NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF ECONOMICS
MSC «**BUSSINESS ADMINISTRATION, ANALYTICS AND INFORMATION SYSTEMS**»

THESIS

Social Impact Analysis of Aging Population

By Leandros Johnson

Supervisor: Professor Ioannis C. Demetriou

Acknowledgement

I would like to thank Professor Ioannis Demetriou, from whom I was taught the concepts of Business Systems and Business Analytics during my postgraduate studies in the Department of Economics, for his most valuable assistance and guidance.

I would also like to thank my companion Maria Orlof, from whom I had emotional support and ideas to improve this survey.

Last but not least, I would like to thank my family here in Athens and in Ioannina, Greece, for their emotional and financial support. Specifically, I like to thank my sister Eugenia Papathemistokleous, which prompted me alongside my companion to continue in postgraduate studies and supported me all the way.

Table of Contents

List of Tables

List of Figures

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία αναφέρεται στο αντικείμενο της γήρανσης του πληθυσμού στην Ευρώπη. Είναι ένα αντικείμενο που απασχολεί κυρίως τις δυτικές οικονομίες του πλανήτη και ιδιαίτερα την Ευρώπη. Ο στόχος της εργασίας είναι να βρει υποκείμενες σχέσεις μεταξύ δεικτών, μέσω αναλυτικών μεθόδων και απώτερο σκοπό την διερμήνευσή τους. Για τον λόγο αυτό, επελέχθησαν οι χώρες Δανία, Γερμανία, Ιρλανδία, Ελλάδα, Ισπανία, Ολλανδία, Αυστρία, Πορτογαλία και Ηνωμένο Βασίλειο ως αντιπροσωπευτικό δείγμα της Ευρωπαϊκής Ένωσης. Αρχικά, δίνεται ένα σύντομο πλαίσιο λειτουργίας του συστήματος συνταξιοδότησης κάθε χώρας. Έπειτα δίνονται οι περιγραφές των δεικτών που θα μας απασχολήσουν. Οι δείκτες περιλαμβάνονται στα πλαίσια της δημογραφίας, της οικονομίας και της θνησιμότητας για να υπάρξει μια σφαιρική άποψη του φαινομένου. Υπάρχει πλούσιο περιεχόμενο από άποψη περιγραφικής στατιστικής ώστε ο αναγνώστης να αντιληφθεί την δομή κάθε δείκτη τοπικά. Στην συνέχεια ακολουθεί κλασσική ανάλυση παλινδρόμησης, ελέγχοντας τις βασικές της υποθέσεις είτε μέσω στατιστικών δεικτών, είτε μέσω οπτικοποίησης διαγραμμάτων. Ακολουθεί η μέθοδος συσταδοποίησης εκεί όπου παρατηρούνται καμπυλότητες προκειμένου να διαχωρίσουμε τις περιόδους με σημεία καμπής και να βρούμε αντιπροσωπευτικές τιμές. Δίνεται ιδιαίτερη έμφαση στην Ελλάδα, καθώς είναι η χώρα που υπέστη τις μεγαλύτερες συνέπειες από την κρίση των χρηματοπιστωτικών αγορών του 2008 και εντέλει στην επικείμενη κρίση δημοσίου χρέους της Ευρώπης , όπου και κατέληξε να είναι πρωταγωνίστρια. Τα εργαλεία που θα αποφέρουν τα επιθυμητά αποτελέσματα είναι το Excel της Microsoft και το προγραμματιστό περιβάλλον για στατιστικούς σκοπούς R, διαμέσου του R studio.

ABSTRACT

The subject of the MSc. Thesis is the aging population of Europe. Aging is the subject that concerns mostly the west economies of the globe and especially Europe. We present a detailed survey to look for underlying relationships among indices and through analytical methods in order to get the ability of interpretation of the results. For this reason, countries from European Union were selected as representative sample and include Denmark, Germany, Ireland, Greece, Spain, Netherlands, Austria, Portugal and United Kingdom. At first, there is a brief overview of operations for each country's pension system. Then, the description of selected indices follows. The indices that will be examined are contained in demographics, economics and mortality as to receive a general overview of the phenomenon. There is a rich content considering the aspect of descriptive statistics in order for the reader to perceive the structure of each examined index locally. A classical regression analysis follows, testing its basic hypotheses either through certain statistical indices or through visual diagrams. Moreover, cluster analysis is deployed where nonlinearity is observed, as to distinguish periods with curving points and to receive representative values of examined indices. Special significance is given to Greece, because this country suffered the most of the consequences since the stock market crash back in 2008 and ended to become the protagonist in European public debt crisis. The tools that will produce the desired results are Excel from Microsoft and the programming environment for statistical learning R, through R - Studio.

Abbreviations

DK	DE	IE	EL	ES	NL	AT	PT	UK
Denmark	Germany	Ireland	Greece	Spain	Netherlands	Austria	Portugal	United Kingdom

CHAPTER 1

Presentation

1.1 Introduction

“In 2003, at the time I made my "Old Europe" comment, the centre of gravity in NATO and Europe had long since shifted to the East. With the former Warsaw Pact countries joining NATO, the alliance has a different mix today. Some people were sensitive about my comment because they thought it was a pejorative way of highlighting demographic realities. Apparently they felt it pointed a white light at a weakness in Europe - an aging population. Europe has come some distance since World War II in becoming Europe.”

By Donald Rumsfeld

It is a fact that the world is changing faster than the past. Though the immediate question is how we are going to support and embrace those changes. Especially for Europe, besides the future of European Union, the subject of an aging population raises serious questions. Aristotle posited “that it is natural for the body to reach its prime around age 35 and to ‘advance’ until about age 50” (Spengler, 1980). Nowadays advance of technology and medical practices have given us a tremendous amount of longevity in life expectancy. On the other end of the spectrum we have a steady fall in fertility rates. Thus the elder community rises as a part of population that needs attendance and care, just like children. In addition we recognise the baby boom of the post-world war II generation, alongside with the increase in life expectancy will complete the aggregation in elderly community.

Moreover, we expect a surge in healthcare consumption due to extensive treatment of the so called baby boom generation. Not to mention, financial issues are entering the stage as we observe the growth of the pension beneficiaries. In most European countries exists mandatory funded pension systems under the management of the corresponding governments. Funded systems under voluntary basis give alternative solutions, though their growth is limited. In most cases, the adaptation of the liquidity of the pension systems comes in a form of change in retiring age or alterations in nominal increases of the amounts given, though those policy reforms are opposed by the people.

Especially for Greece, there is the matter to sustain economic stability and moreover growth through investments. However the diminishing of the working population undermines the economic recovery of the bailout program from European Commission, European Central Bank and International Monetary Fund. (Traa, 2018)

1.2 Pension systems in Europe

Looking at an overview of the pension systems, in Denmark there is a three pillar system. The Nordic country has 1) a state pension, 2) semi-mandatory occupational pensions and 3) personal savings pension. The basic amount from the first pillar is means-tested and the rest supplements are fully funded (Jørgen, 2011).

Similar is the pension system in Germany, with 1) state retirement insurance with mandatory participation, 2) private company plans and 3) personal retirement investments. The first pillar is a redistributive model (Bucher-Koenen, 2011)

In Ireland there are three tiers in the pension system. First tier ensures a minimum standard of living. Second tier comprises of mandatory public or private savings system and the last is a private voluntary savings system. The public system gives to the beneficiaries either 1) a basic flat-rate to those that meet the contribution conditions or either 2) a means-tested pension to those that don't meet the criteria (OECD IE, 2014)

Pension system in Greece has a main compulsory pension provision with funds grouped in certain professions/occupations, an auxiliary provision and a social solidarity grant provision which covers residents of low income. From the main provision there is a basic means-tested pension eligible to those with less than 15 years of contributions. The full contributory period is 40 years. From the year 2021 the statutory age will be adjusted according to changes in life expectancy every three years. The 2010 legislation set a minimum age of retirement at 60 years increasing it up to two years in 2012 and penalties are implemented upon retirement earlier than the minimum (National Actuarial Authority of Greece, 2015).

The Spanish pension systems comprises of a main contributory and professional type, non-contributory and a complementary system. The contributory system is funded from each person as a pay as you go system. The non-contributory system is funded by general taxes and the complementary is privately funded. It is necessary to have 15 years of contributions to get qualification for a pension benefit and the threshold is 65 years with some penalty for early retirement (Innaculada, 2003).

In the Netherlands there is again a three main pillar pension system. The first is a flat-rate state pension (Algemene Ouderdomswet or abbreviated as AOW), the second is occupational schemes and last the individual savings schemes. The AOW state pension is funded from payroll taxes and is linked with the minimum wage. For each year the employee is insured they accrue 2% of the full basic benefit and the current statutory age of retirement is 66. If inhabitants have less than fifty years of Dutch residency and no assets or other source of support there is a means tested social-assistance scheme for them. (OECD NL, 2017)

Austria has only a main defined-benefit public system with different retirement age for sex. For men the statutory age is 65 years and for women is 60, gradually increasing to 65 between 2024 and 2033. The coverage condition to get a benefit is 15 years in the last 30 years of working life or 25 years in the full lifetime. In 2005 a legislation reform reduced the contribution years due to employment to seven and the remaining eight can be reached by e.g. child raising periods. Also the beneficiaries with low earnings receive a means-tested supplement (OECD AT, 2017).

Portugal has a public scheme. It is earnings-related with a means-tested benefit for those of low income. The statutory retirement age is 66 years and two months with adjustments by 2/3 of gains in life expectancy with 65 years as a base and measures the average of the previous two years. In addition, there is the Solidarity Supplement for the Elderly (SSE) aiming to fight poverty among old people. (OECD PT, 2017)

In the United Kingdom a three pillar system exists. The first pillar is the public pension which is separated through a basic state pension and the new State pension. It is a pay as you go scheme through National Insurance contributions and provides a minimum standard of living for tier one. Tier two is connected more to the employee's earnings. The second pillar is related with the private occupational pensions, its aim is the redistribution of an individual's lifetime earnings. The last pillar is about the private savings schemes in a voluntary basis. The statutory retirement age will be 66 years by October 2020 and will continue to rise (IOPS, 2017)

Overall, most pension systems rely upon the contributions of employees or tax revenue for a basic scheme. Legislation is adjusting to the changes of longevity but with slower pace.

This thesis has four more chapters. The first presents the indices that were selected to understand the phenomenon. Summary statistics are presented for each index and for each country. The main visualization tool is the classic box and whiskers plot. Next chapter contains a classical regression analysis with visual means and tables in order to test the basic hypotheses. The following chapter contains cluster analysis with visual figures to portray the separation of data. Last chapter summarizes the conclusions drawn from analysis.

CHAPTER 2

Data Summary

Chapter 2 presents which indices will be used and provides descriptive statistics for each index in local economies. Subchapter 2.1 gives the source of data alongside the description for every index used in this thesis. Subchapter 2.2 categorises the indices into three groups and then deploys box and whiskers plots to portray the main statistical structure of each examined index and for every examined country. Not to mention, comparison tables and diagrams are used to zoom out from local economies and receive a European view.

2.1 Data Description

In order to find connections that portray the impact of aging population we deploy this retrospective study of numerical data, selecting indices and finding elasticities in Greece.

In this chapter we provide an overview of data that are being used. Since the main theme is aging population, as an interesting demographic aspect we used the database from EUROSTAT alongside HELLENIC STATISTICAL AUTHORITY. The sample consists of nine European countries for the period 1991 until 2017 where data exists in each indicator. To approach the social implications connected with the elderly population the following indicators were selected:

I. Potential Support Ratio (PS_R) as the number of persons aged 15 to 64 per every person aged 65 or older. This index was calculated with the divination of the appropriate portions of population for each country respectively, recorded on 1st January each year.

II. Percentage of Aged Population (AP) for the persons aged 65 or over. This index was calculated in a similar way with the previous one.

III. Deaths of persons aged 65 or over (D) with specific causes. The causes include : a) diseases of the respiratory system, b) influenza (including swine flu), c) pneumonia, d) chronic lower respiratory diseases, e) asthma and status asthmaticus, f) other lower respiratory diseases and g) other diseases of the respiratory system (remainder of a)).

IV. Number of Heating Degree Days (HDD). The severity of the cold in a specific time period taking into consideration outdoor temperature and average room temperature (in other words the need for heating). The calculation of HDD relies on the base temperature, defined as the lowest daily mean air temperature not leading to indoor heating. The value of the base temperature depends in principle on several factors associated with the building and the surrounding environment.

By using a general climatological approach, the base temperature is set to a constant value of 15°C in the HDD calculation.

V. Pensions as percentage of Gross Domestic Product (PC_P) and especially about old age pension (partial or other form of pensions are excluded).

VI. Total Unemployment as percentage of active population (PC_TU).

VII. Unemployment of females as percentage of active female population (PC_FU).

VIII. Final consumption expenditure of households by consumption of Electricity, gas and other fuels (Egf_ME) in current prices, million euros.

IX. Final consumption expenditure of households by consumption of Health (H_ME) in current prices, million euros.

2.2 Descriptive Statistics

Summary statistics are presented in each aspect we study per country.

Demographics

The following indices were selected in order to get a glimpse of the population structure: Potential Support Ratio, percentage of Aging Population, Percentage of Total Unemployment and Percentage of Female Unemployment.

Economy

We have collected two indices about consumption in a country's economy. Considering the fact that elder people have increasing need for heat in the winter, coolness in the summer and electricity for medical appliances that are used indoors, we gathered data about electricity, gas and other fuels consumption. The elderly use more often health services, not only for treatments of chronic diseases or other illnesses, but also in form of often routine checks as preventive medical examination requires. This is why health consumption is highly relevant within our subject of study, though both indices refer to general consumption of the country and not just the elderly. We anticipate a strong connection among the fluctuations in both consumption indices and the percentage of aged people. The indices are recorded in million euros with changes in currency where appropriate in current prices.

A third important index refers to pensions. The form analysed is that of pensions as percentage of Gross Domestic Product in order to check the scale from the perspective of each nation's economy. The expectation is to find high correlation with the percentage of aging population. Bellow we present tables with summary statistics about consumption and pensions.

Mortality

In Eurostat there are mortality indices with many and various causes of death. Although the writer does not happen to have medical knowledge, the selection of indices is driven by a simple assumption. Cold weather may cause respiratory diseases. Adults have an adequate immune system and there is a variety of medication to fight diseases from flu to pneumonia. Yet elder people need more support in these situations because their immune system is weakening by the years. For the lack of hospitalized patients' data we selected the number of Deaths for persons aged 65 or over with all kinds of respiratory causes and the number of Heating Degree Days. The last index shows how many days in a country has cold climate with temperature below 15℃ thus leading to the need for indoor heat.

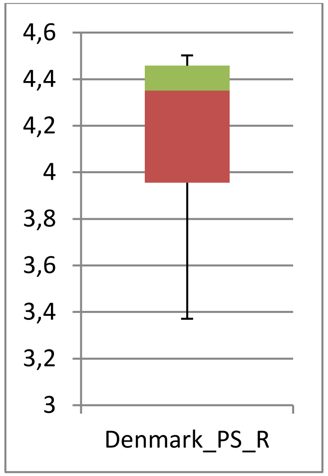


Figure 1: DK Potential Support Ratio

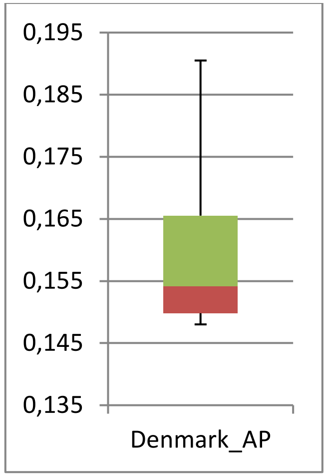


Figure 2: DK percentage of Aging Population

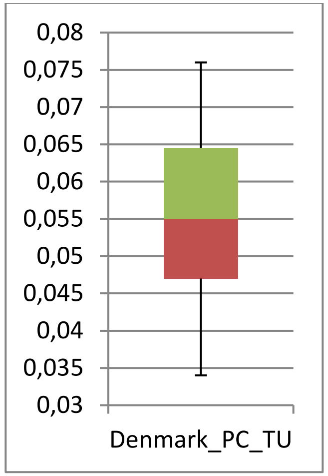


Figure 3: DK Percentage of Total Unemployment

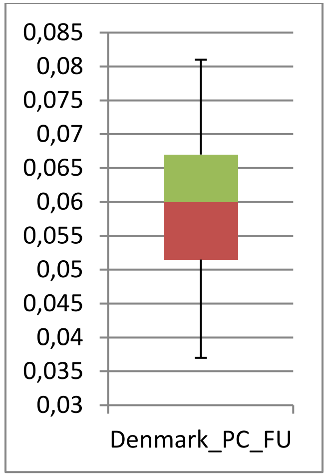


Figure 4: DK Percentage of Female Unemployment

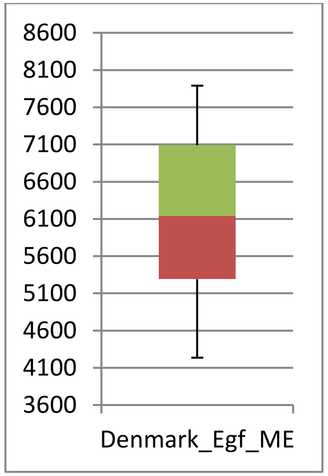


Figure 5: DK Electricity, gas and other Fuels consumption

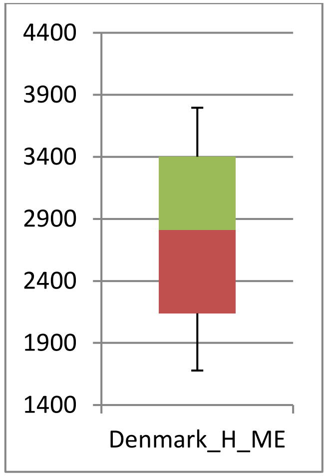


Figure 6: DK Health consumption

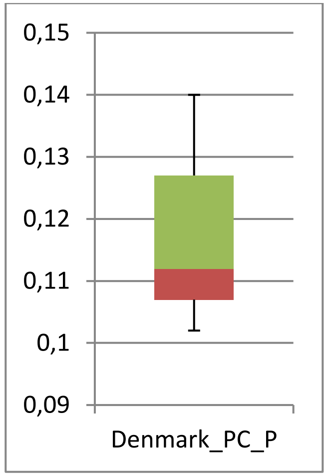


Figure 7: DK Pension percentage

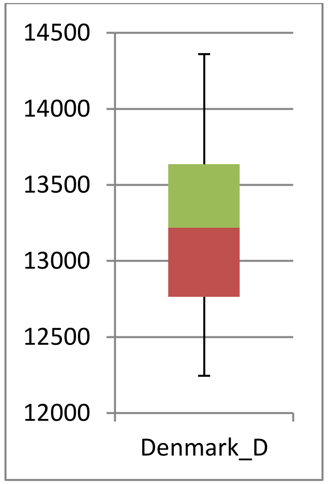


Figure 8: DK Deaths of old people from respiratory diseases

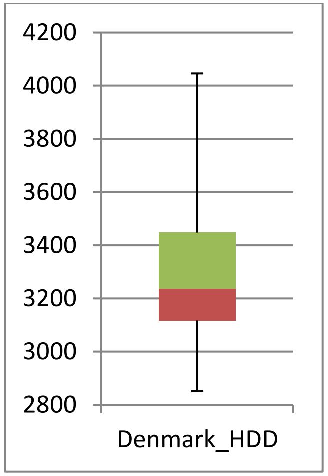


Figure 9: DK Number of Heating Degree Days

An average potential support ratio of 4.172 is observed, which means there are approximately 4 persons aged 15 to 64 for every aged person 65 or over. The aging population has an average of 16% and a maximum value of 19.1% proportionate to whole Denmark's population. Also both unemployment (total and female) are relatively low (0.056 and 0.060 respectively), but female unemployment tends to be higher (compare Figures 3 and 4).

In average, higher electricity and other fuels consumption is observed relatively to health consumption in Denmark, especially around 6 billion euros for the first and 2.8 billion euros for the second index. The mean percentage of pensions from GDP is 0.116. However, Figure 7 illustrates that the majority of data tend to be higher than mean, towards third and fourth quartile.

Denmark has in average 13 thousand deaths of aged people from respiratory diseases and three thousand heating degree days. It is noticeable how small the standard deviation of deaths in terms of scale is. Also, for two decades there are almost 280 thousand deaths of elders.

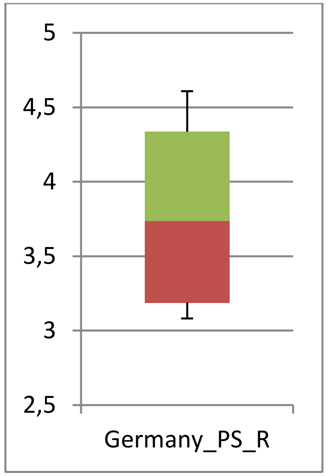


Figure 10: DE Potential Support Ratio

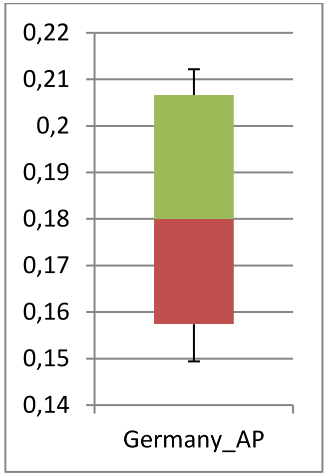


Figure 11: DE percentage of Aging Population

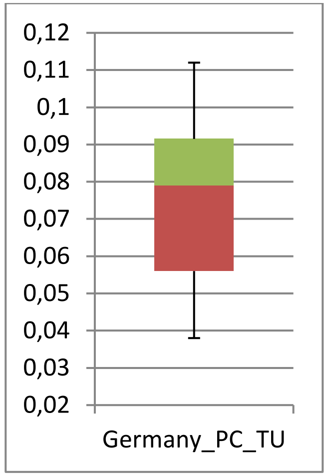


Figure 12: DE Percentage of Total Unemployment

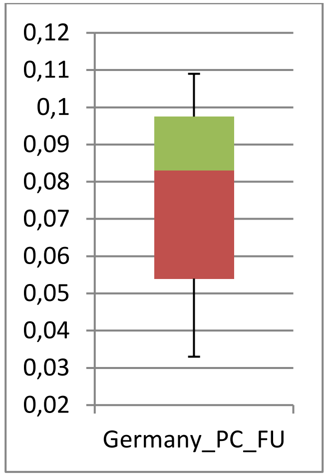


Figure 13: DE Percentage of Female Unemployment

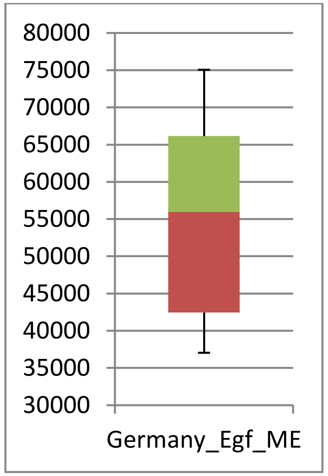


Figure 14: DE Electricity, gas and other Fuels consumption

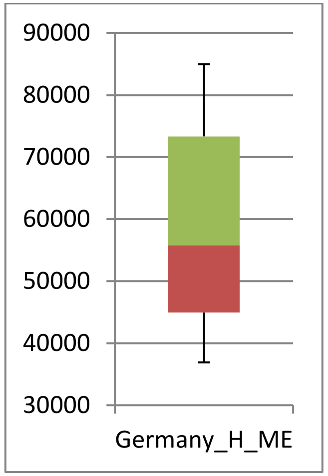


Figure 15: DE Health consumption

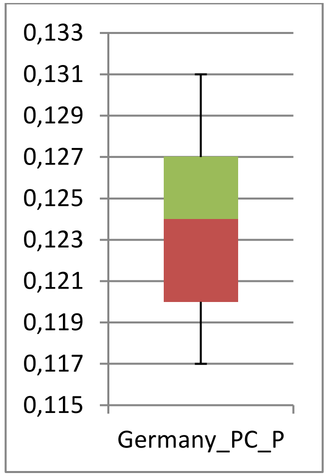


Figure 16: DE Pension percentage

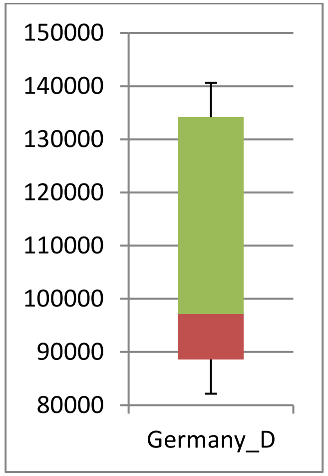


Figure 17: DE Deaths of old people from respiratory diseases

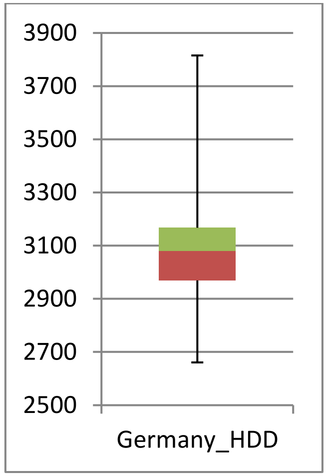


Figure 18: DE Number of Heating Degree Days

Europe’s strongest economy in nominal prices of gross domestic product, namely Germany (Fund, 2018), has an average potential support ratio of 3.778 and an average aging population of 18.1%. The maximum value of the last is 21.2% which is relatively high. On the other hand, total unemployment is on par with female unemployment with an average of 7.6% Germany differentiates.

Health consumption in Germany is approximately 4 billion euros higher than Electricity and other fuels consumption. Noticeable is also the standard deviation in Germany’s indices as we see a big spread in the y-axis at the box and whiskers plots (see Figures 14 and 15). There are also, the results of descriptive statistics for pensions in Germany. In Figure 16 the structure of box and whisker’s plot indicate an even deviation from the mean, which implicates in turn data that regress.

Germany seems to have a big population in elderly. In average we see almost 108 thousand deaths from respiratory diseases and the sum in two decades is approximately 2.3 million deaths. In addition a large part of mortality data belongs to the third quartile (see Figure 17). The number of heating degree days is in average three thousand.

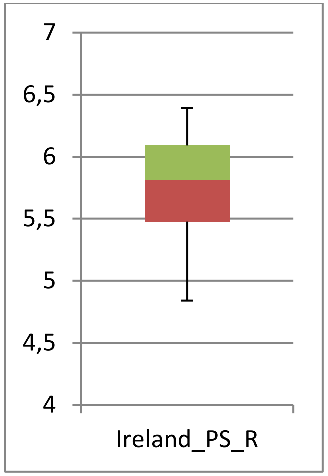


Figure 19: IE Potential Support Ratio

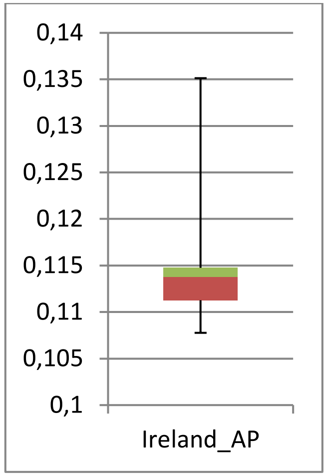


Figure 20: IE percentage of Aging Population

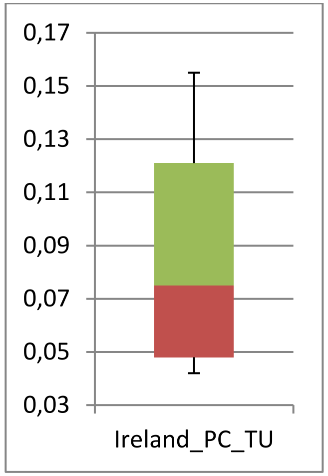


Figure 21: IE Percentage of Total Unemployment

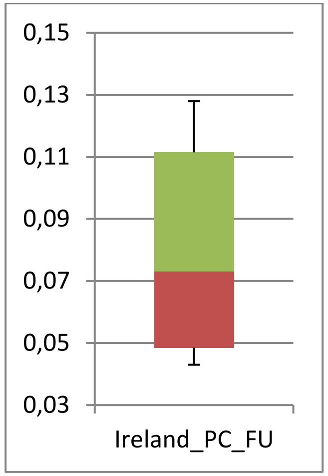


Figure 22: IE Percentage of Female Unemployment

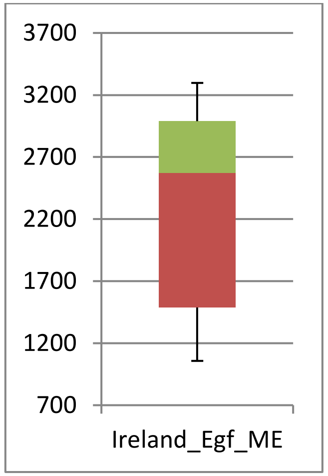


Figure 23: IE Electricity, gas and other Fuels

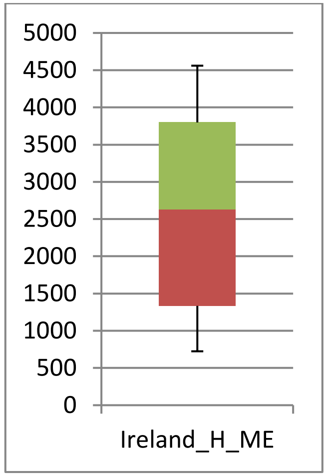


Figure 24: IE Health consumption

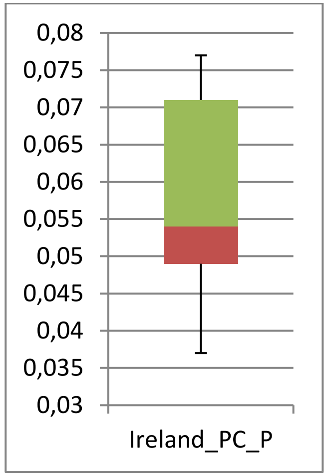


Figure 25: IE Pension percentage

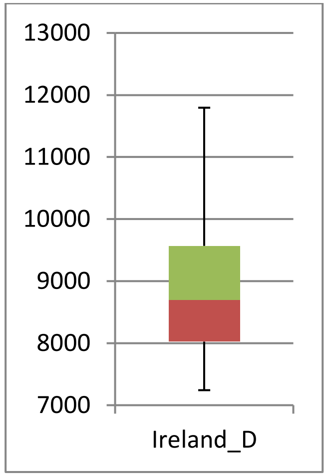


Figure 26: IE Deaths of old people from respiratory diseases

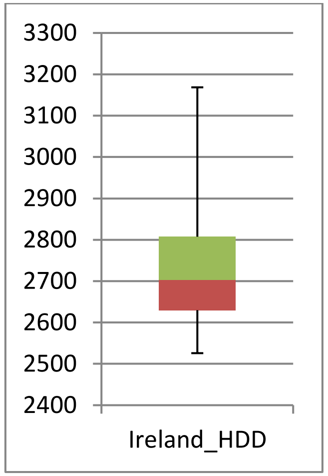


Figure 27: IE Number of Heating Degree Days

Ireland seems to be the champion of demographic statistics in Euro Area from the studied countries. The observed average potential support ratio is 5.73 so almost 6 persons aged 15 to 64 support the beneficiaries of pension system. In addition the elderly are relatively small to population with an average of 11.6% and maximum value of 13.5%. It is noticeable the maximum value of percentage of total unemployment, which is 15.5%.

Few things can be told about consumptions in Ireland. The average level of electricity and other fuels is approximately 2.3 billion euros and the level of health consumption is approximately 2.6 billion euros. More interesting is the comparison of the Figures 23 and 24, as we observe an almost same level of median, but the last two quartiles of health consumption is higher. This gives the impression of a higher trend relatively with the first consumption index. From the main summary statistics we can distinct the mean percentage of Pensions from GDP in Ireland, which is 5.7%. Although skewness is 0.366 and this can be seen also in Figure 25, where values aggregate in the third quartile.

In Ireland there are approximately 2,700 heating degree days in average, also almost nine thousand deaths of aged persons from respiratory diseases. The sum of deaths is 187 thousand for 21 years.

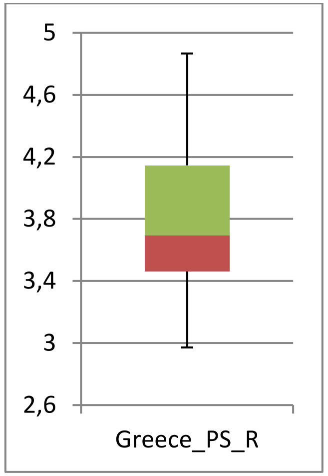


Figure 28: EL Potential Support Ratio

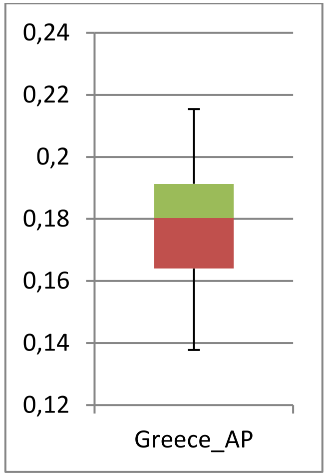


Figure 29: EL percentage of Aging Population

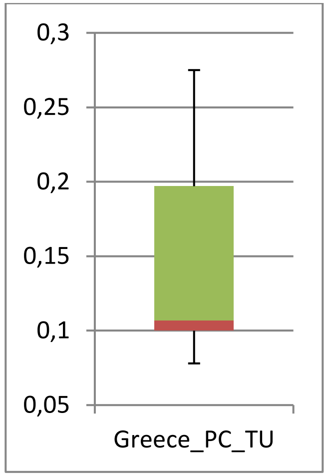


Figure 30: EL Percentage of
Total Unemployment

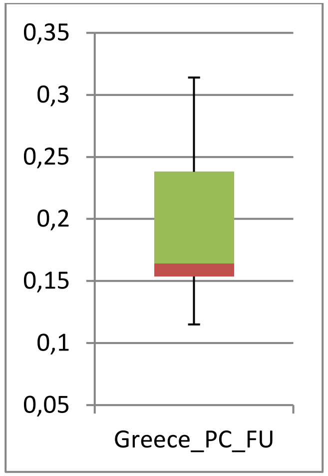


Figure 31: EL Percentage of Female Unemployment

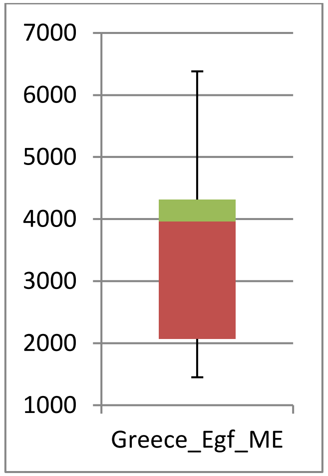


Figure 32: EL Electricity, gas and other Fuels consumption

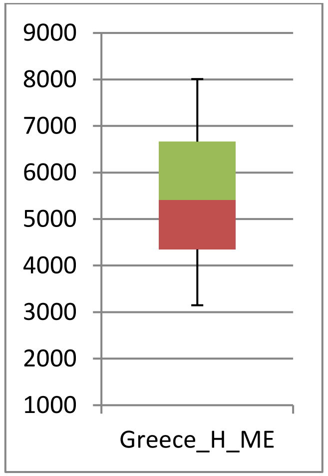


Figure 33: EL Health consumption

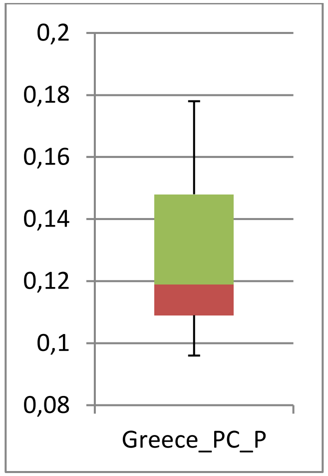


Figure 34: EL Pension percentage

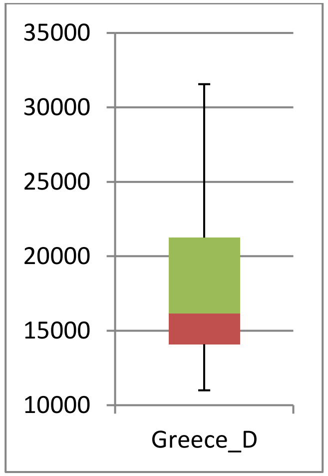


Figure 35: EL Deaths of old people from respiratory diseases

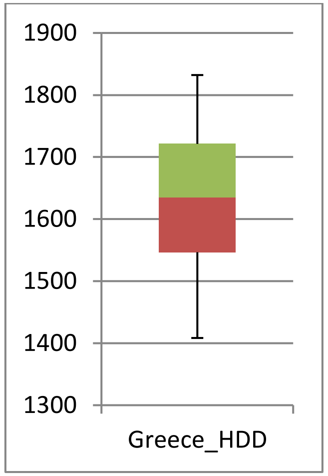


Figure 36: EL Number of Heating Degree Days

Greece has a high aging population index and small potential support ratio. The maximum value of first is 21.5% and the minimum of last is 2.972 and these represent data of year 2017. According to these observations around 3 persons aged 15 to 64 support more than one fifth of the population, excluding children under age of 15.

