



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Στατιστική Ανάλυση για την πρόβλεψη των εκπομπών
αερίων του θερμοκηπίου στην Ευρωπαϊκή Ένωση

Statistical Analysis for predicting emissions of
greenhouse gases in the European Union

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΜΑΡΙΑ ΔΗΜΗΤΡΙΑΔΟΥ

Επιβλέπων: Δημήτρης Φουσακάκης,

Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2015

Περίληψη

Η παρούσα Διπλωματική εργασία εκπονήθηκε στα πλαίσια των σπουδών μου για την απόκτηση του προπτυχιακού διπλώματος της σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών (Σ.Ε.Μ.Φ.Ε.) του Εθνικού Μετσόβιου Πολυτεχνείου (Ε.Μ.Π). Σκοπός της είναι μέσα από την εφαρμογή δύο διαφορετικών μεθόδων ανάλυσης παλινδρόμησης να επιτευχθεί άμεση πρόβλεψη των εκπομπών των αερίων του θερμοκηπίου στην Ευρωπαϊκή Ένωση.

Η υπερθέρμανση του πλανήτη δεν είναι μια νέα μελέτη στην επιστήμη. Κάποιοι λένε ότι είναι μια φυσική διαδικασία και ότι ανέκαθεν υπήρχαν αέρια του θερμοκηπίου, όμως η ποσότητα των αερίων στην ατμόσφαιρα έχει ανεβεί στα ύψη στην πρόσφατη ιστορία. Αν συνεχιστούν αυτές οι τάσεις, οι επιστήμονες φοβούνται ότι οι χειρότερες επιπτώσεις της υπερθέρμανσης του πλανήτη - ακραία καιρικά φαινόμενα, η άνοδος της στάθμης της θάλασσας, των φυτών και των ζώων που έχουν εξαφανιστεί, η όξυνση των ωκεανών, σημαντικές αλλαγές στο κλίμα και η άνευ προηγουμένου κοινωνική αναταραχή - θα είναι αναπόφευκτες.

Σε παγκόσμιο επίπεδο, η παραγωγή των αερίων του θερμοκηπίου είναι μια πηγή σοβαρής ανησυχίας. Αναπνέουμε καθημερινά περίπου 14 κιλά αέρα, πίνουμε 2 κιλά νερό και τρώμε 1,5 κιλό τροφή. Λόγω του αντίκτυπου που έχει η ατμοσφαιρική ρύπανση στην υγεία, την ποιότητα ζωής του ανθρώπου αλλά και στο περιβάλλον γενικότερα, κρίνεται σημαντική η ανάλυση και η πρόβλεψη των επιπέδων των ατμοσφαιρικών ρύπων. Έτσι, μπορεί να ελεγχθεί και με κατάλληλα μέτρα να βελτιωθεί η παρούσα κατάσταση.

Στόχος της παρούσας διπλωματικής εργασίας είναι να προτείνει ένα αξιόπιστο μοντέλο πρόβλεψης των ατμοσφαιρικών ρύπων. Τα δεδομένα για τις εκπομπές των ατμοσφαιρικών ρύπων δεν είναι άμεσα διαθέσιμα για όλα τα έτη. Η Eurostat έχει δεδομένα μέχρι το έτος $t-3$ (όπου t το τρέχον έτος), ενώ άλλες βάσεις δεδομένων όπως η UNFCCC διαθέτει τα δεδομένα αυτά 18 μήνες μετά το τέλος του κάθε έτους. Επομένως, αυτή τη στιγμή στις δύο βάσεις δεδομένων υπάρχουν στοιχεία για τις εκπομπές μέχρι το 2012. Σκοπός είναι η πρόβλεψη των εκπομπών για το 2013 μέσω μοντέλων παλινδρόμησης. Με βάση αυτά τα μοντέλα και έχοντας αργότερα πιο πρόσφατα δεδομένα μπορούν να γίνουν εύκολα εκτιμήσεις και για τα επόμενα χρόνια. Επομένως, όπως γίνεται τώρα πρόβλεψη για το 2013, εύκολα μπορεί να πραγματοποιηθεί αργότερα πρόβλεψη και για το 2014 αλλά και για το 2015 χρησιμοποιώντας τα ίδια μοντέλα.

Σε πρώτο στάδιο, δημιουργήθηκε ένα ξεχωριστό μοντέλο για κάθε ρύπο και για κάθε χώρα, ώστε να είναι πιο ακριβείς οι εκτιμήσεις. Σαν εναλλακτικός τρόπος παρουσιάζεται η ανάλυση παλινδρόμησης με διαστρωματικά στοιχεία χρονολογικών σειρών (Panel data), η οποία έχει το πλεονέκτημα ότι σου επιτρέπει να ελέγξεις τους παράγοντες που μεταβάλλονται μεταξύ των χωρών, αλλά δεν μεταβάλλονται διαχρονικά. Στόχος αυτής της μελέτης είναι η παραγωγή άμεσων προβλέψεων για τις ατμοσφαιρικές εκπομπές του διοξειδίου του άνθρακα (CO_2), του μεθανίου (CH_4) και του υποξειδίου του αζώτου (N_2O), τα οποία αποτελούν τα σημαντικότερα αέρια του θερμοκηπίου, σύμφωνα με την Υπηρεσία Περιβαλλοντικής Προστασίας (EPA).

Στο θεωρητικό μέρος της εργασίας γίνεται εκτενής αναφορά στην κλιματική αλλαγή, το φαινόμενο του θερμοκηπίου και τα αέρια που συμβάλλουν στη δημιουργία αυτού του φαινομένου. Στο Κεφάλαιο 2 γίνεται μια εισαγωγή στην ανάλυση παλινδρόμησης των πολλαπλών γραμμικών μοντέλων και στο επόμενο Κεφάλαιο παρουσιάζεται η μεθοδολογία που χρησιμοποιήθηκε και η επιλογή των επεξηγηματικών μεταβλητών, οι οποίες θα χρησιμοποιηθούν για την διαμόρφωση του κατάλληλου μοντέλου. Στο Κεφάλαιο 4 παρατίθενται τρεις (μία για κάθε αέριο του θερμοκηπίου) από τις αναλύσεις των πολλαπλών γραμμικών μοντέλων που δημιουργήθηκαν και στο Κεφάλαιο 5 ακολουθούν τα

αποτελέσματα των προβλέψεων των εκπομπών κάθε χώρας για κάθε αέριο του θερμοκηπίου. Στο Κεφάλαιο 6 γίνεται μία εισαγωγή στην ανάλυση παλινδρόμησης με διαστρωματικά στοιχεία χρονολογικών σειρών και στο επόμενο Κεφάλαιο παρουσιάζεται η μεθοδολογία που χρησιμοποιήθηκε. Στο Κεφάλαιο 8 παρατίθενται δύο αναλύσεις με τη χρήση των panel data (μία σταθερών επιδράσεων και μία τυχαίων επιδράσεων). Στο 9^ο Κεφάλαιο ακολουθούν τα αποτελέσματα των προβλέψεων και τέλος, στο 10^ο καταγράφονται τα συμπεράσματα και συγκρίνονται οι δύο μέθοδοι που χρησιμοποιήθηκαν.

Abstract

This thesis was produced as part of my studies to obtain my undergraduate degree in National Technical University of Athens's School of Applied Mathematics and Physic Sciences. Its purpose is to achieve now-casting emissions of the greenhouse gases in the European Union through the application of two different regression models.

Global warming is not a new study in science. Some say it is a natural process and that greenhouse gases have always been existed, but the amount of gases in the atmosphere have skyrocketed in recent history. Scientists fear that the worst effects of global warming - extreme weather events, rising sea levels, the extinction of some plants and animals, the intensification of the oceans, significant climate changes and without previous social unrest - will be unavoidable if these trends continue.

Globally, the production of greenhouse gases is a source of serious concern. Every day, we breathe about 14 pounds of air, drink 2 kg of water and eat 1.5 kg of food. Because of the impact the air pollution has on health, the wellbeing and the environment in general, it is important to analyze and predict the levels of air pollutants. Thus, by using suitable measures, the present situation can be controlled and improved.

The objective of this thesis is to propose a reliable forecasting model of atmospheric pollutants. The data on emissions of air pollutants are not available for all the years. Eurostat, has data up to the year $t-3$ (where t is the current year), while other databases such as UNFCCC have this data 18 months after the end of each year. Currently, in the two databases, data for the emissions can be found until 2012. The aim is to forecast emissions for the year 2013 through regression models. Based on these models and with the most recent data, which can be acquired later, estimations can be easily made for the next years. Therefore, as forecast can be done now for 2013, it can be easily accomplished later for 2014 and 2015 using the same models.

At first stage, a separate model was created for each pollutant and for each country, so as to have more precise estimates. As an alternative way, the Panel data analysis (also known as cross-sectional time-series data) was presented, which has the advantage that allows you to control for variables that change across countries but not over time. The aim of this study is to produce direct estimates for air emissions of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), which, according to the Environmental Protection Agency (EPA), comprise the most important greenhouse gases.

In the theoretical part of the work, there is an extensive reference to climate change, the greenhouse effect and the gases that contribute to creating this phenomenon. Chapter 2 is an introduction to the regression analysis of multiple linear models and the next Chapter presents the methodology, that was used and the selection of explanatory variables, that are used to form the appropriate model. In Chapter 4, three (one for each pollutant) analyzes of multiple linear models, which created, are listed and Chapter 5 presents the results of emission forecasts in each country for each greenhouse gas. Chapter 6 is an introduction to Panel data analysis and the next chapter presents the methodology that was used. Chapter 8 lists two analyses using panel data (a fixed effects model and a random effects model). In the 9th Chapter, there are the projection's results and finally, the 10th has the conclusions and the comparison of the two methods that were used.

Περιεχόμενα

Περίληψη	3
Abstract	5
Περιεχόμενα	6
Κατάλογος Διαγραμμάτων	8
Κατάλογος Πινάκων	11
1.Εισαγωγή.....	14
1.1.Κλιματική Αλλαγή-Φαινόμενο του Θερμοκηπίου.....	14
1.2.Αέρια του Θερμοκηπίου	15
2.Ανάλυση Παλινδρόμησης Πολλαπλού Γραμμικού Μοντέλου.....	19
2.1.Έλεγχοι Υποθέσεων	20
2.1.1.Γραμμικότητα	20
2.2.2.Κανονικότητα Σφαλμάτων.....	20
2.2.3.Ομοσκεδαστικότητα	21
2.2.4.Ανεξαρτησία Σφαλμάτων	21
3.Μεθοδολογία	22
3.1.Επιλογή των πιθανών επεξηγηματικών μεταβλητών.....	22
3.1.1.Διοξείδιο του άνθρακα (Carbon Dioxide – CO ₂)	22
3.1.2.Μεθάνιο (Methane-CH ₄)	26
3.1.3.Υποξείδιο του αζώτου (Nitrogenous Oxide – N ₂ O).	32
3.2.Συλλογή δεδομένων	38
3.3.Επιλογή καταλληλότερων επεξηγηματικών μεταβλητών για το μοντέλο.....	40
3.3.1.Πολυσυγγραμμικότητα (VIF)	40
3.3.2.Κριτήριο AIC	41
3.3.3.Κριτήριο BIC	41
3.3.4.Εφαρμογή των κριτηρίων AIC και BIC στην R.....	42
3.3.5.Μέθοδοι L ₁ και L ₂ -Penalized	45
3.3.5.1.Μέθοδος L ₂ - Penalized (Ridge regression).	45
3.3.5.2.Μέθοδος L ₁ - Penalized (LASSO).....	48
3.3.5.3.Cross Validation (CVL).....	49

4.Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης.....	51
4.1.Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές CO ₂ της Δανίας	51
4.2.Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές CH ₄ της Κύπρου	59
4.3.Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές N ₂ O της Ολλανδίας	67
5.Πρόβλεψη	76
6.Ανάλυση Παλινδρόμησης με διαστρωματικά στοιχεία χρονολογικών σειρών (Panel Data)	78
6.1.Μοντέλα Σταθερών Επιδράσεων (Fixed-Effects Models).....	78
6.2.Μοντέλα Τυχαίων Επιδράσεων (Random-Effects Models)	79
7.Μεθοδολογία (Panel Data).....	81
7.1.Κατανομή χωρών σε πίνακες	81
7.2.Έλεγχος Πολυσυγγραμμικότητας	82
7.3.Επιλογή Καλύτερης Μεθόδου	82
7.4.Επιλογή Καταλληλότερων Επεξηγηματικών Μεταβλητών	83
7.5.Διαγνωστικοί Έλεγχοι.....	83
8.Ανάλυση Παλινδρόμησης με Panel Data.....	84
8.1.Ανάλυση Παλινδρόμησης με Panel Data για τις εκπομπές μεθανίου (CH ₄) των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου.....	84
8.2.Ανάλυση Παλινδρόμησης με Panel Data για τις εκπομπές διοξειδίου του άνθρακα των χωρών με πλήρη δεδομένα	94
9.Πρόβλεψη	101
10.Συμπεράσματα	103
Βιβλιογραφία	106
Παράρτημα I	108
Παράρτημα II.....	114
Παράρτημα III	144

Κατάλογος Διαγραμμάτων

1.1 Το φαινόμενο του θερμοκηπίου	14
1.2 Κατανομή συνολικών εκπομπών αερίων του θερμοκηπίου, ΕΕ-28, 2012.....	15
1.3 Γραφική Παράσταση των συνολικών εκπομπών αερίων του θερμοκηπίου (GHG), 1990-2012 και προβλεπόμενες εκπομπές μέχρι το 2020.....	17
3.1 Κατανομή συνολικών εκπομπών CO ₂ της ΕΕ-28 από όλες τις πηγές, το 2012	23
3.2 Κατανομή συνολικών εκπομπών CO ₂ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012.....	24
3.3 Κατανομή συνολικών εκπομπών CO ₂ της ΕΕ-28 από τον τομέα της καύσης των καυσίμων, το 2012.....	25
3.4 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από όλους τους τομείς, το 2012	27
3.5 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από τη γεωργία, το 2012	27
3.6 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από τα αστικά απόβλητα με βάση τον τύπο διαχείρισής τους, από το 1995 μέχρι το 2007	28
3.7 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από τα απόβλητα, το 2012.....	29
3.8 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από την ενέργεια, το 2012.....	29
3.9 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από τα απόβλητα, το 2012.....	30
3.10 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από τα στερεά καύσιμα, το 2012.....	30
3.11 Κατανομή συνολικών εκπομπών CH ₄ της ΕΕ-28 από το πετρέλαιο και το φυσικό αέριο, το 2012.....	31
3.12 Κατανομή συνολικών εκπομπών N ₂ O της ΕΕ-28 από όλες τις πηγές, 1990-2012.....	33
3.13 Κατανομή συνολικών εκπομπών N ₂ O της ΕΕ-28 από όλες τις πηγές, το 2012.....	34
3.14 Κατανομή συνολικών εκπομπών N ₂ O της ΕΕ-28 από τη γεωργία, το 2012.....	35
3.15 Κατανομή συνολικών εκπομπών N ₂ O της ΕΕ-28 από την ενέργεια, το 2012	36
3.16 Κατανομή συνολικών εκπομπών N ₂ O της ΕΕ-28 από την καύση των καυσίμων, το 2012.....	37
3.17 Γεωμετρική Ερμηνεία της μεθόδου παλινδρόμησης κορυφογραμμής.....	47
3.18 Γεωμετρική Ερμηνεία της μεθόδου Lasso (αριστερά) και της μεθόδου παλινδρόμησης κορυφογραμμής (δεξιά).....	49

3.19 Παράδειγμα Cross Validation καμπύλη σφάλματος	50
4.1 Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CO ₂ - Δανία).....	52
4.2 Ίχνος κορυφογραμμής (CO ₂ -Δανία)	54
4.3 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CO ₂ -Δανία)	54
4.4 Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (CO ₂ -Δανία)	55
4.5 Έλεγχος γραμμικότητας (CO ₂ -Δανία).....	56
4.6 Έλεγχος κανονικότητας των υπολοίπων (CO ₂ -Δανία)	57
4.7 Έλεγχος ομοσκεδαστικότητας (CO ₂ -Δανία)	57
4.8 Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (CO ₂ -Δανία) (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)	58
4.9 Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CH ₄ - Κύπρος).....	59
4.10 Ίχνος κορυφογραμμής (CH ₄ -Κύπρος)	62
4.11 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CH ₄ -Κύπρος).....	63
4.12 Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (CH ₄ - Κύπρος)	63
4.13 Έλεγχος γραμμικότητας (CH ₄ -Κύπρος)	65
4.14 Έλεγχος κανονικότητας των υπολοίπων (CH ₄ -Κύπρος).....	65
4.15 Έλεγχος ομοσκεδαστικότητας (CH ₄ -Κύπρος).....	65
4.16 Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (CH ₄ -Κύπρος) (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)	66
4.17 Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (N ₂ O - Ολλανδία).....	68
4.18 Ίχνος κορυφογραμμής (N ₂ O - Ολλανδία)	70
4.19 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (N ₂ O - Ολλανδία).....	71
4.20 Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (N ₂ O - Ολλανδία)	72
4.21 Έλεγχος γραμμικότητας (N ₂ O - Ολλανδία)	73
4.22 Έλεγχος κανονικότητας των υπολοίπων (N ₂ O - Ολλανδία).....	74

4.23 Έλεγχος ομοσκεδαστικότητας (N_2O - Ολλανδία)	74
4.24 Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)	75
5.1 Γραφικές παραστάσεις των προβλέψεων των αερίων του θερμοκηπίου για τις χώρες της ΕΕ-28 (βάσει της πολλαπλής γραμμικής παλινδρόμησης).....	77
8.1 Συνολικές εκπομπές CH_4 ανά χώρα για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012).....	84
8.2 Μέση τιμή εκπομπών CH_4 ανά χώρα για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012)	85
8.3 Συνολικές εκπομπές CH_4 ανά έτος για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012).....	85
8.4 Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CH_4 – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου)	87
8.5 Μέση τιμή εκπομπών CO_2 ανά χώρα για τις χώρες με πλήρη δεδομένα (1995-2012) 94	
8.6 Συνολικές εκπομπές CO_2 ανά έτος για τις χώρες με πλήρη δεδομένα (1995-2012)	94
8.7 Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CO_2 – για τις χώρες με πλήρη δεδομένα)	95
9.1 Γραφικές Παραστάσεις των προβλέψεων των εκπομπών των αερίων του θερμοκηπίου στις χώρες της ΕΕ-28 (βάσει των Panel Data)	102

Κατάλογος Πινάκων

1.1	Συνολικές εκπομπές αερίων του θερμοκηπίου,EE-28, 2012	16
1.2	Δυναμικό θέρμανσης του πλανήτη (GWP) κάθε αερίου του θερμοκηπίου.....	16
3.1	Συνολικές εκπομπές CO ₂ από τον τομέα της Ενέργειας και από τις Ενεργειακές Βιομηχανίες για το 2012.....	25
3.2	Συνολικές εκπομπές CO ₂ από τον τομέα της Ενέργειας και από τις Μεταφορές για το 2012	25
3.3	Συνολικές εκπομπές CO ₂ από τον τομέα της Ενέργειας και από τις Μεταποιητικές βιομηχανίες και κατασκευές για το 2012	25
3.4	Συνολικές εκπομπές CH ₄ από τον τομέα της Ενέργειας και από την Εξόρυξη και επεξεργασία του άνθρακα για το 2012	31
3.5	Συνολικές εκπομπές CH ₄ από τον τομέα της Ενέργειας και από το Φυσικό Αέριο για το 2012	32
3.6	Συνολικές εκπομπές N ₂ O από τον τομέα της Ενέργειας και από τις Ενεργειακές βιομηχανίες για το 2012	36
3.7	Συνολικές εκπομπές N ₂ O από τον τομέα της Ενέργειας και από τις Μεταφορές για το 2012	36
3.8	Συνολικές εκπομπές N ₂ O από τον τομέα της Ενέργειας και από τις Μεταποιητικές βιομηχανίες και κατασκευές για το 2012	37
3.9	Πηγές συλλογής δεδομένων των επεξηγηματικών μεταβλητών.....	51
4.1	Δεδομένα για τις εκπομπές CO ₂ και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Δανίας.....	51
4.2	Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CO ₂ - Δανία).....	51
4.3	Τιμές VIF των επεξηγηματικών μεταβλητών (CO ₂ - Δανία).....	52
4.4	Πίνακας όλων των πιθανών μοντέλων για την πρόβλεψη εκπομπών CO ₂ της Δανίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC και AIC).....	53
4.5	Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CO ₂ -Δανία).....	54
4.6	Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (CO ₂ -Δανία).....	55
4.7	Δεδομένα για τις εκπομπές CH ₄ και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Κύπρου	59
4.8	Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CH ₄ - Κύπρος) .	59

4.9	Τιμές VIF των επεξηγηματικών μεταβλητών (CH ₄ - Κύπρος).....	60
4.10	Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών CH ₄ της Κύπρου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC).....	61
4.11	Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών CH ₄ της Κύπρου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο AIC).....	62
4.12	Τιμές των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CH ₄ -Κύπρος)	63
4.13	Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (CH ₄ -Κύπρος).....	64
4.14	Δεδομένα για τις εκπομπές N ₂ O και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Ολλανδίας.....	67
4.15	Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (N ₂ O - Ολλανδία).....	67
4.16	Τιμές VIF των επεξηγηματικών μεταβλητών (N ₂ O - Ολλανδία).....	68
4.17	Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών N ₂ O της Ολλανδίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο AIC)	68
4.18	Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών N ₂ O της Ολλανδίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC).....	70
4.19	Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (N ₂ O - Ολλανδία).....	71
4.20	Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (N ₂ O - Ολλανδία).....	72
5.1	Προβλέψεις εκπομπών των αερίων του θερμοκηπίου κάθε χώρας για το 2013	76
5.2	Πρόβλεψη εκπομπών των αερίων του θερμοκηπίου της ΕΕ-28 για το 2013	76
8.1	Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CH ₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου).....	86
8.2	Τιμές VIF των επεξηγηματικών μεταβλητών (CH ₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου).....	87
8.3	Τιμές VIF των επεξηγηματικών μεταβλητών μετά την αφαίρεση των μεταβλητών που ευθύνονταν για την πολυσυγγραμμικότητα (CH ₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου).....	88
8.4	Πίνακας των δέκα καταλληλότερων μοντέλων με βάση το BIC για την πρόβλεψη εκπομπών CH ₄ των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC).....	91

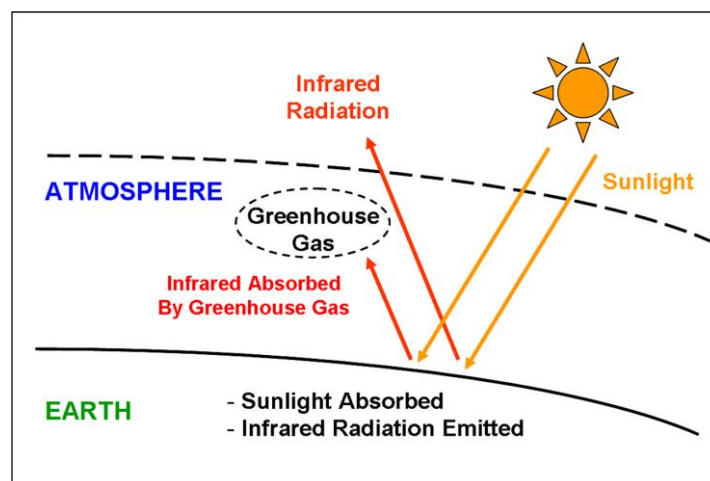
8.5	Πίνακας των δέκα καταλληλότερων μοντέλων με βάση το AIC για την πρόβλεψη εκπομπών CH ₄ των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο AIC).....	91
8.6	Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CO ₂ – για τις χώρες με πλήρη δεδομένα)	95
8.7	Τιμές VIF των επεξηγηματικών μεταβλητών (CO ₂ – για τις χώρες με πλήρη δεδομένα).....	95
8.8	Τιμές VIF των επεξηγηματικών μεταβλητών μετά την αφαίρεση της μεταβλητής που ευθυνόταν για την πολυσυγγραμμικότητα (CO ₂ – για τις χώρες με πλήρη δεδομένα).....	96
8.9	Πίνακας όλων των πιθανών μοντέλων με βάση το BIC και AIC για την πρόβλεψη εκπομπών CO ₂ των με πλήρη δεδομένα (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC και AIC).....	99
9.1	Προβλέψεις εκπομπών αερίων του θερμοκηπίου κάθε χώρας για το 2013 με τη χρήση των panel data.....	101
9.2	Πρόβλεψη εκπομπών αερίων του θερμοκηπίου της ΕΕ-28 για το 2013 με τη χρήση των panel data	100
10.1	Διαφορές των προβλέψεων της πολλαπλής γραμμικής παλινδρόμησης με τις προβλέψεις της παλινδρόμησης με τη χρήση των Panel Data	103
10.2	Εκπομπές των τριών αερίων του θερμοκηπίου στην ΕΕ-28 (1995-2012).....	104
10.3	Προβλέψεις εκπομπών των τριών αερίων του θερμοκηπίου στην ΕΕ-28 για το 2013 με την πολλαπλή γραμμική παλινδρόμηση	104
10.4	Προβλέψεις εκπομπών των τριών αερίων του θερμοκηπίου στην ΕΕ-28 για το 2013 με την χρήση των panel data.....	104

1. Εισαγωγή

1.1. Κλιματική Αλλαγή-Φαινόμενο του Θερμοκηπίου

Η κλιματική αλλαγή είναι πλέον γεγονός. Η Γη μας θερμαίνεται. Η μέση θερμοκρασία της Γης έχει αυξηθεί κατά $1,4^{\circ}\text{F}$ κατά τη διάρκεια του προηγούμενου αιώνα, και αναμένεται να αυξηθεί ακόμη 2 έως $11,5^{\circ}\text{F}$ για τα επόμενα εκατό χρόνια. Λόγω της υπερθέρμανσης του πλανήτη, το 2014 ήταν η θερμότερη χρονιά που έχει καταγραφεί και 10 από τα θερμότερα έτη καταγράφονται όλα μετά το 1998. Μικρές αλλαγές της μέσης θερμοκρασίας του πλανήτη μπορούν να προξενήσουν μεγάλες και δυνητικά επικίνδυνες κλιματικές μεταβολές και καιρικές συνθήκες. Σύμφωνα με τον «Διεθνή Οργανισμό Έρευνας πάνω στον καρκίνο» (International Agency for Research on Cancer – IARC), πολλά καρκινογόνα στοιχεία που βρίσκονται στην ατμόσφαιρα βάζουν καθημερινά την υγεία μας σε κίνδυνο. Ο συγκεκριμένος οργανισμός εξάγει έρευνες για τις αιτίες του καρκίνου, καθώς επίσης και επιδημιολογικές μελέτες για τον καρκίνο σε όλο τον κόσμο και αναπτύσσει στρατηγικές για τον έλεγχο του.

Η θερμοκρασία της Γης είναι μια πράξη εξισορρόπησης. Εξαρτάται από την ισορροπία μεταξύ της ενέργειας που εισέρχεται και εξέρχεται από το σύστημα του πλανήτη. Όταν η εισερχόμενη ενέργεια από τον ήλιο απορροφάται από το σύστημα της Γης, η Γη θερμαίνεται. Όταν η ενέργεια του ήλιου αντανακλάται πίσω στο διάστημα, η Γη αποφεύγει την υπερθέρμανση. Περίπου το 30 τοις εκατό της ακτινοβολίας που χτυπούν τη Γη αντανακλάται πίσω στο διάστημα από τα σύννεφα, τους πάγους και άλλες ανακλαστικές επιφάνειες. Το υπόλοιπο 70 τοις εκατό απορροφάται από τους ωκεανούς, τη Γη και την ατμόσφαιρα, σύμφωνα με τη NASA. Αυτή η ανταλλαγή της εισερχόμενης και εξερχόμενης ακτινοβολίας που θερμαίνει τη Γη αναφέρεται ως το «φαινόμενο του θερμοκηπίου» (Διάγραμμα 1.1), επειδή ένα θερμοκήπιο λειτουργεί με τον ίδιο τρόπο. Πολλοί παράγοντες, τόσο φυσικοί όσο και ανθρωπίνι, μπορούν να προκαλέσουν αλλαγές στο ενεργειακό ισοζύγιο της Γης.



Διάγραμμα 1.1: Το φαινόμενο του θερμοκηπίου

Σε γενικές γραμμές, οι κλιματικές αλλαγές πριν τη Βιομηχανική Επανάσταση το 1700 μπορούν να εξηγηθούν από φυσικά αίτια, όπως οι μεταβολές στην ηλιακή ενέργεια, οι ηφαιστειακές εκρήξεις, και η φυσική εξέλιξη των συγκεντρώσεων των αερίων του θερμοκηπίου. Ωστόσο, κατά τις δύο τελευταίες δεκαετίες ένα αυξανόμενο σώμα επιστημονικών στοιχείων έχει καθιερωθεί που υποδηλώνει ότι οι πιο πρόσφατες αλλαγές στο κλίμα της Γης έχουν προκληθεί από την ανθρωπίνη δραστηριότητα – τις λεγόμενες

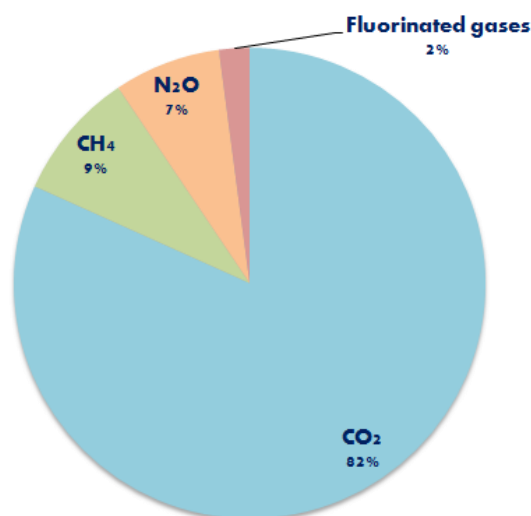
ανθρωπογενείς επιδράσεις. Αυτοί που ευθύνονται, σε μεγάλο βαθμό, για την πρόσφατη αλλαγή του κλίματος, είναι οι ίδιοι που υποφέρουν από τις επιπτώσεις της.

Κατά τη διάρκεια του περασμένου αιώνα, οι ανθρώπινες δραστηριότητες έχουν διαδώσει μεγάλες ποσότητες διοξειδίου του άνθρακα και άλλων αερίων του θερμοκηπίου στην ατμόσφαιρα. Η πλειοψηφία των αερίων του θερμοκηπίου προέρχονται από την καύση ορυκτών καυσίμων για την παραγωγή ενέργειας, αν και η αποψίλωση των δασών, οι βιομηχανικές διεργασίες, και μερικές γεωργικές πρακτικές εκπέμπουν επίσης βλαβερά αέρια στην ατμόσφαιρα. Τα αέρια του θερμοκηπίου λειτουργούν σαν μια κουβέρτα γύρω από τη Γη, παγιδεύοντας την ενέργεια στην ατμόσφαιρα με αποτέλεσμα να την ζεσταίνουν. Αυτό το φαινόμενο ονομάζεται το φαινόμενο του θερμοκηπίου. Η συγκέντρωση των αερίων του θερμοκηπίου μπορεί να αλλάξει το κλίμα της Γης και να οδηγήσει σε επικίνδυνες συνέπειες για την ανθρώπινη υγεία, την ευημερία και τα οικοσυστήματα. Οι επιλογές που θα κάνουμε σήμερα, θα επηρεάσουν την ποσότητα αερίων του θερμοκηπίου που βρίσκονται στην ατμόσφαιρα στο εγγύς μέλλον και για τα επόμενα χρόνια.

Οι κυριότερες πηγές ρύπανσης της ατμόσφαιρας είναι τα καυσαέρια των οχημάτων, οι εκπομπές από εργοστάσια και βιοτεχνίες (κυρίως από την καύση ορυκτών καυσίμων για την παραγωγή ενέργειας), τα εργοστάσια παραγωγής ενέργειας, οι κεντρικές θερμάνσεις των σπιτιών, η καύση απορριμμάτων και γεωργικών υπολειμμάτων, οι αέριες εκπομπές από τη γεωργία και τη κτηνοτροφία και οι φυσιολογικές εκπομπές φυτών και δένδρων.

1.2. Αέρια του Θερμοκηπίου

Τα αέρια στην ατμόσφαιρα που απορροφούν την ακτινοβολία είναι γνωστά ως «αέρια του θερμοκηπίου» (GHG), επειδή είναι σε μεγάλο βαθμό υπεύθυνα για το φαινόμενο του θερμοκηπίου και αποτελούν λιγότερο από το 0,1% της συνολικής ατμόσφαιρας. Τα σημαντικότερα από αυτά, με τα οποία ασχολείται αυτή η μελέτη, είναι το διοξείδιο του άνθρακα (CO₂), που είναι το πιο κοινό, το μεθάνιο (CH₄) και το υποξείδιο του αζώτου (N₂O). Αέρια του θερμοκηπίου εκπέμπονται τόσο από φυσικές διεργασίες όσο και από ανθρώπινες δραστηριότητες. Ο όγκος των εκπομπών έχει αυξηθεί με ταχείς ρυθμούς κατά τα τελευταία 200 χρόνια, αντανακλώντας στην αύξηση του παγκόσμιου πληθυσμού, παράλληλα με την οικονομική ανάπτυξη που έχει οδηγήσει σε υψηλότερα επίπεδα παραγωγής και κατανάλωσης και η δημιουργία μιας παγκόσμιας οικονομίας.



Διάγραμμα 1.2.: Κατανομή συνολικών εκπομπών αερίων του θερμοκηπίου, ΕΕ-28, 2012

Διοξείδιο του Άνθρακα (CO₂)	3.717.511,18	81,73%
Μεθάνιο (CH₄)	401.757,13	8,83%
Υποξείδιο του Αζώτου (N₂O)	333.852,39	7,34%
Φθοριούχα Αέρια (Fluorinated gases)	95.234,34	2,10%
Όλα τα Αέρια του Θερμοκηπίου (All greenhouse gases)	4.548.355,04	100,00%

Πίνακας 1.1: Συνολικές εκπομπές αερίων του θερμοκηπίου, ΕΕ-28, 2012

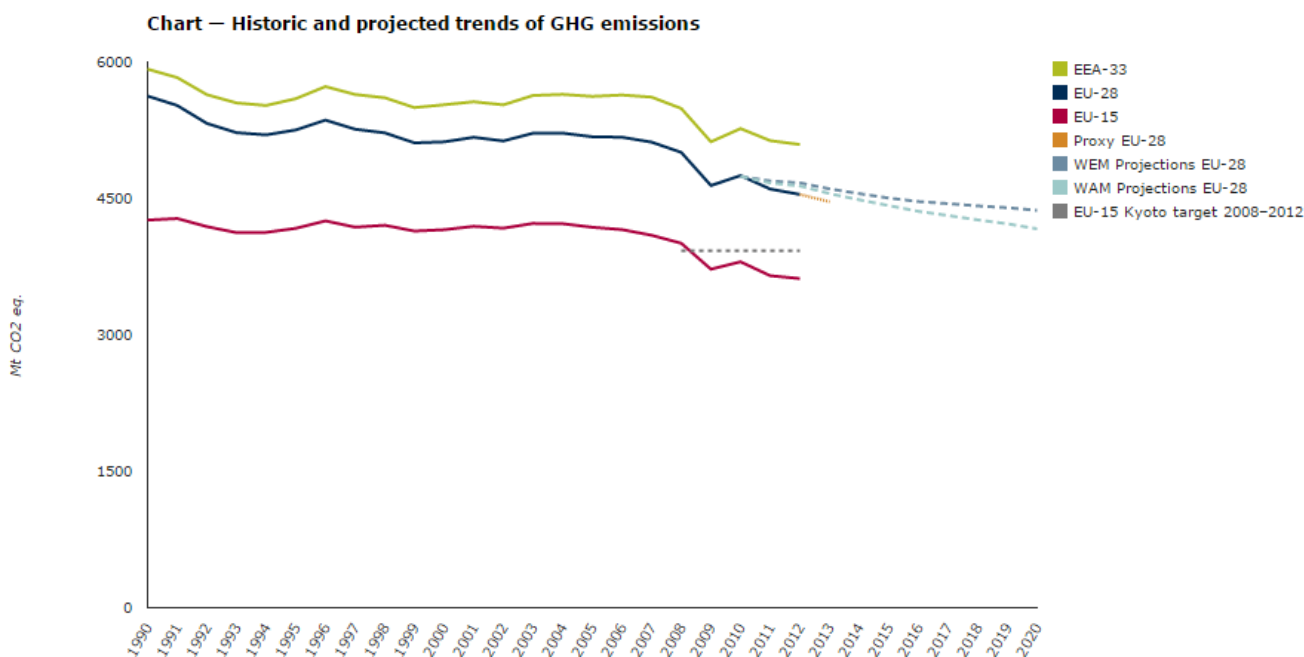
Το διοξείδιο του άνθρακα είναι το σημαντικότερο από τα αέρια μέσα στο πλαίσιο των εκπομπών αερίων του θερμοκηπίου (Διάγραμμα 1.2). Κατείχε το 81,73% των ΕΕ-28, των εκπομπών των αερίων του θερμοκηπίου το 2012, ακολουθούμενο από το μεθάνιο (8,83%), το υποξείδιο του αζώτου (7,34%) και τα φθοριούχα αέρια (2,1%) (Πίνακας 1.1). Κάθε αέριο του θερμοκηπίου έχει διαφορετική συμβολή στην πρόκληση της υπερθέρμανσης του πλανήτη, ανάλογα με τις ιδιότητες της ακτινοβολίας του, το μοριακό του βάρος και το μήκος του χρόνου που παραμένει στην ατμόσφαιρα. Το δυναμικό θέρμανσης του πλανήτη (GWP) κάθε αερίου του θερμοκηπίου ορίζεται σε σχέση με ένα συγκεκριμένο βάρος του διοξειδίου του άνθρακα και για ένα ορισμένο χρονικό διάστημα (100 ετών) (Πίνακας 1.2). Το δυναμικό θέρμανσης του πλανήτη χρησιμοποιείται για τη μετατροπή των εκπομπών άλλων αερίων του θερμοκηπίου σε ισοδύναμα CO₂ - καθιστώντας δυνατή τη σύγκριση των δυνητικών επιπτώσεων των διαφόρων αερίων.

Αέριο του Θερμοκηπίου	Παγκόσμιο Δυναμικό Θέρμανσης (GWP)
Διοξείδιο του Άνθρακα	1
Μεθάνιο	21
Υποξείδιο του Αζώτου	310

Πίνακας 1.2: Δυναμικό θέρμανσης του πλανήτη (GWP) κάθε αερίου του θερμοκηπίου

Παράλληλα με το αυξανόμενο σώμα επιστημονικών στοιχείων που υποστηρίζουν τη θεωρία ότι τα αυξημένα επίπεδα των εκπομπών αερίων του θερμοκηπίου μπορεί να αποδοθούν στην ανθρώπινη δραστηριότητα, ένας μεγάλος αριθμός χωρών έχει αρχίσει να καταβάλλει προσπάθειες για τη μείωση των εκπομπών αερίων του θερμοκηπίου. Η συμφωνία για την μείωση των εκπομπών αερίων του θερμοκηπίου, η γνωστή σαν πρωτόκολλο του Κιότο, είναι μία από τις σημαντικότερες διεθνείς νομοθετικές πράξεις καταπολέμησης των κλιματικών μεταβολών. Περιλαμβάνει τις δεσμεύσεις που έχουν αναλάβει οι εκβιομηχανισμένες χώρες για τον περιορισμό των οικείων εκπομπών ορισμένων αερίων που συμβάλλουν στο φαινόμενο του θερμοκηπίου, υπεύθυνων για τη θέρμανση του πλανήτη. Συνέβαλε επίσης στην περαιτέρω ευαισθητοποίηση του κοινού, παγκοσμίως, στα προβλήματα που συνδέονται με την αλλαγή του κλίματος. Το Πρωτόκολλο του Κιότο αφορά τις εκπομπές έξι αερίων θερμοκηπίου, στα οποία συμπεριλαμβάνονται το διοξείδιο του άνθρακα (CO₂) το μεθάνιο (CH₄) και το υποξείδιο του αζώτου (N₂O). Βάσει του πρωτοκόλλου 15 κράτη μέλη της ΕΕ (ΕΕ-15) πρέπει να μειώσουν τις συλλογικές εκπομπές τους κατά την περίοδο 2008-2012 κατά 8% κάτω από τα επίπεδα του 1990 (Διάγραμμα 1.3).

Η δημοσιοποίηση της τελευταίας έκθεσης της Διακυβερνητικής Επιτροπής του ΟΗΕ για το Κλίμα (IPCC) προειδοποιεί ότι μέχρι το τέλος του αιώνα η μέση επιφανειακή θερμοκρασία θα ανέβει κατά 0,3 έως 4,8 βαθμούς Κελσίου σε σχέση με τα προβιομηχανικά επίπεδα, ανάλογα με τα μέτρα που θα ληφθούν για τη μείωση των εκπομπών. Η αλλαγή του κλίματος είναι μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζει η ανθρωπότητα, και ένας αυξανόμενος αριθμός χωρών επιδιώκουν με συγκεκριμένες ενέργειες, την αντιμετώπιση αυτού του προβλήματος.



Διάγραμμα 1.3: Γραφική Παράσταση των συνολικών εκπομπών αερίων του θερμοκηπίου (GHG), 1990-2012 και προβλεπόμενες εκπομπές μέχρι το 2020.

Η έκθεση του ΕΟΠ «Κλιματική αλλαγή, επιπτώσεις και τρωτότητα στην Ευρώπη 2012» δείχνει ότι όλες οι περιοχές στην Ευρώπη επηρεάζονται από την κλιματική αλλαγή, με ένα ευρύ φάσμα επιπτώσεων στην κοινωνία, τα οικοσυστήματα και την ανθρώπινη υγεία. Παρόλο που δεν γνωρίζουμε πλήρως τον τρόπο με τον οποίο η κλιματική αλλαγή μπορεί να επηρεάσει την ποιότητα του αέρα, και αντίστροφα, πρόσφατη έρευνα δείχνει ότι αυτή η αμοιβαία σχέση μπορεί να είναι πιο ισχυρή από ότι φάνηκε σε προηγούμενες εκτιμήσεις. Στις αξιολογήσεις της από το 2007, η Διακυβερνητική Επιτροπή για την Αλλαγή του Κλίματος, ο διεθνής φορέας που δημιουργήθηκε για να αξιολογήσει την κλιματική αλλαγή, προβλέπει μια επιδείνωση της ποιότητας του ατμοσφαιρικού αέρα των πόλεων στο μέλλον, λόγω της κλιματικής αλλαγής.

