράγοντες που τους δημιούργησαν. Στην περίπτωσή μας κάτι τέτοιο βέβαια ήταν αδύνατο.

Στο αρχικό μοντέλο παρουσιάστηκαν 6 outliers και συγκεκριμένα [βλέπε Παράρτημα Β1], οι περιπτώσεις 28, 41, 75, 185, 229 και 269 του δείγματός μας].

Ξεσκαρτάραμε τους 4 μεγαλύτερους, κατ' απόλυτη τιμή (δηλαδή τους 28, 41, 75, 185), αλλά στη συνέχεια διατηρήσαμε όσους παρουσιάστηκαν εκ νέου μετά τις διάφορες αναθεωρήσεις του μοντέλου.

η. Ελεγχος και αναγνώριση παρατηρήσεων με σημαντική επίδραση στο μοντέλο. Ένα ευσταθές μοντέλο θα πρέπει να μην επηρεάζεται από μεμονωμένα σημεία του δείγματος. Ελέγξαμε την ευστάθεια του μοντέλου, χρησιμοποιώντας την απόσταση Cook με ιδιαίτερα ικανοποιητικά αποτελέσματα.

θ. Ελεγχος της στοχαστικής ανεξαρτησίας των ϵ_i . Ο συνηθέστερος τρόπος είναι, μέσω της ανάλυσης των residuals, η αναζήτηση σειριακών συσχετίσεων (serial correlations) των residuals. Χρησιμοποιήσαμε τους εξής ελέγχους :

- i. Την στατιστική D, των Durbin-Watson.
- ii. Διάγραμμα ελέγχου των αυτοσυσχετίσεων μέχρι και lag15 (autocorrelations plot ή ACF plot).
- iii. Απαραμετρικό έλεγχο διαδρομών Wald-Wolfowitz για διαδρομές θετικών και αρνητικών προσήμων στα residuals.
- iv. Scatterplots των residuals συναρτήσει πάλι των residuals με lag1.
- v. Έλεγχο Bartlett της σειριακής συσχέτισης lag1, με ταυτόχρονη χρήση πίνακα πιθανοτήτων Bonferroni.

Οι παραπάνω έλεγχοι οδηγούν στο συμπέρασμα ότι, ο πίνακας συνδιακύμανσης (covariance matrix) των residuals του διορθωμένου μοντέλου, όπως έχει προκύψει μετά την χρήση weighted least squares [βλέπε στ προηγουμένως], δεν έχει την υποτιθέμενη από το μοντέλο μορφή $\sigma^2 \mathbf{I}$, αλλά έχει την μορφή :

$$V = (V_{ij}) \quad i = 1, \dots, 316 \quad \text{é} \text{á} \text{é} \quad j = 1, \dots, 316 \quad \text{ì} \text{à} \quad V_{ij} = \begin{cases} \sigma^2 & \text{áí} \quad i = j \\ \tilde{\sigma}_1 & \text{áí} \quad |i - j| = 1 \\ \tilde{\sigma}_2 & \text{áí} \quad |i - j| = 2 \\ \tilde{\sigma}_3 & \text{áí} \quad |i - j| = 3 \\ 0 & \text{á} \text{é} \text{á} \text{ö} \text{ï} \text{ñ} \text{á} \text{ó} \text{é} \text{é} \text{ü} \end{cases}$$

όπου $\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3$ οι σειριακές συσχετίσεις lag1, lag2, lag3 που εκτιμήθηκαν σαν $\tilde{\sigma}_1 = 0.225, \tilde{\sigma}_2 = 0.282 \quad \text{é} \text{á} \text{é} \quad \tilde{\sigma}_3 = 0.209$ [βλέπε επίσης Draper και Smith (1981), σελ. 156]. Δηλαδή παρατηρούμε ότι, οι σειριακές αυτοσυσχετίσεις lag1 και lag2, αν και μικρές, δεν μπορούν ν' αγνοηθούν. Βέβαια, η επίδρασή τους πάνω στο μοντέλο είναι μικρή· για παράδειγμα η μη αγνόηση της ρ_1 σημαίνει ότι περίπου το 5.0%

$[(0.225)^2=0.050]$ της διακύμανσης του λάθους στην πρόβλεψη των τριφασικών απωλειών ενός μετασχηματιστή, μπορεί να "εξηγηθεί" από το λάθος στην πρόβλεψη του αμέσως προηγούμενου του.

Η μη διόρθωση των σειριακών συσχετίσεων, όμως, μπορεί να οδηγήσει σε τεχνητή μείωση του τυπικού σφάλματος της πρόβλεψης των τριφασικών καθώς και των τυπικών σφαλμάτων των συντελεστών του μοντέλου, δίνοντας μια ψεύτικη εικόνα ακρίβειας. Αν και στην περίπτωση μας οι μεταβολές αυτές δεν είναι πολύ μεγάλες, θα μπορούσαν να θέσουν ερωτηματικά για ορισμένους, έστω, ελέγχους από αυτούς που κάνουμε στα α έως η . Η συνηθέστερη αιτία που οδηγεί στην παρουσία σειριακών συσχετίσεων είναι η μη εισδοχή στο σύνολο των predictor variables κάποιας ανεξάρτητης μεταβλητής, συσχετισμένης με το χρόνο κατασκευής του μετασχηματιστή. Στην περίπτωση που αναλύσαμε οι μόνες απώλειες που επηρεάζονται με το χρόνο κατασκευής είναι οι απώλειες τις μονοφασικής μέτρησης, αφού οι υπόλοιπες καθορίζονται από την μελέτη. Μεταβλητές που η είσοδός τους στο μοντέλο θα μπορούσε να οδηγήσει σε άρση των σειριακών συσχετίσεων θεωρήσαμε ότι θα ήταν κυρίως :

- i. Οι απώλειες της μαγνητικής λαμαρίνας όπως μετρήθηκαν από την ΕΛΒΗΜ.
- ii. Οι μεταβλητές της ανόπτησης (αφού οι διαδικασίες της ανόπτησης σε μεγάλο βαθμό επηρεάζονται από την ουρά αναμονής στους θαλάμους ανόπτησης κι επομένως από το χρόνο παραγωγής).
- iii. Το πραγματικό βάρος του κάθε πυρήνα (ή έστω, του μετασχηματιστή συνολικά) μετά την πλέξη.

Η άρση των σειριακών συσχετίσεων μέσω "διερεύνησης" του μοντέλου θα οδηγούσε σε σημαντικές βελτιώσεις στην ακρίβεια, αφού η διόρθωσή τους χωρίς την εισαγωγή νέων μεταβλητών δείχνει ότι, το τυπικό σφάλμα στην πρόβλεψη είναι αρκετά μεγαλύτερο απ' ό,τι αρχικά εκτιμά το μοντέλο.

Στην συγκεκριμένη μελέτη της παραγωγής του 1991 επιχειρήσαμε την λύση των σειριακών συσχετίσεων με δύο τρόπους. Πρώτα εισάγαμε τον αριθμό εντολής σαν επί πλέον μεταβλητή. Η μεταβλητή αυτή είναι "τεχνητή" σαν predictor variable, αλλά εν απουσία άλλων συσχετισμένων με το χρόνο μεταβλητών, κρίθηκε αναγκαία για την ελάττωση των σειριακών συσχετίσεων. Αυτό επιβεβαιώθηκε και από την ανάλυση της διασποράς που επακολούθησε. Ο αριθμός εντολής έλυσε το πρόβλημα της αυτοσυσχέτισης lag1 αλλά αύξησε την αυτοσυσχέτιση lag2. Μην έχοντας άλλες δυνατότητες εισαγωγής μεταβλητών εξαρτώμενων από το χρόνο, επιχειρήθηκε η διόρθωση του προβλήματος αυτού να γίνει μέσω της διαδικασίας

Cochrane-Orcutt. Το εκ νέου αναθεωρημένο μοντέλο παρουσιάζει μεγαλύτερα τυπικά σφάλματα (standard errors) στις εκτιμήτριες των παραμέτρων του διανύσματος β , αλλά θεωρείται καλύτερο μοντέλο πρόβλεψης επειδή λύνει το πρόβλημα των σειριακών αυτοσυσχετίσεων. Εναλλακτικά, αλλά με χρήση μεγαλύτερου υπολογιστικού συστήματος από αυτό που χρησιμοποιήθηκε, θα μπορούσε εδώ να επιχειρηθεί είτε η Cholesky decomposition του πίνακα V της συνδιακύμανσης των residuals, είτε η επαναληπτική χρήση της μεθόδου Cochrane-Orcutt που θα οδηγούσε σε μείωση των τυπικών σφαλμάτων.

Στη συνέχεια προχωράμε στην αναλυτική παρουσίαση και τεκμηρίωση των α έως θ του κεφαλαίου αυτού παραθέτοντας ταυτόχρονα και τις σχετικές εκτυπώσεις αναλύσεων, υπολογισμών και ελέγχων. Μετά το τέλος της αναλυτικής παρουσίασης δίνεται και ο τύπος του μοντέλου πρόβλεψης των τριφασικών απωλειών των μετασχηματιστών, στον οποίο καταλήξαμε χρησιμοποιώντας τα δεδομένα της παραγωγής του 1991. Αξίζει τέλος να σημειωθεί ότι οι έκτοτε ουσιώδεις αλλαγές στη διαδικασία εκπόνησης της μελέτης κατασκευής των μετασχηματιστών ή ακόμη και στην παραγωγική διαδικασία στο εργοστάσιο θα οδηγήσουν και σε αλλαγή του προτεινόμενου μοντέλου.