Health consumption in Greece is bigger than consumption of electricity, gas and other fuels as the average of first is around 5 and half billion euros and the second around 3.6 billion euros. There is one contradiction; they share almost the same standard deviation of one and a half billion euros. Greece has high percentage of pensions from GDP in average (0.129). Skewness is also high and this is portrayed in Figure 34 too, since the third and fourth quartiles contain a large part of data.

The average amount of deaths from respiratory diseases in elder people in Greece is 18 thousand and the average number of heating degree days is approximately 1.6 thousands. Not to mention that the sum of dead persons is 380 thousand for two decades, a relatively big amount connected with the heating degree days but is expected from the country with the highest percentage of aging population.

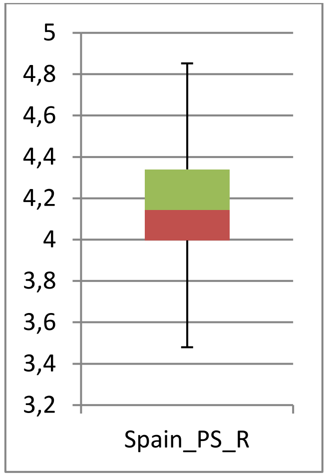


Figure 37: ES Potential Support Ratio

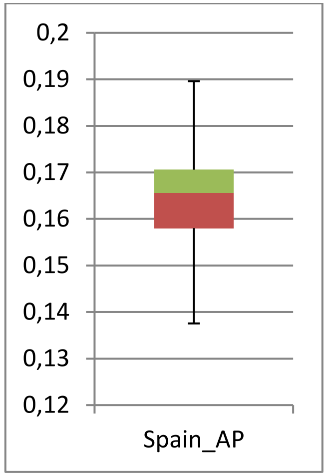


Figure 38: ES percentage of Aging Population

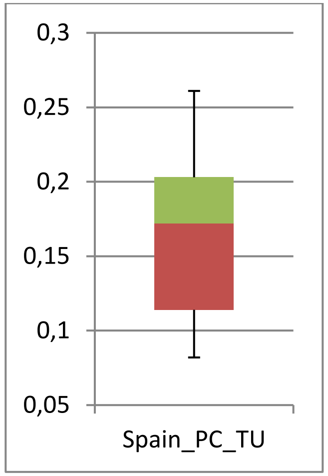


Figure 39: ES Percentage of Total Unemployment

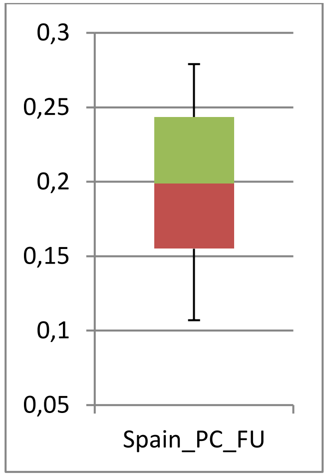


Figure 40: ES Percentage of Female Unemployment

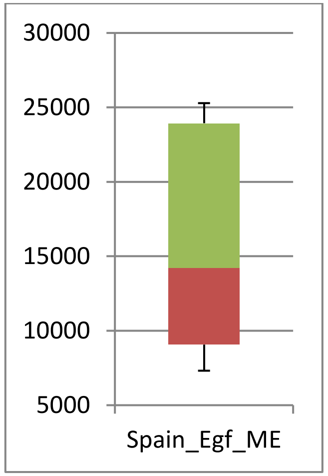


Figure 41: ES Electricity, gas and other Fuels consumption

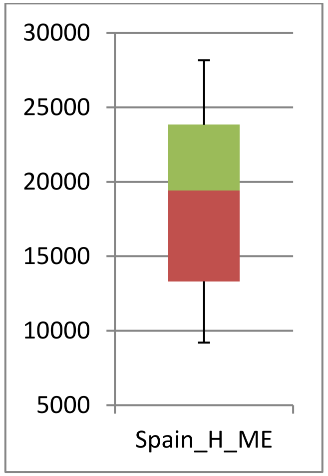


Figure 42: ES Health consumption

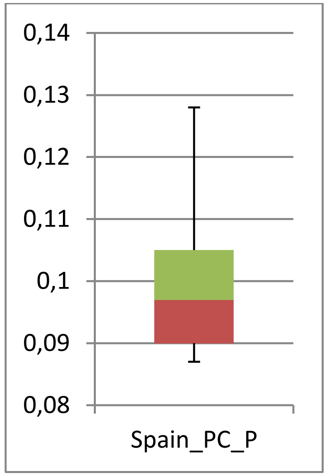


Figure 43: ES Pension percentage

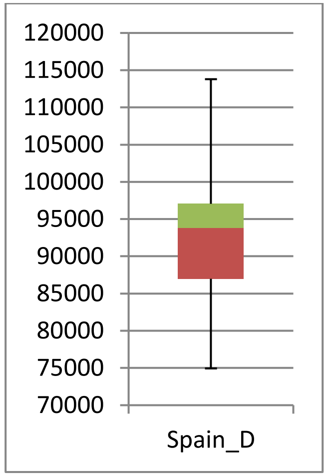


Figure 44: ES Deaths of old people from respiratory diseases

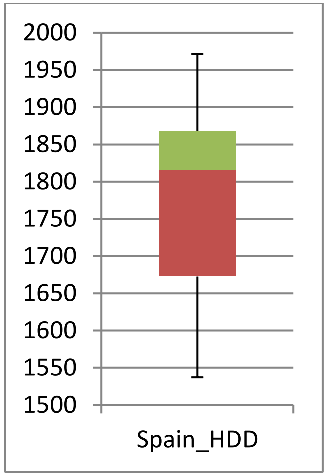


Figure 45: ES Number of Heating Degree Days

Spain’s demographic summary statistics show interesting values. The average of potential support ratio is 4.151 and of aging population is 16.5%. More interesting is the higher female unemployment when compared to total unemployment.

In average we see higher consumption in health than in Electricity and other fuels in Spain. Another observation we can make is that the standard deviation in Electricity consumption is higher indicating bigger changes in through time compared to health.The average percentage of pensions from GDP is 0.101 in Spain. Although is not as high as the previous examined values there is much skewness in data. The last can be illustrated in Figure 43, because maximum value has much distance from Q3.

Spain belongs to the Mediterranean so it is natural to expect low number of heating degree days as is seen in Greece. Here seems the average number is almost 1,800 days. Also the average number of deaths from respiratory causes is around 93 thousand and the sum is almost two million for 21 years.

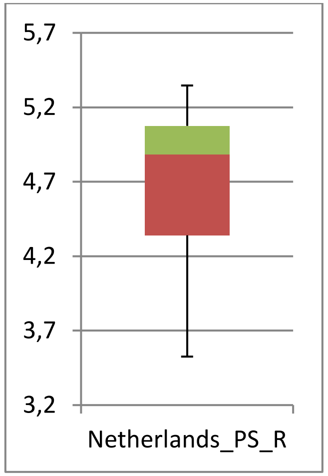


Figure 46: NL Potential Support Ratio

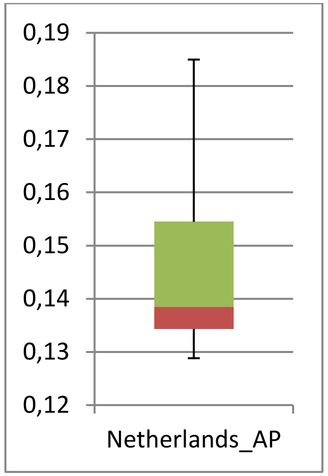


Figure 47: NL percentage of Aging Population

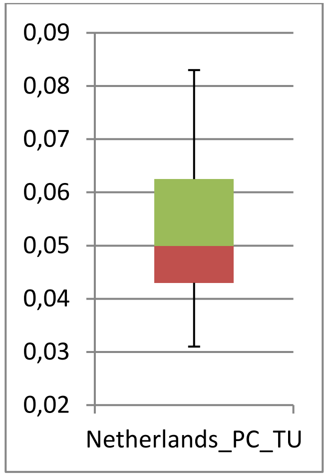


Figure 48: NL Percentage of Total Unemployment

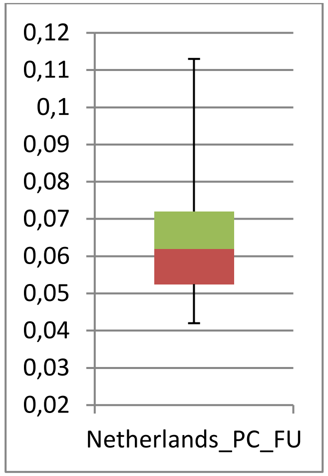


Figure 49: NL Percentage of Female Unemployment

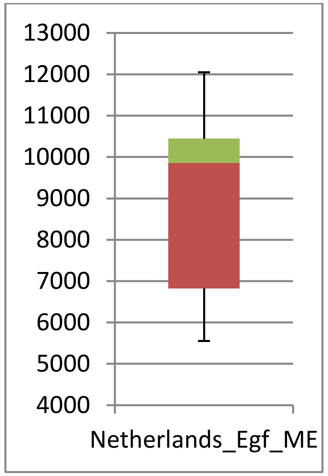


Figure 50: NL Electricity, gas and other Fuels consumption

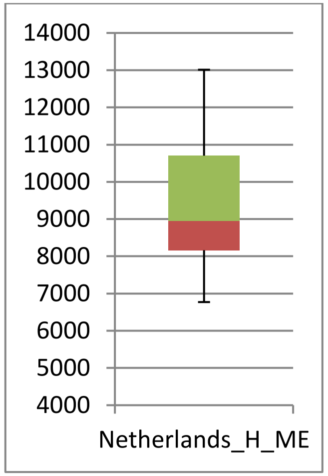


Figure 51: NL Health consumption

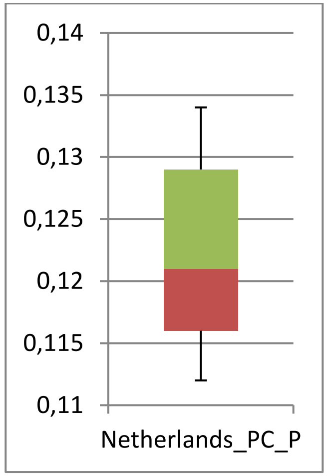


Figure 52: NL Pension percentage

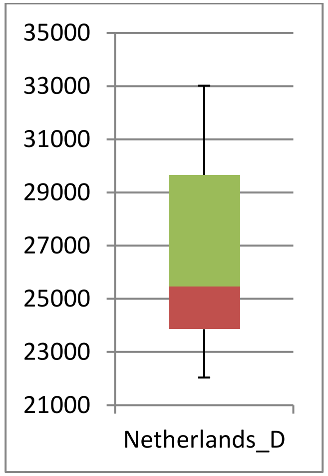


Figure 53: NL Deaths of old people from respiratory diseases

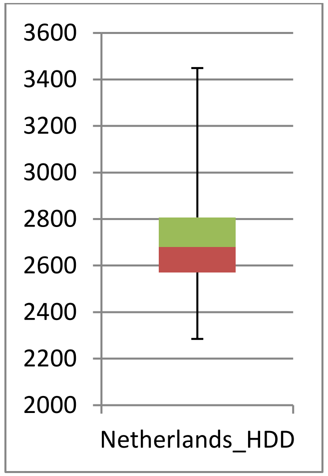


Figure 54: NL Number of Heating Degree Days

Netherland has better results in average, thus the potential support ratio is observed at the level of 4.66, aging population is 14.6%, total unemployment is at the level of 5.4% and female unemployment is at 6.5%.

Both consumptions in Netherlands are approximately on the same average level with the standard deviation in electricity and other fuels consumption being higher than health consumption. The difference in structure can be seen also in the Figures 50 and 51 where the interquartile of Electricity consumption is bigger and many data belong to the second quartile. The percentage of pensions from GDP has a large third quartile so the variations tend higher.

The Netherlands have in average 2.7 thousand heating degree days and 26.7 thousand deaths of persons aged 65 or over from respiratory diseases. The sum of deaths for two decades is 560 thousand.

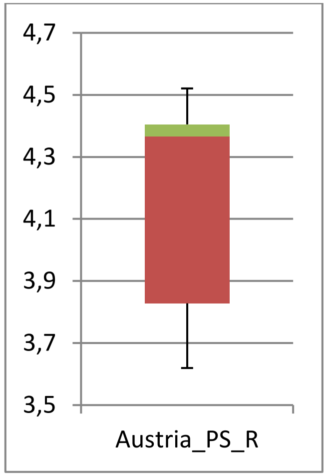


Figure 55: AT Potential Support Ratio

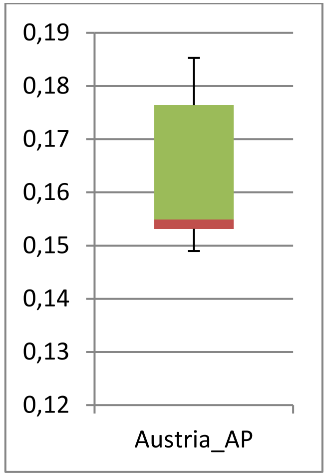


Figure 56 AT percentage of Aging Population

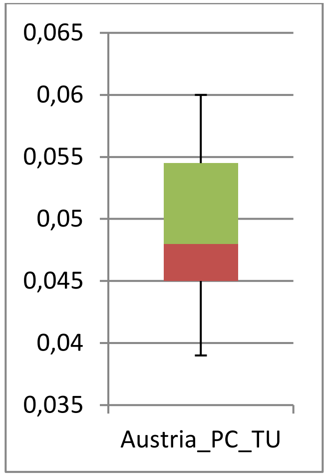


Figure 57: AT Percentage of
Total Unemployment

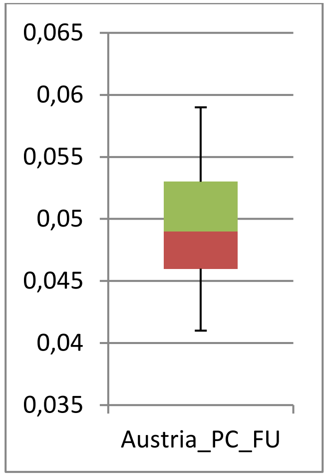


Figure 58: AT Percentage of Female Unemployment

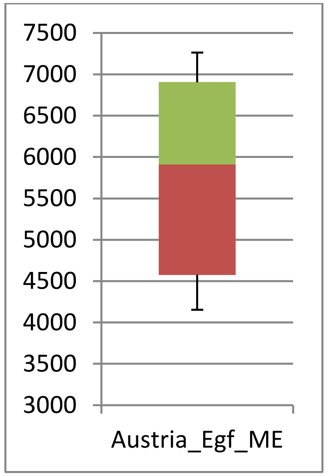


Figure 59: AT Electricity, gas and other Fuels consumption

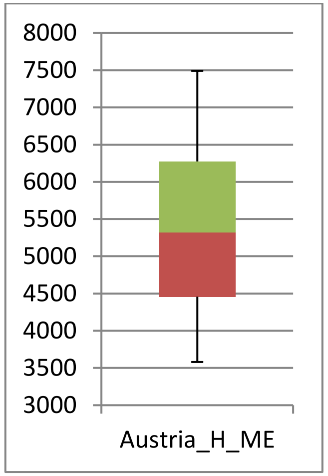


Figure 60: AT Health consumption

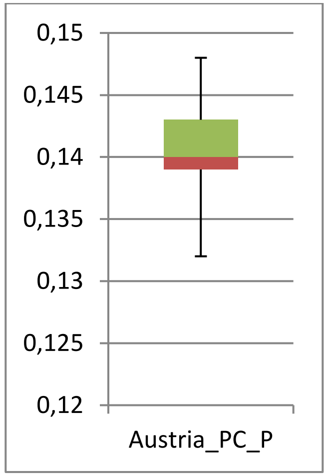


Figure 61: AT Pension percentage

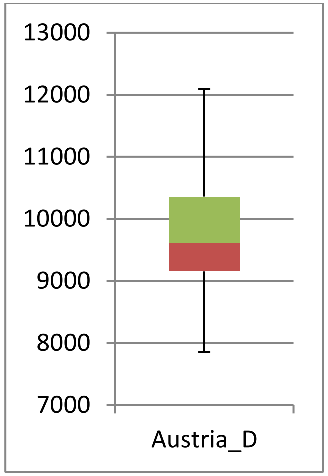


Figure 62: AT Deaths of old people from respiratory diseases

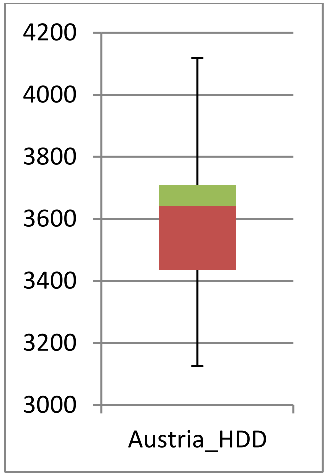


Figure 63: AT Number of Heating Degree Days

Austria has an average level of potential support ratio at 4.147, aging population at 16.4% and same average level of total and female unemployment at 4.9%, which are good numbers. Though the minimum value of first index is 3.62 and the maximum of the second is 18.5% the opposite direction of.

Austria has in average 5.5 billion euros both in health and electricity and other fuels consumption. Not to mention the overall structure of data is almost the same as is seen in Figures 59 and 60. The percentage of Pensions from GDP in Austria is 0.140 in average and Figure 61 illustrates the structure, which gives a narrow interquartile.

Austria has 3.6 thousand heating degree days in average and 9.6 thousand deaths of aged people with respiratory problems. The sum of dead persons is almost 202.5 thousand, a relatively small amount concerning the many heating degree days

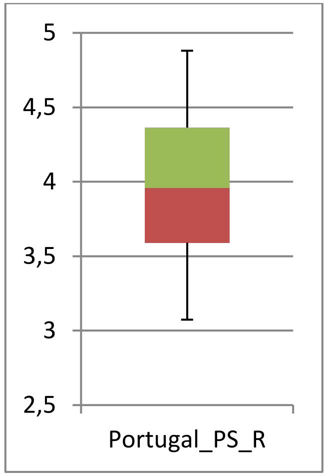


Figure 64: PT Potential Support Ratio

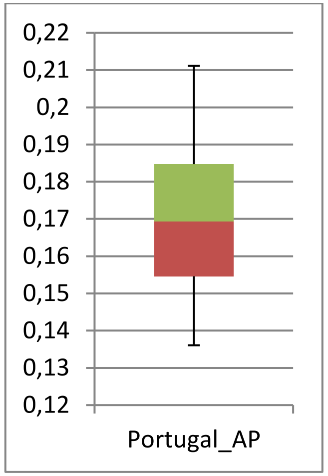


Figure 65: PT percentage of Aging Population

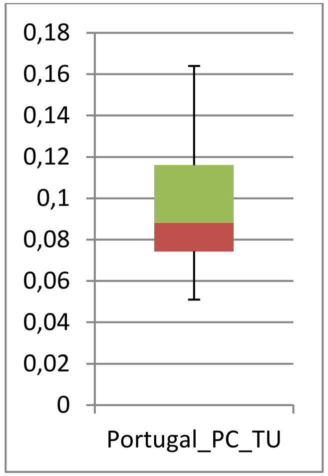


Figure 66: PT Percentage of Total Unemployment

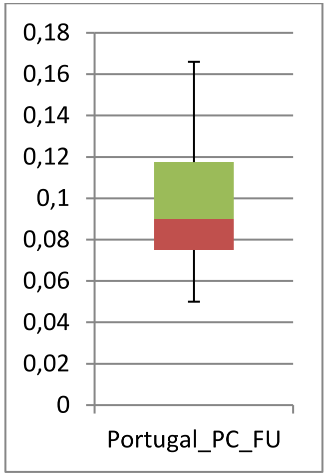


Figure 67: PT Percentage of Female Unemployment

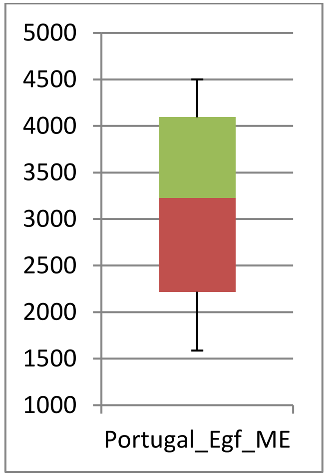


Figure 68: PT Electricity, gas and other Fuels consumption

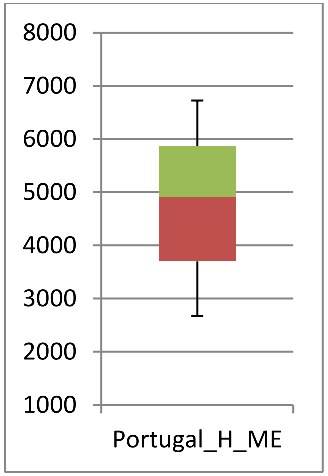


Figure 69: PT Health consumption

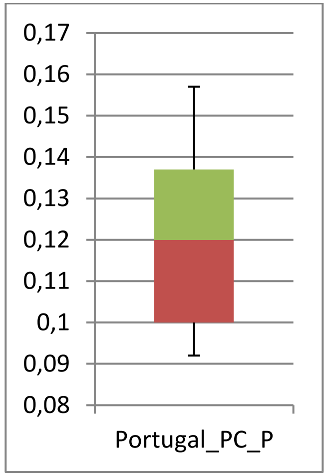


Figure 70: PT Pension percentage

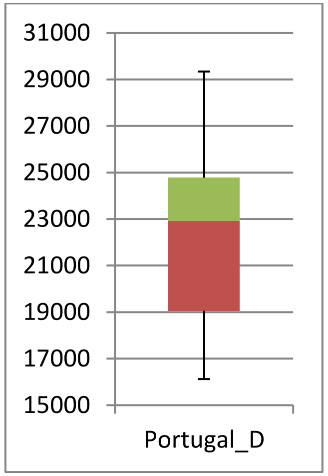


Figure 71: PT Deaths of old people from respiratory diseases

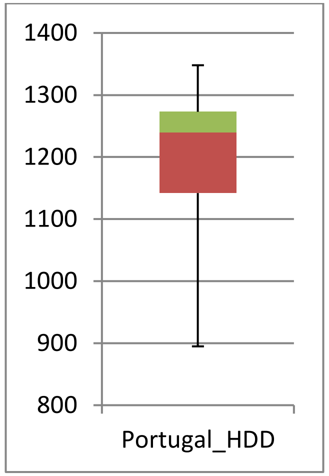


Figure 72: PT Number of Heating Degree Days

Portugal has some similarities with other southern European countries. This translates to low average potential support ratio and high average percentage of aging population, as the first is observed at 3.969 and the second at 17.1%. Also, the unemployment is high in average, but the female one is slightly higher (see Figure 66 and 67).

Health consumption in Portugal is higher than electricity and other fuels in average around 1.6 billion euros. From the box and whiskers plots a bigger interquartile is seen in electricity and other fuels consumption compared to health. (See Figures 68 and 69).The average percentage of Pensions from GDP in Portugal is 0.120. Figure 70 illustrates the structure of data with the vast majority located in the interquartile.

Portugal has a low number of heating degree days, which is 1.2 thousand in average. Also in average there are 22.3 thousand deaths of aged persons from diseases of the respiratory system and the sum of it is approximately 470 thousand dead persons.

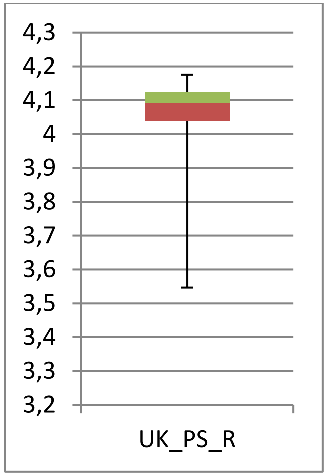


Figure 73: UK Potential Support Ratio

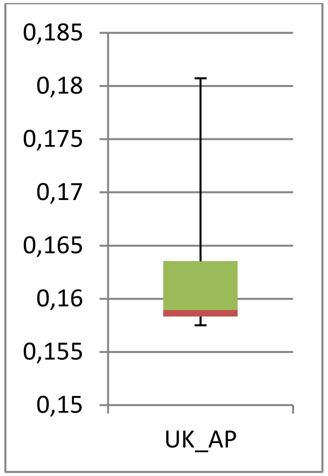


Figure 74: UK percentage of Aging Population

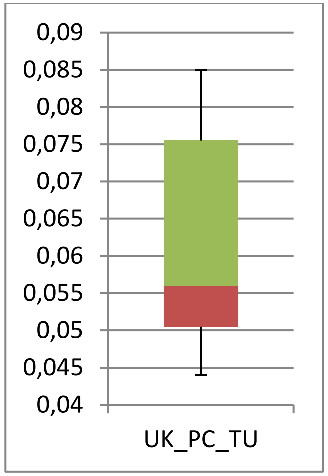


Figure 75: UK Percentage of Total Unemployment

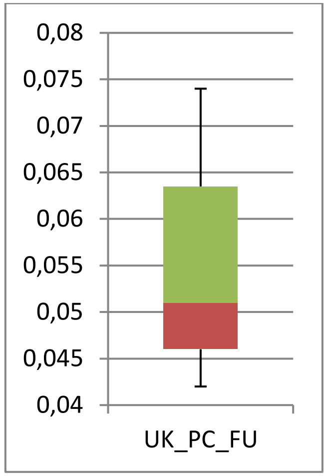


Figure 76: UK Percentage of Female Unemployment

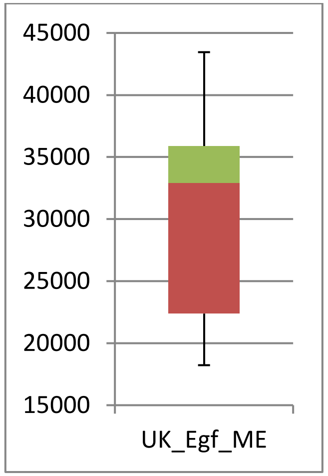


Figure 77: UK Electricity, gas and other Fuels consumption

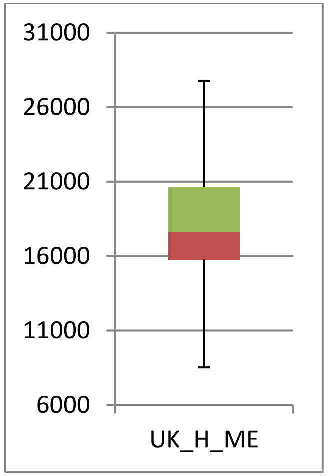


Figure 78: UK Health consumption

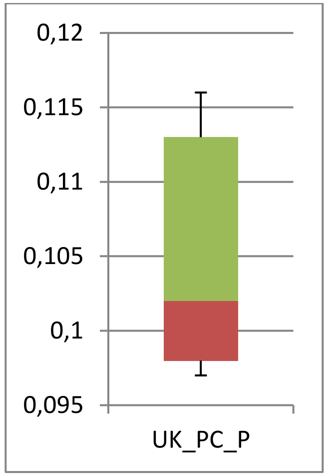


Figure 79: UK Pension percentage

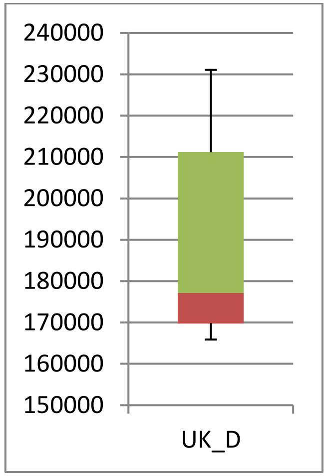


Figure 80: UK Deaths of old people from respiratory diseases

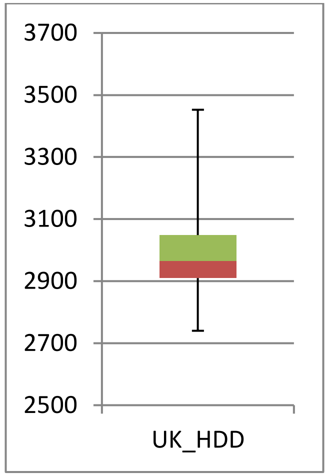


Figure 81: UK Number of Heating Degree Days

United Kingdom has average potential support ratio at 4.019 and aging population at 16.3%. It is quite interesting that the interquartile of UK’s PS_R is above 4 (see Figure 73), this shows that for the most part of time United Kingdom had strong support of a labour force either from births or immigration.

It seems that electricity and other fuels consumption is exceptionally high in average compared to health consumption in United Kingdom. The first index is on the level of 30 billion euros and the second on the level of 18 billion euros. There is also concentration of data in the second quartile in the first index (see Figure 77). For the United Kingdom, the percentage of Pensions from GDP is 0.105 in average. Skewness is 0.441 and combined with Figure 79, may indicate potential curve, since third quartile takes much place in the box and whisker’s plot.

We expect a great number of heating degree days in United Kingdom and in average there are almost exactly three thousand days. The average number of deaths for old age people is 186.4 thousand from respiratory diseases and the sum is almost four million dead people. Not to mention that the majority of mortality data concentrate in the third quartile (see Figure 80).

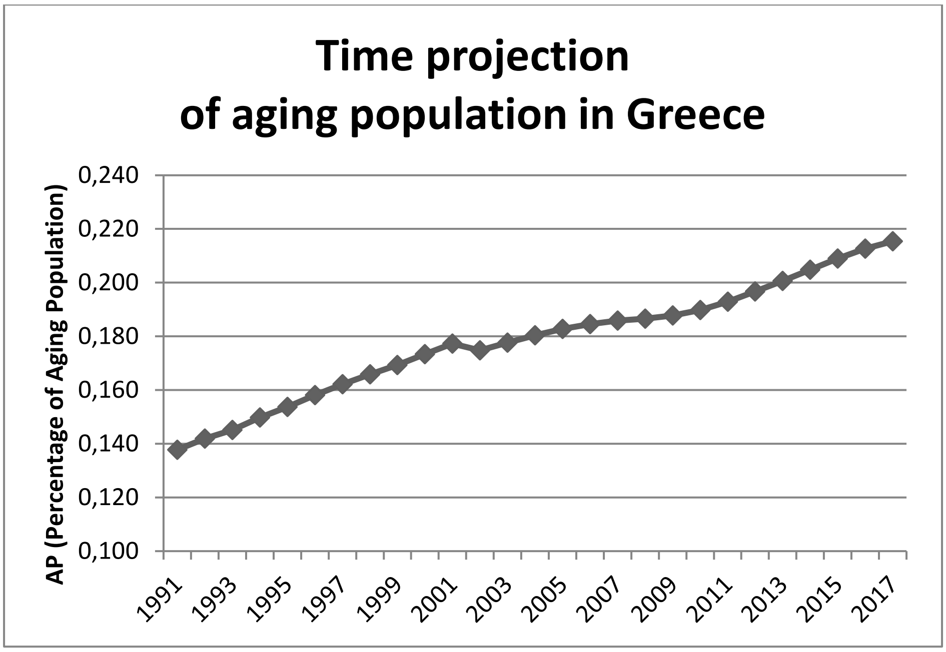


Figure 82: Time projection of aging population in Greece

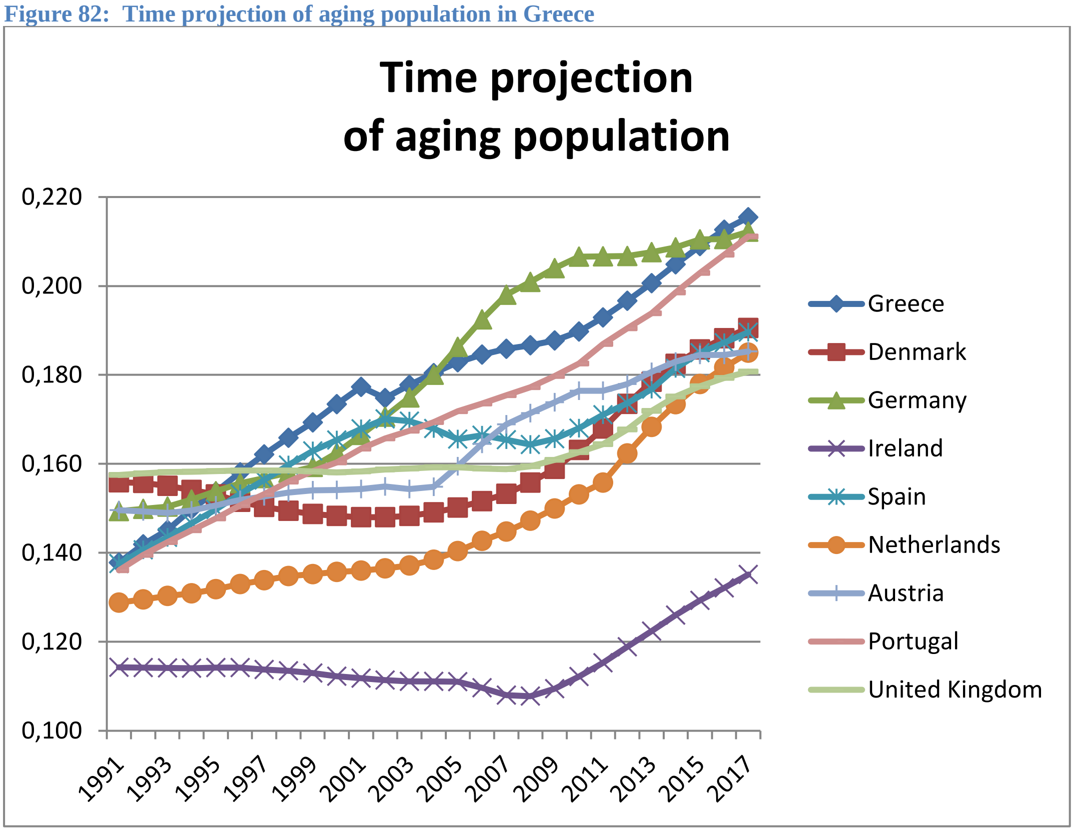


Figure 83: Time projection of aging population comparison

In comparison, from the nine countries in this study we find Ireland has the less aged people proportionate to her population. Of course the last decade the trend is rising complying with the norm of the rest of Europe. On the other hand, Greece, Germany and Portugal have surpassed the threshold of 20% and the rest that of 18%.

In the Figure 84 we get an outlook of potential support ratio structure for each of the nine studied countries. Keeping in mind that the minimum value is the last year's data (aka 2017) a convergence is seen for almost all countries in par with the convergence in aging population.

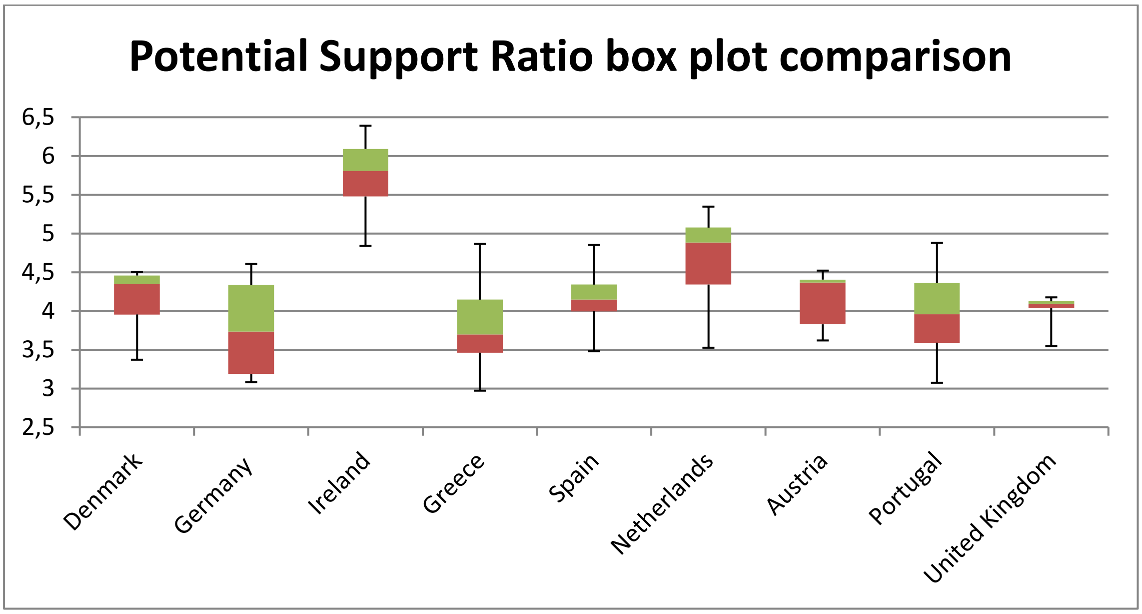


Figure 84: Potential Support Ratio box plot comparison

Table 1: Correlation matrix for Potential Support Ratio among the countries

PS_R	*Denmark*	*Germany*	*Ireland*	*Greece*	*Spain*	*Netherlands*	*Austria*	*Portugal*	*United Kingdom*
Denmark	1								
Germany	0.744614	1							
Ireland	0.681121	0.034776	1						
Greece	0.730074	0.946903	0.11126	1					
Spain	0.743965	0.849903	0.258928	0.967222	1				
Netherlands	0.941796	0.894696	0.45692	0.915189	0.908211	1			
Austria	0.898473	0.935587	0.314033	0.894172	0.833024	0.96600806	1		
Portugal	0.812822	0.96206	0.206278	0.98984	0.955444	0.958207194	0.940649	1	
United Kingdom	0.927518	0.555353	0.823708	0.64843	0.743812	0.864916917	0.744931	0.713847	1

From Table 1 we can extract information about the connections among the studied countries in terms of potential support ratio. For example, Greece has exceptionally high correlation with almost all the countries except Ireland and the United Kingdom. United Kingdom on the other hand has high correlation with Denmark, Ireland, the Netherlands, Austria and Portugal.