Είναι πολύ πιθανό ότι το μεγαλύτερο μέρος της αύξησης της θερμοκρασίας από τα μέσα του 20ού αιώνα οφείλεται στην παρατηρούμενη αύξηση των συγκεντρώσεων των αερίων του θερμοκηπίου, ως αποτέλεσμα των εκπομπών που προκαλούν ανθρώπινες δραστηριότητες. Η παγκόσμια θερμοκρασία έχει αυξηθεί κατά περίπου 0,8 °C τα τελευταία 150 χρόνια και αναμένεται να αυξηθεί περαιτέρω. Αν η θερμοκρασία της γης αυξηθεί περισσότερο από 2 °C σε σχέση με τα προβιομηχανικά επίπεδα, θα αυξηθεί ο κίνδυνος των επικίνδυνων αλλαγών για τα παγκόσμια ανθρώπινα και φυσικά συστήματα. Στη σύμβαση-πλαίσιο των Ηνωμένων Εθνών για τις κλιματικές μεταβολές (UNFCCC) έχει αναγνωριστεί ο στόχος περιορισμού της αύξησης της παγκόσμιας μέσης θερμοκρασίας σε σχέση με την προβιομηχανική εποχή κάτω των 2 °C.

Πώς μπορεί να επιτευχθεί αυτό; Οι παγκόσμιες εκπομπές αερίων θερμοκηπίου πρέπει να σταθεροποιηθούν σε αυτήν τη δεκαετία και να μειωθούν κατά 50% συγκριτικά με τα επίπεδα του 1990 έως το 2050. Λαμβάνοντας υπόψη τις απαραίτητες προσπάθειες που καταβάλλουν οι αναπτυσσόμενες χώρες, η ΕΕ υποστηρίζει τον στόχο για μείωση των εκπομπών αερίων θερμοκηπίου κατά 80% έως 95% έως το 2050 (συγκριτικά με το 1990). Ακόμη και αν οι πολιτικές και οι προσπάθειες για τη μείωση των εκπομπών αποδειχθούν αποτελεσματικές, κάποιος βαθμός κλιματικής αλλαγής είναι αναπόφευκτος. Συνεπώς, είναι επίσης αναγκαίες στρατηγικές και δράσεις για την προσαρμογή στις επιπτώσεις της. Ένας συνδυασμός των διάφορων τομέων της οικονομίας (μεταφορές, ενέργεια, βιομηχανία,

γεωργία) που εκπέμπουν πολλά και διάφορα χημικά συστατικά σχετίζονται με αυτά τα περιβαλλοντικά προβλήματα, και για αυτό το λόγο, οι πολιτικές για τη μείωση των εκπομπών εξετάζουν όλο και περισσότερο διάφορα προβλήματα ρύπανσης από κοινού, χρησιμοποιώντας πολλές προσεγγίσεις. Τα κράτη μέλη πρέπει να συμμορφωθούν με τα πρότυπα ποιότητας αέρα της ΕΕ γρήγορα και να μειώσουν τις εκπομπές ατμοσφαιρικών ρύπων.

2. Ανάλυση Παλινδρόμησης Πολλαπλού Γραμμικού Μοντέλου

Η εξέταση της συμπεριφοράς ενός φαινομένου (εικομπές αερίων του θερμοκηπίου) με στόχο την πρόβλεψη της μελλοντικής εξέλιξης του, βασίζεται στην ανάλυση και επεξεργασία αριθμού παρατηρήσεων των παραγόντων που επηρεάζουν το συγκεκριμένο φαινόμενο. Ο τρόπος με τον οποίο συμπεριφέρεται μια μεταβλητή (μεταβλητή απόκρισης ή εξαρτημένη μεταβλητή), καθορίζεται από τη συμπεριφορά άλλων μεταβλητών (επεξηγηματικές ή ανεξάρτητες μεταβλητές). Μία από τις μεθόδους προβλέψεων που χρησιμοποιείται είναι η ανάλυση παλινδρόμησης.

Εφόσον οι τιμές των επεξηγηματικών μεταβλητών είναι διαθέσιμες, σκοπός της ανάλυσης μας είναι να προσδιοριστεί η σχέση μεταξύ των μεταβλητών αυτών και της μεταβλητής απόκρισης για τη δημιουργία προβλέψεων. Το μοντέλο παλινδρόμησης χρησιμοποιείται ώστε να μοντελοποιηθεί η σχέση μεταξύ μίας μεταβλητής απόκρισης Y και p επεξηγηματικών μεταβλητών X_1, X_2, \dots, X_p . Μοντέλα παλινδρόμησης που περιέχουν δύο ή περισσότερες επεξηγηματικές μεταβλητές ονομάζονται μοντέλα πολλαπλής παλινδρόμησης (multiple regression models).

Το πολλαπλό γραμμικό μοντέλο έχει την πιο κάτω μορφή:

$$Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

ή ισοδύναμα:

$$E(Y_i | X_{1i}, \dots, X_{pi}) = \alpha + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$$

όπου: α η σταθερά του μοντέλου

Y_i η μεταβλητή απόκρισης, $i=1,2,\dots,n$

X_{it} η t -επεξηγηματική μεταβλητή, $t=1,2,\dots,p$, $i=1,2,\dots,n$

p ο αριθμός των επεξηγηματικών μεταβλητών

n το μέγεθος του δείγματος

β_t ο συντελεστής της t -επεξηγηματικής μεταβλητής, $t=1,2,\dots,p$

ε_i το τυχαίο σφάλμα που ακολουθεί την κανονική κατανομή $N(0, \sigma^2)$, $i=1,2,\dots,n$, όπου η διασπορά σ^2 είναι άγνωστη και ως εκτιμήτρια της χρησιμοποιούμε την αμερόληπτη

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \equiv S^2$$

Οι συντελεστές $\beta_1, \beta_2, \beta_3, \dots, \beta_p$ και η σταθερά α είναι οι παράμετροι που επιδιώκουμε να εκτιμήσουμε ώστε να προσδιορίσουμε τη σχέση μεταξύ των μεταβλητών. Μία μονάδα αύξησης της επεξηγηματικής μεταβλητής X_t , μεταβάλλει τη μέση τιμή της μεταβλητής απόκρισης κατά β_t , εφόσον οι υπόλοιπες επεξηγηματικές μεταβλητές παραμένουν σταθερές. Ο σταθερός όρος α αντιπροσωπεύει τη μέση τιμή της μεταβλητής απόκρισης όταν όλες οι επεξηγηματικές μεταβλητές παίρνουν την τιμή μηδέν. Πρακτικά καλούμαστε να προσαρμόσουμε μία «ευθεία» στα δεδομένα μας εκτιμώντας τους συντελεστές αυτής της ευθείας μέσω των δεδομένων μας.

Για την εκτίμηση των συντελεστών αυτών χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων (Ordinary least squares-OLS). Σύμφωνα με την OLS, οι εκτιμήσεις των συντελεστών προκύπτουν από την ελαχιστοποίηση της ποσότητας

$$\sum_{i=1}^n \varepsilon_i^2,$$

όπου ε_i , $i=1, \dots, n$ τα τυχαία σφάλματα που παριστάνουν τις κατακόρυφες αποστάσεις των σημείων (x_i, y_i) από την ευθεία ελαχίστων τετραγώνων. Οι εκτιμήτριες των ελαχίστων τετραγώνων (least square estimators) είναι γραμμικές, αμερόληπτες και ακολουθούν την κανονική κατανομή.

Επιπρόσθετα, το γενικό γραμμικό μοντέλο πρέπει να τηρεί συγκεκριμένες προϋποθέσεις, οι οποίες παρουσιάζονται αναλυτικότερα στο Κεφάλαιο 2.1.

1. Γραμμικότητα
2. Κανονικότητα Σφαλμάτων
3. Ομοσκεδαστικότητα
4. Ανεξαρτησία Σφαλμάτων

2.1. Έλεγχοι Υποθέσεων

Αφού έχουμε καταλήξει σε ένα μοντέλο, μέσα από τις μεθόδους και τα κριτήρια που περιγράφηκαν πιο πριν, οφείλουμε να ελέγξουμε αν ικανοποιούνται οι προϋποθέσεις του γενικού γραμμικού μοντέλου.

2.1.1. Γραμμικότητα

Στην πολλαπλή γραμμική παλινδρόμηση, ο έλεγχος της γραμμικότητας γίνεται μέσω των διαγραμμάτων διασποράς των σημείων (x_{ti}, p_{ti}) , όπου x_{ti} οι παρατηρήσεις και p_{ti} τα μερικά υπόλοιπα. Αναλυτικότερα, πρέπει να ελέγξουμε αν η επεξηγηματική μεταβλητή X_{ti} συνδέεται γραμμικά με την δεσμευμένη μέση τιμή της Y_i μεταβλητής απόκρισης, αν όλες οι άλλες επεξηγηματικές μεταβλητές συνδέονται γραμμικά με την δεσμευμένη τιμή της Y_i . Δηλαδή, έχοντας τη σχέση:

$$y_i \approx \hat{\alpha} + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_{t-1} x_{t-1,i} + p_t(x_{ti}) + \hat{\beta}_{t+1} x_{t+1,i} + \dots + \hat{\beta}_p x_{pi}$$

πρέπει να δείξουμε ότι η συνάρτηση p_t είναι γραμμική.

Η σχέση που μας δίνει τα υπόλοιπα είναι η:

$$y_i \approx \hat{\alpha} + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_p x_{pi} + \varepsilon_i$$

Παρατηρούμε ότι αν αφαιρέσουμε την δεύτερη σχέση από την πρώτη έχουμε:

$$p_t(x_{ti}) \approx \hat{\beta}_t x_{ti} + \varepsilon_i \equiv P_{ti}$$

Οπότε ελέγχοντας τη σχέση των μερικών υπολοίπων P_{ti} με τις παρατηρήσεις x_{ti} μέσω διαγραμμάτων ελέγχουμε τη γραμμικότητα κάθε μεταβλητής και κατ' επέκταση του γενικού γραμμικού μοντέλου.

2.1.2. Κανονικότητα Σφαλμάτων

Κανονικότητα εμφανίζεται όταν τα σφάλματα της παλινδρόμησης ακολουθούν την κανονική κατανομή. Η κανονικότητα των σφαλμάτων ελέγχεται μέσω των QQplots για τα υπόλοιπα. Για να δεχτούμε την υπόθεση της κανονικότητας πρέπει τα σημεία που εμφανίζονται στο διάγραμμα να έχουν καλή προσαρμογή πάνω στην ευθεία που σχηματίστηκε. Επίσης, εκτός από τον γραφικό έλεγχο, υπάρχει και ένας αριθμητικός έλεγχος, το Jarque Bera test. Το Jarque Bera test μετρά την απόκλιση μας από την κανονικότητα. Η μηδενική υπόθεση είναι η ύπαρξη κανονικότητας έναντι της εναλλακτικής υπόθεσης για μη ύπαρξη κανονικότητας. Στην R, εκτελούμε τον έλεγχο χρησιμοποιώντας το πακέτο «Moments», και παίρνουμε ως αποτέλεσμα μια p-τιμή. Αν η p-τιμή είναι μεγαλύτερη από 0.05 δεν απορρίπτουμε την υπόθεση της κανονικότητας.

2.1.3. Ομοσκεδαστικότητα

Η ομοσκεδαστικότητα, δηλαδή η κοινή διασπορά των σφαλμάτων, ελέγχεται μέσω της γραφικής παράστασης των υπολοίπων συναρτήσει των προβλεπόμενων τιμών της Y_i . Τα ζεύγη αυτών των τιμών δεν πρέπει να εμφανίζουν κάποιο συστηματικό τρόπο συμπεριφοράς. Εκτός, από τον γραφικό έλεγχο, υπάρχει δυνατότητα να γίνει και αριθμητικός έλεγχος μέσω του πακέτου `lmtest` της R. Ο έλεγχος αυτός ονομάζεται τεστ Breusch-Pagan. Όταν ισχύει η μηδενική υπόθεση τότε η διακύμανση των σφαλμάτων είναι σταθερή έναντι της εναλλακτικής όπου ισχύει η ετεροσκεδαστικότητα. Στην περίπτωση που η p -τιμή, που εμφανίζει η εκτέλεση του `bptest` στην R, είναι μεγαλύτερη από 0.05 δεν απορρίπτουμε την υπόθεση της ομοσκεδαστικότητας. Ο έλεγχος της ομοσκεδαστικότητας είναι σημαντικός διότι σε περίπτωση ετεροσκεδαστικότητας, η μέθοδος ελαχίστων τετραγώνων δεν δίνει τις πιο αποτελεσματικές εκτιμήτριες και υπάρχουν άλλες αμερόληπτες εκτιμήτριες με μικρότερες διακυμάνσεις. Επίσης, οι διακυμάνσεις των β είναι μεροληπτικές, για αυτό η ομοσκεδαστικότητα κρίνεται απαραίτητη προϋπόθεση για την επιλογή ενός καλού μοντέλου.

2.1.4. Ανεξαρτησία Σφαλμάτων

Για δύο οποιοσδήποτε παρατηρήσεις, τα υπόλοιπα πρέπει να είναι ανεξάρτητα. Η υπόθεση αυτή μπορεί να ελεγχθεί μέσα από ένα διάγραμμα υπολοίπων ως προς τον αριθμό των παρατηρήσεων. Στο διάγραμμα δεν πρέπει να παρουσιάζεται κάποια σχέση και τα υπόλοιπα να συμπεριφέρονται τυχαία. Εκτός από το διαγραμματικό έλεγχο, καλό θα ήταν να γίνεται και αριθμητικός έλεγχος με το κριτήριο Durbin-Watson. Το Durbin-Watson test του πακέτου `lmtest` της R, έχει ως μηδενική υπόθεση την ύπαρξη ανεξαρτησίας. Αν η p -τιμή που θα εμφανίσει ως αποτέλεσμα είναι μικρότερη από 0.05 απορρίπτουμε την υπόθεση της ανεξαρτησίας.

3. Μεθοδολογία

Η ανάλυση παλινδρόμησης πραγματοποιήθηκε μέσα από τα ακόλουθα στάδια:

- i. Επιλογή των πιθανών επεξηγηματικών μεταβλητών
- ii. Συλλογή δεδομένων
- iii. Επιλογή καταλληλότερων επεξηγηματικών μεταβλητών για το μοντέλο
- iv. Έλεγχοι υποθέσεων
- v. Πρόβλεψη

3.1. Επιλογή των πιθανών επεξηγηματικών μεταβλητών

Τα μοντέλα παλινδρόμησης που χρησιμοποιούνται σε αυτή τη διπλωματική εργασία είναι βασισμένα σε ιστορικά δεδομένα. Μέσα από τη μελέτη διαφόρων δημοσιεύσεων της Eurostat, αλλά και της UNFCCC, και με τη βοήθεια της έρευνας που διεξήγαγε σχετικά η εταιρία Agilis, καθορίστηκαν οι πιθανές επεξηγηματικές μεταβλητές του μοντέλου, δηλαδή οι πιθανοί παράγοντες που επηρεάζουν τις εκπομπές των ατμοσφαιρικών ρύπων.

Σε αυτό το κεφάλαιο παρουσιάζονται τα αέρια και οι πηγές που συμβάλλουν στις ατμοσφαιρικές εκπομπές του καθενός μαζί με τις σχετικές μεταβλητές που θα συμπεριληφθούν στα μοντέλα. Σημαντικοί θεωρήθηκαν οι παράγοντες που συμβάλλουν τουλάχιστον 10% στις συνολικές εκπομπές κάθε ρύπου. Στην παρούσα μελέτη, εξετάζεται μόνο η ρύπανση που προέρχεται από ανθρωπογενείς πηγές, παρόλο που η ρύπανση μπορεί να οριστεί ευρύτερα σε άλλα πλαίσια. Διαφορετικοί ρύποι προέρχονται από διαφορετικές πηγές και έχουν διαφορετικές επιπτώσεις. Παρακάτω, παρουσιάζονται αναλυτικά οι βασικοί ρύποι και οι πηγές τους.

3.1.1. Διοξείδιο του άνθρακα (Carbon Dioxide – CO₂)

Το διοξείδιο του άνθρακα (CO₂), σε μία συγκέντρωση κοντά στο 0,04%, είναι η ένωση που επηρεάζει περισσότερο την κλιματική αλλαγή στην ατμόσφαιρα. Η συμβολή του CO₂ στο φαινόμενο του θερμοκηπίου υπολογίζεται στο 60%. Οι ατμοσφαιρικές συγκεντρώσεις του έχουν αυξηθεί κατά περίπου ένα τρίτο τα τελευταία 200 χρόνια. Η ίδια η ατμόσφαιρα λειτουργεί ως δεξαμενή στον παγκόσμιο κύκλο του άνθρακα. Κατά τη διάρκεια του φυσικού κύκλου του διοξειδίου του άνθρακα, η ποσότητα CO₂ στην ατμόσφαιρά διατηρείται σε ισορροπία. Μέσω της αναπνοής και της αποσύνθεσης των φυτών αλλά και των ηφαιστειακών εκρήξεων, απελευθερώνεται φυσικό CO₂ στην ατμόσφαιρα, όπου παραμένει για 100 περίπου χρόνια. Απομακρύνεται πάλι από την ατμόσφαιρα, μέσω της φωτοσύνθεσης των φυτών και μέσω της διάλυσης του στον ωκεανό. Η ποσότητα του φυσικά παραγόμενου CO₂ εξισορροπείται σχεδόν απόλυτα από την ποσότητα που αφαιρείται με φυσικό τρόπο.

Οι ανθρωπογενείς δραστηριότητες, όμως, έχουν επίδραση σε αυτό το ισοζύγιο, και είναι αυτές που σε μεγάλο βαθμό ευθύνονται για την υπερθέρμανση του πλανήτη. Αυτή η ασύμμετρη αύξηση συγκέντρωσης αρχικά συνδέεται με την καύση των ορυκτών καυσίμων. Η οξείδωση του άνθρακα, που αποθηκεύεται στα καύσιμα σε CO₂, είναι η διαδικασία που απελευθερώνει ενέργεια, γι' αυτό και η παραγωγή ενέργειας και οι εκπομπές CO₂ είναι άρρηκτα συνδεδεμένες διαδικασίες. Το 1896 ο Svante Arrhenius ήταν ο πρώτος που προέβλεψε, στη δημοσιευμένη μελέτη του “On the influence of carbonic acid in the air upon the temperature of the ground”, πως η έκλυση διοξειδίου του άνθρακα λόγω της καύσης ορυκτών καυσίμων θα δημιουργούσε πλανητική υπερθέρμανση.

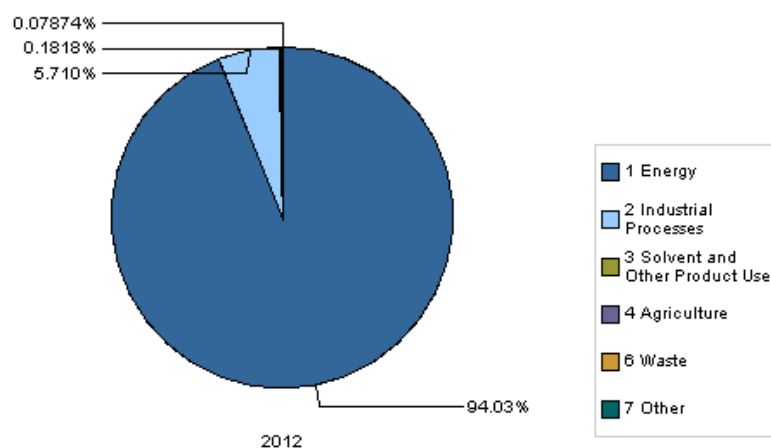
Σήμερα έχει υπολογισθεί από το σύνολο των ανθρωπίνων δραστηριοτήτων πως κατά την καύση των ορυκτών καυσίμων για παραγωγή ενέργειας παράγεται το 70-75% των

ειλυόμενων ποσοτήτων διοξειδίου του άνθρακα. Το υπόλοιπο 20-25% του αυξημένου ρυθμού έκλυσης οφείλεται στην αποψίλωση των δασών και στις εκπομπές από τις εξατμίσεις των μεταφορικών μέσων, σύμφωνα με έρευνα που δημοσιεύθηκε από το Πανεπιστήμιο Duke. Η διαδικασία αυτή απελευθερώνει περίπου ένα δισεκατομμύριο τόνους διοξειδίου του άνθρακα στην ατμόσφαιρα κάθε χρόνο, σύμφωνα με την 2010 Παγκόσμια εκτίμηση των δασικών πόρων.

Οι εκπομπές CO₂ επιταχύνονται προς τα πάνω από δεκαετία σε δεκαετία. Για τα τελευταία δέκα χρόνια (2005 - 2014), ο μέσος ετήσιος ρυθμός αύξησης είναι 2,11 ppm (parts per million). Αυτός ο ρυθμός αύξησης είναι περισσότερο από το διπλάσιο της αύξησης στη δεκαετία του 1960. Πριν από τη βιομηχανική επανάσταση του 19ου αιώνα, ο παγκόσμιος μέσος όρος εκπομπών CO₂ ήταν περίπου 280 ppm. Έρευνες δείχνουν ότι η ατμοσφαιρική συγκέντρωση του CO₂ αυξάνεται με επιταχυνόμενο ρυθμό επειδή τα ορυκτά καύσιμα καίγονται σε αυξημένο ρυθμό και μειώνεται η απορρόφηση του CO₂ από τον ωκεανό.

Οι ανθρωπογενείς πηγές των εκπομπών CO₂ προέρχονται κυρίως από τον ακόλουθο τομέα:

A. Ενέργεια (Energy) (αντιπροσωπεί το 94,03% του συνόλου εκπομπών CO₂ της ΕΕ-28 από όλες τις πηγές, το 2012 (Διάγραμμα 3.1))



Διάγραμμα 3.1. Κατανομή συνολικών εκπομπών CO₂ της ΕΕ-28 από όλες τις πηγές, το 2012

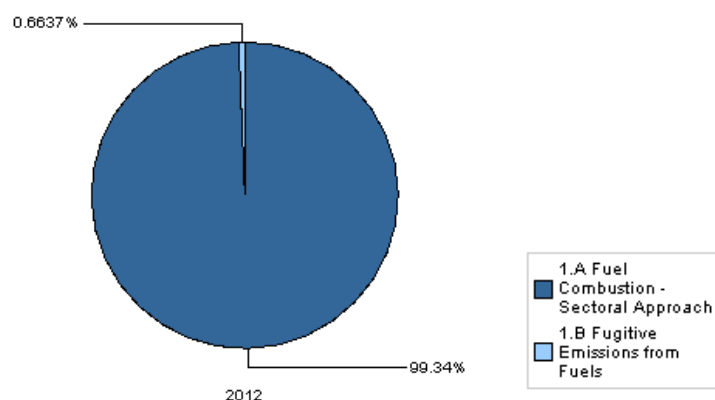
A. Ενέργεια (Energy)

Η καύση των καυσίμων (Fuel Combustion) (αντιπροσωπεί το 99,34% του συνόλου των εκπομπών CO₂ της ΕΕ-28 από την ενέργεια, το 2012) είναι η κύρια πηγή των εκπομπών CO₂ από τον τομέα της ενέργειας (Διάγραμμα 3.2). Οι εκπομπές προέρχονται τόσο από την καύση στον τομέα της ενέργειας, όπου το καύσιμο χρησιμοποιείται ως πηγή θερμότητας στη διύλιση ή την παραγωγή ηλεκτρικής ενέργειας, καθώς και από την καύση στην τελική κατανάλωση του καυσίμου ή δευτερογενών προϊόντων του. Τα αυτοκίνητά μας, τα φορτηγά, οι μονάδες παραγωγής ενέργειας και άλλες βιομηχανικές εγκαταστάσεις, χρειάζονται ενέργεια. Σχεδόν όλα τα οχήματα και οι εγκαταστάσεις χρησιμοποιούν κάποια μορφή καύσιμου και το καταναλώνουν για να παραχθεί ενέργεια.

Ο τομέας παραγωγής ηλεκτρισμού έχει την μεγαλύτερη ευθύνη παγκοσμίως όσον αφορά την κλιματική αλλαγή. Παράγει το 37% των ανθρωπογενών εκπομπών διοξειδίου του άνθρακα (CO₂), ειλύοντας 23 δισεκατομμύρια τόνους διοξειδίου του άνθρακα (CO₂) τον χρόνο, δηλαδή περισσότερους από 700 εκατομμύρια τόνους το δευτερόλεπτο. Οι εκπομπές CO₂ από την κατανάλωση καυσίμων εξαρτώνται κυρίως από την περιεκτικότητα του καυσίμου σε άνθρακα. Ο γαιάνθρακας (άνθρακας) είναι το πιο πλούσιο ορυκτό σε άνθρακα

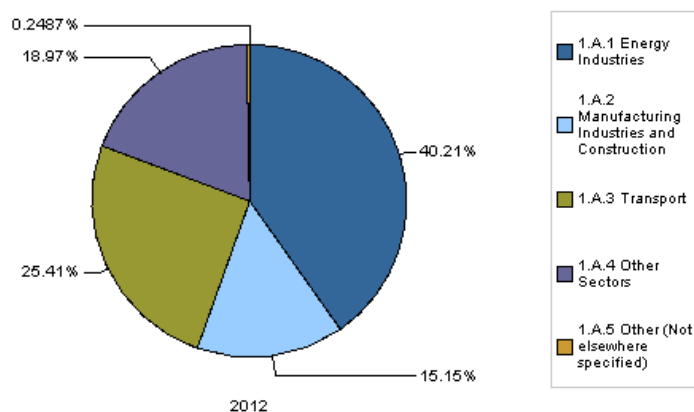
από όλα τα άλλα ορυκτά καύσιμα (π.χ. πετρέλαιο, φυσικό αέριο). Η καύση του παράγει 70% περισσότερο διοξείδιο του άνθρακα ανά μονάδα ενέργειας, σε σχέση με το φυσικό αέριο. Ο λιγνίτης είναι ο πιο «βρώμικος» από όλα τα είδη γαιάνθρακα και η καύση του παράγει τις περισσότερες εκπομπές CO₂ ανά μονάδα παραγόμενης ενέργειας. Αξίζει να σημειωθεί ότι η Ελλάδα εξακολουθεί να επιμένει στο λιγνίτη για την παραγωγή ηλεκτρισμού και το 2009 πάνω από το 55% του ηλεκτρισμού μας προήλθε από το συγκεκριμένο καύσιμο.

Επιπλέον, τα ορυκτά καύσιμα χρησιμοποιούνται και για σκοπούς πέραν της ενέργειας, όπως για παράδειγμα την παραγωγή αμμωνίας από φυσικό αέριο ή ασφάλτου από πετρέλαιο. Ορισμένες από αυτές τις εφαρμογές έχουν ως αποτέλεσμα την αποθήκευση του διοξειδίου του άνθρακα. Μέρος του άνθρακα που αποθηκεύεται μπορεί να οξειδώνεται γρήγορα όπως συμβαίνει με τον άνθρακα από την παραγωγή λιπασμάτων, λιπαντικών, απορρυπαντικών και πτητικών οργανικών διαλυτών.



Διάγραμμα 3.2. Κατανομή συνολικών εκπομπών CO₂ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012

Ως εκ τούτου, οι εκπομπές CO₂ πηγάζουν κυρίως από τις ενεργειακές βιομηχανίες, τις μεταφορές και τις μεταποιητικές βιομηχανίες και κατασκευές (Διάγραμμα 3.3):



Διάγραμμα 3.3. Κατανομή συνολικών εκπομπών CO₂ της ΕΕ-28 από τον τομέα της καύσης των καυσίμων, το 2012

- **Ενεργειακές βιομηχανίες (Energy Industries)** (αντιπροσωπεύουν το 39,95% του συνόλου των εκπομπών CO₂ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012 (Πίνακας 3.1))

1.A.1. Ενεργειακές Βιομηχανίες	1.396.222,19
1. Ενέργεια	3.495.198,68

Πίνακας 3.1. Συνολικές εκπομπές CO₂ από τον τομέα της Ενέργειας και από τις Ενεργειακές Βιομηχανίες για το 2012.

Η καύση των ορυκτών καυσίμων αντιπροσωπεύει το σημαντικότερο μερίδιο των εκπομπών της ενεργειακής βιομηχανίας. Στις ενεργειακές βιομηχανίες, τα ορυκτά καύσιμα είναι ταυτόχρονα και πρώτες ύλες για τις διαδικασίες μετατροπής, καθώς και πηγές ενέργειας για τη λειτουργία αυτών των διαδικασιών.

- **Μεταφορές (Transport)** (αντιπροσωπεύουν το 25,25% του συνόλου των εκπομπών CO₂ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012 (Πίνακας 3.2))

1.A.3. Μεταφορές	882.388,60
1. Ενέργεια	3.495.198,68

Πίνακας 3.2. Συνολικές εκπομπές CO₂ από τον τομέα της Ενέργειας και από τις Μεταφορές για το 2012.

Το διοξείδιο του άνθρακα κατέχει τη συντριπτική πλειοψηφία των εκπομπών από τον τομέα των μεταφορών. Στον τομέα αυτό η κατανάλωση ενέργειας εξαρτάται σχεδόν αποκλειστικά από τα εισαγόμενα ορυκτά καύσιμα, σε μεγάλο βαθμό με τη μορφή αργού πετρελαίου, που μετατρέπεται σε μια σειρά προϊόντων από πετρέλαιο και ντίζελ μέχρι βενζίνη, κηροζίνη ή υγραέριο. Οι εκπομπές CO₂ από τη μεταφοράς είναι αποτέλεσμα της καύσης αυτών των καυσίμων.

- **Μεταποιητικές βιομηχανίες και κατασκευές (Manufacturing Industries and Construction)** (αντιπροσωπεύουν το 15,05% του συνόλου των εκπομπών CO₂ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012 (Πίνακας 3.3))

1.A.2. Μεταποιητικές βιομηχανίες και κατασκευές	526.051,02
1. Ενέργεια	3.495.198,68

Πίνακας 3.3. Συνολικές εκπομπές CO₂ από τον τομέα της Ενέργειας και από τις Μεταποιητικές βιομηχανίες και κατασκευές για το 2012.

Στις μεταποιητικές και κατασκευαστικές βιομηχανίες τα καύσιμα χρησιμοποιούνται ως πηγή ενέργειας για τη μετατροπή των πρώτων υλών σε προϊόντα. Αυτό περιλαμβάνει την καύση των καυσίμων για την παραγωγή ηλεκτρικής ενέργειας και θερμότητας στις βιομηχανίες για δική τους χρήση.

- ✓ Λόγω του ότι οι εκπομπές CO₂ σχετίζονται κυρίως με την καύση ορυκτών καυσίμων, θα μπορούσε να θεωρηθεί η ύπαρξη σύνδεσης ανάμεσα στις εκπομπές και στην ακαθάριστη εγχώρια κατανάλωση (Gross Inland Consumption) ορυκτών καυσίμων, όπως λιθάνθρακα (hard coal) και λιγνίτη (lignite), φυσικού αερίου (natural gas), αργού πετρελαίου (crude oil) και προϊόντων πετρελαίου (petroleum products). Συνεπώς η ακαθάριστη εγχώρια κατανάλωση αυτών των καυσίμων είναι μια πιθανή επεξηγηματική μεταβλητή για τον υπολογισμό των εκπομπών CO₂ που προέρχονται από τον ενεργειακό τομέα.

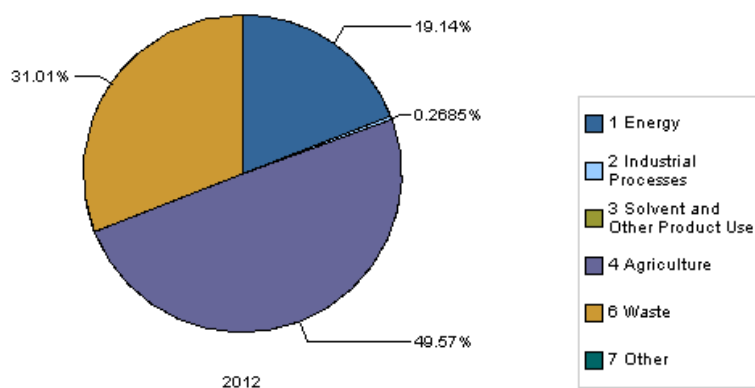
3.1.2. Μεθάνιο (Methane-CH₄)

Το μεθάνιο είναι ο δεύτερος πιο σημαντικός παράγοντας που συμβάλλει στην υπερθέρμανση του πλανήτη, μεταξύ των ανθρωπογενών αερίων μετά το διοξείδιο του άνθρακα. Ενδείκνυται ότι θα γίνει στο μέλλον το κυρίαρχο αέριο του θερμοκηπίου, λόγω του ότι ένας τόνος μεθανίου ισοδυναμεί με 21 τόνους CO₂ σε χρονικό διάστημα 100 χρόνων. Οι εκπομπές μεθανίου προκύπτουν τόσο από φυσικές όσο και ανθρωπογενείς πηγές, συμπεριλαμβανομένων των αποβλήτων, της ενέργειας και της γεωργίας. Σε παγκόσμιο επίπεδο, η μεγαλύτερη ανθρωπογενής συμβολή στις εκπομπές CH₄ προέρχεται από την κτηνοτροφία, ακολουθούμενη από τις διαρροές από την παραγωγή φυσικού αερίου και την εξόρυξη του άνθρακα, αλλά και την καλλιέργεια ρυζιού. Συνδέεται περισσότερο με την παραγωγή τροφίμων και την πληθυσμιακή αύξηση.

Από τη στιγμή που θα απελευθερωθεί στην ατμόσφαιρα έχει διάρκεια ζωής περίπου δώδεκα χρόνια. Παρόλο που θεωρείται σχετικά βραχύβιο αέριο, η διάρκεια ζωής του εξακολουθεί να είναι αρκετά μεγάλη ώστε να μπορεί να μεταφερθεί σε άλλες περιοχές. Το 2009, μια ομάδα Βρετανών και Γερμανών ερευνητών πραγματοποίησε έρευνα στα ανοικτά των ακτών της Νορβηγίας για να παρατηρήσει το μεθάνιο να απελευθερώνεται από τον βυθό που «έλιωνε». Σε περιοχές κοντά στους πόλους, ένα τμήμα της μάζας εδάφους ή του βυθού είναι μόνιμα παγωμένο. Σύμφωνα με κάποιες εκτιμήσεις, το στρώμα αυτό, γνωστό ως στρώμα αενάως παγωμένων εδαφών, περιέχει τη διπλάσια ποσότητα άνθρακα από αυτή που υπάρχει επί του παρόντος στην ατμόσφαιρα. Υπό θερμότερες συνθήκες, ο άνθρακας αυτός μπορεί να απελευθερωθεί από βιομάζα σε αποσύνθεση, είτε ως διοξείδιο του άνθρακα είτε ως μεθάνιο.

Στην Ευρώπη, η κατάταξη των σημαντικότερων πηγών είναι διαφορετική δεδομένου ότι ορισμένες από αυτές τις πηγές είτε δεν συμβάλλουν σημαντικά στις εκπομπές της Ευρώπης είτε είναι ήδη υπό έλεγχο. Η μεγαλύτερη συνεισφορά προέρχεται από τους ακόλουθους τομείς (Διάγραμμα 3.4):

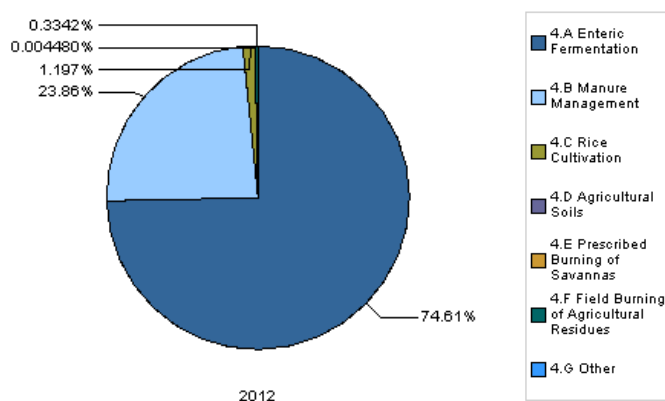
- A. Γεωργία (Agriculture)** (αντιπροσωπεύοντας το 49,57% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από όλες τις πηγές το 2012)
- B. Απόβλητα (Waste)** (αντιπροσωπεύοντας το 31,01% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από όλες τις πηγές το 2012)
- C. Ενέργεια (Energy)** (αντιπροσωπεύοντας το 19,14% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από όλες τις πηγές το 2012)



Διάγραμμα 3.4. Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από όλους τους τομείς, το 2012

A. Γεωργία (Agriculture)

Η γεωργία είναι η κύρια πηγή εκπομπών μεθανίου ως αποτέλεσμα της εντερικής ζύμωσης και διαχείρισης των ζωικών λιπασμάτων(κοπριά) (Διάγραμμα 3.5):



Διάγραμμα 3.5. Κατανομή συνολικών εκπομπών CH₄ της EE-28 από τη γεωργία, το 2012

- **Εντερική ζύμωση (Enteric Fermentation)** (αντιπροσωπεύοντας το 74,61% του συνόλου των εκπομπών CH₄ της EE-28 από την γεωργία το 2012)

Οι εκπομπές μεθανίου (CH₄) από γεωργικές πηγές είναι κυρίως από την εντερική ζύμωση κατά τη διάρκεια της διαδικασίας της πέψης των μηρυκαστικών. Τα ζώα που θεωρούνται κύριοι παραγωγοί μεθανίου είναι τα βοοειδή, τα πρόβατα και οι χοίροι. Η ποσότητα των εκπομπών προέρχεται από το μεγάλο πληθυσμό τους και το υψηλό ποσοστό των εκπομπών μεθανίου λόγω του πεπτικού τους συστήματος. Κάθε αγελάδα κατά την πέψη της τροφής της εκπέμπει 200 έως 400 g μεθανίου ημερησίως. Έχει εκτιμηθεί ότι, σε παγκόσμια κλίμακα, τα εντερικά αέρια οικιακών ζώων συνεισφέρουν ετησίως περί τα 80 έως 100 εκατομμύρια τόνους μεθανίου.

- **Διαχείριση των ζωικών λιπασμάτων (Manure Management)** (αντιπροσωπεύοντας το 23,86% του συνόλου των εκπομπών CH₄ της EE-28 από την γεωργία το 2012)

Η διαχείριση των ζωικών λιπασμάτων (αποθήκευση ή εφαρμογή τους στη γη) είναι η δεύτερη πιο σημαντική πηγή εκπομπών μεθανίου από τη γεωργία. Οι πιο σημαντικές εκπομπές παράγονται από τη διαχείριση των υγρών απόβλητων αφού τείνει να παράγει περισσότερο CH₄ από την διαχείριση των στερεών. Κατά την αναερόβια αποσύνθεση του οργανικού υλικού της κοπριάς στην επεξεργασία υγρής κοπριάς σε δεξαμενές ή τεχνητές λίμνες εκλύονται σημαντικά ποσά μεθανίου. Αντίθετα, η κοπριά που προστίθεται ως βελτιωτικό εδάφους στις καλλιέργειες δεν εκλύει σημαντικά ποσά μεθανίου.

- ✓ Ο αριθμός των ζώων, δηλαδή των βοοειδών (cattle population), των χοίρων (pig population) και των προβάτων (sheep population), μπορεί να χρησιμοποιηθεί ως επεξηγηματική μεταβλητή για την εκτίμηση του όγκου των εκπομπών μεθανίου.

B. Απόβλητα (Waste)

Οι εκπομπές αερίων του θερμοκηπίου της EE-28 που προέρχονται από τα απόβλητα αυτά εκτιμάται ότι είναι 140.802,82 τόνους σε ισοδυναμία CO₂ το 2012. Από το ποσό αυτό, το μεγαλύτερο μέρος προκύπτει από τους χώρους υγειονομικής ταφής απορριμμάτων (που εκπέμπουν μεθάνιο) (Διάγραμμα 3.6), ενώ η διαχείριση των οικιακών και βιομηχανικών λυμάτων είναι επίσης μια σχετικά σημαντική πηγή. Τα απόβλητα είναι η δεύτερη σημαντικότερη πηγή εκπομπών CH₄, που προκύπτει από τις δραστηριότητες

διαχείρισης απορριμμάτων. Οι δραστηριότητες αυτές συνήθως περιλαμβάνουν τη διάθεση των στερεών αποβλήτων και την επεξεργασία των λυμάτων:

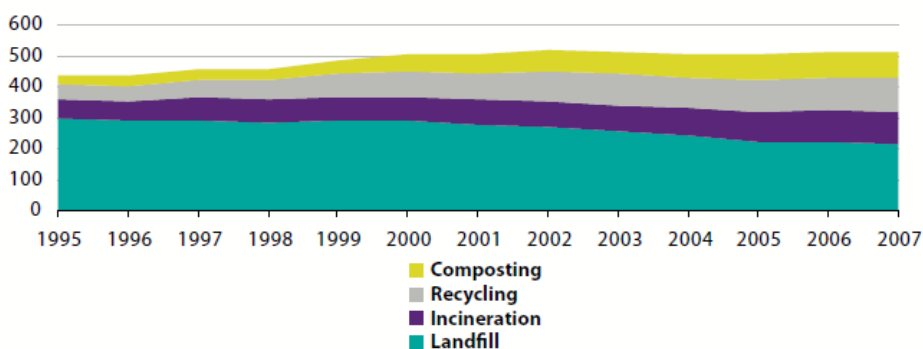
- **Διαχείριση απορριμμάτων (Solid waste disposal)** (αντιπροσωπεύοντας το 82,41% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από τα απόβλητα, το 2012)

Η επεξεργασία και διάθεση των αστικών, βιομηχανικών και άλλων στερεών αποβλήτων οδηγούν σε σημαντικές ποσότητες των εκπομπών μεθανίου. Στην πραγματικότητα, η διάθεση στερεών αποβλήτων στην γη είναι η κύρια πηγή εκπομπών μεθανίου από τη διαχείριση των αποβλήτων. Στις αναπτυσσόμενες χώρες, η πλειοψηφία των εκπομπών προκύπτει λόγω των χαμηλών προτύπων για τις τεχνολογίες επεξεργασίας και διάθεσης λυμάτων, ενώ στις βιομηχανικές χώρες η κυρίαρχη πηγή εκπομπών είναι οι ΧΥΤΑ (Χώροι Υγειονομικής Ταφής), όπου το μεθάνιο παράγεται ως αποτέλεσμα της αποδόμησης της οργανικής ύλης υπό αναερόβιες (χωρίς οξυγόνο) συνθήκες. Η διασπώμενη οργανική ύλη στα αστικά στερεά απόβλητα σχετίζεται με τη λάσπη και τα βιομηχανικά απόβλητα όπως τρόφιμα, χαρτί, ξύλο, υφάσματα, κ.λπ.

Σχεδόν τα δύο τρίτα (62,4%) των αστικών αποβλήτων στην ΕΕ-28 αντιμετωπίζονταν μέσω της υγειονομικής ταφής, το 1995. Το ποσοστό αυτό μειώθηκε από 296 kg ανά κάτοικο στα 157 kg μέχρι το 2012. Το επίπεδο των εκπομπών μεθανίου από τη διάθεση στερεών αποβλήτων είναι στενά συνδεδεμένο με τη σύνθεση των αποβλήτων, οι οποίες μπορεί να διαφέρουν σημαντικά μεταξύ των χωρών και των περιφερειών. Η περιεκτικότητα σε υγρασία είναι ένας άλλος σημαντικός παράγοντας που συνδέεται με την παραγωγή μεθανίου εντός των εγκαταστάσεων υγειονομικής ταφής. Κάτω από αναερόβιες συνθήκες, οι θερμοκρασίες συχνά φθάνουν άνω των 30 ° C και διατηρούνται ανεξάρτητα από τη θερμοκρασία της επιφάνειας, και περαιτέρω προωθούν την ανάπτυξη βακτηρίων.

