
Μοντέλο 1 (όχι όπως λέμε "πρώτο μοντέλο", αλλά όπως λέμε "προσχέδιο").

Κώστας Παπαδόπουλος (γιατί μία τέτοια αποτυχία χρειάζεται να μπορεί να αποδοθεί κάπου)

Πειραιάς, 2030

Contents

1	Περιγραφή μοντέλου	2
2	Περιορισμοί	2
3	Αναλυτική περιγραφή μοντέλου	2
4	Οδηγίες εκτέλεσης πειραμάτων	3
4.1	Κατέβασμα αρχείων	3
4.2	Εγκατάσταση βιβλιοθηκών	3
4.2.1	pip install	4
4.2.2	Βιβλιοθήκες	4
4.2.3	Βάση	4
4.2.4	Παραμετροποίηση	4
5	Πειράματα	5
6	Αποτελέσματα και σχολιασμός	5
6.1	Γραμμική παλινδρόμηση	5
6.2	Γραμμική παλινδρόμηση σε διαφορετικές περιόδους	8
6.3	Σύγκριση αποτελεσμάτων με τα πραγματικά	9
6.4	Κατά κεφαλή αποτελέσματα	10

1 Περιγραφή μοντέλου

Σε αυτό το πρώτο δοκιμαστικό μοντέλο πρόκειται να επιχειρήσουμε να μοιράσουμε τις δωρεάν άδειες μίας χρονιάς σε επίπεδο χωρών, προσπαθώντας να είμαστε κάπως "δίκαιοι". Θα χρησιμοποιήσουμε δεδομένα για τον πληθυσμό, την παραγωγή διοξειδίου του άνθρακα και το ακαθάριστο προϊόν της κάθε χώρας.

Αρχικά, υπολογίζουμε τον "συντελεστή αξιοποίησης διοξειδίου", ο οποίος δείχνει το πόσο καλά η κάθε χώρα μπορεί να αξιοποιήσει το διοξείδιο το οποίο παράγει στο να αυξήσει το ακαθάριστο προϊόν της. Με άλλα λόγια, αν μία χώρα παράγοντας παραπάνω διοξείδιο επιτυγχάνει να αυξήσει και το ακαθάριστο προϊόν της, αυτό σημαίνει πως το χρησιμοποιεί αρκετά αποδοτικά.

Στη συνέχεια πρόκειται να ορίσουμε ένα άνω όριο για κάθε χώρα, το οποίο να ορίζεται ως το γινόμενο των περσινών της αδειών επί ένα συντελεστή ο οποίος εξαρτάται από τη μεταβολή του GDP και τον πληθυσμό της χώρας.

Στη συνέχεια, τέλος το μόνο που μένει είναι να υπολογίσουμε την μοιρασιά, έχοντας υπολογίσει τους περιορισμούς μας. Αυτό γίνεται χρησιμοποιώντας γραμμικό προγραμματισμό.

2 Περιορισμοί

Αρχικά, πολλά από τα δεδομένα που έχουμε και πρόκειται να χρησιμοποιήσουμε δεν είναι ακριβή. Για παράδειγμα, τα δεδομένα τα οποία έχουν να κάνουν με το πρώην σοβιετικό μπλοκ είναι σε μεγάλο βαθμό εκτιμήσεις. Επίσης, για τη Σλοβακία μας λείπουν εντελώς κάποια δεδομένα.

Στο κομμάτι της γραμμικής παλινδρόμησης είναι που γίνεται πάρτι. Τα δύο αυτά δεδομένα δεν έχουν κάποια γραμμική σχέση μεταξύ τους. Το ακαθάριστο προϊόν μίας χώρας εξαρτάται από πολλούς παράγοντες. Επομένως, ο μόνος τρόπος να βρούμε μία σχέση μεταξύ τους, θα ήταν αν γνωρίζαμε πως αυτή η σχέση ήταν σταθερή και εάν είχαμε δεδομένα για πολλά χρόνια, ώστε όλοι οι άλλοι παράμετροι στατιστικά να μην επηρεάζουν το αποτέλεσμα. Κανένα από τα δύο κριτήρια δεν ισχύει, και αυτό φαίνεται στα αποτελέσματα και στο πόσο "κακή" βγαίνει η γραμμική παλινδρόμηση. Όμως, κάνοντας την υπόθεση πως όλες οι χώρες της Ευρώπης είχαν ένα συγκρίσιμο ρυθμό βελτίωσης και παίρνοντας 40 χρόνια εκπαίδευσης, ίσως αν μην είναι παντελώς παράλογο. Όσο όμως και να εθελotuφλώ, δεν είναι καλή ιδέα.