More details in comparison are found in the Tables 2 – 5 for demography.

Table 2: Main summary statistics for Potential Support Ratio comparison

PS_R	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	4.172	0.074	4.350	0.384	3.371	4.502
Germany	3.778	0.110	3.736	0.570	3.082	4.608
Ireland	5.753	0.082	5.810	0.425	4.840	6.391
Greece	3.797	0.102	3.696	0.531	2.972	4.867
Spain	4.151	0.070	4.145	0.362	3.480	4.852
Netherlands	4.666	0.109	4.882	0.565	3.526	5.347
Austria	4.147	0.063	4.367	0.328	3.620	4.521
Portugal	3.969	0.100	3.956	0.521	3.074	4.880
United Kingdom	4.019	0.036	4.094	0.188	3.547	4.176

Table 3: Main summary statistics for percentage of Aging Population comparison

AP	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	0.160	0.003	0.154	0.014	0.148	0.191
Germany	0.181	0.05	0.180	0.024	0.149	0.212
Ireland	0.116	0.001	0.114	0.007	0.108	0.135
Greece	0.178	0.004	0.180	0.022	0.138	0.215
Spain	0.165	0.003	0.166	0.013	0.138	0.190
Netherlands	0.146	0.003	0.138	0.017	0.129	0.185
Austria	0.164	0.003	0.155	0.013	0.149	0.185
Portugal	0.171	0.004	0.169	0.021	0.136	0.211
United Kingdom	0.163	0.001	0.159	0.007	0.158	0.181

Table 4: Main summary statistics for percentage of Total Unemployment comparison

PC_TU	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	0.056	0.003	0.055	0.012	0.034	0.076
Germany	0.076	0.004	0.079	0.021	0.038	0.112
Ireland	0.087	0.008	0.075	0.040	0.042	0.155
Greece	0.144	0.014	0.107	0.067	0.078	0.275
Spain	0.164	0.012	0.172	0.056	0.082	0.261
Netherlands	0.054	0.003	0.050	0.014	0.031	0.083
Austria	0.049	0.001	0.048	0.006	0.039	0.060
Portugal	0.094	0.007	0.088	0.033	0.051	0.164
United Kingdom	0.061	0.003	0.056	0.013	0.044	0.085

From the Table 4 we confirm which country had more impact through the financial crisis of 2008. Greece and Spain have exceptionally high average total unemployment over 10%, also Ireland alongside Portugal are close to this threshold. All these countries have undergone bailout programs with the support of International Monetary Fund. The changes in the unemployment can be seen from the Maximum column compared with Standard Deviation column, cause with high fluctuation there is high value in each mentioned column.

Table 5: Main summary statistics for percentage of Female Unemployment

PC_FU	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	0.060	0.002	0.060	0.012	0.037	0.081
Germany	0.076	0.005	0.083	0.024	0.033	0.109
Ireland	0.080	0.007	0.073	0.032	0.043	0.128
Greece	0.192	0.013	0.164	0.062	0.115	0.314
Spain	0.196	0.011	0.199	0.053	0.107	0.279
Netherlands	0.065	0.004	0.062	0.018	0.042	0.113
Austria	0.049	0.001	0.049	0.005	0.041	0.059
Portugal	0.096	0.007	0.090	0.033	0.050	0.166
United Kingdom	0.055	0.002	0.051	0.011	0.042	0.074

Table 6: Main summary statistics for Electricity, gas and other fuels consumption comparison

Egf_ME	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	6,137.761	233.945	6,142.900	1,121.963	4,235.500	7,889.600
Germany	54,981.904	2,622.587	56,002.000	12,577.487	37,040.000	75,061.000
Ireland	2,296.517	169.036	2,573.700	810.668	1,057.100	3,297.300
Greece	3,607.478	314.918	3,962.300	1,510.292	1,450.100	6,381.400
Spain	15,722.087	1,485.581	14,229.000	7,124.597	7,315.900	25,287.000
Netherlands	8,877.635	448.271	9,854.000	2,149.832	5,553.100	12,050.000
Austria	5,768.000	238.541	5,909.800	1,144.004	4,153.400	7,262.600
Portugal	3,136.635	210.964	3,227.500	1,011.748	1,587.400	4,501.000
United Kingdom	29,955.830	1,706.220	32,925.900	8,182.745	18,225.400	43,451.700

Table 7: Main summary statistics for Health consumption comparison

H_ME	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	2,764.287	144.436	2,811.300	692.689	1,677.900	3,796.000
Germany	58,956.343	3,286.379	55,756.000	15,760.919	36,909.500	84,976.000
Ireland	2,595.883	278.499	2,630.200	1,335.634	724.200	4,560.400
Greece	5,581.183	314.989	5,412.900	1,510.633	3,148.300	8,008.800
Spain	18,741.213	1,286.700	19,414.000	6,170.799	9,203.300	28,172.000
Netherlands	9,395.365	359.748	8,950.000	1,725.290	6,770.000	13,013.000
Austria	5,398.374	250.271	5,325.800	1,200.258	3,580.600	7,489.800
Portugal	4,799.587	271.540	4,912.900	1,302.262	2,675.500	6,725.000
United Kingdom	18,073.417	1,081.534	17,645.800	5,186.856	8,525.700	27,776.100

From the Tables 6 and 7 we conclude that Ireland has the less consumption in million euros current prices from the studied countries in average. Germany has the highest consumption in average.

Table 8: Main summary statistics in Pensions comparison

PC_P	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	0.116	0.003	0.112	0.012	0.102	0.140
Germany	0.124	0.001	0.124	0.004	0.117	0.131
Ireland	0.057	0.003	0.054	0.013	0.037	0.077
Greece	0.129	0.006	0.119	0.028	0.096	0.178
Spain	0.101	0.003	0.097	0.013	0.087	0.128
Netherlands	0.122	0.001	0.121	0.007	0.112	0.134
Austria	0.140	0.001	0.140	0.004	0.132	0.148
Portugal	0.120	0.005	0.120	0.022	0.092	0.157
United Kingdom	0.105	0.002	0.102	0.007	0.097	0.116

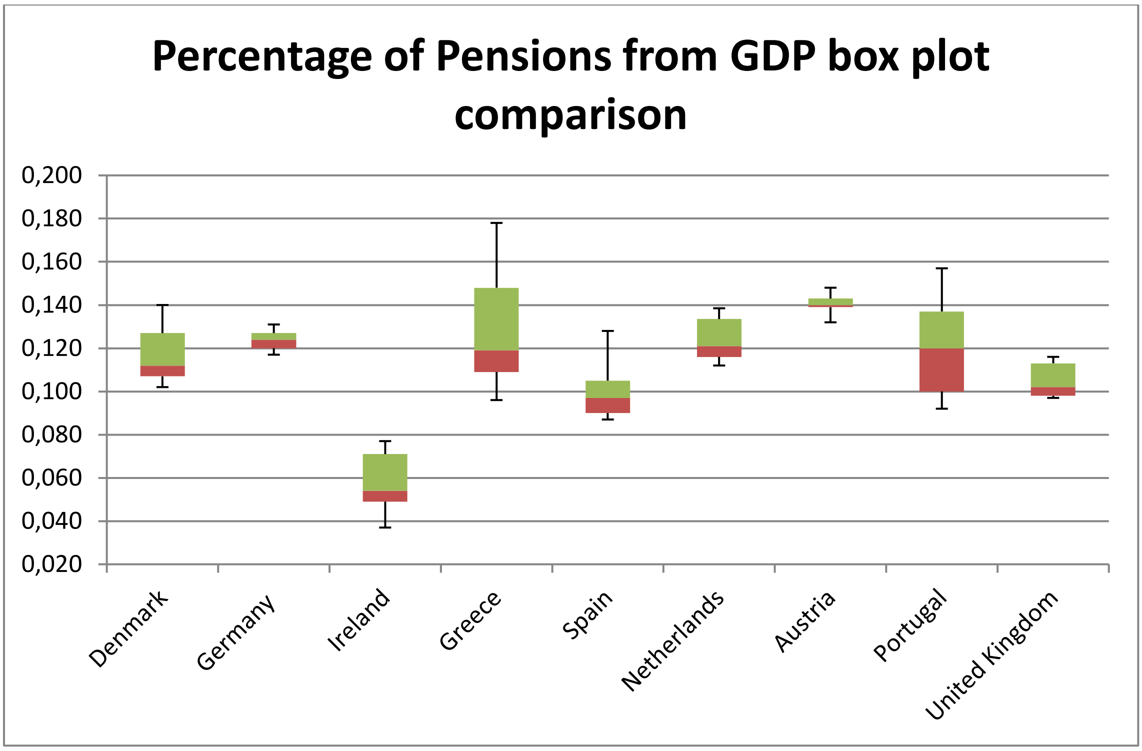


Figure 85: Percentage of Pensions from GDP box plot comparison

Table 8 combined with Figure 85 gives a comparable overview of pensions for every studied country. It is apparent that Ireland has the least percentage of pensions its economy gives. On the other hand, Greece and Portugal are in the first two places in the rank, with both countries containing a wider spread for the fluctuation of the index.

Table 9: Main summary statistics in deaths index comparison

D	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*	*Sum*
Denmark	13,207.381	126.656	13,220.000	580.411	12,245.000	14,360.000	277,355.000
Germany	107,782.762	5,303.593	97,217.000	24,304.116	82,160.000	154,456.000	2,263,438.000
Ireland	8,908.286	269.017	8,700.000	1,232.789	7,242.000	11,795.000	187,074.000
Greece	18,100.000	1,195.412	16,165.000	5,478.065	10,999.000	31,561.000	380,100.000
Spain	92,812.381	2,056.170	93,797.000	9,422.554	74,940.000	113,792.000	1,949,060.000
Netherlands	26,677.429	725.712	25,474.000	3,325.628	22,041.000	33,016.000	560,226.000
Austria	9,641.476	239.937	9,611.000	1,099.528	7,857.000	12,092.000	202,471.000
Portugal	22,341.000	816.215	22,918.000	3,740.369	16,123.000	29,341.000	469,161.000
United Kingdom	186,427.238	4,701.293	177,095.000	21,544.030	165,875.000	231,073.000	3,914,972.000

From Table 9 is concluded that Ireland has the least number of deaths of old aged people from diseases in the respiratory system. Not to mention, it has the least accrued amount of dead people in 21 years of data.

Table 10: Main summary statistics in heating degree days comparison

HDD	*Mean*	*Standard Error*	*Median*	*Standard Deviation*	*Minimum*	*Maximum*
Denmark	3,304.011	59.415	3,235.750	284.947	2,850.870	4,046.260
Germany	3,093.182	52.839	3,080.550	253.406	2,660.860	3,815.500
Ireland	2,736.083	30.235	2,702.730	145.001	2,525.790	3,168.560
Greece	1,631.950	27.304	1,634.520	130.944	1,408.340	1,832.160
Spain	1,771.557	28.015	1,816.320	134.354	1,537.130	1,971.670
Netherlands	2,722.888	54.266	2,680.270	260.250	2,284.960	3,449.160
Austria	3,593.645	45.773	3,640.030	219.520	3,124.730	4,118.120
Portugal	1,195.703	25.957	1,239.350	124.486	894.850	1,348.000
United Kingdom	3,000.430	35.223	2,965.940	168.923	2,739.930	3,452.490

From Table 10 is concluded that Portugal has the least average number of heating degree days and Austria has the highest average.

CHAPTER 3

Regression analysis

Aim of regression analysis is to receive a mathematical model that describes a linear relationship among variables. Even if that model is found, does not automatically explain the behaviour of fluctuations. For this reason, the linear model needs several hypotheses tested through particular statistical indices. Chapter three provides information on the testing with visuals and tables.

The method we will use upon the data is the simple linear regression with the form of $y = \beta_0 + \beta_1 x + \varepsilon$

where

y = dependent variable
x = independent variable
β_0 = y-intercept
β_1 = slope of the line
ε = error variable

The aim with the means of Ordinary Linear Regression is to test the hypotheses

$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

If the null hypothesis is true, then there is no linear relationship between y and x variables. On the other hand, if the alternative hypothesis is true there exists a linear relationship which can be used for inference and forecasting. (Keller, 2014)

Linear regression has four hypotheses:

- Linear function between dependent and independent variables
- Residuals have a constant variance
- Residuals follow the normal distribution
- Residuals are independent

This method can be applied through R programming, particularly through R-Studio.The dataset from excel can be converted into .csv form so we can work it in R studio.

After reading the csv tables that contain the dataset for every index, the correlation function is applied to see the relation between the indices. Then the glm function is applied, for example using AP_DK as y and PS_R_DK as x. The glm function stands for generalized linear models, which gives regression results. Looking at those, the estimate of Intercept is found 0.30, meaning that regardless of variation in PS_R_DK the AP_DK index will be set initially at 30%. Then the estimate of PS_R_DK is - 0.035, meaning that for every 1 added unit in PS_R_DK it will reduce the AP_DK index for - 0.0.35 units. Especially, it is inferred that the estimate of coefficient for PS_R_DK is statistically significant as is seen in p-value = < 2e-16. This translates that the possibility the estimate is wrong is almost zero, thus its rank is three stars to achieve highest significance. Not to mention that R squared can be found through the equation: $R^2 = 1 - (Residual\ Deviance\ /\ Null\ Deviance)$, which in our case is 0.992 meaning that the variation of AP_DK is explained by 99.2% from the linear model.

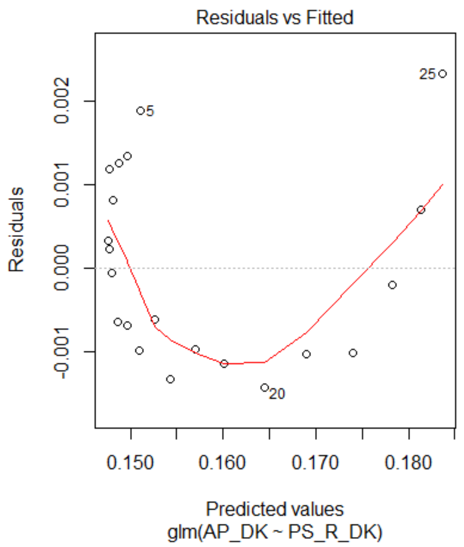


Figure 86: Residuals vs Fitted plot for AP-PS_R DK

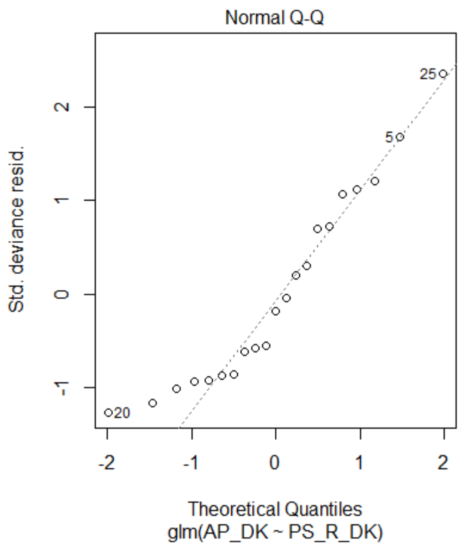


Figure 87: Fitted line of normalized residuals for AP-PS_R DK

From Figure 86 the line with red colour shows the mean of the residuals for every fitted value. A good linear model would be a red line to spread in the middle without curves and around zero in y axis. Here the line curves but the residuals are gathered near the mean red line, which indicate a slight nonlinearity. Another part of our normality hypothesis for residuals appears in Figure 87, which shows a slight good fit. As for the independence of residuals, it can be tested with a Durbin-Watson test. This statistic determines whether a relationship exists between consecutive residuals. It ranges from zero to four with ideal parameter 2 to indicate no first order autocorrelation. The 0.277 DW result indicates a strong positive first order autocorrelation with p-value less than a=0.05 showing a rejection of the null hypothesis.

To inform the reader all the following analyses include the years 1995 to 2015 for reasons of uniformity.

Table 11: Elements of linear model for AP - PS_R comparison ($AP = \beta_0 + \beta_1 \times PS\_R$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of PS_R (β_1)	P-value of PS_R coeff.	R squared	DW test
DK	-0.996	0.305	-0.035	< 2E-16	0.992	0.277
DE	-0.998	0.344	-0.043	< 2E-16	0.995	0.165
IE	-0.952	0.201	-0.015	3.33E-11	0.906	0.146
EL	-0.998	0.335	-0.041	< 2E-16	0.996	0.373
ES	-0.996	0.313	-0.036	< 2E-16	0.993	0.515
NL	-0.996	0.286	-0.030	< 2E-16	0.993	0.288
AT	-0.998	0.333	-0.041	< 2E-16	0.996	0.715
PT	-0.997	0.332	-0.041	< 2E-16	0.994	0.238
UK	-0.974	0.314	-0.038	1.06E-13	0.949	0.168

Table 11 shows indicative results from the glm function in R and for every country studied. It is inferred that a strong negative relation exists between Potential Support Ratio and the percentage of Aging Population. Though every coefficient in each country's model is significant at α = 5%, there is a violation of independence of the residuals if we observe the Durbin-Watson test result. All residuals have positive first order autocorrelation.

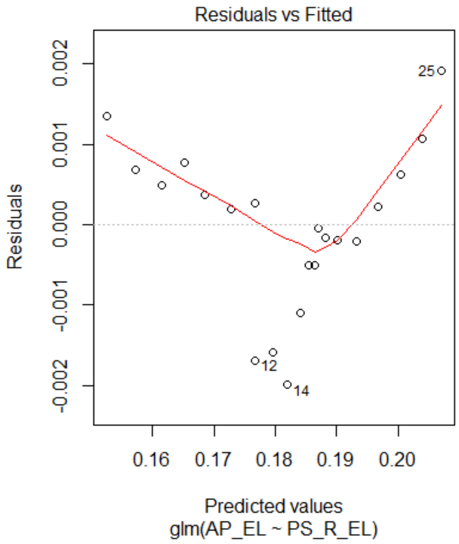


Figure 88: Residual vs Fitted plot for AP - PS_R EL

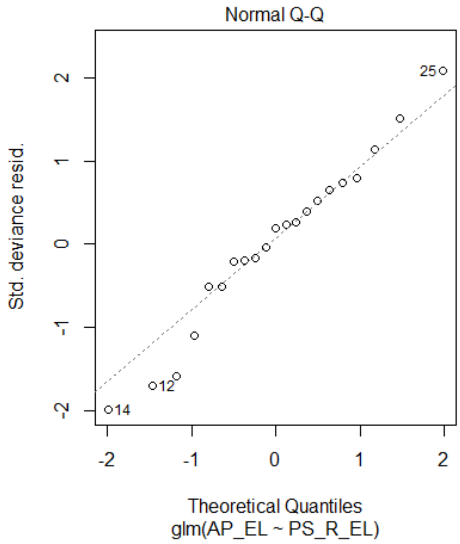


Figure 89: Fitted line of normalized residuals for AP - PS_R EL

In Figure 88 is observed a curved red line against most residuals with a slight nonlinearity, also the Figure 89 indicates that the normality in residuals applies for the linear model in Greece.

Table 12: Elements of linear model for D - HDD comparison ($D = \beta_0 + \beta_1 \times HDD$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of HDD (β_1)	P-value of HDD coeff.	R squared	DW test
DK	-0.175	14,356.353	-0.346	9.68E-09	0.031	1.030
DE	-0.053	122,860.315	-4.858	8.21E-01	0.003	0.278
IE	-0.417	18,205.595	-3.395	6.02E-02	0.174	0.564
EL	-0.601	59,050.882	-24.989	3.95E-03	0.361	0.893
ES	0.186	69,674.130	12.990	4.20E-01	0.034	1.043
NL	-0.043	281,325,511.000	-0.532	8.53E-01	0.002	0.700
AT	-0.232	13,713.992	-1.129	3.12E-01	0.054	0.794
PT	0.306	11,448.632	9.074	1.77E-01	0.809	0.471
UK	0.057	165,100.000	7.084	8.05E-01	0.003	0.474

It is inferred at Table 12 that there is no linearity between Deaths of aged people 65 or over and the number of Heating Degree Days. The correlation column has values that tend towards zero with the exception of Greece (-0.601). The p-value of the β_1 coefficient (here for HDD) show insignificance at α=5% for the majority. The R squared column appears to have similar results with the exception of Greece and mostly Portugal (0.361 and 0.809 respectively), meaning that the linear model in Portugal explains the 80% of variations in the number of Deaths. As for the independence of residuals, the DW test column informs on the rejection to this hypothesis. Only Denmark and Spain have DW values over one, thus tending to the ideal of independence.

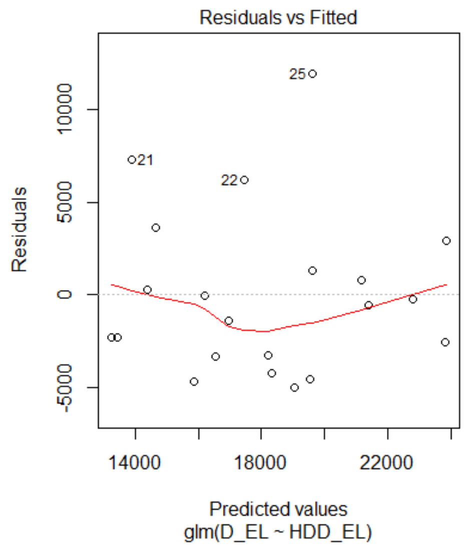


Figure 90: Residual vs Fitted plot for D - HDD EL

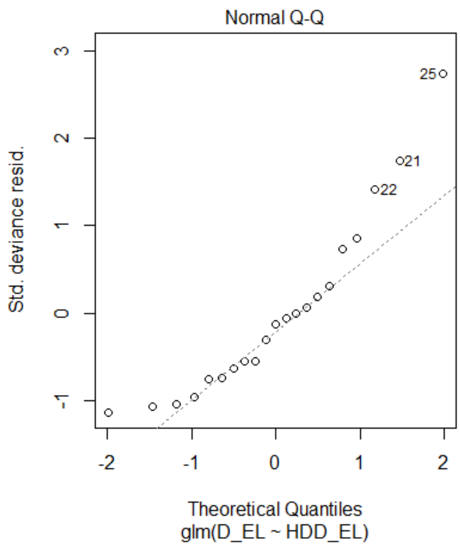


Figure 91: Fitted line of normalized residuals for D - HDD EL

Figures 90 and 91 similarly present the non-forecasting value of the linear model in Greece for Deaths and the number of Heating Degree Days. The residuals differentiate much from the mean red line that shows skewedness and parts of them diverge from normality line. On the other hand, the red line in Figure 90 indicates possible linearity.

Table 13: Elements of linear model for PC_P - AP comparison ($PC_P = \beta_0 + \beta_1 \times AP$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of AP (β_1)	P-value of AP coeff.	R squared	DW test
DK	0.928	-0.027	0.908	1.31E-09	0.862	1.283
DE	-0.502	0.142	-0.100	2.05E-02	0.252	0.602
IE	0.245	-0.008	0.577	2.85E-01	0.060	0.277
EL	0.927	-0.185	1.729	1.50E-09	0.860	0.360
ES	0.618	-0.065	0.990	2.84E-03	0.382	0.141
NL	0.472	0.088	0.233	3.06E-02	0.223	0.329
AT	0.517	0.111	0.182	1.64E-02	0.267	0.571
PT	0.979	-0.110	1.336	1.39E-14	0.958	0.909
UK	0.761	-0.042	0.907	6.22E-05	0.579	0.593

First conclusions drawn from Table 13 are the non-existence of linear relations between percentage of Pensions from GDP and percentage of Aging Population among the countries. The exception to this appears on Denmark, Greece and Portugal according to correlation column. These three countries have high R squared value (86% or over) thus an adequate forecast opportunity. In contrast, DW test column indicates positive first order autocorrelation among residuals with the exception of Denmark having 1.283 dw value, yet still positive autocorrelation.

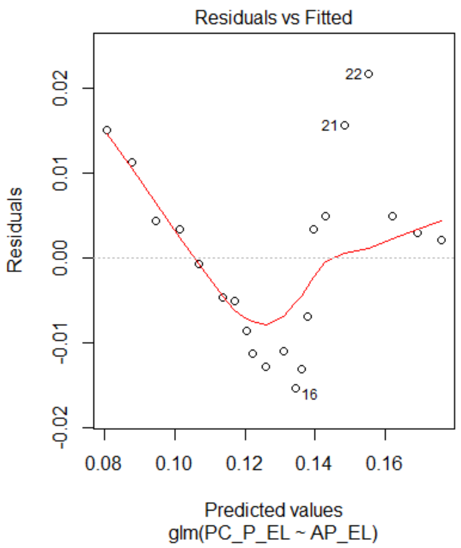


Figure 92: Residual vs Fitted plot for PC_P - AP EL

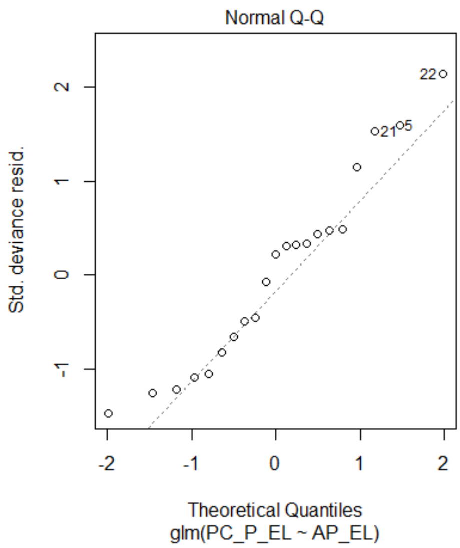


Figure 93: Fitted line of normalized residuals for PC_P - AP EL

Figure 92 shows that the majority of mean red line is near the residuals and Figure 93 shows a small tendency towards normality of them for the linear model in Greece.

Table 14: Elements of linear model for PC_TU - AP comparison ($AP = \beta_0 + \beta_1 \times PC_{TU}$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of PC_TU (β_1)	P-value of PC_TU coeff.	R squared	DW test
DK	0.625	0.124	0.600	2.44E-03	0.391	0.312
DE	-0.620	0.240	-0.706	2.70E-03	0.385	0.136
IE	0.486	0.108	0.065	2.55E-02	0.236	0.227
EL	0.709	0.159	0.165	3.19E-04	0.503	0.141
ES	0.283	0.160	0.040	2.15E-01	0.080	0.124
NL	0.310	0.130	0.291	1.71E-01	0.096	0.068
AT	0.528	0.110	1.122	1.40E-02	0.278	0.303
PT	0.845	0.136	0.395	1.42E-06	0.714	0.389
UK	0.188	0.157	0.086	4.14E-01	0.035	0.107

The table 14 present's results of linear modelling for the percentage of Total Unemployment and percentage of Aging Population indices. The majority of studied countries show nonlinearity. Exceptions to this are Greece and Portugal with significant positive correlation, though the R squared values differentiate. In Greece, the model explains the 50% of variations in percentage of Aging Population and in Portugal the 71.4%. On the other hand, all residuals indicate positive first order autocorrelation.

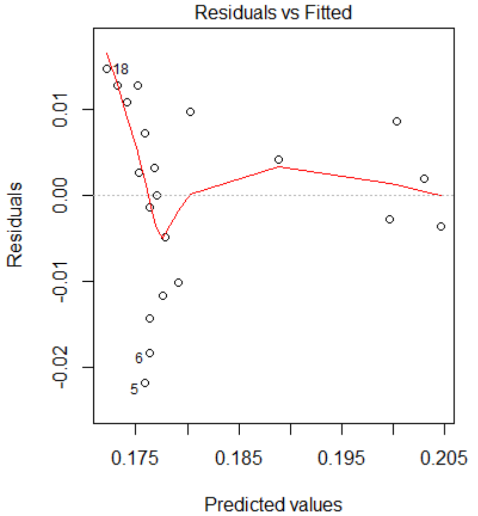


Figure 94: Residual vs Fitted plot for PC_TU - AP EL

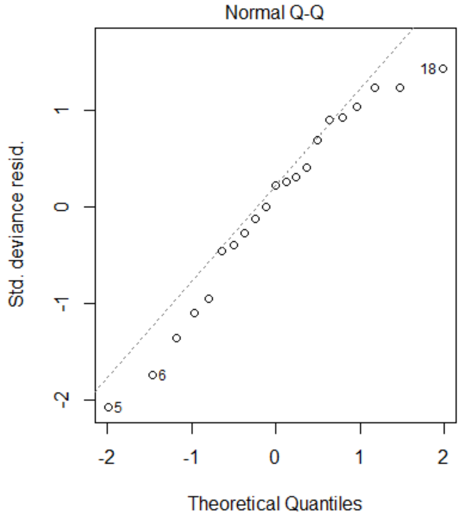


Figure 95: Fitted line of normalized residuals for PC_TU - AP EL

Figure 94 presents skewedness since parts of residuals have higher negative value than positive and mean red line curves sharply. On the other hand Figure 95 indicates tendency to normality for the linear model in Greece.

Table 15: Elements of linear model for PC_FU - AP comparison ($AP = \beta_0 + \beta_1 \times PC_{FU}$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of PC_FU (β_1)	P-value of PC_FU coeff.	R squared	DW test
DK	0.451	0.130	0.459	4.00E-02	0.204	0.127
DE	-0.749	0.245	-0.757	9.43E-05	0.561	0.172
IE	0.535	0.107	0.090	1.25E-02	0.286	0.237
EL	0.631	0.153	0.158	2.15E-03	0.399	0.122
ES	-0.028	0.168	-0.004	9.03E-01	0.001	0.098
NL	-0.073	0.150	-0.053	7.55E-01	0.005	0.048
AT	0.298	0.133	0.651	1.90E-01	0.089	0.117
PT	0.817	0.137	0.377	6.32E-06	0.667	0.321
UK	0.384	0.150	0.210	8.57E-02	0.147	0.161

Table 15 presents a similar view compared to Table 50 since the data are about female unemployment. Here the correlation column indicates weak relation between percentage of Female Unemployment and percentage of Aging Population, with a difference in Germany and Portugal. For Germany there is a negative correlation (-0.749) and the model explains 56.10% of variations in independent variable. On the other end of the spectrum, Portugal has positive correlation (0.817) and R squared value of 66.7%. Again all the Durbin-Watson test results show a positive first order autocorrelation of the residuals.

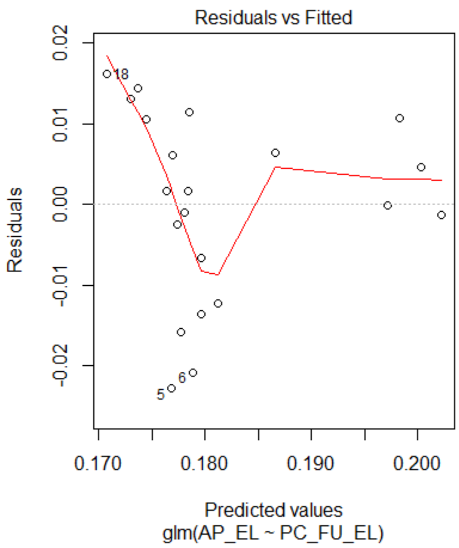


Figure 96: Residual vs Fitted plot for PC_FU - AP EL

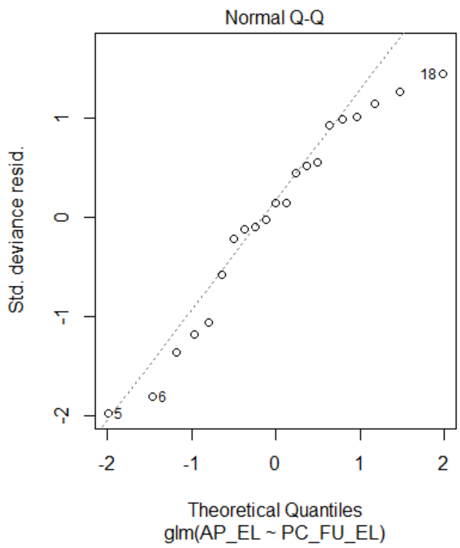


Figure 97: Fitted line of normalized residuals for PC_FU - AP EL

The linear model in Greece for female unemployment and aging population looks unsupportive in linearity as is shown in Figure 96. The mean red line curves twice and residuals differentiate with variations. Figure 97 also shows a weak tendency to normality.

Table 16: Elements of linear model for Egf_ME - AP comparison (Egf_ME $= \beta_0 + \beta_1 \times AP$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of AP (β_1)	P-value of AP coeff	R squared	DW test
DK	0.768	-5,195.000	71,307.000	4.77E-05	0.590	0.306
DE	0.972	-51,575.000	572,707.000	2.00E-13	0.945	1.913
IE	0.358	-3,808.000	53,013.000	1.11E-01	0.128	0.088
EL	0.737	-10,525.000	77,473.000	1.38E-04	0.543	0.580
ES	0.769	-91,353.000	636,477.000	4.56E-05	0.592	0.167
NL	0.791	-9,597.000	125,580.000	1.93E-05	0.626	0.308
AT	0.949	-8,979.000	88,972.000	5.58E-11	0.901	1.048
PT	0.979	-7,315.000	59,842.000	1.18E-14	0.959	0.447
UK	0.785	-149,828.000	1,105,790.000	2.51E-05	0.616	0.335

Table 16 holds the linear modelling results for Electricity, gas and other fuels consumption in relation with the percentage of Aging Population. All countries show positive correlation, either weak or strong. The models that present more forecasting value are those of Germany, Austria and Portugal, since they have strong positive correlation (over 0.95) and R squared value high (over 90%). What comes as a surprise is that DW test result for Germany is 1.913, meaning that the residuals are almost independent among them.

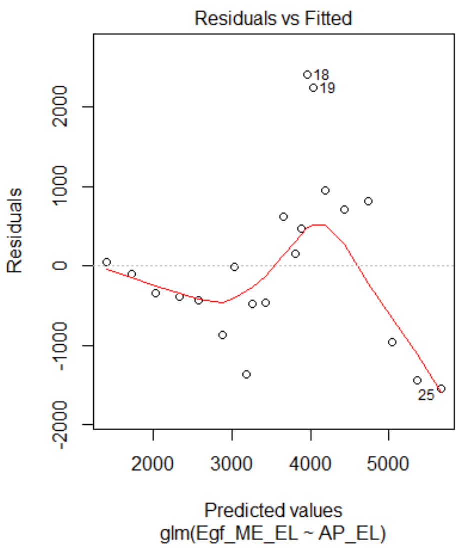


Figure 98: Residual vs Fitted plot for Egf_ME - AP EL

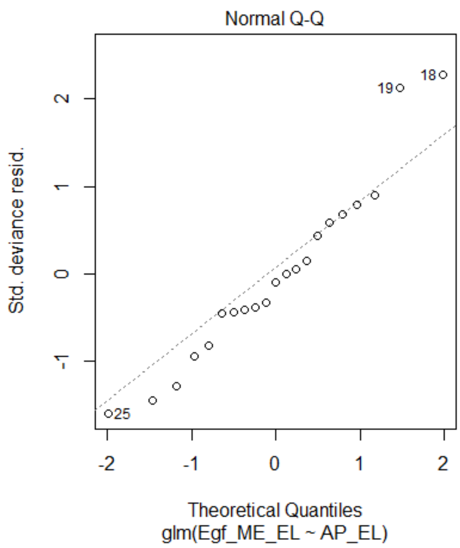


Figure 99: Fitted line of normalized residuals for Egf_ME - AP EL

Figure 98 indicates nonlinearity for the linear model in Greece, since the mean red line curves twice and from Figure 99 is drawn a weak normality of the residuals.