Οι εκπομπές μεθανίου από τις εγκαταστάσεις υγειονομικής ταφής χαρακτηρίζεται από μια καθυστέρηση μεταξύ των αποβλήτων που διατίθενται και των εκπομπών που παράγονται. Τα απόβλητα αρχίζουν να παράγουν μεθάνιο μετά από μια περίοδο άνω του ενός έτους, και για την παραγωγή φυσικού αερίου έπειτα από περίπου τέσσερα έως δέκα χρόνια μετά από την υγειονομική ταφή, ενώ τα αέρια μπορούν να συνεχίσουν να διαδίδονται έως και 50 ή 60 χρόνια. Αυτή η διαδικασία είναι σε άμεση αντίθεση με την αποτέφρωση των αποβλήτων, όπου τα αέρια εκπέμπονται κατά τη διάρκεια της επεξεργασίας των αποβλήτων.

Figure 2.50: Municipal waste by type of treatment, EU-27 (kg per capita)



Διάγραμμα 3.6. Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από τα αστικά απόβλητα με βάση τον τύπο διαχείρισής τους, από το 1995 μέχρι το 2007

Ως αποτέλεσμα, η ποσότητα των στερεών αποβλήτων που εναποτίθεται στη γη είναι συναφείς με τις ακόλουθες μεταβλητές (Διάγραμμα 3.7):

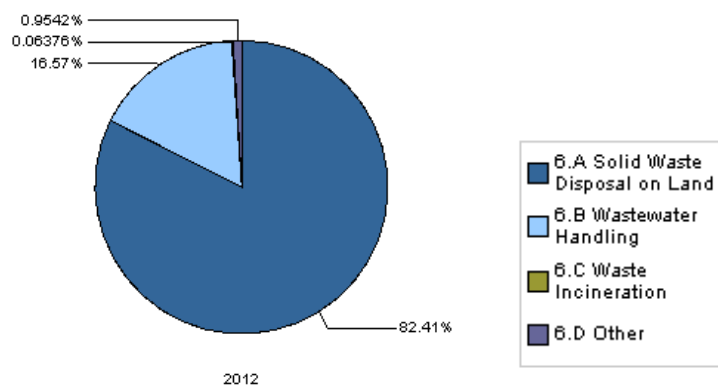
- **Στερεά απόβλητα που εναποτίθενται επί ή εντός του εδάφους (Solid waste deposited onto or into land):** Το μεθάνιο παράγεται από την αποδόμηση της οργανικής ύλης. Ως εκ τούτου, στοιχεία για την ποσότητα των στερεών αποβλήτων που αφορούν τις εξής κατηγορίες : ζωικά και φυτικά υπολείμματα (εκτός από ζωικά υπολείμματα από την παρασκευή τροφίμων και προϊόντων, και τα ζωικά περιττώματα), οικιακά απόβλητα, κοινή λάσπη θεωρούνται πιθανές μεταβλητές.

- **Αστικά απόβλητα σε χώρους υγειονομικής ταφής (Municipal waste landfilled):** Αυτά τα απόβλητα προέρχονται κατά κανόνα από νοικοκυριά και από άλλες μικρές επιχειρήσεις, εστιατόρια και γραφεία. Αποτελούνται από υλικά, συμπεριλαμβανομένων, μεταξύ άλλων, το χαρτί, τα πλαστικά, τα τρόφιμα, τις οικιακές συσκευές και άλλα μη-επικίνδυνα υλικά.

- ✓ Ως εκ τούτου, τα αστικά απόβλητα που έχουν επεξεργαστεί μέσω υγειονομικής ταφής (municipal waste) είναι μια πιθανή επεξηγηματική μεταβλητή για την εκτίμηση του επιπέδου των εκπομπών μεθανίου.
- ✓ Επιπρόσθετα, ο αστικός πληθυσμός (urban population), ο συνολικός πληθυσμός (total population) αλλά και το ακαθάριστο εγχώριο προϊόν (GDP) μπορούν να χρησιμοποιηθούν ως πιθανές μεταβλητές για την εκτίμηση των εκπομπών από τη διάθεση των στερεών αποβλήτων.
- **Διαχείριση υγρών αποβλήτων (Wastewater Handling)** (που αντιστοιχεί στο 16,57% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από τα απόβλητα, το 2012)

Η διαχείριση των υγρών αποβλήτων, με υψηλό οργανικό περιεχόμενο, υπό σε αναερόβιες συνθήκες προκαλεί μεγάλες ποσότητες εκπομπών μεθανίου. Αν τα λύματα επεξεργάζονται αερόβια, δεν προκύπτουν εκπομπές. Οι μέθοδοι διαχείριση υγρών αποβλήτων διαφέρουν μεταξύ των ανεπτυγμένων και των αναπτυσσόμενων χωρών. Στις περισσότερες αναπτυγμένες χώρες, τα λύματα επεξεργάζονται αερόβια χωρίς εκπομπές μεθανίου ή στις κεντρικές αερόβιες μονάδες επεξεργασίας λυμάτων και λιμνοθάλασσες, όπου οι εκπομπές είναι μικρές. Σε πολλές αναπτυσσόμενες χώρες, τα λύματα συχνά εναποτίθενται κατευθείαν στο νερό ή επεξεργάζονται σε αναερόβιες συνθήκες.

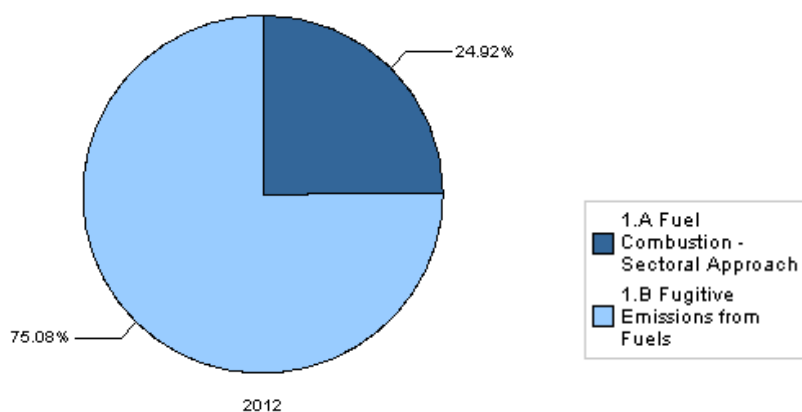
- ✓ Δεδομένου ότι η παρούσα μελέτη εξετάζει τις ευρωπαϊκές χώρες όπου το μεγαλύτερο μέρος των λυμάτων υποβάλλεται σε επεξεργασία υπό αερόβιες συνθήκες, οι εκπομπές υπολογίζονται ως συνάρτηση του πληθυσμού (total population).



Διάγραμμα 3.7. Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από τα απόβλητα, το 2012

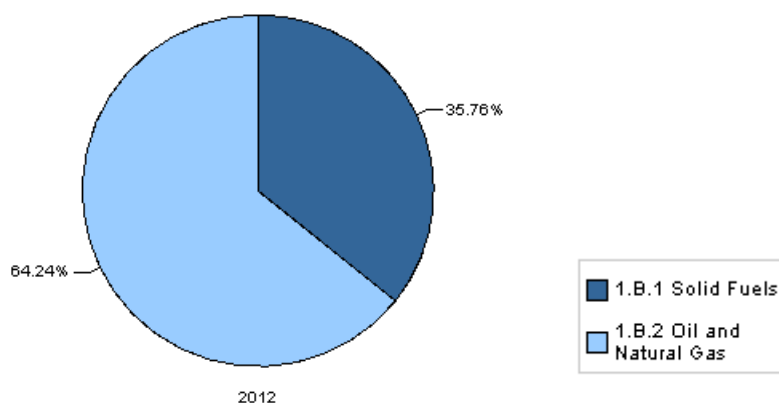
Γ. Ενέργεια (Energy)

Οι εκπομπές του μεθανίου και του υποξειδίου του αζώτου συχνά προκύπτουν από τις διαφεύγουσες εκπομπές (Fugitive Emissions), δηλαδή εκπομπές που οφείλονται σε διαρροές του εξοπλισμού, εξατμιστικές διεργασίες και απώλειες λόγω του ανέμου κατά την παραγωγή, μεταποίηση, μετάδοση, αποθήκευση και χρήση των καυσίμων. Κατά την περίοδο 1990-2012, οι εκπομπές αυτές έχουν ελαττωθεί σχεδόν κατά το ήμισυ (48,3%) και το μερίδιο των διάχυτων εκπομπών στο σύνολο των εκπομπών αερίων του θερμοκηπίου μειώθηκαν από 2,8% σε 1,8% στην ΕΕ-28.



Διάγραμμα 3.8: Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από την ενέργεια, το 2012

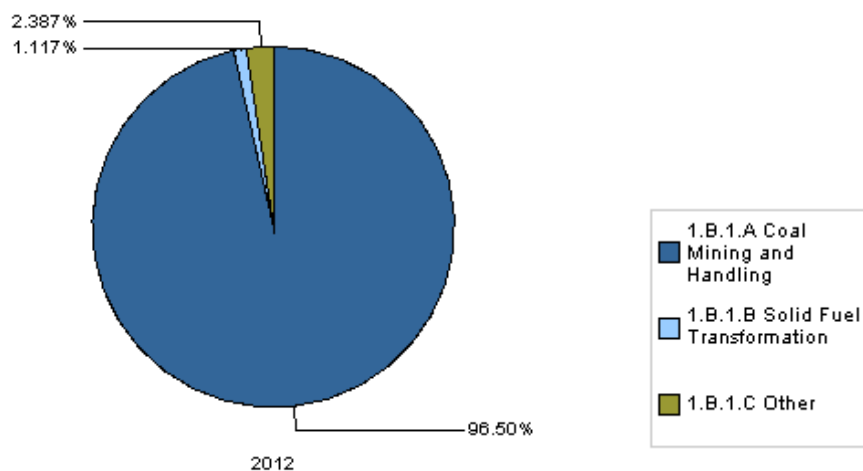
- **Διαφεύγουσες εκπομπές από καύσιμα (Fugitive Emissions from Fuels)** (αντιπροσωπεύουν το 75,08% του συνόλου των εκπομπών της ΕΕ-28 του CH₄ από την ενέργεια το 2012 (Διάγραμμα 3.8)) είναι η κύρια πηγή εκπομπών CH₄ από τον ενεργειακό τομέα. Οι διαφεύγουσες εκπομπές από καύσιμα προέρχονται είτε από τα ορυκτά καύσιμα είτε από το πετρέλαιο και το φυσικό αέριο (Διάγραμμα 3.9).



Διάγραμμα 3.9: Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από τα απόβλητα, το 2012

Κύριες πηγές τους είναι η εξόρυξη και επεξεργασία άνθρακα (Διάγραμμα 3.10) και φυσικού αερίου (Διάγραμμα 3.11).

- **Εξόρυξη και επεξεργασία του άνθρακα** (αντιστοιχεί στο 25,9% του συνόλου των εκπομπών του CH₄ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012 (Πίνακας 3.4))



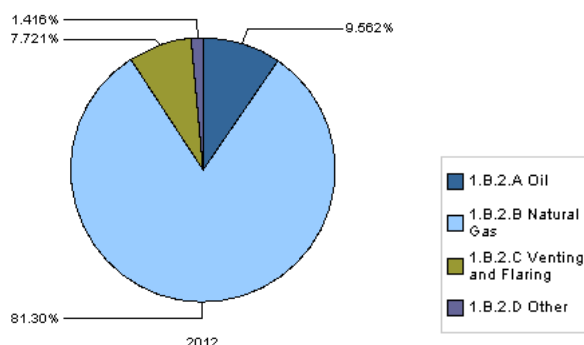
Διάγραμμα 3.10.: Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από τα στερεά καύσιμα, το 2012

1.B.1.A. Εξόρυξη και επεξεργασία του άνθρακα	19.750,19
1.Ενέργεια	76.239,95

Πίνακας 3.4. Συνολικές εκπομπές CH₄ από τον τομέα της Ενέργειας και από την Εξόρυξη και επεξεργασία του άνθρακα για το 2012.

Η διαδικασία σχηματισμού του άνθρακα παράγει μεθάνιο, το οποίο απελευθερώνεται στην ατμόσφαιρα όταν ο άνθρακας εξορύσσεται. Η απελευθέρωση μεθανίου είναι υψηλότερη κατά την υπόγεια εξόρυξη. Επιπλέον, υπάρχουν και εκπομπές από τις μετέπειτα δραστηριότητες, όπως η επεξεργασία, η μεταφορά και η χρήση του άνθρακα.

- ✓ Δεδομένου ότι συμβαίνουν πιο διάχυτες εκπομπές στα ορυχεία πιθανή μεταβλητή που μπορεί να χρησιμοποιηθεί είναι η πρωτογενής παραγωγή λιθάνθρακα και λιγνίτη (primary production of hard coal and lignite).
- **Φυσικό αέριο** (αντιστοιχεί στο 39,21% του συνόλου των εκπομπών CH₄ της ΕΕ-28 από τον τομέα της ενέργειας, το 2012 (Πίνακας 3.5))



Διάγραμμα 3.11.: Κατανομή συνολικών εκπομπών CH₄ της ΕΕ-28 από το πετρέλαιο και το φυσικό αέριο, το 2012

1.B.2.B. Φυσικό αέριο	29.895,11
1.Ενέργεια	76.239,95

Πίνακας 3.5. Συνολικές εκπομπές CH₄ από τον τομέα της Ενέργειας και από το Φυσικό Αέριο για το 2012.

Κατά τη διάρκεια της παραγωγής του φυσικού αερίου παρουσιάζονται εκπομπές μεθανίου. Ωστόσο, οι απώλειες του φυσικού αερίου κατά τη μεταφορά και την τελική χρήση είναι επίσης μια σημαντική πηγή εκπομπών μεθανίου. Οι κατευθυντήριες γραμμές IPCC (Intergovernmental Panel on Climate Change) προτείνουν ως συντελεστές εκπομπής για τις απώλειες κατά τη μεταφορά και διανομή, το μεθάνιο που χάνεται ανά μονάδα φυσικού αερίου που καταναλώνεται για τις χώρες της Δυτικής Ευρώπης και ανά μονάδα παραγόμενου αερίου για τις χώρες της Ανατολικής Ευρώπης. Σε αυτό το πλαίσιο, οι ακόλουθες μεταβλητές κρίνονται κατάλληλες:

- ✓ Πρωτογενής παραγωγή φυσικού αερίου (Primary production of natural gas): η οποία καλύπτει το μεθάνιο που εκπέμπεται κατά τη διάρκεια της παραγωγής του φυσικού αερίου.
- ✓ Ακαθάριστη εγχώρια κατανάλωση φυσικού αερίου (Gross inland consumption of natural gas): Η ακαθάριστη εγχώρια κατανάλωση ενέργειας αντιπροσωπεύει την ποσότητα της ενέργειας που απαιτείται για να ικανοποιήσει την εσωτερική ζήτηση μιας χώρας ή μιας περιοχής. Η μεταβλητή θεωρείται κατάλληλη, δεδομένου ότι οι χώρες της Δυτικής Ευρώπης κυρίως καταναλώνουν και δεν παράγουν αέριο.
- ✓ Απώλειες διανομής (Distribution Losses): Οι απώλειες διανομής αποτελούνται από απώλειες κατά τη μεταφορά, καθώς εκπέμπεται μεθάνιο κατά τη διανομή του φυσικού αερίου στους καταναλωτές.

3.1.3. Υποξείδιο του αζώτου (Nitrogenous Oxide – N₂O)

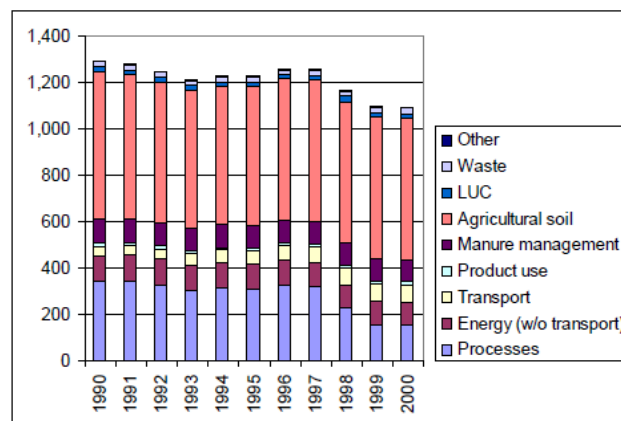
Το υποξείδιο του αζώτου αποτελεί αέριο του θερμοκηπίου, έχοντας την ικανότητα να απορροφά την υπέρυθη ακτινοβολία 310 φορές εντονότερα απ' όσο το διοξείδιο του άνθρακα. Έρχεται τρίτο σε σειρά σπουδαιότητας από άποψη συμβολής αέριων ουσιών ανθρωπογενούς προέλευσης μετά το διοξείδιο του άνθρακα και το μεθάνιο, αντιπροσωπεύοντας περίπου το 6% της συνολικής συνεισφοράς του ανθρώπου στην αύξηση των αερίων θερμοκηπίου.

Το υποξείδιο του αζώτου (N₂O) είναι πλέον η πιο ισχυρή χημική ένωση που καταστρέφει το όζον στο ανώτερο στρώμα της ατμόσφαιρας, σύμφωνα με μια νέα αμερικανική επιστημονική μελέτη. Οι ανθρώπινες δραστηριότητες είναι υπεύθυνες για το ένα τρίτο περίπου των συνολικών εκπομπών υποξειδίου του αζώτου στην ατμόσφαιρα. Τα υπόλοιπα δύο τρίτα παράγονται με φυσικό τρόπο, όταν τα βακτήρια στο έδαφος και τους ωκεανούς διασπούν τις ενώσεις που περιέχουν άζωτο. Το εκπεμπόμενο αέριο ανυψώνεται στη στρατόσφαιρα, όπου το μεγαλύτερο μέρος του διασπάται από την ηλιακή ακτινοβολία σε αβλαβή μόρια αζώτου και οξυγόνου. Όμως ένα τμήμα του δεν διασπάται και μπορεί να παραμείνει επί εκατοντάδες χρόνια στην ατμόσφαιρα, όπου αντιδρά με υψηλής ενέργειας άτομα οξυγόνου και παράγει μια πιο καταστροφική ουσία, το μονοξείδιο του αζώτου, το οποίο καταστρέφει το όζον. Κατά συνέπεια, αρκετά μικρές συγκεντρώσεις N₂O είναι επαρκείς για να το κάνουν ένα σημαντικό αέριο του θερμοκηπίου. Σήμερα, η μέτρηση, ο έλεγχος της συγκέντρωσης του N₂O και ο περιορισμός των εκπομπών του επιβάλλονται,

επειδή δρα τόσο ως αέριο θερμοκηπίου, όσο και ως αέριο καταστροφικό για τη στιβάδα του όζοντος.

Αν και σε ίχνη, το υποξείδιο του αζώτου πάντοτε αποτελούσε κανονικό συστατικό της ατμόσφαιρας. Ωστόσο, η συγκέντρωσή του στην ατμόσφαιρα αυξάνεται γραμμικά με τον χρόνο με ρυθμό 0,2% έως 0,3%, λόγω των ανθρωπίνων δραστηριοτήτων. Σε μια σχετικά πρόσφατη, του 2010, έκθεση της Υπηρεσίας Προστασίας του Περιβάλλοντος (Environmental Protection Agency) εκτιμάται πως η ετησίως εκλυόμενη ποσότητα N₂O στην ατμόσφαιρα είναι κατά μέσο όρο 18,8 εκατομμύρια τόνοι, από τα οποία 12,1 εκατομμύρια τόνοι (το 64%) προέρχεται από φυσικές πηγές, αν και τα δεδομένα αλλάζουν από χρόνο σε χρόνο. Χονδρικά, το 1/3 του ετησίως εκλυόμενου στην ατμόσφαιρα N₂O οφείλεται στις δραστηριότητες του ανθρώπου. Από τις ανθρωπογενείς πηγές N₂O, κυριότερη είναι η γεωργία, με ετήσια συνεισφορά 2 έως 4 εκατομμυρίων τόνων. Παράγεται από βακτηρίδια τόσο κατά τις διαδικασίες νιτροποίησης (nitrification), δηλαδή της οξείδωσης της αμμωνίας σε νιτρώδη και στη συνέχεια σε νιτρικά, όσο και κατά τις διαδικασίες από-νιτροποίησης (de-nitrification), δηλαδή της μετατροπής του αμμωνιακού αζώτου σε νιτρικά. Έτσι, μεγάλες ποσότητες N₂O παράγονται κατά τη βακτηριακή διάσπαση μέρους των τεραστίων ποσοτήτων αμμωνιακών ή αζωτούχων λιπασμάτων, όπως επίσης και των απορριμμάτων των εκτρεφόμενων ζώων.

Ο αγροτικός τομέας –κυρίως λόγω της χρήσης αζωτούχων λιπασμάτων- είναι η κύρια πηγή υποξειδίου του αζώτου και ακολουθούν οι τομείς της ενέργειας και της βιομηχανίας. Το υποξείδιο του αζώτου είναι σε μεγάλο βαθμό αποτέλεσμα των βιολογικών διεργασιών που συμβαίνουν στο έδαφος σε μεγάλες περιοχές (Διάγραμμα 3.12). Το μεγάλο χρονικό διάστημα ατμοσφαιρικής παραμονής του (περίπου 120 χρόνια) επιτρέπει μόνο μικρές ποσότητες κατά τη διάρκεια των συγκεντρώσεων, τα οποία είναι δύσκολο να εντοπιστούν από τις μετρήσεις. Δεδομένου ότι οι διεργασίες του εδάφους είναι επίσης ελάχιστα κατανοητές, η αβεβαιότητα που συνδέεται με ένα σημαντικό μέρος των εκπομπών είναι συνεπώς πολύ υψηλή. Κατά συνέπεια, η τρέχουσα πρόβλεψη των εκπομπών υποξειδίου του αζώτου επικεντρώνεται κυρίως στις πηγές που έχουν γίνει αρκετά κατανοητές.

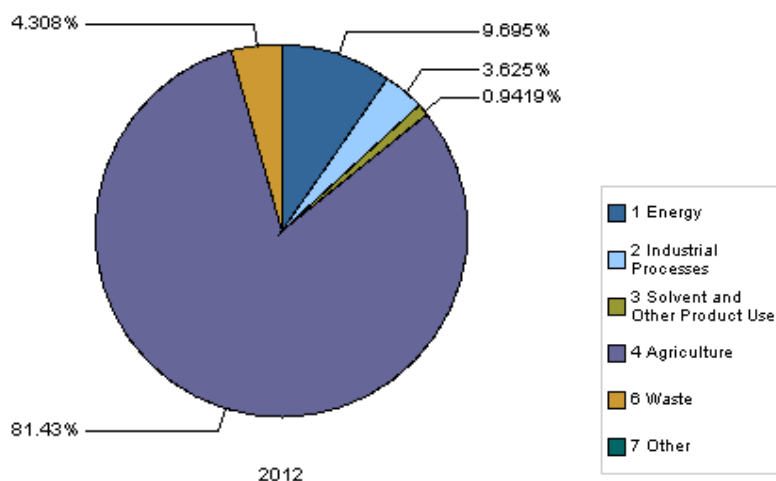


Διάγραμμα 3.12.: Κατανομή συνολικών εκπομπών N₂O της ΕΕ-28 από όλες τις πηγές, 1990-2012

Η μεγαλύτερη συμβολή στις εκπομπές N₂O προέρχεται από τους ακόλουθους τομείς (Διάγραμμα 3.13):

A. Γεωργία (Agriculture) (αντιπροσωπεύοντας το 81,43% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από όλες τις πηγές, το 2012)

B. Ενέργεια (Energy) (αντιπροσωπεύοντας το 9,7% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από όλες τις πηγές, το 2012)



Διάγραμμα 3.13.: Κατανομή συνολικών εκπομπών N₂O της ΕΕ-28 από όλες τις πηγές, το 2012

A. Γεωργία (Agriculture)

Οι μικροβιακές διεργασίες στο έδαφος και στην κοπριά (διαδικασίες νιτροποίησης και απο-νιτροποίησης) θεωρούνται οι κύριες πηγές του το υποξειδίου του αζώτου στην Ευρώπη. Μια βασική παράμετρος που επηρεάζει την ποσότητα των εκπομπών N₂O είναι η διαθεσιμότητα του αζώτου. Οι κατευθυντήριες γραμμές για την υποβολή εκθέσεων των εκπομπών της IPCC προτείνει ένα ενιαίο συντελεστή εκπομπών που σχετίζεται με την είσοδο του αζώτου. Η αβεβαιότητα του υπολογισμού των εκπομπών, με βάση την προσέγγιση αυτή εκτιμήθηκε σε δύο τάξεις μεγέθους (1997).

Οι εκπομπές αερίων θερμοκηπίου από γεωργικές πρακτικές είναι κατά κύριο λόγο με τη μορφή υποξειδίου του αζώτου (N₂O) που προκύπτει από την εφαρμογή των λιπασμάτων και κοπριάς, είτε με τη μορφή του μεθανίου (CH₄), όπως προαναφέρθηκε. Η γεωργία είναι η κύρια πηγή εκπομπών N₂O:

- **Γεωργικά εδάφη (Agricultural soils)** (αντιπροσωπεύοντας το 88,71% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από τη γεωργία, το 2012 (Διάγραμμα 3.14))

Μια βασική παράμετρος στις διεργασίες του εδάφους είναι η διαθεσιμότητα του αζώτου στο έδαφος, η οποία είναι η μόνη παράμετρος που περιλαμβάνεται στην προσέγγιση της IPCC. Πιθανές πηγές εισροής αζώτου στο έδαφος, προέρχονται από τρεις οδούς (GAINS, 2005):

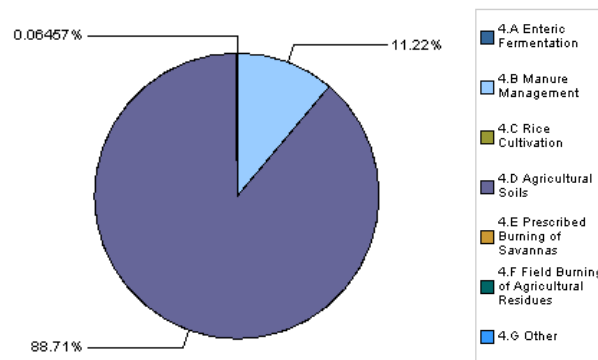
- ✓ **Εφαρμογή μεταλλικών λιπασμάτων (Mineral fertilizer application):**
Αναφέρεται για το λίπασμα που καταναλώνεται, συχνά καταγράφεται ως πωλήσεις λιπασμάτων και / ή ως παραγωγή λιπάσματος του αζώτου (N) που εφαρμόζεται στα

εδάφη. Ως εκ τούτου, η αζωτούχα κατανάλωση λιπασμάτων (nitrogenous consumption of fertilizers) και η αζωτούχα παραγωγή λιπασμάτων (nitrogenous fertilizers production) θεωρούνται κατάλληλες επεξηγηματικές μεταβλητές για την εκτίμηση των εκπομπών N₂O που προέρχονται από την εφαρμογή λιπασμάτων αζώτου. Παρόλα αυτά τα δεδομένα για τους πιο πάνω παράγοντες δεν είναι διαθέσιμα για όλες τις χρονικές περιόδους και έχουν πολλά κενά. Για αυτό, ως εναλλακτική προσεγγιστική μεταβλητή μπορεί να χρησιμοποιηθεί η ακαθάριστη προστιθέμενη αξία για την παραγωγή χημικών ουσιών και προϊόντων (Gross Value Added for Manufacture of chemicals and chemical products)

- ✓ **Εκτρεφόμενα ζώα και ζωικά απόβλητα (Farm animals and animal waste):** Αναφέρεται στο ποσό των ζωικών αποβλήτων που εναποτίθεται σε βοσκοτόπους, εκτάσεις και μάντρες. Μπορεί να υπολογιστεί από τον αριθμό των ζώων κάθε είδους. Ο αριθμός των ζώων (cattle population, pig population, sheep population, goat population) μπορεί να χρησιμοποιηθεί ως επεξηγηματική μεταβλητή για τον υπολογισμό των εκπομπών N₂O.
- ✓ **Υπολείμματα καλλιεργειών (Crop residues):** Αναφέρεται στο ποσό της περιεκτικότητας σε άζωτο των υπολειμμάτων των καλλιεργειών. Εκτιμάται από τα στατιστικά στοιχεία της απόδοσης της παραγωγής των καλλιεργειών. Μια πιθανή επεξηγηματική μεταβλητή για την εκτίμηση των εκπομπών N₂O είναι η φυτική παραγωγή (Crop production).
- **Διαχείριση των ζωικών αποβλήτων (Manure management)** (αντιπροσωπεύοντας το 11,22% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από τη γεωργία, το 2012)

Οι εκπομπές υποξειδίου του αζώτου από τα απόβλητα των ζώων σχετίζονται με την ποσότητα των ζωικών αποβλήτων που εκκρίνεται σε στάβλους (εσωτερική έκκριση). Εκπομπές από τα ζωικά απόβλητα θα πρέπει να αντιμετωπίζονται ξεχωριστά από τις εκπομπές που προκύπτουν από τη χρήση των στερεών ζωικών αποβλήτων στο έδαφος. Ωστόσο, οι ίδιες διεργασίες (μικροβιακή νιτροποίηση και από-νιτροποίηση) είναι υπεύθυνες για το σχηματισμό του N₂O κατά την αποθήκευση των αποβλήτων αυτών.

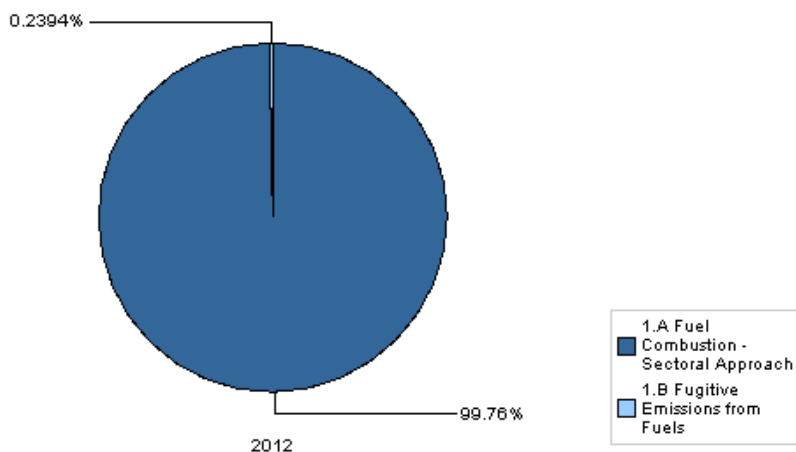
- ✓ Για την κατηγορία αυτή μπορούμε να χρησιμοποιήσουμε τον πληθυσμό των ζώων, δηλαδή των βοοειδών (cattle population), χοίρων (pig population), προβάτων (sheep population) και κατσιών (goat population) ως μεταβλητή για τον υπολογισμό του όγκου των εκπομπών N₂O.



Διάγραμμα 3.14.: Κατανομή συνολικών εκπομπών N₂O της ΕΕ-28 από τη γεωργία, το 2012

B. Ενέργεια (Energy)

Η καύση των καυσίμων (αντιστοιχεί στο 99,76% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από την ενέργεια για το 2012) είναι η κύρια πηγή των εκπομπών N₂O από τον τομέα της ενέργειας (Διάγραμμα 3.15). Οι εκπομπές υποξειδίου του αζώτου από την καύση των καυσίμων κυρίως πηγάζουν από τις ενεργειακές βιομηχανίες και τις μεταφορές (Διάγραμμα 3.16).



Διάγραμμα 3.15: Κατανομή συνολικών εκπομπών N₂O της ΕΕ-28 από την ενέργεια, το 2012

- **Ενεργειακές βιομηχανίες (Energy Industries)** (αντιπροσωπεύουν 30,53% του συνόλου των εκπομπών της ΕΕ-28 του N₂O από την ενέργεια το 2012 (Πίνακας 3.6))

1.A.1. Ενεργειακές βιομηχανίες	9.875,60
1.Ενέργεια	32.345,53

Πίνακας 3.6. Συνολικές εκπομπές N₂O από τον τομέα της Ενέργειας και από τις Ενεργειακές βιομηχανίες για το 2012.

Οι εκπομπές N₂O προκύπτουν κατά την παραγωγή του δημόσιου ηλεκτρισμού και θερμότητας στις ενεργειακές βιομηχανίες. Η καύση των ορυκτών καυσίμων για την παραγωγή τόσο της θερμότητας όσο και της ηλεκτρικής ενέργειας οδηγεί σε σημαντικές ποσότητες των εκπομπών N₂O.

- **Μεταφορές (Transport)** (αντιπροσωπεύουν το 29,21% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από την ενέργεια το 2012 (Πίνακας 3.7))

1.A.3. Μεταφορές	9.448,21
1.Ενέργεια	32.345,53

Πίνακας 3.7. Συνολικές εκπομπές N₂O από τον τομέα της Ενέργειας και από τις Μεταφορές για το 2012.

Οι εκπομπές N₂O από τις μεταφορές είναι πιο δύσκολο να εκτιμηθούν με ακρίβεια, επειδή εξαρτώνται από τα χαρακτηριστικά της τεχνολογίας των οχημάτων και της λειτουργίας και των καυσίμων. Για παράδειγμα, οι εξελίξεις στους καταλύτες έχουν οδηγήσει σε αλλαγές στις εκπομπές N₂O από την παλιά μέχρι την νέα γενιά των καταλυτών.

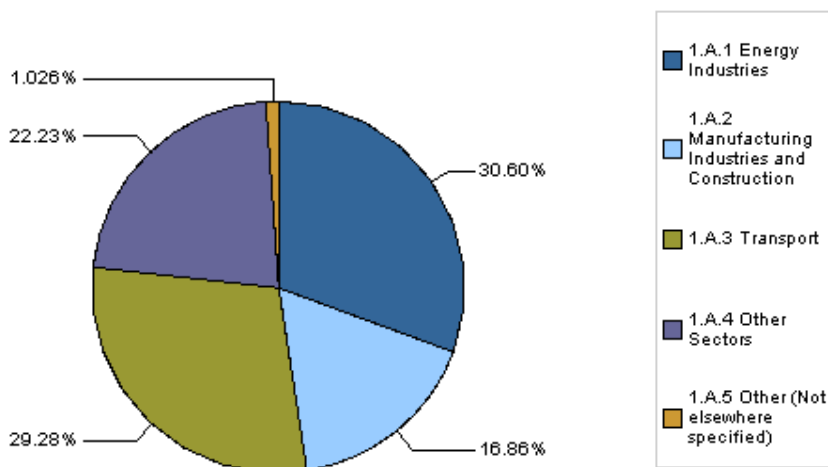
- **Μεταποιητικές βιομηχανίες και κατασκευές (Manufacturing industries and constructions)** (αντιπροσωπεύουν το 16,82% του συνόλου των εκπομπών του N₂O της ΕΕ-28 από την ενέργεια το 2012 (Πίνακας 3.8))

1.A.2. Μεταποιητικές βιομηχανίες και κατασκευές	5.441,09
1.Ενέργεια	32.345,53

Πίνακας 3.8. Συνολικές εκπομπές N₂O από τον τομέα της Ενέργειας και από τις Μεταποιητικές βιομηχανίες και κατασκευές για το 2012.

Στις μεταποιητικές βιομηχανίες, οι πρώτες ύλες μετατρέπονται σε προϊόντα και τα καύσιμα χρησιμοποιούνται ως πηγή ενέργειας στις κατασκευαστικές βιομηχανίες για την μετατροπή αυτή των πρώτων υλών σε προϊόντα. Αυτό περιλαμβάνει την καύση του καυσίμου για την παραγωγή ηλεκτρικής ενέργειας και θερμότητας στις βιομηχανίες για δική τους χρήση.

- ✓ Για όλες τις παραπάνω κατηγορίες, οι εκπομπές N₂O μπορούν να υπολογιστούν χρησιμοποιώντας την μεταβλητή ακαθάριστη εγχώρια κατανάλωση (gross inland consumption) ορυκτών καυσίμων, δηλαδή λιθάνθρακα (hard coal), λιγνίτη (lignite, φυσικού αερίου (natural gas), αργού πετρελαίου (crude oil) και προϊόντων πετρελαίου (petroleum products).



Διάγραμμα 3.16: Κατανομή συνολικών εκπομπών N₂O της ΕΕ-28 από την καύση των καυσίμων, το 2012

3.2. Συλλογή δεδομένων

Αφού γνωρίζουμε τις μεταβλητές που θα εξετάσουμε στην περαιτέρω ανάλυσή μας, συλλέγουμε τις τιμές των μεταβλητών αυτών για το χρονικό διάστημα 1995-2012. Η συλλογή αυτή έγινε μέσα από τις βάσεις δεδομένων της Eurostat, της FAOSTAT, της UNFCCC, της KNOEMA και της World Bank. Αναλυτικότερα στοιχεία υπάρχουν στον Πίνακα 3.9.

Επεξηγηματική Μεταβλητή (Explanatory Variable)	Μονάδα Μέτρησης της Μεταβλητής	Πηγή
Αριθμός Βοοειδών (Cattle Population)	Ακριβής Αριθμός Ζώων	FAOSTAT
Αριθμός Χοίρων (Pig Population)	Ακριβής Αριθμός Ζώων	FAOSTAT
Αριθμός Προβάτων (Sheep Population)	Ακριβής Αριθμός Ζώων	FAOSTAT
Αστικά απόβλητα σε χώρους υγειονομικής ταφής (Municipal Waste Landfill)	Χιλιάδες Τόνοι	Eurostat
Συνολικός Πληθυσμός (Total Population)	Ακριβής Αριθμός Ατόμων	Eurostat
Αστικός Πληθυσμός (Urban Population)	Ακριβής Αριθμός Ατόμων	World Bank
Ακαθάριστο Εγχώριο Προϊόν (GDP)	Εκατομμύρια Ευρώ	Eurostat
Πρωτογενής Παραγωγή Λιθάνθρακα (Primary Production of Hard Coal)	Χιλιάδες Τόνοι	Eurostat
Πρωτογενής Παραγωγή Λιγνίτη (Primary Production of Lignite)	Χιλιάδες Τόνοι	Eurostat

Ακαθάριστη Εγχώρια Κατανάλωση Φυσικού Αερίου (Gross Inland Consumption of Natural Gas)	Χιλιάδες Τόνοι (ισοδυναμίας)	Eurostat
Πρωτογενής Παραγωγή Φυσικού Αερίου (Primary Production of Natural Gas)	Χιλιάδες Τόνοι (ισοδυναμίας)	Eurostat
Απώλειες Διανομής (Distribution Losses)	Χιλιάδες Τόνοι (ισοδυναμίας)	Eurostat
Αριθμός Κατσικών (Goat Population)	Ακριβής Αριθμός Ζώων	FAOSTAT
Φυτική Παραγωγή (Crop Production)	Ποσότητα Προϊόντων	FAOSTAT
ΑΠΑ για την παραγωγή χημικών ουσιών και προϊόντων (GVA for Manufacture of Chemicals and Chemical Products)	Εκατομμύρια Ευρώ	KNOEMA
Ακαθάριστη Εγχώρια Κατανάλωση Λιθάνθρακα (Gross Inland Consumption of Hard Coal)	Χιλιάδες Τόνοι	Eurostat
Ακαθάριστη Εγχώρια Κατανάλωση Λιγνίτη (Gross Inland Consumption of Lignite)	Χιλιάδες Τόνοι	Eurostat
Ακαθάριστη Εγχώρια Κατανάλωση Προϊόντων Πετρελαίου (Gross Inland Consumption of Petroleum Products)	Χιλιάδες Τόνοι	Eurostat
Ακαθάριστη Εγχώρια Κατανάλωση Αργού Πετρελαίου (Gross Inland Consumption of Crude Oil)	Χιλιάδες Τόνοι	Eurostat

Πίνακας 3.9: Πηγές συλλογής δεδομένων των επεξηγηματικών μεταβλητών

3.3. Επιλογή καταλληλότερων επεξηγηματικών μεταβλητών για το μοντέλο

Ένα από τα πλέον σημαντικότερα προβλήματα της γραμμικής παλινδρόμησης είναι η επιλογή ενός υποσυνόλου από επεξηγηματικές μεταβλητές, έτσι ώστε να υπάρχει εξοικονόμηση κόστους κατά την πρόβλεψη της μεταβλητής απόκρισης (μικρός αριθμός επεξηγηματικών μεταβλητών) χωρίς να προκύπτει μεγάλη απώλεια στην αποτελεσματικότητα του μοντέλου. Η επιλογή του μοντέλου μπορεί να χρησιμοποιηθεί και για μελλοντικές προβλέψεις εφόσον τα δεδομένα είναι αντιπροσωπευτικά των συνθηκών υπό τις οποίες θα γίνει η πρόβλεψη. Η εφαρμογή μιας διαδικασίας επιλογής του καταλληλότερου υποσυνόλου επεξηγηματικών μεταβλητών απαιτεί την χρήση στατιστικών πακέτων, καθώς ο όγκος των πράξεων είναι τεράστιος για να μπορέσει να γίνει με το χέρι. Στις μέρες μας υπάρχει πληθώρα στατιστικών πακέτων, όπως το MINITAB, το SPSS και η R, την οποία επέλεξα να χρησιμοποιήσω για τη συγκεκριμένη ανάλυση των δεδομένων.

Στην παρούσα εργασία, για την επιλογή του κατάλληλου μοντέλου, εφαρμόστηκαν αρχικά τα κριτήρια πληροφορίας (AIC, BIC). Τα κριτήρια όμως αυτά δεν λαμβάνουν υπόψη τις τυχόν συσχετίσεις που μπορεί να υπάρχουν μεταξύ των επεξηγηματικών μεταβλητών (φαινόμενο πολυσυγγραμμικότητας). Έτσι, στις περιπτώσεις που τα μοντέλα που είχαν επιλεγεί με βάση τα κριτήρια AIC και BIC αντιμετώπιζαν πρόβλημα, εφαρμόστηκαν δύο μέθοδοι ποινικοποιημένης πιθανοφάνειας, η Lasso και η Ridge Regression.

3.3.1. Πολυσυγγραμμικότητα (VIF)

Όταν υπάρχει έντονη συσχέτιση μεταξύ δύο ή περισσότερων επεξηγηματικών μεταβλητών, έχουμε το φαινόμενο της πολυσυγγραμμικότητας (multicollinearity). Η παρουσία αυτού του φαινομένου οδηγεί σε αυξημένα τυπικά σφάλματα των εκτιμώμενων συντελεστών $\hat{\beta}$ και κατά συνέπεια δυσκολεύει την εκτίμηση της επίδρασης της κάθε επεξηγηματικής μεταβλητής στην μεταβλητή απόκρισης, αφού τα διαστήματα εμπιστοσύνης των αντίστοιχων συντελεστών θα είναι μεγάλα σε εύρος. Επίσης, σε τέτοιες περιπτώσεις είναι δύσκολος ο εντοπισμός των στατιστικά σημαντικών μεταβλητών.