Ένας ακόμα περιορισμός ο οποίος προέκυψε εκ των υστέρων είναι το ότι συνειδητοποίησα πως είμαι επιεικώς άθλιος (ήλπιζα πως ήμουν μέτριος) στο να γράφω αναγνώσιμο κώδικα, επομένως απλά δεν μπορούσα καν να με κάνω debug. Αυτό το πρόβλημα ίσως ενισχύθηκε από την python η οποία με έκανε να πιστεύω πως έχω κάνει κάτι σωστά, ενώ εκείνη έκανε άσχετα πράγματα.

3 Αναλυτική περιγραφή μοντέλου

Ας ορίσουμε τα παρακάτω:

- N Το σύνολο των χωρών που συμμετέχουν και i η κάθε χώρα που συμμετέχει στο ETS. Η Σλοβακία έμεινε εκτός γιατί δεν έχω όλα τα data.
- A το συνολικό allocation free αδειών για αυτή τη χρονιά.
- A_i το ζητούμενο allocation της κάθε χώρας.
- $slope_i$ Ο συντελεστής αξιοποίησης διοξειδίου, δηλαδή την κλίση της γραμμικής παλινδρόμησης μεταξύ GDP και BAU emissions.

- $free_i$ free allocation of last year of country i
- $free_{x,i}$ free allocation of last year of country i of year x
- $GDP_{x,i}$ GDP per capita of country i of year x
- $\overline{GDP_x}$ average GDP of year x
- pop_i population of country i
- $Utility_i = slope_i \times A_i$
- $W4_i$ ο συντελεστής που επηρεάζει το άνω όριο των αδειών που μπορεί να πάρει μία χώρα.

Το πρόβλημα γραμμικού προγραμματισμού το οποίο επιχειρούμε να λύσουμε είναι το:

- $maximize(\sum_{i \in N} Utilities_i)$
- c.t. $A_i \leq W4_i \times free_i$
- $A_i \geq free_i \times 0.5$
- $\sum_{i \in N} A_i = A$

Ο υπολογισμός του $w4$ γίνεται με τον παρακάτω τρόπο:

$$K_i = (GDP_{x,i} - \overline{GDP_x}) - (GDP_{x-1,i} - \overline{GDP_{x-1}})$$

$$w4_i = \frac{\frac{pop_i}{max(pop_j)} - \frac{K_i}{max(K_j)}}{4} + 1, \forall j \in N$$

4 Οδηγίες εκτέλεσης πειραμάτων

Αυτό δεν έχει γίνει πολύ αποδοτικά μέχρι τώρα...

4.1 Κατέβασμα αρχείων

Θα χρειαστεί να κατεβάσετε τον κώδικα ο οποίος βρίσκεται εδώ: [Diplomatiki kwrap model](#)

Μαζί με τον κώδικα, θα υπάρχουν και δύο αρχεία csv τα οποία περιλαμβάνουν τα δεδομένα για τον πληθυσμό και για το GDP διαφόρων χωρών. Τα αρχεία αυτά θα πρέπει να τοποθετηθούν στον ίδιο φάκελο μαζί με τον κώδικα.