Table 17: Elements of linear model for H_ME - AP comparison ($H\_ME = \beta_0 + \beta_1 \times AP$)

Countries	Correlation	Coeff. of Intercept (β_0)	Coeff. of AP (β_1)	P-value of AP coeff.	R squared	DW test
DK	0.817	-4,109.000	42,965.000	6.58E-06	0.665	0.104
DE	0.962	-62,003.000	643,982.000	3.87E-12	0.925	0.259
IE	0.492	-10,147.000	110,122.000	2.34E-02	0.242	0.067
EL	0.498	-4,007.000	52,680.000	2.17E-02	0.248	0.209
ES	0.815	-75,994.000	562,396.000	6.91E-06	0.664	0.169
NL	0.182	5,943.000	22,595.000	4.30E-01	0.033	0.800
AT	0.944	-8,716.000	84,716.000	1.31E-10	0.891	0.226
PT	0.928	-7,768.000	71,822.000	1.42E-09	0.861	0.285
UK	0.749	-77,308.000	583,557.000	9.47E-05	0.561	0.248

Table 18 indicates a positive relation between the Health consumption and the percentage of Aging Population. The correlation column shows strong positivity for the majority with exceptions, such as the Netherlands (0.182). The variations in health consumption due to the linear model with percentage of aging population as independent variable are best explained in Germany, Austria and Portugal (R squared over 85%). On the other hand, dw test results indicate dependency among residuals with positive first order autocorrelation.

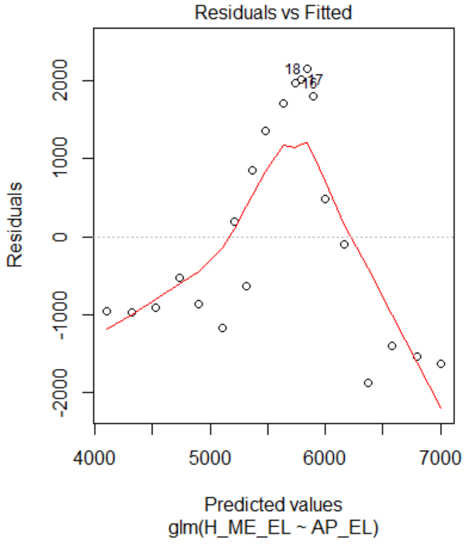


Figure 100: Residual vs Fitted plot for H_ME - AP EL

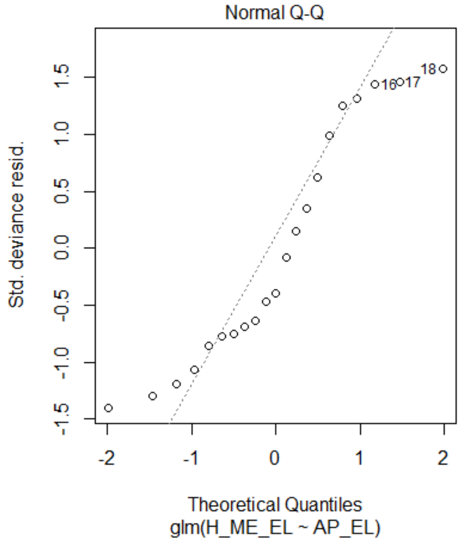


Figure 101: Fitted line of normalized residuals for H_ME - AP EL

The linear model in Greece for health consumption and the percentage of aging population shows signs of nonlinearity in Figure 100 as the mean red line of residuals curves sharply. Also, the normality is weak as is presented in Figure 101.

In conclusion, regression analysis is limited to few relations among studied indices and especially fewer countries (Mostly Portugal). The hypothesis that is violated the most is the independence of the residuals as was seen in Durbin - Watson test results, which only Germany achieved it and for the linear model of the Electricity, gas and other fuels consumption with the percentage of Aging Population relation.

CHAPTER 4

Cluster analysis

Although it initially the data had shown a linear tendency then after a point data regress. This destroys linearity therefore we change our point of view from classical regression analysis to clustering of the data, in order to investigate for underlying relations or groupings.

K-means cluster analysis is suitable because it is a simple approach to partition the dataset into K number of distinct, non-overlapping clusters with two hypotheses. First, every observation belongs to at least one of the defined clusters. Second, every observation does not belong to more than one cluster. Every cluster has a centroid, a point that represents the mean of observations using the Euclidian distance (most commonly) among them for calculation, thus we are able to find what characteristic defines each particular cluster.

The mathematical algorithm that achieves this is:

$$\underset{C_1, \dots, C_k}{minimize} \left\{ \sum_{k=1}^{K} W(C_k) \right\}$$

This means that, for every cluster C_k minimize the amount of distance that observations differ within the cluster, thus assigning those with the closest proximity into a distinct cluster. The first step is to assign random initial centroids and recalculate them until minimization is achieved (James, Witten, Hastie, & Tibshirani, 2015).

A problem that emerges is the definition of the right number of clusters to partition the data. For reasons of simplicity the writer chose another method of clustering. The Hierarchical cluster analysis is a method, which has the advantage to work bottom - up. This translates into an algorithm that assigns every observation as its own cluster and then adding to that cluster the closest observation in terms of proximity. The steps later conclude to fusion of entire clusters towards only one, the top. There is a term called linkage to describe the way to calculate which observation or cluster should be fused. The complete linkage is a type that approaches maximal intercluster dissimilarities. For example, we have cluster A and cluster B and we calculate pairwise the dissimilarity between observations from cluster A to observations from cluster B in order to find the maximum among them. If the complete linkage between cluster A and B is lower than the one between cluster B and C then the fusion of A and B takes place. In R we can view the dendrogram created through this method.

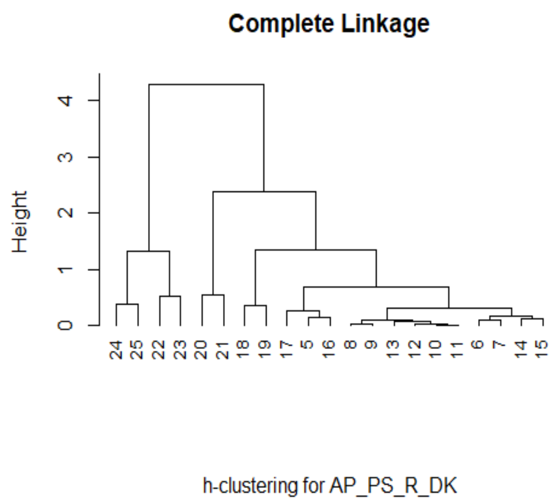


Figure 102: Hierarchical clustering dendrogram for AP – PS_R DK

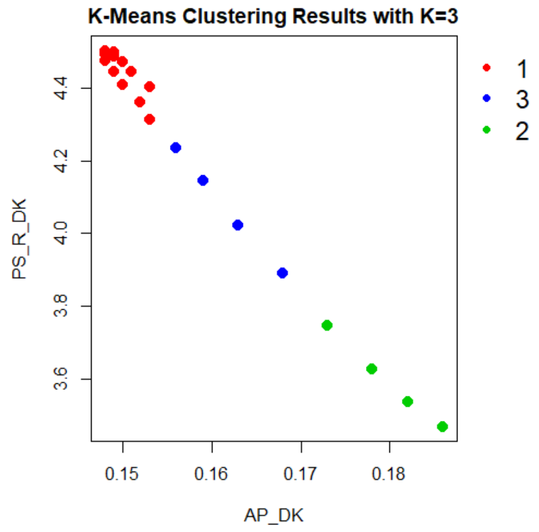


Figure 103: K-means cluster separation in scatter plot for AP - PS_R DK

In Figure 102 is presented how the dendrogram, as a result from hierarchical clustering, unfolds from top to bottom. Observation 17, 5 and 16 are close in proximity and they can create a single cluster. Observation 18 and 19 are similarly closer but when they are about to fuse into a bigger branch we see a difference in height thus indicating dissimilarities with other clusters before fusion. If we decide to cut the dendrogram in height level 2 we get three clusters that share inner characteristics and are separated with less inner dissimilarities. Then we can apply the K-means cluster method with specified K = 3.

Figure 103 illustrates the clusters in the scatter plot for AP – PS_R in Denmark. The decision to specify K = 3 through the hierarchical clustering results in uniform groups as viewed. This process can be applied to the rest models we seen in subsection 3.1 for regression analysis.

Data considering the relation among Potential Support Ratio and percentage of Aging Population are separated into three distinct clusters with K-means. The first includes over a decade (1995 -2007) with average AP 0.150 and PS_R 4.45. The next is involved for the years 2008 – 2011 with average AP 0.162 and 4.072 PS_R (around the time with the immediate ramifications of the economic crisis). The last includes years 2012 – 2015 with average AP 0.180 and PS_R 3.594. The reader can distinguish this way a demographic change in Denmark without the aid of political or economic media. Later on, there can be research to find the reasons that separate those periods.

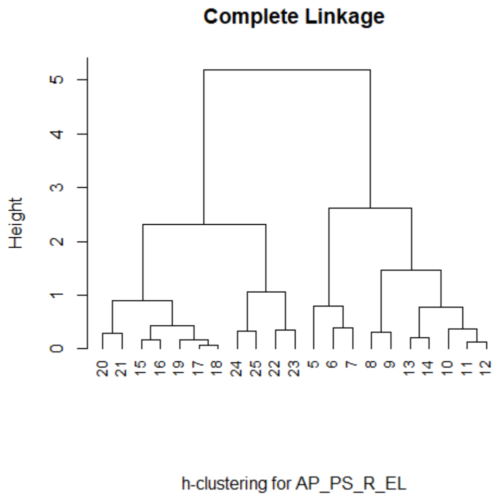


Figure 104: Hierarchical clustering dendrogram for AP – PS_R EL

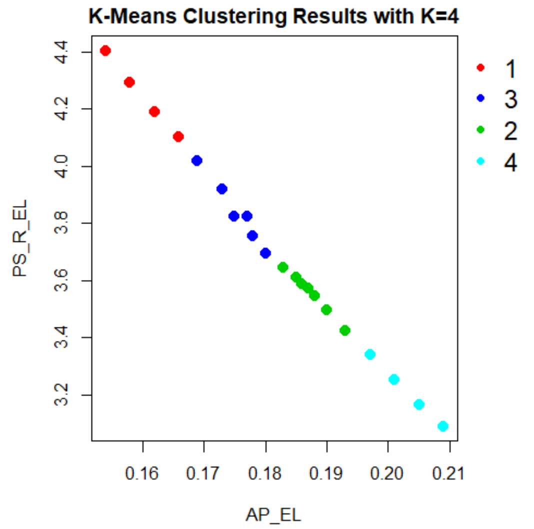


Figure 105: K-means cluster separation in scatter plot for AP - PS_R EL

Figure 104 indicates for the relation of Potential Support Ratio and the percentage of Aging Population four clusters if we cut the dendrogram in height level 2. Figure 105 illustrates the K-means clustering with K=4.

The first cluster includes the period from 1995 - 1998 with mean AP = 0.16 and PS_R = 4.247. The next period - cluster involves the years 1999 - 2004 with mean AP equal to 0.175 and PS_R equal to 3.839. The following cluster includes the years 2005 - 2011 with AP = 0.187 and PS_R 3.555. The last cluster separates the years 2012 - 2015 with 20.3% percentage of aging population and 3.212 persons 64 or younger per one 65 or over person. The reader should know the facts that in 1999 Greece suffered from a major crash in stock market, in 2002 euro as currency was introduced to the nation, 2004 were the Olympic Games in Athens and in 2012 the second bailout program was ratified as the economic crisis had worsen.

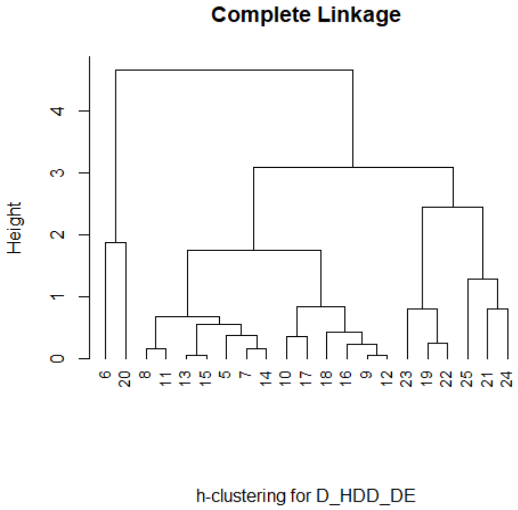


Figure 106: Hierarchical clustering dendrogram for D – HDDD DE

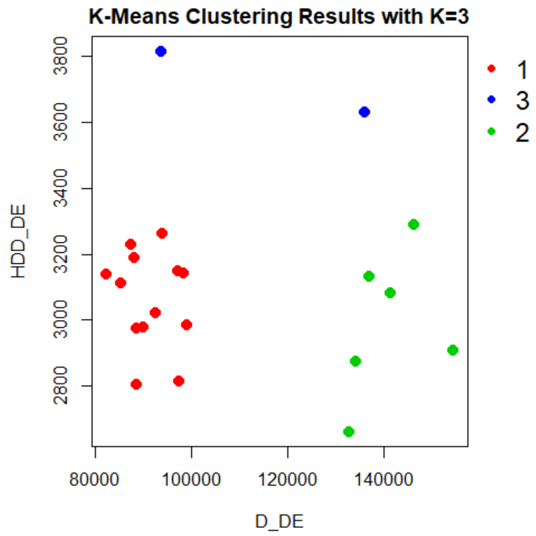


Figure 107: K-means cluster separation in scatter plot for D – HDD DE

In Figure 106 it is observed that hierarchical clustering for the relation of Deaths from respiratory diseases and the number of Heating Degree Days in Germany distinguishes three major clusters. Then K-means is applied to the same relation with K equals three and it is illustrated in Figure 107. The cluster number three shows two outliers from the data.

The cluster with the outliers includes the years 1996 and 2010. Though the centroid for deaths is 114,802 and for HDD 3,722.920 we can tell, looking the actual data, that in 1996 Germany had a very cold year (maximum HDD) but similar number of deaths as in cluster 1. 2010 was the second coldest year but the number of deaths coincided with the centroid from cluster 2 as is seen in Figure 107.

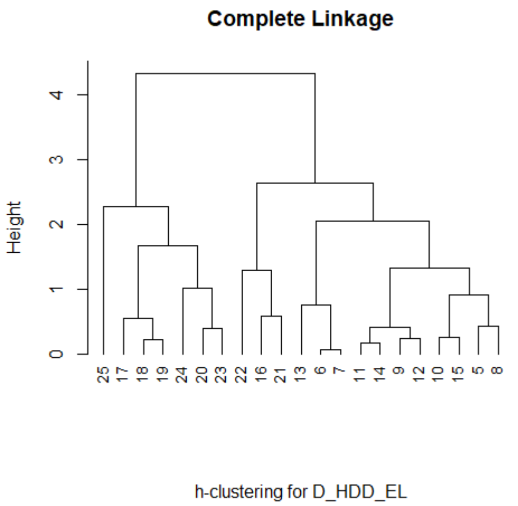


Figure 108: Hierarchical clustering dendrogram for D – HDDD EL

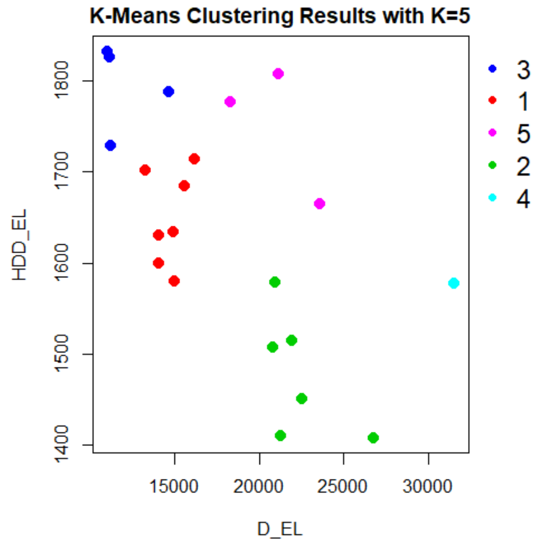


Figure 109: K-means cluster separation in scatter plot for D – HDD EL

From Figure 108 initially seems good to cut the dendrogram from height level three to receive two major clusters from the hierarchical tree. However, data for Deaths of old people from respiratory diseases and the number of Heating Degree Days in Greece present more complexity in the scatter plot. In order to clarify the separation, the dendrogram can be cut at level 1.7 and five clusters can separate the data as is seen in Figure 109 with the help of K-means clustering.

Cluster three includes the years 1995 – 1997 and 2003 with 12,002 dead persons in average alongside 1,793.168 number of heating degree days. Cluster one involves the years 1998 – 2005 (excluding 2003) with more dead persons in average (14,734) and slightly less heating degree days (1,649.234). Cluster five contains the years 2006, 2011 and 2012 with higher number of dead aged people in average (21,020) and number of heating degree days close to cluster three (1,749.340) in average. Cluster two includes the years 2007 - 2010 and 2013 - 2014 with slightly higher number of dead old persons, although the number of heating degree days is much fewer in average (22,388 and 1,478.378 respectively). Last, there is an outlier as a sole cluster. Year 2015 has peak number of dead persons (over 31 thousand) and 1,577.850 number of heating degree days. This concludes that there is no relation between the two indices for Greece.

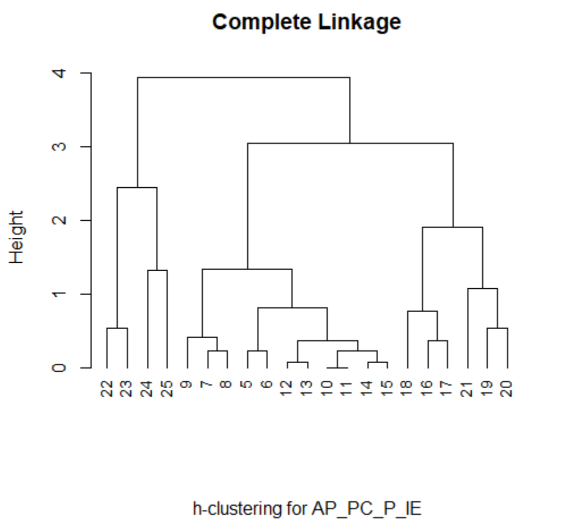


Figure 110: Hierarchical clustering dendrogram for AP - PC_P IE

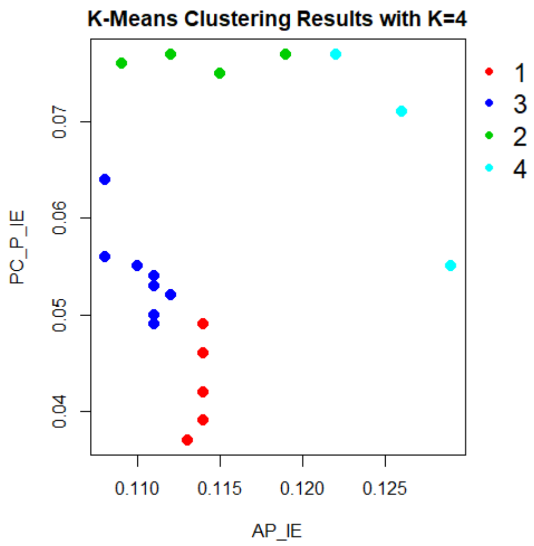


Figure 111: K-means cluster separation in scatter plot for AP - PC_P IE

A dendrogram is deployed in Figure 110 for the relation between percentage of Aging Population and the percentage of Pensions from GDP in Ireland. The tree can be cut at height level 2 in order to have four clusters. In Figure 111, the separation of data can be observed into four distinct clusters through K-means method with K = 4.

Using K-means with K equals four we have the cluster #1 for the period 1995 - 1999, this has average AP = 0.114 and PC_P = 0.043. Cluster #3 involves the years 2000 - 2008 with an average AP equal to 0.110, improved compared to first cluster, and PC_P equal to 0.054. Then the second cluster includes years 2009 - 2012 with slightly worse AP (0.114) and PC_P steadily rising (0.076). The last cluster involves the years 2013 - 2015 with AP = 0.126 and PC_P 0.068 in average, an improvement on par with cluster #2.

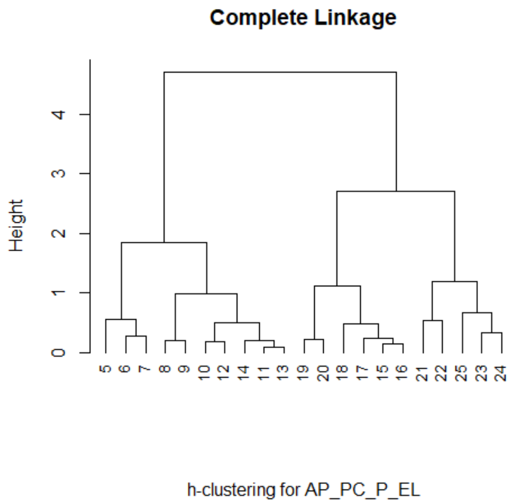


Figure 112: Hierarchical clustering dendrogram for AP - PC_P EL

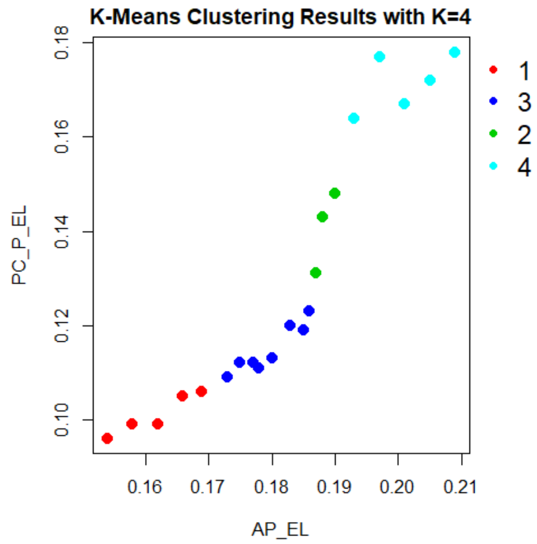


Figure 113: K-means cluster separation in scatter plot for AP - PC_P EL

Figure 112 shows the hierarchical dendrogram cluster result for the percentage of Pensions from GDP and the percentage of aging population in Greece. It is apparent that the tree can be cut at height level 1.5 and the suitable number of clusters is four. Figure 113 illustrates the K-means clustering separation with K equals four and indeed there is some sort of uniformity.

Cluster one starts in 1995 and ends in 1999 with AP = 0.162 and PC_P = 0.101 in average. Later, cluster three begins in 2000 and finishes in 2007 with both the means of indices higher. The difference is in cluster two, because AP is slightly higher than previous cluster (0.188) but there is a surge in PC_P than before (0.141) in average values. Last cluster contains the years 2011 - 2015 and returns with the positive relation among the two indices.

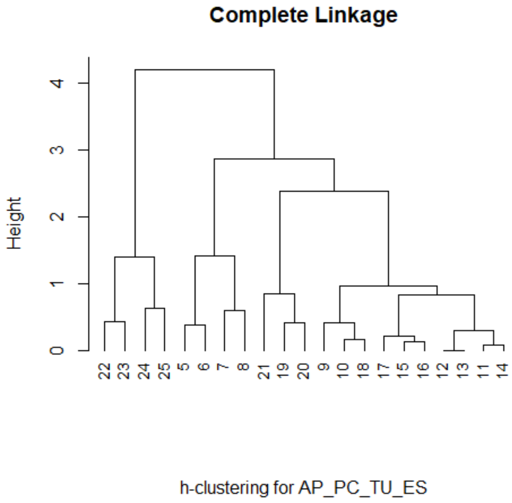


Figure 114: Hierarchical clustering dendrogram for AP - PC_TU ES

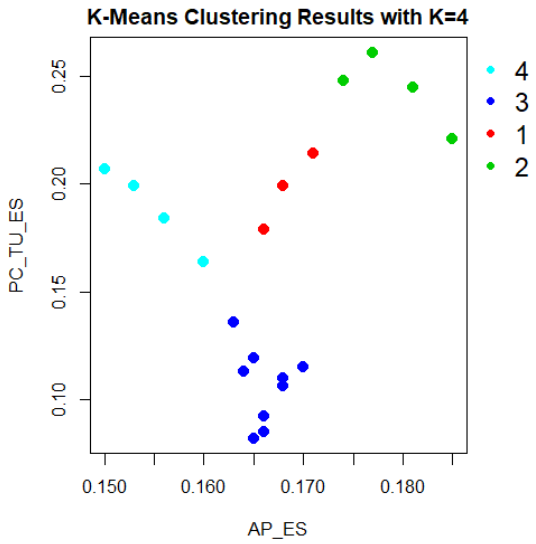


Figure 115: K-means cluster separation in scatter plot for AP - PC_TU ES

In Figure 114 is clear that we can cut the dendrogram in height level 2. Thus, we can have four clusters as result from hierarchical clustering in the relation between percentage of Aging Population and the percentage of Total Unemployment in Spain. The Figure 115 indicates the separation through K-means clustering with K = 4.

Cluster four begins in 1995 and ends in 1998 with average AP equal to 0.155and PC_TU equal to 0.189. Cluster three starts in 1999 till 2008 with improved average PC_TU (0.107) and higher AP (0.167). Cluster one involves the years 2009 - 2011 with slightly worse AP (0.168) and higher PC_TU (0.197) in average. Last cluster includes the years 2012 - 2015 with AP equal to 0.179 and PC_TU equal to 0.244. One could asses that total unemployment affects the aging population with some delay but without the ability to alter the direction in the growth of elderly.

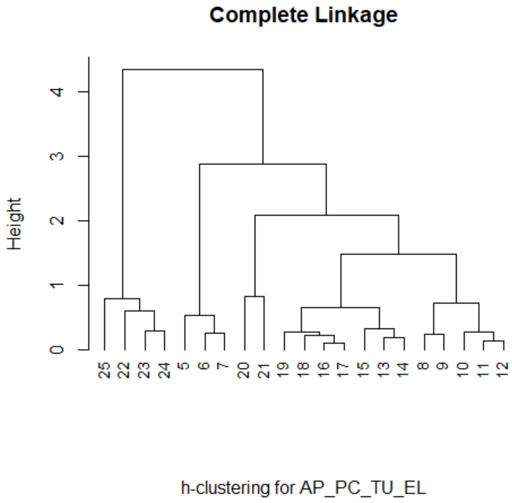


Figure 116: Hierarchical clustering dendrogram for AP - PC_TU EL

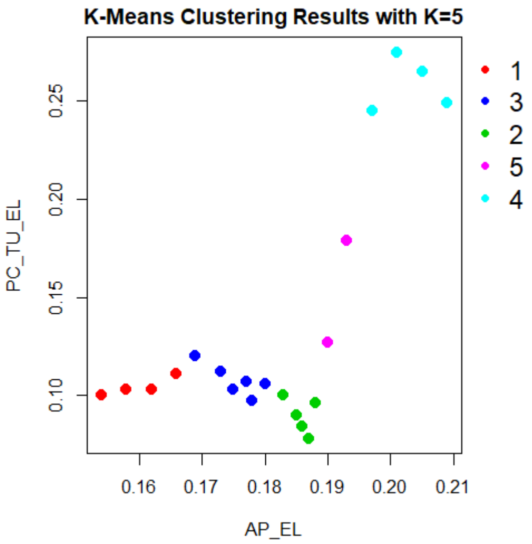


Figure 117: K-means cluster separation in scatter plot for AP - PC_TU EL

Figures 116 shows a dendrogram, that stems from hierarchical clustering for the relation between percentage of Aging Population and percentage of Total unemployment in Greece. It is apparent that the suitable number or clusters is five as we cut the tree in height level one. In Figure 117 is drawn how the data are separated through K-means clustering with K equals five.

Cluster one includes the years 1995 - 1998 with AP = 0.16 and PC_TU = 0.104. Cluster three begins in 1999 until 2004, where both indices slightly rise. Cluster two starts in 2005 and ends in 2009, with improvement in average total unemployment (9%) but still growth is observed in aging population (18.6%). Cluster five includes only 2010 and 2011 with a sudden surge in average total unemployment (15.3%) and steadily risen aging population (19.1%). Last cluster involves the years 2012 - 2015 with both indices at average peak. Thus, total unemployment could be excluded as factor in Greek aging population growth.

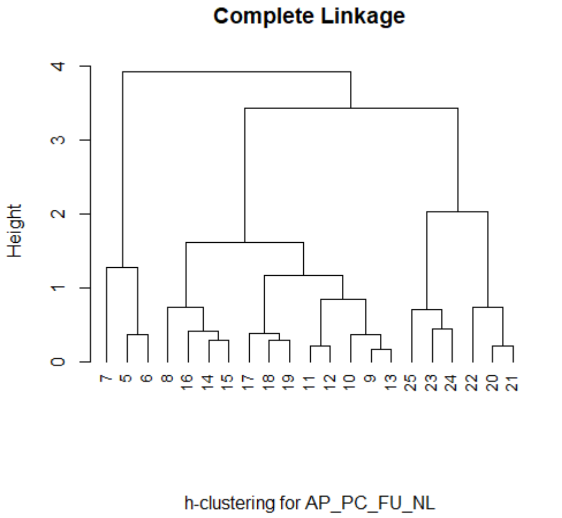


Figure 118: Hierarchical clustering dendrogram for AP - PC_FU NL

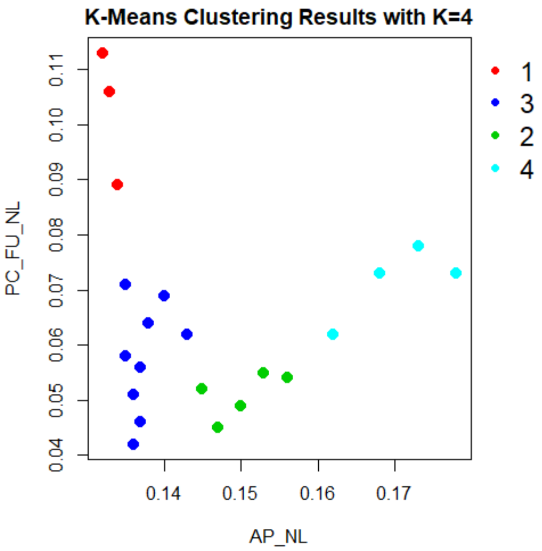


Figure 119: K-means cluster separation in scatter plot for AP - PC_FU NL

Figure 118 deploys a dendrogram from hierarchical clustering for the percentage of Aging Population along with the percentage of Female Unemployment for the Netherlands. Height level two indicates the tree can be cut into four clusters. Then, in Figure 119 K-means clustering is applied with K equals four.

The first cluster starts in 1995 until 1997, with average AP = 0.133 and PC_FU = 0.103 for the Netherlands. Cluster three begins in 1998 and finishes in 2006. In this period average female unemployment is significantly improved at 5.8% and aging population almost stable (0.137). Cluster two involves the years 2002 - 2011 with average AP grown at 0.15 and slightly improved PC_FU at 0.051 in average. Last cluster includes the years 2012 until 2015 with both indices worse in average. An assessment of this image is that if female unemployment improves, then growth of aging population is delayed at best in Netherlands.

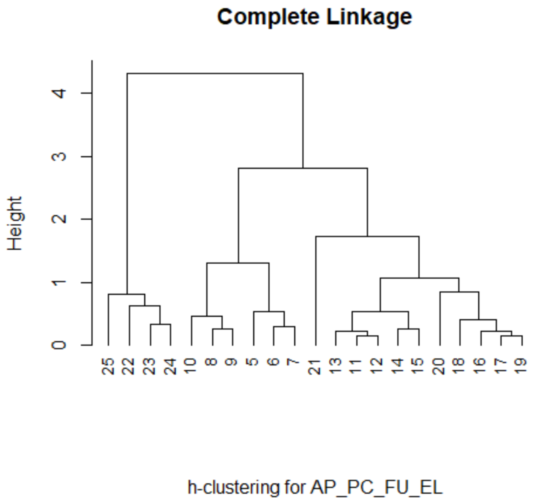


Figure 120: Hierarchical clustering dendrogram for AP - PC_FU EL

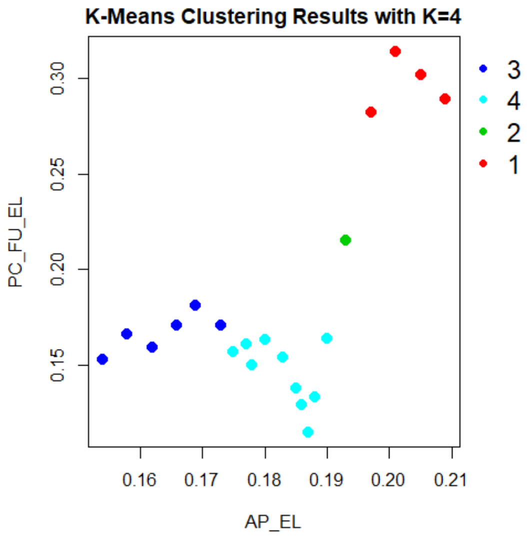


Figure 121: K-means cluster separation in scatter plot for AP - PC_FU EL

Figure 120 deploys a dendrogram as result from hierarchical clustering. The analysis concerns the percentage of Aging Population and the percentage of Female Unemployment in Greece. There can be four clusters if the tree is cut at height level 1.5. Then Figure 121 portrays a K-means clustering separation of the data with K = 4.

Cluster three starts in 1995 until 2000 with AP = 0.164 and PC_FU 0.167 in average. Next cluster involves the years 2001 - 2010 with lower female unemployment in average (0.146), though there is higher percentage of aging population in average (0.183). Cluster two contains an outlier for the year 2011, where there is a sudden surge in PC_FU (0.215) and slightly higher AP (0.193). Last cluster includes the years 2012 - 2015 with both indices increased. This may implicate no direct relation between the two.

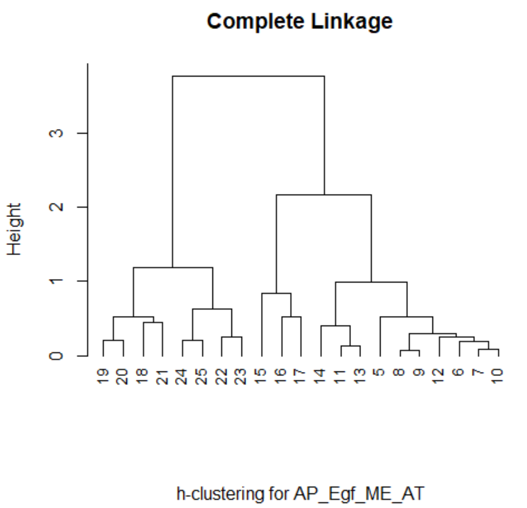


Figure 122: Hierarchical clustering dendrogram for AP – Egf_ME AT

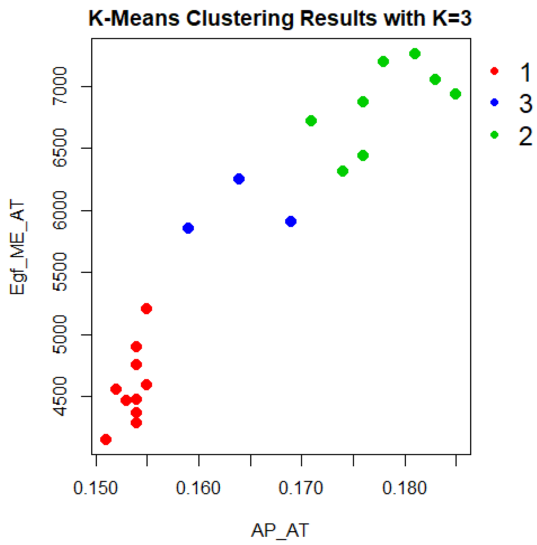


Figure 123: K-means cluster separation in scatter plot for AP – Egf_ME AT

In Figure 122 a dendrogram is drawn through hierarchical clustering in data of Austria that include the percentage of Aging Population and the consumption for Electricity, gas and other fuels in million euro. Height level two indicates that three clusters can be made from cutting the tree. Applying K-means with K=3 on the data, the separation is illustrated in Figure 123.

First cluster starts in 1995 and ends in 2004 with 4,575.680 million euros in Egf_ME in average and with 15.4 % average AP. Cluster three continues through period 2005 up to2007 with a surge in average electricity and other fuels consumption (6,006.200) and 16.4% average percentage for elderly population. Last cluster includes the years 2008 – 2015 and the growth in average AP evolves (0.178) with a slight increase in average Egf_ME (6,849.425). Considering the inflation in oil prices for the period of cluster three, the growth in aging population goes hand in hand with the increase in electricity and other fuels consumption.