Για να εντοπίσουμε την πολυσυγγραμμικότητα χρησιμοποιούμε τον παράγοντα μεγέθυνσης διασποράς (variance inflation factor (VIF)), ο οποίος έχει την τιμή:

$$VIF = 1 / (1 - R_j^2).$$

Ο R_j^2 είναι ο συντελεστής προσδιορισμού μιας γραμμικής παλινδρόμησης με μεταβλητή απόκριση την X_j και επεξηγηματικές μεταβλητές όλες τις άλλες μεταβλητές X_i , $i \neq j$. Ο δείκτης R_j^2 δηλαδή εκφράζει κατά πόσο η X_j μπορεί να προβλεφθεί από τις υπόλοιπες επεξηγηματικές μεταβλητές.

Αναλυτικότερα, γνωρίζουμε ότι η εκτιμημένη διασπορά της παραμέτρου $\hat{\beta}_j$ δίνεται από τη σχέση:

$$\widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{(n-1)S_{X_j}^2} \times \frac{1}{1-R_j^2}$$

Παρατηρούμε ότι όσο αυξάνεται ο όρος $1 / (1 - R_j^2)$, και κατ' επέκταση το VIF, αυξάνεται και η διασπορά. Επιπρόσθετα, η τετραγωνική ρίζα του VIF μας λέει κατά πόσο αυξάνεται το τυπικό σφάλμα, σε σύγκριση με αυτό που θα ήταν αν η επεξηγηματική μεταβλητή ήταν ασυσχέτιστη με τις υπόλοιπες. Τιμές $VIF > 10$ θεωρούνται ως ένδειξη πολυσυγγραμμικότητας. Παρακάτω παρουσιάζονται δύο μέθοδοι που μπορούν να αντιμετωπίσουν αυτό το πρόβλημα (Κεφάλαιο 3.3.5).

3.3.2. Κριτήριο AIC

Το κριτήριο AIC (Akaike's information criterion) αναπτύχθηκε από τον Akaike (1971) και αποτελεί ένα μέτρο με το οποίο μπορούμε να ελέγξουμε την καταλληλότητα των επεξηγηματικών μεταβλητών στο γραμμικό μοντέλο. Με βάση αυτό το κριτήριο, βέλτιστο είναι το μοντέλο που ελαχιστοποιεί το AIC.

Η τιμή του AIC υπολογίζεται από τον τύπο :

$$AIC = 2d - 2 \ln L$$

Όπου: d ο αριθμός των παραμέτρων του γραμμικού μοντέλου ($d=p+2$)

L η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας για το εκτιμημένο μοντέλο.

Έστω $\{y_i, x_{i1}, x_{i2}, \dots, x_{ip}\}$, $i=1,2,3, \dots, n$, n σύνολα δεδομένων τα οποία έχουν παρατηρηθεί για την μεταβλητή απόκρισης Y και τις p επεξηγηματικές μεταβλητές $\{X_1, X_2, \dots, X_p\}$.

Η συνάρτηση πιθανοφάνειας για το πολλαπλό γραμμικό μοντέλο έχει τη μορφή:

$$L = f(y | \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right)$$

και η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας γίνεται:

$$\ell = \ln(L) = -\frac{n \ln(2\pi)}{2} - \frac{n \ln(\sigma^2)}{2} - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

όπου $y = (y_1, y_2, \dots, y_n)'$ είναι το $n \times 1$ διάνυσμα τιμών της μεταβλητής απόκρισης Y .

X ένας πίνακας (πίνακας σχεδιασμού) με $p+1$ στήλες και n γραμμές

$\beta = (\alpha, \beta_1, \dots, \beta_p)'$ είναι το $(p+1) \times 1$ διάνυσμα των παραμέτρων.

$\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$ το $n \times 1$ διάνυσμα των τυχαίων σφαλμάτων.

Ο πίνακας σχεδιασμού X έχει την ακόλουθη μορφή:

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & & x_{2p} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

Από τις πιο πάνω σχέσεις μπορούμε να συμπεράνουμε ότι πάντοτε η εισαγωγή περισσότερων παραμέτρων στο μοντέλο συνεπάγεται με αύξηση της λογαριθμοποιημένης συνάρτησης πιθανοφάνειας ℓ και έτσι βελτιώνει την προσαρμογή του μοντέλου, ανεξάρτητα αν είναι στατιστικά σημαντικές ή όχι. Ωστόσο, το AIC δεν ενδιαφέρεται μόνο για την καλή προσαρμογή, αλλά περιλαμβάνει μια αύξουσα συνάρτηση του αριθμού των εκτιμώμενων παραμέτρων d , έτσι ώστε να λαμβάνεται υπόψη και ο αριθμός παραμέτρων που προστίθενται στο μοντέλο. Επομένως, εμφανίζει το βέλτιστο μοντέλο με όσο το δυνατόν μικρότερο αριθμό παραμέτρων.

3.3.3. Κριτήριο BIC

Το κριτήριο BIC (Bayesian information criterion) αναπτύχθηκε από τον Schwarz (1978) και έχει την ονομασία Bayesian επειδή είναι κριτήριο πληροφοριών με βάση τη Bayesian Μέθοδο. Αποτελεί ένα ακόμη μέτρο με το οποίο μπορούμε να ελέγξουμε την καταλληλότητα των

επεξηγηματικών μεταβλητών στο γραμμικό μοντέλο. Με βάση αυτό το κριτήριο, το μοντέλο που ελαχιστοποιεί το BIC είναι και το βέλτιστο.

Η τιμή του BIC υπολογίζεται από τον τύπο :

$$BIC = d \ln n - 2 \ln L$$

Όπου: **d** ο αριθμός των παραμέτρων του γραμμικού μοντέλου.

L η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας για το εκτιμημένο μοντέλο, όπως παρουσιάζεται στο Κεφάλαιο 3.3.2.

n το πλήθος του δείγματος

Ο σκοπός της παραμέτρου d στον πιο κάτω τύπο είναι ίδιος όπως στο AIC κριτήριο, δηλαδή να εμφανίσει το βέλτιστο μοντέλο με όσο το δυνατόν μικρότερο αριθμό παραμέτρων. Η διαφορά μεταξύ των δύο κριτηρίων βρίσκεται στον όρο ποινής. Το BIC εφαρμόζει ένα πιο αυστηρό όρο ποινής ($d \ln n$) από το AIC ($2d$) και ως εκ τούτου αποθαρρύνει την εισαγωγή επιπρόσθετων παραμέτρων σε μεγαλύτερο βαθμό και τείνει να εξάγει μικρότερα μοντέλα. Για $n \geq 8$ συνεπάγεται ότι $d \ln n > 2d$. Όταν τον n είναι μεγάλο, οι δύο μέθοδοι εξάγουν διαφορετικά αποτελέσματα. Από πρακτική άποψη, το AIC θα μπορούσε να χρησιμοποιηθεί, όταν ο κύριος στόχος της εφαρμογής του μοντέλου είναι η πρόβλεψη, δηλαδή για να διαμορφώσουμε ένα μοντέλο που θα εξάγει αποτελεσματικά τις νέες προβλέψεις. Το BIC θα μπορούσε να χρησιμοποιηθεί, όταν ο κύριος στόχος της εφαρμογής του μοντέλου είναι περιγραφικός, δηλαδή για να διαμορφώσουμε ένα μοντέλο που θα περιλαμβάνει τους σημαντικότερους παράγοντες που επηρεάζουν το αποτέλεσμα.

3.3.4. Εφαρμογή των κριτηρίων AIC και BIC στην R

Όπως προαναφέρθηκε η επιλογή του καλύτερου υποσυνόλου επεξηγηματικών μεταβλητών, που θα συμπεριληφθούν στο μοντέλο, γίνεται με βάση τα κριτήρια πληροφωρίας AIC και BIC. Για αυτή την επιλογή, χρειάζεται να γίνει πλήρης απαρίθμηση (full-enumeration) όλων των συνδυασμών των επεξηγηματικών μεταβλητών. Συγκεκριμένα, χρειάζεται να εξετάσουμε όλους τους 2^p συνδυασμούς (όπου p το πλήθος των επεξηγηματικών μεταβλητών του μοντέλου), ώστε να βρεθεί ο καταλληλότερος, δηλαδή εκείνος που ελαχιστοποιεί τις τιμές των AIC και BIC κριτηρίων. Για την εφαρμογή των δύο κριτηρίων πληροφωρίας, χρησιμοποιήθηκε ο πιο κάτω κώδικας:

```
integer.base.b<-function(x, b=2){
  xi <- as.integer(x)
  if(any(is.na(xi) | ((x-xi)!=0)))
    print(list(ERROR="x not integer", x=x))
  N <- length(x)
  xMax <- max(x)
  ndigits <- (floor(logb(xMax, base=2))+1)
  Base.b <- array(NA, dim=c(N, ndigits))
  for(i in 1:ndigits) {#i <- 1
    Base.b[, ndigits-i+1] <- (x %% b)
    x <- (x %/% b)
  }
  if(N ==1) Base.b[1, ] else Base.b
}
```

```

full_enumeration.linear_BIC_AIC<-function(x,y)
{
  p<-ncol(dedomena) # number of explanatatory variables
  n<-nrow(dedomena) # sample size
  colnames(dedomena)<-paste("X",1:p,sep="")
  namesg<-colnames(dedomena)
  X<-cbind(1, dedomena) # design matrix

# # of models in full enumeration
  nmodels<-2^p
  result<-matrix(0, nrow=nmodels, ncol=p+2)
  gamma<-integer.base.b(0:(nmodels-1))
  gamma<-cbind(1,gamma)
  bic<-BIC(lm(y~1))
  aic<-AIC(lm(y~1))
  result[1,]<-c(gamma[1,-1],bic)
  for (i in 2:nmodels) {
    Xg<-as.matrix(X[,gamma[i]==1], nrow=n)
    xg<-Xg[,-1]
    bic<-BIC(lm(y~xg))
    aic<-AIC(lm(y~xg))
    result[i,]<-c(gamma[i,-1],bic,aic)
  }

# sort results according to BIC

  index <- order(result[,p+1], decreasing=FALSE)
  result<-result[index,]
  colnames(result)<-c(namesg,"BIC","AIC")
  print(result)

# sort results according to AIC

  index <- order(result[,p+2], decreasing=FALSE)
  result<-result[index,]
  colnames(result)<-c(namesg,"BIC","AIC")
  print(result)
}

```

Η συνάρτηση `integer.base.b` δημιουργεί διανύσματα $\vec{\gamma}=(\gamma_1,\gamma_2,\dots,\gamma_p)$, όπου τα γ_i , $i=1,\dots,p$ παίρνουν την τιμή 1 ή 0, p είναι ο αριθμός των επεξηγηματικών μεταβλητών. Το διάνυσμα $\vec{\gamma}=(0,0,\dots,0)$ αντιστοιχεί στο μοντέλο που περιλαμβάνει μόνο τη σταθερά, χωρίς επεξηγηματικές μεταβλητές. Το διάνυσμα $\vec{\gamma}=(1,1,\dots,1)$ αντιστοιχεί στο μοντέλο που περιλαμβάνει όλες τις επεξηγηματικές μεταβλητές. Γενικότερα, κάθε διάνυσμα $\vec{\gamma}$ αντιστοιχεί στο μοντέλο που περιλαμβάνει τις επεξηγηματικές μεταβλητές X_i για κάθε i για το οποίο ισχύει $\gamma_i=1$ στο συγκεκριμένο διάνυσμα $\vec{\gamma}$. Στην περίπτωση μας, η παράμετρος \mathbf{x} αντιστοιχεί σε ένα διάνυσμα μεγέθους όσο το πλήθος των πιθανών συνδυασμών των επεξηγηματικών μεταβλητών στο μοντέλο με τιμές από 0 μέχρι 2^{p-1} , δηλαδή $\mathbf{x}=(0,1,2,\dots,2^{p-1})$. Η παράμετρος $b=2$ αντιστοιχεί στο δυαδικό σύστημα, δηλαδή στη χρήση μόνο των τιμών 0 και 1 για τη δημιουργία των διανυσμάτων. Ο πρώτος έλεγχος εξετάζει αν η

παράμετρος \mathbf{x} που εισάχθηκε στο μοντέλο έχει ακέραιες τιμές και σε αντίθετη περίπτωση τυπώνει σφάλμα. Η παράμετρος N είναι το μέγεθος του διανύσματος \mathbf{x} , δηλαδή 2^p , η $xMax$ παίρνει την μέγιστη τιμή του διανύσματος \mathbf{x} ($xMax=2^{p-1}$) και το $ndigits$ αντιστοιχεί στο πλήθος των επεξηγηματικών μεταβλητών:

$$ndigits = \text{floor}(\log_2 2^{p-1} + 1) = p$$

Ο Base.b είναι ένας πίνακας με $N=2^p$ γραμμές και $ndigits=p$ στήλες. Στην περίπτωση τεσσάρων επεξηγηματικών μεταβλητών ο πίνακας αυτός θα έχει τη μορφή:

```

0 0 0 0
0 0 0 1
0 0 1 0
0 0 1 1
0 1 0 0
0 1 0 1
0 1 1 0
0 1 1 1
1 0 0 0
1 0 0 1
1 0 1 0
1 0 1 1
1 1 0 0
1 1 0 1
1 1 1 0
1 1 1 1

```

όπου οι στήλες αντιπροσωπεύουν τις επεξηγηματικές μεταβλητές και οι γραμμές τα διανύσματα $\vec{\gamma}$, δηλαδή όλους τους συνδυασμούς των επεξηγηματικών μεταβλητών.

Με τον κώδικα επανάληψης:

```

for(i in 1:ndigits) {#i <- 1
  Base.b[, ndigits-i+1] <- (x %% b)
  x <- (x %/% b)}

```

δημιουργούνται οι στήλες αυτές του πίνακα, ξεκινώντας από τα δεξιά προς τα αριστερά και με τη βοήθεια των πράξεων `mod (%%)` και `div(%/%)`.

Αν το N είναι 1, δηλαδή δεν έχουμε επεξηγηματικές μεταβλητές, αλλά μόνο τη σταθερά τότε ο πίνακας αυτός έχει μόνο μία γραμμή:

```

if(N == 1) Base.b[1, ] else Base.b .

```

Έχοντας λοιπόν αυτό τον πίνακα, όπου οι γραμμές του αποτελούν τα διανύσματα $\vec{\gamma}$ και κατ'επέκταση όλους τους συνδυασμούς των επεξηγηματικών μεταβλητών στο μοντέλο, τρέχουμε την συνάρτηση `full_enumeration.linear_BIC_AIC` που μας εμφανίζει τα BIC και AIC για όλους τους συνδυασμούς των επεξηγηματικών μεταβλητών.

Ο πίνακας `result` είναι ένας πίνακας $2^p \times (p+2)$. Οι p πρώτες στήλες αντιστοιχούν στις επεξηγηματικές μεταβλητές και οι δύο τελευταίες στις τιμές των κριτηρίων BIC και AIC. Ο `gamma` είναι ο πίνακας με γραμμές τα διανύσματα $\vec{\gamma}$. Επιπλέον όλες οι τιμές της 1^{ης} στήλης

του πίνακα, που αντιστοιχεί στην σταθερά, ισούνται με 1, καθώς κάθε μοντέλο, με οποιονδήποτε συνδυασμό επεξηγηματικών μεταβλητών, περιλαμβάνει την σταθερά.

Στη συνέχεια σε κάθε γραμμή i του πίνακα results αποθηκεύονται στις p πρώτες θέσεις οι τιμές του αντίστοιχου διανύσματος \vec{y} και στις δύο τελευταίες τα BIC και AIC για το μοντέλο με επεξηγηματικές μεταβλητές τις X_i , όπου $i : \gamma_i=1$.

Τέλος, με τα δύο τελευταία μέρη της συνάρτησης full_enumeration.linear_BIC_AIC μπορούμε να ταξινομήσουμε τα αποτελέσματά μας με βάση τα κριτήρια BIC και AIC αντίστοιχα.

3.3.5. Μέθοδοι L_1 και L_2 -Penalized

Στο πολλαπλό γραμμικό μοντέλο είναι πιθανόν να υπάρχει πολυσυγγραμμικότητα, δηλαδή δύο ή περισσότερες επεξηγηματικές μεταβλητές, να είναι έντονα συσχετισμένες μεταξύ τους (Κεφάλαιο 3.3.1.). Το γεγονός αυτό, πολλές φορές προκαλεί αστάθεια στους εκτιμητές και μπορεί να δημιουργήσει πρόβλημα στη σωστή εκτίμηση του μοντέλου. Η παρουσία του φαινομένου αυτού δυσκολεύει την επιλογή των στατιστικά σημαντικών μεταβλητών, καθώς σημαντικές επεξηγηματικές μεταβλητές θα εμφανίζονται ως μη σημαντικές. Τα τυπικά σφάλματα των συντελεστών είναι αρκετά υψηλά και οι συντελεστές πιθανόν να έχουν αντίθετα πρόσημα από τα αναμενόμενα. Επίσης, οι επιδράσεις μεταξύ των επεξηγηματικών μεταβλητών που συσχετίζονται μεταξύ τους είναι δύσκολο να διαχωριστούν και ως εκ τούτου, δεν θα είμαστε σε θέση να τις εκτιμήσουμε.

Σε τέτοιες περιπτώσεις, όπου δύο ή περισσότερες μεταβλητές συσχετίζονται σε μεγάλο βαθμό, μπορούμε να επιλέξουμε μία από αυτές για να κρατήσουμε στο μοντέλο αφού θα μου δώσουν ουσιαστικά την ίδια πληροφορία. Παρόλα αυτά είναι προτιμότερο και πιο αξιόπιστο να γίνεται επιλογή του καταλληλότερου μοντέλου βάσει κάποιων μεθόδων. Δύο από τις σημαντικότερες μεθόδους ποινικοποιημένης πιθανοφάνειας (penalized methods), που αντιμετωπίζουν, δηλαδή, το πρόβλημα της πολυσυγγραμμικότητας είναι η Lasso (L_1 -penalized) και η παλινδρόμηση κορυφογραμμής (Ridge Regression – L_2 -penalized). Η Lasso χρησιμοποιώντας μια μη κυρτή ποινή, κάνει αυτόματα επιλογή των σημαντικών επεξηγηματικών μεταβλητών και μηδενίζει τους συντελεστές των υπόλοιπων (μη σημαντικών) μεταβλητών, σε αντίθεση με την παλινδρόμηση κορυφογραμμής που μόνο μειώνει τους συντελεστές των μη σημαντικών μεταβλητών κοντά στο μηδέν.

3.3.5.1. Μέθοδος L_2 - Penalized (Ridge regression)

Η πρώτη τεχνική συρρίκνωσης που εφαρμόστηκε είναι η Παλινδρόμηση Κορυφογραμμής (Ridge regression), η οποία εισήχθη από τους Hoerl και Kennard (1970). Καλείται έτσι, καθώς τα μαθηματικά που χρησιμοποιούνται σχετίζονται με τη μέθοδο της Ridge-ανάλυσης. Η μέθοδος αυτή όπως προαναφέρθηκε έχει στόχο να μειώσει τους συντελεστές των μη σημαντικών μεταβλητών κοντά στο μηδέν. Με αυτόν τον τρόπο μπορεί να κάνει πιο απλή την επιλογή των επεξηγηματικών μεταβλητών.

Όταν υπάρχει πρόβλημα πολυσυγγραμμικότητας, η μέθοδος των ελαχίστων τετραγώνων αποτυγχάνει. Αυτό συμβαίνει λόγω του ότι σε τέτοια περίπτωση κάποιες στήλες στον πίνακα σχεδιασμού X είναι γραμμικά εξαρτημένες. Αποτέλεσμα αυτού είναι ο πίνακας $X'X$ να είναι ιδιάζων ($\det(X'X)=0$) και τότε ο πίνακας $(X'X)^{-1}$ να μην υπάρχει. Εφόσον η εκτιμήτρια των ελαχίστων τετραγώνων υπολογίζεται μέσα από τη σχέση:

$$\hat{\beta} = (X'X)^{-1} X'y$$

προκύπτουν μη σταθερές εκτιμήσεις των παραμέτρων του μοντέλου και η εκτιμήτρια $\hat{\beta}$ δεν είναι αμερόληπτη εκτιμήτρια της β εκτιμήτριας των συντελεστών παλινδρόμησης, δηλαδή δεν ισχύει η σχέση :

$$E(\hat{\beta}) = \beta$$

Η εκτιμήτρια $\hat{\beta}_{ridge}$ είναι μεροληπτική εκτιμήτρια της β , και αποδεικνύεται εύκολα ότι είναι ένας γραμμικός συνδυασμός της εκτιμήτριας των ελαχίστων τετραγώνων, αφού:

$$\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1} X'y = (X'X + \lambda I)^{-1} (X'X)\beta = z_k \beta$$

όπου: λ ρυθμιστική παράμετρος (παράμετρος μεροληψίας).

Η εκτιμήτρια $\hat{\beta}_{ridge}$ υπολογίζεται από την πιο πάνω σχέση, ώστε να ελαχιστοποιεί το ποινικοποιημένο άθροισμα τετραγώνων των υπολοίπων (Penalized Residual Sum of Squares (PRSS)) που δίνεται από τη σχέση:

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

και ισοδυναμεί με την ελαχιστοποίηση του:

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2$$

με τον περιορισμό:

$$\sum_{j=1}^p \beta_j^2 < c \quad (\text{για κάποιο } c > 0).$$

Για $\lambda > 0$, υπάρχει λύση, ακόμη και αν η ο πίνακας $X'X$ δεν είναι αντιστρέψιμος, δίνοντάς μας λύσεις και στην περίπτωση όπου $p > n$ (το πλήθος των επεξηγηματικών μεταβλητών είναι μεγαλύτερο από το μέγεθος του δείγματος).

Ένα μεγάλο πλεονέκτημα της ανάλυσης κορυφογραμμής είναι το γράφημα του ίχνους κορυφογραμμής. Το γράφημα έχει μία καμπύλη για κάθε συντελεστή και παρουσιάζει την εκτιμώμενη τιμή κάθε συντελεστή έναντι του λ . Στόχος είναι να βρεθεί μια τιμή του λ , η οποία δίνει ένα σύνολο συντελεστών με μέσο τετραγωνικό σφάλμα μικρότερο από αυτό της ευθείας ελαχίστων τετραγώνων. Η τιμή του λ για την οποία οι συντελεστές έχουν σταθεροποιηθεί δίνει τις εκτιμήσεις των συντελεστών. Κύριο πρόβλημα είναι η επιλογή του λ , καθώς:

- Για κάθε λ , έχουμε μια διαφορετική εκτίμηση για τους συντελεστές.
- Οι τιμές του λ είναι καταχωρισμένες σε μία ενιαία γραμμή στο γράφημα, και δεν μπορείς πάντα να διακρίνεις εύκολα την ακριβή τιμή του.
- Η λ χαράζει μια πορεία λύσεων (μια διαδρομή για κάθε συντελεστή που απεικονίζεται με μία γραμμή για κάθε επεξηγηματική μεταβλητή), οπότε στην περίπτωση πολλών επεξηγηματικών μεταβλητών στο μοντέλο, δύσκολα διακρίνονται αυτές οι διαδρομές.

Για την επιλογή κατάλληλου λ χρησιμοποιούμε την μέθοδο K-fold Cross Validation, όπου επιλέγουμε την τιμή με το ελάχιστο σφάλμα πρόβλεψης (Κεφάλαιο 3.3.5.3.).

Επιπλέον, εφαρμόζοντας τη μέθοδο της παλινδρόμησης κορυφογραμμής η διασπορά $V(\hat{\beta}_{ridge})$ είναι μικρότερη της $V(\hat{\beta})$:

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

$$V(\hat{\beta}_{ridge}) = \sigma^2 (X'X + \lambda I_p)^{-1} X'X (X'X + \lambda I_p)^{-1}.$$

Κατά συνέπεια το μέσω τετραγωνικό σφάλμα (ΜΤΣ) της $\hat{\beta}_{ridge}$ μπορεί να είναι πολύ μικρότερο από το ΜΤΣ της $\hat{\beta}$, καθώς όπως φαίνεται στην πιο κάτω σχέση το ΜΤΣ έχει θετική γραμμική συσχέτιση με τη διασπορά:

$$MTS(\hat{\theta}) = V(\hat{\theta}) + bias^2$$

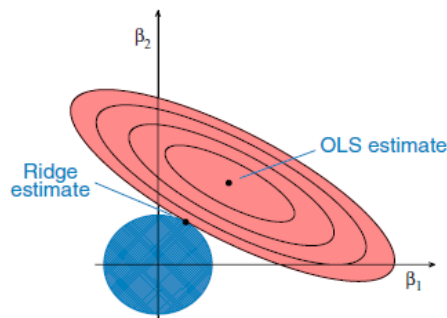
Οι εκτιμητές της παλινδρόμησης κορυφογραμμής δεν είναι αμερόληπτοι καθώς:

$$\begin{aligned}\hat{\beta}_\lambda^{\text{ridge}} &= (\mathbf{I}_p + \lambda \mathbf{R}^{-1}) \hat{\beta}^{\text{ls}} \quad \text{και} \\ \mathbb{E}(\hat{\beta}_\lambda^{\text{ridge}}) &= \mathbb{E}\{(\mathbf{I}_p + \lambda \mathbf{R}^{-1}) \hat{\beta}^{\text{ls}}\} \\ &= (\mathbf{I}_p + \lambda \mathbf{R}^{-1}) \beta \\ &\quad (\text{if } \lambda \neq 0) \\ &\neq \beta\end{aligned}$$

όπου $\mathbf{R} = \mathbf{X}'\mathbf{X}$.

Άρα για κάθε $\lambda > 0$, οι εκτιμητές κορυφογραμμής είναι μεροληπτικοί.

Γεωμετρική ερμηνεία της Παλινδρόμησης Κορυφογραμμής:



Διάγραμμα 3.17. Γεωμετρική Ερμηνεία της μεθόδου παλινδρόμησης κορυφογραμμής.

Οι ελλείψεις αντιστοιχούν στα αθροίσματα των τετραγώνων των υπολοίπων (RSS), η εσωτερική έλλειψη αντιστοιχεί σε μικρότερα RSS και το RSS ελαχιστοποιείται στις εκτιμήσεις των ελαχίστων τετραγώνων (OLS estimates) (Διάγραμμα 3.17). Για $p=2$ (δύο επεξηγηματικές μεταβλητές στο μοντέλο), ο περιορισμός της παλινδρόμησης κορυφογραμμής αντιστοιχεί σε έναν κύκλο ($\sum_{j=1}^p \beta_j^2 < c$). Η μέθοδος αυτή, προσπαθεί να ελαχιστοποιήσει το μέγεθος της έλλειψης και του κύκλου ταυτόχρονα. Η εκτίμηση της κορυφογραμμής δίνεται από το σημείο στο οποίο αγγίζουν η έλλειψη και ο κύκλος.

Υπάρχει ένας «συμβιβασμός» μεταξύ του όρου ποινής (λ) και του RSS. Ίσως ένα μεγάλο β μας δώσει ένα καλύτερο άθροισμα τετραγώνων υπολοίπων, αλλά τότε θα αυξήσει τον όρο ποινής. Αυτός είναι ο λόγος που μπορεί να προτιμήσουμε μικρότερα β με χειρότερο άθροισμα τετραγώνων των υπολοίπων. Από την άποψη της βελτιστοποίησης, ο όρος ποινής ισοδυναμεί με τον περιορισμό του β (η νόρμα του να είναι μικρότερη από μία σταθερά c). Υπάρχει μια αντιστοιχία μεταξύ λ και c . Όσο μεγαλύτερη είναι η τιμή της παραμέτρου λ , τόσο το β_j θα τείνει κοντά στο μηδέν. Για $\lambda = 0$, παίρνουμε τις λύσεις της μεθόδου των ελαχίστων τετραγώνων και για $\lambda \uparrow \infty$, έχουμε $\hat{\beta}_{\text{ridge}} = 0$ (δηλαδή, στο μοντέλο περιλαμβάνεται μόνο η σταθερά).

Ένα βασικό μειονέκτημα της μεθόδου, ειδικά στην περίπτωση που το μοντέλο αποτελείται από μεγάλο αριθμό παραμέτρων, είναι το γεγονός ότι το μοντέλο δεν γίνεται εύκολα ερμηνεύσιμο.

3.3.5.2. Μέθοδος L_1 - Penalized (LASSO)

Αξίζει να αναφέρουμε ότι, η παλινδρόμηση κορυφογραμμής δεν χρησιμοποιείται άμεσα στην πράξη, όμως έχει δημιουργήσει ένα εντελώς νέο τομέα της έρευνας για μεθόδους ποινικοποιημένης πιθανοφάνειας. Η πιο δημοφιλής προσέγγιση είναι η Lasso (L_1 -Penalized). Η μέθοδος Lasso (Least Absolute Shrinkage and Selection Operator) προτάθηκε από τον Tibshirani (1996) και θεωρείται από τις καλύτερες μεθόδους επιλογής κατάλληλου μοντέλου. Σκοπός της είναι να συρρικνώσει κάποιους συντελεστές και να θέσει τους υπόλοιπους σε μηδέν και ως εκ τούτου να προσπαθεί να κρατήσει τα καλά χαρακτηριστικά της επιλογής υποσυνόλου των επεξηγηματικών μεταβλητών. Θεωρείται αποτελεσματική, διότι έχει την ιδιότητα να εκτελεί ταυτόχρονα επιλογή μεταβλητών και συρρίκνωση του μοντέλου. Αξίζει να σημειωθεί ότι όταν οι επεξηγηματικές μεταβλητές είναι ισχυρά συσχετισμένες, η Lasso επιλέγει μόνο μία μεταβλητή από το σύνολο.

Σε αντίθεση με την παλινδρόμηση κορυφογραμμής, η Lasso, αντί να χρησιμοποιεί την Ευκλείδεια (L_2) νόρμα στον περιορισμό της, χρησιμοποιεί την L_1 -νόρμα:

$$\text{Ridge: } \sum_{j=1}^p (\beta_j)^2 < c \qquad \text{Lasso: } \sum_{j=1}^p |\beta_j| < c.$$

Είναι μια μικρή αλλά σημαντική αλλαγή, διότι με αυτόν τον τρόπο μερικοί συντελεστές μπορούν να συρρικνωθούν ακριβώς στο μηδέν.

Οι εκτιμήσεις των συντελεστών της Lasso προκύπτουν ελαχιστοποιώντας τη σχέση:

$$\min_{\beta} (Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^p |\beta_j|,$$

όπου παρατηρούμε ότι η L_2 -ποινή της παλινδρόμησης κορυφογραμμής ($\sum_{i=1}^p \beta_i^2$) αντικαταστάθηκε από την L_1 -ποινή ($\sum_{j=1}^p |\beta_j|$).

Η παράμετρος c στον περιορισμό ελέγχει το μέγεθος της συρρίκνωσης, στο οποίο υπόκεινται οι συντελεστές. Έστω ότι

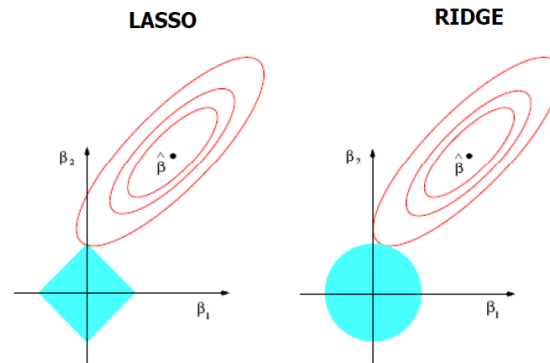
$$c_0 = \sum_{j=1}^p |\hat{\beta}_{LSj}|,$$

δηλαδή, το c_0 ισούται με το άθροισμα των απόλυτων τιμών των εκτιμήσεων των ελαχίστων τετραγώνων. Τιμές μεταξύ $0 < c < c_0$ προκαλούν συρρίκνωση λύσεων προς το μηδέν. Αν, για παράδειγμα, $c = c_0/2$, η μέση συρρίκνωση των συντελεστών των ελαχίστων τετραγώνων είναι 50%. Όταν το λ είναι επαρκώς μεγάλο, μερικοί από τους συντελεστές γίνονται ακριβώς μηδέν, καταλήγοντας σε ένα μοντέλο με μικρό αριθμό επεξηγηματικών μεταβλητών.

Όσον αφορά τη σχέση μεταξύ του λ , της διασποράς και της μεροληψίας έχουμε ότι όσο αυξάνεται το λ , αυξάνεται η μεροληψία και μειώνεται η διασπορά. Όπως και στην παλινδρόμηση κορυφογραμμής για την επιλογή κατάλληλου λ χρησιμοποιούμε την μέθοδο K-fold Cross Validation, όπου επιλέγουμε την τιμή με το ελάχιστο σφάλμα πρόβλεψης (Κεφάλαιο 3.3.5.3).

Σε αντίθεση με την παλινδρόμηση κορυφογραμμής, δεν υπάρχει αναλυτική λύση για τη Lasso, διότι η λύση δεν είναι γραμμική ως προς το Y . Ολόκληρη η διαδρομή των εκτιμήσεων Lasso για όλες τις τιμές του λ μπορεί να υπολογιστεί αποτελεσματικά μέσω της παλινδρόμησης LARS (Least Angle Regression) (Efron 2003).

Γεωμετρική ερμηνεία της Lasso:



Διάγραμμα 3.18. Γεωμετρική Ερμηνεία της μεθόδου Lasso (αριστερά) και της μεθόδου παλινδρόμησης κορυφογραμμής (δεξιά).

Η Lasso εκτελεί L_1 -συρρίκνωση, ώστε να υπάρχουν "γωνιές" στο περιορισμό, ο οποίος σε δύο διαστάσεις, ο περιορισμός είναι της μορφής $|\beta_1| + |\beta_2| \leq c$ και αντιστοιχεί σε ένα διαμάντι (Διάγραμμα 3.18). Εάν το άθροισμα των τετραγώνων (η έλλειψη) "χτυπήσει" μία από αυτές τις γωνίες, τότε ο συντελεστής που αντιστοιχεί στον άξονα αυτό συρρίκνώνεται στο μηδέν. Όσο το πλήθος των επεξηγηματικών μεταβλητών p αυξάνεται, οι γωνίες στο πολυδιάστατο διαμάντι αυξάνονται, και έτσι είναι πολύ πιθανό κάποιοι συντελεστές να μηδενίζονται. Ως εκ τούτου, η Lasso εκτελεί ταυτόχρονα συρρίκνωση και αποτελεσματική επιλογή υποσυνόλου και είναι πολύ χρήσιμη όταν το πλήθος των επεξηγηματικών μεταβλητών p είναι αρκετά μεγάλο.

3.3.5.3. Cross Validation (CVL)

Η επιλογή κατάλληλης ρυθμιστικής παραμέτρου λ είναι καθοριστικής σημασίας για τις μεθόδους που αναπτύξαμε πιο πάνω (Κεφάλαια 3.3.5.1 & 3.3.5.2). Η μέθοδος Cross Validation (διασταυρωμένης επικύρωσης) είναι ένας απλός τρόπος για να εκτιμήσουμε το σφάλμα πρόβλεψης.

Για να επιλέξουμε το λ με βάση την Cross Validation πρέπει να χωρίσουμε τα δεδομένα σε δύο σύνολα. Το ένα θα χρησιμοποιηθεί για την εκτίμηση της ρυθμιστικής παραμέτρου λ (σύνολο εκπαίδευσης - training fold). Το δεύτερο σύνολο θα χρησιμοποιηθεί για τον έλεγχο της ικανότητας πρόβλεψης του μοντέλου (σύνολο ελέγχου - test fold). Ουσιαστικά κάνουμε μια αξιολόγηση του πόσο καλά έχουμε «εκπαιδεύσει» το μοντέλο μας. Όπως είναι φυσικό, με αυτό τον διαχωρισμό προκύπτουν κάποια προβλήματα:

1. Στην περίπτωση που δεν έχουμε πολλά δεδομένα, με το διαχωρισμό, το κάθε σύνολο θα έχει ακόμη πιο λίγα δεδομένα.
2. Έστω ότι έχουμε επαρκή δεδομένα, κάθε διαφορετικός διαχωρισμός δίνει διαφορετικές λύσεις, οπότε υπάρχει πρόβλημα στο πώς θα διαχωρίσουμε αυτά τα δεδομένα.
3. Το μέγεθος του κάθε δείγματος (πόσο μεγάλο πρέπει να είναι;).

Για να αντιμετωπιστούν αυτά τα προβλήματα, η K-fold cross validation χωρίζει τα δεδομένα σε K-μέρη (τα λεγόμενα folds), χρησιμοποιεί τα δεδομένα στα K-1 folds για τη δημιουργία του μοντέλου και ελέγχει την πρόβλεψη του μοντέλου με βάση τα δεδομένα στο K-fold που απέμεινε. Επαναλαμβάνει αυτή τη διαδικασία για όλα τα πιθανά σύνολα ελέγχου και υπολογίζει το μέσο σφάλμα πρόβλεψης.

Για μεγάλο K η εκτίμηση του σφάλματος είναι αρκετά ακριβής αλλά με μεγάλες αποκλίσεις και επίσης αυξάνετε το υπολογιστικό κόστος. Για μικρό K μειώνεται το πλήθος των επαναλήψεων και το υπολογιστικό κόστος, και το εκτιμώμενο σφάλμα θα είναι μεγαλύτερο από το πραγματικό αλλά με μικρότερες αποκλίσεις. Η συνηθέστερη τιμή για το K είναι το 10 (10-fold Cross Validation).

Αναλυτικότερα, η διαδικασία K-fold cross validation έχει ως εξής:

1. Διαχωρίζει το σετ των δεδομένων σε K υποσύνολα αυστηρά ίσου μεγέθους, F_1, \dots, F_k .
2. Σε κάθε επανάληψη ($k = 1, \dots, K$), το k-fold λειτουργεί ως σύνολο ελέγχου, $(x_i, y_i), i \in F_k$, ενώ τα υπόλοιπα K-1 folds αποτελούν το σύνολο εκπαίδευσης $(x_i, y_i), i \notin F_k$.
3. Σε κάθε επανάληψη δημιουργεί ένα μοντέλο με βάση τα δεδομένα των K-1 folds και ελέγχει την πρόβλεψη του μοντέλου με βάση τα δεδομένα στο k-fold. Δηλαδή, για κάθε τιμή της ρυθμιστικής παραμέτρου $\lambda \in \{\lambda_1, \dots, \lambda_k\}$, υπολογίζει την εκτίμηση \hat{f}_λ^{-k} , όπου \hat{f}_λ^{-k} η προβλεπόμενη τιμή του y_i με βάση το σύνολο εκπαίδευσης, και καταγράφει το συνολικό σφάλμα βάσει του συνόλου ελέγχου:

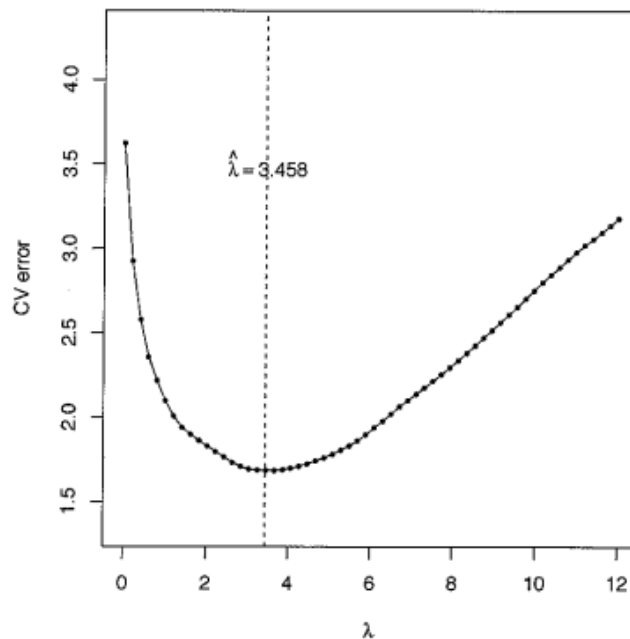
$$e_k(\lambda) = \sum_{i \in F_k} (y_i - \hat{f}_\lambda^{-k}(x_i))^2$$

4. Για κάθε ρυθμιστική παράμετρο λ , υπολογίζει το μέσο σφάλμα επί όλων των folds.

$$CV(\lambda) = \frac{1}{n} \sum_{k=1}^K e_k(\lambda) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in F_k} (y_i - \hat{f}_\lambda^{-k}(x_i))^2$$

Μετά από αυτή τη διαδικασία, έχουμε μία Cross Validation καμπύλη σφάλματος $CV(\theta)$, όπως αυτή που φαίνεται στο Διάγραμμα 3.19. Έτσι μπορούμε να επιλέξουμε την ρυθμιστική παράμετρο που ελαχιστοποιεί αυτή την καμπύλη.

$$\hat{\lambda} = \underset{\lambda \in \{\lambda_1, \dots, \lambda_k\}}{\operatorname{argmin}} CV(\lambda)$$



Διάγραμμα 3.19. Παράδειγμα Cross Validation καμπύλη σφάλματος

Συμπερασματικά, η Cross Validation εξυπηρετεί ως ένα μέτρο ικανότητας πρόβλεψης του μοντέλου. Έχοντας την βέλτιστη τιμή για το λ , μπορούμε να εκτελέσουμε τις μεθόδους ποινικοποιημένης πιθανοφάνειας για την εύρεση του καλύτερου μοντέλου.

4. Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης

Για την πρόβλεψη των εκπομπών των τριών αερίων του θερμοκηπίου κάθε χώρας χρειάστηκε να δημιουργηθούν 84 μοντέλα (28 χώρες x 3 αέρια του θερμοκηπίου). Για λόγους εξοικονόμησης χώρου, παραθέτω μόνο τρεις αναλύσεις (μία για κάθε αέριο του θερμοκηπίου) μαζί με τις προβλέψεις των μοντέλων τους. Για τα υπόλοιπα ακολούθησε η ίδια μεθοδολογία. Τα 84 μοντέλα που διαμορφώθηκαν βρίσκονται στο Παράρτημα Ι και τα αποτελέσματά των προβλέψεων τους παρουσιάζονται στο Κεφάλαιο 5.