4.2.1 ΕΠΙΛΟΓΗ ΚΑΛΥΤΕΡΟΥ ΥΠΟΣΥΝΟΛΟΥ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ

Κατ' αρχήν εξετάζουμε την περίπτωση όπου έχουμε πίνακα σχεδιασμού X χωρίς όρους βαθμού μεγαλύτερου του πρώτου.

- α. Οι μεταβλητές που εισήχθησαν αρχικά στο μοντέλο ήταν οι εξής :
 1. E_u
 2. G
 3. Οι απώλειες της μονοφασικής μέτρησης για τον κάθε πυρήνα που απαρτίζει κάθε μετασχηματιστή
 4. Η ένταση B του μαγνητικού πεδίου στο οποίο γίνεται η μονοφασική μέτρηση (μεταβλητή INTENSIT).
 5. Ο προμηθευτής της μαγνητικής λαμαρίνας

Οι παρατηρήσεις που έχουμε να κάνουμε σε αυτό το σημείο είναι :

i. Οι απώλειες της μονοφασικής μέτρησης και ο προμηθευτής της πρώτης ύλης υποχρεωτικά θεωρήθηκαν σαν κατηγορικές και όχι σαν αριθμητικές μεταβλητές. Έτσι ο πραγματικός αριθμός στηλών του πίνακα σχεδιασμού **X** είναι 27.

ii. Οι μεταβλητές της βάσης δεδομένων που δεν εισάγονται είναι :

1. Η ονομαστική ισχύς του μετασχηματιστή (μεταβλητή ONISXYS). Αυτό επειδή ο κατ' αρχήν έλεγχος συγγραμικότητας εντόπισε ότι η μεταβλητή αυτή ήταν γραμμικός συνδυασμός άλλων μεταβλητών που ήδη εισάγονταν.

2. Οι μεταβλητές F1, BAROS11, BAROS12 διότι είχε θεωρηθεί, από το τμήμα μελετών της ΕΛΒΗΜ, ότι δεν περιέχουν ιδιαίτερα σοβαρή πληροφορία για τις απώλειες οι οποίες ήταν εκφρασμένες σε w/kg. Πάντως την ανάγκη ή όχι εισαγωγής και των μεταβλητών αυτών εξετάζουμε στην συνέχεια.

3. Η μεταβλητή D1 διότι παρατηρήθηκε ότι στους μετασχηματιστές των 50 KVA είχαμε πάρει σαν D1 το D. Το λάθος οφειλόταν στην αλλαγή εκτέλεσης της μελέτης την περίοδο εκείνη από το εργοστάσιο (δηλαδή αντί για 8 πυρήνες τελικά χρησιμοποιήθηκαν 4), πράγμα που δεν έγινε αντιληπτό κατά το πέρασμα των μετρήσεων στη βάση δεδομένων.

Το R^2 είναι ικανοποιητικό (0.982), αλλά οι έλεγχοι ανάλυσης της διασποράς απορρίπτουν το μοντέλο αυτό, υποδεικνύοντας την ανάγκη μείωσης του αριθμού των μεταβλητών [βλέπε Παράρτημα B1]. Το MEAN SQUARE ERROR (MSE) είναι 216.263.

β. Ακολουθεί επομένως η διαδικασία εξαγωγής μεταβλητών μία προς μία από το σύνολο των predictor variables. Κριτήριο εξαγωγής μιας μεταβλητής είναι ο στατιστικός έλεγχος της υπόθεσης ότι η επίδρασή της είναι ασήμαντη στο μοντέλο, ο οποίος γίνεται μέσω των F-test της ανάλυσης της διασποράς. Επίσης κριτήριο είναι το τυπικό λάθος της πρόβλεψης των απωλειών που στους σχετικούς πίνακες [βλέπε Παράρτημα B2], είναι η τετραγωνική ρίζα του MSE, καθώς και το STANDARD ERROR των συντελεστών, που η ελάττωσή του συνεπάγεται καλύτερη εκτίμησή τους. Η διαδικασία αυτή (stepwise backward elimination) οδήγησε σε απαλειφή τις απώλειες της μονοφασικής μέτρησης του πρώτου πυρήνα (LOSSES11) η οποία τελικά δεν έγινε επειδή τη θεωρούμε σαν a priori σημαντική μεταβλητή.

Άρα τελικά το μοντέλο έμεινε αμετάβλητο κι έτσι είχαμε το ίδιο R^2 (0.982) καθώς επίσης και το ίδιο MSE (216.263).

γ. Ακολούθως στο αρχικό σετ μεταβλητών αντικαταστήσαμε την μεταβλητή "INTENSIT" με την μεταβλητή "ONISXYS" κι επαναλάβαμε όσα περιγράφονται στα α και β προηγουμένως. Με την ONISXYS στη θέση της INTENSIT το καλύτερο σετ μεταβλητών είναι : Eu, G, Onisxys, Απώλειες μονοφασικής, προμηθευτής. Για το σετ αυτό παίρνουμε $R^2=0.967$ και $MSE=395.452$. Παρατηρούμε ότι αυτό το σετ μεταβλητών είναι κατά πολύ χειρότερο από αυτό του β οπότε και αποφασίσαμε να παραμείνουμε στο σετ του β. Κατόπιν στο σετ μεταβλητών του β προστέθηκε η μεταβλητή F1. Αμέσως διαπιστώθηκε, με τη χρήση ελέγχου F-test , ότι αυτή είναι περιττή κι έτσι απαλείφθηκε. Τέλος, πήραμε το μεγαλύτερο σύνολο μεταβλητών χωρίς συγγραμμικότητες, που να περιέχουν τις μεταβλητές BAROS11, BAROS12 (δηλαδή τα βάρη του μικρού και του μεγάλου πυρήνα αντίστοιχα). Χρησιμοποιώντας και πάλι την τεχνική της backwards stepwise elimination καταλήξαμε ότι με αυτή την αφετηρία το καλύτερο σετ ανεξάρτητων μεταβλητών ήταν : Eu, G, BAROS11, Απώλειες της μονοφασικής μέτρησης, προμηθευτής. Για το σετ αυτό, μετά την ανάλυση, πήραμε περίπου τα ίδια αποτελέσματα με το προηγούμενο σετ. Αποφασίσαμε τελικά να παραμείνουμε, και να χρησιμοποιήσουμε, το σετ μεταβλητών που προέκυψε από το β.

δ. Στην συνέχεια ελέγχθηκε η "διάσταση" του μοντέλου, δηλαδή ο αριθμός των ανεξάρτητων μεταβλητών, σύμφωνα με την συνιστώμενη από τον Leo Breimann τεχνική [Βλέπε Wilkinson, Leland (1990)]. Με γεννήτρια τυχαίων αριθμών πήραμε δείγμα 256 παρατηρήσεων (80% του συνόλου του δείγματος) και σε αυτές προσαρμόσαμε ένα νέο γραμμικό μοντέλο με τις ίδιες ανεξάρτητες μεταβλητές [βλέπε Παράρτημα Β3] που έδωσε $R_{80\%}^2=0.985$. Σύμφωνα με το μοντέλο αυτό έγινε πρόβλεψη των απωλειών για τις υπόλοιπες 64 παρατηρήσεις (20% του δείγματος) και στην συνέχεια υπολογίστηκε το R^2 μεταξύ της πρόβλεψης και των πραγματικών απωλειών για το 20%. Είχαμε $R_{20\%}^2=0.984$. Σε περίπτωση που το μοντέλο περιελάμβανε περιττές μεταβλητές η διαφορά των δύο R^2 θα ήταν δραματική. Στην περίπτωσή μας παρατηρούμε ότι τα δύο R^2 σχεδόν συμπίπτουν οπότε δεχόμαστε τον έλεγχο.

ε. Τέλος προχωρήσαμε στη διερεύνηση της ύπαρξης τυχόν αλληλεπιδράσεων μεταξύ των μεταβλητών καθώς και μη γραμμικών όρων στον πίνακα σχεδιασμού **X**. Τη διερεύνηση αυτή την κάναμε αφού βγάλαμε τους 4 outliers, έχοντας πεισθεί ότι η παρουσία τους δεν οφείλεται στο μοντέλο [βλέπε κεφάλαιο 4.2, μέρος ζ]. Προστέθηκαν διαδοχικά στο μοντέλο και εξετάσθηκε η αλληλεπίδραση των :

- i. Eu * G
- ii. Eu * INTENSIT
- iii. G * INTENSIT

καθώς και ύπαρξη παραγόντων της μορφής :

iv. $(Eu)^2$

v. $(G)^2$

vi. $(INTENSIT)^2$

Για κάθε μία από τις προηγούμενες έξη περιπτώσεις έγινε *stepwise backward elimination* για να βρεθεί το καλύτερο μοντέλο που θα περιείχε και όρους της παραπάνω μορφής. Δε θα παραθέσουμε τις σχετικές πολυσέλιδες εκτυπώσεις, αλλά το συμπέρασμα είναι ότι, παρόντων των μεταβλητών που έχουν επιλεγεί από το μέρος β. παραπάνω, όλες οι αλληλεπιδράσεις και όροι βαθμού μεγαλύτερου του 1 είναι ασήμαντοι. Επομένως παραμένουμε στο μοντέλο με ανεξάρτητες μεταβλητές τις Eu , G , $INTENSIT$, απώλειες της μονοφασικής μέτρησης, προμηθευτή της πρώτης ύλης.