4.2 Εγκατάσταση βιβλιοθηκών

Για την εκτέλεση του αρχείου χρειάζεται python 3.7, μαζί με τα πακέτα matplotlib, scipy, pandas, numpy και mysql.connector

Αν χρησιμοποιείτε το notebook, τότε αυτές οι βιβλιοθήκες θα μουν αυτόματα. Για να συμβεί αυτό θα πρέπει να χρησιμοποιείτε python 3.7, αλλιώς θα αποτύχει η εγκατάσταση. Αυτές είναι οι εντολές για να τα εγκαταστήσετε στον υπολογιστή σας, εάν δεν χρησιμοποιήσετε το notebook:

4.2.1 pip install

Κατεβάζουμε το script get-pip.py από το: <https://bootstrap.pypa.io/get-pip.py> και το αποθηκεύουμε στον ίδιο φάκελο που βρίσκεται η εγκατάσταση της python. Στην συνέχεια εκτελούμε την παρακάτω εντολή για να εκτελεστεί το script.

```
python get-pip.py
```

4.2.2 Βιβλιοθήκες

Έπειτα, μπορείτε να τρέξετε τις παρακάτω εντολές για να εγκαταστήσετε τις υπόλοιπες βιβλιοθήκες:

```
pip install numpy
pip install pandas
pip install matplotlib
pip install scipy
pip install mysql.connector
```

4.2.3 Βάση

Για να τρέξει αυτό θα χρειαστεί την βάση του ets εγκατεστημένη σε mysql στον υπολογιστή, καθώς από εκεί αντλεί τα δεδομένα για την κατανάλωση (και για τα ονόματα των χωρών).

4.2.4 Παραμετροποίηση

Αφού έχει συμβεί αυτό, μπορείτε να ανοίξετε το notebook ή το αρχείο python και να αλλάξετε την τιμή των μεταβλητών στην αρχή του προγράμματος:

```
# DEFINITIONS

# DATABASE
db='eu_ets '
use='root '
passwor=''
hos='localhost '

# Diagrams of data
Need_diagrams_of_random_data = False
Countrys_number_for_diagrams = 10 # Country for data check
# 0 Austria, 1 Belgium, 2 Bulgaria, 3 Croatia, 4 Cyprus,
# 5 Czech Republic, 6 Denmark, 7 Estonia, 8 Finland, 9 France,
# 10 Germany, 11 Greece, 12 Hungary, 13 Ireland, 14 Italy,
# 15 Latvia, 16 Lithuania, 17 Luxembourg, 18 Malta, 19 Netherlands,
# 20 Poland, 21 Portugal, 22 Romania, 23 Slovenia, 24 Spain,
# 25 Sweden, 26 United Kingdom

# Training of the linear regration of GDP over CO2 emissions
```

```
# Data available from 1960 to 2020
first_year_of_calculation = 1960
last_year_of_calculation = 2020
Need_diagrams_of_linear_regration = False # True if you want
to see the diagrams of the linear regression

Prediction_year = 2018 # must be between 2005 and 2018
```

Το πρώτο κομμάτι αφορά στη βάση και τα credentials που απαιτούνται για να συνδεθεί κάποιος στη βάση.

Το δεύτερο κομμάτι έχει να κάνει έναν μικρό έλεγχο για το αν κατάφερε το πρόγραμμα να διαβάσει τα csv.

Το τρίτο μέρος έχει να κάνει με την περίοδο για την οποία θα γίνει η εκπαίδευση της γραμμικής παλινδρόμησης και το έτος για το οποίο θα γίνει η πρόβλεψη.

5 Πειράματα

Ακόμα και με αυτό το μοντέλο μπορούμε να κάνουμε διάφορα μικρά πράγματα για να βγάλουμε κάποιο συμπέρασμα.

1. Σχολιασμός του κατά πόσο είναι λογικά τα αποτελέσματα της γραμμικής παλινδρόμησης.
2. Σύγκριση συντελεστών αξιοποίησης διοξειδίου με διαφορετικές χρονιές εκπαίδευσης.
3. Σύγκριση αποτελεσμάτων σε σχέση με την πραγματική απόδοση που προέκυψε την επόμενη χρονιά.
4. Σχολιασμός των ανά κεφαλή αποτελεσμάτων.
5. Σύγκριση αποτελεσμάτων για άλλες χρονιές.