4.1. Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές CO₂ της Δανίας

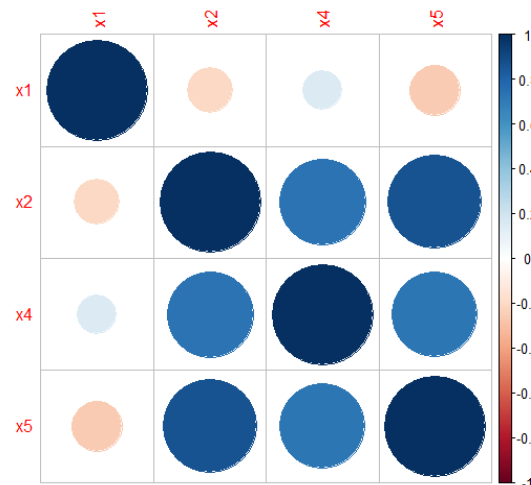
	CO ₂ (CARBON DIOXIDE)	GROSS INLAND CONSUMPTION OF NATURAL GAS	GROSS INLAND CONSUMPTION OF HARD COAL	GROSS INLAND CONSUMPTION OF LIGNITE	GROSS INLAND CONSUMPTION OF PETROLEUM PRODUCTS	GROSS INLAND CONSUMPTION OF CRUDE OIL
1995	62061.45	132736	11046	0	8939	9798
1996	75470.74	156362	15052	0	9925	10656
1997	66145.31	164628	11181	0	9548	8525
1998	62070.48	178861	9445	0	9532	7831
1999	59517.20	188025	7721	0	9312	8091
2000	55277.29	186269	6681	0	9027	8294
2001	57049.27	193890	7029	0	9119	8151
2002	56568.32	193738	6941	0	8726	7906
2003	61760.44	195134	9563	0	8268	8250
2004	56278.25	194007	7372	0	8262	8008
2005	52638.28	184195	6327	0	8103	7659
2006	60608.86	189937	9207	0	8249	7911
2007	55860.25	170052	7939	0	8237	7656
2008	52332.93	170536	6871	0	8036	7345
2009	49936.20	163097	6804	0	7332	7545
2010	50419.96	185140	6519	0	7635	7142
2011	45475.43	155640	5549	0	7174	6505
2012	40798.82	145886	4259	0	6962	7372

Πίνακας 4.1. Δεδομένα για τις εκπομπές CO₂ και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Δανίας

- Όπως φαίνεται από τον Πίνακα 4.1, η μεταβλητή Gross Inland Consumption of Lignite δεν περιλαμβάνεται στην ανάλυση, λόγω των τιμών της.
- Ελέγχουμε και τη συσχέτιση μεταξύ των επεξηγηματικών μεταβλητών αριθμητικά αλλά και γραφικά. Όσο πιο κοντά στην απόλυτη μονάδα είναι οι τιμές, τόσο πιο έντονη θα είναι η μεταξύ τους συσχέτιση. Αν είναι μηδέν τότε είναι ασυσχέτιστες.

	C.Natural Gas	C.Hard Coal	C.Petroleum Prod.	C.Crude Oil
C.Natural Gas	1	-0,203601	0,1575437	-0,2588137
C.Hard Coal	-0,203601	1	0,7322342	0,8626443
C.Petroleum Prod.	0,1575437	0,7322342	1	0,7267419
C.Crude Oil	-0,2588137	0,8626443	0,7267419	1

Πίνακας 4.2. Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CO₂ - Δανία)



Διάγραμμα 4.1. Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CO₂ - Δανία)

Variables	VIF
C.Natural Gas	1,50238
C.Hard Coal	4,39601
C.Petroleum Prod.	3,26601
C.Crude Oil	4,80452

Πίνακας 4.3. Τιμές VIF των επεξηγηματικών μεταβλητών (CO₂ - Δανία)

Παρόλο που στον Πίνακα 4.2 και στο Διάγραμμα 4.1 φαίνεται να υπάρχει κάποια συσχέτιση μεταξύ των μεταβλητών Gross Inland Consumption of Hard Coal (X₂) και Gross Inland Consumption of Crude Oil (X₅), από τις τιμές των VIF στον Πίνακα 4.3 μπορούμε να θεωρήσουμε ότι δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας, καθώς είναι όλες μικρότερες από 10.

- Προσαρμογή του γενικού γραμμικού μοντέλου στην R:

Call:

```
lm(formula = x0 ~ x1 + x2 + x4 + x5)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-1214.2 -364.4  128.6  356.3  594.3
```

Coefficients:

```
Estimate Std. Error t value Pr(> |t|)
(Intercept) -2.320e+02 2.324e+03 -0.100 0.922
x1           8.888e-02 8.889e-03 9.999 1.80e-07 ***
x2           2.422e+00 1.124e-01 21.549 1.48e-11 ***
x4           2.308e+00 2.824e-01 8.171 1.77e-06 ***
x5           2.805e-01 3.138e-01 0.894 0.388
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 554.5 on 13 degrees of freedom

Multiple R-squared: 0.9962, Adjusted R-squared: 0.995

F-statistic: 851.1 on 4 and 13 DF, p-value: 1.397e-15

- Στη συνέχεια εφαρμόζω τον κώδικα (Κεφάλαιο 3.3.4) που μου δίνει το καλύτερο από τα 16(=2⁴) πιθανά μοντέλα, με βάση τα κριτήρια AIC και BIC.

	x1	x2	x4	x5	BIC	AIC
1	x	x	x		288,198	283,746
2	x	x	x	x	290,015	284,673
3	x	x		x	319,778	315,326
4	x	x			321,092	317,530
5		x	x		324,303	320,742
6		x	x	x	326,045	321,593
7		x			341,695	339,024
8		x		x	344,436	340,874
9			x	x	350,831	347,270
10	x		x	x	351,984	347,532
11			x		355,395	352,724
12	x			x	356,989	353,428
13	x		x		358,048	354,486
14				x	360,659	357,988
15					378,743	376,962
16	x				381,508	378,837

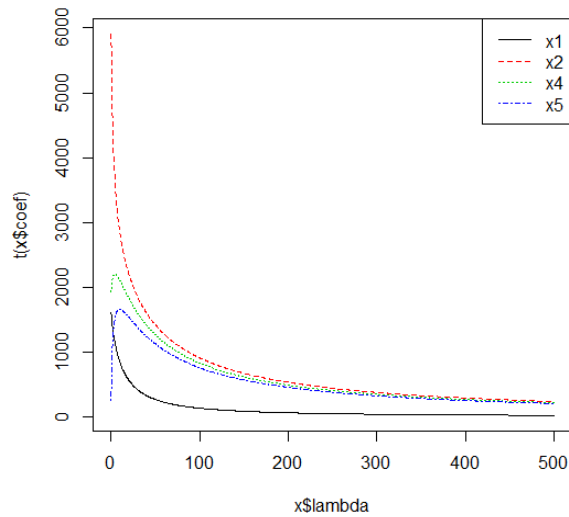
Πίνακας 4.4. Πίνακας όλων των πιθανών μοντέλων για την πρόβλεψη εκπομπών CO₂ της Δανίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC και AIC)

Με βάση και τα δύο κριτήρια καταλήγω στο ίδιο αποτέλεσμα, όπου το βέλτιστο μοντέλο περιλαμβάνει τις μεταβλητές Gross Inland Consumption of Natural Gas(X₁), Gross Inland Consumption of Hard Coal(X₂) και Gross Inland Consumption of Petroleum Products(X₄) (Πίνακας 4.4).

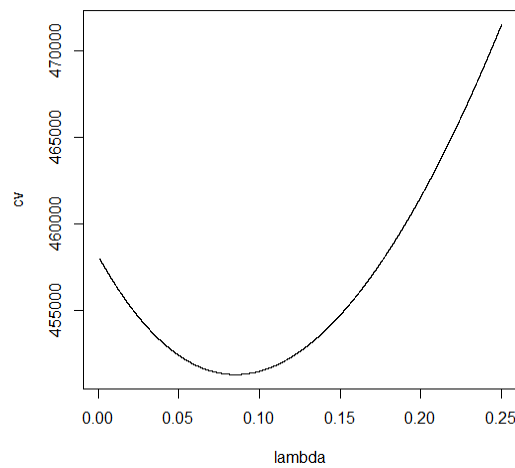
- Ακολούθως, εφαρμόζω και τις μεθόδους της Παλινδρόμησης Κορυφογραμμής (Ridge Regression) και Lasso:

Παλινδρόμηση Κορυφογραμμής (Ridge Regression):

Στο Διάγραμμα 4.2 παρουσιάζεται το ίχνος κορυφογραμμής, δηλαδή η γραφική παράσταση των συντελεστών για τις διάφορες τιμές του λ . Λόγω της ύπαρξης πολυσυγγραμμικότητας, η αστάθεια των συντελεστών παλινδρόμησης είναι εμφανής. Για κάποια τιμή του λ , οι εκτιμήσεις σταθεροποιούνται. Στο Διάγραμμα 4.3 παρουσιάζεται η 10-fold Cross Validation καμπύλη σφάλματος. Η τιμή του λ στην οποία ελαχιστοποιείται αυτή η καμπύλη ($\lambda = 0,08509571$) αντιστοιχεί στο μικρότερο μέσο τετραγωνικό σφάλμα.



Διάγραμμα 4.2. Ίχνος κορυφογραμμής (CO₂-Δανία)



Διάγραμμα 4.3. 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CO₂-Δανία)

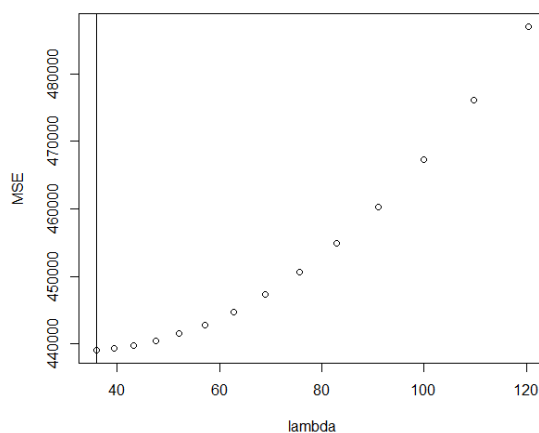
x1	0,0882
x2	2,3781
x4	2,3264
x5	0,3643

Πίνακας 4.5. Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CO₂-Δανία)

Από τον πίνακα των συντελεστών (Πίνακας 4.5) παρατηρούμε ότι με την Παλινδρόμηση Κορυφογραμμής η μεταβλητή Gross Inland Consumption of Natural Gas (X_1) με συντελεστή 0,0882 θεωρείται η λιγότερο σημαντική. Παρόλα αυτά κανένας συντελεστής δεν πλησιάζει έντονα στο μηδέν, επομένως μπορούμε να θεωρήσουμε ότι το καταλληλότερο μοντέλο περιλαμβάνει και τις τέσσερις μεταβλητές.

Lasso:

Χρησιμοποιώντας τη συνάρτηση 10-folds CVL για τη μέθοδο Lasso παίρνουμε την τιμή του λ στην οποία αντιστοιχεί το μικρότερο μέσο τετραγωνικό σφάλμα, $\lambda=32,7078$ (Διάγραμμα 4.4).



Διάγραμμα 4.4. Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (CO₂-Δανία)

x1	0,0858
x2	2,4120
x4	2,3416
x5	0,2299

Πίνακας 4.6. Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (CO₂-Δανία)

Από τον πίνακα των συντελεστών (Πίνακας 4.6) παρατηρούμε ότι η Lasso δεν μηδενίζει κανένα συντελεστή, παρόλο που οι δύο από αυτούς, των Gross Inland Consumption of Natural Gas (X_1) και Gross Inland Consumption of Crude Oil (X_5) είναι αρκετά κοντά στο μηδέν.

Συνοπτικά, με βάση τα κριτήρια AIC και BIC το βέλτιστο μοντέλο περιλαμβάνει τις μεταβλητές Gross Inland Consumption of Natural Gas (X_1), Gross Inland Consumption of Hard Coal (X_2) και Gross Inland Consumption of Petroleum Products (X_4), ενώ η μέθοδος της παλινδρόμησης κορυφογραμμής και η Lasso καταλήγουν και στις τέσσερις μεταβλητές ως σημαντικές. Στον Πίνακα 4.4 παρατηρώ ότι το μοντέλο με όλες τις μεταβλητές είναι το δεύτερο καλύτερο με βάση τα AIC και BIC. Επομένως, θεωρώ προτιμότερο να κρατήσω και τις τέσσερις μεταβλητές στο μοντέλο μου.

- Προσαρμογή του μοντέλου που επιλέχθηκε στην R

Call:

```
lm(formula = x0 ~ x1 + x2 + x4 + x5)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1214.2	-364.4	128.6	356.3	594.3

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.320e+02	2.324e+03	-0.100	0.922
x1	8.888e-02	8.889e-03	9.999	1.80e-07 ***
x2	2.422e+00	1.124e-01	21.549	1.48e-11 ***
x4	2.308e+00	2.824e-01	8.171	1.77e-06 ***
x5	2.805e-01	3.138e-01	0.894	0.388

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 554.5 on 13 degrees of freedom

Multiple R-squared: 0.9962, Adjusted R-squared: 0.995

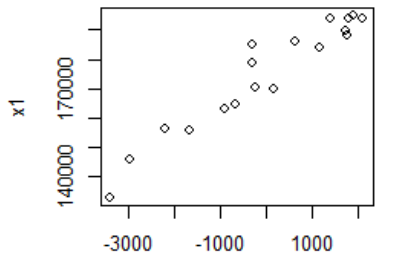
F-statistic: 851.1 on 4 and 13 DF, p-value: 1.397e-15

- 95% διαστήματα εμπιστοσύνης

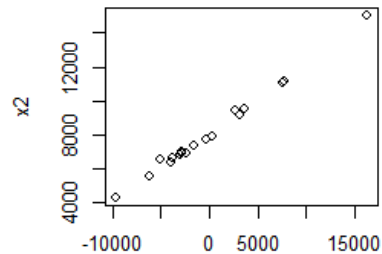
	2.5 %	97.5 %
(Intercept)	-5.252134e+03	4788.1359961
x1	6.967311e-02	0.1080782
x2	2.179316e+00	2.6649781
x4	1.697466e+00	2.9177235
x5	-3.974114e-01	0.9583859

- Τέλος, ελέγχουμε και τις προϋποθέσεις του γενικού γραμμικού μοντέλου

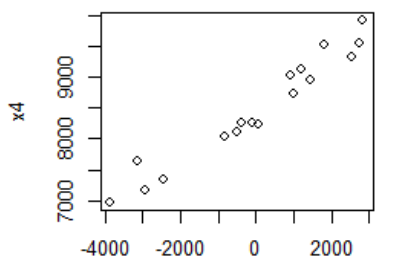
- Γραμμικότητα



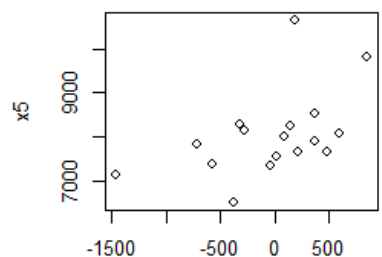
residuals(results_teliko, "partial")[, 1]



residuals(results_teliko, "partial")[, 2]



residuals(results_teliko, "partial")[, 3]

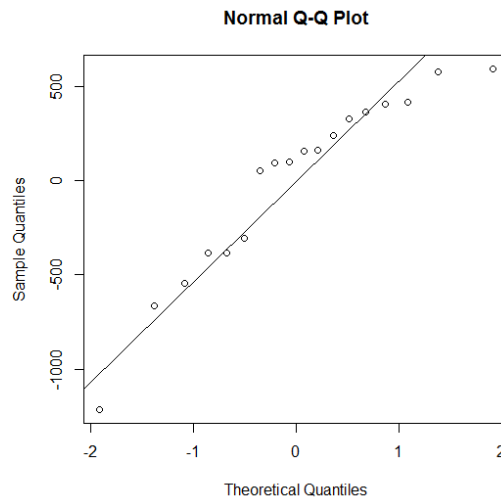


residuals(results_teliko, "partial")[, 4]

Διάγραμμα 4.5. Έλεγχος γραμμικότητας (CO₂-Δανάια)

Όπως φαίνεται από τις γραφικές παραστάσεις των επεξηγηματικών μεταβλητών ως προς τα μερικά υπόλοιπα (partial residuals), η υπόθεση της γραμμικότητας είναι λογική (Διάγραμμα 4.5).

○ **Κανονικότητα των υπολοίπων**

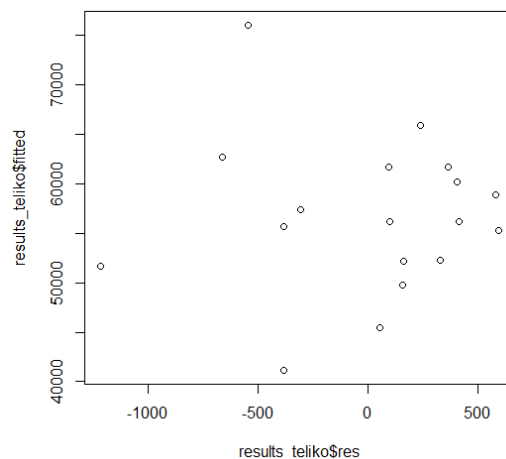


Διάγραμμα 4.6. Έλεγχος κανονικότητας των υπολοίπων (CO₂-Δανία)

Από τη γραφική παράσταση των κανονικοποιημένων υπολοίπων (Διάγραμμα 4.6) φαίνεται να έχουν ικανοποιητική προσαρμογή των σημείων στην ευθεία.

Jarque-Bera Normality Test: JB = 2.5732, p-value = 0.2762 > 0.05 άρα δεν απορρίπτουμε την κανονικότητα.

○ **Ομοσκεδαστικότητα**

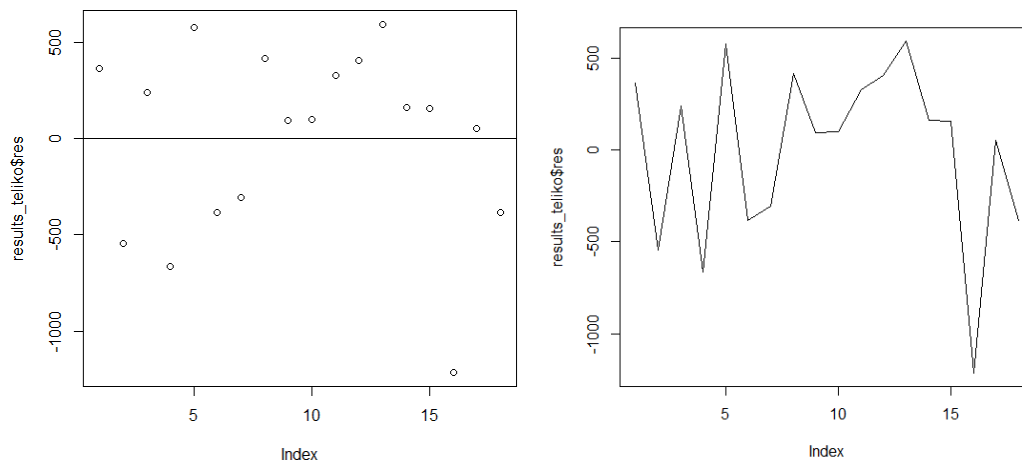


Διάγραμμα 4.7. Έλεγχος ομοσκεδαστικότητας (CO₂-Δανία)

Από τη γραφική παράσταση των υπολοίπων (residuals (res)) συναρτήσεως των προβλεπόμενων τιμών (fitted) (Διάγραμμα 4.7) δεν μπορούμε να σιγουρευτούμε για την ύπαρξη ομοσκεδαστικότητας οπότε προχωράμε στον έλεγχο Breusch-Pagan.

Breusch-Pagan test: BP = 0.84344, df = 4, p-value = 0.9325 άρα δεν απορρίπτουμε την ομοσκεδαστικότητα.

○ Ανεξαρτησία μεταξύ των υπολοίπων



Διάγραμμα 4.8. Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (CO₂-Δανία) (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)

Durbin-Watson test: DW = 2.3337, p-value = 0.4792 > 0.05 αποδέχομαι ανεξαρτησία υπολοίπων

Από τις γραφικές παραστάσεις των υπολοίπων (residuals (res)) (Διάγραμμα 4.8) αλλά και τον έλεγχο Durbin-Watson καταλήγουμε στο να μην απορρίψουμε την ανεξαρτησία μεταξύ των υπολοίπων.

Καταλήγουμε στο εξής μοντέλο:

$$y = -232 + 0.0889 * CNaturalGas + 2.422 * CHardCoal + 2.308 * CPetroleumProducts + 0.2805 * CrudeOil$$

Αντικαθιστώντας τα στοιχεία του 2013 στις επεξηγηματικές μεταβλητές, **το αποτέλεσμα της πρόβλεψης των εκπομπών διοξειδίου του άνθρακα είναι 42650.24 Gg**. Το διάστημα ατομικής πρόβλεψης (95% διάστημα εμπιστοσύνης) είναι (41202.18,44098.3).

```
> predict(results_teliko,list(x1<-138833, x2<-5363, x4<-6754, x5<-7016),int="p");
      fit      lwr      upr
42650.24 41202.18 44098.3
```

4.2. Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές CH₄ της Κύπρου

	CH ₄ (METHANE)	CATTLE POPULATION	PIG POPULATION	SHEEP POPULATION	MUNICIPAL WASTE Landfill disposal	TOTAL POPULATION (on 1 January)	URBAN POPULATION	GDP	PRIMARY PRODUCTION OF HARD COAL	PRIMARY PRODUCTION OF LIGNITE	GROSS INLAND CONSUMPTION OF NATURAL GAS	PRIMARY PRODUCTION OF NATURAL GAS	DISTRIBUTION LOSSES
1995	1119.30	64364	356200	255000	387	645399	581990	7011.7	0	0	0	0	0
1996	1131.60	68129	374100	250000	389	656333	595213	7302.6	0	0	0	0	0
1997	1145.85	70088	399530	252000	398	666313	608219	7771.2	0	0	0	0	0
1998	1154.77	62426	414788	250000	406	675215	621145	8452.9	0	0	0	0	0
1999	1166.36	55841	431300	240000	413	682862	634211	9107.7	0	0	0	0	0
2000	1185.95	54074	418500	233000	423	690497	647548	9949	0	0	0	0	0
2001	1246.57	54123	408400	246000	442	697549	661201	10719.6	0	0	0	0	0
2002	1282.81	53613	451300	296600	450	705539	673117	11081.1	0	0	0	0	0
2003	1287.13	58335	491400	294000	467	713720	684256	11654.2	0	0	0	0	0
2004	1293.63	58868	488101	264600	482	722893	694922	12596	0	0	0	0	0
2005	1275.12	60532	470504	278984	489	733067	704926	13598.2	0	0	0	0	0
2006	1276.70	57799	429717	268874	499	744013	714164	14670.5	0	0	0	0	0
2007	1265.75	56109	452644	272192	512	757916	722713	15901.5	0	0	0	0	0
2008	1273.71	55769	464932	267308	531	776333	730686	17157.1	0	0	0	0	0
2009	1276.37	54097	463263	300200	540	796930	738250	16853.5	0	0	0	0	0
2010	1294.73	54715	463729	328864	490	819140	745550	17406	0	0	0	0	0
2011	1241.37	56915	439236	355944	461	839751	752597	17878	0	0	0	0	0
2012	1303.68	56924	394706	346836	451	862011	759373	17720.2	0	0	0	0	0

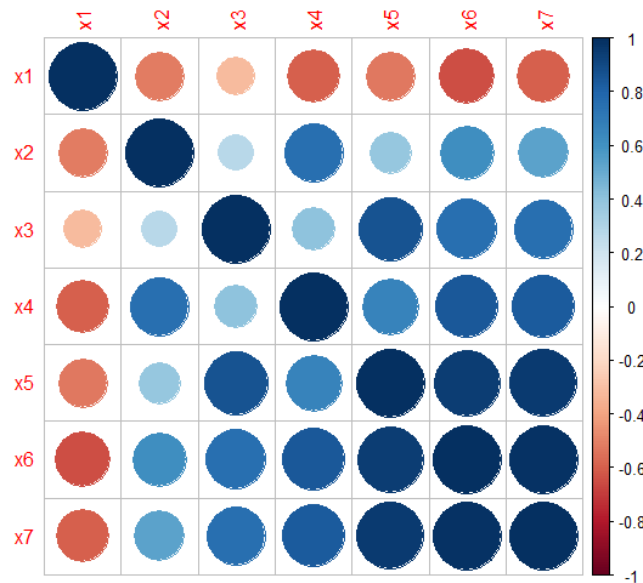
Πίνακας 4.7. Δεδομένα για τις εκπομπές CH₄ και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Κύπρου

- Αφαιρούμε τις 5 τελευταίες μεταβλητές από την ανάλυσή μας, καθώς όπως φαίνεται στον Πίνακα 4.7 παίρνουν μόνο μηδενικές τιμές.
- Ελέγχουμε και τη συσχέτιση μεταξύ των επεξηγηματικών μεταβλητών αριθμητικά αλλά και γραφικά. Όσο πιο κοντά στην απόλυτη μονάδα είναι οι τιμές, τόσο πιο έντονη θα είναι η μεταξύ τους συσχέτιση. Αν είναι μηδέν τότε είναι ασυσχέτιστες.

	Cattle	Pig	Sheep	Waste
Cattle	1	-0.5188897	-0.3159884	-0.5949188
Pig	-0.5188897	1	0.2799366	0.7534626
Sheep	-0.3159884	0.2799366	1	0.4015699
Waste	-0.5949188	0.7534626	0.4015699	1
T.Population	-0.5272313	0.3865339	0.8630027	0.6683158
U.Population	-0.6449997	0.6111258	0.7577167	0.8421000
GDP	-0.5962856	0.5320177	0.7543975	0.8386296

	T.Population	U.Population	GDP
Cattle	-0.5272313	-0.6449997	-0.5962856
Pig	0.3865339	0.6111258	0.5320177
Sheep	0.8630027	0.7577167	0.7543975
Waste	0.6683158	0.8421000	0.8386296
T.Population	1	0.9456133	0.9590143
U.Population	0.9456133	1	0.9846052
GDP	0.9590143	0.9846052	1

Πίνακας 4.8. Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CH₄ - Κύπρος)



Διάγραμμα 4.9. Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (CH₄- Κύπρος)

Variables	VIF
Cattle	2,1461
Pig	4,7166
Sheep	6,3678
Waste	21,3914
T.Population	83,2275
U.Population	77,0617
GDP	132,8746

Πίνακας 4.9. Τιμές VIF των επεξηγηματικών μεταβλητών (CH₄- Κύπρος)

Από τους Πίνακες 4.8 και 4.9 αλλά και από το Διάγραμμα 4.9, βλέπουμε ότι υπάρχει πρόβλημα πολυσυγγραμμικότητας, καθώς πολλές από τις μεταβλητές συσχετίζονται μεταξύ τους σε μεγάλο βαθμό. Συγκεκριμένα, οι μεταβλητές Total Population, Urban Population και GDP φαίνεται να προκαλούν το μεγαλύτερο πρόβλημα έχοντας τιμές συσχέτισης μεγαλύτερες από 0.90 και VIF τιμές πολύ μεγαλύτερες από 10. Κατά συνέπεια, θα ήταν προτιμότερο να μην συμπεριλαμβάνονται και οι τρεις επεξηγηματικές μεταβλητές στο μοντέλο καθώς δίνουν την ίδια πληροφορία.

- Προσαρμογή του πολλαπλού γραμμικού μοντέλου στην R

```

Call:
lm(formula = x0 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7)

Residuals:
    Min       1Q   Median       3Q      Max
-29.1779 -11.1358  0.8468  11.7060  21.5372

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.927e+02  4.940e+02  -0.795  0.44509
x1           -5.038e-04  1.440e-03  -0.350  0.73369
x2           -2.982e-04  2.733e-04  -1.091  0.30081
x3            5.236e-04  3.344e-04   1.566  0.14849
x4            6.371e-01  4.621e-01   1.379  0.19804
x5           -5.934e-04  6.790e-04  -0.874  0.40263
x6            3.217e-03  7.414e-04   4.338  0.00147 **
x7           -3.242e-02  1.416e-02  -2.289  0.04510 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.73 on 10 degrees of freedom
Multiple R-squared:  0.9443,    Adjusted R-squared:  0.9054
F-statistic: 24.24 on 7 and 10 DF,  p-value: 1.668e-05

```

- Εφαρμόζουμε τον κώδικα (Κεφάλαιο 3.3.4) που μας δίνει το καλύτερο από τα 128 ($=2^7$) πιθανά μοντέλα, με βάση τα κριτήρια AIC και BIC. Για λόγους εξοικονόμησης χώρου, παραθέτω δύο πίνακες με μόνο τα δέκα καταλληλότερα μοντέλα, με βάση το κάθε κριτήριο. Τα μοντέλα αυτά κρίνονται καταλληλότερα λόγω του ότι έχουν μικρότερες τιμές AIC και BIC από τα υπόλοιπα μοντέλα.

	x1	x2	x3	x4	x5	x6	x7	BIC	AIC
1				x		x	x	166,974	162,522
2			x	x		x	x	168,201	162,859
3		x	x	x		x	x	169,693	163,460
4			x		x	x	x	168,993	163,651
5					x	x	x	168,227	163,775
6		x		x		x	x	169,229	163,887
7		x	x	x	x	x	x	171,209	164,086
8			x	x	x	x	x	170,319	164,086
9				x	x	x	x	169,822	164,480
10	x			x		x	x	169,863	164,521

Πίνακας 4.10. Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών CH₄ της Κύπρου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC)

	x1	x2	x3	x4	x5	x6	x7	BIC	AIC
1				x		x	x	166,974	162,522
2			x	x		x	x	168,201	162,859
3					x	x	x	168,227	163,775
4			x		x	x	x	168,993	163,651
5		x		x		x	x	169,229	163,887
6		x	x	x		x	x	169,693	163,460
7				x	x	x	x	169,822	164,480
8	x			x		x	x	169,863	164,521
9			x	x	x	x	x	170,319	164,086
10	x		x	x		x	x	170,857	164,625

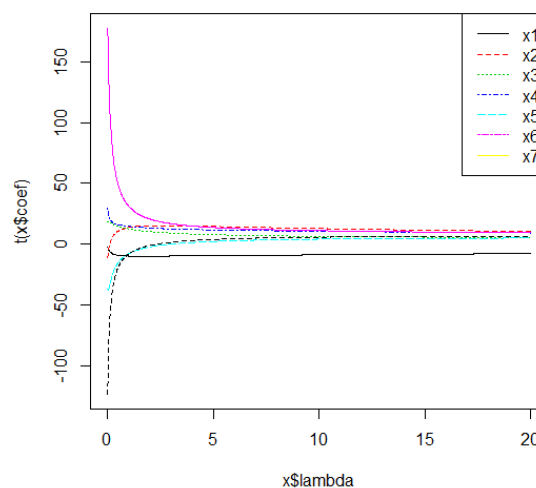
Πίνακας 4.11. Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών CH₄ της Κύπρου (ταξινομημένα ξεινώντας από το μοντέλο με το μικρότερο AIC)

Όπως μπορούμε να δούμε από τους Πίνακες 4.10 και 4.11, τα κριτήρια καταλήγουν στο ίδιο μοντέλο ως καταλληλότερο με μεταβλητές τις Waste (X₁), Urban Population (X₆) και GDP (X₇).

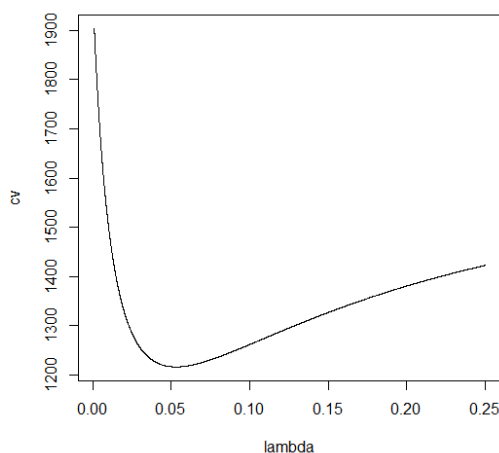
- Ακολουθώντας, εφαρμόζουμε και τις μεθόδους Παλινδρόμηση Κορυφογραμμής (Ridge Regression) και Lasso:

Παλινδρόμηση Κορυφογραμμής (Ridge Regression):

Στο Διάγραμμα 4.10 παρουσιάζεται η πορεία των συντελεστών της παλινδρόμησης κορυφογραμμής, καθώς η συγκεκριμένη μέθοδος οδηγεί στη συρρίνωσή τους προς το μηδέν. Στο Διάγραμμα 4.11 παρουσιάζεται η καμπύλη σφάλματος 10-folds CVL. Η τιμή του λ στην οποία ελαχιστοποιείται αυτή η καμπύλη ($\lambda = 0,05289679$) αντιστοιχεί στο μικρότερο μέσο τετραγωνικό σφάλμα.



Διάγραμμα 4.10. Τυχνος κορυφογραμμής (CH₄-Κύπρος)



Διάγραμμα 4.11. 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CH₄-Κύπρος)

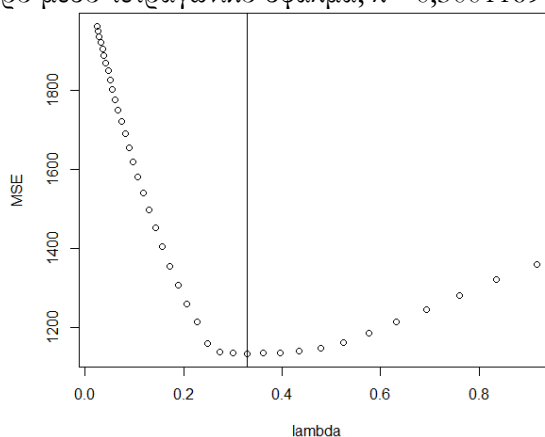
x1	-0.0009848381
x2	-0.0001023725
x3	0.0005223839
x4	0.4956460508
x5	-0.0005847081
x6	0.0024513272
x7	-0.0214399094

Πίνακας 4.12. Τιμές των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (CH₄-Κύπρος)

Από τον πίνακα των συντελεστών (Πίνακας 4.12) παρατηρούμε ότι με την Ridge Regression η μεταβλητή Waste (X₄) με συντελεστή 0.4956, ο οποίος αποκλίνει από το μηδέν, θεωρείται η πιο σημαντική. Επίσης οι μεταβλητές Urban Population (X₆) και GDP (X₇) μπορούμε να πούμε ότι είναι σημαντικές, ενώ οι συντελεστές των υπόλοιπων μεταβλητών πλησιάζουν κατά πολύ το μηδέν.

Lasso:

Χρησιμοποιώντας τη συνάρτηση 10-folds CVL παίρνουμε την τιμή του λάμδα στην οποία αντιστοιχεί το μικρότερο μέσο τετραγωνικό σφάλμα, $\lambda = 0,3001169$ (Διάγραμμα 4.12).



Διάγραμμα 4.12 Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (CH₄-Κύπρος)

x1	-7.288732e-04
x2	-2.420259e-05
x3	4.009375e-04
x4	3.931546e-01
x5	-5.116714e-04
x6	2.437842e-03
x7	-2.069561e-02

Πίνακας 4.13. Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (CH₄-Κύπρος)

Από τον πίνακα των συντελεστών (Πίνακας 4.13) παρατηρούμε ότι η Lasso δεν μηδενίζει κανένα συντελεστή, αλλά πολλοί από αυτούς πλησιάζουν στο μηδέν. Οι συντελεστές που δεν πλησιάζουν έντονα το μηδέν είναι ανήκουν στις μεταβλητές Waste (X₄), Urban Population (X₆) και GDP (X₇).

Συνοπτικά, με βάση τα κριτήρια AIC και BIC, αλλά και τη μέθοδο της παλινδρόμησης κορυφογραμμής και τη Lasso, το βέλτιστο μοντέλο περιλαμβάνει τις μεταβλητές Waste (X₄), Urban Population (X₆) και GDP (X₇). Επομένως, επιλέγω αυτές τις τρεις μεταβλητές για το μοντέλο μου.

- Ανάλυση παλινδρόμησης του βέλτιστου μοντέλου που επιλέχτηκε:

Call:

```
lm(formula = x0 ~ x4 + x6 + x7)
```

Residuals:

Min	1Q	Median	3Q	Max
-25.2751	-15.2362	0.7459	15.3311	25.7033

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.598e+02	2.341e+02	-2.391	0.031411 *
x4	4.846e-01	1.791e-01	2.706	0.017061 *
x6	2.929e-03	4.718e-04	6.208	2.28e-05 ***
x7	-3.363e-02	6.795e-03	-4.949	0.000214 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.98 on 14 degrees of freedom

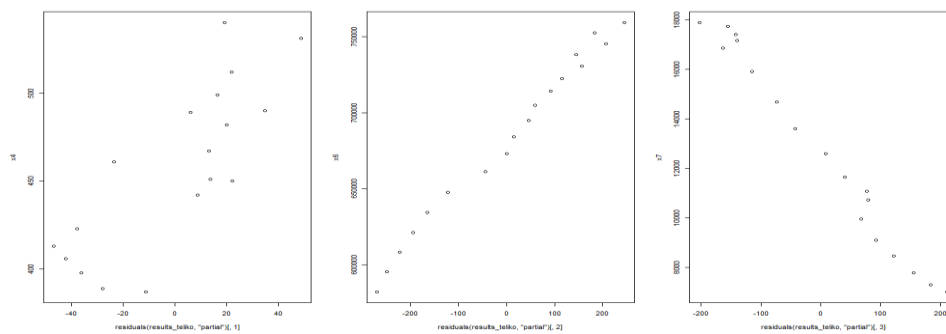
Multiple R-squared: 0.9279, Adjusted R-squared: 0.9125

F-statistic: 60.08 on 3 and 14 DF, p-value: 3.074e-08

- 95% διαστήματα εμπιστοσύνης

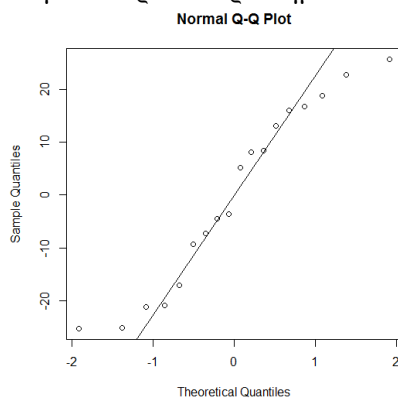
	2.5 %	97.5 %
(Intercept)	-1.061930e+03	-57.622059430
x4	1.004770e-01	0.868694606
x6	1.916990e-03	0.003940759
x7	-4.820455e-02	-0.019055219

- Προϋποθέσεις του γενικού γραμμικού μοντέλου
 - Η υπόθεση της **γραμμικότητας** είναι λογική με βάση τις γραφικές παραστάσεις των επεξηγηματικών μεταβλητών ως προς τα μερικά υπόλοιπα (Διάγραμμα 4.13).



Διάγραμμα 4.13. Έλεγχος γραμμικότητας (CH₄-Κύπρος)

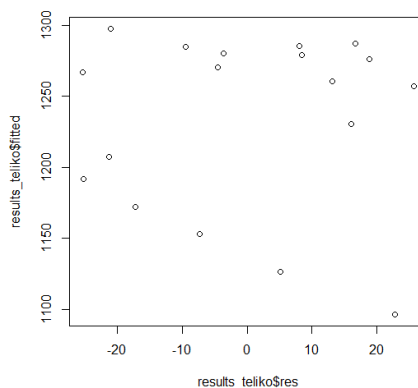
- Η **κανονικότητα των υπολοίπων**, όπως φαίνεται πολύ καθαρά από το Διάγραμμα 4.14 είναι λογική και η ευθεία που σχηματίζεται είναι προσαρμοσμένη στα περισσότερα σημεία.



Διάγραμμα 4.14. Έλεγχος κανονικότητας υπολοίπων (CH₄-Κύπρος)

Jarque-Bera Normality Test: JB = 1.4174, p-value = 0.4923 > 0.05 άρα δεν απορρίπτουμε την υπόθεση κανονικότητας.

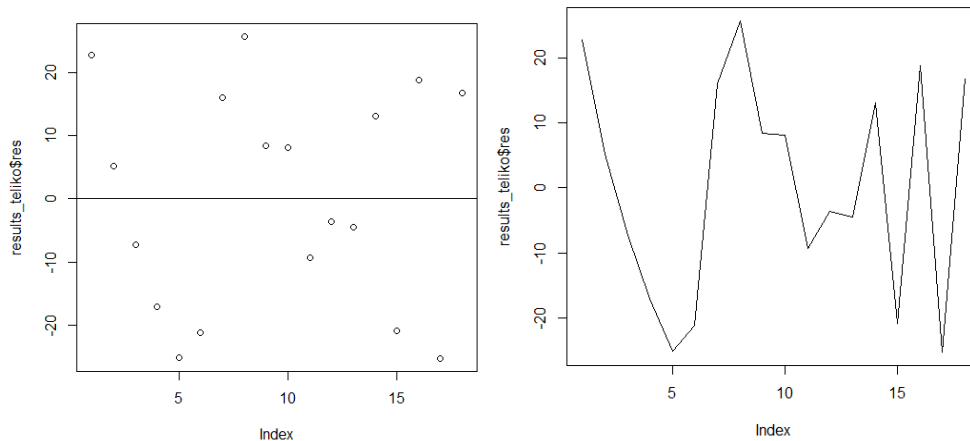
- Με το Διάγραμμα 4.15 παρουσιάζονται τα υπόλοιπα (residuals (res)) (άξονας x) συναρτήσει των προβλεπόμενων τιμών (fitted) (άξονας y). Παρατηρούμε τα ζεύγη των τιμών να μην εμφανίζουν κάποιο συστηματικό τρόπο συμπεριφοράς, οπότε η υπόθεση της **ομοσκεδαστικότητας** είναι λογική.



Διάγραμμα 4.15. Έλεγχος ομοσκεδαστικότητας (CH₄-Κύπρος)

Breusch-Pagan test: BP = 2.5895, df = 3, p-value = 0.4593 δεν απορρίπτουμε την υπόθεση ομοσκεδαστικότητας.

- Στο Διάγραμμα 4.16 βλέπουμε ότι υπάρχει **ανεξαρτησία μεταξύ των υπολοίπων**, γιατί δεν παρουσιάζεται κάποια σχέση μεταξύ τους και τα υπόλοιπα συμπεριφέρονται τυχαία.



Διάγραμμα 4.16. Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)

Durbin-Watson test: DW = 1.8893, p-value = 0.1347 > 0.05 άρα δεν απορρίπτουμε την ανεξαρτησία των υπολοίπων.