4.2.2 ΕΛΕΓΧΟΙ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ ΤΟΥ ΜΟΝΤΕΛΟΥ

Οι έλεγχοι έγιναν μετά την αφαίρεση των 4 outliers 28, 41, 75, 185. Συνολικά έγινε :

α. Χρήση της τεχνικής των επαναλήψεων. Η ιδέα είναι ότι υπάρχουν ομάδες μετρήσεων που αντιστοιχούν στις ίδιες ακριβώς τιμές των ανεξαρτήτων μεταβλητών (για παράδειγμα έχουμε μια ομάδα από 3 μετασχηματιστές, που ενώ ήταν καθ' όλα ίδιοι όσον αφορά τις ανεξάρτητες μεταβλητές, έδωσαν διαφορετικές μετρήσεις τριφασικών απωλειών). Η διασπορά τριφασικής μέσα στην ομάδα θεωρείται ότι οφείλεται σε τυχαίους παράγοντες (*pure error*) και επομένως αποτελεί μια εκτιμήτρια της άγνωστης σ^2 . Αν τώρα υποθεθεί ότι το μοντέλο προσαρμόζεται καλά, δε θα υπάρχει μεροληψία (*bias*) στις διακυμάνσεις των μετρήσεων ανάμεσα στις ομάδες κι επομένως και από εκεί μπορούμε να έχουμε μια εκτιμήτρια της σ^2 . Αρα ο έλεγχος της καλής προσαρμογής μπορεί να γίνει ελέγχοντας αν η διαφορά των δύο εκτιμητριών οφείλεται σε τυχαίους παράγοντες. Πρόκειται για *F-test* που γίνεται σπάζοντας το *SSR* (*Sum of Squares of the Residuals*). Εμείς από τα *repeats* βρήκαμε 42 ομάδες και πήραμε :

	DF	SS	MS	F-Ratio
Residual	289	48877.542	169.126	
Lack of Fit	232	39243.671	169.153	

Pure Error	57	9633.871	169.015	1.00081
------------	----	----------	---------	---------

Εφόσον η τιμή του F-Ratio είναι περίπου ίση με 1 τότε δεχόμαστε ότι οι δύο εκτιμήτριες εκτιμούν την ίδια σ^2 και άρα το μοντέλο δείχνει καλή προσαρμογή.

β. F-test for overall regression. Στο Παράρτημα B2 παρουσιάζεται το μοντέλο όπως προέκυψε μετά την αφαίρεση των outliers. Έχουμε ότι $R^2=0.984$ και $MSE=188.246$. Παρατηρούμε ότι το πηλίκο Regression M.S./Residuals MS ξεπερνάει κατά πολύ το τετραπλάσιο της $F(25,289,0991)$ κι επομένως η προσαρμογή του μοντέλου θεωρείται πολύ καλή [βλέπε κριτήριο του J. Wetz, όπως παρατίθεται στους Draper & Smith (1981), σελ. 93].

4.2.3 ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΣΤΗΝ ΚΑΤΑΝΟΜΗ ΤΩΝ RESIDUALS

Στο παράρτημα B6 δίνονται οι έλεγχοι της υπόθεσης ότι τα residuals είναι κανονικά κατανομημένα. Γενικά η ανάλυση των residuals είναι από τα κεντρικότερα ζητήματα στην υιοθέτηση ενός γραμμικού μοντέλου [βλέπε Draper & Smith (1981) κεφάλαιο 3, επίσης Seber (1977) κεφάλαιο 6, και τέλος Cook & Weisberg κεφάλαιο 2]. Ελέγχουμε την κανονικότητα χρησιμοποιώντας :

α. Ιστόγραμμα των Residuals. Ο έλεγχος είναι οπτικός και δεν εμφανίζεται αισθητή απόκλιση από την κανονική κατανομή.

β. Normal plot. Έχουμε βασικά ευθεία κατά μήκος της διαγωνίου με πολύ μικρές αποκλίσεις στα άνω και κάτω άκρα που δε θεωρούνται σημαντικές.

γ. Απαραμετρικός έλεγχος Kolmogorov-Smirnov για την κανονική κατανομή σε απλό δείγμα. Χρησιμοποιούμε μία παραλλαγή του ελέγχου αυτού που ελέγχει γενικά την κανονικότητα της κατανομής μιας μεταβλητής και όχι ειδικά την περίπτωση $N(0,1)$ (δηλαδή την τυπική κανονική με μέσο 0 και διασπορά 1). Βλέπε σχετικά Lilliefors (1967). Η p-value είναι 0.101 κι επομένως η υπόθεση της κανονικής κατανομής των residuals γίνεται δεκτή.

4.2.4 ΕΛΕΓΧΟΣ ΟΜΟΙΟΓΕΝΕΙΑΣ ΤΗΣ ΔΙΑΣΠΟΡΑΣ

Εχουμε ήδη αναφερθεί σε μία από τις βασικές υποθέσεις της Πολυμεταβλητής Γενικής Γραμμικής Υπόθεσης (Multivariate General Linear Hypothesis ή MGLH) : Οτι και οι 320 εξαρτημένες μεταβλητές έχουν κοινή διασπορά. Η υπόθεση αυτή πρέπει υποχρεωτικά να ελέγχεται για κάθε γραμμικό μοντέλο που χρησιμοποιούμε. Ετσι για το καλύτερο μέχρι στιγμής μοντέλο [βλέπε Παράρτημα B2] προχωρούμε στον έλεγχο σύμφωνα με το διάγραμμα των residuals συναρτήσε :

- α. Της πρόβλεψης του μοντέλου
- β. Κάθε μιας από τις ανεξάρτητες μεταβλητές

Τα διαγράμματα παρουσιάζονται στο Παράρτημα B2 και για το σκεπτικό που παίρνουμε κάθε ένα από αυτά βλέπε Draper & Smith (1981). Πριν τον έλεγχο εξαιρέσαμε, όπως και προηγουμένως, τους 4 outliers.

Το διάγραμμα α. έχει την κλασσική μορφή < που, ως γνωστόν [Draper & Smith (1981) κεφάλαιο 3], δείχνει ότι η υπόθεση της σταθερής (κοινής) διασποράς παραβιάζεται. Τα διαγράμματα β δείχνουν ότι η μεταβολή της διασποράς γίνεται συναρτήσε των μεταβλητών : Eu, G, INTENSIT, SUPPLIER. Οι απώλειες των μονοφασικών μοιάζει να μην επηρεάζουν σημαντικά τη διασπορά.

Για τη διόρθωση του προβλήματος αυτού πρέπει να χρησιμοποιηθεί η μέθοδος των Weighted Least Squares. Τα δεδομένα του δείγματος, δηλαδή, πρέπει να ομαδοποιηθούν σε ομάδες όπου η διασπορά των τριφασικών απωλειών διατηρείται σταθερή μέσα στην κάθε ομάδα. Αυτό ουσιαστικά σημαίνει ότι θεωρούμε ότι ο πίνακας συνδιακύμανσης των τριφασικών απωλειών V , αντί της μορφής $V=\sigma^2I$, έχει την μορφή :

$$V_{ij} = \begin{cases} \sigma_e^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Πριν προχωρήσουμε στην, αρκετά κουραστική, διαδικασία ομαδοποίησης των στοιχείων, ελέγχουμε την ακρίβειά της με εμπειρικά μέσα ειλημμένης εκτίμησής μας για ανομοιογένεια της διασποράς, με τον έλεγχο Bartlett [βλέπε συνόψιση των διαδοχικών βελτιώσεων του ελέγχου αυτού στο Bartlett (1947)]. Ο έλεγχος αυτός, σε περίπτωση δείγματος κανονικά κατανεμημένου, θεωρείται σαν ο ισχυρότερος που υπάρχει για το πρόβλημά μας [βλέπε Gartside (1972)]. Ελέγχουμε την ομοιογένεια κατά μήκος της έντασης του πεδίου που γίνεται η μονοφασική μέτρηση [παράρτημα B2] και απορρίπτουμε την υπόθεση της σταθερής διασποράς με σημαντικό αποτέλεσμα (χ^2 score=104.823 με DF=4 και p=0). Εδώ σημειώνουμε ότι,

ως γνωστόν, τα residuals αυτά καθ' εαυτά δεν είναι ποτέ ανεξάρτητα κατανεμημένα [βλέπε Belsley et al (1980), Anscombe & Tukey (1963), καθώς και Draper & Smith (1981)] και μάλιστα, αν R_j είναι το j -οστό residual, ισχύει : $\text{var}(R_j) = \sigma^2(I-h_j)$, όπου h_j είναι το j -οστό διαγώνιο στοιχείο του πίνακα $\mathbf{H} := \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Αυτός είναι ο λόγος που παίρνουμε διάγραμμα των residuals συναρτήσει της πρόβλεψης, όπως και ο λόγος που στον έλεγχο Bartlett ομαδοποιούμε τριφασικές απώλειες και όχι residuals. Αυτό δε δημιουργεί πάντως πρόβλημα στην εξέταση διαγραμμάτων των residuals επειδή θεωρείται [βλέπε Anscombe & Tukey] ότι όταν το πηλίκο (βαθμοί ελευθερίας των residuals) / (αριθμός των residuals) είναι μεγάλο (δηλαδή κοντά στην μονάδα), τα διαγράμματα αυτά δίνουν ακριβή εικόνα των πιθανών προβλημάτων. Στην περίπτωσή μας το πηλίκο αυτό είναι : $\frac{289}{316} = 0.91$ κι έτσι δεν ανησυχούμε.

4.2.5 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΤΗΣ ΔΙΑΣΠΟΡΑΣ

Προσπαθούμε να ομαδοποιήσουμε τις παρατηρήσεις των τριφασικών απωλειών και, χρησιμοποιώντας την τεχνική του γραμμικού μοντέλου, να κατασκευάσουμε κάποιο μοντέλο πρόβλεψης της διασποράς από τις ανεξάρτητες μεταβλητές. Δηλαδή ουσιαστικά υποθέτουμε ότι αν η παρατήρηση t είναι της μορφής :

$$(Y_t, \mathbf{X}_t), t=1, \dots, 316, \text{ τότε } \text{Var}(Y_t) = \sigma^2(\mathbf{X}_t)$$

Η υπόθεσή μας προκύπτει από τους ελέγχους που έγιναν προηγουμένως.