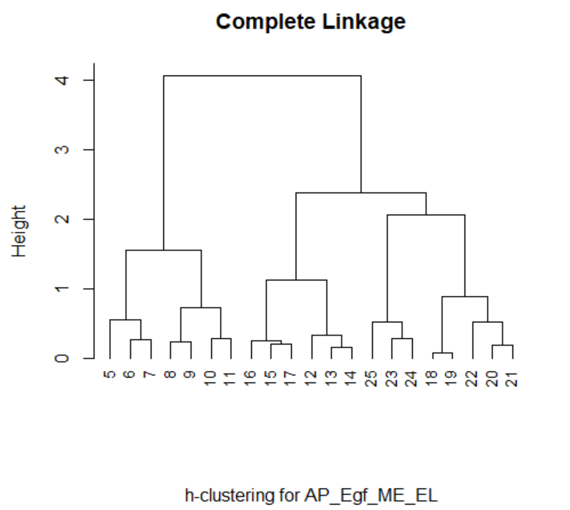


Figure 124: Hierarchical clustering dendrogram for AP – Egf_ME EL

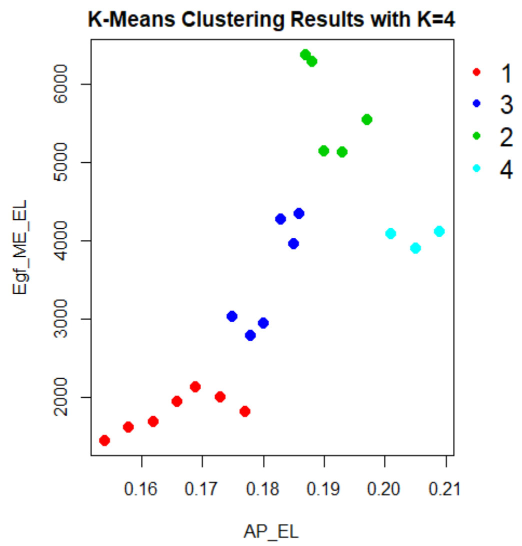


Figure 125: K-means cluster separation in scatter plot for AP - Egf_ME EL

Figure 124 shows the dendrogram as result from hierarchical clustering for the relation between percentage of Aging Population and consumption of Electricity, gas and other fuels in million euros in Greece. At height level 2 there can be four major clusters. The next step is to apply K-means clustering with K equals four as is drawn in Figure 139.

First cluster begins in 1995 and finishes in 2001. The centroids of this cluster are 0.166 for AP and 1,807.543 for Egf_ME. Cluster three starts in 2002 and ends in 2007 with elevated aging population (0.181) and even more in electricity and other fuels consumption (3,558.300) in average. Cluster two involves the years 2008 – 2012 with higher AP (0.191) and a surge in Egf_ME (5,700.060) in average. Last cluster includes the years 2013 - 2015 with peak in average AP equal to 20.5% and significantly lower consumption in Egf_ME (3,601.263). Overall it seems that growth in elderly signals the rise in consumption for electricity and other fuels, but the cluster #4 regresses from that direction, so other factors should be considered as well.

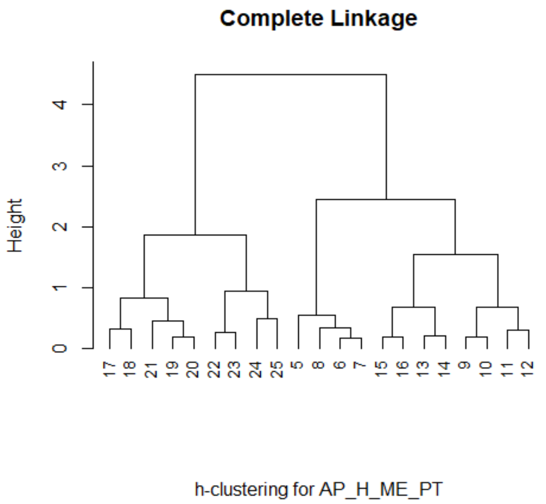


Figure 126: Hierarchical clustering dendrogram for AP - H_ME PT

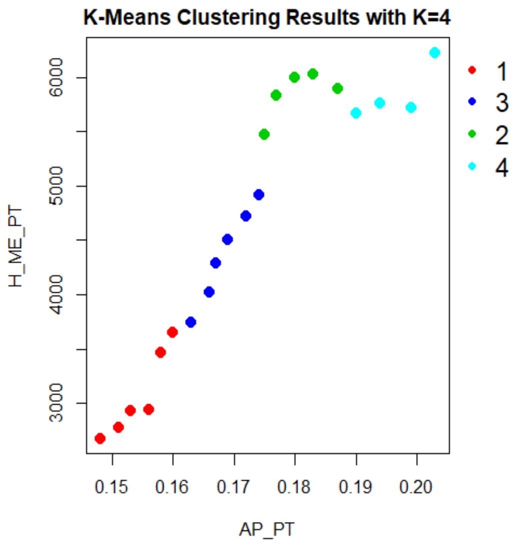


Figure 127: K-means cluster separation in scatter plot for AP - H_ME PT

Figure 126 unfolds the dendrogram as result from hierarchical clustering. The method is applied for the relation between the percentage of Aging Population and the consumption in Health in million euros for Portugal. If the tree is cut in height level 1.7 there can be four distinct clusters. In Figure 127 is illustrated the separation through K-means clustering with K = 4.First cluster starts in 1995 then ends in 2000 with average AP equals 0.154 and average H_ME 3,073.183. Cluster three begins in 2001 and finishes in 2006 with elevated percentage of elderly population and also elevated consumption in health (0.169 and 4,367.183 respectively). Cluster two involves the years 2007 – 2011 with both indices higher (AP = 0.18 and H_ME = 5.844.540). Last cluster includes the years 2012 – 2015 and although AP grew there was a subtle reduction in H_ME. The reader should mind that Portugal had undergone a bailout program from 2011 – 2014 that would explain budget cuts in health. (Axel, 2016)

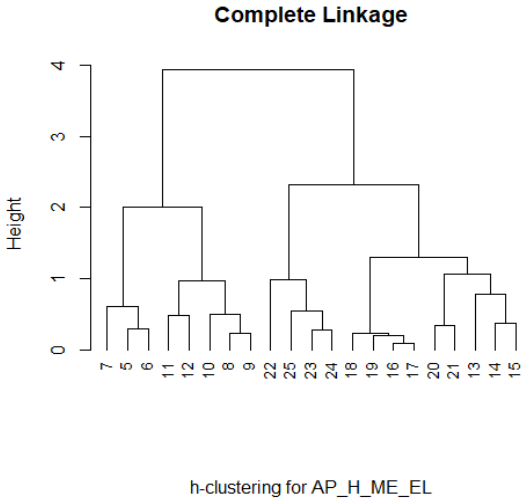


Figure 128: Hierarchical clustering dendrogram for AP - H_ME EL

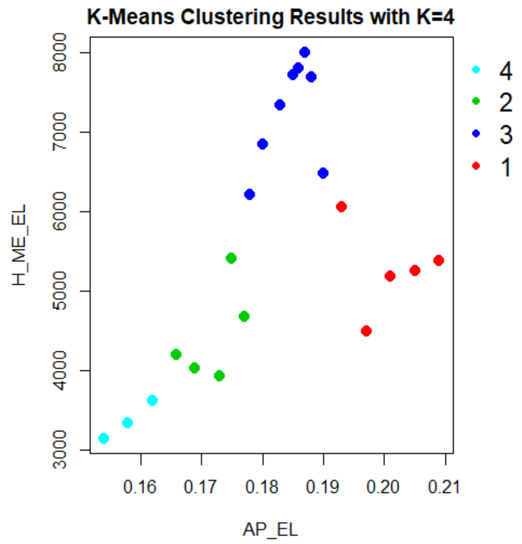


Figure 129: K-means cluster separation in scatter plot for AP - H_ME EL

Figure 128 shows how the dendrogram unfolds from hierarchical clustering. It is applied in the dataset that contains the percentage of Aging Population and the consumption of Health in million euros for Greece. Since the tree can be cut at height level 1.5 then K-means is applied with K = 4 as is illustrated in Figure 129.

Cluster four starts in1995 until 1997 with 15.8% aging population and 3,373.033 health consumption in average. Cluster two begins in 1998 and finishes in 2002 with both indices higher. Next cluster involves the years 2003 - 2010 with 0.185 AP and a dramatic surge in H_ME (7,265.925) in average. Last cluster includes the years 2011 – 2015 with even higher AP (0.201) and diminishing H_ME (5,272.000). The image is similar with Portugal due to the fact that Greece had also undergone two bailout programs for its debt crisis since 2010. Only the budget cuts here seem more drastic in terms of volume.

CHAPTER 5

Conclusions

The population of persons aged 65 or over is steadily rising in Europe. This is a fact that was overviewed in chapter 2 through descriptive statistics. The sole country that still has better results in relative demographic matters is Ireland with median Potential Support Ratio 5.810. This translates that almost six persons aged 14 to 64 can sustain with their contributions the pension and health system of the nation. Greece is in more critical condition. Although both countries were affected from the recession in 2008, the Greek nation is still on the way to recovery and it's probably the main reason why many people in productive age decided to emigrate. Alongside Germany and Portugal, those three countries have reached over 20% of elderly population.

On the other end of the spectrum, Germany is indisputably the biggest economy in Europe and differentiates enough from Greece and Portugal in terms of unemployment. This may indicate at first glance that unemployment (total or female) does not affect the change of pace in the growth of elder population. For example, Austria has the least of mean percentage of Total Unemployment (0.049) and as is seen in Figure 83 has higher percentage of aging population than Ireland and other countries. This assessment can be verified through chapter 3 and 4 with the regression and cluster analysis, that there is no direct relation at least. However, unemployment has other indirect ramifications that are not studied on this paper, such as birth rates or the mean initial age to give birth and even the sustainability of health and insurance system.

Chapter 1 gave a general description of the pension systems for the studied countries. All of them have at least a compulsory state system, which is either funded through contributions of employee's or tax money. In this paper, pensions are examined as a percentage of Gross Domestic Product in order to get a glimpse of the relation with the aging population. For most of the countries seems no linear. This can be connected with legislation acts, which redefine the statutory age for a pension beneficiary and even penalties are designed for early retirement. Chapter 2 showed that Greece and Portugal expend the most for pensions in terms of GDP partition and Ireland the least. If the reader is to look for a linear relation with the aging population, chapter 3 pinpoints Denmark, Greece and Portugal for that matter, since R square values are high even though nonlinearity exists at some point in data.

Another economic aspect is that of consumption, particularly that of Electricity, gas and other fuels and of Health. There is much difference in terms of scale among the studied countries and it indicates dissimilarities in economic power. It was anticipated to receive positive correlation results alongside with the percentage of aging population, specifically with health consumption. Instead, chapter 3 indicates fluctuations with some exceptions. In electricity and other fuels consumption there is linear relation with percentage of elderly for Germany, Austria and Portugal. The same applies with the health consumption. Greece could not have a linear model due to many budget cuts and reforms that stem from three bailout programs for its debt crisis. This concurs with cluster analysis, which distinguishes the periods that those programs were in effect. However, positive correlation persists and that is why many studies conclude that health industries have room for growth.

Besides health, mortality was examined in this survey. What was found is that deaths from respiratory diseases for people aged 65 or over are unrelated with the number of heating degree days. What is more surprising is the fact that for some countries regression analysis produces negative coefficients, which translates that the more heating degree days the less are the deaths from respiratory diseases in elderly. Great exception to this is Portugal with positive coefficient for HDD and R square value little over 80%. We can assess that other factors influence those deaths and not the lack of temperature for indoor heating for most studied countries. Perhaps data for hospitalized patients with respiratory diseases could be more related to the number of heating degree days, but that would be a subject for another researcher.

Finaly, cluster analysis distincts the periods where events like the economic crisis occurred or even bailout programs were in effect. This is most obvious in Portugal and Greece were socioeconomic events had immersive ramifications. On the other hand, it is more difficult to distinguish periods with deaths from respiratory diseases related to the number of heating degree days. As is seen in chapter 4, there are no continuing chronologialy periods in mortality area but overlapping.

If the reader wants more content that was omitted in the main body of this survey, Appendix A offers visualization info alongside extensive summary statistics and Appendix B the code written in R.

Bibliography

Axel, B. (2016, September 22). *IMF says Portugal bailout only "qualified success", leaving unfinished business.* Retrieved from www.reuters.com: https://www.reuters.com/article/eurozone-portugal-imf-idUSL9N13C00P

Bourdelais, P. (1999). Demographic aging: A notion to revisit. *The History of the Family*, 31-50.

Bucher-Koenen, T. &. (2011). Financial literacy and retirement planning in Germany. *Journal of Pension Economics and Finance*, 565-584.

David E. Bloom, A. B.-S. (2011, May). Population Aging: Facts, Challenges, and Responses. *Program on the Global Demography of Aging*. Harvard Initiative for Global Health.

Economist. (2011, April 7). *"Pensions: 70 or Bust".* Retrieved from www.economist.com: http://www.economist.com/node/18529505?Story_ID=18529505

Fund, I. M. (2018, October). *World Economic Outlook*. Retrieved from www.imf.org: https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/EU

Getzen, T. E. (1992). Population Aging and the Growth of Health Expenditures. *Journal of Gerontology*, Volume 47, Issue 3, Pages S98–S104,.

Hope, K. (2018, August 16). *Financial Times Europe*. Retrieved November 2018, from https://www.ft.com: https://www.ft.com/content/24866436-9f9f-11e8-85da-eeb7a9ce36e4

Innaculada, V. M. (2003). *The Spanish Pension System: The possibility of introducing notional defiined contribution accounts.* EconWPA Public Economics.

IOPS. (2017, December). *Resources.* Retrieved from www.iopsweb.org: http://www.iopsweb.org/resources/IOPS-Profile-UK-2017.pdf

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An Introduction to Statistical Learning with applications in R.* New York: Springer.

Jørgen, G. A. (2011). *Denmark: The Silent Revolution towards a Multipillar Pension System.* Oxford: Oxford University Press.

Keller, G. (2014). *Statistics for Management and Economics.* Cengage Learning.

National Actuarial Authority of Greece. (2015). Greek Pension System Fiche. *Ageing Projections Exercise*. National Actuarial Authority.

OECD AT. (2017). *Pensions at a glance: Austria.* Retrieved from www.oecd.org: https://www.oecd.org/els/public-pensions/PAG2017-country-profile-Austria.pdf

OECD IE. (2014). *Reviews of Pensions Systems: Ireland.* OECD Publishing.

OECD NL. (2017). *Pension at a glance: Netherlands.* Retrieved from www.oecd.org: https://www.oecd.org/els/public-pensions/PAG2017-country-profile-Netherlands.pdf

OECD PT. (2017). *Pensions at a glance: Portugal.* Retrieved from www.oecd.org: https://www.oecd.org/els/public-pensions/PAG2017-country-profile-Portugal.pdf

Spengler, R. L. (1980). *The economics of individual and population aging.* Cambridge: Press Syndicate of the University of Cambridge.

Traa, B. (2018, September 2). *business*. Retrieved from www.ekathimerini.com: http://www.ekathimerini.com/232223/article/ekathimerini/business/bob-traa-demographic-slowdown-has-taken-toll-on-greek-potential-growth

Turner, A. (2009). Population ageing: what should we worry about? London, UK: Philosophical Transactions of the Royal Society.

Table of Authorities

Books

Journal Articles

Reports

APPENDIX A

A1: Summary statistics for all variables in Denmark

	DK_PS_R	*DK _AP*	*DK _D*	*DK _HDD*	*DK _PC_P*	*DK _PC_TU*	*DK _PC_FU*	*DK _Egf_ME*	*DK _H_ME*
Mean	4.172	0.160	13,207.381	3,304.011	0.116	0.056	0.060	6,137.761	2,764.287
Standard Error	0.074	0.003	126.656	59.415	0.003	0.003	0.002	233.945	144.436
Median	4.350	0.154	13,220.000	3,235.750	0.112	0.055	0.060	6,142.900	2,811.300
Mode	#N/A	#N/A	#N/A	#N/A	0.112	0.052	0.075	#N/A	#N/A
Standard Deviation	0.384	0.014	580.411	284.947	0.012	0.012	0.012	1,121.963	692.689
Sample Variance	0.147	0.000	336,877.348	81,194.560	0.000	0.000	0.000	1,258,800.172	479,818.598
Kurtosis	-0.344	0.046	-0.762	1.796	-0.994	-0.898	-0.604	-1.161	-1.471
Skewness	-1.091	1.209	0.136	1.236	0.660	0.019	-0.105	-0.146	-0.078
Range	1.131	0.042	2,115.000	1,195.390	0.038	0.042	0.044	3,654.100	2,118.100
Minimum	3.371	0.148	12,245.000	2,850.870	0.102	0.034	0.037	4,235.500	1,677.900
Maximum	4.502	0.191	14,360.000	4,046.260	0.140	0.076	0.081	7,889.600	3,796.000
Sum	112.634	4.315	277,355.000	75,992.250	2.437	1.288	1.380	141,168.500	63,578.600
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.955	0.150	12,764.000	3,117.310	0.107	0.047	0.052	5,290.100	2,139.350
Q3	4.458	0.166	13,636.000	3,447.570	0.127	0.065	0.067	7,090.550	3,401.500

A2: Summary statistics for all variables in Germany

	DE_PS_R	DE_AP	DE_D	DE_HDD	DE_PC_P	DE_PC_TU	DE_PC_FU	DE_Egf_ME	DE_H_ME
Mean	3.778	0.181	107,782.762	3,093.182	0.124	0.076	0.076	54,981.904	58,956.343
Standard Error	0.110	0.005	5,303.593	52.839	0.001	0.004	0.005	2,622.587	3,286.379
Median	3.736	0.180	97,217.000	3,080.550	0.124	0.079	0.083	56,002.000	55,756.000
Mode	#N/A	#N/A	#N/A	#N/A	0.125	0.086	0.101	#N/A	#N/A
Standard Deviation	0.570	0.024	24,304.116	253.406	0.004	0.021	0.024	12,577.487	15,760.919
Sample Variance	0.325	0.001	590,690,036.890	64,214.774	0.000	0.000	0.001	158,193,187.408	248,406,571.175
Kurtosis	-1.723	-1.784	-1.183	2.491	-1.182	-0.953	-1.205	-1.565	-1.357
Skewness	0.154	-0.018	0.772	1.212	0.034	-0.308	-0.442	-0.126	0.195
Range	1.526	0.063	72,296.000	1,154.640	0.014	0.074	0.076	38,021.000	48,066.500
Minimum	3.082	0.149	82,160.000	2,660.860	0.117	0.038	0.033	37,040.000	36,909.500
Maximum	4.608	0.212	154,456.000	3,815.500	0.131	0.112	0.109	75,061.000	84,976.000
Sum	102.019	4.891	2,263,438.000	71,143.190	2.600	1.748	1.759	1,264,583.800	1,355,995.900
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.188	0.157	88,607.000	2,969.350	0.120	0.056	0.054	42,442.600	44,929.000
Q3	4.336	0.207	134,170.000	3,167.435	0.127	0.092	0.098	66,134.000	73,343.000

A3: Summary statistics for all variables in Ireland

	IE_PS_R	IE_AP	IE_D	IE_HDD	IE_PC_P	IE_PC_TU	IE_PC_FU	IE_Egf_ME	IE_H_ME
Mean	5.753	0.116	8,908.286	2,736.083	0.057	0.087	0.080	2,296.517	2,595.883
Standard Error	0.082	0.001	269.017	30.235	0.003	0.008	0.007	169.036	278.499
Median	5.810	0.114	8,700.000	2,702.730	0.054	0.075	0.073	2,573.700	2,630.200
Mode	#N/A	#N/A	#N/A	#N/A	0.077	0.047	0.047	#N/A	#N/A
Standard Deviation	0.425	0.007	1,232.789	145.001	0.013	0.040	0.032	810.668	1,335.634
Sample Variance	0.181	0.000	1,519,769.714	21,025.282	0.000	0.002	0.001	657,183.082	1,783,918.276
Kurtosis	-0.504	1.781	0.086	2.179	-1.092	-1.369	-1.613	-1.682	-1.530
Skewness	-0.459	1.601	0.927	1.106	0.366	0.427	0.336	-0.247	0.053
Range	1.550	0.027	4,553.000	642.770	0.040	0.113	0.085	2,240.200	3,836.200
Minimum	4.840	0.108	7,242.000	2,525.790	0.037	0.042	0.043	1,057.100	724.200
Maximum	6.391	0.135	11,795.000	3,168.560	0.077	0.155	0.128	3,297.300	4,560.400
Sum	155.324	3.120	187,074.000	62,929.920	1.206	2.002	1.830	52,819.900	59,705.300
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	5.477	0.111	8,024.000	2,629.835	0.049	0.048	0.049	1,487.650	1,334.250
Q3	6.091	0.115	9,568.000	2,807.425	0.071	0.121	0.112	2,988.950	3,797.650

A4: Summary statistics for all variables in Greece

	EL_PS_R	*EL_AP*	*EL_D*	*EL_HDD*	*EL_PC_P*	*EL_PC_TU*	*EL_PC_FU*	*EL_Egf_ME*	*EL_H_ME*
Mean	3.797	0.178	18,100.000	1,631.950	0.129	0.144	0.192	3,607.478	5,581.183
Standard Error	0.102	0.004	1,195.412	27.304	0.006	0.014	0.013	314.918	314.989
Median	3.696	0.180	16,165.000	1,634.520	0.119	0.107	0.164	3,962.300	5,412.900
Mode	#N/A	#N/A	#N/A	#N/A	0.099	#N/A	0.171	#N/A	#N/A
Standard Deviation	0.531	0.022	5,478.065	130.944	0.028	0.067	0.062	1,510.292	1,510.633
Sample Variance	0.282	0.000	30,009,201.600	17,146.363	0.001	0.004	0.004	2,280,982.021	2,282,012.032
Kurtosis	-0.628	-0.672	0.163	-0.942	-1.002	-0.685	-0.683	-0.952	-1.087
Skewness	0.411	-0.186	0.714	-0.166	0.715	1.021	0.920	0.187	0.105
Range	1.895	0.078	20,562.000	423.820	0.082	0.197	0.199	4,931.300	4,860.500
Minimum	2.972	0.138	10,999.000	1,408.340	0.096	0.078	0.115	1,450.100	3,148.300
Maximum	4.867	0.215	31,561.000	1,832.160	0.178	0.275	0.314	6,381.400	8,008.800
Sum	102.520	4.817	380,100.000	37,534.840	2.704	3.301	4.410	82,972.000	128,367.200
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.462	0.164	14,096.000	1,546.595	0.109	0.100	0.154	2,068.350	4,349.150
Q3	4.146	0.191	21,256.000	1,721.185	0.148	0.197	0.238	4,314.600	6,663.950

A5: Summary statistics for all variables in Spain

	ES_PS_R	*ES_AP*	*ES_D*	*ES_HDD*	*ES_PC_P*	*ES_PC_TU*	*ES_PC_FU*	*ES_Egf_ME*	*ES_H_ME*
Mean	4.151	0.165	92,812.381	1,771.557	0.101	0.164	0.196	15,722.087	18,741.213
Standard Error	0.070	0.003	2,056.170	28.015	0.003	0.012	0.011	1,485.581	1,286.700
Median	4.145	0.166	93,797.000	1,816.320	0.097	0.172	0.199	14,229.000	19,414.000
Mode	#N/A	#N/A	#N/A	#N/A	0.094	0.199	0.251	#N/A	#N/A
Standard Deviation	0.362	0.013	9,422.554	134.354	0.013	0.056	0.053	7,124.597	6,170.799
Sample Variance	0.131	0.000	88,784,528.848	18,051.108	0.000	0.003	0.003	50,759,887.815	38,078,757.945
Kurtosis	-0.261	-0.160	0.174	-1.122	-0.071	-1.295	-1.203	-1.728	-1.403
Skewness	0.073	-0.241	0.100	-0.399	1.092	0.111	-0.118	0.239	-0.130
Range	1.373	0.052	38,852.000	434.540	0.041	0.179	0.172	17,971.100	18,968.700
Minimum	3.480	0.138	74,940.000	1,537.130	0.087	0.082	0.107	7,315.900	9,203.300
Maximum	4.852	0.190	113,792.000	1,971.670	0.128	0.261	0.279	25,287.000	28,172.000
Sum	112.066	4.451	1,949,060.000	40,745.810	2.113	3.762	4.506	361,608.000	431,047.900
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.996	0.158	86,982.000	1,672.485	0.090	0.114	0.155	9,068.500	13,319.000
Q3	4.340	0.171	97,048.000	1,867.700	0.105	0.203	0.244	23,940.000	23,828.000

A6: Summary statistics for all variables in Netherlands

	NL_PS_R	*NL_AP*	*NL_D*	*NL_HDD*	*NL_PC_P*	*NL_PC_TU*	*NL_PC_FU*	*NL_Egf_ME*	*NL_H_ME*
Mean	4.666	0.146	26,677.429	2,722.888	0.122	0.054	0.065	8,877.635	9,395.365
Standard Error	0.109	0.003	725.712	54.266	0.001	0.003	0.004	448.271	359.748
Median	4.882	0.138	25,474.000	2,680.270	0.121	0.050	0.062	9,854.000	8,950.000
Mode	#N/A	#N/A	#N/A	#N/A	0.130	0.037	0.062	#N/A	#N/A
Standard Deviation	0.565	0.017	3,325.628	260.250	0.007	0.014	0.018	2,149.832	1,725.290
Sample Variance	0.319	0.000	11,059,804.057	67,729.904	0.000	0.000	0.000	4,621,779.507	2,976,624.155
Kurtosis	-0.603	-0.097	-1.235	2.498	-1.213	-0.665	1.660	-1.405	-0.652
Skewness	-0.799	1.072	0.473	1.326	0.326	0.414	1.344	-0.356	0.401
Range	1.821	0.056	10,975.000	1,164.200	0.022	0.052	0.071	6,496.900	6,243.000
Minimum	3.526	0.129	22,041.000	2,284.960	0.112	0.031	0.042	5,553.100	6,770.000
Maximum	5.347	0.185	33,016.000	3,449.160	0.134	0.083	0.113	12,050.000	13,013.000
Sum	125.978	3.955	560,226.000	62,626.420	2.560	1.243	1.486	204,185.600	216,093.400
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	4.340	0.134	23,858.000	2,570.620	0.116	0.043	0.053	6,825.100	8,163.400
Q3	5.075	0.154	29,657.000	2,807.135	0.129	0.063	0.072	10,442.500	10,703.000

A7: Summary statistics for all variables in Austria

	AT_PS_R	*AT_AP*	*AT_D*	*AT_HDD*	*AT_PC_P*	*AT_PC_TU*	*AT_PC_FU*	*AT_Egf_ME*	*AT_H_ME*
Mean	4.147	0.164	9,641.476	3,593.645	0.140	0.049	0.049	5,768.000	5,398.374
Standard Error	0.063	0.003	239.937	45.773	0.001	0.001	0.001	238.541	250.271
Median	4.367	0.155	9,611.000	3,640.030	0.140	0.048	0.049	5,909.800	5,325.800
Mode	#N/A	#N/A	#N/A	#N/A	0.139	0.047	0.046	#N/A	#N/A
Standard Deviation	0.328	0.013	1,099.528	219.520	0.004	0.006	0.005	1,144.004	1,200.258
Sample Variance	0.107	0.000	1,208,962.362	48,189.129	0.000	0.000	0.000	1,308,745.214	1,440,620.293
Kurtosis	-1.560	-1.504	-0.056	0.511	-0.331	-1.059	-0.883	-1.754	-1.036
Skewness	-0.417	0.458	0.115	0.236	-0.061	0.001	0.227	-0.095	0.105
Range	0.901	0.036	4,235.000	993.390	0.016	0.021	0.018	3,109.200	3,909.200
Minimum	3.620	0.149	7,857.000	3,124.730	0.132	0.039	0.041	4,153.400	3,580.600
Maximum	4.521	0.185	12,092.000	4,118.120	0.148	0.060	0.059	7,262.600	7,489.800
Sum	111.959	4.420	202,471.000	82,653.840	2.949	1.128	1.133	132,664.000	124,162.600
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.828	0.153	9,153.000	3,435.180	0.139	0.045	0.046	4,577.100	4,457.350
Q3	4.405	0.176	10,359.000	3,708.375	0.143	0.055	0.053	6,903.950	6,274.100

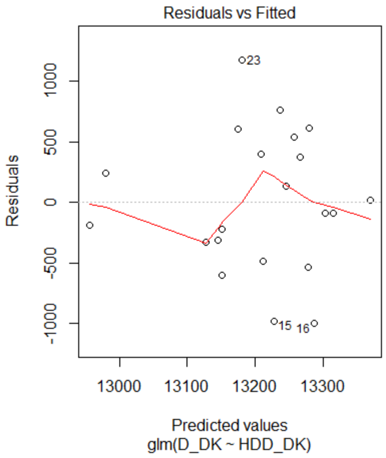
A8: Summary statistics for all variables in Portugal

	PT_PS_R	*PT_AP*	*PT_D*	*PT_HDD*	*PT_PC_P*	*PT_PC_TU*	*PT_PC_FU*	*PT_Egf_ME*	*PT_H_ME*
Mean	3.969	0.171	22,341.000	1,195.703	0.120	0.094	0.096	3,136.635	4,799.587
Standard Error	0.100	0.004	816.215	25.957	0.005	0.007	0.007	210.964	271.540
Median	3.956	0.169	22,918.000	1,239.350	0.120	0.088	0.090	3,227.500	4,912.900
Mode	#N/A	#N/A	#N/A	#N/A	0.095	0.051	0.088	#N/A	#N/A
Standard Deviation	0.521	0.021	3,740.369	124.486	0.022	0.033	0.033	1,011.748	1,302.262
Sample Variance	0.271	0.000	13,990,361.900	15,496.804	0.000	0.001	0.001	1,023,634.673	1,695,885.917
Kurtosis	-0.963	-0.807	-0.948	0.568	-1.263	-0.317	-0.379	-1.504	-1.336
Skewness	0.005	0.243	0.100	-0.946	0.288	0.678	0.546	-0.149	-0.305
Range	1.806	0.075	13,218.000	453.150	0.065	0.113	0.116	2,913.600	4,049.500
Minimum	3.074	0.136	16,123.000	894.850	0.092	0.051	0.050	1,587.400	2,675.500
Maximum	4.880	0.211	29,341.000	1,348.000	0.157	0.164	0.166	4,501.000	6,725.000
Sum	107.157	4.606	469,161.000	27,501.160	2.528	2.169	2.210	72,142.600	110,390.500
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	3.591	0.155	19,061.000	1,141.965	0.100	0.075	0.075	2,220.000	3,699.500
Q3	4.364	0.185	24,780.000	1,272.875	0.137	0.116	0.118	4,092.800	5,865.300

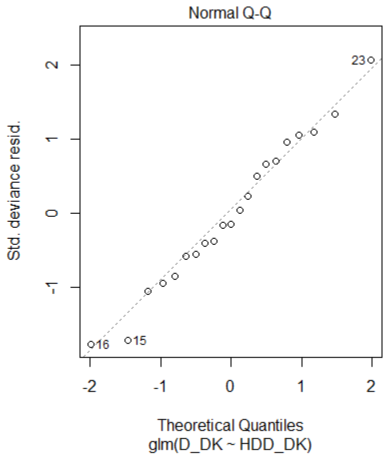
A9: Summary statistics for all variables in United Kingdom

	UK_PS_R	*UK_AP*	*UK_D*	*UK_HDD*	*UK_PC_P*	*UK_PC_TU*	*UK_PC_FU*	*UK_Egf_ME*	*UK_H_ME*
Mean	4.019	0.163	186,427.238	3,000.430	0.105	0.061	0.055	29,955.830	18,073.417
Standard Error	0.036	0.001	4,701.293	35.223	0.002	0.003	0.002	1,706.220	1,081.534
Median	4.094	0.159	177,095.000	2,965.940	0.102	0.056	0.051	32,925.900	17,645.800
Mode	#N/A	#N/A	#N/A	#N/A	0.097	0.079	0.043	#N/A	#N/A
Standard Deviation	0.188	0.007	21,544.030	168.923	0.007	0.013	0.011	8,182.745	5,186.856
Sample Variance	0.035	0.000	464,145,246.790	28,535.091	0.000	0.000	0.000	66,957,322.553	26,903,475.692
Kurtosis	1.288	1.012	-0.834	1.607	-1.558	-1.300	-1.046	-1.548	-0.122
Skewness	-1.630	1.566	0.857	1.227	0.441	0.523	0.585	0.084	0.141
Range	0.629	0.023	65,198.000	712.560	0.019	0.041	0.032	25,226.300	19,250.400
Minimum	3.547	0.158	165,875.000	2,739.930	0.097	0.044	0.042	18,225.400	8,525.700
Maximum	4.176	0.181	231,073.000	3,452.490	0.116	0.085	0.074	43,451.700	27,776.100
Sum	108.506	4.394	3,914,972.000	69,009.880	2.196	1.410	1.260	688,984.100	415,688.600
Count	27.000	27.000	21.000	23.000	21.000	23.000	23.000	23.000	23.000
Q1	4.039	0.158	169,863.000	2,909.880	0.098	0.051	0.046	22,382.500	15,770.600
Q3	4.125	0.164	211,172.000	3,048.475	0.113	0.076	0.064	35,888.800	20,628.900

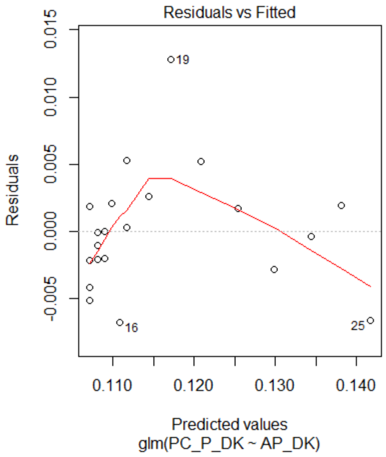
A10: Residual vs Fitted plot for D - HDD DK



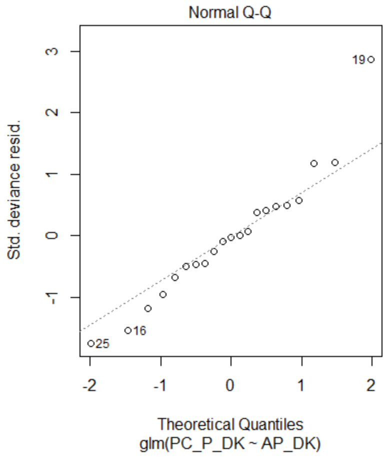
A11: Fitted line of normalized residuals for D - HDD DK



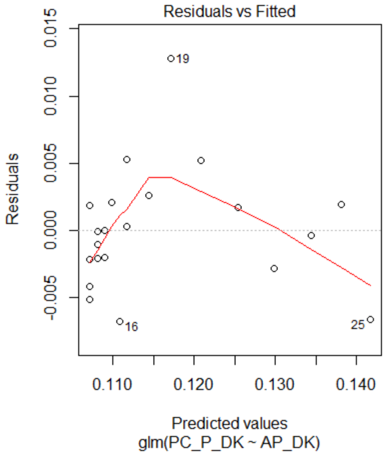
A12: Residual vs Fitted plot for AP - PC_P DK



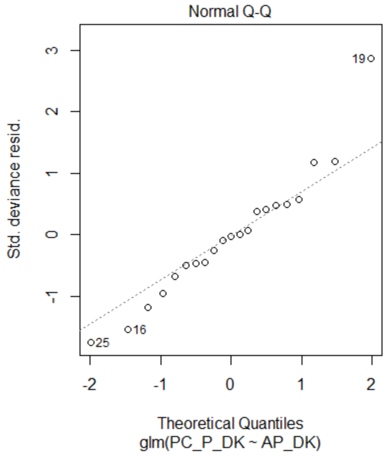
A13: Fitted line of normalized residuals for AP - PC_P DK



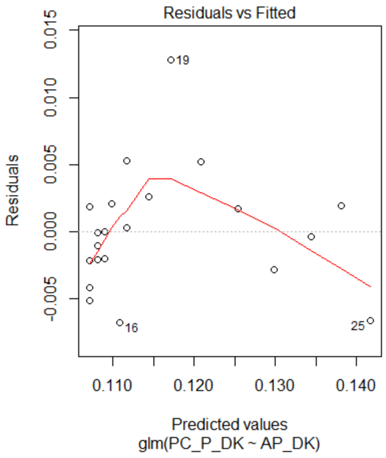
A14: Residual vs Fitted plot for AP - PC_P DK



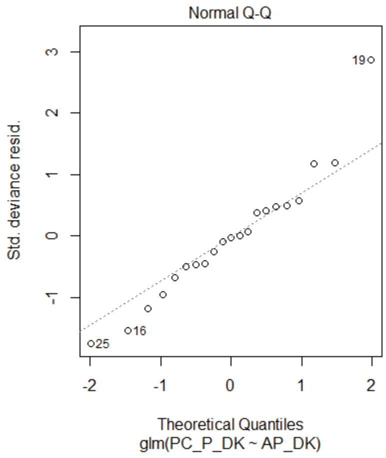
A15: Fitted line of normalized residuals for AP - PC_P DK



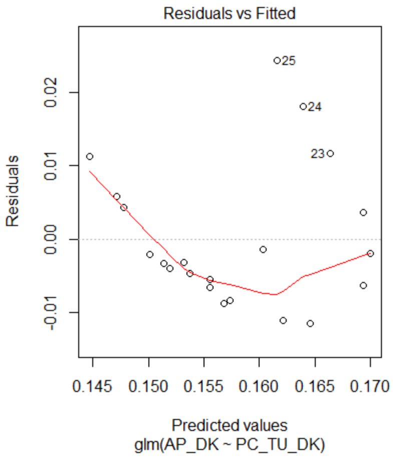
A16: Residual vs Fitted plot for AP - PC_P DK