Με την πιο πάνω διαδικασία καταλήγουμε στο εξής μοντέλο:

$$y = -559.8 + 0.4846 * \text{Waste} + 0.0029 * \text{UrbanPopulation} - 0.0336 * \text{GDP}$$

Αντικαθιστώντας τα στοιχεία του 2013 στις επεξηγηματικές μεταβλητές, **το αποτέλεσμα της πρόβλεψης των εκπομπών υποξειδίου του αζώτου είναι 1333.994 Gg. Το διάστημα ατομικής πρόβλεψης της y (95% διάστημα εμπιστοσύνης) είναι (1272.484, 1395.504).**

```
> predict(results_teliko, list(x4 < -423, x6 < -766099, x7 < -16503.7), int = "p");
```

fit	lwr	upr
1333.994	1272.484	1395.504

4.3. Ανάλυση Πολλαπλής Γραμμικής Παλινδρόμησης για τις εκπομπές N₂O της Ολλανδίας

	N ₂ O (NITROUS OXIDE)	CATTLE POPULATION	PIG POPULATION	SHEEP POPULATION	GOAT POPULATION	CROP PRODUCTION	GROSS VALUE ADDED FOR MANUFACTURE OF CHEMICALS AND CHEMICAL PRODUCTS	GROSS INLAND CONSUMPTION OF HARD COAL	GROSS INLAND CONSUMPTION OF LIGNITE	GROSS INLAND CONSUMPTION OF NATURAL GAS	GROSS INLAND CONSUMPTION OF PETROLEUM PRODUCTS	GROSS INLAND CONSUMPTION OF CRUDE OIL
1995	19891.44	4654000	14100000	1674000	73000	19716993	7587.8	13865	33	1445057	25320	62491
1996	19827.63	4557000	13900000	1627000	73000	20673599	7214.4	13528	31	1591837	24901	63492
1997	19530.03	4411000	13700000	1465000	110000	20144081	7198.8	13551	43	1497446	25741	63392
1998	18806.25	4283000	12800000	1394000	130000	16258531	6855.4	13923	20	1488174	26269	64726
1999	18084.84	4206000	13600000	1401000	145000	21671905	6842	11940	35	1448532	27122	61292
2000	17411.94	4070000	13100000	1308000	165000	21187105	7403	12706	46	1465758	27929	61788
2001	16348.21	4047000	12800000	1296000	190000	19177818	7673	13513	57	1502488	28866	62134
2002	15503.35	3858000	11500000	1186000	232000	19912902	8195	13479	52	1499084	28923	57515
2003	15290.14	3759000	11100000	1185000	268000	19207065	8397	13810	39	1507292	30559	59944
2004	15742.97	3767000	11100000	1236000	290000	20960129	9357	13590	45	1538877	31015	62063
2005	15459.9	3799000	11200000	1363000	300000	19561124	9663	12932	50	1479370	31554	62622
2006	15328.02	3749000	11200000	1376000	310000	18247998	10334	12424	35	1436488	31146	61033
2007	13583.5	3763000	11600000	1369000	340000	19154795	10832	13536	58	1396378	34141	55162
2008	9702.21	3890000	11900000	1213000	355000	19509666	10414	12883	25	1456052	33316	55856
2009	9453.12	3967000	12186450	1116610	374180	20672065	7333	11997	28	1469048	32416	54431
2010	9345.67	3975190	12254970	1129500	352830	19585002	9654	12096	30	1645772	33482	57228
2011	9281.21	3885350	12429140	1088490	380350	20895679	10731	11909	38	1436709	32180	55715
2012	9060.8	3879250	12233650	1042760	396730	19748892	10455	13108	40	1381760	33043	55170

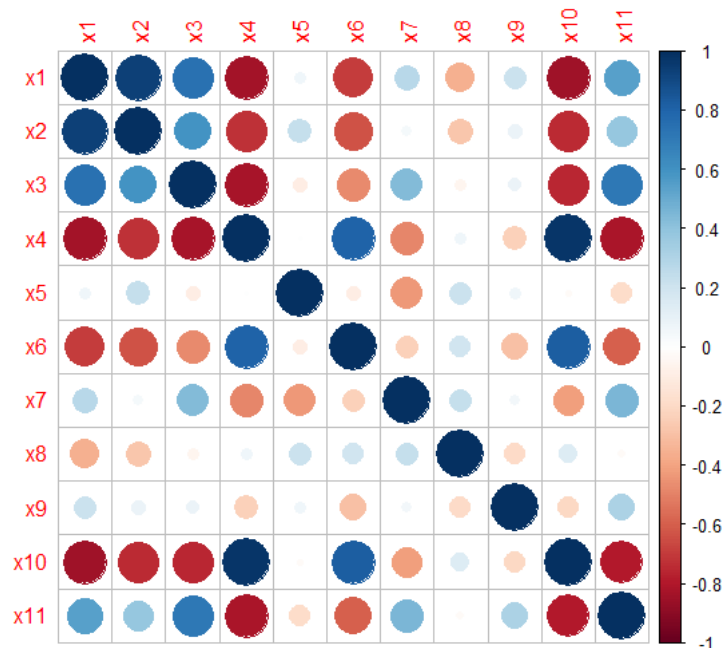
Πίνακας 4.14. Δεδομένα για τις εκπομπές N₂O και τις επεξηγηματικές μεταβλητές για το χρονικό διάστημα 1995-2012 της Ολλανδίας

- Ελέγχουμε και τη συσχέτιση μεταξύ των επεξηγηματικών μεταβλητών, που παρουσιάζονται στον Πίνακα 4.14, αριθμητικά αλλά και γραφικά. Όσο πιο κοντά στην απόλυτη μονάδα είναι οι τιμές, τόσο πιο έντονη θα είναι η μεταξύ τους συσχέτιση. Αν είναι μηδέν τότε είναι ασυσχέτιστες.

	Cattle	Pig	Sheep	Goat	Crop Prod.	GVA Chem.
Cattle	1	0.93058996	0.74664133	-0.831637316	0.061553102	-0.69545657
Pig	0.9305900	1	0.59144555	-0.729074739	0.238339635	-0.63995577
Sheep	0.7466413	0.59144555	1	-0.825648569	-0.090365067	-0.47274831
Goat	-0.8316373	-0.72907474	-0.82564857	1	0.009092407	0.80725714
Crop Prod.	0.0615531	0.23833964	-0.09036507	0.009092407	1	-0.09174491
GVA Chem.	-0.6954566	-0.63995577	-0.47274831	0.807257141	-0.091744914	1
C.Hard Coal	0.2708519	0.04505651	0.43473973	-0.481441567	-0.432023459	-0.23135726
C.Lignite	-0.3587428	-0.27158194	-0.05818835	0.060250416	0.213450449	0.19112882
C.Natural Gas	0.2119379	0.08553360	0.08276992	-0.235775005	0.060294352	-0.29464019
C.Petroleum Prod.	-0.8412883	-0.74145946	-0.75422383	0.973042797	-0.021032524	0.82947972
C.Crude Oil	0.5426797	0.38903627	0.71634611	-0.818898310	-0.189446605	-0.59829195

	C.Hard Coal	C.Lignite	C.Natural Gas	C.Petroleum Prod.	C.Crude Oil
Cattle	0.27085189	-0.35874278	0.21193794	-0.84128825	0.54267971
Pig	0.04505651	-0.27158194	0.08553360	-0.74145946	0.38903627
Sheep	0.43473973	-0.05818835	0.08276992	-0.75422383	0.71634611
Goat	-0.48144157	0.06025042	-0.23577501	0.97304280	-0.81889831
Crop Prod.	-0.43202346	0.21345045	0.06029435	-0.02103252	-0.18944661
GVA Chem.	-0.23135726	0.19112882	-0.29464019	0.82947972	-0.59829195
C.Hard Coal	1	0.23207664	0.05204971	-0.41674776	0.45890960
C.Lignite	0.23207664	1	-0.19403620	0.14164860	-0.02573213
C.Natural Gas	0.05204971	-0.19403620	1	-0.20381510	0.31910485
C.Petroleum Prod.	-0.41674776	0.14164860	-0.20381510	1	-0.79797603
C.Crude Oil	0.45890960	-0.02573213	0.31910485	-0.79797603	1

Πίνακας 4.15. Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (N₂O - Ολλανδία)



Διάγραμμα 4.17. Διάγραμμα συσχέτισης των επεξηγηματικών μεταβλητών (N₂O - Ολλανδία)

Variables	VIF
Cattle	110,9650
Pig	65,4185
Sheep	23,6547
Goat	123,8853
Crop Prod.	2,1369
GVA Chem.	6,7611
C.Hard Coal	7,3917
C.Lignite	3,2181
C.Natural Gas	3,8198
C.Petroleum Prod.	68,6454
C.Crude Oil	6,3273

Πίνακας 4.16. Τιμές VIF των επεξηγηματικών μεταβλητών (N₂O - Ολλανδία)

Από το Διάγραμμα 4.17 και τον Πίνακα 4.15, παρατηρούμε ότι μερικές από τις μεταβλητές συσχετίζονται μεταξύ τους και από τις τιμές VIF στον Πίνακα 4.16 επιβεβαιώνουμε την ύπαρξη πολυσυγγραμμικότητας.

- Προσαρμογή του πολλαπλού γραμμικού μοντέλου στην R

Call:

```
lm(formula = x0 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11)
```

Residuals:

```
Min    1Q  Median    3Q   Max
-767.46 -112.40  -0.19  283.40  493.12
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.027e+04	1.496e+04	2.024	0.0894
x1	-5.027e-03	5.469e-03	-0.919	0.3934
x2	-2.478e-04	1.181e-03	-0.210	0.8408
x3	8.116e-03	4.080e-03	1.989	0.0938
x4	-1.662e-02	1.466e-02	-1.133	0.3003
x5	9.006e-05	1.735e-04	0.519	0.6223
x6	-2.357e-01	2.579e-01	-0.914	0.3960
x7	3.239e-01	5.630e-01	0.575	0.5860
x8	1.765e+01	2.443e+01	0.722	0.4972
x9	-2.903e-03	4.502e-03	-0.645	0.5428
x10	-3.678e-01	3.959e-01	-0.929	0.3887
x11	2.028e-01	1.075e-01	1.887	0.1080

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 604.3 on 6 degrees of freedom

Multiple R-squared: 0.9916, Adjusted R-squared: 0.9762

F-statistic: 64.44 on 11 and 6 DF, p-value: 2.568e-05

- Εφαρμόζουμε τον κώδικα (Κεφάλαιο 3.3.4) που μας δίνει το καλύτερο από τα 2048 ($=2^{11}$) πιθανά μοντέλα, με βάση τα κριτήρια AIC και BIC. Για λόγους εξοικονόμησης χώρου, παραθέτω μόνο τα δέκα καταλληλότερα μοντέλα με βάση το κάθε κριτήριο.

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	BIC	AIC
1	x		x	x		x	x	x		x	x	292,715	283,812
2		x	x	x	x			x	x		x	291,835	283,821
3	x		x	x		x	x	x	x	x	x	294,476	284,682
4		x	x	x	x	x		x	x		x	293,688	284,785
5		x	x	x				x	x		x	292,006	284,883
6	x		x	x			x	x		x	x	292,906	284,893
7		x	x	x			x	x	x		x	293,006	284,992
8	x		x	x			x			x	x	292,216	285,093
9	x	x	x	x				x	x	x	x	294,042	285,138
10	x	x	x	x		x	x			x	x	294,100	285,197

Πίνακας 4.17. Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών N₂O της Ολλανδίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο AIC)

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	BIC	AIC
1		x	x	x	x			x	x		x	291,835	283,821
2		x	x	x				x	x		x	292,006	284,883
3	x		x	x			x			x	x	292,216	285,093
4	x		x	x						x	x	292,255	286,022
5	x		x	x				x		x	x	292,520	285,397
6	x		x	x						x		292,535	287,193
7	x		x	x		x	x	x		x	x	292,715	283,812
8		x	x	x	x				x		x	292,788	285,665
9	x		x	x			x			x		292,832	286,599
10	x		x	x			x	x		x	x	292,906	284,893

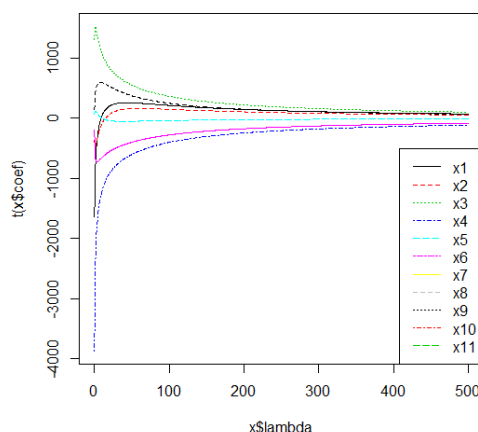
Πίνακας 4.18. Πίνακας των δέκα καταλληλότερων μοντέλων για την πρόβλεψη εκπομπών N_2O της Ολλανδίας (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC)

Με βάση το κριτήριο AIC, το καταλληλότερο μοντέλο περιλαμβάνει τις μεταβλητές Cattle (X_1), Sheep (X_3), Goat (X_4), GVA for Manufacture of Chemicals and Chemical Products (X_6), Gross Inland Consumption of Hard Coal (X_7), Gross Inland Consumption of Lignite (X_8), Gross Inland Consumption of Petroleum Products (X_{10}) και Gross Inland Consumption of Crude Oil (X_{11}) (Πίνακας 4.17). Με βάση το κριτήριο BIC, το καταλληλότερο μοντέλο περιλαμβάνει τις μεταβλητές Pig (X_2), Sheep (X_3), Goat (X_4), Crop Production (X_5), Gross Inland Consumption of Lignite (X_8), Gross Inland Consumption of Natural Gas (X_9) και Gross Inland Consumption of Crude Oil (X_{11}) (Πίνακας 4.18).

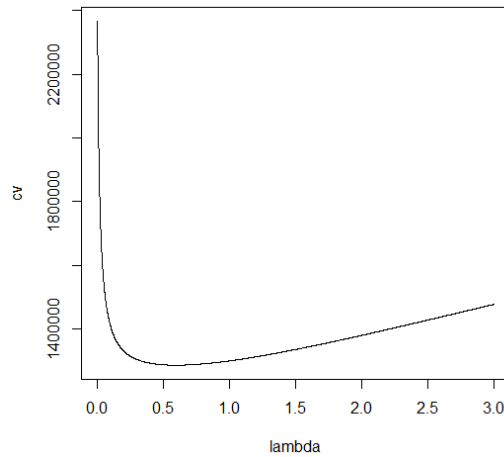
Έχοντας δύο διαφορετικά βέλτιστα μοντέλα από τα κριτήρια AIC και BIC, μπορούμε να χρησιμοποιήσουμε τις δύο επόμενες μεθόδους ποινικοποιημένης πιθανοφάνειας (Παλινδρόμηση Κορυφογραμμής και Lasso), οι οποίες αντιμετωπίζουν και το πρόβλημα της πολυσυγγραμμικότητας, ώστε να εξάγουμε καλύτερα συμπεράσματα.

Παλινδρόμηση Κορυφογραμμής (Ridge Regression):

Αρχικά χρησιμοποιούμε τη μέθοδο της Παλινδρόμησης Κορυφογραμμής (Ridge Regression) και προσπαθούμε να βρούμε το βέλτιστο λ με τη μέθοδο 10-fold CVL (Cross Validation), που αντιστοιχεί στο μικρότερο μέσο τετραγωνικό σφάλμα. Στο Διάγραμμα 4.18 παρουσιάζεται η πορεία των συντελεστών της παλινδρόμησης κορυφογραμμής ενώ συρρικνώνονται προς το μηδέν. Στο Διάγραμμα 4.19 παρουσιάζεται η τιμή του λ ($\lambda = 0.5930618$), στην οποία ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα της 10-folds CVL.



Διάγραμμα 4.18. Έργο κορυφογραμμής (N_2O - Ολλανδία)



Διάγραμμα 4.19. 10-fold Cross Validation καμπύλη σφάλματος για τη μέθοδο της Παλινδρόμησης Κορυφογραμμής ($\mathbf{N}_2\mathbf{O}$ - Ολλανδία)

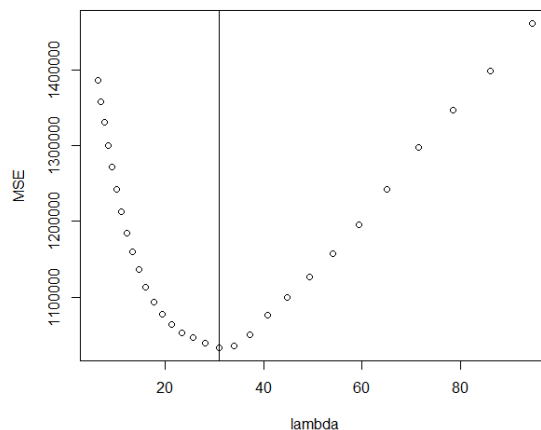
x1	-0,0022685
x2	-0,0005297
x3	0,0071230
x4	-0,0126105
x5	0,0000836
x6	-0,2917064
x7	0,2903617
x8	32,1970200
x9	-0,0037250
x10	-0,3305348
x11	0,2664540

Πίνακας 4.19. Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο της Παλινδρόμησης Κορυφογραμμής ($\mathbf{N}_2\mathbf{O}$ - Ολλανδία)

Από τον πίνακα των συντελεστών (Πίνακας 4.19) παρατηρούμε ότι με την Ridge Regression η μεταβλητή Gross Inland Consumption of Lignite (X_8) με συντελεστή 32.197, ο οποίος αποκλίνει κατά πολύ από το μηδέν, θεωρείται η πιο σημαντική. Επίσης οι μεταβλητές Goat (X_4), GVA for Manufacture of Chemicals and Chemical Products (X_6), Gross Inland Consumption of Hard Coal (X_7), Gross Inland Consumption of Petroleum Products (X_{10}) και Gross Inland Consumption of Crude Oil (X_{11}) μπορούμε να πούμε ότι είναι σημαντικές, ενώ οι συντελεστές των υπόλοιπων μεταβλητών πλησιάζουν αρκετά το μηδέν.

Lasso:

Χρησιμοποιώντας τη συνάρτηση 10-folds CVL παίρνουμε την τιμή του λάμδα στην οποία αντιστοιχεί το μικρότερο μέσο τετραγωνικό σφάλμα, $\lambda = 30.94488$ (Διάγραμμα 4.20).



Διάγραμμα 4.20. Γραφική παράσταση των μέσων τετραγωνικών σφαλμάτων ως προς το λ (N_2O - Ολλανδία)

```

x1      -3.429430e-04
x2      -9.867491e-04
x3       4.880138e-03
x4      -2.372360e-02
x5       5.704635e-05
x6      -1.138946e-01
x7       .
x8       3.196184e+01
x9      -4.813407e-03
x10     -1.145013e-01
x11     2.506747e-01

```

Πίνακας 4.20. Εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών με τη μέθοδο Lasso (N_2O - Ολλανδία)

Από τον πίνακα των συντελεστών (Πίνακας 4.20) παρατηρούμε ότι η Lasso μηδενίζει τον συντελεστή της μεταβλητής Gross Inland Consumption of Hard Coal (X_7). Οι συντελεστές που δεν πλησιάζουν έντονα το μηδέν ανήκουν στις μεταβλητές Goat (X_4), GVA for Manufacture of Chemicals and Chemical Products (X_6), Gross Inland Consumption of Petroleum Products (X_{10}) και Gross Inland Consumption of Crude Oil (X_{11}).

Συνοπτικά, με την εφαρμογή καθεμιάς μεθόδου καταλήγουμε στα πιο κάτω αποτελέσματα:

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
AIC	x		x	x		x	x	x		x	x
BIC		x	x	x	x			x	x		x
R.R				x		x	x	x		x	x
L				x		x		x		x	x

Οι μεταβλητές Goat (X_4), GVA for Manufacture of Chemicals and Chemical Products (X_6), Gross Inland Consumption of Lignite (X_8), Gross Inland Consumption of Petroleum Products (X_{10}) και Gross Inland Consumption of Crude Oil (X_{11}) κρίνονται οι σημαντικότερες με βάση τη Lasso όμως παρουσιάζονται επίσης σημαντικές και με βάση το κριτήριο AIC και τη Ridge Regression. Επομένως, επιλέγω αυτές τις μεταβλητές για το μοντέλο μου.

- Ανάλυση παλινδρόμησης του βέλτιστου μοντέλου που επιλέχτηκε

Call:

```
lm(formula = x0 ~ x4 + x6 + x8 + x10 + x11)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1767.8	-524.0	-6.6	554.2	1416.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.162e+03	1.279e+04	-0.638	0.5352
x4	-2.623e-02	1.124e-02	-2.334	0.0378 *
x6	1.710e-01	3.237e-01	0.528	0.6071
x8	5.400e+01	2.617e+01	2.064	0.0614 .
x10	5.454e-02	4.085e-01	0.134	0.8960
x11	4.072e-01	1.345e-01	3.028	0.0105 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1070 on 12 degrees of freedom

Multiple R-squared: 0.9474, Adjusted R-squared: 0.9255

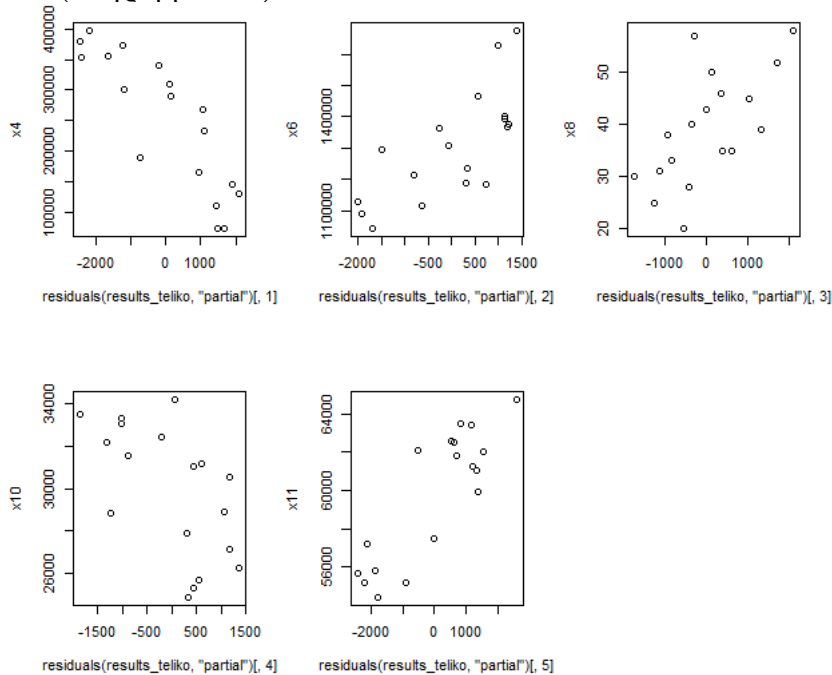
F-statistic: 43.23 on 5 and 12 DF, p-value: 2.897e-07

- 95% Διαστήματα Εμπιστοσύνης

	2.5 %	97.5 %
(Intercept)	-3.602155e+04	1.969703e+04
x4	-5.071974e-02	-1.742197e-03
x6	-5.344119e-01	8.763318e-01
x8	-3.013591e+00	1.110166e+02
x10	-8.355618e-01	9.446386e-01
x11	1.141778e-01	7.001273e-01

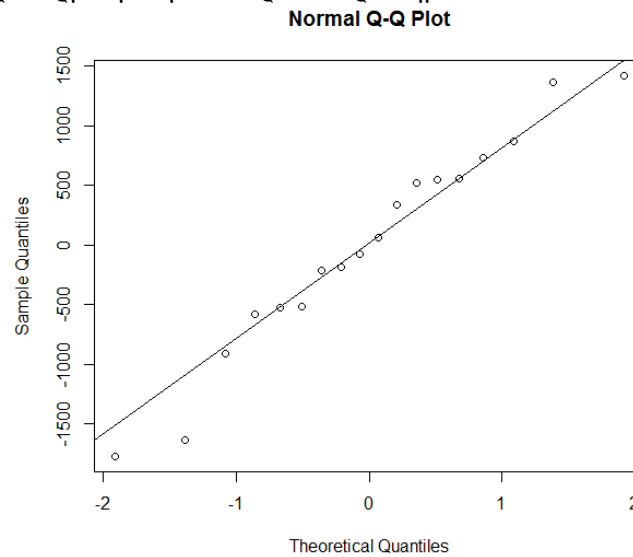
- Τέλος, ελέγχουμε και τις προϋποθέσεις του γενικού γραμμικού μοντέλου

- Η υπόθεση της **γραμμικότητας** είναι λογική με βάση τις γραφικές παραστάσεις των επεξηγηματικών μεταβλητών ως προς τα μερικά υπόλοιπα (Διάγραμμα 4.21).



Διάγραμμα 4.21. Έλεγχος γραμμικότητας (N₂O – Ολλανδία)

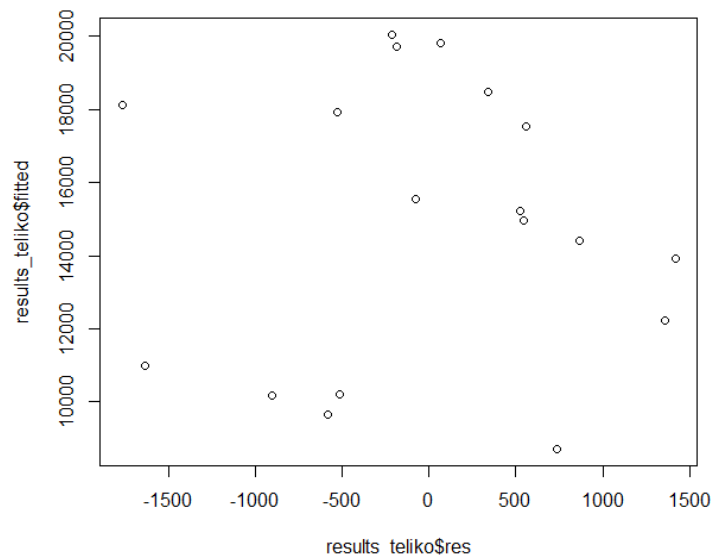
- Η κανονικότητα των υπολοίπων, όπως φαίνεται πολύ καθαρά από το Διάγραμμα 4.22 είναι λογική και η ευθεία που σχηματίζεται είναι προσαρμοσμένη στα περισσότερα σημεία.



Διάγραμμα 4.22. Έλεγχος κανονικότητας των υπολοίπων (N₂O - Ολλανδία)

Jarque-Bera Normality Test: JB = 0.5411, p-value = 0.763 > 0.05 άρα δεν απορρίπτουμε την υπόθεση κανονικότητας.

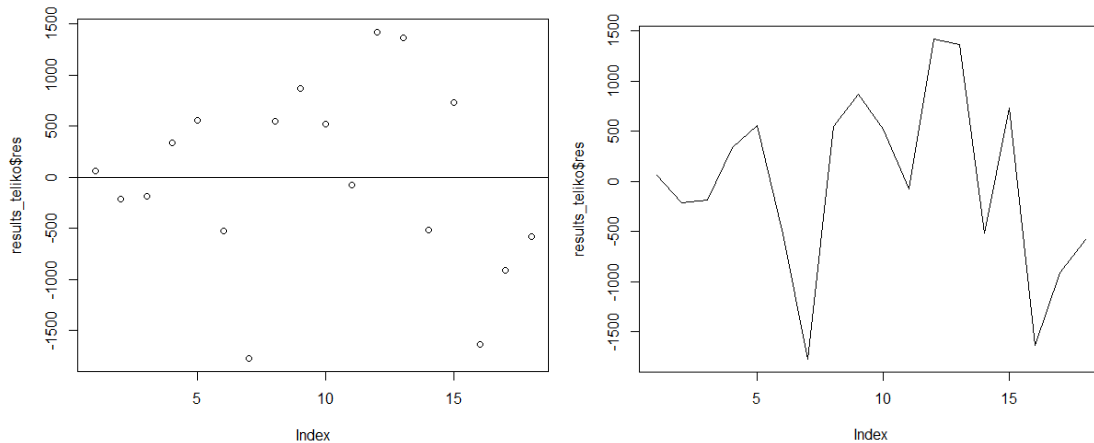
- Με το Διάγραμμα 4.23 παρουσιάζονται τα υπόλοιπα (residuals (res)) (άξονας x) συναρτήσει των προβλεπόμενων τιμών (fitted) (άξονας y). Παρατηρούμε τα ζεύγη των τιμών να μην εμφανίζουν κάποιο συστηματικό τρόπο συμπεριφοράς, οπότε η υπόθεση της **ομοσκεδαστικότητας** είναι λογική.



Διάγραμμα 4.23. Έλεγχος ομοσκεδαστικότητας (N₂O - Ολλανδία)

Breusch-Pagan test: BP = 5.5759, df = 5, p-value = 0.3497 > 0.05 άρα δεν απορρίπτουμε την υπόθεση της ομοσκεδαστικότητας.

- Στο Διάγραμμα 4.24 βλέπουμε ότι υπάρχει **ανεξαρτησία μεταξύ των υπολοίπων**, γιατί δεν παρουσιάζεται κάποια σχέση μεταξύ τους και τα υπόλοιπα συμπεριφέρονται τυχαία.



Διάγραμμα 4.24. Έλεγχος ανεξαρτησίας μεταξύ των υπολοίπων (N₂O - Ολλανδία) (αριστερά: σημειακή απεικόνιση, δεξιά: συνεχής απεικόνιση)

Durbin-Watson test: DW = 1.6463, p-value = 0.05209 > 0.05 άρα δεν απορρίπτουμε την υπόθεση ανεξαρτησίας των υπολοίπων.

Καταλήγουμε στο εξής μοντέλο:

$$y = -8162 - 0.026 * Goat + 0.17 * GVAchemicals + 5.4 * CLignite + 0.055 * CPetroleumProducts + 0.41 * CCrudeOil$$

Αντικαθιστώντας τα στοιχεία του 2013 στις επεξηγηματικές μεταβλητές, **το αποτέλεσμα της πρόβλεψης των εκπομπών υποξειδίου του αζώτου είναι 7519.389 Gg**. Το 95% διάστημα εμπιστοσύνης για την ατομική πρόβλεψη είναι (4254.291, 10784.49).

```
>predict(results_teliko,list(x4<-412550,x6<-9822.45,x8<-31,x10<-32013,
x11<-52570),int="p");
      fit      lwr      upr
7519.389 4254.291 10784.49
```

5. Πρόβλεψη

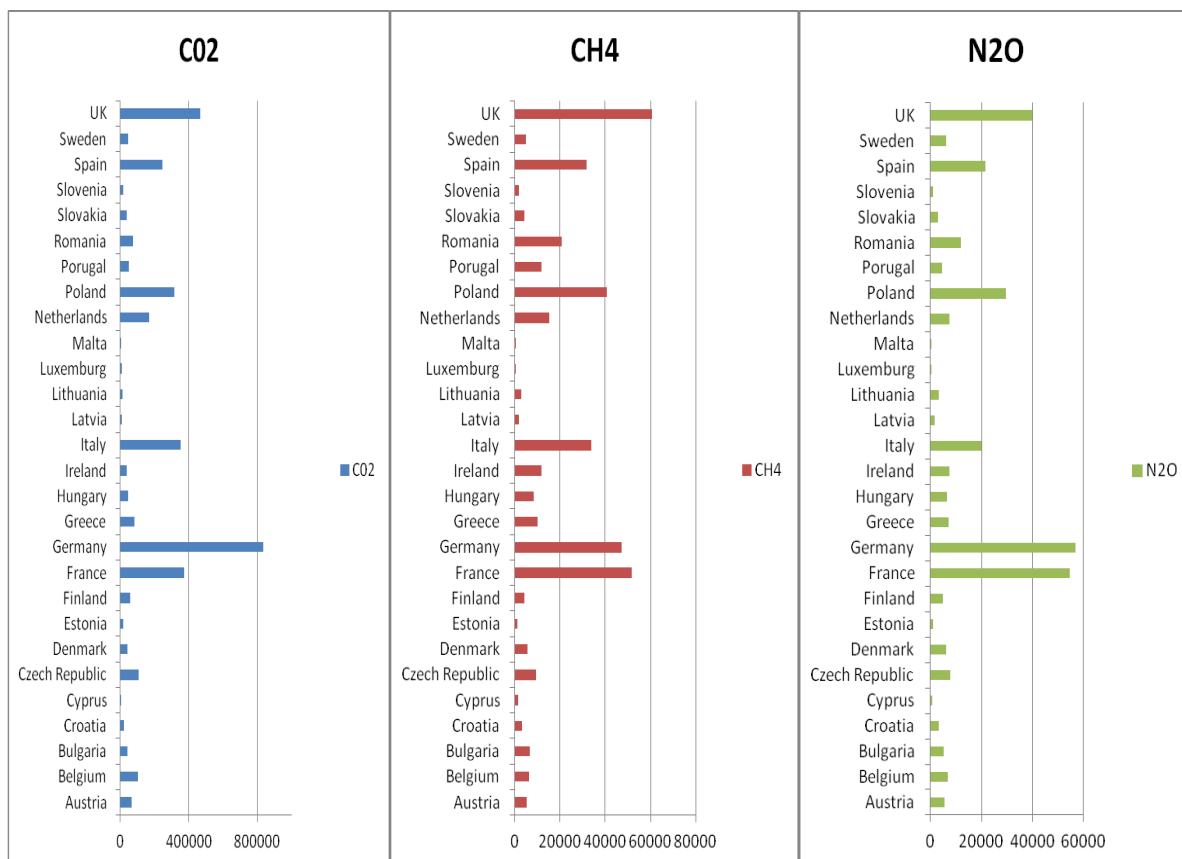
Αφού έχουμε καταλήξει στα 84 βέλτιστα μοντέλα που ικανοποιούν όλες τις προϋποθέσεις του πολλαπλού γραμμικού μοντέλου (Παράρτημα Ι), μπορούμε να αντικαταστήσουμε τις τιμές των επεξηγηματικών μεταβλητών με βάση τα δεδομένα που έχουμε για το 2013, ώστε να παράγουμε την πρόβλεψη για τις εκπομπές του 2013. Οι πίνακες για τα δεδομένα του 2013 βρίσκονται στο τέλος του Παραρτήματος ΙΙ. Αντικαθιστώντας τις τιμές των επεξηγηματικών μεταβλητών σε όλα τα μοντέλα που επιλέξαμε έχουμε τις προβλέψεις για κάθε αέριο του θερμοκηπίου και για κάθε χώρα (Πίνακας 5.1). Έτσι προσθέτοντας αυτές τις τιμές των προβλέψεων έχουμε τις προβλέψεις για τους τρεις ρύπους στην ΕΕ-28 για το 2013 (Πίνακας 5.2.).

	C02	CH4	N2O
Austria	66666,71	5118,085	5405,701
Belgium	104596,3	6354,873	6725,397
Bulgaria	43253,95	6530,132	5255,775
Croatia	19375,81	3137,2	3258,816
Cyprus	6377,979	1333,994	569,7256
Czech Republic	106662,8	9363,248	7655,302
Denmark	42650,24	5593,904	6281,747
Estonia	16271,95	942,2349	1016,414
Finland	56214,29	4064,906	4994,96
France	371736,7	51576,5	54679,3
Germany	833945,5	47383,45	56844,18
Greece	81859,39	9935,592	7157,274
Hungary	43562,94	8138,115	6469,095
Ireland	38101,27	11816,33	7541,877
Italy	354334,6	33755,43	20073,49
Latvia	7458,589	1590,826	1741,949
Lithuania	12842,81	2808,561	3176,152
Luxemburg	10203,38	430,2881	472,5265
Malta	2416,887	132,2412	95,70572
Netherlands	167300,1	15225,84	7519,389
Poland	317222,5	40675,72	29646,79
Porugal	49372,67	11731,98	4474,761
Romania	72806,97	20859,57	11805,64
Slovakia	37585,13	4152,109	3021,763
Slovenia	15255,68	1795,511	1085,624
Spain	244912,6	31857,01	21692,15
Sweden	46850,17	4880,167	5979,274
UK	467044,9	60632,78	39929,01

Πίνακας 5.1: Προβλέψεις εκπομπών των αερίων του θερμοκηπίου κάθε χώρας για το 2013

	C02	CH4	N2O
EU-28	3636883	401816,6	324569,8

Πίνακας 5.2. Πρόβλεψη εκπομπών των αερίων του θερμοκηπίου της ΕΕ-28 για το 2013



Διάγραμμα 5.1. Γραφικές παραστάσεις των προβλέψεων των αερίων του θερμοκηπίου για τις χώρες της ΕΕ-28 (βάσει της πολλαπλής γραμμικής παλινδρόμησης)

Παρατηρώντας το Διάγραμμα 5.1 αλλά και τον Πίνακα 5.1 βλέπουμε ότι το Ηνωμένο Βασίλειο, η Ισπανία, η Πολωνία, η Ιταλία, η Γερμανία και η Γαλλία κατέχουν τις υψηλότερες θέσεις στις εκπομπές αερίων του θερμοκηπίου. Οι παραπάνω χώρες είναι και οι πιο πολυπληθείς στην ΕΕ-28, οπότε δικαιολογούνται σε κάποιο βαθμό οι υψηλές ποσότητες αερίων που εκπέμπουν στην ατμόσφαιρα. Η Γερμανία βρίσκεται πρώτη σε εκπομπές διοξειδίου του άνθρακα και με μεγάλη διαφορά σε σχέση με τη Γαλλία που είναι στη δεύτερη θέση. Το Ηνωμένο Βασίλειο κατέχει την πρώτη θέση σε εκπομπές μεθανίου ενώ όσον αφορά τις εκπομπές υποξειδίου του αζώτου, και πάλι, η Γερμανία κατέχει την πρώτη θέση, με μικρή διαφορά από τη Γαλλία.

6. Ανάλυση Παλινδρόμησης με διαστρωματικά στοιχεία χρονολογικών σειρών (Panel Data)

Τα διαστρωματικά δεδομένα χρονολογικών σειρών, τα λεγόμενα δεδομένα Panel Data, είναι ένα σύνολο δεδομένων, στο οποίο η συμπεριφορά των οντοτήτων παρατηρείται σε διάφορες χρονικές περιόδους (έτη). Με τα δεδομένα αυτά μπορούμε να ελέγξουμε τους παράγοντες που μεταβάλλονται μεταξύ των στρωμάτων, αλλά δε μεταβάλλονται διαχρονικά, καθώς και τους παράγοντες που είναι μη μετρήσιμοι, όπως πολιτισμικοί παράγοντες, και έτσι δεν μπορούμε να τους περιλάβουμε στην παλινδρόμηση χρησιμοποιώντας την πολλαπλή παλινδρόμηση.

Η χρήση της μεθοδολογίας των panel data έχει πολλά πλεονεκτήματα σε σχέση με την αντίστοιχη της πολλαπλής ανάλυσης παλινδρόμησης. Η ύπαρξη panel data υποδηλώνει από μόνη της ότι οι χώρες έχουν διαφορετικά χαρακτηριστικά και διατηρεί την ατομική ανομοιογένεια (individual heterogeneity). Η χρήση απλών χρονολογικών σειρών αγνοεί αυτή τη διαφορετικότητα μεταξύ τους γεγονός που μπορεί να οδηγήσει σε μεροληπτικές εκτιμήσεις. Επίσης, χρησιμοποιώντας τα panel data έχουμε λιγότερη συγγραμμικότητα μεταξύ των επεξηγηματικών μεταβλητών, περισσότερους βαθμούς ελευθερίας και περισσότερη αποτελεσματικότητα στις εκτιμήσεις. Τα μοντέλα που βασίζονται αποκλειστικά σε στοιχεία χρονολογικών σειρών υποφέρουν σημαντικά από την ύπαρξη έντονης πολυσυγγραμμικότητας (multicollinearity).

Στα πλαίσια αυτής της διπλωματικής για την ανάλυση των panel data εξέτασα δύο τεχνικές μεθόδους ώστε να επιλέξω για κάθε σύνολο δεδομένων την πιο αποτελεσματική. Οι δύο αυτές τεχνικές είναι η παλινδρόμηση Σταθερών Επιδράσεων (Fixed Effects) και η παλινδρόμηση Τυχαίων Επιδράσεων (Random Effects).

6.1. Μοντέλα Σταθερών Επιδράσεων (Fixed-Effects Models)

Η παλινδρόμηση Σταθερών Επιδράσεων (FE) χρησιμοποιείται κυρίως όταν ενδιαφερόμαστε μόνο για την ανάλυση των επιδράσεων των επεξηγηματικών μεταβλητών που μεταβάλλονται με την πάροδο του χρόνου. Η FE διερευνά τη σχέση μεταξύ της μεταβλητής απόκρισης και των επεξηγηματικών μεταβλητών μέσα σε μια οντότητα (χώρα, πρόσωπο, εταιρεία κ.λπ.). Κάθε οντότητα έχει τα δικά της μοναδικά χαρακτηριστικά που μπορούν να επηρεάσουν ή όχι τις επεξηγηματικές μεταβλητές.

Όταν χρησιμοποιείται η FE υποθέτουμε ότι κάτι μέσα στην οντότητα μπορεί να επηρεάσει ή να προαλέσει μεροληψία στους συντελεστές των επεξηγηματικών μεταβλητών και πρέπει να το ελέγξουμε. Η FE αφαιρεί την επίδραση αυτών των χρονικά αμετάβλητων χαρακτηριστικών ώστε να μπορεί να εκτιμηθεί η καθαρή επίδραση των συντελεστών πρόβλεψης στη μεταβλητή απόκρισης.

Μια άλλη σημαντική υπόθεση του μοντέλου FE είναι ότι τα εν λόγω χρονικά αναλλοίωτα χαρακτηριστικά είναι μοναδικά για κάθε οντότητα και δεν θα πρέπει να συσχετιστούν με άλλα ατομικά χαρακτηριστικά. Κάθε οντότητα είναι διαφορετική, ως εκ τούτου ο όρος σφάλματος και η σταθερά μίας οντότητας (η οποία καταγράφει τα ατομικά χαρακτηριστικά) δεν θα πρέπει να συσχετίζονται με των άλλων οντοτήτων. Εάν οι όροι σφάλματος συσχετίζονται, τότε η FE δεν είναι κατάλληλη, δεδομένου ότι τα συμπεράσματα μπορεί να μην είναι σωστά και το μοντέλο θα πρέπει να διαμορφωθεί διαφορετικά, πιθανώς

με τη χρήση των τυχαίων επιδράσεων. Αυτό είναι το κύριο σκεπτικό για τον έλεγχο Hausman (που παρουσιάζεται στο Κεφάλαιο 7.3). Εν κατακλείδι, υποθέτουμε ότι η διακύμανση μεταξύ των οντοτήτων είναι σταθερή και για αυτό μας ενδιαφέρουν κυρίως οι επιδράσεις μέσα σε κάθε οντότητα.

Τα μοντέλα σταθερών επιδράσεων έχουν την πιο κάτω μορφή:

$$Y_{it} = \alpha + \beta_1 X_{1,it} + \dots + \beta_k X_{k,it} + \gamma_2 E_2 + \dots + \gamma_n E_n + u_{it}$$

όπου: γ_m είναι η άγνωστη σταθερά της οντότητας E_m , $m=2, \dots, n$.

E_m είναι η m -οντότητα (δεδομένου ότι είναι δυαδική, θα πρέπει $n-1$ οντότητες να περιλαμβάνονται στο μοντέλο).

Y_{it} είναι η μεταβλητή απόκρισης,

όπου $i =$ οντότητα ($i=1, \dots, n$) και $t =$ χρόνος ($t=1, \dots, T$).

$X_{j,it}$ είναι η j -επεξηγηματική μεταβλητή, $i=1, \dots, n$, $t=1, \dots, T$.

β_j είναι ο συντελεστής της επεξηγηματικής μεταβλητής $X_{j,it}$.

α είναι η σταθερά.

u_{it} είναι ο όρος σφάλματος, που ακολουθεί την κανονική κατανομή $N(0, \sigma_u^2)$, $i=1, \dots, n$, $t=1, \dots, T$. Η διασπορά σ_u^2 είναι άγνωστη και εκτιμάται μέσα από τη σχέση:

$$\hat{\sigma}_u^2 = \frac{1}{\underbrace{nT - k - n}_{\text{Βαθμοί ελευθερίας}}} \underbrace{\sum_{i=1}^n \sum_{t=1}^T \hat{u}_{it}^2}_{\substack{\text{SSR}(\text{άθροισμα} \\ \text{τετραγώνων} \\ \text{λόγω} \\ \text{παλινδρόμησης)}}$$

όπου n ο αριθμός των οντοτήτων, T η χρονική περίοδος και k ο αριθμός των παραμέτρων (εκτός της σταθεράς).