Κατ' αρχήν χρησιμοποιώντας τις 42 ομάδες από repeats [βλέπε και κεφάλαιο 4.2.2 προηγουμένως καθώς και Draper & Smith (1981), παράδειγμα 2.11] προσαρμόσαμε ένα γραμμικό μοντέλο πρόβλεψης της διασποράς ανά ομάδα repeats. Το αρχικό σετ των ανεξάρτητων μεταβλητών συμπεριελάμβανε τις Eu, G, INTENSIT, SUPPLIER και τις απώλειες της μονοφασικής μέτρησης. Κάνοντας backwards stepwise elimination οδηγηθήκαμε σε καλύτερο μοντέλο με μεταβλητές μονάχα τις απώλειες της μονοφασικής μέτρησης. Τα γεγονότα ότι :

- i. Απαιτήθηκε υπερβολικός αριθμός βημάτων.
- ii. Στο σετ των ανεξάρτητων μεταβλητών παρέμειναν μονάχα οι απώλειες της μονοφασικής μέτρησης, που από τα διαγράμματα του κεφαλαίου 4.2.4 γνωρίζαμε ότι δεν επηρέαζαν σημαντικά τη διασπορά.
- iii. Δεν έγινε δυνατό να απαλείψουμε τις συγγραμμικότητες μέσα στις απώλειες των μονοφασικών μετρήσεων.

τελικά μας οδήγησαν στην αναζήτηση νέου μοντέλου.

Δοκιμάσαμε επομένως ένα μοντέλο που εξ αρχής αποκλείει ότι οι απώλειες της μονοφασικής μέτρησης περικλείουν πληροφορία για τη διασπορά των απωλειών της τριφασικής μέτρησης. Σ' αυτό συνηγορούν :

- i. Τα διαγράμματα του κεφαλαίου 4.2.4
- ii. Το εξής σκεπτικό : Ανάμεσα σε δύο διαφορετικές μελέτες, οι διακυμάνσεις απωλειών των μονοφασικών μετρήσεων (δηλαδή οι κατανομές σε A, B+, κ.λ.π) πιστεύουμε ότι δε θα διαφέρουν. Επίσης δε θα διαφέρει ουσιωδώς και η διαδικασία της πλέξης (Θεωρώντας ότι η διαφορά στην μελέτη δεν επηρεάζει την συμπεριφορά των συνεργείων πλέξης). Όμως, οι διακυμάνσεις στην μέτρηση των τριφασικών απωλειών μπορεί να οφείλονται στους εξής παράγοντες : στη διαφορά μελέτης, στην σύνθεση του μετασχηματιστή ανάλογα με τις απώλειες της μονοφασικής, στις διαφορές της διαδικασίας πλέξης.

Αλλά, αν οι διακυμάνσεις των μονοφασικών απωλειών και οι διακυμάνσεις στη διαδικασία της πλέξης δε διαφέρουν ουσιωδώς ανάμεσα στις δύο ομάδες-μελέτες, ο μόνος λόγος που εξηγεί τη διαφορά στις διασπορές των απωλειών τριφασικής μέτρησης ανάμεσα στις δύο ομάδες είναι η διαφορά στην μελέτη.

Οι 9 ομάδες στοιχείων που διαμορφώνονται είναι :

α/α	SUPPLIER	INTENSIT	S ²	ONISXYS
1	1	12500	50.586	50
2	2	12500	50.309	50
3	1	15500	207.702	160
4	2	15500	234.598	160
5	3	15500	475.292	160
6	1	13200	125.005	100
7	1	14600	222.027	250
8	2	14600	116.950	250
9	2	14900	251.414	250

Η 8η ομάδα αποτελείται από 5 μονάχα μέλη (στα 316). Οι προσπάθειες κατασκευής μοντέλου πρόβλεψης της διασποράς με χρήση και των 9 ομάδων δεν οδήγησαν σε ικανοποιητικά αποτελέσματα. Παρατηρήσαμε ότι αν αγνοούσαμε την 8η ομάδα, μπορούσαμε να πάρουμε ένα καλό μοντέλο ($R^2=0.995$). Η χρήση του μοντέλου αυτού οδηγούσε σε αφύσικα μεγάλη πρόβλεψη της διασποράς ειδικά για την 8η ομάδα και μόνον ($S^2=234.187$ αντί της εκτιμώμενης από το δείγμα $S^2=116.950$).

Εδώ έπρεπε να αποφασίσουμε : ή το δείγμα ήταν πολύ μικρό και μεροληπούσε δίνοντας εκτίμηση της διασποράς μικρότερη της κανονικής ή κάτι ιδιαίτερο συμβαίνει ειδικά στις μετρήσεις της ομάδας αυτής που βελτιώνει τη διασπορά. Αναζητήθηκαν αιτίες που θ' απομόνωναν την ομάδα αυτή από τις υπόλοιπες (ανατρέχοντας στους αριθμούς των coils κ.λ.π) ανεπιτυχώς. Πιθανόν η διαφορά να οφειλόταν σε ιδιαίτερα καλές συνθήκες ανόπτησης για την ομάδα αυτή. Πιθανότερη, παρ' όλα αυτά, παραμένει η εξήγηση τη μεροληψίας που ενισχύεται και από τα γραφήματα της διασποράς που έγιναν. Τελικά, αποφασίσθηκε ν' αγνοηθεί η 8η ομάδα στο μοντέλο πρόβλεψης της διασποράς [βλέπε Παράρτημα Β4]. Αν έγινε λάθος αυτό θα είναι προς την συντηρητική κατεύθυνση, διότι τότε το μοντέλο θα δίνει πρόβλεψη των τριφασικών απωλειών με μεγαλύτερη ακρίβεια απ' όση εμείς πιστεύουμε ότι έχουμε.

Στο σημείο αυτό κάνουμε μερικές παρατηρήσεις σε σχέση με τη διασπορά των τριφασικών απωλειών που αξίζει να γίνουν. Πραγματικά, το μοντέλο δίνει αξιόλογα αποτελέσματα για τη βαρύτητα των προμηθευτών πρώτης ύλης καθώς και της μελέτης όσον αφορά την αύξηση ή ελάττωση της διασποράς των τριφασικών απωλειών. Μεγάλη διασπορά σημαίνει μειωμένη δυνατότητα πρόβλεψης των απωλειών της τριφασικής μέτρησης κι επομένως αυξημένο κόστος, αφού για να είναι η εταιρεία μέσα στις απαιτήσεις του πελάτη (π.χ. ΔΕΗ), θα χρειαστεί όσο αυξάνει η διασπορά τόσο να κατεβάζει την μέση τιμή των απωλειών της τριφασικής μέτρησης, ώστε να διατηρεί το ίδιο ποσοστό μετασχηματιστών μέσα στα προδιαγεγραμμένα όρια απωλειών. Το μοντέλο στ' οποίο καταλήξαμε είναι το ακόλουθο :

$$\text{VARIANCE}(S^2) = -539.0021 + 3.14023 \times Eu + 0.04256 \times \text{INTENSIT} + C(\text{SUPPLIER})$$

$$\begin{aligned} \text{όπου : } C(\text{SUPPLIER}) = & -91.27283 \text{ } \alpha \text{ί } \text{ι } \delta \eta \text{ί} \zeta \epsilon \alpha \delta \delta \rho \delta \alpha \beta \acute{\iota} \alpha \acute{\epsilon} \zeta \text{ USINOR} \\ & -78.05157 \text{ } \alpha \text{ί } \text{ι } \delta \eta \text{ί} \zeta \epsilon \alpha \delta \delta \rho \delta \alpha \beta \acute{\iota} \alpha \acute{\epsilon} \zeta \text{ ARMCO} \\ & 169.3244 \text{ } \alpha \text{ί } \text{ι } \delta \eta \text{ί} \zeta \epsilon \alpha \delta \delta \rho \delta \alpha \beta \acute{\iota} \alpha \acute{\epsilon} \zeta \text{ THYSSEN} \end{aligned}$$

Το μοντέλο αυτό έχει, όπως έχει ήδη αναφερθεί, $R^2=0.995$ και $MSE=296.274$. Παρατηρούμε ότι η χρησιμοποίηση μαγνητικού σιδήρου από την THYSSEN (Γερμανική) περίπου διπλάσιάζει την συνεισφορά του προμηθευτή στην αύξηση της διασποράς. Αυτό θ' άλλαζε αν η ΕΛΒΗΜ έπαιε να χρησιμοποιεί μαγνητική λαμαρίνα από την THYSSEN.