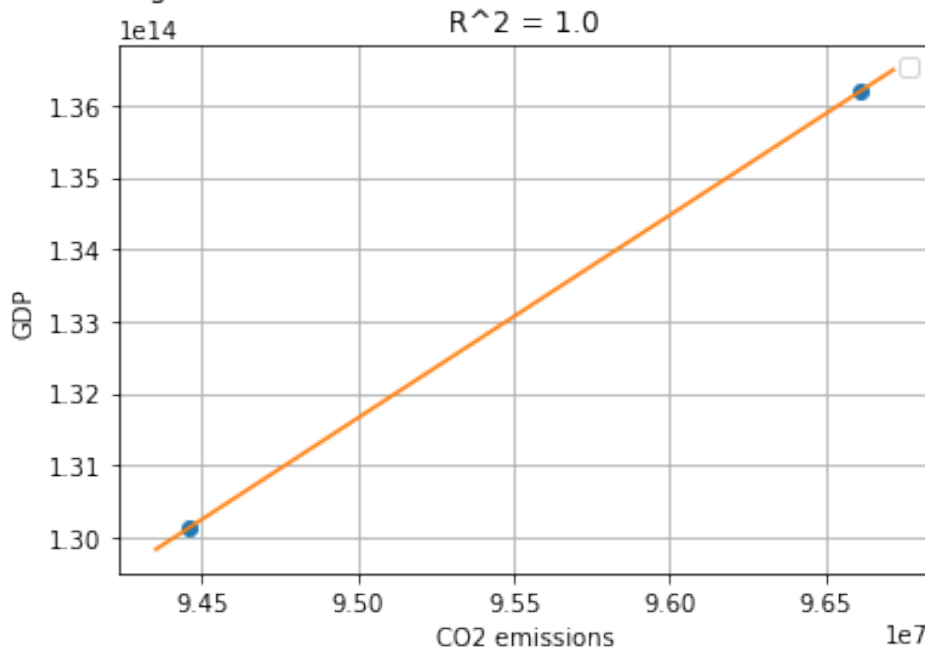
Το πρώτο πείραμα που μπορούμε να κάνουμε είναι το να δούμε τι συμβαίνει όταν αλλάζουμε τα έτη στα οποία "εκπαιδεύεται" η γραμμική παλινδρόμηση για να δούμε καθαρά και μόνο εάν μας φαίνεται λογικό το αποτέλεσμά της. Επίσης, είναι ενδιαφέρον και μόνο το να δούμε εάν η υπόθεση το ότι μέσα σε βάθος χρόνου θα ξεκαθαρίζει λίγο το πεδίο. Τα υπόλοιπα νομίζω είναι αρκετά λογικά.

6 Αποτελέσματα και σχολιασμός

6.1 Γραμμική παλινδρόμηση

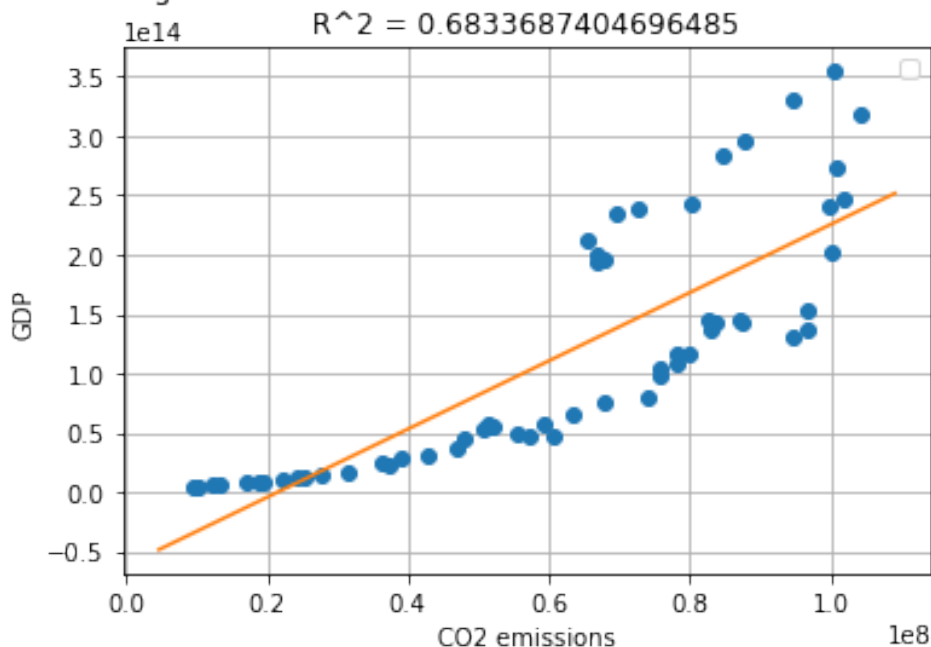
Η παλινδρόμηση που θα κάνουμε ουσιαστικά ψάχνει την καλύτερη ευθεία η οποία αντιπροσωπεύει τα στοιχεία τα οποία ορίζονται από τα σημεία (Διοξείδιο του Άνθρακα, GDP) για κάθε χρονιά, εδώ βλέπουμε το αποτέλεσμα μόνο για 2 σημεία.