Στην περίπτωση αυτή η μεταβλητή Country (χώρα) αντιπροσωπεύει τις οντότητες και η μεταβλητή Year (έτος) αντιπροσωπεύει τη μεταβλητή του χρόνου (t).

6.2. Μοντέλα Τυχαίων Επιδράσεων (Random-Effects Models)

Στο μοντέλο Τυχαίων Επιδράσεων (RE), σε αντίθεση με το μοντέλο σταθερών επιδράσεων, η διαφορά μεταξύ των οντοτήτων θεωρείται ότι είναι τυχαία και ασυσχέτιστη με τις επεξηγηματικές μεταβλητές που περιλαμβάνονται στο μοντέλο.

Το μοντέλο τυχαίων επιδράσεων έχει τη μορφή:

$$Y_{it} = \beta_1 X_{1,it} + \dots + \beta_k X_{k,it} + \alpha + u_{it} + \varepsilon_{it}$$

Όπου: α είναι η σταθερά.

Y_{it} είναι η μεταβλητή απόκρισης,

όπου $i =$ οντότητα ($i=1, \dots, n$) και $t =$ χρόνος ($t=1, \dots, T$).

$X_{j,it}$ είναι η j -επεξηγηματική μεταβλητή, $i=1, \dots, n$, $t=1, \dots, T$.

β_j είναι ο συντελεστής της επεξηγηματικής μεταβλητής $X_{j,it}$.

u_{it} είναι ο όρος σφάλματος εντός των οντοτήτων, που ακολουθεί την κανονική κατανομή $N(0, \sigma_u^2)$, $i=1, \dots, n$, $t=1, \dots, T$.

ε_{it} είναι ο όρος σφάλματος μεταξύ των οντοτήτων, που ακολουθεί την κανονική κατανομή $N(0, \sigma_\varepsilon^2)$, $i=1, \dots, n$, $t=1, \dots, T$.

Συμβολίζουμε το άθροισμα των όρων σφάλματος:

$$v_{it} = u_{it} + \varepsilon_{it}.$$

Οι διασπορές σ_u^2 και σ_ε^2 είναι άγνωστες και εκτιμώνται μέσα από τις σχέσεις:

$$\hat{\sigma}_v^2 = \frac{1}{nT - (k+1)} \sum_{i=1}^n \sum_{t=1}^T \hat{v}_{it}^2$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{nT(T-1)/2 - (k+1)} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_\varepsilon^2$$

όπου n ο αριθμός των οντοτήτων, T η χρονική περίοδος και k ο αριθμός των παραμέτρων (εκτός της σταθεράς).

Η κύρια διάκριση μεταξύ σταθερών και τυχαίων επιδράσεων εστιάζεται στο αν οι επιδράσεις των μη μετρήσιμων παραγόντων ενσωματώνουν στοιχεία, τα οποία συσχετίζονται με τις επεξηγηματικές μεταβλητές στο μοντέλο. Αν πιστεύουμε ότι οι διαφορές μεταξύ των οντοτήτων έχουν κάποια επίδραση στις μεταβλητές απόκρισης, τότε θα πρέπει να χρησιμοποιήσουμε τις τυχαίες επιδράσεις.

Στο μοντέλο RE υποθέτουμε ότι όρος του σφάλματος της οντότητας δεν συσχετίζεται με τις επεξηγηματικές μεταβλητές, κάτι που επιτρέπει στις χρονικά αμετάβλητες μεταβλητές να παίξουν ένα ρόλο ως επεξηγηματικές μεταβλητές. Σε αυτά τα μοντέλα θα πρέπει να συγκεκριμενοποιήσουμε τα ατομικά χαρακτηριστικά που μπορούν ή όχι να επηρεάσουν τις επεξηγηματικές μεταβλητές. Το πρόβλημα με αυτό είναι ότι ορισμένες μεταβλητές μπορεί να μην είναι διαθέσιμες, ως εκ τούτου οδηγούμαστε στην ύπαρξη μεροληψίας στο μοντέλο.

Όταν χρησιμοποιούμε το μοντέλο FE ενδιαφερόμαστε μόνο για την εξαγωγή αποτελεσμάτων που αφορούν το συγκεκριμένο δείγμα, ενώ με το μοντέλο RE μπορούμε να εξάγουμε πληροφορίες για ολόκληρο τον πληθυσμό από όπου προήλθε το δείγμα. Εάν το μέγεθος της επίδρασης σε σχέση με την διασπορά μεταξύ των οντοτήτων είναι μεγάλο τότε επιλέγουμε την ανάλυση παλινδρόμησης των σταθερών επιδράσεων (FE).

7. Μεθοδολογία (Panel Data)

Χρησιμοποιώντας τα panel data, αντιμετωπίζουμε τα δεδομένα που έχουμε για κάθε χώρα ως ένα σύνολο δεδομένων, δηλαδή στην ανάλυση ενός αερίου του θερμοκηπίου αντί για 28 διαφορετικούς πίνακες έχουμε έναν ενιαίο. Δημιουργώντας, όμως, έναν ενιαίο πίνακα για κάθε αέριο παρατηρούμε ότι δεν έχουν όλες οι χώρες τις ίδιες επεξηγηματικές μεταβλητές. Για παράδειγμα, για τις εκπομπές μεθανίου, έχουμε ως επεξηγηματική μεταβλητή την πρωτογενή παραγωγή λιθάνθρακα, λιγνίτη και φυσικού αερίου. Η μεταβλητή αυτή για κάποιες χώρες, όπως η Κύπρος, έχει μόνο μηδενικές τιμές, συνεπώς είναι προτιμότερο να διαχωρίσω τον πίνακα αυτό με βάση τις μεταβλητές για τις οποίες διαθέτει ή όχι δεδομένα η κάθε χώρα. Ακόμα κι έτσι, η ανάλυση είναι λιγότερο χρονοβόρα σε σχέση με την πολλαπλή ανάλυση παλινδρόμησης.

Για την ανάλυση με χρήση των panel data ακολουθούμε την πιο κάτω διαδικασία:

1. Κατανομή χωρών σε πίνακες.
2. Έλεγχος πολυσυγγραμμικότητας.
3. Επιλογή καλύτερης μεθόδου (OLS, FE, RE).
4. Επιλογή καταλληλότερων επεξηγηματικών μεταβλητών.
5. Διαγνωστικοί Έλεγχοι.
6. Πρόβλεψη.

7.1. Κατανομή χωρών σε πίνακες

Για την ανάλυση των εκπομπών διοξειδίου του άνθρακα δημιουργήθηκαν δύο πίνακες:

- i. Ο πρώτος πίνακας περιλαμβάνει τις χώρες που δεν παράγουν λιγνίτη.
- ii. Ο δεύτερος πίνακας περιλαμβάνει τις υπόλοιπες χώρες, οι οποίες έχουν δεδομένα για όλες τις επεξηγηματικές μεταβλητές.

Για την ανάλυση των εκπομπών μεθανίου δημιουργήθηκαν τέσσερις πίνακες:

- i. Ο πρώτος πίνακας περιλαμβάνει τις χώρες που δεν παράγουν καύσιμα (λιγνίτη, λιθάνθρακα και φυσικό αέριο).
- ii. Ο δεύτερος πίνακας περιλαμβάνει τις χώρες που δεν παράγουν φυσικό αέριο και λιθάνθρακα, αλλά παράγουν λιγνίτη. (Ο λιγνίτης θεωρείται από τους πιο βρώμικους γαιάνθρακες, και κατ' επέκταση σημαντικά στις εκπομπές μεθανίου. Έτσι θεωρήσα ότι είναι καλύτερα να μην συμπεριλάβω αυτές τις χώρες στον προηγούμενο πίνακα, καθώς θα διαγραφόταν η μεταβλητή που αντιπροσωπεύει την παραγωγή του λιγνίτη.)
- iii. Ο τρίτος πίνακας περιλαμβάνει τις χώρες που δεν παράγουν λιθάνθρακα.
- iv. Ο τέταρτος πίνακας περιλαμβάνει τις υπόλοιπες χώρες, οι οποίες παράγουν όλα τα καύσιμα.

Για την ανάλυση των εκπομπών υποξειδίου του αζώτου δημιουργήθηκαν τρεις πίνακες:

- i. Ο πρώτος περιλαμβάνει τις χώρες για τις οποίες δεν είχαμε καθόλου δεδομένα για την μεταβλητή GVA for Manufacture of Chemicals and Chemical Products.
- ii. Ο δεύτερος πίνακας περιλαμβάνει τις χώρες οι οποίες δεν παράγουν λιγνίτη.
- iii. Ο τρίτος πίνακας περιλαμβάνει τις υπόλοιπες χώρες όπου έχουμε πλήρη δεδομένα.

Αξιίζει να σημειωθεί ότι για την πρόβλεψη των εκπομπών μεθανίου της Κροατίας επιλέχθηκε το μοντέλο που διαμορφώθηκε σύμφωνα με την πρώτη μέθοδο (της ανάλυσης παλινδρόμησης πολλαπλού γραμμικού μοντέλου), λόγω του ότι ήταν η μοναδική χώρα όπου τα δεδομένα για την επεξηγηματική μεταβλητή “Municipal Waste Landfilled” δεν ήταν διαθέσιμα.

7.2. Έλεγχος Πολυσυγγραμμικότητας

Αρχικά εξετάζω τις επεξηγηματικές μεταβλητές μου ως προς την συσχέτιση μεταξύ τους. Ο έλεγχος γίνεται μέσω του πακέτου `corrplot` της R και των τιμών VIF (Κεφάλαιο 3.3.1.) των επεξηγηματικών μεταβλητών. Αν υπάρχει έντονη συσχέτιση μεταξύ κάποιων επεξηγηματικών μεταβλητών επιλέγω να κρατήσω μία από αυτές ώστε να διορθωθεί το πρόβλημα πολυσυγγραμμικότητας. Μετά από αυτό το βήμα θα παραμείνουν στο μοντέλο οι επεξηγηματικές μεταβλητές που δεν έχουν ισχυρή συσχέτιση μεταξύ τους.

7.3. Επιλογή Καλύτερης Μεθόδου

Στη συνέχεια χρησιμοποιώ την μέθοδο ελαχίστων τετραγώνων (OLS). Η μέθοδος αυτή με βοηθά να αποκτήσω μία αρχική εικόνα των δεδομένων μου και της σημαντικότητας των επεξηγηματικών μεταβλητών που υπάρχουν στο μοντέλο. Η μέθοδος των ελαχίστων τετραγώνων όπως αναφέρθηκε νωρίτερα αγνοεί την ανομοιογένεια (heterogeneity) των οντοτήτων ή του χρόνου.

Οι σταθερές επιδράσεις κάθε χώρας μπορούν να μοντελοποιηθούν χρησιμοποιώντας ψευδομεταβλητές. Με αυτόν τον τρόπο τα χαρακτηριστικά των χωρών αντικατοπτρίζονται επαρκώς από τον καθορισμό διαφορετικών σταθερών συντελεστών για κάθε χώρα. Ως εκ τούτου, μια κοινή διαμόρφωση αυτού του μοντέλου, το οποίο είναι γνωστό ως μοντέλο ελαχίστων τετραγώνων με ψευδομεταβλητές (Least square dummy variable model (LSDV)) είναι η εξής:

$$Y_{it} = \sum_{i=1}^n \gamma_i E_i + \sum_{j=1}^k \beta_j X_{j,it} + u_{it}$$

Όπου: E_i είναι η i -οντότητα, $i=1, \dots, N$.

γ_i είναι η άγνωστη σταθερά για κάθε οντότητα D_i , $i=1, \dots, N$.

Y_{it} είναι η μεταβλητή απόκρισης,

όπου i = οντότητα ($i=1, \dots, N$) και t = χρόνος ($t=1, \dots, T$).

$X_{j,it}$ είναι η j -επεξηγηματική μεταβλητή, $j=1, \dots, k$, $i=1, \dots, n$, $t=1, \dots, T$.

β_j είναι ο συντελεστής της επεξηγηματικής μεταβλητής $X_{j,it}$.

u_{it} είναι ο όρος σφάλματος, που ακολουθεί την κανονική κατανομή $N(0, \sigma_u^2)$, $i=1, \dots, n$, $t=1, \dots, T$.

Το μοντέλο ελαχίστων τετραγώνων με ψευδομεταβλητές παρέχει έναν καλό τρόπο για να καταλάβουμε τις σταθερές επιδράσεις. Στο μοντέλο OLS, η επίδραση της j -επεξηγηματικής μεταβλητής $X_{j,it}$ εμπεριέχει τις διαφορές μεταξύ των χωρών, ενώ στο μοντέλο LSDV εκτιμάται η καθαρή επίδραση της μεταβλητής αυτής.

Η R διαθέτει έναν έλεγχο το `pFtest` με τον οποίο μπορούμε να αποφασίσουμε αν η καταλληλότερη μέθοδος είναι η μέθοδος των ελαχίστων τετραγώνων ή αν είναι προτιμότερο να προσθέσουμε ψευδομεταβλητές. Ο έλεγχος αυτός εμφανίζει μια p -τιμή, η οποία αν είναι μικρότερη από 0.05 τότε το μοντέλο σταθερών επιδράσεων είναι μια πιο καλή επιλογή.

Στη συνέχεια μπορούμε να εκτιμήσουμε το μοντέλο τυχαίων επιδράσεων, ώστε να το συγκρίνουμε με το μοντέλο σταθερών επιδράσεων και να καταλήξουμε στην επιλογή του καταλληλότερου. Η ερμηνεία των συντελεστών σε αυτό το μοντέλο είναι δύσκολη, δεδομένου ότι περιλαμβάνει τόσο τις επιδράσεις εντός των οντοτήτων αλλά και τις επιδράσεις μεταξύ των οντοτήτων. Σε αυτή την περίπτωση το συνολικό άθροισμα των

τετραγώνων (Total Sum of Squares) αντιπροσωπεύει τη μέση επίδραση του \mathbf{X} πάνω στο \mathbf{Y} όταν αλλάζει το \mathbf{X} στο πέρασμα του χρόνου και μεταξύ των χωρών κατά μία μονάδα.

Για να αποφασίσουμε μεταξύ των σταθερών ή τυχαία επιδράσεων μπορούμε να εκτελέσουμε το Hausman Test όπου η μηδενική υπόθεση είναι ότι το προτιμότερο μοντέλο είναι των τυχαίων επιδράσεων έναντι της εναλλακτικής των σταθερών επιδράσεων. Ελέγχει βασικά αν τα σφάλματα συσχετίζονται με τις επεξηγηματικές μεταβλητές, η μηδενική υπόθεση είναι ότι δεν είναι συσχετισμένα. Εάν η p -τιμή ελέγχου είναι μικρότερη από 0.05, τότε χρησιμοποιούμε τις σταθερές επιδράσεις.

Ένας εναλλακτικός έλεγχος που μπορούμε να χρησιμοποιήσουμε είναι το Breusch-Pagan Lagrange multiplier (LM) test. Το τεστ LM μας βοηθά να αποφασίσουμε μεταξύ μιας παλινδρόμησης τυχαίων επιδράσεων και μιας απλής OLS παλινδρόμησης. Η μηδενική υπόθεση στο τεστ LM είναι ότι οι διακυμάνσεις των οντοτήτων είναι μηδέν, δηλαδή δεν υπάρχει καμία σημαντική μεταξύ των χωρών. Αν η p -τιμή είναι μεγαλύτερη από 0.05 τότε δεν απορρίπτουμε την μηδενική υπόθεση και καταλήγουμε στο συμπέρασμα ότι το μοντέλο των τυχαίων επιδράσεων δεν είναι κατάλληλο. Σε τέτοια περίπτωση, σημαίνει ότι δεν υπάρχει ένδειξη για σημαντικές διαφορές από χώρα σε χώρα, για αυτό μπορεί να χρησιμοποιηθεί η απλή OLS παλινδρόμηση.

7.4. Επιλογή Καταλληλότερων Επεξηγηματικών Μεταβλητών

Αφού καταλήξουμε στην καταλληλότερη μέθοδο μεταξύ των σταθερών και τυχαίων επιδράσεων πρέπει να επιλέξουμε ένα υποσύνολο μεταβλητών, το οποίο θα μας δώσει τις καλύτερες εκτιμήσεις. Η επιλογή γίνεται με βάση τα κριτήρια AIC (Κεφάλαιο 3.3.2) και BIC (Κεφάλαιο 3.3.3). Εφαρμόζοντας τον κώδικα που προαναφέρθηκε στο Κεφάλαιο 3.3.4 καταλήγουμε στο βέλτιστο μοντέλο.

7.5. Διαγνωστικοί Έλεγχοι

Αφού έχουμε καταλήξει σε ένα βέλτιστο μοντέλο, προχωράμε στον έλεγχο της ομοσκεδαστικότητας. Για να εξετάσουμε την ομοσκεδαστικότητα εκτελούμε το Breusch-Pagan test. Η μηδενική υπόθεση είναι η ύπαρξη ομοσκεδαστικότητας, οπότε p -τιμές μικρότερες από 0.05 υποδεικνύουν ύπαρξη ετεροσκεδαστικότητας. Αν εντοπιστεί ετεροσκεδαστικότητα, μπορούμε να χρησιμοποιήσουμε τον πίνακα συνδιασποράς robust για να αντιμετωπιστεί το πρόβλημα.

Η `vcovHC` συνάρτηση της R υπολογίζει τρεις εκτιμητές συνδιακύμανσης για αντιμετώπιση της ετεροσκεδαστικότητας:

- "white1" – εφαρμόζεται όταν υπάρχει γενική ετεροσκεδαστικότητα, αλλά χωρίς να υπάρχει αυτοσυσχέτιση και συνιστάται για τις τυχαίες επιδράσεις.
- "white2" - είναι ίδιος με τον "white1" περιορισμένος σε μια κοινή διακύμανση μέσα στις ομάδες και συνιστάται για τυχαίες επιδράσεις.
- "arellano" -εφαρμόζεται όταν υπάρχει ετεροσκεδαστικότητα αλλά και αυτοσυσχέτιση και συνιστάται για σταθερές επιδράσεις.

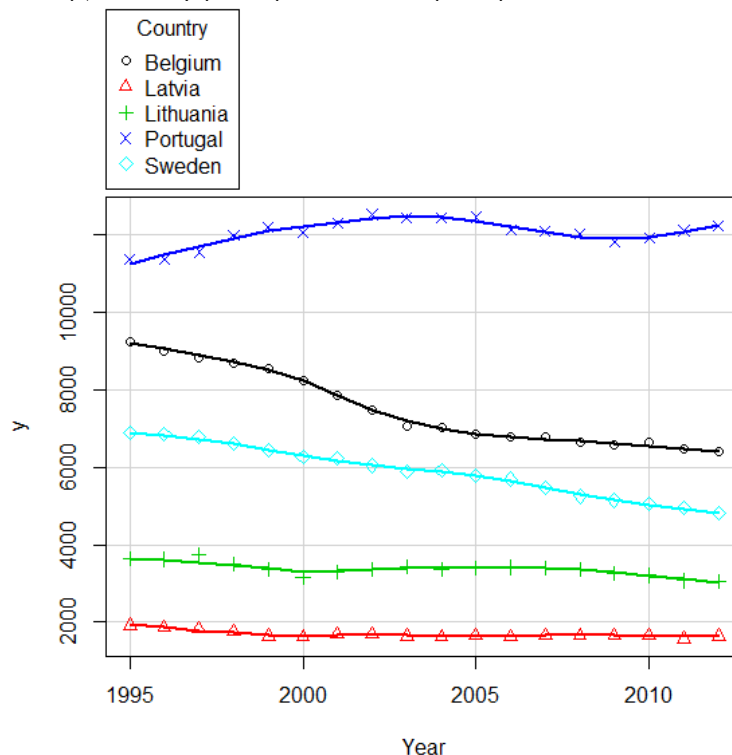
8. Ανάλυση Παλινδρόμησης με Panel Data

Για την πρόβλεψη των εκπομπών των τριών αερίων του θερμοκηπίου κάθε χώρας χρειάστηκε να δημιουργηθούν 10 μοντέλα (2 για τις εκπομπές CO₂, 5 για τις εκπομπές CH₄, 3 για τις εκπομπές N₂O). Για λόγους εξοικονόμησης χώρου, παραθέτω μόνο δύο αναλύσεις (μία με Fixed Effects και μία με Random Effects) μαζί με τις προβλέψεις των μοντέλων τους. Για τα υπόλοιπα ακολουθήθηκε η ίδια μεθοδολογία. Τα 10 μοντέλα που διαμορφώθηκαν βρίσκονται στο Παράρτημα Ι και τα αποτελέσματά των προβλέψεων τους παρουσιάζονται στο Κεφάλαιο 9.

8.1. Ανάλυση Παλινδρόμησης με Panel Data για τις εκπομπές μεθανίου (CH₄) των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου

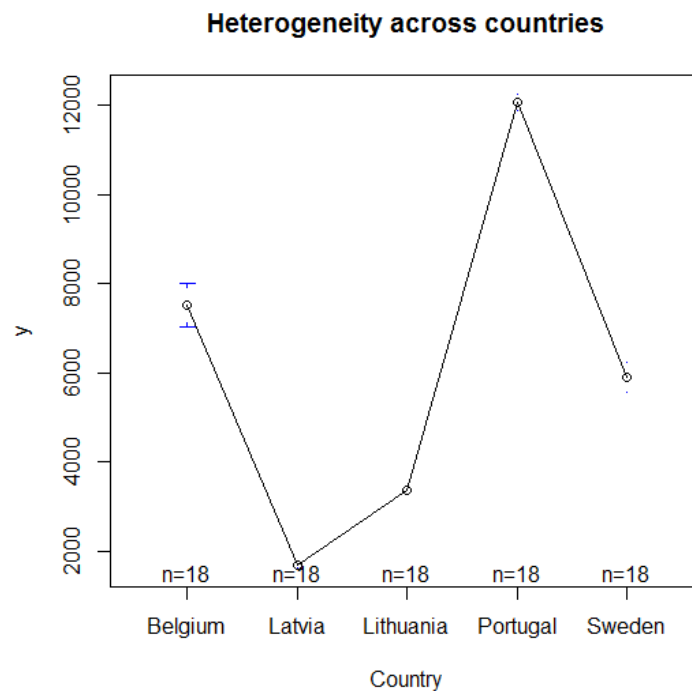
Παρατηρώντας τον πίνακα που δημιουργήσαμε βλέπουμε ότι εκτός από το Βέλγιο, οι υπόλοιπες χώρες εκτός από το φυσικό αέριο, δεν παράγουν ούτε λιθάνθρακα. Το Βέλγιο, παρόλο που έχει τιμές για την μεταβλητή Primary Production of Hard Coal, δεν μας δημιουργεί πρόβλημα να το κατατάξουμε στην ίδια ομάδα, καθώς η συγκεκριμένη μεταβλητή συσχετίζεται σε μεγάλο βαθμό με τις υπόλοιπες, οπότε ακόμη κι αν δεν βγει σημαντική δεν θα επηρεάσει σημαντικά το αποτέλεσμα της συγκεκριμένης χώρας.

Μέσα από αυτό το Διάγραμμα 8.1 μπορούμε να δούμε τη συμπεριφορά των εκπομπών μεθανίου για κάθε χώρα από το 1995 μέχρι το 2012. Παρατηρούμε ότι η Πορτογαλία είναι η μόνη από αυτές τις χώρες που παρουσιάζει άνοδο και επίσης, στο γενικό σύνολο διατηρεί πιο υψηλές τιμές στις εκπομπές μεθανίου.



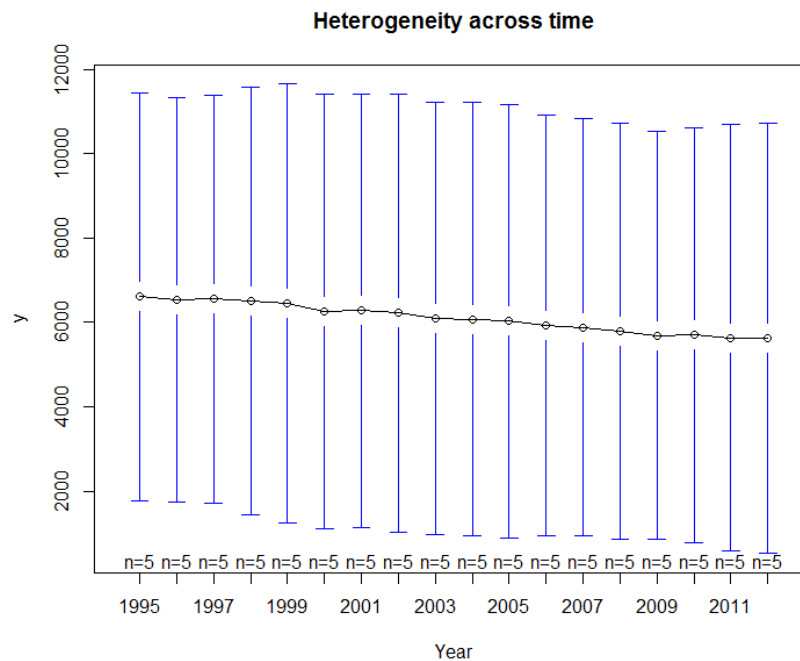
Διάγραμμα 8.1. Συνολικές εκπομπές CH₄ ανά χώρα για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012)

Στο Διάγραμμα 8.2 που ακολουθεί παρατηρούμε την ανομοιογένεια που υπάρχει από χώρα σε χώρα, δηλαδή την ύπαρξη μη παρατηρούμενων μεταβλητών που δεν αλλάζουν με το πέρασμα του χρόνου. Τα σημεία αντιστοιχούν στη μέση τιμή των εκπομπών κάθε χώρας και η μπλε κάθετη γραμμή υποδεικνύει το 95% διάστημα εμπιστοσύνης γύρω από τον μέσο όρο.



Διάγραμμα 8.2. Μέση τιμή εκπομπών CH₄ ανά χώρα για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012)

Στο Διάγραμμα 8.3 παρατηρούμε την ανομοιογένεια με την πάροδο του χρόνου. Μπορούμε να δούμε ότι οι εκπομπές στο σύνολό τους μειώνονται χρόνο με το χρόνο.



Διάγραμμα 8.3. Συνολικές εκπομπές CH₄ ανά έτος για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου (1995-2012)

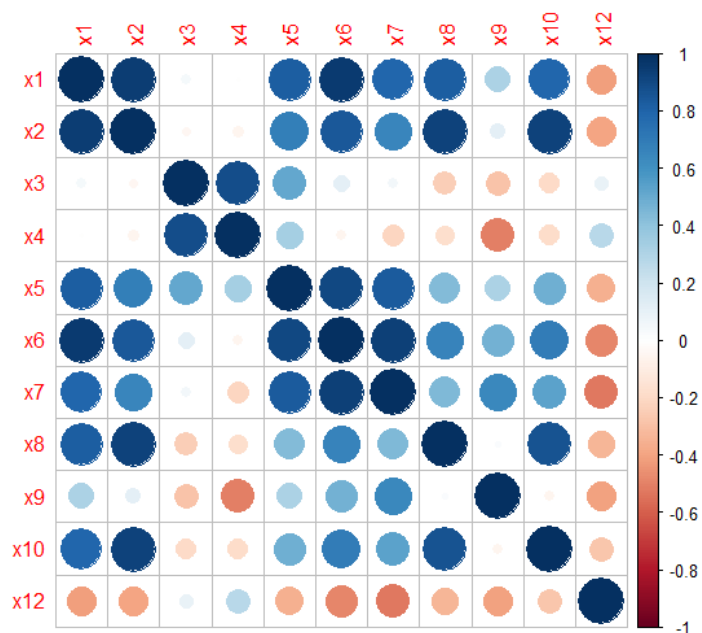
Έλεγχος Πολυσυγγραμμικότητας

Από τον Πίνακα 8.1 αλλά και το Διάγραμμα 8.4 παρατηρούμε ότι υπάρχει έντονη συσχέτιση μεταξύ αρκετών επεξηγηματικών μεταβλητών. Οι μεταβλητές που δημιουργούν το σημαντικότερο πρόβλημα, μπορούν να γίνουν γνωστές κυρίως από τις τιμές VIF (Πίνακας 8.2).

	Cattle	Pig	Sheep	Waste	T.Population	U.Population
Cattle	1	0.94633418	0.04832910	-0.005646217	0.8285158	0.95018140
Pig	0.946334182	1	-0.04110270	-0.059409807	0.6846192	0.84885551
Sheep	0.048329100	-0.04110270	1	0.882678569	0.5186395	0.11957568
Waste	-0.005646217	-0.05940981	0.88267857	1	0.3340386	-0.05290155
T.Population	0.828515804	0.68461919	0.51863948	0.334038562	1	0.90637693
U.Population	0.950181395	0.84885551	0.11957568	-0.052901553	0.9063769	1
GDP	0.799791245	0.65419513	0.05219932	-0.216188825	0.8314353	0.93695226
P.Hard Coal	0.821154298	0.92256997	-0.24420819	-0.177756296	0.4316464	0.66015182
P.Lignite	0.319041455	0.11276779	-0.28355101	-0.509393335	0.3114519	0.47370765
C.Natural Gas	0.795605133	0.92005915	-0.19612740	-0.187631155	0.4886742	0.69875912
D.Losses	-0.417203996	-0.39704854	0.09515758	0.278843988	-0.3547890	-0.48240876

	GDP	P.Hard Coal	P.Lignite	C.Natural Gas	D.Losses
Cattle	0.79979125	0.82115430	0.31904145	0.79560513	-0.41720400
Pig	0.65419513	0.92256997	0.11276779	0.92005915	-0.39704854
Sheep	0.05219932	-0.24420819	-0.28355101	-0.19612740	0.09515758
Waste	-0.21618882	-0.17775630	-0.50939334	-0.18763116	0.27884399
T.Population	0.83143525	0.43164642	0.31145193	0.48867421	-0.35478902
U.Population	0.93695226	0.66015182	0.47370765	0.69875912	-0.48240876
GDP	1	0.44914043	0.64072419	0.53131150	-0.52259613
P.Hard Coal	0.44914043	1	0.02273277	0.86996704	-0.33640389
P.Lignite	0.64072419	0.02273277	1	-0.05119825	-0.40250273
C.Natural Gas	0.53131150	0.86996704	-0.05119825	1	-0.27676409
D.Losses	-0.52259613	-0.33640389	-0.40250273	-0.27676409	1

Πίνακας 8.1. Πίνακας με τις τιμές συσχέτισης των επεξηγηματικών μεταβλητών (CH₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου)



Διάγραμμα 8.4. Διάγραμμα συσχέτισης των εξηγηματικών μεταβλητών (CH₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου)

Variables	VIF
Cattle	160,726
Pig	199,486
Sheep	85,860
Waste	37,911
T.Population	869,208
U.Population	1.252,370
GDP	86,763
P.Hard Coal	18,372
P.Lignite	12,196
C.Natural Gas	29,157
D.Losses	1,890

Πίνακας 8.2. Τιμές VIF των εξηγηματικών μεταβλητών (CH₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου)

Έτσι αφαιρώντας τις Cattle (X_1), Pig (X_2), Sheep (X_3), Total Population (X_5) και Urban Population (X_6) από το μοντέλο αντιμετωπίζουμε το πρόβλημα πολυσυγγραμμικότητας. Αναλυτικότερα, οι μεταβλητές Cattle (X_1), Pig (X_2) έχουν έντονη συσχέτιση με την Primary Production of Hard Coal (X_8), η Sheep (X_3) έχει συσχετίζεται με τη Waste (X_4) και οι Total Population (X_5) και Urban Population (X_6) έχουν ισχυρή συσχέτιση με την GDP (X_7). Η μη ύπαρξη συσχέτισης μεταξύ των μεταβλητών που έμειναν στο μοντέλο είναι εμφανή από τις τιμές των VIF στον Πίνακα 8.3.

Variables	VIF
Waste	2,2009
GDP	6,3842
P.Hard Coal	4,9627
P.Lignite	5,7389
C.Natural Gas	9,1522
D.Losses	1,5381

Πίνακας 8.3. Τιμές VIF των επεξηγηματικών μεταβλητών μετά την αφαίρεση των μεταβλητών που ευθύνονταν για την πολυσυγγραμμικότητα (CH₄ – για τις χώρες χωρίς πρωτογενή παραγωγή φυσικού αερίου)

Στη συνέχεια χρησιμοποιώ την μέθοδο ελαχίστων τετραγώνων (OLS). Μέσα από τις εντολές:

```
> ols <-lm(y ~ x4+x7+x8+x9+x10+x12, data=data)
> summary(ols)
```

παίρνω τα πιο κάτω αποτελέσματα για τις εκτιμήσεις των συντελεστών των επεξηγηματικών μεταβλητών.

Call:

```
lm(formula = y ~ x4 + x7 + x8 + x9 + x10 + x12, data = data)
```

Residuals:

```
   Min      1Q  Median      3Q      Max
-1449.54 -306.08  -20.45   310.42  2126.86
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.983e+02  2.109e+02  -2.837  0.00572 **
x4           3.129e+00  8.928e-02  35.048 < 2e-16 ***
x7           1.466e-02  1.294e-03  11.336 < 2e-16 ***
x8           8.509e-02  3.929e-02   2.165  0.03323 *
x9           7.610e-01  5.985e-01   1.271  0.20713
x10          1.308e-03  9.507e-04   1.376  0.17243
x12         -3.714e-02  1.457e-01  -0.255  0.79941
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 621.2 on 83 degrees of freedom

Multiple R-squared: 0.9731, Adjusted R-squared: 0.9711

F-statistic: 499.7 on 6 and 83 DF, p-value: < 2.2e-16

Ακολουθώς, εφαρμόζω τη μέθοδο LSDV για τις σταθερές επιδράσεις:

```
> fixed.dum <-lm(y~ x4 + x7 + x8 + x9 + x10 + x12+ factor(Country) - 1, data=data)
> summary(fixed.dum)
```

Call:

```
lm(formula = y ~ x4 + x7 + x8 + x9 + x10 + x12 + factor(Country) -
    1, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-548.84	-105.55	8.48	121.59	522.67

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
x4	5.114e-01	1.389e-01	3.683	0.000421	***
x7	-6.720e-03	1.325e-03	-5.070	2.56e-06	***
x8	2.008e-01	2.873e-02	6.991	7.78e-10	***
x9	2.247e-01	3.326e-01	0.675	0.501362	
x10	4.384e-03	9.178e-04	4.777	8.07e-06	***
x12	1.090e-01	6.191e-02	1.761	0.082051	.
factor(Country) Belgium	4.752e+03	7.766e+02	6.120	3.40e-08	***
factor(Country) Latvia	1.156e+03	1.277e+02	9.048	7.84e-14	***
factor(Country) Lithuania	2.348e+03	2.174e+02	10.799	< 2e-16	***
factor(Country) Portugal	1.078e+04	5.986e+02	18.006	< 2e-16	***
factor(Country) Sweden	7.228e+03	5.285e+02	13.678	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 247.6 on 79 degrees of freedom

Multiple R-squared: 0.9989, Adjusted R-squared: 0.9988

F-statistic: 6727 on 11 and 79 DF, p-value: < 2.2e-16

Για να συγκρίνουμε τις δύο μεθόδους αποθηκεύουμε τις εκτιμήσεις του μοντέλου σταθερών επιδράσεων με την πιο κάτω εντολή:

```
> fixed <- plm(y ~ x4 + x7 + x8 + x9 + x10 + x12, data=data, index=c("Country",
    "Year"), model="within")
```

και εκτελώ των έλεγχο pFtest

```
> pFtest(fixed, ols)
```

F test for individual effects

data: y ~ x4 + x7 + x8 + x9 + x10 + x12

F = 110.85, df1 = 4, df2 = 79, **p-value < 2.2e-16**

alternative hypothesis: significant effects

Η p-τιμή του ελέγχου είναι μικρότερη από 0.05 επομένως απορρίπτουμε τη μηδενική υπόθεση και από αυτές τις δύο επιλογές διαλέγουμε το μοντέλο των σταθερών επιδράσεων (FE).

Στον επόμενο στάδιο ελέγχουμε αν είναι προτιμότερο να χρησιμοποιήσουμε τις τυχαίες αντί τις σταθερές επιδράσεις.

Αποθηκεύουμε τους συντελεστές των επεξηγηματικών μεταβλητών με την πιο κάτω εντολή:
 > random <- plm(y ~ x4+x7+x8+x9+x10+x12, data=data, index=c("Country", "Year"),
 model="pooling")
 > summary(random)

Oneway (individual) effect Pooling Model

Call:

```
plm(formula = y ~ x4 + x7 + x8 + x9 + x10 + x12, data = data,
     model = "pooling", index = c("Country", "Year"))
```

Balanced Panel: n=5, T=18, N=90

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-1450.0	-306.0	-20.4	310.0	2130.0

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	-5.9828e+02	2.1088e+02	-2.8371	0.00572	**
x4	3.1291e+00	8.9282e-02	35.0481	< 2e-16	***
x7	1.4664e-02	1.2937e-03	11.3355	< 2e-16	***
x8	8.5085e-02	3.9293e-02	2.1654	0.03323	*
x9	7.6097e-01	5.9851e-01	1.2714	0.20713	
x10	1.3084e-03	9.5066e-04	1.3763	0.17243	
x12	-3.7145e-02	1.4571e-01	-0.2549	0.79941	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1.189e+09

Residual Sum of Squares: 32033000

R-Squared : 0.97306

Adj. R-Squared : 0.89738

F-statistic: 499.651 on 6 and 83 DF, p-value: < 2.22e-16

Στη συνέχεια εκτελούμε τον έλεγχο Lagrange Multiplier Test - (Breusch-Pagan):

```
> plmtest(pool, type=c("bp"))
```

Lagrange Multiplier Test - (Breusch-Pagan)

data: y ~ x4 + x7 + x8 + x9 + x10 + x12

chisq = 0.88748, df = 1, **p-value = 0.3462**

alternative hypothesis: significant effects

Η p-τιμή είναι μεγαλύτερη από 0.05 οπότε η απλή OLS παλινδρόμηση κρίνεται καλύτερη επιλογή από την παλινδρόμηση τυχαίων επιδράσεων.

Με βάση τους πιο πάνω ελέγχους επιλέγουμε να χρησιμοποιήσουμε το μοντέλο των σταθερών επιδράσεων (FE). Ακόμη και αν η απλή OLS παλινδρόμηση θεωρήθηκε προτιμότερη από την παλινδρόμηση τυχαίων επιδράσεων, το μοντέλο των σταθερών επιδράσεων (FE) βάσει του ελέγχου pFtest είναι μια πιο καλή επιλογή.

Μετέπειτα, πρέπει να επιλέξουμε ένα υποσύνολο μεταβλητών που θα είναι όσο το δυνατόν μικρότερο αλλά παράλληλα θα μας δίνει καλή προσαρμογή για το μοντέλο. Με βάση τα κριτήρια BIC και AIC, παίρνουμε τους πιο κάτω πίνακες:

	x4	x7	x8	x9	x10	x12	LTV	LTH	PRG	SW	BIC	AIC
1	x	x	x		x		x	x	x	x	128,5649	126,0651
2	x	x	x		x	x	x	x	x	x	128,5836	125,8338
3	x	x	x	x	x		x	x	x	x	128,8784	126,1286
4	x	x	x	x	x	x	x	x	x	x	128,9817	125,9820
5	x	x	x		x		x		x	x	129,5001	127,2502
6		x	x		x	x	x	x	x	x	129,5082	127,0084
7	x	x	x	x	x		x		x	x	129,5574	127,0576
8		x	x		x		x	x	x	x	129,6518	127,4020
9	x	x	x		x	x	x	x	x		129,7996	127,2998
10	x	x	x		x	x	x		x	x	129,8781	127,3783

Πίνακας 8.4. Πίνακας των δέκα καταλληλότερων μοντέλων με βάση το BIC για την πρόβλεψη εκπομπών CH₄ των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC)

	x4	x7	x8	x9	x10	x12	LTV	LTH	PRG	SW	BIC	AIC
1	x	x	x		x	x	x	x	x	x	128,5836	125,8338
2	x	x	x	x	x	x	x	x	x	x	128,9817	125,9820
3	x	x	x		x		x	x	x	x	128,5649	126,0651
4	x	x	x	x	x		x	x	x	x	128,8784	126,1286
5		x	x		x	x	x	x	x	x	129,5082	127,0084
6	x	x	x	x	x		x		x	x	129,5574	127,0576
7		x	x	x	x	x	x	x	x	x	129,9575	127,2077
8	x	x	x	x	x	x	x		x	x	129,9742	127,2244
9	x	x	x		x		x		x	x	129,5001	127,2502
10	x	x	x		x	x	x	x	x		129,7996	127,2998

Πίνακας 8.5. Πίνακας των δέκα καταλληλότερων μοντέλων με βάση το AIC για την πρόβλεψη εκπομπών CH₄ των χωρών χωρίς πρωτογενή παραγωγή φυσικού αερίου (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο AIC)

Παρατηρούμε ότι μικρότερο BIC επιτυγχάνεται με την επιλογή των μεταβλητών Waste (X₄), GDP (X₇), Primary Production of hard Coal (X₈) και Gross Inland Consumption of Natural Gas (X₁₀) (Πίνακας 8.4). Όμως, το καλύτερο μοντέλο με βάση το AIC, περιλαμβάνει τις μεταβλητές Waste (X₄), GDP (X₇), Primary Production of hard Coal (X₈), Gross Inland Consumption of Natural Gas (X₁₀) και Distribution Losses (X₁₂) (Πίνακας 8.5). Οι τιμές των AIC και BIC δεν διαφέρουν πολύ από τα δύο μοντέλα, οπότε προτιμώ να επιλέξω το μοντέλο με τις λιγότερες μεταβλητές.

Οπότε το μοντέλο μας διαμορφώνεται ως εξής:

```
> fixed <- plm(y ~x4+x7+x8+x10, data=data, index=c("Country", "Year"),
  model="within")
> summary(fixed)
```

Oneway (individual) effect Within Model

Call:

```
plm(formula = y ~ x4 + x7 + x8 + x10, data = data, model = "within",
  index = c("Country", "Year"))
```

Balanced Panel: n=5, T=18, N=90

Residuals :

```
  Min. 1st Qu.  Median 3rd Qu.  Max.
-565.0 -105.0   4.2  121.0  549.0
```

Coefficients :

```
  Estimate Std. Error t-value Pr(>|t|)
x4  0.53008602  0.13648628  3.8838 0.000209 ***
x7 -0.00674956  0.00132245 -5.1038 2.157e-06 ***
x8  0.20162893  0.02868924  7.0280 5.953e-10 ***
x10 0.00472702  0.00090336  5.2327 1.285e-06 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 27410000

Residual Sum of Squares: 5111200

R-Squared : 0.81353

Adj. R-Squared : 0.73217

F-statistic: 88.344 on 4 and 81 DF, p-value: < 2.22e-16

Ο συντελεστής κάθε χώρας είναι:

```
> fixef(fixed)
  Belgium  Latvia  Lithuania Portugal  Sweden
4595.920 1217.619 2387.758 10751.118 7367.887
```

Έλεγχος Ομοσκεδαστικότητας (Breusch-Pagan test):

```
> bptest(y ~x4+x7+x8+x10 + factor(Country), data = data, studentize=F)
```

Breusch-Pagan test

```
data: y ~ x4 + x7 + x8 + x10 + factor(Country)
```

BP = 44.286, df = 8, **p-value = 5.023e-07**

Η p-τιμή ελέγχου είναι μικρότερη από 0.05 οπότε απορρίπτεται η υπόθεση της ομοσκεδαστικότητας.

Για την αντιμετώπιση της ετεροσκεδαστικότητας μπορώ να χρησιμοποιήσω του εκτιμητές συνδιακύμανσης robust (arellano).