4.2.6 ΔΙΟΡΘΩΣΗ ΜΟΝΤΕΛΟΥ ΜΕ ΤΗΝ ΤΕΧΝΙΚΗ ΤΩΝ WEIGHTED

LEAST SQUARES

Από την μέχρι τώρα ανάλυση έχουμε καταλήξει, [βλέπε κεφάλαιο 4.2.4 προηγουμένως] ότι αν η τριφασική μέτρηση των απωλειών της t-στής μέτρησης ανήκει στην κ ομάδα σε σχέση με την ομαδοποίηση των διασπορών, και την συμβολίσουμε με Y_t , θα έχουμε :

$$\text{Var}(Y_t) = \sigma_{\epsilon}^2$$

Ας θεωρήσουμε τώρα την t παρατήρηση. Σύμφωνα με το μοντέλο αυτή θα έχει την μορφή :

$$Y_t = b_0 + b_1 * Eu + b_2 * G + b_3 * INTENSIT + b_4 (LOSSES11) + b_5 (LOSSES12) + b_6 (LOSSES13) + b_7 (LOSSES14) + b_8 (SUPPLIER) + \hat{a}_t$$

όπου οι 26 σταθερές πρέπει να εκτιμηθούν. Ο τρόπος που δουλεύουμε είναι ο εξής : Εισάγουμε μία ακόμη ανεξάρτητη μεταβλητή, την : $W_t := \sqrt{\frac{1}{\sigma_{\epsilon}^2}}$ και παίρνουμε το μοντέλο:

$$W_t Y_t = b_0 Y_t + b_1 W_t * Eu_t + b_2 W_t * G_t + b_3 W_t * INTENSIT_t + W_t * b_4 (LOSSES11) + W_t * b_5 (LOSSES12) + W_t * b_6 (LOSSES13) + W_t * b_7 (LOSSES14) + W_t * b_8 (SUPPLIER_t) + W_t * \hat{a}_t$$

Παρατηρούμε ότι τώρα $\text{Var}(W_t Y_t) = 1$. Αρα αρκεί να εκτιμήσουμε τις $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$ για το νέο μοντέλο, τ' οποίο δεν παραβιάζει την απαίτηση της ομοιογενούς κατανομής των διασπορών. Παρατηρούμε την σημασία που παίζει το μοντέλο πρόβλεψης της διασποράς, αφού αυτό δίνει τιμές στην W_t .

Για την κατασκευή του Weighted Least Sum Squares (WLSS) μοντέλου θέσαμε $W_t = \frac{1}{\sqrt{S_{\epsilon}^2}}$, όπου S_{ϵ}^2 οι τιμές του πίνακα του κεφαλαίου 4.2.5 για τις 9

ομάδες στοιχείων. Στο Παράρτημα Β6 δίνεται αναλυτικά το μοντέλο αυτό καθώς και οι υπολογισμοί όλων των στατιστικών ελέγχων που θα είναι χρήσιμοι στη συνέχεια. Το μοντέλο στ' οποίο καταλήξαμε έχει αρκετά καλό R^2 (0.952) και $MSE = 1.016$. Επίσης διαπιστώσαμε ότι περνάει τον έλεγχο ομοιγένειας των διασπορών του Bartlett με πάρα πολύ καλά αποτελέσματα (χ^2 score = 2.648 \hat{a} DF = 8, $p = 0.954$), και τέλος η υπόθεση της κανονικής κατανομής γίνεται δεκτή. Γενικά δεν παρουσιάστηκαν προβλήματα από την άποψη όσων έχουν εκτεθεί μέχρι εδώ. Βέβαια παρουσιάζει το πρόβλημα της ισχυρής σειριακής συσχέτισης (serial correlation).

4.2.7 ΕΥΡΕΣΗ ΤΩΝ OUTLIERS ΚΑΙ ΠΑΡΑΤΗΡΗΣΕΩΝ ΜΕ ΣΗΜΑΝΤΙΚΗ ΕΠΙΔΡΑΣΗ ΣΤΟ ΜΟΝΤΕΛΟ

Το θέμα έχει ήδη συζητηθεί στις παραγράφους ζ και η του κεφαλαίου 4.2 ("Πολυμεταβλητή Γενική Γραμμική Υπόθεση"). Αυθαίρετα αφαιρέσαμε τις παρατηρήσεις 28, 41, 75, 185 [βλέπε Παράρτημα Β1] από το αρχικά καλύτερο μοντέλο μας, πριν τη χρήση Weighted Least Squares (WLS). Τα studentized residuals ήταν 3.482, 3.068, 3.477, 2.969 αντίστοιχα. Για τον έλεγχο της ευστάθειας του μοντέλου, δηλαδή την μη ύπαρξη παρατηρήσεων που η απαλειφή τους από το δείγμα θα οδηγούσε σε ουσιώδη τροποποίηση του μοντέλου, χρησιμοποιήσαμε την απόσταση Cook [βλέπε Cook (1977), Cook & Weisberg (1992), καθώς και Draper & Smith (1981), κεφάλαιο 3]. Τα αποτελέσματα ήταν εξαιρετικά ικανοποιητικά, αφού στη χειρότερη περίπτωση η απόσταση Cook δε ξεπεράσε την τιμή 0.8 (θα προβληματιζόμασταν για τιμές πάνω από 2). Σχετικά, βλέπε Παράρτημα Β3 (εκτυπώσεις residuals με αποστάσεις Cook για το μοντέλο πριν τη χρήση WLS) και το Παράρτημα Β6 για διαγράμματα των αποστάσεων Cook συναρτήσει της πρόβλεψης στο WLS μοντέλο.

4.2.8 ΕΛΕΓΧΟΙ ΤΗΣ ΣΤΟΧΑΣΤΙΚΗΣ ΑΝΕΞΑΡΤΗΣΙΑΣ ΤΩΝ ϵ_i

Όπως έχουμε ήδη αναφέρει στο κεφάλαιο 4.2, παράγραφος θ., οι σχετικοί έλεγχοι γίνονται μέσω της ανάλυσης των residuals. Το συνηθέστερο πρόβλημα που παρουσιάζεται κατά την ανάλυση είναι η ανακάλυψη σειριακών συσχετίσεων, όπου δηλαδή κάθε residual εμπεριέχει ένα στατιστικά αξιόλογο μέρος πληροφορίας για το επόμενο του (lag1), ή για το μεθεπόμενο του (lag2), κ.ο.κ. Η αγνόηση του προβλήματος οδηγεί στη διατύπωση υπεραισιόδοξων εκτιμήσεων για το σφάλμα

της πρόβλεψης, σε μη αμερόληπτες εκτιμήτριες των συντελεστών του μοντέλου, και τέλος θέτει υπό αμφισβήτηση όσους ελέγχους έχουν προηγηθεί. Τέτοιο πρόβλημα αντιμετωπίζουν συχνά στα μοντέλα τους πρόβλεψης οι οικονομήτρες, λόγω της ιδιαίτερης φύσης των προβλημάτων που τους απασχολούν.

Θεωρείται ότι, η παρουσία στατιστικά σημαντικών σειριακών συσχετίσεων υποδηλώνει την αγνόηση από το μοντέλο κάποιας ουσιώδους μεταβλητής συσχετισμένης με το χρόνο. Περισσότερα στο θέμα αυτό αναφέρθηκαν ήδη στο κεφάλαιο 4.2 κι επίσης αναφέρονται παρακάτω. Χρησιμοποιήθηκαν οι εξής έλεγχοι :

α. Εξέταση της τιμής της στατιστικής D των Durbin-Watson [βλέπε Durbin-Watson (1950, 1951, 1952) καθώς και Draper & Smith (1981)]. Βρίσκουμε [Παράρτημα B6] 1.538, τιμή χαμηλότερη του κάτω ορίου κι επομένως απορρίπτεται η υπόθεση ότι για $s=1,2,\dots$, ταυτόχρονα, $\rho_s=0$ όπου $s=1,2,\dots$, ο αριθμός του Lag και ρ_s η αντίστοιχη σειριακή συσχέτιση.

β. Διαγράμματα ελέγχου των δειγματικών αυτοσυσχετίσεων μέχρι και Lag15 (ACF plot). Βλέπουμε [Παράρτημα B6] ότι $\rho_1=0.225$, $\rho_2=0.282$ και $\rho_3=0.209$ είναι έξω από τα ανώτερα επιτρεπτα όρια που θα επέτρεπαν να γίνει δεκτή η υπόθεση ότι αυτές έχουν τιμή 0. (Τα όρια είναι $2*SE$, δηλαδή 0.112, 0.118 και 0.126 αντίστοιχα). Επίσης φαίνεται ότι όλες οι υπόλοιπες σειριακές αυτοσυσχετίσεις είναι στατιστικά σημαντικές.

γ. Απαραμέτρικός έλεγχος Wald-Wolfowitz [ο σχετικός έλεγχος παρουσιάστηκε αρχικά στο Swed & Eisenhart(1943)] για τον εντοπισμό διαδρομών που υποδηλώνουν έλλειψη τυχαιότητας στην ακολουθία των residuals (κι επομένως σειριακές συσχετίσεις). Στο Παράρτημα B6 ο σχετικός έλεγχος αποδέχεται την τυχαιότητα σε επίπεδο ελέγχου $\alpha=0.01$ αλλά την απορρίπτει σε επίπεδο ελέγχου $\alpha=0.05$.

δ. Στο Παράρτημα B6 παρουσιάζεται scatterplot των residuals συναρτήσει των residuals με lag1. Οπτικά δεν είναι σαφής η ύπαρξη γραμμικής σχέσης που θα υποδήλωνε σειριακή συσχέτιση lag1.

ε. Τέλος προχωρήσαμε σε Bartlett X^2 -έλεγχο της υπόθεσης $H_0: \rho_1=0$ με ταυτόχρονη χρήση πίνακα πιθανοτήτων Bonferroni. Το αποτέλεσμα είναι στατιστικά σημαντικό (απορρίπτεται η υπόθεση H_0 με επίπεδο σημαντικότητας $p=0.001$).

Το συμπέρασμα όπου καταλήγουμε από τους παραπάνω ελέγχους είναι ότι επιτρέπεται να θεωρήσουμε $\rho_s=0$ για $s>3$, αλλά όχι $s=1$ ή 2 ή 3 . Δηλαδή ο πίνακας

συνδιακύμανσης V έχει τελικά, μετά την εφαρμογή WLS, την μορφή που δίνεται στο κεφάλαιο 4.2, παράγραφος θ. Επομένως απαιτείται η λήψη διορθωτικών μέτρων στο μοντέλο του Παραρτήματος Β2.