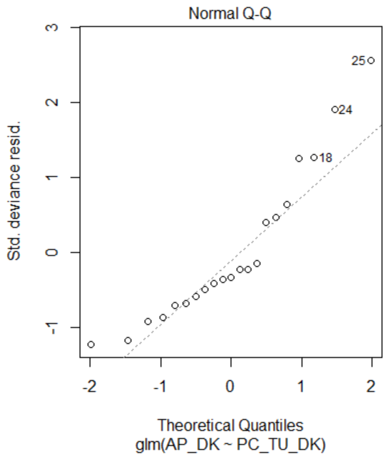
A17: Fitted line of normalized residuals for AP - PC_P DK



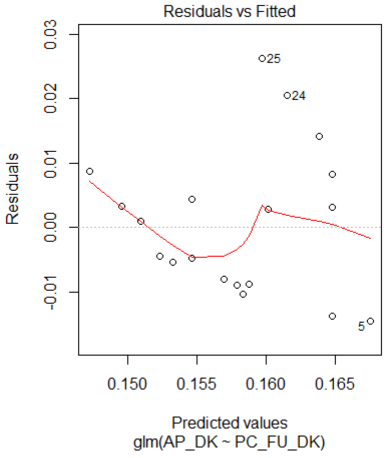
A18: Residual vs Fitted plot for AP - PC_TU DK



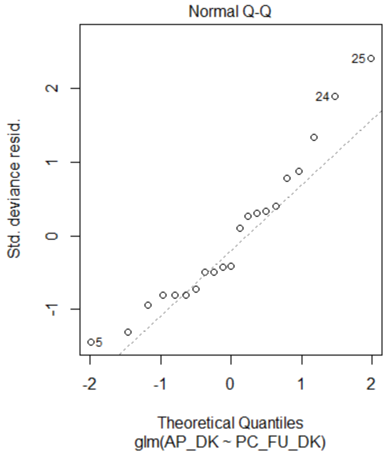
A19: Fitted line of normalized residuals for AP - PC_TU DK



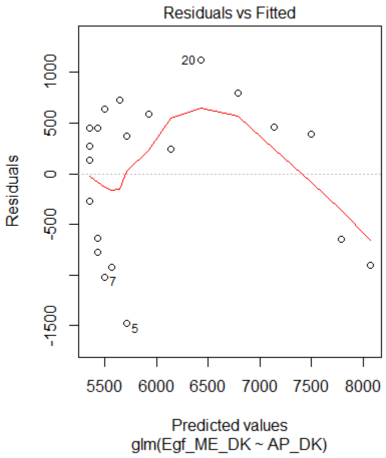
A20: Residual vs Fitted plot for AP - PC_FU DK



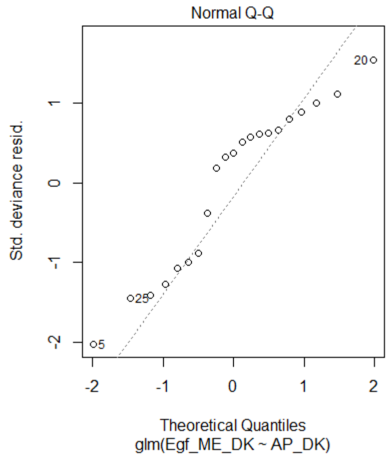
A21: Fitted line of normalized residuals for AP - PC_FU DK



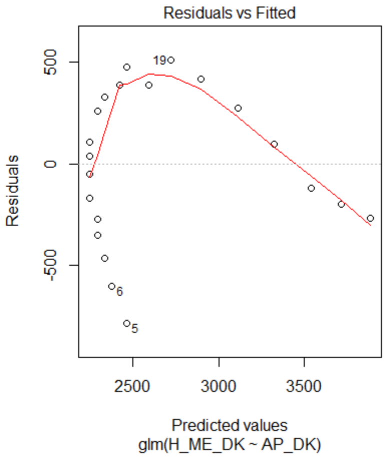
A22: Residual vs Fitted plot for AP – Egf_ME DK



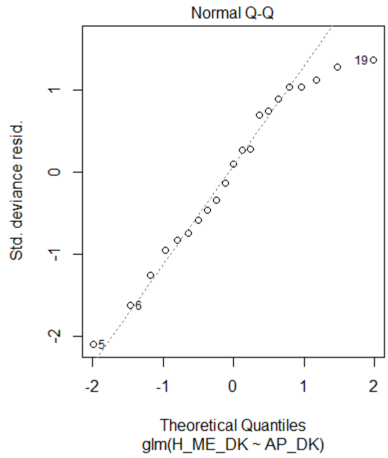
A23: Fitted line of normalized residuals for AP – Egf_ME DK



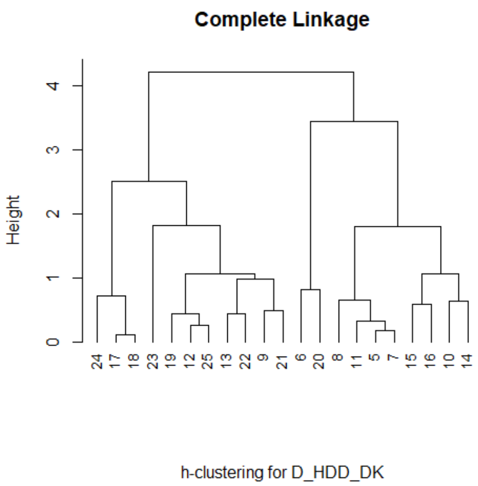
A24: Residual vs Fitted plot for AP – H_ME DK



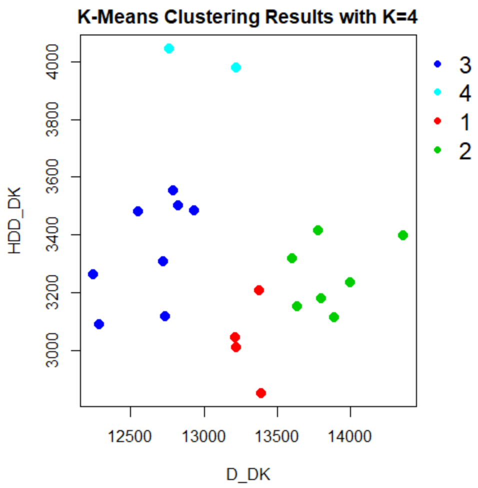
A25: Fitted line of normalized residuals for AP – H_ME DK



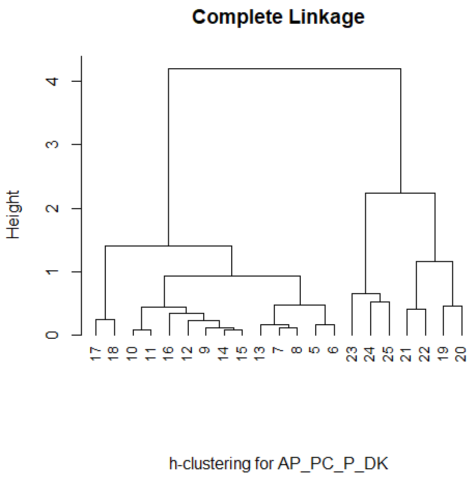
A26: Hierarchical clustering dendrogram for D – HDDD DK



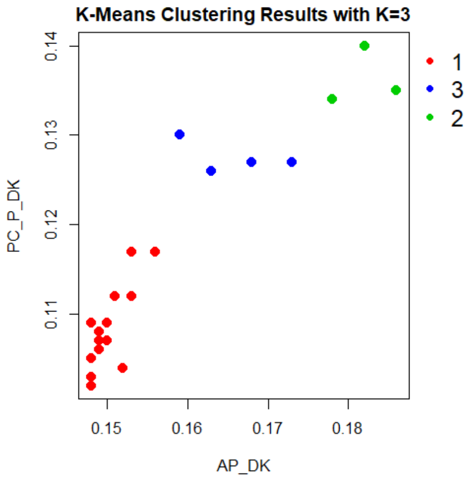
A27: K-means cluster separation in scatter plot for D - HDD DK



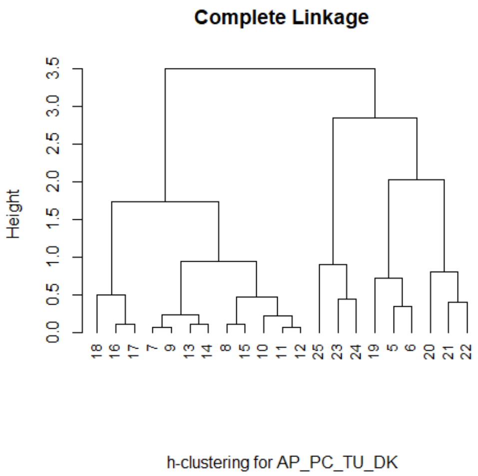
A28: Hierarchical clustering dendrogram for AP – PC_P DK



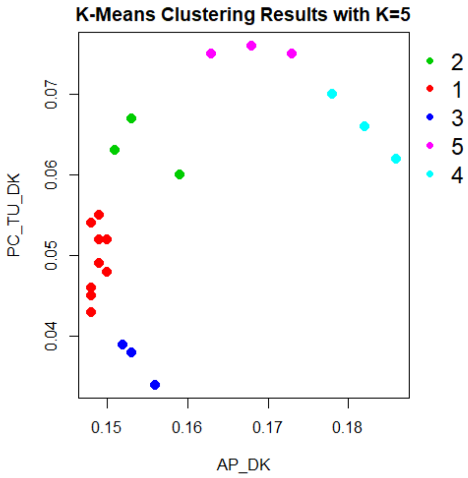
A29: K-means cluster separation in scatter plot for AP – PC_P DK



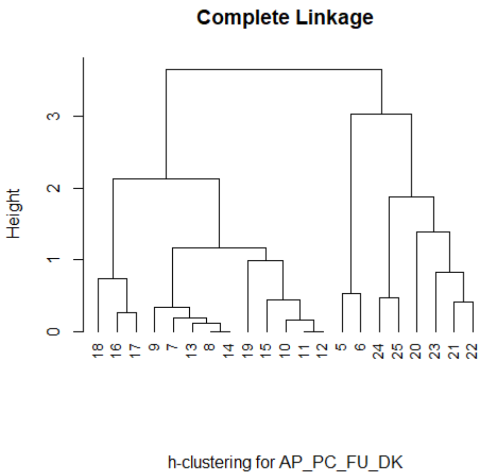
A30: Hierarchical clustering dendrogram for AP – PC_TU DK



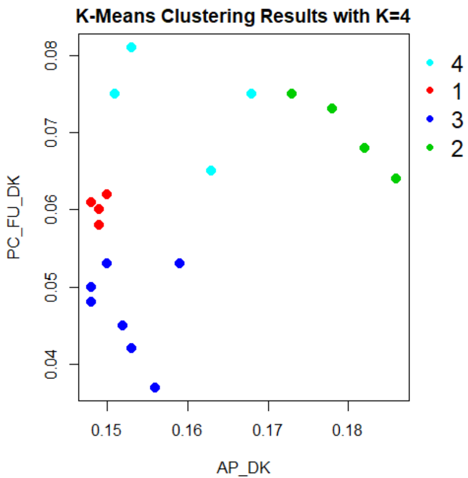
A31: K-means cluster separation in scatter plot for AP – PC_TU DK



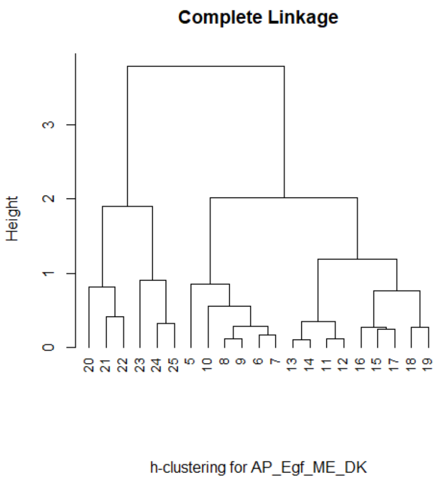
A32: Hierarchical clustering dendrogram for AP – PC_FU DK



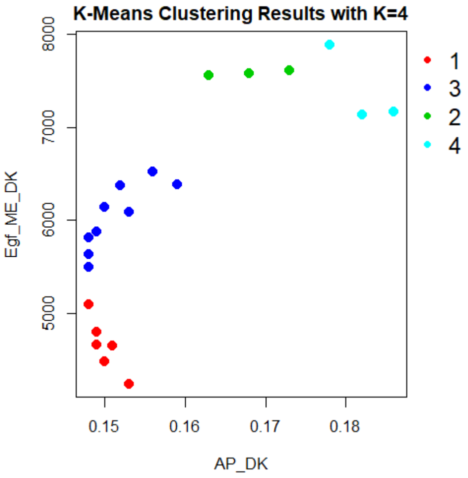
A33: K-means cluster separation in scatter plot for AP – PC_FU DK



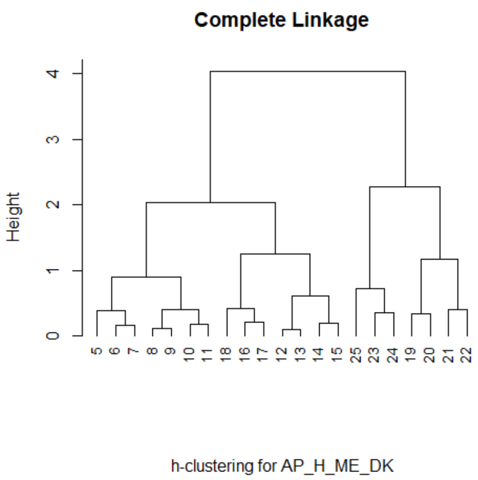
A34: Hierarchical clustering dendrogram for AP – Egf_ME DK



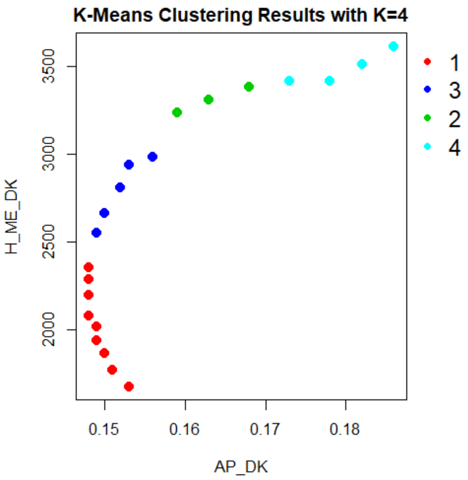
A35: K-means cluster separation in scatter plot for AP – Egf_ME DK



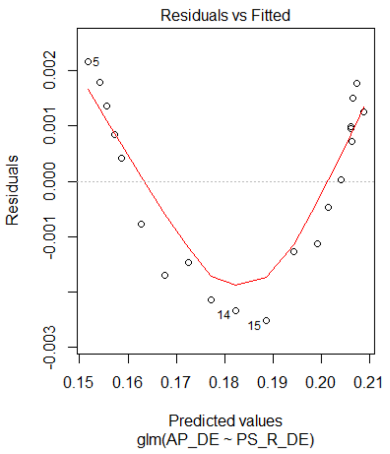
A36: Hierarchical clustering dendrogram for AP – H_ME DK



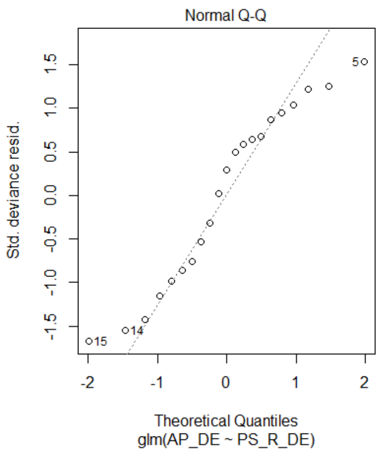
A37: K-means cluster separation in scatter plot for AP – H_ME DK



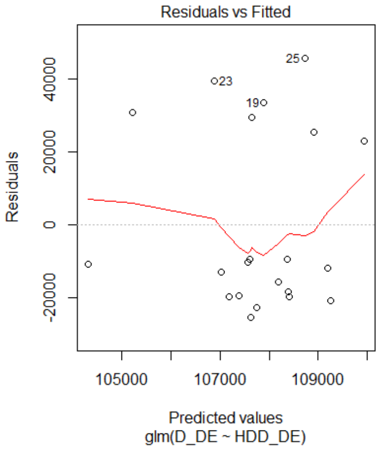
A38: Residual vs Fitted plot for AP – PS_R DE



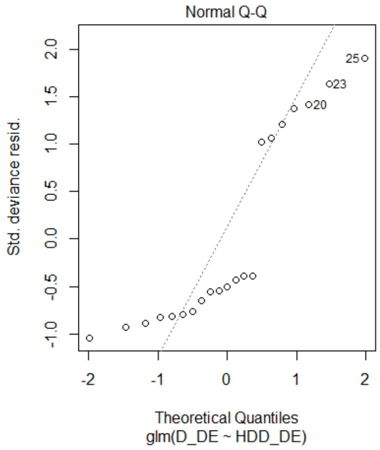
A39: Fitted line of normalized residuals for AP – PS_R DE



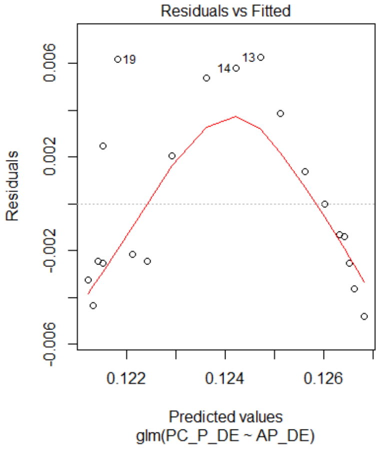
A40: Residual vs Fitted plot for D - HDD DE



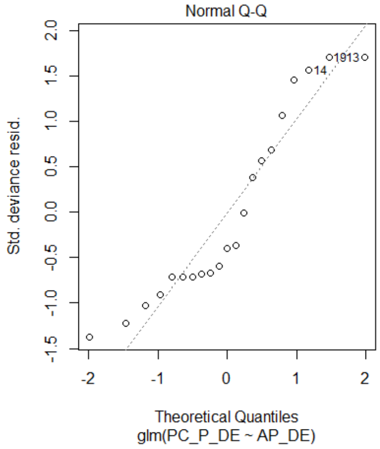
A41: Fitted line of normalized residuals for D - HDD DE



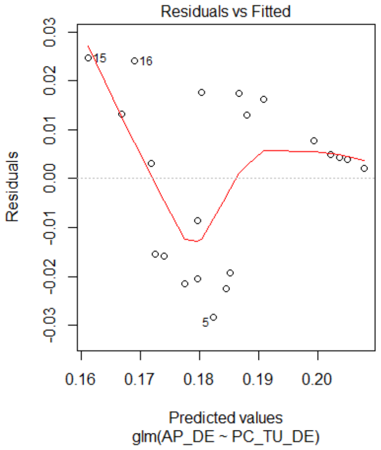
A42: Residual vs Fitted plot for AP – PC_P DE



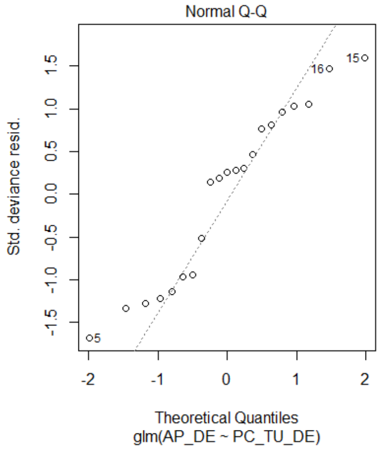
A43: Fitted line of normalized residuals for AP – PC_P DE



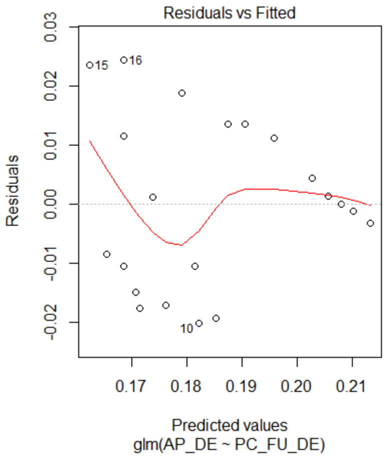
A44: Residual vs Fitted plot for AP – PC_TU DE



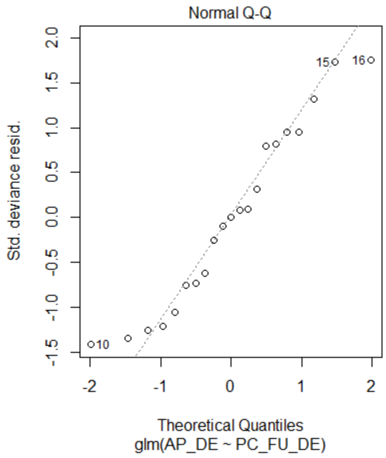
A45: Fitted line of normalized residuals for AP – PC_TU DE



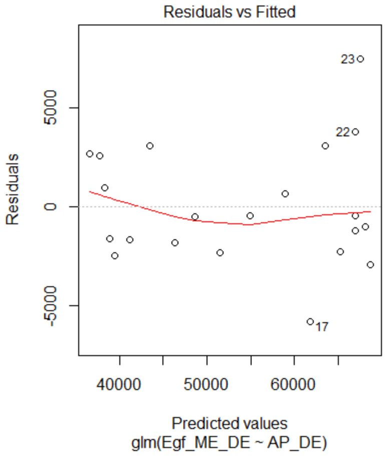
A46: Residual vs Fitted plot for AP – PC_FU DE



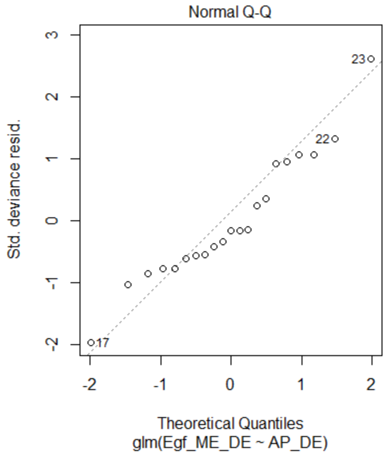
A47: Fitted line of normalized residuals for AP – PC_FU DE



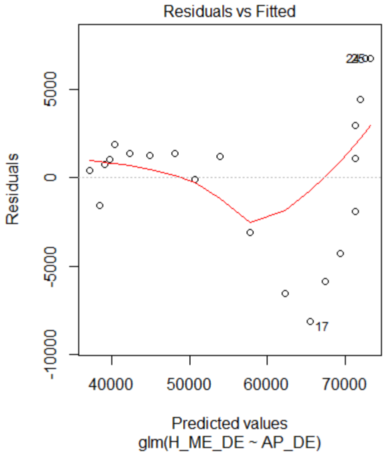
A48: Residual vs Fitted plot for AP – Egf_ME DE



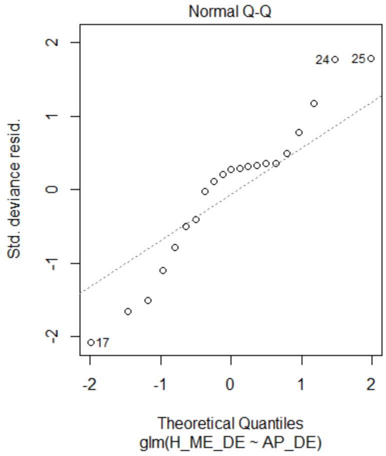
A49: Fitted line of normalized residuals for AP – Egf_ME DE



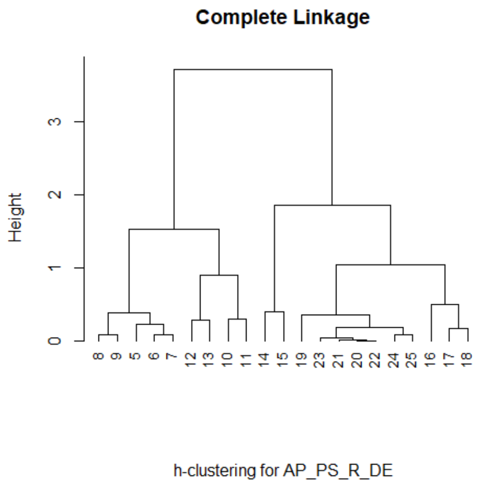
A50: Residual vs Fitted plot for AP – H_ME DE



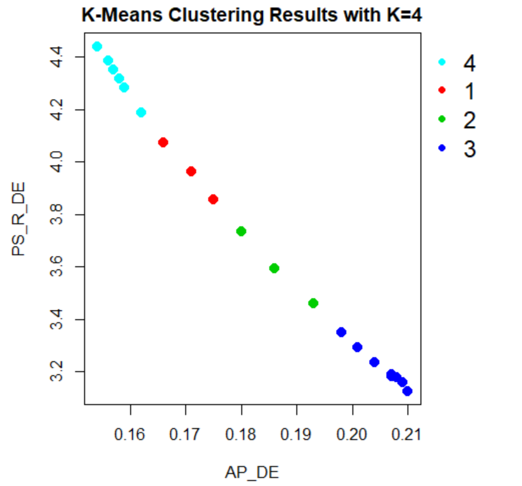
A51: Fitted line of normalized residuals for AP – H_ME DE



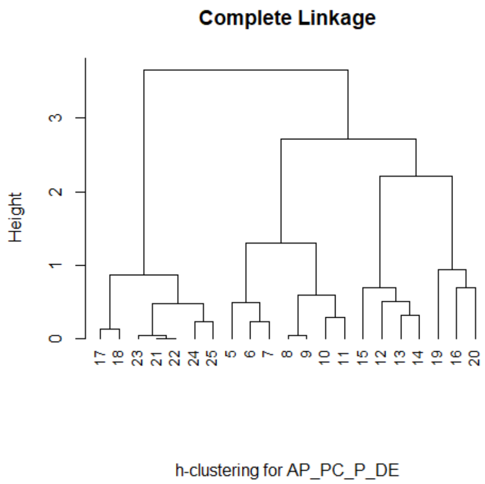
A52: Hierarchical clustering dendrogram for AP – PS_R DE



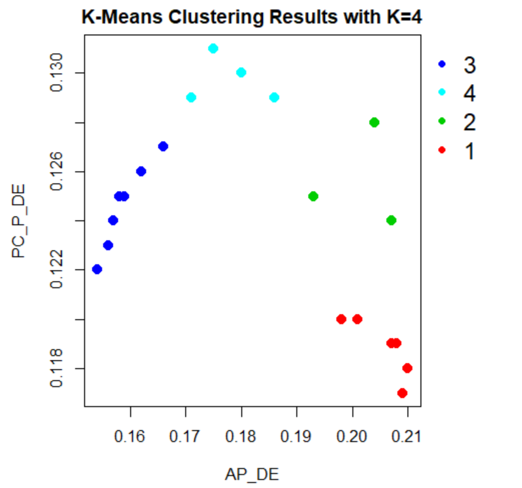
A53: K-means cluster separation in scatter plot for AP – PS_R DE



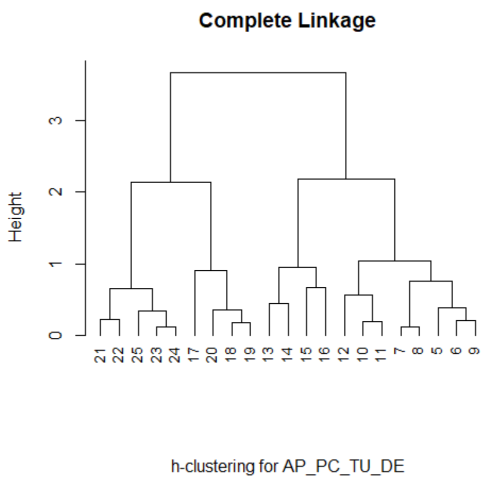
A54: Hierarchical clustering dendrogram for AP – PC_P DE



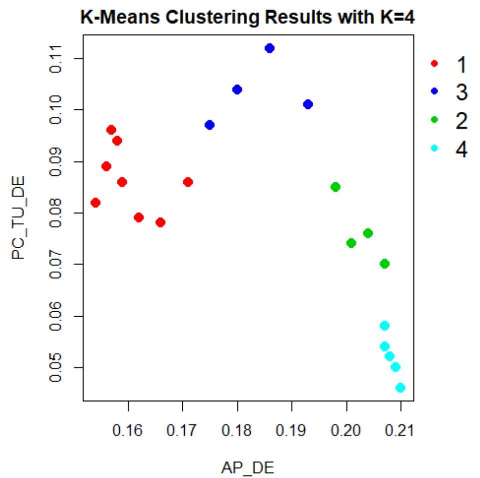
A55: K-means cluster separation in scatter plot for AP – PC_P DE



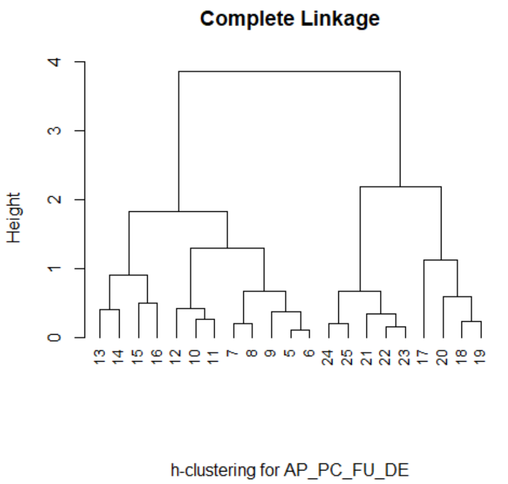
A56: Hierarchical clustering dendrogram for AP – PC_TU DE



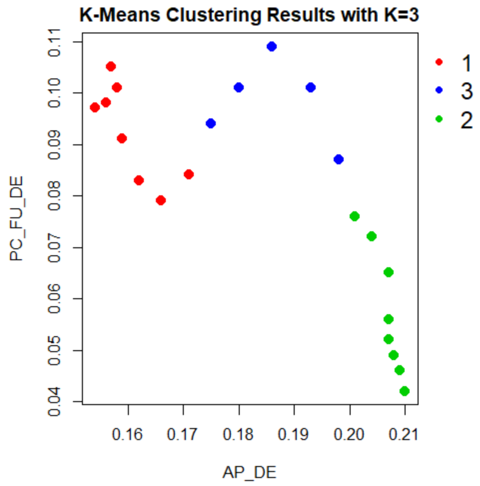
A57: K-means cluster separation in scatter plot for AP – PC_TU DE



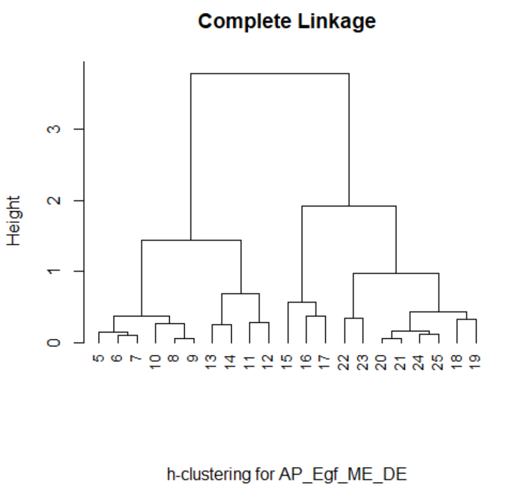
A58: Hierarchical clustering dendrogram for AP – PC_FU DE



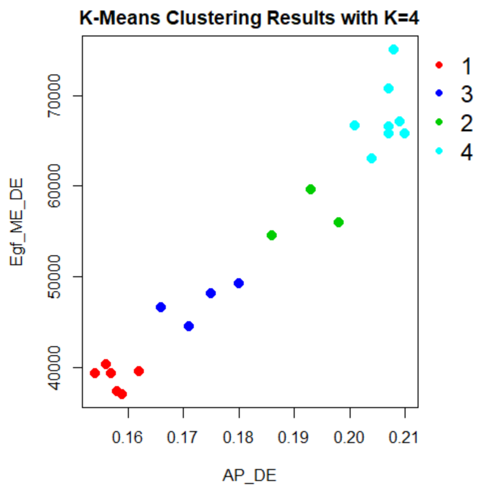
A59: K-means cluster separation in scatter plot for AP – PC_FU DE



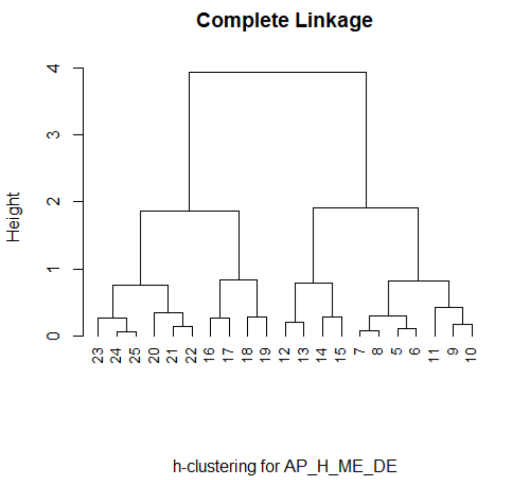
A60: Hierarchical clustering dendrogram for AP – Egf_ME DE



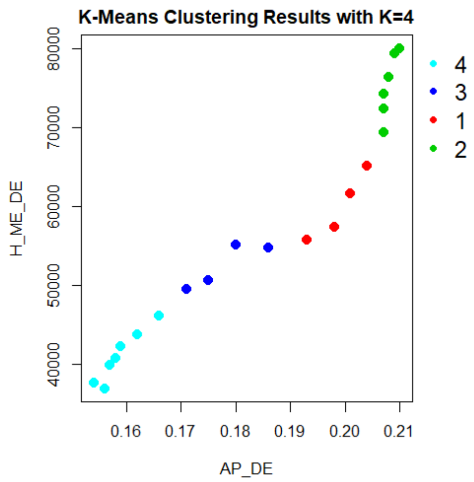
A61: K-means cluster separation in scatter plot for AP – Egf_ME DE



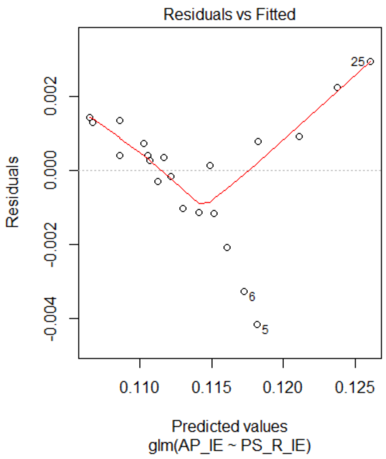
A62: Hierarchical clustering dendrogram for AP – H_ME DE



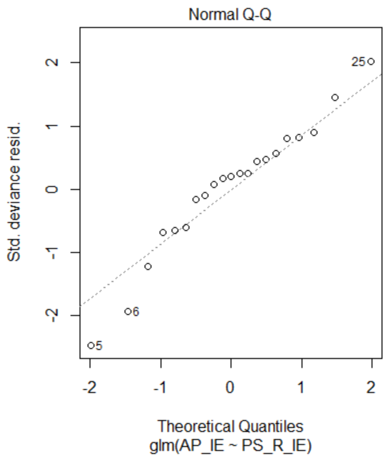
A63: K-means cluster separation in scatter plot for AP – H_ME DE



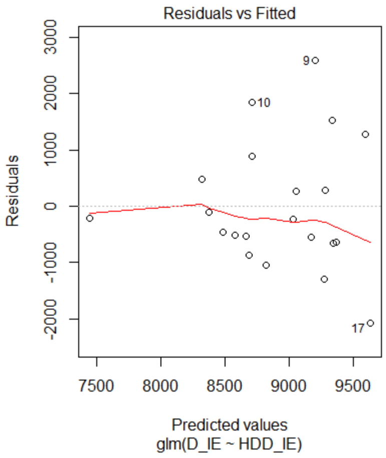
A64: Residual vs Fitted plot for AP – PS_R IE



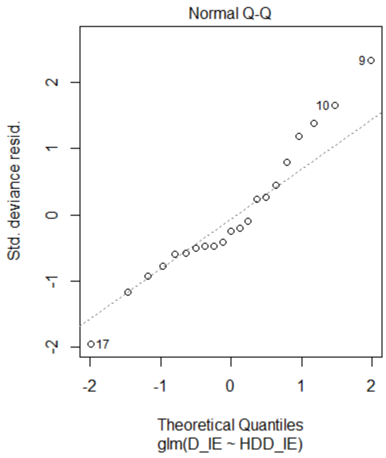
A65: Fitted line of normalized residuals for AP – PS_R IE



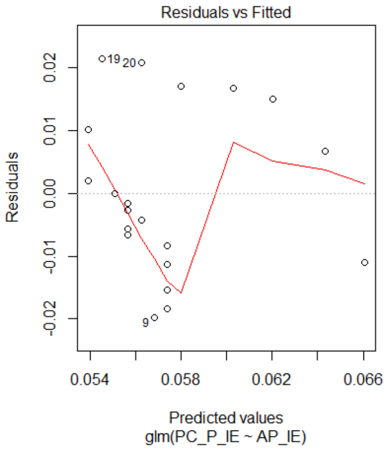
A66: Residual vs Fitted plot for D - HDD IE