```
>coefest(fixed, vcovHC(fixed, method = "arellano"))
```

```
Estimate Std. Error t value Pr(> |t|)
x4  0.5300860  0.2623382  2.0206 0.0466234 *
x7 -0.0067496  0.0022294 -3.0275 0.0033063 **
x8  0.2016289  0.0505469  3.9889 0.0001447 ***
x10 0.0047270  0.0023358  2.0238 0.0462925 *
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Μετά από την πιο πάνω διαδικασία καταλήγουμε στο μοντέλο σταθερών επιδράσεων:

$$Y=0.0530*x4-0.0067*x7+0.2016*x8+0.0047*x10+CountryFixedEffects$$

Country Fixed Effects:

Belgium	Latvia	Lithuania	Portugal	Sweden
4595.920	1217.619	2387.758	10751.118	7367.887

Αντικαθιστώντας τα στοιχεία του 2013 στις επεξηγηματικές μεταβλητές, το αποτέλεσμα της πρόβλεψης των εκπομπών μεθανίου του Βελγίου, της Λετονίας, της Λιθουανίας, της Πορτογαλίας και της Σουηδίας είναι :

$$Y_{\text{Belgium}} = 0.0530*46-0.0067*382692+0.2016*4641+0.0047*602704+4595.920$$

$$=5822.028 \text{ Gg}$$

$$Y_{\text{Latvia}} = 0.0530*521-0.0067*23372.1+0.2016*123+0.0047*50438+1217.619$$

$$=1599.26 \text{ Gg}$$

$$Y_{\text{Lithuania}} = 0.0530*798-0.0067*34631.2+0.2016*375+0.0047*90624+2387.758$$

$$=3081.01 \text{ Gg}$$

$$Y_{\text{Portugal}} = 0.0530*2320-0.0067*165690+0.2016*4449+0.0047*157251+10751.118$$

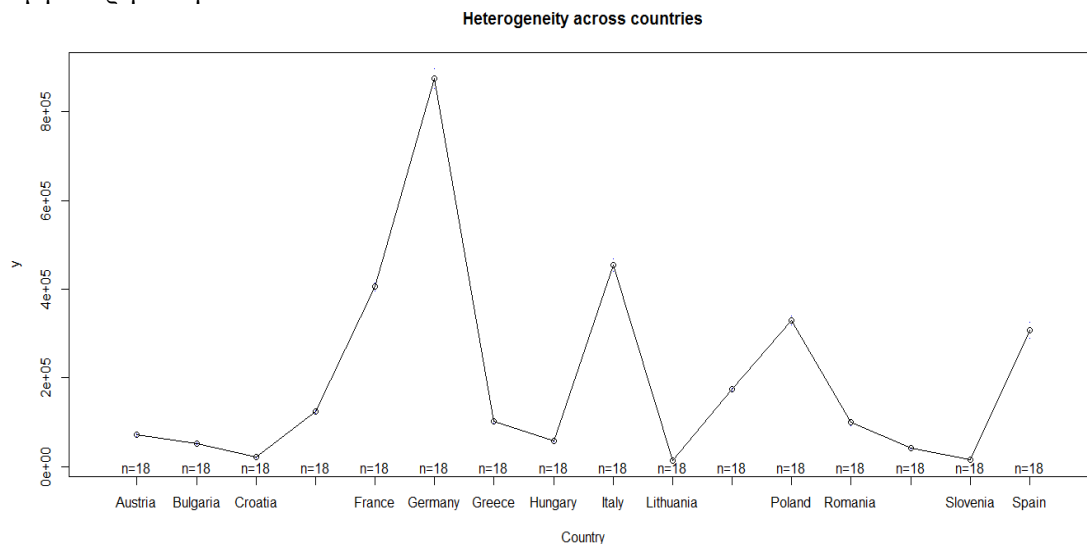
$$=12502.95 \text{ Gg}$$

$$Y_{\text{Sweden}} = 0.0530*28-0.0067*420849.1+0.2016*2901+0.0047*39996+7367.887$$

$$=5316.153 \text{ Gg}$$

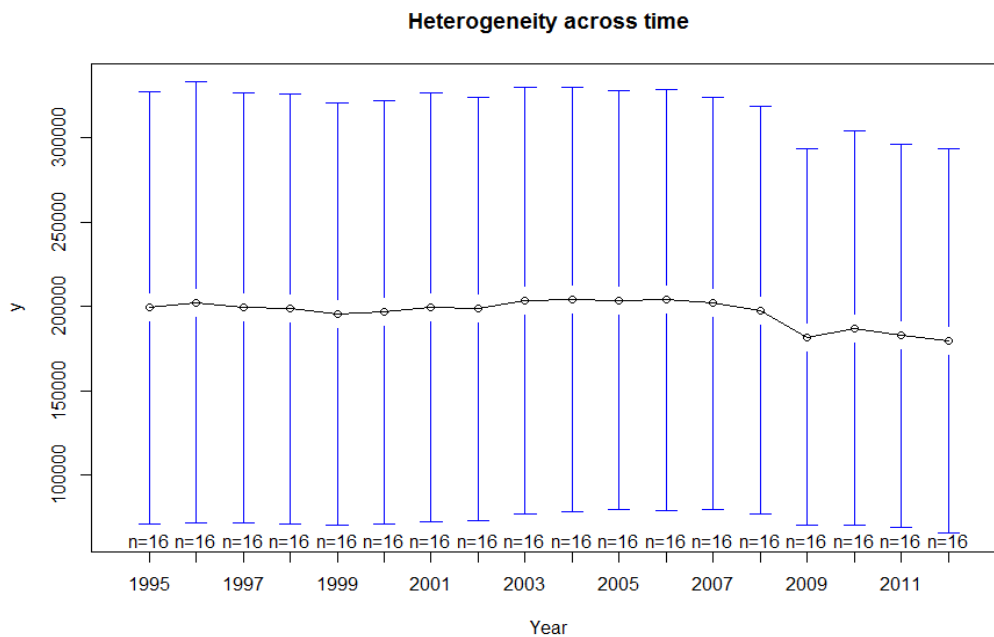
8.2. Ανάλυση Παλινδρόμησης με Panel Data για τις εκπομπές διοξειδίου του άνθρακα των χωρών με πλήρη δεδομένα

Στο Διάγραμμα 8.5 που ακολουθεί παρατηρούμε την ανομοιογένεια που υπάρχει από χώρα σε χώρα και να συγκρίνουμε τις διαφορές ανάμεσα στις εκπομπές διοξειδίου του άνθρακα των διαφόρων χωρών. Η Γερμανία, ανάμεσα σε αυτές τις χώρες, κατέχει την υψηλότερη θέση.



Διάγραμμα 8.5. Μέση τιμή εκπομπών CO₂ ανά χώρα για τις χώρες με πλήρη δεδομένα (1995-2012)

Στο Διάγραμμα 8.6 παρατηρούμε την ανομοιογένεια με την πάροδο του χρόνου. Μπορούμε να δούμε ότι οι εκπομπές στο σύνολό τους έχουν μειωθεί από το 1995 στο 2012.



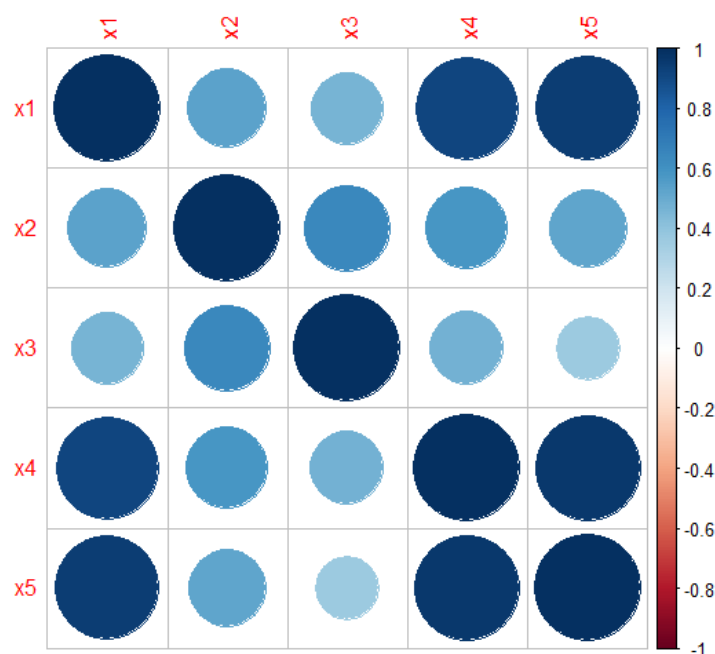
Διάγραμμα 8.6. Συνολικές εκπομπές CO₂ ανά έτος για τις χώρες με πλήρη δεδομένα (1995-2012)

Έλεγχος Πολυσυγγραμμικότητας

Από τον Πίνακα 8.6 και το Διάγραμμα 8.7 παρατηρούμε ότι υπάρχει έντονη συσχέτιση μεταξύ τριών εξηγηματικών μεταβλητών, των Gross Inland Consumption of Natural Gas, Gross Inland Consumption of Petroleum Products και Gross Inland Consumption of Crude Oil.

	C.Natural Gas	C.Hard Coal	C.Lignite	C.Petroleum Prod.	C.Crude Oil
C.Natural Gas	1	0.5397095	0.4642071	0.9126320	0.9423089
C.Hard Coal	0.5397095	1	0.6403134	0.5871241	0.5232714
C.Lignite	0.4642071	0.6403134	1	0.4723656	0.3630982
C.Petroleum Prod.	0.9126320	0.5871241	0.4723656	1	0.9606218
C.Crude Oil	0.9423089	0.5232714	0.3630982	0.9606218	1

Πίνακας 8.6. Πίνακας με τις τιμές συσχέτισης των εξηγηματικών μεταβλητών (CO₂ – για τις χώρες με πλήρη δεδομένα)



Διάγραμμα 8.7. Διάγραμμα συσχέτισης των εξηγηματικών μεταβλητών (CO₂ – για τις χώρες με πλήρη δεδομένα)

Variables	VIF
C.Natural Gas	10,7675
C.Hard Coal	2,0671
C.Lignite	2,2847
C.Petroleum Prod.	17,2804
C.Crude Oil	26,7358

Πίνακας 8.7. Τιμές VIF των εξηγηματικών μεταβλητών (CO₂ – για τις χώρες με πλήρη δεδομένα)

Αφαιρώντας την μεταβλητή Gross Inland Consumption of Crude Oil(X_5), η οποία έχει την μεγαλύτερη τιμή VIF (Πίνακας 8.7), βλέπουμε ότι οι υπόλοιπες τιμές μειώνονται αισθητά και δεν υπάρχουν πλέον σοβαρές ενδείξεις για ύπαρξη πολυσυγγραμμικότητας (Πίνακας 8.8). Επομένως δεν χρειάζεται να αφαιρέσω την Gross Inland Consumption of Petroleum Products (X_4), παρόλο που συσχετίζεται με την Gross Inland Consumption of Hard Coal (X_1).

Variables	VIF
C.Natural Gas	6,0454
C.Hard Coal	2,0624
C.Lignite	1,7542
C.Petroleum Prod.	6,4740

Πίνακας 8.8. Τιμές VIF των επεξηγηματικών μεταβλητών μετά την αφαίρεση της μεταβλητής που ευθυνόταν για την πολυσυγγραμμικότητα (CO_2 – για τις χώρες με πλήρη δεδομένα)

Στη συνέχεια χρησιμοποιώ την μέθοδο ελαχίστων τετραγώνων (OLS). Μέσα από τις εντολές:

```
> ols <-lm(y ~ x1+x2+x3+x4, data=data)
> summary(ols)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-38185  -3886   2064   5226  44105

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.788e+03  8.819e+02  -3.161  0.00174 **
x1           5.233e-02  1.788e-03  29.259 < 2e-16 ***
x2           2.203e+00  3.824e-02  57.606 < 2e-16 ***
x3           9.427e-01  1.946e-02  48.447 < 2e-16 ***
x4           3.178e+00  4.518e-02  70.342 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11110 on 283 degrees of freedom
Multiple R-squared:  0.9976,    Adjusted R-squared:  0.9976
F-statistic: 2.935e+04 on 4 and 283 DF,  p-value: < 2.2e-16
```


Ακολουθώντας εφαρμόζω τη μέθοδο LSDV για τις σταθερές επιδράσεις:

```
>fixed.dum <-lm(y~ x1+x2+x3+x4+ factor(Country) - 1, data=data)
>summary(fixed.dum)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4 + factor(Country) - 1, data = data)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-27116.6 -1238.2      6.7   1048.8 26735.8
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
x1	4.671e-02	2.125e-03	21.983	< 2e-16	***
x2	2.417e+00	8.176e-02	29.568	< 2e-16	***
x3	9.335e-01	1.005e-01	9.285	< 2e-16	***
x4	3.002e+00	6.870e-02	43.700	< 2e-16	***
factor(Country) Austria	5.204e+03	1.650e+03	3.154	0.001794	**
factor(Country) Bulgaria	-5.164e+03	3.045e+03	-1.696	0.091088	.
factor(Country) Croatia	2.125e+03	1.202e+03	1.768	0.078275	.
factor(Country) Czech Republic	1.435e+04	4.928e+03	2.911	0.003903	**
factor(Country) France	1.753e+04	7.329e+03	2.391	0.017473	*
factor(Country) Germany	2.415e+04	2.112e+04	1.144	0.253848	
factor(Country) Greece	-1.226e+04	6.585e+03	-1.862	0.063731	.
factor(Country) Hungary	1.282e+03	2.040e+03	0.628	0.530272	
factor(Country) Italy	3.690e+04	8.236e+03	4.481	1.10e-05	***
factor(Country) Lithuania	5.460e+02	1.179e+03	0.463	0.643667	
factor(Country) Netherlands	-1.700e+04	4.171e+03	-4.077	6.02e-05	***
factor(Country) Poland	-1.327e+04	8.107e+03	-1.636	0.102934	
factor(Country) Romania	-2.515e+02	3.907e+03	-0.064	0.948714	
factor(Country) Slovakia	4.968e+03	1.375e+03	3.614	0.000361	***
factor(Country) Slovenia	1.270e+03	1.249e+03	1.017	0.310124	
factor(Country) Spain	-5.363e+03	5.477e+03	-0.979	0.328310	

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4849 on 268 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9997

F-statistic: 5.448e+04 on 20 and 268 DF, p-value: < 2.2e-16

Για να συγκρίνουμε τις δύο μεθόδους αποθηκεύουμε τις εκτιμήσεις του μοντέλου σταθερών επιδράσεων με την πιο κάτω εντολή:

```
>fixed <- plm(y ~x1+x2+x3+x4, data=data, index=c("Country", "Year"),
  model="within")
```

Και εκτελώ των έλεγχο pFtest:

```
> pFtest(fixed, ols)
```

F test for individual effects

data: y ~ x1 + x2 + x3 + x4

F = 81.145, df1 = 15, df2 = 268, **p-value < 2.2e-16**

alternative hypothesis: significant effects

Η p-τιμή του ελέγχου είναι μικρότερη από 0.05, επομένως απορρίπτουμε τη μηδενική υπόθεση και από αυτές τις δύο επιλογές διαλέγουμε το μοντέλο των σταθερών επιδράσεων (FE).

Στον επόμενο στάδιο ελέγχουμε αν είναι προτιμότερο να χρησιμοποιήσουμε τις τυχαίες αντί τις σταθερές επιδράσεις.

Αποθηκεύουμε τους συντελεστές των επεξηγηματικών μεταβλητών με την πιο κάτω εντολή:

```
> random <- plm(y ~x1+x2+x3+x4, data=data, index=c("Country", "Year"),
  model="random")
> summary(random)
```

Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:

```
plm(formula = y ~ x1 + x2 + x3 + x4, data = data, model = "random",
  index = c("Country", "Year"))
```

Balanced Panel: n=16, T=18, N=288

Effects:

	var	std.dev	share
idiosyncratic	23511311	4849	0.143
individual	140776135	11865	0.857
theta:	0.9041		

Residuals :

Min.	1st Qu.	Median	3rd Qu.	Max.
-23600	-1230	151	1110	29600

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	-4.5829e+02	3.6934e+03	-0.1241	0.9013
x1	4.8669e-02	1.8382e-03	26.4772	<2e-16 ***
x2	2.3946e+00	7.3964e-02	32.3757	<2e-16 ***
x3	9.4829e-01	6.1771e-02	15.3517	<2e-16 ***
x4	3.0811e+00	5.6850e-02	54.1970	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 2.2351e+11

Residual Sum of Squares: 6693600000

R-Squared : 0.97005

Adj. R-Squared : 0.95321

F-statistic: 2291.67 on 4 and 283 DF, p-value: < 2.22e-16

Στη συνέχεια εκτελούμε τον έλεγχο Hausman Test:

```
> phtest(fixed,random)
```

Hausman Test

data: y ~ x1 + x2 + x3 + x4

chisq = 5.7937, df = 4, **p-value = 0.2151**

alternative hypothesis: one model is inconsistent

Η p-τιμή ελέγχου είναι μεγαλύτερη από 0.05, επομένως δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση και καταλήγουμε στην επιλογή του μοντέλο των τυχαίων επιδράσεων (RE).

Μετέπειτα, πρέπει να επιλέξουμε ένα υποσύνολο μεταβλητών που θα είναι όσο το δυνατόν μικρότερο αλλά παράλληλα θα μας δίνει καλή προσαρμογή μοντέλου. Με βάση τα κριτήρια BIC και AIC παίρνουμε τον πιο κάτω πίνακα:

	x1	x2	x3	x4	BIC	AIC
1	x	x	x	x	6.211,923	6.189,945
2		x	x	x	6.607,316	6.589,001
3	x	x		x	6.848,312	6.829,997
4	x		x	x	6.938,830	6.920,515
5		x		x	6.948,951	6.934,299
6			x	x	6.985,999	6.971,347
7	x	x	x		7.046,329	7.028,014
8	x	x			7.144,562	7.129,910
9	x		x		7.258,573	7.243,921
10	x			x	7.274,564	7.259,912
11				x	7.297,993	7.287,004
12	x				7.445,720	7.434,731
13		x	x		7.620,397	7.605,745
14		x			7.661,268	7.650,279
15			x		7.749,119	7.738,130
16					7.926,036	7.918,710

Πίνακας 8.9. Πίνακας όλων των πιθανών μοντέλων με βάση το BIC και AIC για την πρόβλεψη εκπομπών CO₂ των με πλήρη δεδομένα (ταξινομημένα ξεκινώντας από το μοντέλο με το μικρότερο BIC και AIC)

Παρατηρούμε ότι μικρότερο BIC και AIC επιτυγχάνεται με την επιλογή και των τεσσάρων μεταβλητών (Πίνακας 8.9).

Μένει να ελέγξουμε την ύπαρξη ομοσκεδαστικότητας:

```
> bptest(y ~ x1+x2+x3+x4 + factor(Country), data = data, studentize=F)
```

Breusch-Pagan test

```
data: y ~ x1 + x2 + x3 + x4 + factor(Country)
```

BP = 585.99, df = 19, **p-value < 2.2e-16**

Η p-τιμή ελέγχου είναι μικρότερη από 0.05 οπότε απορρίπτεται η υπόθεση της ομοσκεδαστικότητας.

Για την αντιμετώπιση της ετεροσκεδαστικότητας μπορώ να χρησιμοποιήσω του εκτιμητές συνδιακύμανσης robust (white1).

```
> coeftest(random, vcovHC(random, method = "white1"))
```

```
Estimate Std. Error t value Pr(> |t|)
```

```
(Intercept) -4.5829e+02 2.9268e+03 -0.1566 0.8757
```

```
x1 4.8669e-02 3.4544e-03 14.0890 <2e-16 ***
```

```
x2 2.3946e+00 8.6379e-02 27.7225 <2e-16 ***
```

```
x3 9.4829e-01 8.4337e-02 11.2441 <2e-16 ***
```

```
x4 3.0811e+00 9.8505e-02 31.2785 <2e-16 ***
```

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Μετά από την πιο πάνω διαδικασία καταλήγουμε στο μοντέλο τυχαίων επιδράσεων:

$$Y = -458.29 + 0.0487 * CNaturalGas + 2.3946 * CHardCoal + 0.9483 * CLignite + 3.0811 * CPetroleumProducts$$

Αντικαθιστώντας τα στοιχεία του 2013 στις επεξηγηματικές μεταβλητές, το αποτέλεσμα της πρόβλεψης των εκπομπών διοξειδίου του άνθρακα της Αυστρίας, της Κροατίας, της Γαλλίας, της Ελλάδας, της Βουλγαρίας, της Τσεχίας, της Γερμανίας, της Ουγγαρίας, της Ιταλίας, της Λιθουανίας, της Ολλανδίας, της Πολωνίας, της Ρουμανίας της Σλοβακίας, της Σλοβενίας και της Ισπανίας είναι :

$$Y_{Austria} = -458.29 + 0.0487 * 293566 + 2.3946 * 4876 + 0.9483 * 13 + 3.0811 * 12075 \\ = \mathbf{62721.95 \text{ Gg}}$$

$$Y_{Croatia} = -458.29 + 0.0487 * 95537 + 2.3946 * 1083 + 0.9483 * 56 + 3.0811 * 3220 \\ = \mathbf{16759 \text{ Gg}}$$

$$Y_{France} = -458.29 + 0.0487 * 1633196 + 2.3946 * 18991 + 0.9483 * 146 + 3.0811 * 76568 \\ = \mathbf{360555.7 \text{ Gg}}$$

$$Y_{Greece} = -458.29 + 0.0487 * 135497 + 2.3946 * 302 + 0.9483 * 54386 + 3.0811 * 11492 \\ = \mathbf{93841.08 \text{ Gg}}$$

$$Y_{Bulgaria} = -458.29 + 0.0487 * 99977 + 2.3946 * 1830 + 0.9483 * 28719 + 3.0811 * 3698 \\ = \mathbf{47417.46 \text{ Gg}}$$

$$Y_{CzechRepublic} = -458.29 + 0.0487 * 290832 + 2.3946 * 7316 + 0.9483 * 38936 + 3.0811 * 8405 \\ = \mathbf{94034.37 \text{ Gg}}$$

$$Y_{Germany} = -458.29 + 0.0487 * 3051546 + 2.3946 * 66136 + 0.9483 * 182491 + 3.0811 * 109635 \\ = \mathbf{817277.5 \text{ Gg}}$$

$$Y_{Hungary} = -458.29 + 0.0487 * 322601 + 2.3946 * 1015 + 0.9483 * 9662 + 3.0811 * 5699 \\ = \mathbf{44394.46 \text{ Gg}}$$

$$Y_{Italy} = -458.29 + 0.0487 * 2402667 + 2.3946 * 21627 + 0.9483 * 5 + 3.0811 * 56950 \\ = \mathbf{343738.5 \text{ Gg}}$$

$$Y_{Lithuania} = -458.29 + 0.0487 * 90624 + 2.3946 * 375 + 0.9483 * 1 + 3.0811 * 2457 \\ = \mathbf{12421.48 \text{ Gg}}$$

$$Y_{Netherlands} = -458.29 + 0.0487 * 1391534 + 2.3946 * 13049 + 0.9483 * 31 + 3.0811 * 32013 \\ = \mathbf{197178.1 \text{ Gg}}$$

$$Y_{Poland} = -458.29 + 0.0487 * 574674 + 2.3946 * 72085 + 0.9483 * 65934 + 3.0811 * 22715 \\ = \mathbf{332637 \text{ Gg}}$$

$$Y_{Romania} = -458.29 + 0.0487 * 410052 + 2.3946 * 1660 + 0.9483 * 24971 + 3.0811 * 8267 \\ = \mathbf{72624.77 \text{ Gg}}$$

$$Y_{Slovakia} = -458.29 + 0.0487 * 201571 + 2.3946 * 166040540 + 0.9483 * 2730 + 3.0811 * 3505 \\ = \mathbf{32447.76 \text{ Gg}}$$

$$Y_{Slovenia} = -458.29 + 0.0487 * 28966 + 2.3946 * 457 + 0.9483 * 4030 + 3.0811 * 2357 \\ = \mathbf{13129.55 \text{ Gg}}$$

$$Y_{Spain} = -458.29 + 0.0487 * 1092028 + 2.3946 * 20633 + 0.9483 * 0 + 3.0811 * 49787 \\ = \mathbf{255496.1 \text{ Gg}}$$

9. Πρόβλεψη

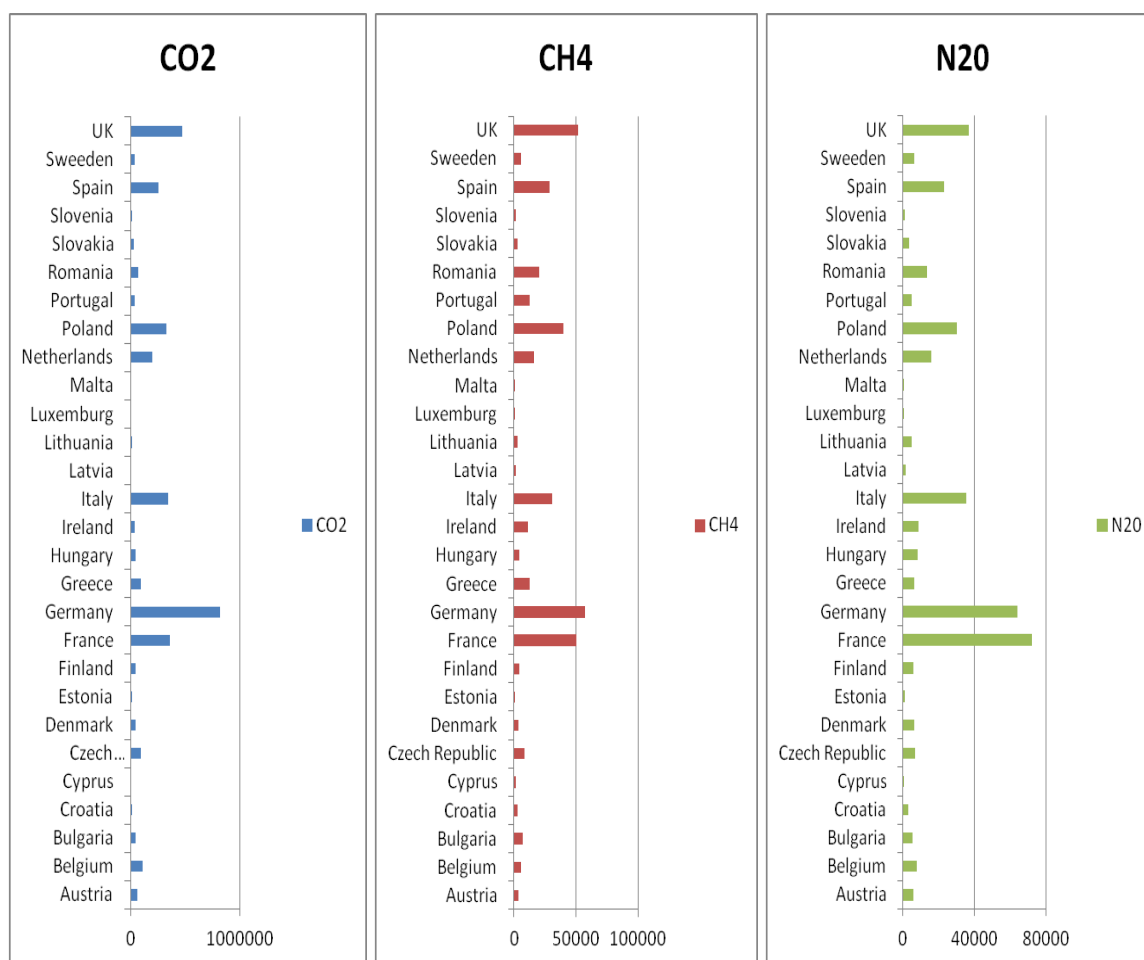
Αφού έχουμε καταλήξει στα βέλτιστα μοντέλα για κάθε σύνολο δεδομένων panel (Παράρτημα Ι) μπορούμε να αντικαταστήσουμε τις τιμές των επεξηγηματικών μεταβλητών με βάση τα δεδομένα που έχουμε ώστε να παράγουμε την πρόβλεψη για τις εκπομπές του 2013. Αντικαθιστώντας τις τιμές των μεταβλητών σε όλα τα μοντέλα που επιλέξαμε έχουμε τις προβλέψεις για κάθε αέριο του θερμοκηπίου και για κάθε χώρα (Πίνακας 9.1). Έτσι προσθέτοντας αυτές τις τιμές των προβλέψεων έχουμε τις προβλέψεις για τους τρεις ρύπους στην ΕΕ-28 για το 2013 (Πίνακας 9.2).

	CO2	CH4	N2O
Austria	62721,95	3803,115	5758,421
Belgium	113668,2	5822,028	7639,588
Bulgaria	47417,46	6963,44	5330,616
Croatia	16759	3137,2	3017,254
Cyprus	5083,109	1321,825	702,7977
Czech Republic	94034,37	8754,495	6649,003
Denmark	46232,59	3539,491	6332,479
Estonia	16488,42	910,9924	861,7448
Finland	49678,12	4087,08	5764,887
France	360555,7	50459,62	72125,34
Germany	817277,5	57263,35	63943,9
Greece	93841,08	12709,55	6105,572
Hungary	44394,46	4298,573	8158,014
Ireland	36549,63	11173,3	8574,725
Italy	343738,5	30932,82	35535,89
Latvia	7051,964	1599,262	1546,862
Lithuania	12421,48	3081,01	4924,225
Luxemburg	11523,79	125,5712	530,6894
Malta	2190,597	82,14838	44,75671
Netherlands	197178,1	16192,26	15540,2
Poland	332637	39945,14	30055,72
Portugal	37144,67	12502,95	4687,194
Romania	72624,77	20221,67	13375,76
Slovakia	32447,76	2499,877	3523,533
Slovenia	13129,55	1214,506	1168,052
Spain	255496,1	28494,28	22932,02
Sweden	40346,91	5316,153	6451,109
UK	472351,1	51537,44	36798,33

Πίνακας 9.1: Προβλέψεις εκπομπών αερίων του θερμοκηπίου κάθε χώρας για το 2013 με τη χρήση των panel data

	CO2	CH4	N2O
EU-28	3634984	387989,1	378078,7

Πίνακας 9.2: Πρόβλεψη εκπομπών αερίων του θερμοκηπίου της ΕΕ-28 για το 2013 με τη χρήση των panel data



Διάγραμμα 9.1 Γραφικές Παραστάσεις των προβλέψεων των εκπομπών των αερίων του θερμοκηπίου στις χώρες της ΕΕ-28 (βάσει των Panel Data)

Παρατηρώντας το Διάγραμμα 9.1 αλλά και τον Πίνακα 9.1 βλέπουμε μια συμφωνία με τις προβλέψεις που έγιναν με βάση την πολλαπλή γραμμική παλινδρόμηση καθώς οι Γαλλία, Γερμανία, Ιταλία, Πολωνία, Ισπανία και Ηνωμένο Βασίλειο είναι οι χώρες με τις υψηλότερες εκπομπές αερίων του θερμοκηπίου, με βάση την ανάλυση με χρήση των Panel Data. Οι εκπομπές διοξειδίου του άνθρακα της Γερμανίας και εδώ διαφέρουν κατά πολύ σε σχέση με τις άλλες χώρες. Όμως υπάρχει κάποιες διαφοροποιήσεις. Στις εκπομπές υποξειδίου του αζώτου η Γαλλία κατέχει την πρώτη θέση ενώ το Ηνωμένο Βασίλειο (το οποίο με βάση την πολλαπλή ανάλυση παλινδρόμησης κατείχε την πρώτη θέση στις εκπομπές N₂O) κατέχει την Τρίτη θέση. Επιπρόσθετα, όσον αφορά τις εκπομπές μεθανίου η Γερμανία βρίσκεται στην πρώτη θέση ενώ με μικρή διαφορά το Ηνωμένο Βασίλειο είναι στη δεύτερη και η Γαλλία στην τρίτη θέση. Παρόλα αυτά οι διαφοροποιήσεις ανάμεσα στις δύο προβλέψεις είναι μικρές και όχι πολύ σημαντικές, καθώς οι τιμές των προβλεπόμενων εκπομπών των δύο αναλύσεων είναι πολύ κοντά μεταξύ τους.

10. Συμπεράσματα

Συγκρίνοντας τις προβλέψεις των δύο μεθόδων παίρνουμε τον πιο κάτω πίνακα διαφορών.

	CO2	CH4	N2O
Austria	3944,756	1314,97	-352,72
Belgium	-9071,9	532,8454	-914,191
Bulgaria	-4163,51	-433,308	-74,8405
Croatia	2616,812	0	241,5616
Cyprus	1294,87	12,16878	-133,072
Czech Republic	12628,43	608,753	1006,299
Denmark	-3582,35	2054,413	-50,7323
Estonia	-216,474	31,24249	154,6692
Finland	6536,17	-22,1736	-769,927
France	11181,01	1116,878	-17446
Germany	16668,04	-9879,9	-7099,72
Greece	-11981,7	-2773,96	1051,702
Hungary	-831,524	3839,542	-1688,92
Ireland	1551,641	643,0322	-1032,85
Italy	10596,09	2822,608	-15462,4
Latvia	406,6254	-8,43626	195,0873
Lithuania	421,3346	-272,449	-1748,07
Luxemburg	-1320,41	304,7169	-58,1629
Malta	226,2903	50,09282	50,94901
Netherlands	-29878	-966,417	-8020,81
Poland	-15414,5	730,5809	-408,926
Portugal	12228	-770,969	-212,433
Romania	182,1999	637,8966	-1570,12
Slovakia	5137,365	1652,232	-501,77
Slovenia	2126,13	581,0051	-82,428
Spain	-10583,5	3362,728	-1239,87
Sweden	6503,261	-435,986	-471,835
UK	-5306,21	9095,34	3130,682
	CO2	CH4	N2O
EU-28	1898,963	13827,45	-53508,9

Πίνακας 10.1: Διαφορές των προβλέψεων της πολλαπλής γραμμικής παλινδρόμησης με τις προβλέψεις της παλινδρόμησης με τη χρήση των Panel Data

Παρατηρώντας τον Πίνακα 10.1, βλέπουμε ότι δεν υπάρχουν σημαντικές διαφορές ανάμεσα στις προβλέψεις της πολλαπλής γραμμικής παλινδρόμησης με τις προβλέψεις της παλινδρόμησης με τη χρήση panel data. Οι κύριες διαφορές παρατηρούνται στις προβλέψεις των εκπομπών του υποξειδίου του αζώτου, όπου με τη χρήση των panel data φαίνεται πως οι τιμές αυτές στις περισσότερες χώρες είναι πιο υψηλές σε σχέση με τις τιμές βάσει της πρώτης μεθόδου.

Όσον αφορά τις προβλέψεις για τις εκπομπές των αερίων του θερμοκηπίου στην ΕΕ-28 (Πίνακες 10.3 & 10.4), συμπεραίνουμε ότι οι εκπομπές του διοξειδίου του άνθρακα

το 2013 είχαν μειωθεί, όπως φαίνονται από τις προβλέψεις και των δύο μεθόδων. Οι προβλέψεις των εκπομπών των δύο άλλων αερίων του θερμοκηπίου, παρουσιάζεται μια διαφοροποίηση μεταξύ των δύο μεθόδων. Χρησιμοποιώντας την πολλαπλή γραμμική παλινδρόμηση οι εκπομπές του μεθανίου για το 2013 παρουσιάζουν αύξηση, ενώ χρησιμοποιώντας την παλινδρόμηση με panel data οι εκπομπές αυτές παρουσιάζουν μείωση. Βλέποντας την συμπεριφορά των εκπομπών του συγκεκριμένου αερίου στην ΕΕ-28 από το 1995-2012 (Πίνακας 10.2), καταλαβαίνουμε ότι είναι πιο φυσικό και αναμενόμενο το αποτέλεσμα με βάση τη χρήση των panel data, όπου οι εκπομπές μειώνονται. Επίσης, σύμφωνα με την Eurostat, η ταχύτητα με την οποία μειώνονται οι εκπομπές του μεθανίου στην ΕΕ-28 από το 1990 μέχρι το 2012 είναι 36,8%. Επομένως, καταλήγουμε στο να πιστεύουμε ότι η πρόβλεψη με τη χρήση των panel data, είναι πιο αξιόπιστη από ότι με την αρχική μέθοδο.

Επιπρόσθετα, όσον αφορά τις εκπομπές του υποξειδίου του αζώτου, παρατηρούμε ότι σύμφωνα με την μέθοδο της πολλαπλής γραμμικής παλινδρόμησης οι εκπομπές του μειώνονται ενώ με βάση τη δεύτερη μέθοδο παρουσιάζουν αύξηση. Βλέποντας τις εκπομπές της ΕΕ-28 από το 1995 μέχρι το 2012, παρατηρούμε ότι υπάρχει μείωση χρόνο με το χρόνο, με εξαίρεση το 1996 και το 2004. Σύμφωνα με την Eurostat, η ταχύτητα με την οποία οι εκπομπές του υποξειδίου του αζώτου μειώνονται είναι 33,7 %. Επομένως, σε αυτή την περίπτωση η αρχική μέθοδος έδωσε καλύτερες προβλέψεις.

ΕΕ-28	CO2	CH4	N2O
1995	4169432.30	545786.20	467557.83
1996	4270285.08	541792.49	473761.24
1997	4179310.81	530790.19	471658.15
1998	4175864.74	517075.33	449282.59
1999	4109800.09	505223.35	426519.76
2000	4135979.68	495133.32	423114.68
2001	4205763.60	484854.37	415931.85
2002	4179713.50	477945.70	405008.28
2003	4276585.73	468286.42	399854.61
2004	4286727.15	453096.28	404224.73
2005	4262335.86	444050.30	395168.82
2006	4274361.36	438177.88	382654.27
2007	4223517.54	430410.78	382361.18
2008	4122865.76	425262.14	373405.40
2009	3787629.14	415757.85	352348.19
2010	3907815.82	408801.12	343093.56
2011	3767424.03	401007.67	340891.73
2012	3717116.79	398233.64	333639.25

Πίνακας 10.2: Εκπομπές των τριών αερίων του θερμοκηπίου στην ΕΕ-28 (1995-2012)

	CO2	CH4	N2O
EU-28	3636883	401816,6	324569,8

Πίνακας 10.3: Προβλέψεις εκπομπών των τριών αερίων του θερμοκηπίου στην ΕΕ-28 για το 2013 με την πολλαπλή γραμμική παλινδρόμηση

	CO2	CH4	N2O
EU-28	3634984	387989,1	378078,7

Πίνακας 10.4: Προβλέψεις εκπομπών των τριών αερίων του θερμοκηπίου στην ΕΕ-28 για το 2013 με την χρήση των panel data

Παρόλο που η πρόβλεψη για τις εκπομπές του υποξειδίου του αζώτου δεν ήταν τόσο καλή, η ανάλυση παλινδρόμησης με χρήση panel data κρίνεται προτιμότερη. Η ανάλυση πολλαπλής γραμμικής παλινδρόμησης είναι πολύ χρονοβόρα, ειδικά όταν χρειάζεται να διαμορφωθούν πολλά μοντέλα. Ενώ με τη χρήση των panel data, η διαδικασία γίνεται ευκολότερη και χωρίς να έχει σημαντικές διαφορές από τα αποτελέσματα της πιο αναλυτικής μεθόδου. Επίσης, το γεγονός ότι με τη χρήση των panel data, λαμβάνονται υπόψη τα διαφορετικά χαρακτηριστικά μεταξύ των χωρών μας κάνει να πιστεύουμε ότι είναι καταλληλότερη μέθοδος για την ανάλυση αυτών των δεδομένων.

Όπως προαναφέρθηκε, η πρόβλεψη των εκπομπών του διοξειδίου του άνθρακα, του μεθανίου και του υποξειδίου του αζώτου είναι μεγάλης σημασίας. Γνωρίζοντας τους παράγοντες στους οποίους οφείλεται η αύξηση ή αντίστοιχα η μείωση των εκπομπών, μπορούμε να αντιμετωπίσουμε πιο αποτελεσματικά το πρόβλημα. Οι εκπομπές πολλών ατμοσφαιρικών ρύπων έχουν μειωθεί σημαντικά τις τελευταίες δεκαετίες με αποτέλεσμα τη βελτίωση της ποιότητας της ατμόσφαιρας. Ωστόσο, οι συγκεντρώσεις ατμοσφαιρικών ρύπων παραμένουν σε υπερβολικά υψηλό επίπεδο και τα προβλήματα ποιότητας του αέρα εξακολουθούν να υπάρχουν. Σημαντικό ποσοστό του ευρωπαϊκού πληθυσμού κατοικεί σε περιοχές, ιδιαίτερα σε πόλεις, όπου παρουσιάζονται υπερβάσεις στα πρότυπα ποιότητας του αέρα. Η ατμοσφαιρική ρύπανση είναι ένα πρόβλημα όλης της Ευρώπης, και αυτό γιατί οι ατμοσφαιρικοί ρύποι που εκλύονται σε μία χώρα ενδέχεται να μεταφερθούν στην ατμόσφαιρα και να επιδεινώσουν ή να καταστήσουν κακή την ποιότητα του αέρα σε μια άλλη περιοχή.

Στα πλαίσια αυτής της διπλωματικής εργασίας, μέσα από τις έρευνες και τις στατιστικές που έχω μελετήσει κατάλαβα πόσο σημαντική είναι η ποιότητα του αέρα στη ζωή μας. Ο Παγκόσμιος Οργανισμός Υγείας κατατάσσει την ατμοσφαιρική ρύπανση γενικότερα, και τα αιωρούμενα σωματίδια πιο συγκεκριμένα, τον μεγαλύτερο ενιαίο κίνδυνο περιβαλλοντικής υγείας στον κόσμο. Περίπου 600.000 θάνατοι στην Ευρώπη το 2012 οφείλονται στη ρύπανση του αέρα, η πλειοψηφία των οποίων (482.000) προκλήθηκε από την ατμοσφαιρική ρύπανση. Οι ζωές μας συνδέονται με το κλίμα. Μπορούμε να ευαισθητοποιηθούμε και να λάβουμε τα δικά μας μέτρα.