4.2.9 ΔΙΟΡΘΩΣΗ ΤΩΝ ΣΕΙΡΙΑΚΩΝ ΣΥΣΧΕΤΙΣΕΩΝ

Σε περίπτωση στατιστικά σημαντικών σειριακών συσχετίσεων το πρώτο που επιχειρεί κάποιος είναι να εισάγει μια επιπλέον μεταβλητή (ή και περισσότερες) που να είναι συσχετισμένη με το χρόνο και που, ταυτόχρονα, να μην οδηγεί σε oversaturation του μοντέλου, δηλαδή σε παραβίαση των αρχικών ελέγχων καλής προσαρμογής και ανάλυσης της διασποράς που ήδη έχουν πραγματοποιηθεί. Οι μόνες διαθέσιμες στην περίπτωση μας ήταν ο αριθμός εντολής (μεταβλητή ORDER) και ο αύξων αριθμός στην παρτίδα του coil της μαγνητικής λαμαρίνας (μεταβλητή NOCOIL). Η χρήση τους οπωσδήποτε είναι πολύ ελλειπές υποκατάστατο χρήσης ουσιαστών μεταβλητών συσχετισμένων με το χρόνο. Τελικά, και μετά από σχετική επεξεργασία, αφαιρέσαμε την μεταβλητή ORDER από το σύνολο των ανεξάρτητων μεταβλητών επειδή, αν και συνεισέφερε ελαφρά στη βελτίωση των αυτοσυσχετίσεων, χαλούσε την κανονικότητα των residuals. Έτσι κι αλλιώς ήταν μια μεταβλητή χωρίς "φυσική ερμηνεία", που είχε προστεθεί διότι δεν υπήρχαν άλλες συσχετισμένες με το χρόνο μεταβλητές.

Στο Παράρτημα Β5 παρουσιάζονται κατά σειρά τα εξής :

- α. Το νέο μοντέλο όπου στο προηγούμενο σετ των ανεξάρτητων μεταβλητών (Eu, G, INTENSIT, LOSSES11, LOSSES12, LOSSES13, LOSSES14, SUPPLIER) έχουν προστεθεί ο αριθμός coil στην παρτίδα λαμαρίνας που χρησιμοποιήθηκε για τον συγκεκριμένο μετασχηματιστή (NOCOIL).
- β. Κατάλογος των outliers. Απορρίπτονται με σχετικά αυθαίρετο τρόπο 4 από αυτούς (οι μεγαλύτερες ή σχεδόν ίσες με 3 τιμές).
- γ. Backward Stepwise Elimination που οδηγεί στην απόρριψη του NOCOIL από τις ανεξάρτητες μεταβλητές.

Παρατηρούμε επομένως ότι καταλήγουμε στο μοντέλο του Παραρτήματος Β2 απ' όπου ξεκινήσαμε.

Στο Παράρτημα Β6 έχουμε τη διόρθωση του μοντέλου μέσω WLS για την ομογενοποίηση των διασπορών καθώς και τους σχετικούς ελέγχους. Ολα παρουσιάζονται ομαλά εκτός από το πρόβλημα του ACF graph. Έχουμε αυτοσυσχετίσεις lag1, lag2 και lag3 που δεν μπορούν ν' αγνοηθούν, ενώ οι αυτοσυσχετίσεις μεγαλύτερης τάξης μπορούν να θεωρηθούν ασήμαντες.

Επομένως προχωράμε στο "τελευταίο καταφύγιο" για την άρση των αυτοσυσχετίσεων, που είναι η μέθοδος Cochrane-Orcutt [βλέπε σχετικά Cochrane-Orcutt (1949)]. Ο τρόπος λειτουργίας της μεθόδου είναι ο εξής : Εστω το μοντέλο

$$Y_t = b_0 + b_1 X_{t1} + \dots + b_k X_{tk} + \hat{a}_t, \hat{a}_t \approx N(0, \sigma^2)$$

όπου $\text{cov}(\hat{a}_t, \hat{a}_{t+s}) \neq 0$. Στην περίπτωση μας $s=1 \ \& \ 2^*$. Εστω ρ_s η συσχέτιση lags. Τότε, σχηματίζουμε το μοντέλο (για $s=1 \ \& \ 2$).

$$\tilde{O}_t - \tilde{\eta}_1 * \tilde{O}_{t-1} - \tilde{\eta}_2 * \tilde{O}_{t-2} = b_0 * (1 - \tilde{\eta}_1 - \tilde{\eta}_2) + b_1 * (X_{t1} - \tilde{\eta}_1 * X_{t-1,1} - \tilde{\eta}_2 * X_{t-2,1}) +$$

$$\dots + b_k * (X_{tk} - \tilde{\eta}_1 * X_{t-1,k} - \tilde{\eta}_2 * X_{t-2,k}) + \hat{a}_t$$

$$Z_t = Y_t - \tilde{\eta}_1 * \tilde{O}_{t-1} - \tilde{\eta}_2 * \tilde{O}_{t-2}$$

Ας ονομάσουμε

$$W_{ti} = (X_{ti} - \tilde{\eta}_1 * X_{t-1,i} - \tilde{\eta}_2 * X_{t-2,i}) \quad i = 1, \dots, k$$

Ο αλγόριθμος :

Βήμα 1 : Αντικατάστησε τα ρ_s με τις δειγματικές εκτιμήτριές τους

Βήμα 2: Χρησιμοποίησε γραμμικό μοντέλο για να προσαρμόσεις τις ανεξάρτητες μεταβλητές W_i , $i=1, \dots, k$, στην Z

Βήμα 3: Χρησιμοποίησε στο αρχικό σου μοντέλο (Y και X_i) τις εκτιμήτριες των β από το Βήμα 2.

συνήθως οδηγεί στην απαλειφή των σειριακών συσχετίσεων.

* *Διάφοροι έλεγχοι για την ύπαρξη αυτοσυσχετίσεων lag3, lag2 και lag1. Η μέθοδος Cochrane-Orcutt για lag1 & lag2. Έλεγχος για την ύπαρξη αυτοσυσχετίσεων lag1 & lag2. Η μέθοδος Cochrane-Orcutt για lag1 & lag2. Η μέθοδος Cochrane-Orcutt για lag1 & lag2. Η μέθοδος Cochrane-Orcutt για lag1 & lag2.*

Υπάρχει και επαναληπτική, ακριβέστερη εκδοχή του αλγόριθμου, η οποία όμως δεν ήταν δυνατό να χρησιμοποιηθεί λόγω περιορισμών στην υπολογιστική μας ισχύ. Στο Παράρτημα Β7 γίνεται η εκτίμηση των παραμέτρων β μέσω της μεθόδου Cochran-Orcutt. Παρατηρούμε το υψηλό R^2 (0.982) και το ικανοποιητικό $MSE=212.91$. Επίσης, στο Παράρτημα Β7 υπάρχουν :

- i. Πίνακας των Residuals του μοντέλου
- ii. Διαγράμματα των Residuals με κάθε ανεξάρτητη μεταβλητή του μοντέλου
- iii. Απαραμετρικός έλεγχος κανονικότητας των Residuals (Kolmogorov-Smirnov/Lilliefors)
- iv. Απαραμετρικός έλεγχος διαδρομών Wald-Wolfowitz για διαδρομές θετικών και αρνητικών προσήμων στα residuals.
- v. Διάγραμμα ελέγχου των αυτοσυσχετίσεων (autocorrelations plot ή ACF plot)

Τέλος δίνεται το τελικό μοντέλο που συνάγεται από τα στοιχεία της παραγωγής μετασχηματιστών για το 1991, καθώς και ο τύπος για τη πρόβλεψη των απωλειών των τριφασικών μετρήσεων όπως προκύπτει μετά την εξάλειψη όλων των προβλημάτων.

$$LOSSESF = -644.28165 + 4.03346 * Eu + 1.15608 * G + 0.0332 * INTENSIT + C_{11} + C_{12} + C_{13} + C_{14} + C_S$$

όπου οι σταθερές C_{11} , C_{12} , C_{13} , C_{14} , C_S υπολογίζονται ως εξής :

- $C_{11} = 0.80283$ αν ο πρώτος πυρήνας καταμετρήθηκε σαν Α στην μονοφασική μέτρηση
 -2.31728 αν ο πρώτος πυρήνας καταμετρήθηκε σαν Β+ στην μονοφασική μέτρηση
 -5.61942 αν ο πρώτος πυρήνας καταμετρήθηκε σαν Β στην μονοφασική μέτρηση
 -0.43501 αν ο πρώτος πυρήνας καταμετρήθηκε σαν C+ στην μονοφασική μέτρηση

-1.34372 αν ο πρώτος πυρήνας καταμετρήθηκε σαν C στην μονοφασική μέτρηση

0 αν ο πρώτος πυρήνας καταμετρήθηκε σαν D στην μονοφασική μέτρηση

$C_{12} =$ -5.60681 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν A στην μονοφασική μέτρηση

-6.56223 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν B+ στην μονοφασική μέτρηση

-5.45390 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν B στην μονοφασική μέτρηση

-5.72362 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν C+ στην μονοφασική μέτρηση

-0.06036 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν C στην μονοφασική μέτρηση

0 αν ο δεύτερος πυρήνας καταμετρήθηκε σαν D στην μονοφασική μέτρηση

$C_{13} =$ -11.98336 αν ο τρίτος πυρήνας καταμετρήθηκε σαν A στην μονοφασική μέτρηση