Linear Regression of CO2 emissions and GDP for Greece in 2000-2001



Το επόμενο βήμα είναι το να δούμε πώς τα πηγαίνει αν του βάλουμε όλα τα δεδομένα, όπου εμφανώς υπάρχει πρόβλημα καθώς σε πολλές χώρες έχουμε λίγο περίεργα δεδομένα:

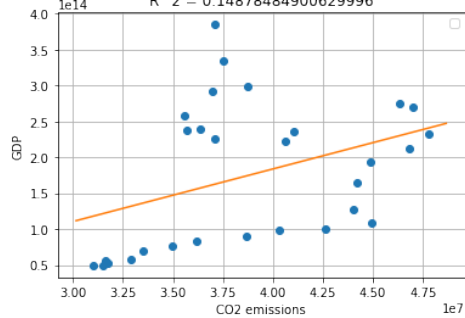
Linear Regression of CO2 emissions and GDP for Greece in 1960-2018



Οπότε αναγκαστικά θα πρέπει να αξιοποιήσουμε μικρότερο μέρος του data set. Κοιτώντας τα αποτελέσματα, η περίοδος 1990-2018 δείχνει αρκετά καλή. Παρακάτω βλέπετε την χειρότερη και την καλύτερη χώρα ως προς το πόσο καλή είναι η παλινδρόμηση.

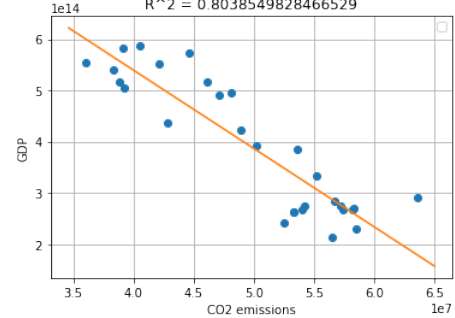
Το συνολικό αποτέλεσμα είναι το παρακάτω, αφού πρώτα κανονικοποιηθεί στο διάστημα $(-1, 1)$ και πολλαπλασιαστεί με -1 , επειδή το πακέτο του simplex από default λύνει προβλήματα

Linear Regression of CO2 emissions and GDP for Ireland in 1990-2018
 $R^2 = 0.14878484900629996$



(a)

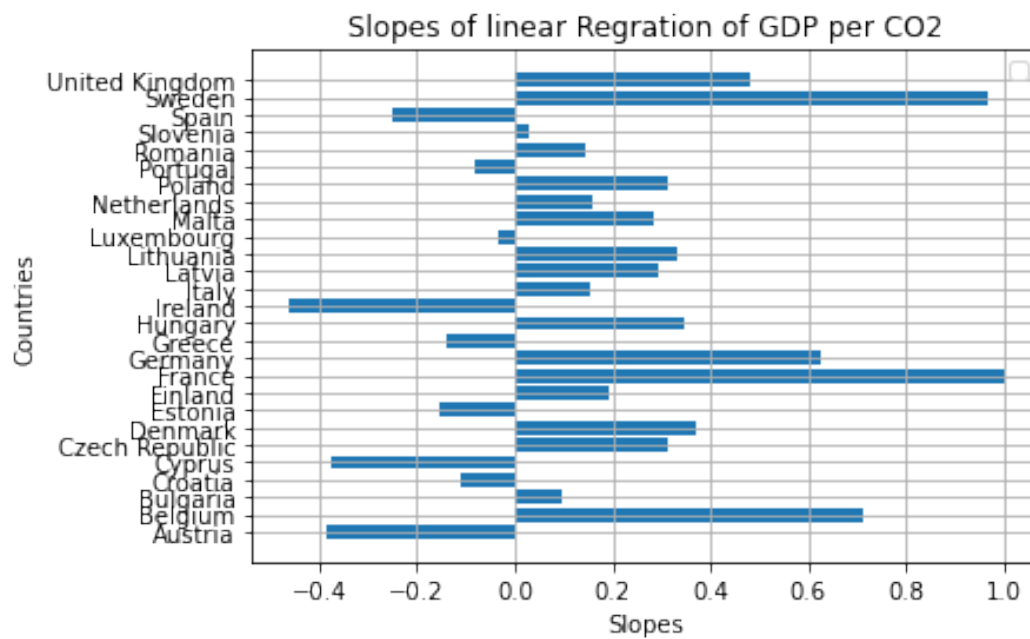
Linear Regression of CO2 emissions and GDP for Sweden in 1990-2018
 $R^2 = 0.8038549828466529$