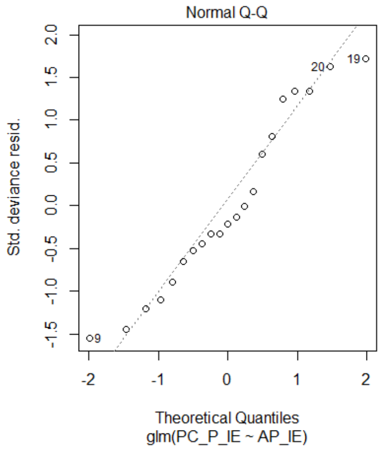
A67: Fitted line of normalized residuals for D - HDD IE



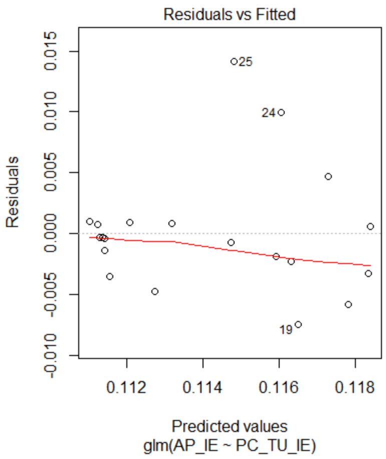
A68: Residual vs Fitted plot for AP – PC_P IE



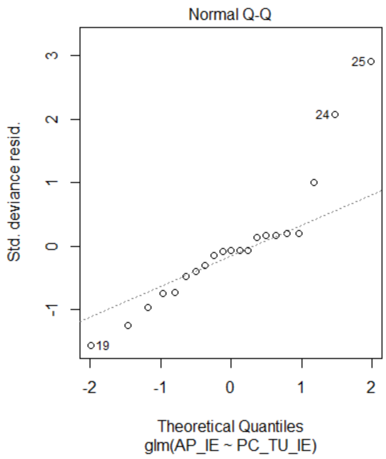
A69: Fitted line of normalized residuals for AP – PC_P IE



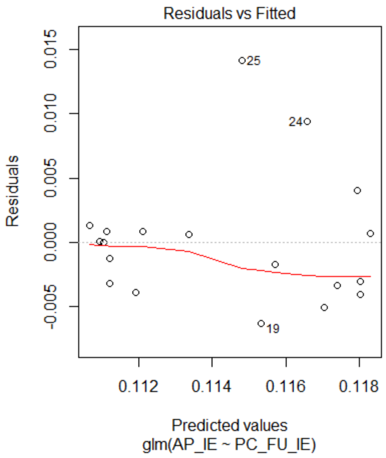
A70: Residual vs Fitted plot for
AP – PC_TU IE



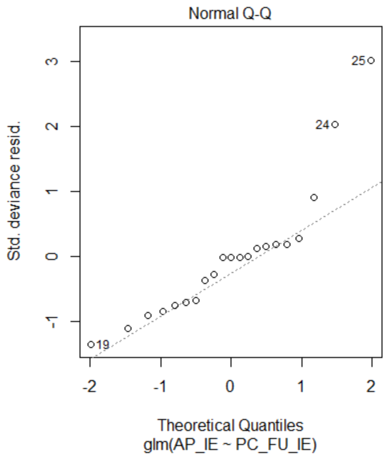
A71: Fitted line of normalized residuals
for AP – PC_TU IE



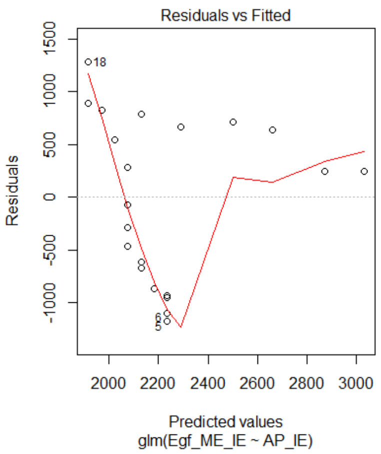
A72: Residual vs Fitted plot for
AP – PC_FU IE



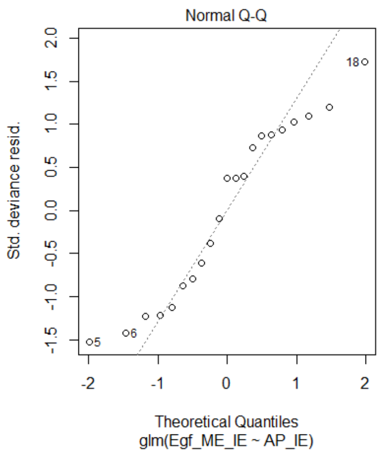
A73: Fitted line of normalized residuals
for AP – PC_FU IE



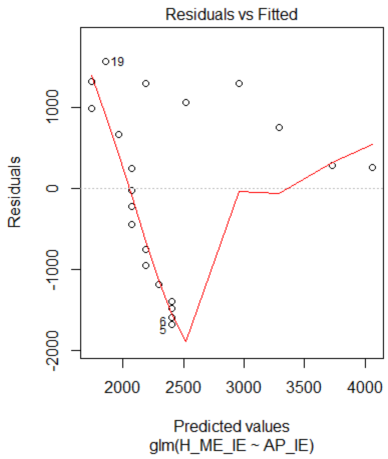
A74: Residual vs Fitted plot for
AP – Egf_ME IE



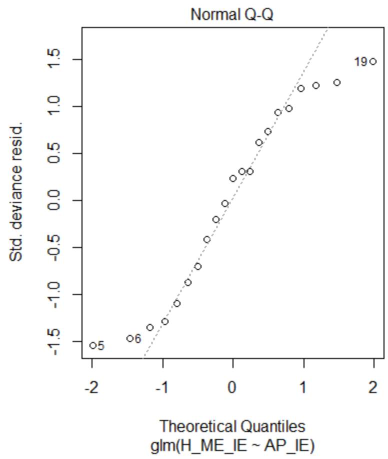
A75: Fitted line of normalized residuals
for AP – Egf_ME IE



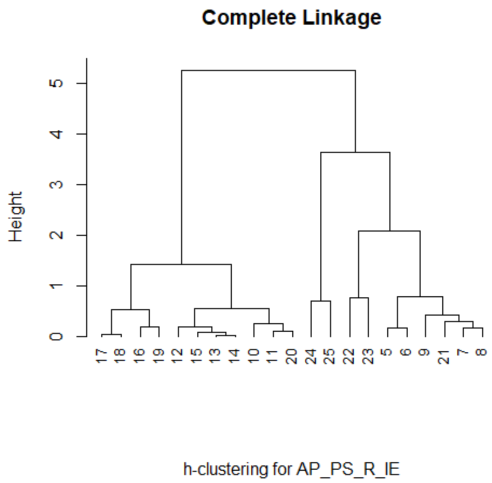
A76: Residual vs Fitted plot for AP – H IE



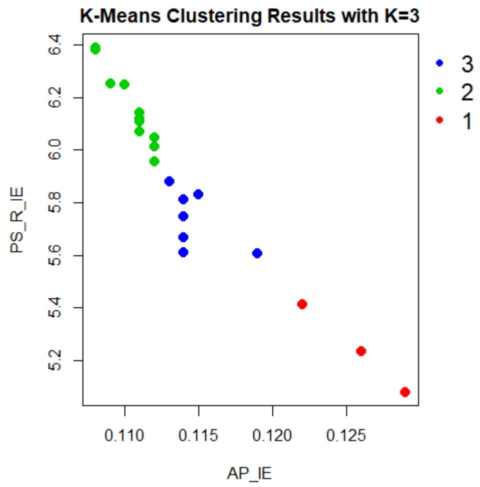
A77: Fitted line of normalized residuals for AP – H IE



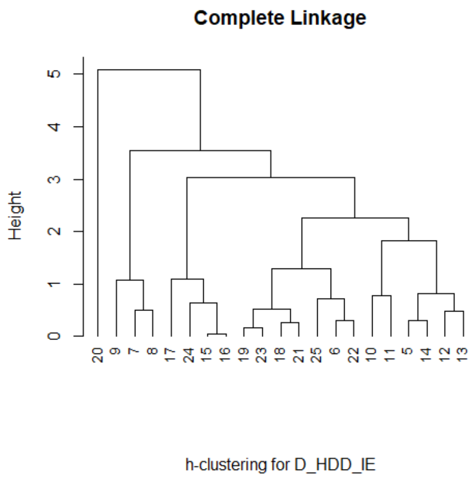
A78: Hierarchical clustering dendrogram for AP – PS_R IE



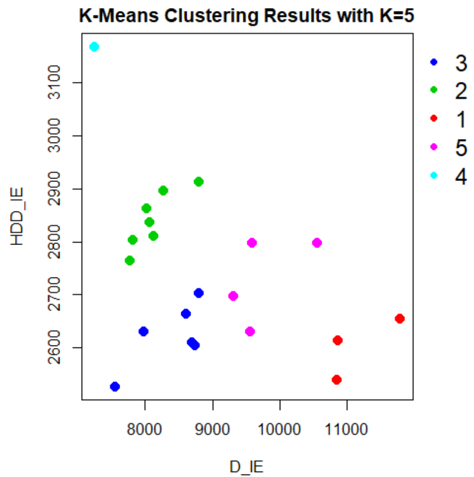
A79: K-means cluster separation in scatter plot for AP – PS_R IE



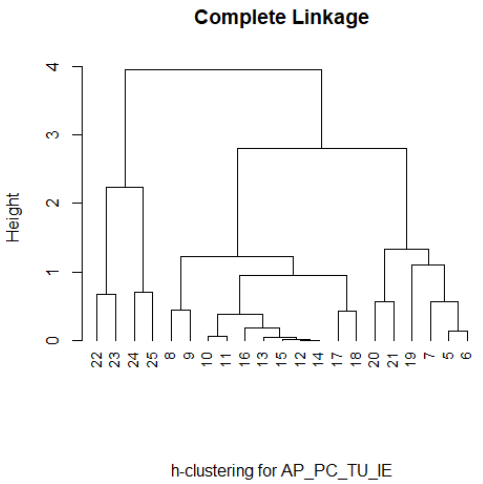
A80: Hierarchical clustering dendrogram for D - HDD IE



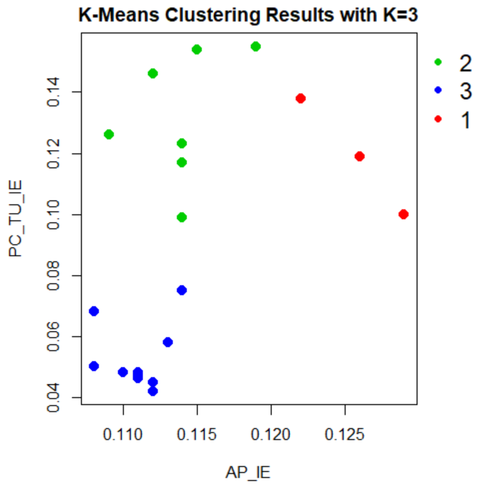
A81: K-means cluster separation in scatter plot for D - HDD IE



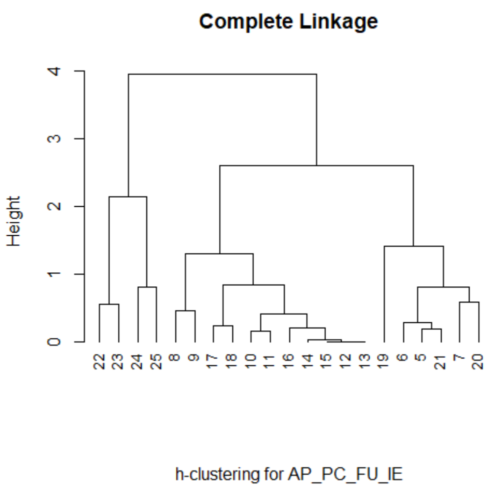
A82: Hierarchical clustering dendrogram for AP – PC_TU IE



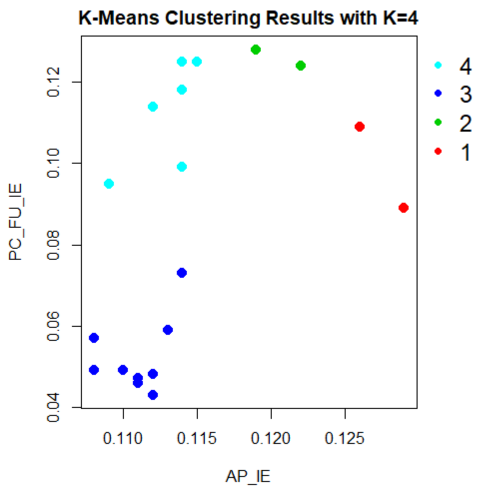
A83: K-means cluster separation in scatter plot for AP – PC_TU IE



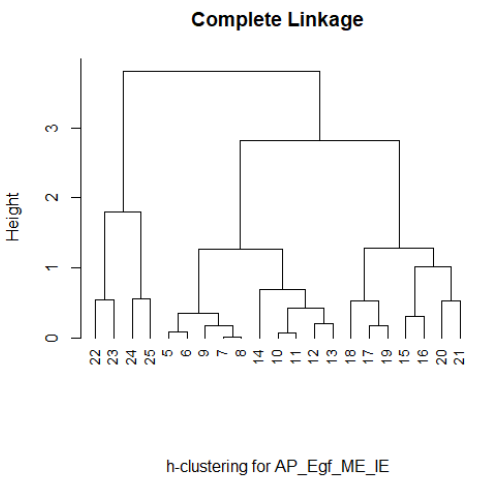
A84: Hierarchical clustering dendrogram for AP – PC_FU IE



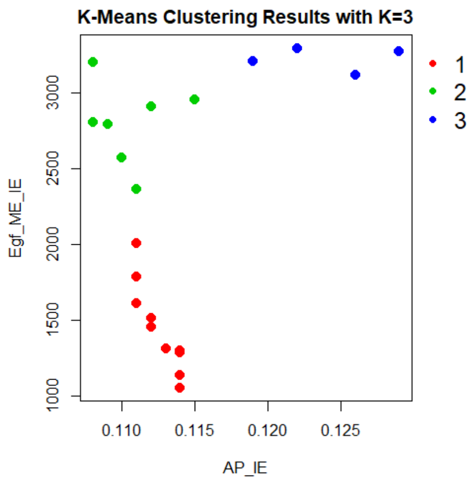
A85: K-means cluster separation in scatter plot for AP – PC_FU IE



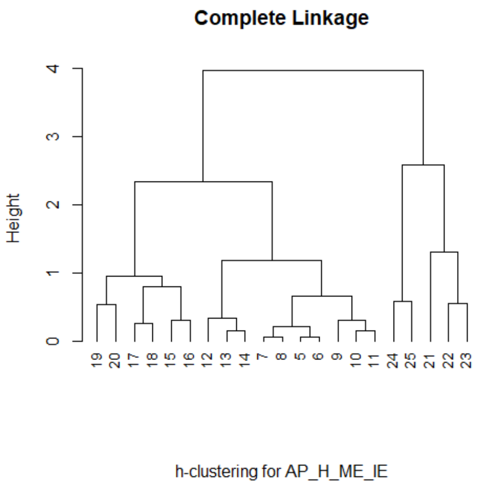
A86: Hierarchical clustering dendrogram for AP – Egf_ME IE



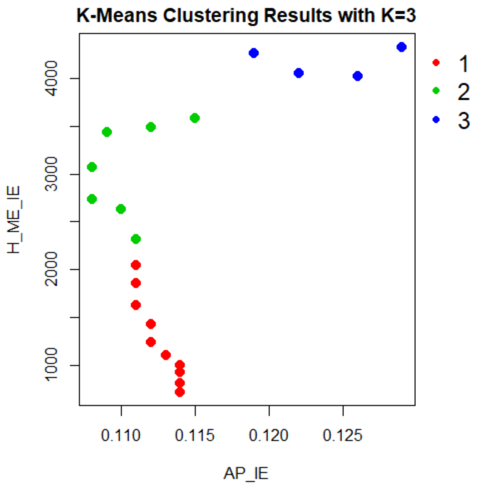
A87: K-means cluster separation in scatter plot for AP – Egf_ME IE



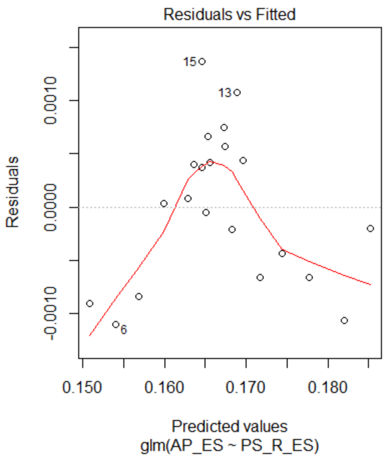
A88: Hierarchical clustering dendrogram for AP – H_ME IE



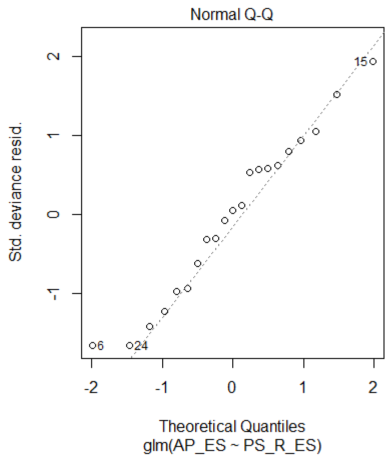
A89: K-means cluster separation in scatter plot for AP – H_ME IE



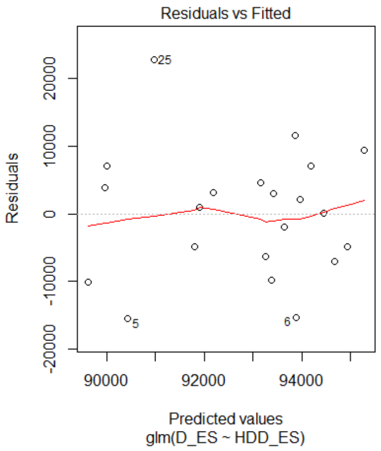
A90: Residual vs Fitted plot for AP – PS_R ES



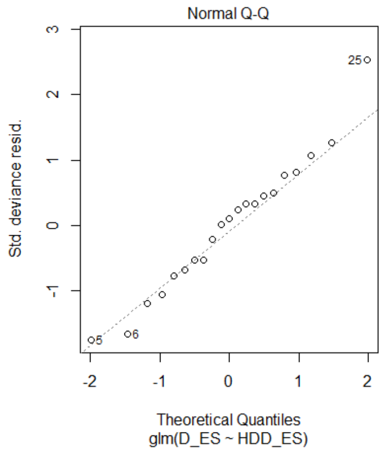
A91: Fitted line of normalized residuals for AP – PS_R ES



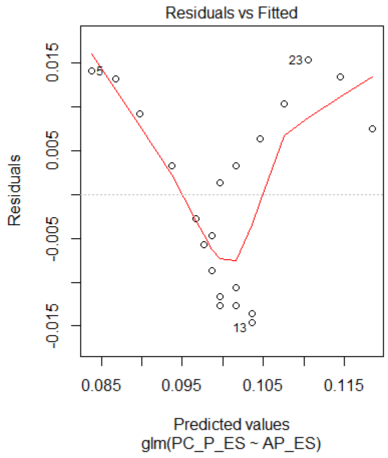
A92: Residual vs Fitted plot for D - HDD ES



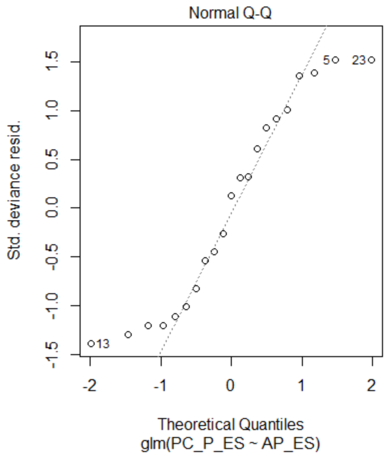
A93: Fitted line of normalized residuals for D - HDD ES



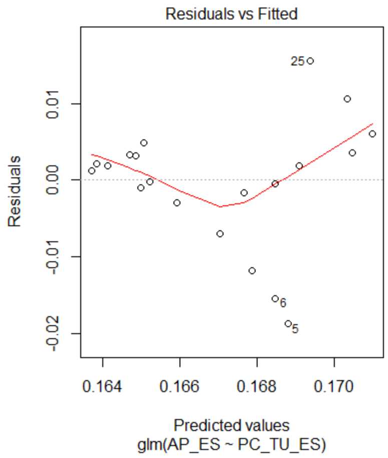
A94: Residual vs Fitted plot for AP - PC_P ES



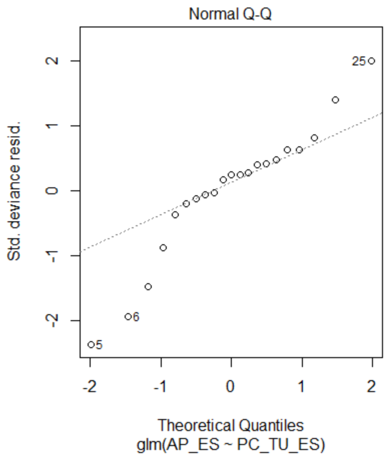
A95: Fitted line of normalized residuals for AP - PC_P ES



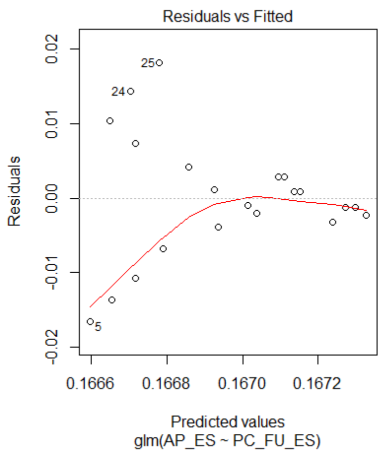
A96: Residual vs Fitted plot for AP - PC_TU ES



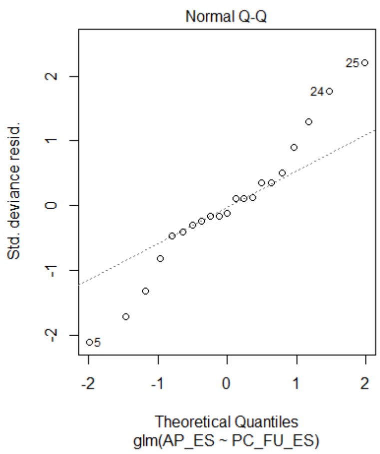
A97: Fitted line of normalized residuals for AP - PC_TU ES



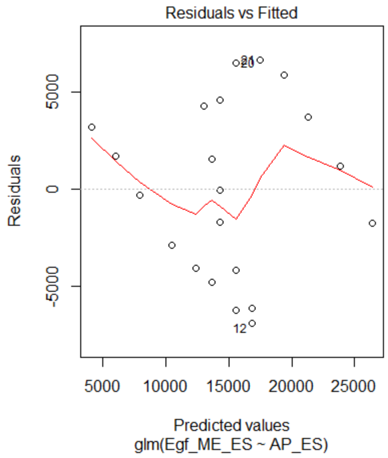
A98: Residual vs Fitted plot for AP - PC_FU ES



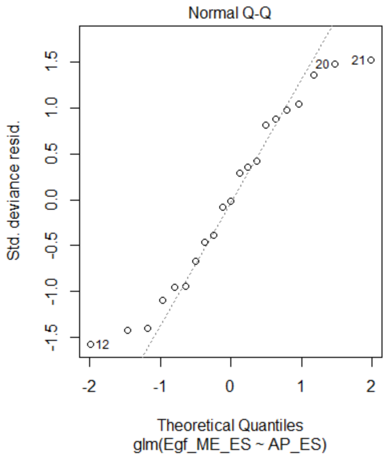
A99: Fitted line of normalized residuals for AP - PC_FU ES



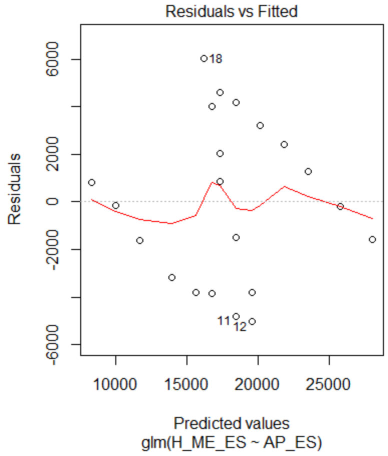
A100: Residual vs Fitted plot for AP – Egf_ME ES



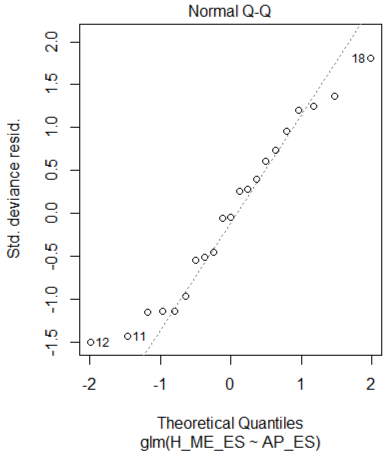
A101: Fitted line of normalized residuals for AP – Egf_ME ES



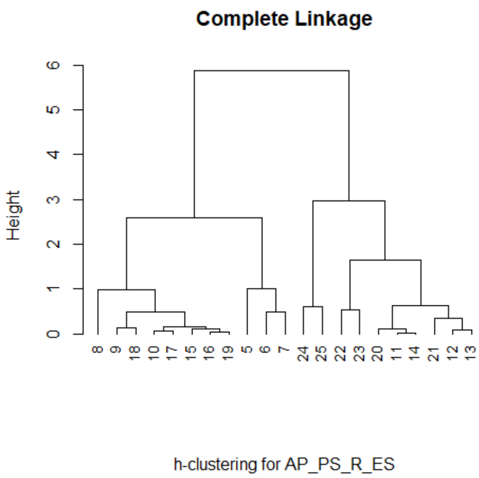
A102: Residual vs Fitted plot for AP – H_ME ES



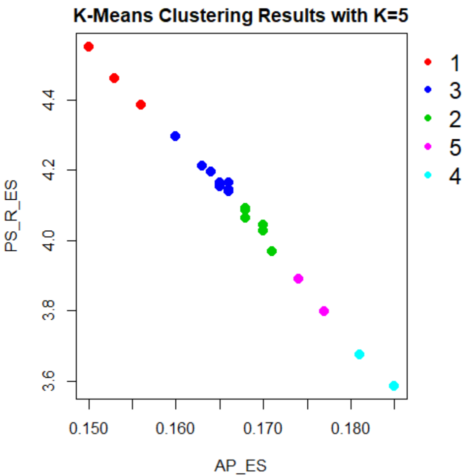
A103: Fitted line of normalized residuals for AP – H_ME ES



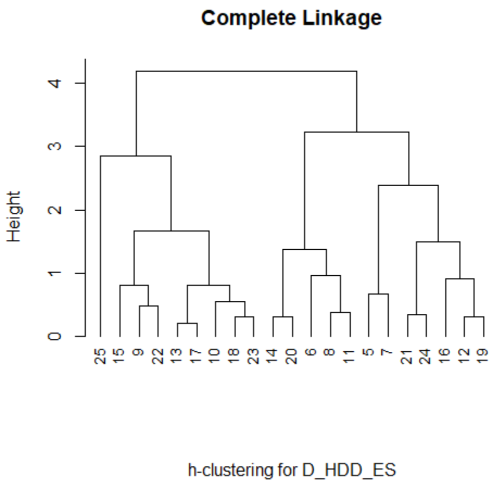
A104: Hierarchical clustering dendrogram for AP – PS_R ES



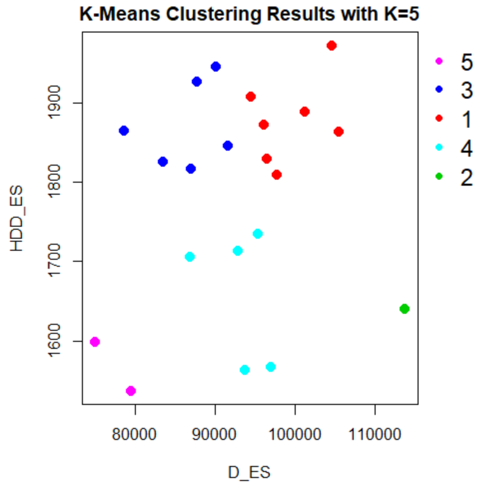
A105: K-means cluster separation in scatter plot for AP – PS_R ES



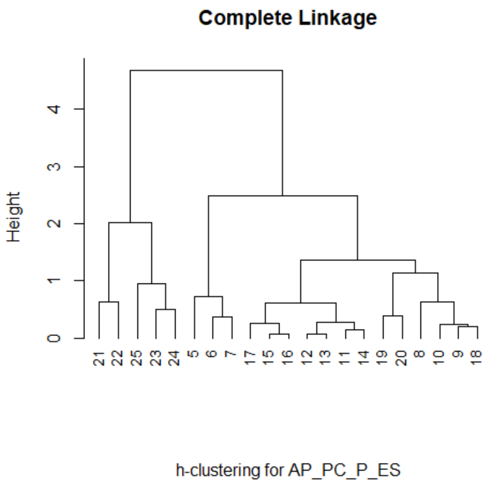
A106: Hierarchical clustering dendrogram for D - HDD ES



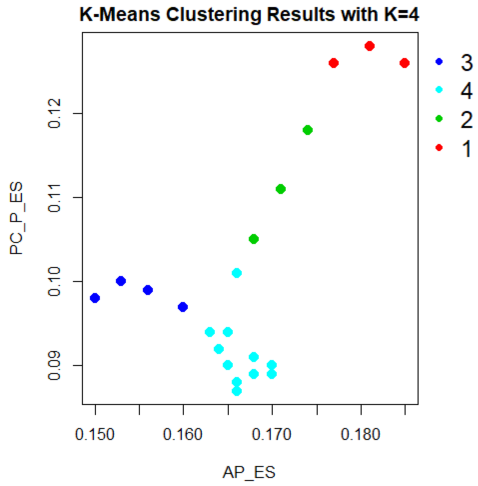
A107: K-means cluster separation in scatter plot for D - HDD ES



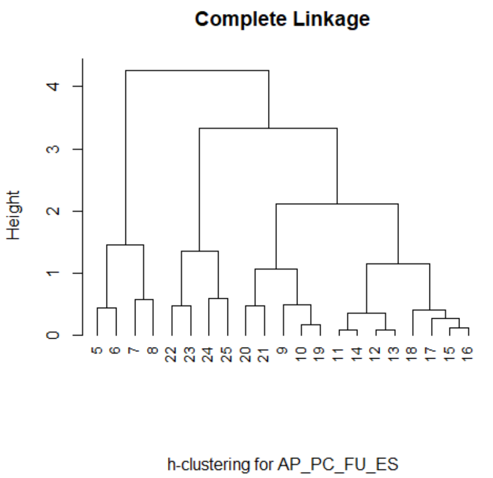
A108: Hierarchical clustering dendrogram for AP – PC_P ES



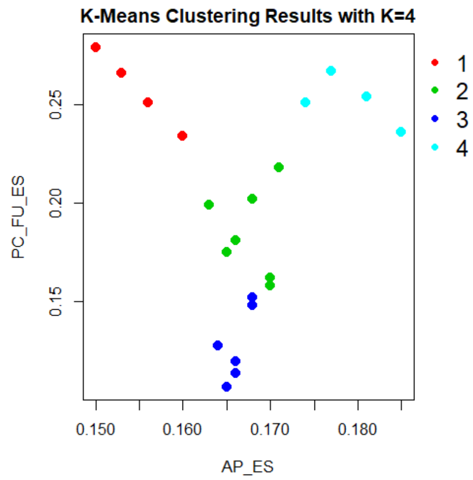
A109: K-means cluster separation in scatter plot for AP – PC_P ES



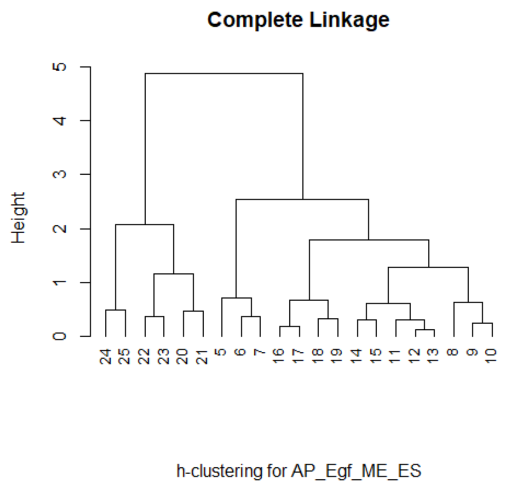
A110: Hierarchical clustering dendrogram for AP – PC_FU ES



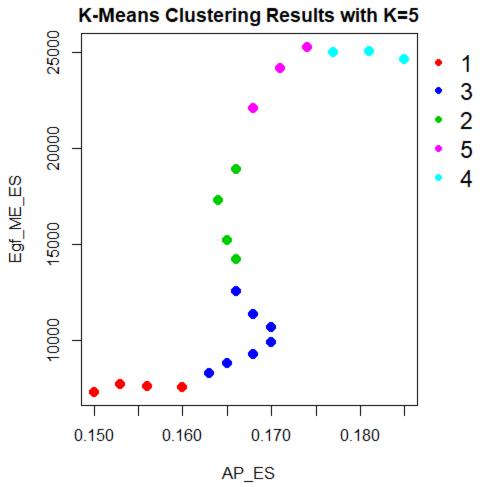
A111: K-means cluster separation in scatter plot for AP – PC_FU ES



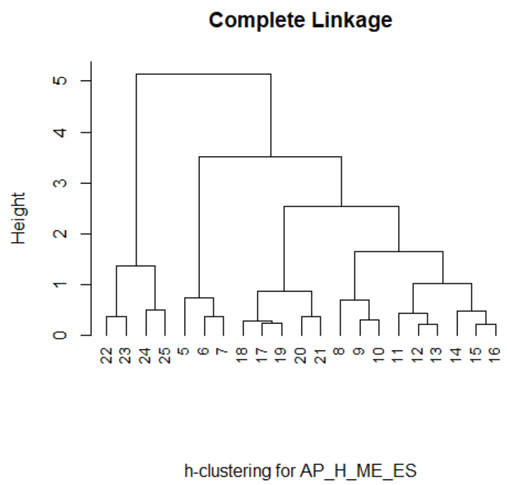
A112: Hierarchical clustering dendrogram for AP – Egf_ME ES



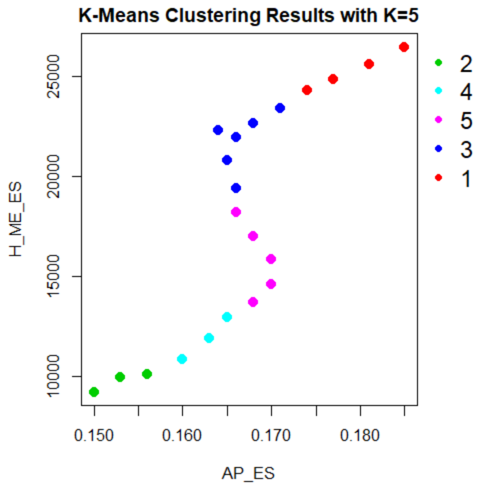
A113: K-means cluster separation in scatter plot for AP – Egf_ME ES



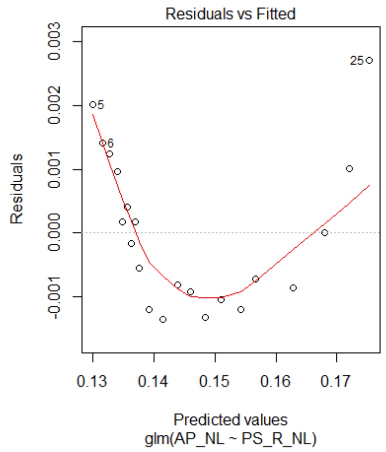
A114: Hierarchical clustering dendrogram for AP – H_ME ES



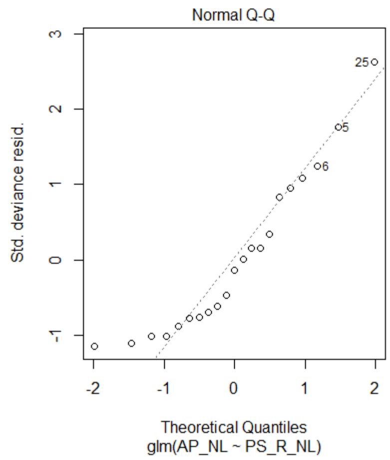
A115: K-means cluster separation in scatter plot for AP – H_ME ES



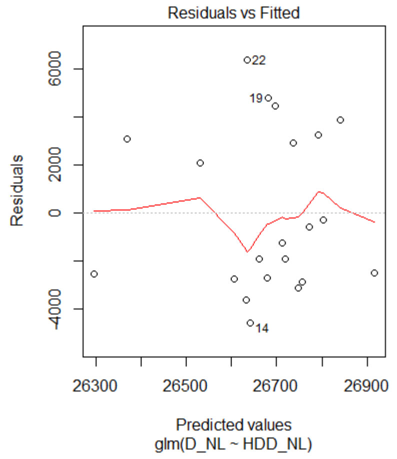
A116: Residual vs Fitted plot for AP – PS_R NL