-4.98722 αν ο τρίτος πυρήνας καταμετρήθηκε σαν B+ στην μονοφασική μέτρηση

-4.93916 αν ο τρίτος πυρήνας καταμετρήθηκε σαν B στην μονοφασική μέτρηση

-4.11898 αν ο τρίτος πυρήνας καταμετρήθηκε σαν C+ στην μονοφασική μέτρηση

-0.66414 αν ο τρίτος πυρήνας καταμετρήθηκε σαν C στην μονοφασική μέτρηση

0 αν ο τρίτος πυρήνας καταμετρήθηκε σαν D στην μονοφασική μέτρηση

$C_{14} =$ -17.52769 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν A στην μονοφασική μέτρηση

-5.68121 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν B+ στην μονοφασική μέτρηση

-5.03654 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν B στην μονοφασική μέτρηση

-6.55401 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν C+ στην μονοφασική μέτρηση

-1.93805 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν C στην μονοφασική μέτρηση

0 αν ο τέταρτος πυρήνας καταμετρήθηκε σαν D στην μονοφασική μέτρηση

$C_5 =$ -26.44678 αν ο προμηθευτής είναι η USINOR
-36.33447 αν ο προμηθευτής είναι η ARMCO
0 αν ο προμηθευτής είναι η THYSSEN

4.3 ΕΛΕΓΧΟΙ ΠΡΟΒΛΕΨΗΣ ΑΠΩΛΕΙΩΝ ΤΟΥ ΜΟΝΤΕΛΟΥ ΠΡΟΣ ΤΗΝ ΠΡΟΒΛΕΨΗ ΑΠΩΛΕΙΩΝ ΤΗΣ ΜΕΛΕΤΗΣ

Σύμφωνα με τα στοιχεία του Παραρτήματος Β8* προχωρήσαμε στην σύγκριση των προβλεπομένων απωλειών από το μοντέλο, και των προβλεπομένων απωλειών από την μελέτη. Η ιδέα είναι η εξής: Παίρνουμε ένα δείγμα μετασχηματιστών που κατασκευάστηκαν σε μία χρονική περίοδο. Για κάθε μετασχηματιστή ορίζουμε σαν: $R_1 =$ Απώλειες Τριφασικής του Μετασχηματιστή - Προβλεπόμενες απώλειες από το μοντέλο.

$R_2 =$ Απώλειες Τριφασικής του Μετασχηματιστή - Προβλεπόμενες απώλειες από την μελέτη.

Ας ονομάσουμε $\mu_1 = E(R_1)$ την μέση τιμή της τυχαίας μεταβλητής R_1 και, αντίστοιχα $\mu_2 = E(R_2)$ την μέση τιμή της τυχαίας μεταβλητής R_2 .

Οποιαδήποτε "καλή" πρόβλεψη (είτε μοντέλο είτε μελέτη) θα έπρεπε να δίνει $\mu_1 = 0$ και $\mu_2 = 0$. Αυτό θα σήμαινε ότι αποκλίσεις από τις προβλεπόμενες απώλειες οφείλονται στην επίδραση τυχαίων παραγόντων. Δηλαδή, ότι δεν υπάρχει κάποιος μόνιμος παράγων που η επίδρασή του να δημιουργεί συστηματική απόκλιση ανάμεσα στο παρατηρούμενο και στο προβλεπόμενο.

Επομένως προχωράμε σε στατιστικό έλεγχο της υπόθεσης (1) $H_{01}: \mu_1 = 0$ εναντίον της $H_{11}: \mu_1 \neq 0$ και της $H_{02}: \mu_2 = 0$ εναντίον της $H_{12}: \mu_2 \neq 0$, ώστε να δούμε αν το δείγμα υποστηρίζει, και πόσο τις παραπάνω υποθέσεις. Εγινε αρχικά έλεγχος με δείγμα που αφορούσε τα στοιχεία του 1991, και κατόπιν έλεγχος με δείγμα που αφορούσε στοιχεία του 1992.

I. Για το δείγμα του 1991 κάναμε δύο ειδών ελέγχους.

* Όι RESID1 άιάόΥñάάάέ όδά residuals δϊδ δñϊέγδδϊδί άδϋ δϊ ιϊϊδΎεϊ, άίβ δϊ RESID2 άιάόΥñάάάέ όδά residuals δϊδ δñϊέγδδϊδί άδϋ όςϊ ιάεΎδς.

A. Απαραμετρικός Έλεγχος Προσήμου

Επειδή στο δείγμα έχουμε 9 ομάδες με διαφορετική συνάρτηση κατανομής πιθανότητας των τριφασικών απωλειών για την κάθε ομάδα (τουλάχιστον λόγω των διαφορών στη διασπορά από ομάδα σε ομάδα), κι επειδή θέλουμε να μην σπάσουμε το δείγμα (μεγάλο δείγμα σημαίνει μεγάλη ακρίβεια), προχωράμε στο απαραμετρικό sign test. Εάν είναι m_1 η διάμεσος της τυχαίας μεταβλητής R_1 (δηλαδή το m_1 ορίζεται από την σχέση $P(R_1 \leq m_1) = 1/2$) και m_2 η διάμεσος της τυχαίας μεταβλητής R_2 το sign test ελέγχει τις υποθέσεις:

$$\begin{array}{llll} & H_{01}: \mu_1=0 & \text{vs} & H_{11}: \mu_1 \neq 0 \\ \text{και} & H_{02}: \mu_2=0 & \text{vs} & H_{02}: \mu_2 \neq 0 \end{array}$$

Αν οι R_1 και R_2 είναι κανονικά κατανεμημένες, τότε ο έλεγχος αυτός είναι ταυτόχρονα και έλεγχος των μέσων (διάμεσος και μέσος συμπίπτουν). Χρησιμοποιούμε τον έλεγχο αυτό κυρίως επειδή είναι εφαρμόσιμος σε καταστάσεις όπου η κατανομή της R_1 (και R_2) αλλάζει από παρατήρηση σε παρατήρηση χωρίς όμως να αλλάζει η διάμεσος, όπως ακριβώς συμβαίνει στο πρόβλημά μας.

Τα αποτελέσματα είναι:

Για την R_1 : Επίπεδο σημαντικότητας του ελέγχου $\alpha=0.16$.

Η υπόθεση $\mu_1=0$ γίνεται δεκτή για $\alpha=0.01$ ή $\alpha=0.05$

Για την R_2 : Επίπεδο σημαντικότητας του ελέγχου $\alpha=0$

Η υπόθεση $\mu_2=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Στηριγμένοι στο sign test, επομένως, δεν μπορούμε να απορρίψουμε την υπόθεση ότι το προτεινόμενο μοντέλο πρόβλεψης δίνει "καλές" προβλέψεις. Αντίθετα, το ίδιο test συμπεραίνει ότι η πρόβλεψη απωλειών της μελέτης είναι ανεπαρκής.

B. Έλεγχος Μέσου Κανονικής με Αγνώστη Διασπορά

Για μεγάλο δείγμα τα R_1 είναι "σχεδόν" ανεξάρτητα κανονικά κατανεμημένα. Άρα, αν δε σπάσουμε το δείγμα μας, μπορούμε να προχωρήσουμε σε αρκετά ακριβή έλεγχο της (1) χρησιμοποιώντας t-test. Πρόβλημα και πάλι εδώ είναι οι διαφορετικές διασπορές από ομάδα σε ομάδα. Λύνουμε το πρόβλημα αυτό θεωρώντας τα $R'_{1i} := R_{1i} W_{\hat{\sigma}}$ και $R'_{2i} := R_{2i} W_{\hat{\sigma}}$, όπου $W_{\hat{\sigma}}$ είναι το weight της ομάδας στην οποία ανήκει η παρατήρηση i , και χρησιμοποιώντας σε δείγμα τα R'_1 και R'_2 .

Τα αποτελέσματα του ελέγχου αυτού είναι:

Για την R_1 : Επίπεδο σημαντικότητας του ελέγχου $p=0.479$

Η υπόθεση $\mu_1=0$ γίνεται δεκτή

Για την R_2 : Επίπεδο σημαντικότητας του ελέγχου $p=0$

Η υπόθεση $\mu_2=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Δηλαδή και το t-test επιβεβαιώνει τα συμπεράσματα του sign test.

II. Για το δείγμα του 1992 κάναμε τους ίδιους ελέγχους.

A. Απαραμετρικός Έλεγχος Προσήμου

Χρησιμοποιήθηκε η ίδια θεωρία με τον προηγούμενο αντίστοιχα έλεγχο. Για αυτό θα προχωρήσουμε απευθείας στα αποτελέσματα. Έχουμε λοιπόν:

Για την R_1 : Επίπεδο σημαντικότητας του ελέγχου $p=0$.

Η υπόθεση $m_1=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Για την R_2 : Επίπεδο σημαντικότητας του ελέγχου $p=0$

Η υπόθεση $m_2=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Τα αποτελέσματα του test δείχνουν ότι το μοντέλο δε δίνει καλές προβλέψεις για το δείγμα του 1992. Αυτό μπορεί να οφείλεται στο ότι υπήρξε κάποια αλλαγή στην παραγωγική διαδικασία από το 1991 στο 1992. Στηριζόμενοι στο sign test παρατηρούμε ότι το μοντέλο πρόβλεψης απορρίπτεται. Όσον αφορά στην μελέτη, το test επιβεβαιώνει ό,τι προέκυψε από τους αντίστοιχους ελέγχους για το δείγμα του 1991. Δηλαδή, ότι η μελέτη είναι ανεπαρκής στην πρόβλεψη των τριφασικών απωλειών.