(b)

Figure 1: 1990-2018

ελαχιστοποίησης:



6.2 Γραμμική παλινδρόμηση σε διαφορετικές περιόδους

Εάν επαναλάβουμε την ίδια διαδικασία με διαφορετικά χρονικά όρια, βλέπουμε πως τα αποτελέσματα μας δεν είναι καθόλου σταθερά. Συγκεκριμένα προκύπτουν τα παρακάτω:

	Countries	1990-2018	1990-1995	1996-2000	2001-2005	2006-2010	2011-2018	STD
0	Austria	-0.386373	-0.300944	-1.000000	-0.177957	0.511066	-0.028052	0.487002
1	Belgium	0.710211	-0.981619	-0.375648	0.125047	0.398831	0.185486	0.495952
2	Bulgaria	0.095389	-0.054882	0.010018	-0.033222	0.076099	0.011437	0.044989
3	Croatia	-0.109171	NaN	-0.000162	-0.107769	0.008670	-0.090000	NaN
4	Cyprus	-0.374203	-0.271643	-0.037223	-0.103245	-0.649242	-0.116239	0.220688
5	Czech Republic	0.314417	0.076508	0.005662	0.002625	0.248875	-0.069284	0.108307
6	Denmark	0.370961	-0.137012	-0.079914	0.052500	0.420575	-0.055479	0.199911
7	Estonia	-0.153719	NaN	0.030850	-0.037361	-0.083944	0.005675	NaN
8	Finland	0.193489	0.017265	-0.003902	0.004660	0.290604	-0.081580	0.127382
9	France	1.000000	0.427059	0.013766	-1.000000	1.000000	-0.166314	0.662595
10	Germany	0.625841	0.935909	-0.507288	0.272008	0.322178	0.175570	0.459273
11	Greece	-0.138951	-0.612982	0.111079	-0.265703	-0.086015	-0.157174	0.239351
12	Hungary	0.345865	0.150863	-0.011981	-0.014956	0.008546	-0.077711	0.075682
13	Ireland	-0.463062	-0.876665	-0.282326	-0.341093	-0.419367	-1.000000	0.295237
14	Italy	0.155040	-0.549719	0.366593	-0.244901	0.088634	-0.133672	0.309038
15	Latvia	0.293974	NaN	0.082761	-0.273341	-0.031125	0.003893	NaN
16	Lithuania	0.333928	NaN	0.033125	-0.096499	-0.171961	0.034606	NaN
17	Luxembourg	-0.032425	0.272098	-0.023682	-0.077298	0.967725	0.063443	0.382620
18	Malta	0.283703	0.025373	0.037622	-0.080797	0.237315	0.100034	0.104394
19	Netherlands	0.159818	-1.000000	-0.122034	-0.406353	-0.150966	0.350360	0.440979
20	Poland	0.311858	0.528936	0.021453	-0.081648	0.206411	-0.059199	0.226773
21	Portugal	-0.081758	-0.494217	-0.019018	-0.029542	0.204510	0.016215	0.231010
22	Romania	0.141991	-0.018361	0.001348	-0.038991	0.041666	0.055850	0.035693
23	Slovenia	0.030192	NaN	-0.024271	-0.329005	-0.318765	-0.049929	NaN
24	Spain	-0.249049	-0.456168	0.028481	-0.137329	0.050248	-0.099686	0.181500
25	Sweden	0.965809	0.518429	-0.244299	0.141169	-0.211682	-0.151970	0.288357
26	United Kingdom	0.480917	0.399661	0.802238	-0.210460	-0.696836	0.024668	0.512525