B. Έλεγχος Μέσου Κανονικής με Αγνώστη Διασπορά

Επίσης χρησιμοποιήθηκε η ίδια θεωρία με τον προηγούμενο αντίστοιχα έλεγχο. Αρα τ' αποτελέσματα που παίρνουμε είναι:

Για την R_1 : Επίπεδο σημαντικότητας του ελέγχου $p=0$.

Η υπόθεση $\mu_1=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Για την R_2 : Επίπεδο σημαντικότητας του ελέγχου $p=0$

Η υπόθεση $\mu_2=0$ απορρίπτεται για οποιοδήποτε επίπεδο ελέγχου

Επίσης σ' αυτό τον έλεγχο φαίνεται ότι το μοντέλο δεν επιβεβαιώνεται από τα στοιχεία του 1992. Παρατηρούμε [βλέπε Παράρτημα B8] ότι η t-value για το μοντέλο

είναι $11.402 > 0$. Αυτό σημαίνει ότι ο έλεγχος κάνει δεκτή την υπόθεση ότι ο μέσος είναι θετικός, άρα βγαίνει το συμπέρασμα πως οι προβλεπόμενες απώλειες του μοντέλου είναι χαμηλότερες από τις πραγματικές. Οσον αφορά στην μελέτη, το t -test επιβεβαιώνει το $sign$ test.

Τελικά, έχουμε τα εξής:

i. Οσον αφορά στο δείγμα του 1991, το μοντέλο πρόβλεψης έχει προέλθει από το ίδιο δείγμα οπότε δεν περιμέναμε να το απορρίψουν οι έλεγχοι. Επομένως δεν μπορούμε ακόμα να συμπεράνουμε ότι το μοντέλο είναι καλό. Παρ' όλα αυτά, τα συμπεράσματα για την πρόβλεψη της μελέτης είναι στατιστικά σημαντικά, αφού με έμφαση απορρίπτουν την καλή προβλεψιμότητά της. Δεδομένου του μεγάλου μεγέθους του δείγματος, της χρονικής διάρκειας στην οποία αυτό αναφέρεται (ολόκληρο έτος, κι επομένως δεν υπάρχουν προβλήματα εποχιακότητας) και της ταυτόσημης απάντησης και των δύο ελέγχων, το συμπέρασμα είναι ότι οι χρησιμοποιούμενες μελέτες είναι στατιστικά ανεπαρκείς στην πρόβλεψη των πραγματικών τριφασικών απωλειών.

ii. Οσον αφορά στο δείγμα του 1992, παρατηρούμε ότι το μοντέλο πρόβλεψης δε δίνει καλές εκτιμήσεις των τριφασικών απωλειών. Αυτό μπορεί να οφείλεται σε κάποια αλλαγή στην παραγωγική διαδικασία, και όπως είναι γνωστό ένα στατιστικό μοντέλο πρόβλεψης είναι ευαίσθητο σε τέτοιου είδους αλλαγές. Το πιθανότερο όμως είναι το εξής, που δεν αποκλείει και την προηγούμενη εξήγηση,: το $weighted$ μοντέλο που χρησιμοποιήσαμε στους ελέγχους στηρίζεται στο μοντέλο πρόβλεψης της διασποράς που είχε προέλθει από το δείγμα του 1991.

Το δείγμα του 1991 αποτελούνταν από 9 σημεία (το κάθε ένα εκπροσωπούσε και μία ομάδα με σταθερή διασπορά). Θεωρούμε τον αριθμό των σημείων αυτών ανεπαρκή για ένα καλό μοντέλο πρόβλεψης της διασποράς, και, κατά πάσα πιθανότητα εκεί οφείλεται η ανεπάρκεια του παρόντος μοντέλου να ερμηνεύσει τα στοιχεία του 1992.

Το συμπέρασμα εδώ είναι ότι, για την πραγματική του εφαρμογή στην παραγωγική διαδικασία, το μοντέλο οφείλει να υφίσταται περιοδικές αναθεωρήσεις, με τον ίδιο ακριβώς τρόπο, και γι' ανάλογους λόγους εξ αιτίας των οποίων, αναθεωρούνται και τα στατιστικά διαγράμματα ελέγχου, διαφόρων παραγωγικών διαδικασιών. Θα πετυχαίνουμε έτσι, να "πιάνει" το μοντέλο αλλαγές στην τεχνολογία παραγωγής, τροποποιήσεις στην ποιότητα της πρώτης ύλης, κλπ. καθώς και, να "αυτοβελτιώνεται", μια και θα βελτιώνεται με τον τρόπο αυτό το μοντέλο πρόβλεψης

της διασποράς, που πάνω του στηρίζεται το τελικό μοντέλο πρόβλεψης των τριφασικών απωλειών.

4.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

1. Η εργασία που έγινε εξάντλησε τις δυνατότητες πρόβλεψης των τριφασικών απωλειών που δίνουν το σύνολο των στοιχείων όπως κρατιούνται από την ΕΛΒΗΜ (περίοδος 1991). Παρ' όλα αυτά υπάρχουν ελλείψεις που φαίνονται κυρίως από τις σειριακές συσχετίσεις. Είναι δυνατή η σημαντική βελτίωση του μοντέλου αλλά υπό τον όρο ότι θα παίρνονται μετρήσεις όλων των μεταβλητών που χρησιμοποιούνται, ώστε να έχουμε πραγματικές τιμές.

2. Το παρόν μοντέλο εξηγεί πάνω από 98% των παρουσιαζομένων απωλειών στην τριφασική μέτρηση και άρα είναι ικανοποιητικό. Πιστεύουμε όμως, ότι κάτω από την υπόθεση του 1 παραπάνω, μπορούμε να πετύχουμε εντυπωσιακή βελτίωση του τυπικού σφάλματος.

3. Εμμεσα το παρόν μοντέλο υποθέτει ότι μακροσκοπικά η παραγωγική διαδικασία της ΕΛΒΗΜ βρίσκεται σε κατάσταση ελέγχου, λόγω κυρίως του τρόπου εφαρμογής της τεχνικής WLS. Αυτό δεν είναι απολύτως ακριβές, όπως υποδηλώνει η παρουσία outliers αλλά και η εμπειρία μας από το γρέζι και τη διαδικασία της πλέξης. Επίσης πρέπει να παρατηρήσουμε ότι, ενώ η μέθοδος ανάλυσης από την οποία προέκυψε το μοντέλο πρόβλεψης των τριφασικών απωλειών είναι σωστή, το μοντέλο θα πρέπει ν' αναπροσαρμοστεί. Είναι ακόμη φανερό ότι όσον αφορά στα στοιχεία του 1991, το μοντέλο πρόβλεψης της διασποράς ήταν σωστό, αλλά, -όπως φαίνεται από τους ελέγχους πρόβλεψης των απωλειών του μοντέλου, με χρήση στοιχείων του 1992- είναι αναγκαία η αναπροσαρμογή του. Επομένως θα έπρεπε να τυποποιηθούν και να συστηματοποιηθούν οι ανάλογες διαδικασίες ελέγχου.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Μηλολιδάκης Κωνσταντίνος : Ελεγχος Ποιότητας, Πολυτεχνείο Κρήτης, Χανιά, 1987.
2. Μηλολιδάκης Κωνσταντίνος : Εισαγωγή στην Στατιστική, Πολυτεχνείο Κρήτης, Χανιά, 1988.
3. Wilkinson, Leland : SYSTAT: The system for Statistics, Evanston IL, SYSTAT, Inc., 1990.
4. N. Draper & H. Smith : Applied Regression Analysis, Wiley, NY, 1981 (2nd ed).
5. G.A.F. Seber : Linear Regression Analysis, Wiley, NY, 1971.
6. R.D. Cook & S. Weisberg : Residuals & Influence in Regression, Chapman & Hall, NY, 1992.
7. Lilliefors, HW : On the Colmogorov-Smirnov test for normality with mean and variance unknown, Journal of the American Statistical Assosiation, 64, pp. 399-402, 1967.
8. Bartlett, MS : Multivariate Analysis, Journal of the Royal Statistical Society, Series B, 9, pp. 176-197, 1947.

9. P.S. Gartside : A study of Methods for Comparing Several Variances, *Journal of the American Statistical Assosiation*, 67, pp. 342-346, 1972.
10. D. Belsley & E. Kuh & R. Welsch : *Regression Diagnostics; Identifying Influential Data and Sources of Collinearity*, Wiley, NY, 1980.
11. Anscombe, F.J. & J.W. Tukey : The Examination of Analysis of residuals, *Technometrics*, 5, pp. 141-160, 1963.
12. R.D. Cook : Detection of Influential Observations in Linear Regression, *Technometrics*, 19, pp. 15-18, 1977.
13. J. Durbin & G.S. Watson : Testing for Serial Correlation in Least Squares Regression, I, *Biometrika*, 37, pp. 409-428, 1950.
14. J. Durbin & G.S. Watson : Testing for Serial Correlation in Least Squares Regression, II, *Biometrika*, 38, pp. 159-178, 1951.
15. J. Durbin & G.S. Watson : Testing for Serial Correlation in Least Squares Regression, III, *Biometrika*, 58, pp. 1-19, 1971.
16. F. Swed & C. Eisenhart : Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *Annals of Mathematical Statistics*, 14, pp. 66-87, 1943.
17. D. Cochrane & G. Orcutt : Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Assosiation*, 44, pp. 42-61, 1949.
18. J. Kmenta : *Elements of Econometrics*, Macmillan, NY, 1971.
19. S. Chatterjee & B. Price : *Regression Analysis by Example*, Wiley, NY, 1977.