Table 1: Slopes for different periods

6.3 Σύγκριση αποτελεσμάτων με τα πραγματικά

Σε αυτό το σημείο θα δούμε το πόσο διαφορετικά είναι τα αποτελέσματα της προσομοίωσης με τα πραγματικά. Συγκεκριμένα θα χρησιμοποιήσουμε τα δεδομένα του 2017 για να προσομοιώσουμε το 2018. Οπότε προκύπτουν τα παραπάνω. Θα δοκιμάσουμε επίσης να κάνουμε την προσομοίωση και χρησιμοποιώντας τον W4 και χρησιμοποιώντας μία στανταρ τιμή 1.5 x τα περσινά.

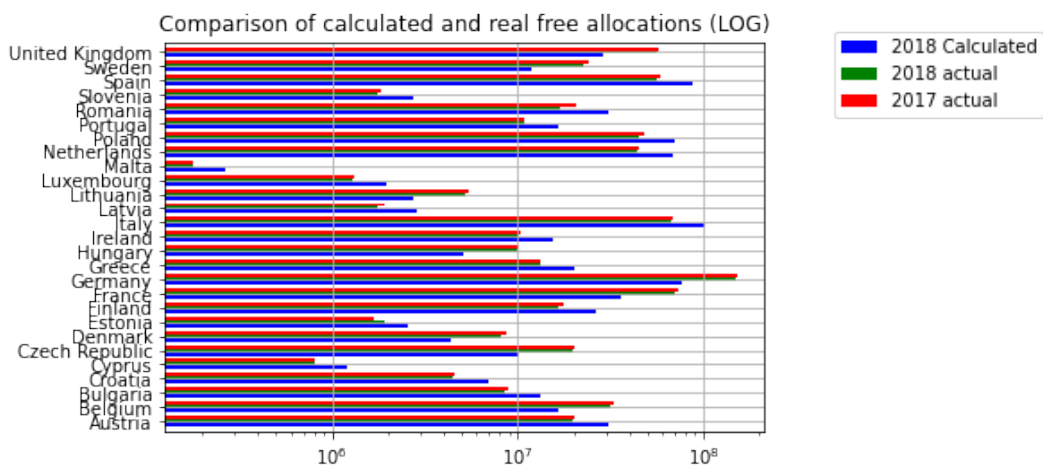


Figure 2: Χρησιμοποιώντας συντελεστή 1.5 για όλες τις χώρες

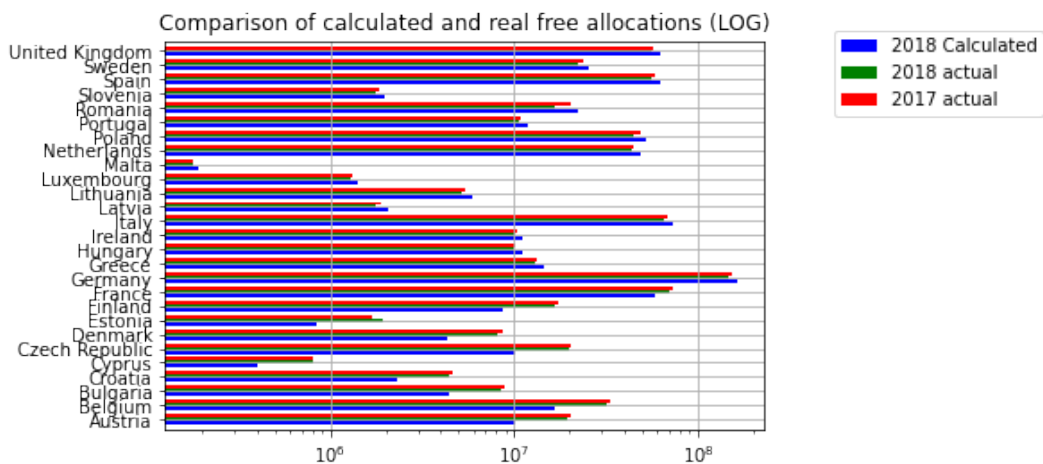


Figure 3: Χρησιμοποιώντας συντελεστή W4 για όλες τις χώρες

6.4 Κατά κεφαλή αποτελέσματα

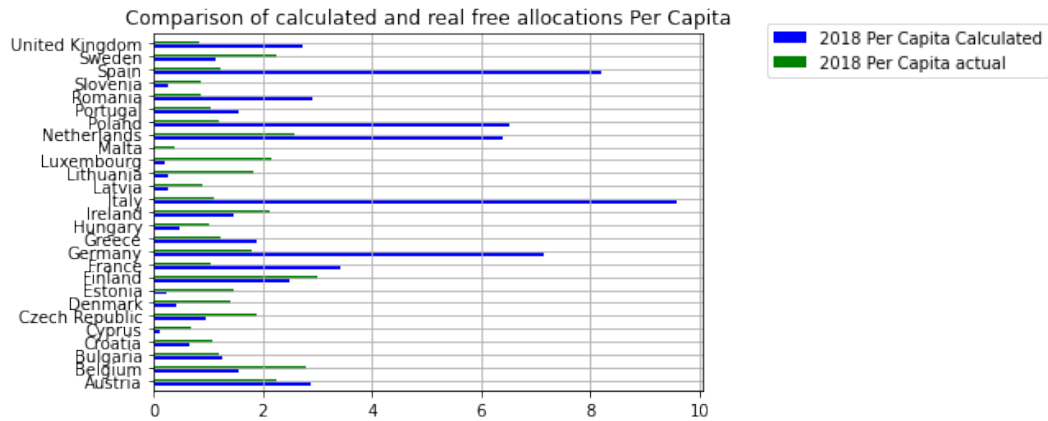


Figure 4: Χρησιμοποιώντας συντελεστή 1.5 για όλες τις χώρες

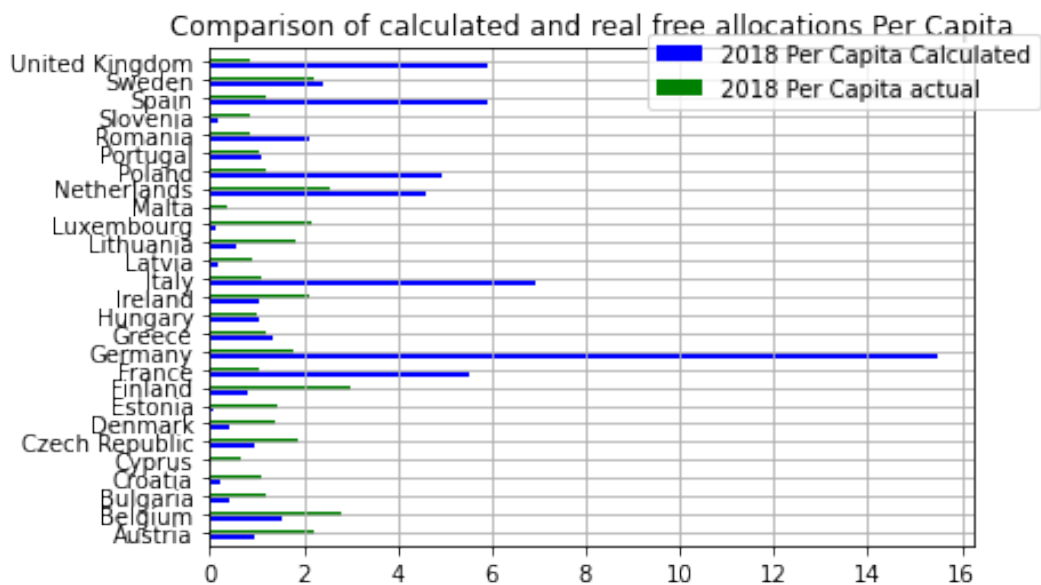


Figure 5: Χρησιμοποιώντας συντελεστή W4 για όλες τις χώρες