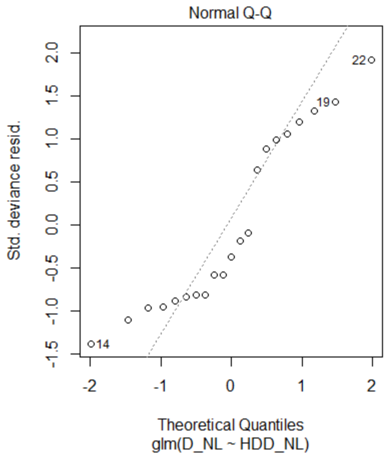
A117: Fitted line of normalized residuals for AP – PS_R NL



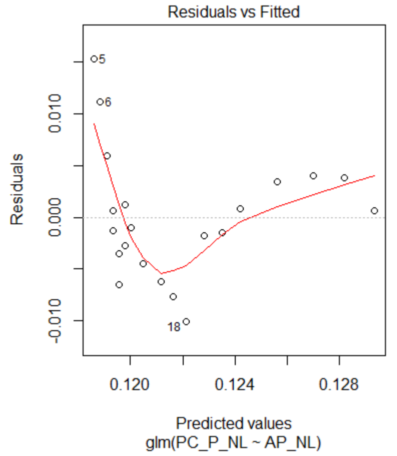
A118: Residual vs Fitted plot for D - HDD NL



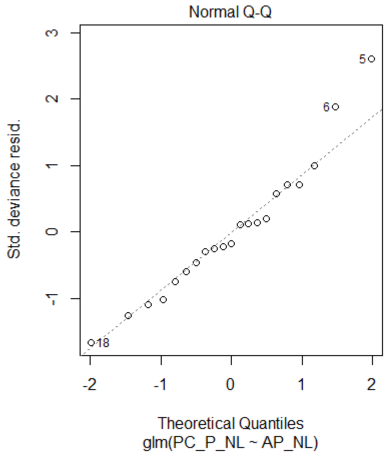
A119: Fitted line of normalized residuals for D - HDD NL



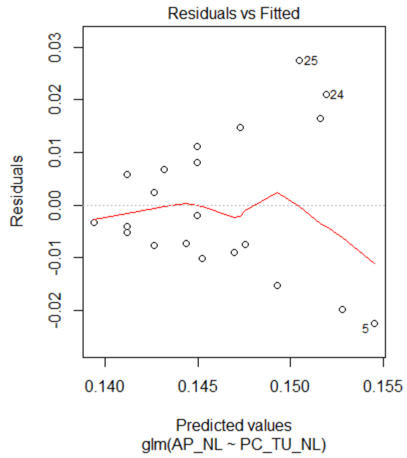
A120: Residual vs Fitted plot for AP – PC_P NL



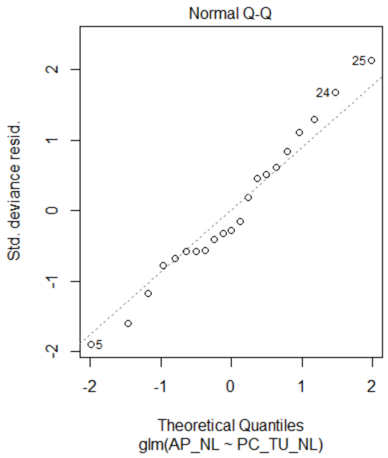
A121: Fitted line of normalized residuals for AP – PC_P NL



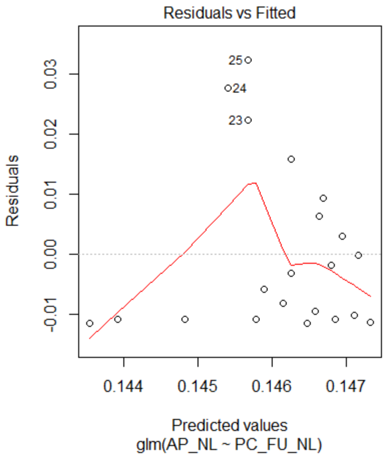
A122: Residual vs Fitted plot for AP – PC_TU NL



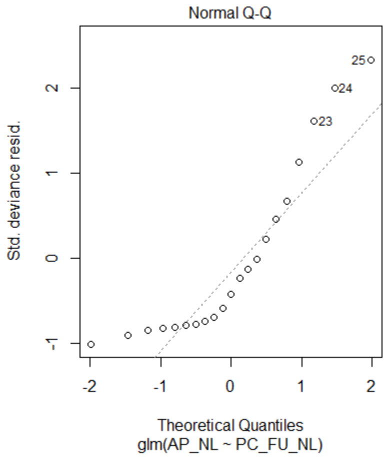
A123: Fitted line of normalized residuals for AP – PC_TU NL



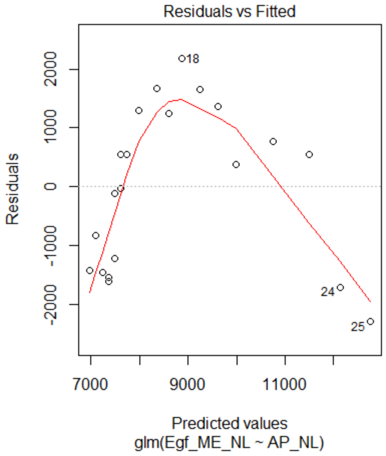
A124: Residual vs Fitted plot for AP – PC_FU NL



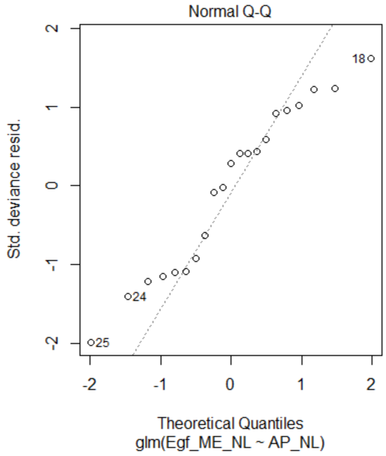
A125: Fitted line of normalized residuals for AP – PC_FU NL



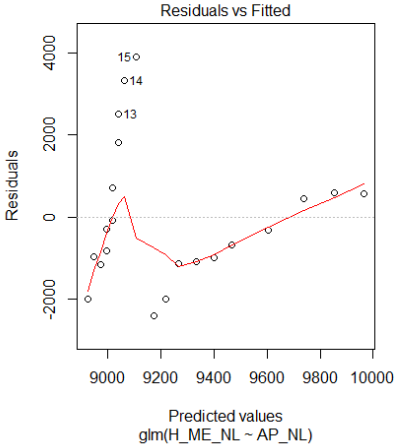
A126: Residual vs Fitted plot for AP – Egf_ME NL



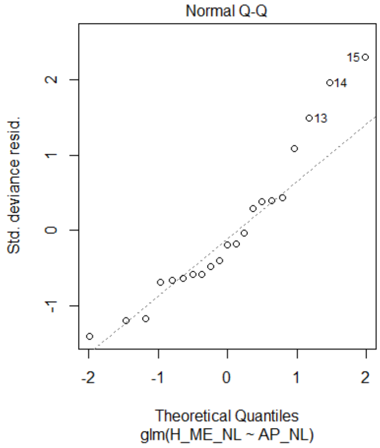
A127: Fitted line of normalized residuals for AP – Egf_ME NL



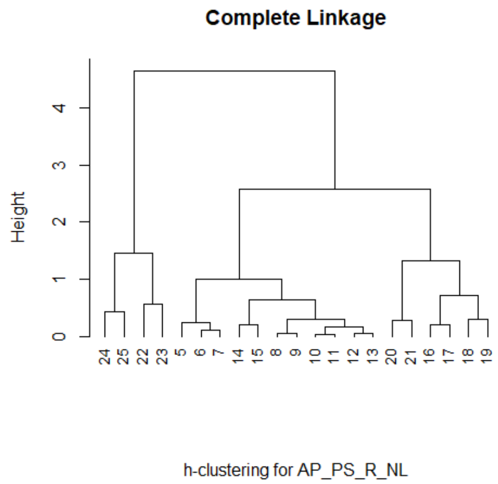
A128: Residual vs Fitted plot for AP – H_ME NL



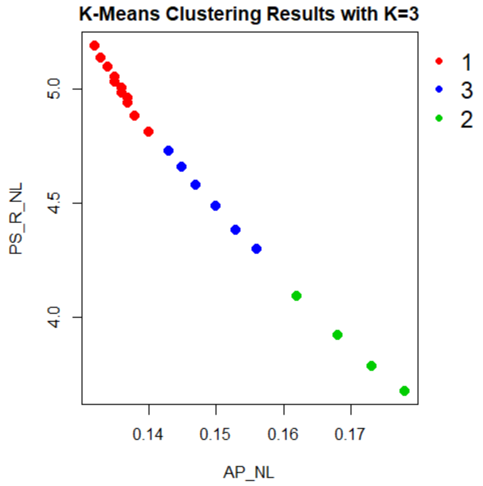
A129: Fitted line of normalized residuals for AP – H_ME NL



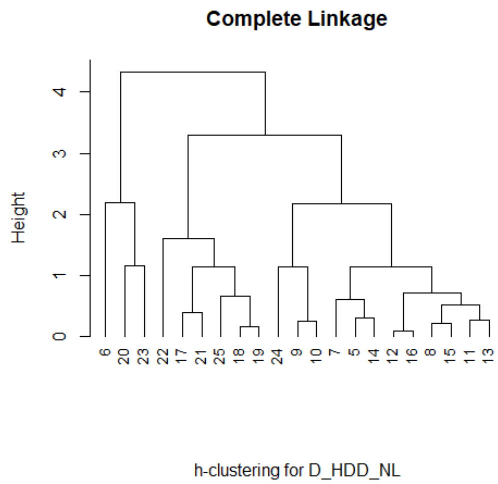
A130: Hierarchical clustering dendrogram for AP – PS_R NL



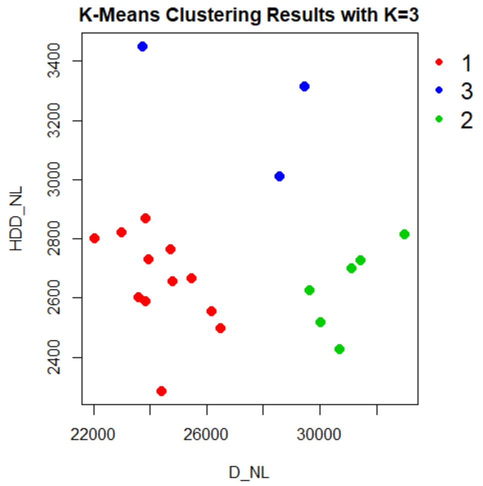
A131: K-means cluster separation in scatter plot for AP – PS_R NL



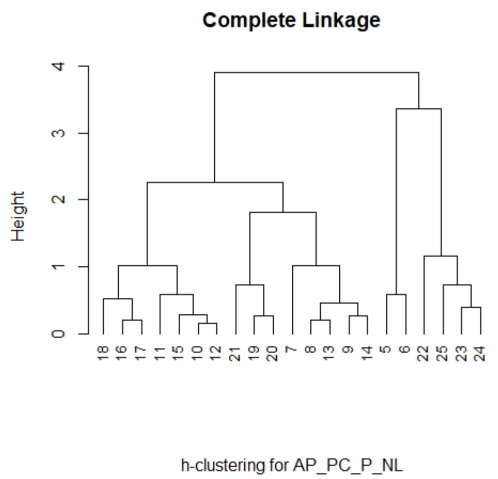
A132: Hierarchical clustering dendrogram for D - HDD NL



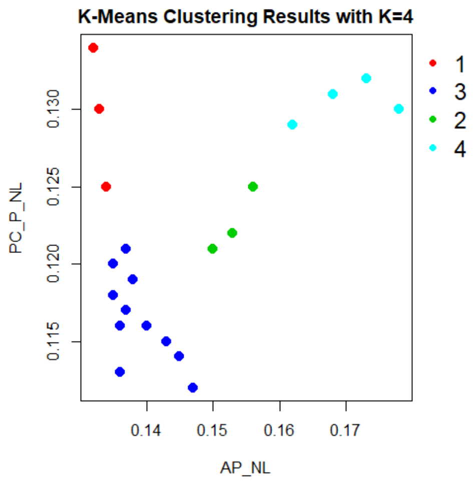
A133: K-means cluster separation in scatter plot for D - HDD NL



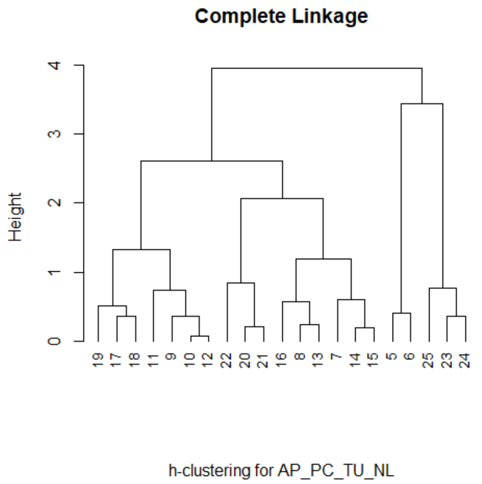
A134: Hierarchical clustering dendrogram for AP – PC_P NL



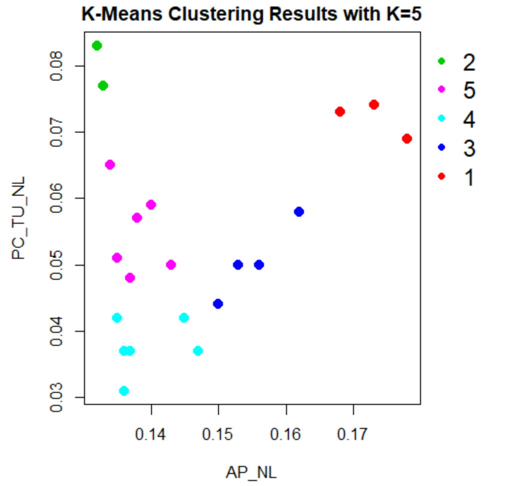
A135: K-means cluster separation in scatter plot for AP – PC_P NL



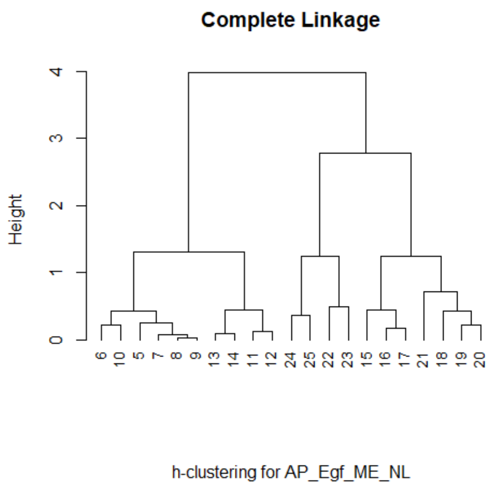
A136: Hierarchical clustering dendrogram for AP – PC_TU NL



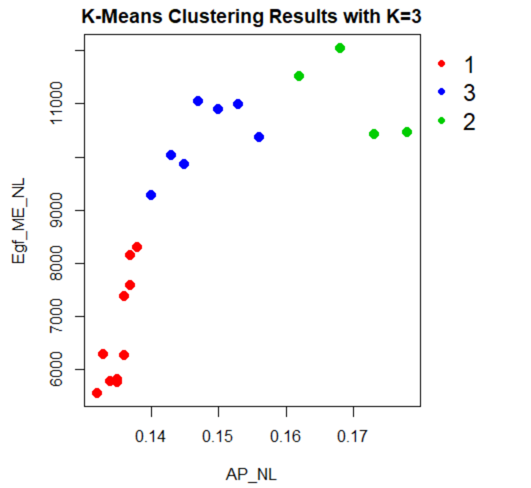
A137: K-means cluster separation in scatter plot for AP – PC_TU NL



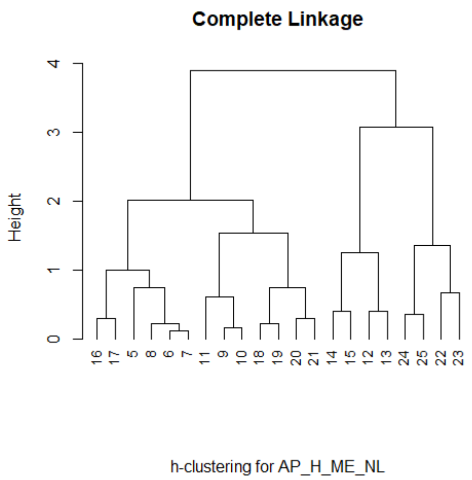
A138: Hierarchical clustering dendrogram for AP – Egf_ME NL



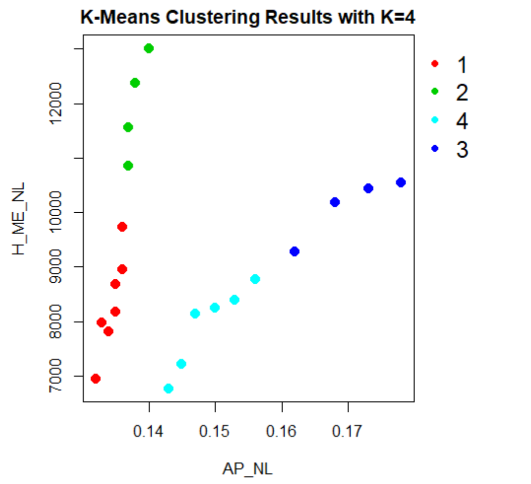
A139: K-means cluster separation in scatter plot for AP – Egf_ME NL



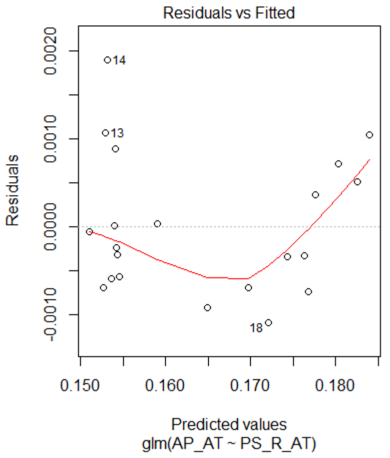
A140: Hierarchical clustering dendrogram for AP – H_ME NL



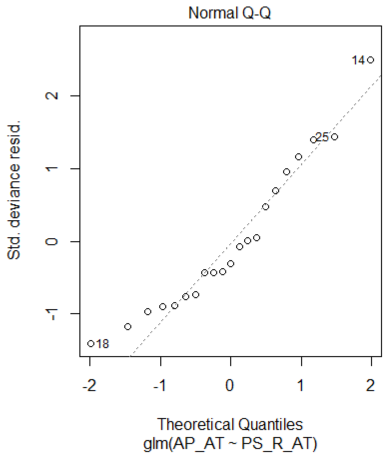
A141: K-means cluster separation in scatter plot for AP – H_ME NL



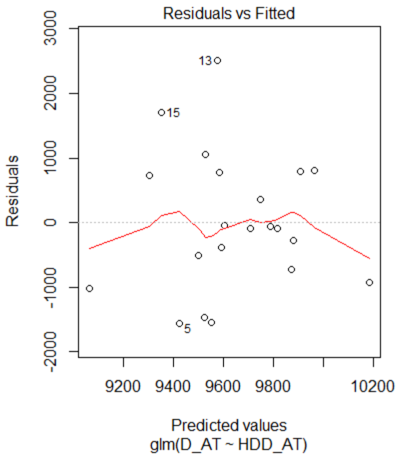
A142: Residual vs Fitted plot for AP – PS_R AT



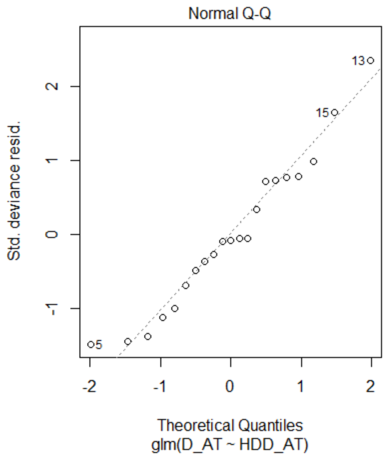
A143: Fitted line of normalized residuals for AP – PS_R AT



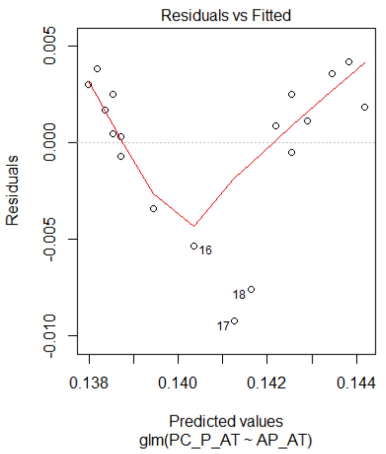
A144: Residual vs Fitted plot for D - HDD AT



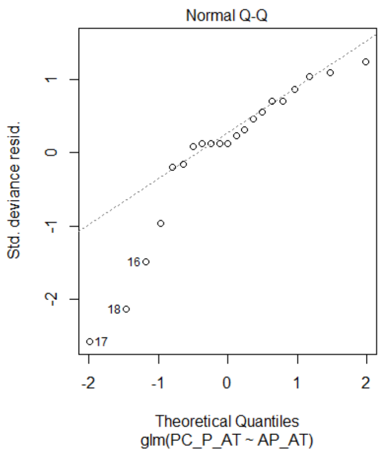
A145: Fitted line of normalized residuals for D - HDD AT



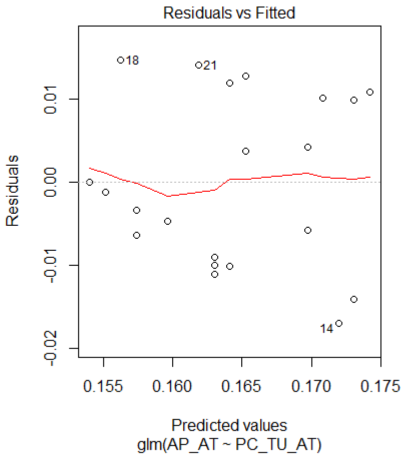
A146: Residual vs Fitted plot for AP – PC_P AT



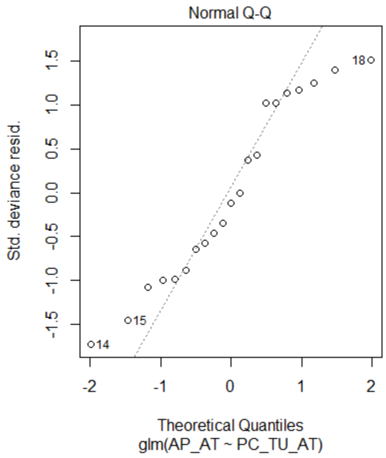
A147: Fitted line of normalized residuals for AP – PC_P AT



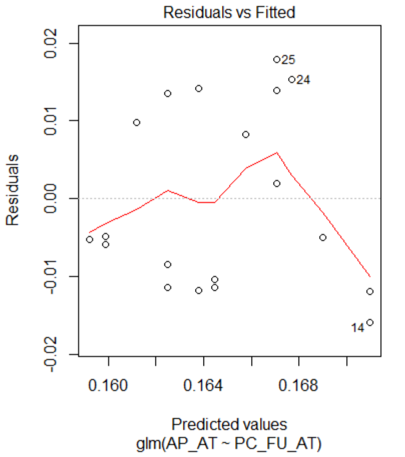
A148: Residual vs Fitted plot for AP – PC_TU AT



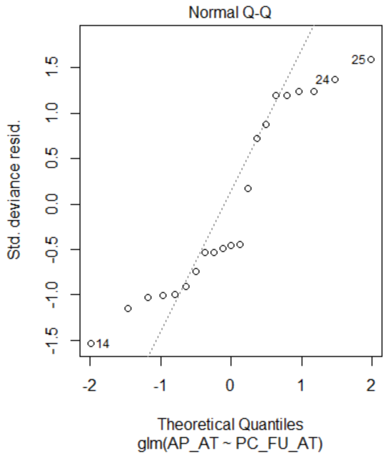
A149: Fitted line of normalized residuals for AP – PC_TU AT



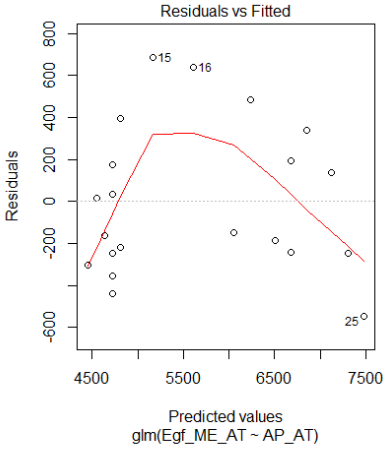
A150: Residual vs Fitted plot for AP – PC_FU AT



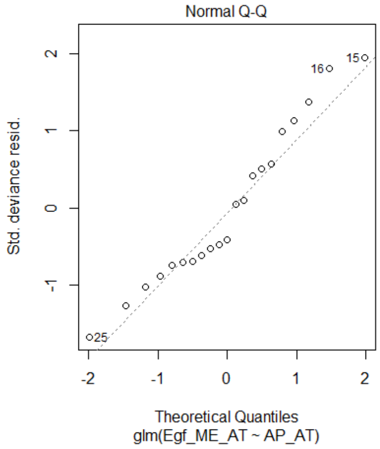
A151: Fitted line of normalized residuals for AP – PC_FU AT



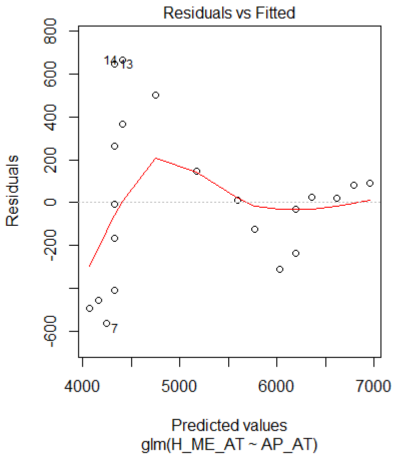
A152: Residual vs Fitted plot for AP – Egf_ME AT



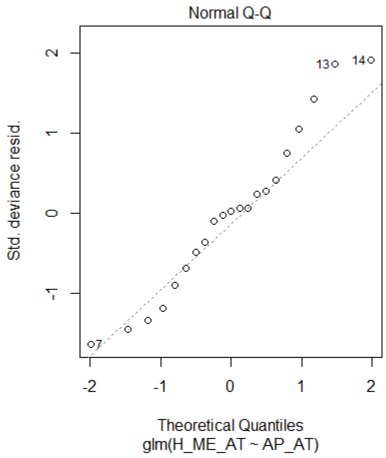
A153: Fitted line of normalized residuals for AP – Egf_ME AT



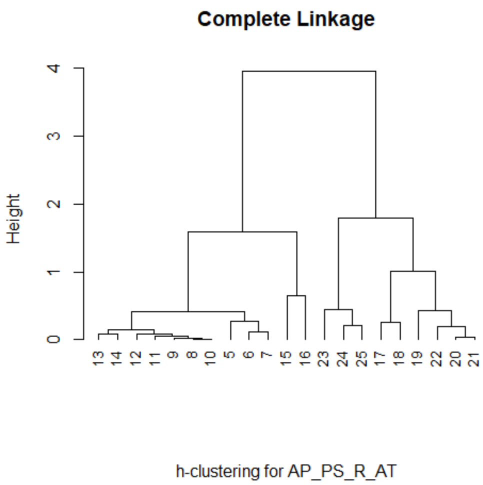
A154: Residual vs Fitted plot for AP – H_ME AT



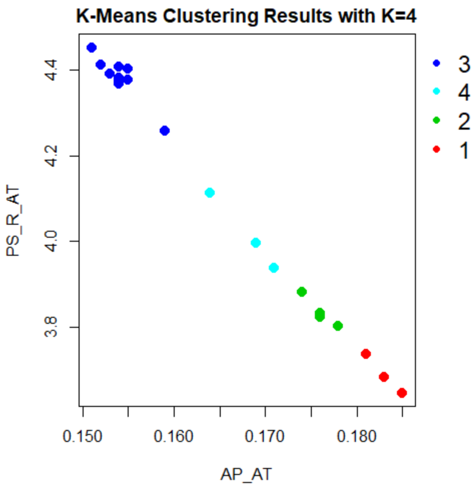
A155: Fitted line of normalized residuals for AP – H_ME AT



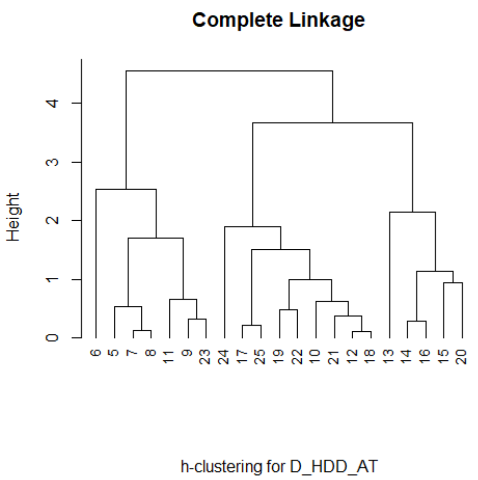
A156: Hierarchical clustering dendrogram for AP – PS_R AT



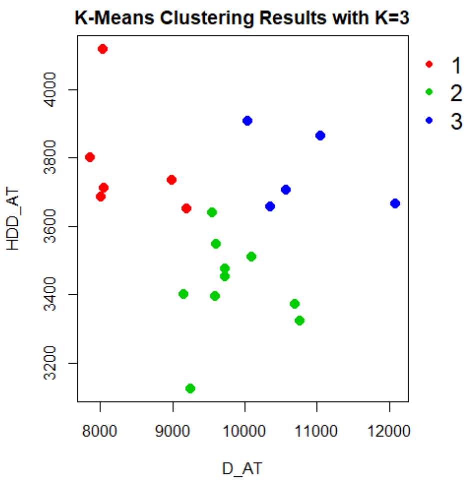
A157: K-means cluster separation in scatter plot for AP – PS_R AT



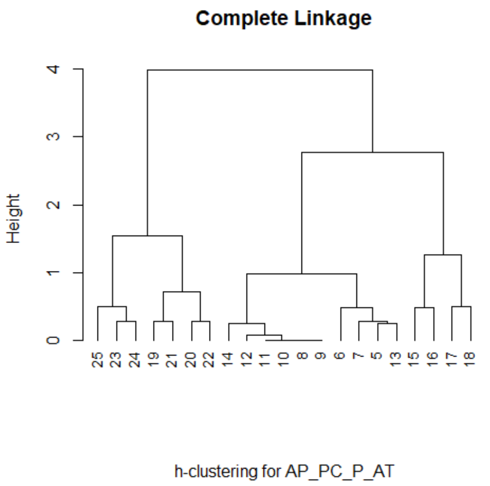
A158: Hierarchical clustering dendrogram for D - HDD AT



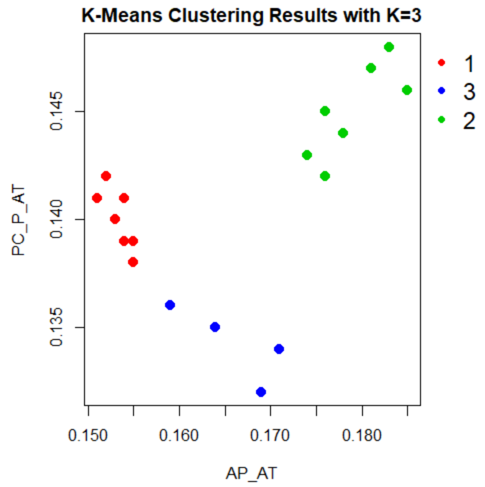
A159: K-means cluster separation in scatter plot for D - HDD AT



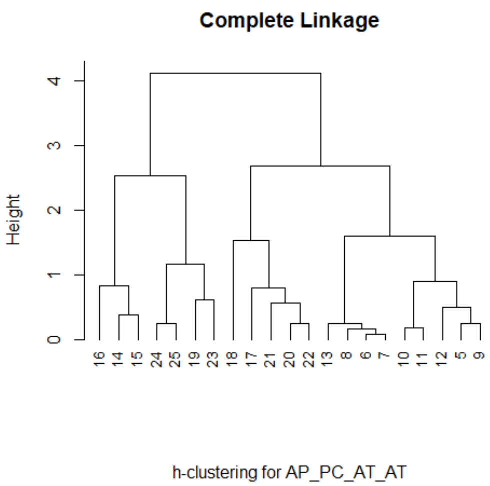
A160: Hierarchical clustering dendrogram for AP – PC_P AT



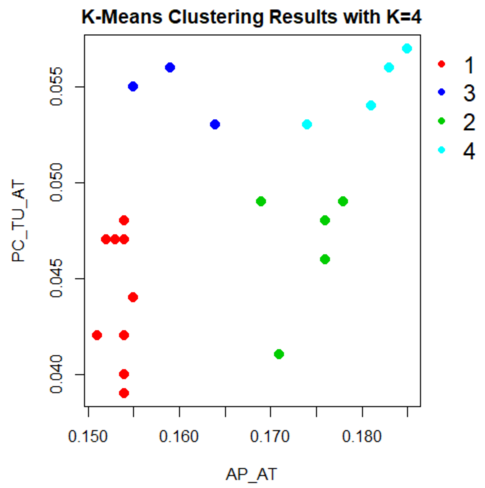
A161: K-means cluster separation in scatter plot for AP – PC_P AT



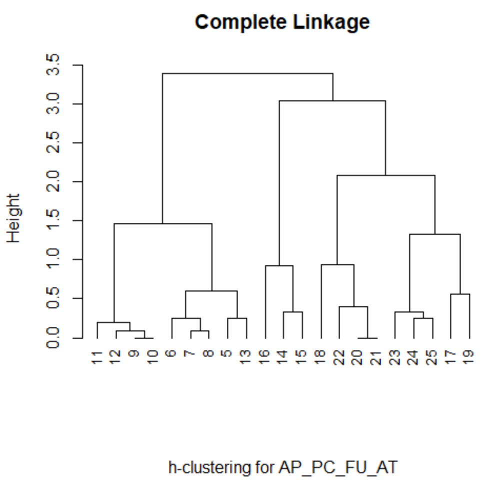
A162: Hierarchical clustering dendrogram for AP – PC_TU AT



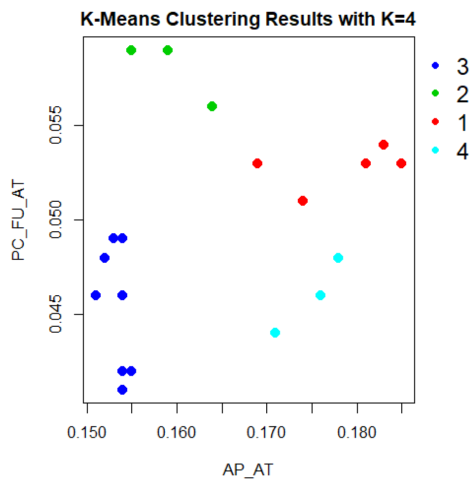
A163: K-means cluster separation in scatter plot for AP – PC_TU AT



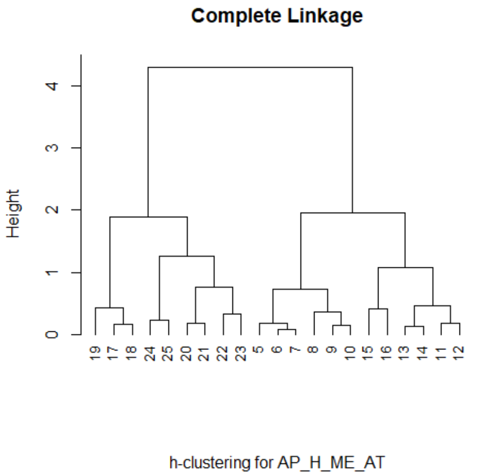
A164: Hierarchical clustering dendrogram for AP – PC_FU AT



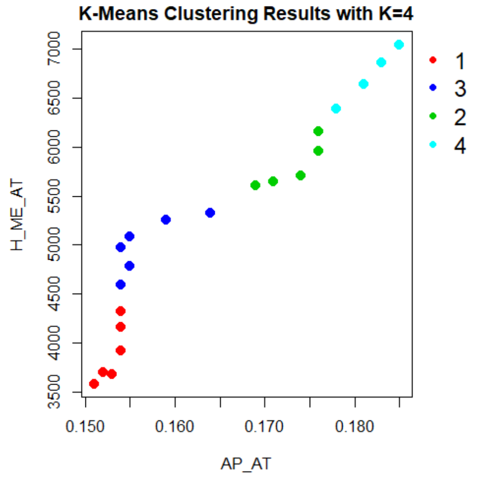
A165: K-means cluster separation in scatter plot for AP – PC_FU AT



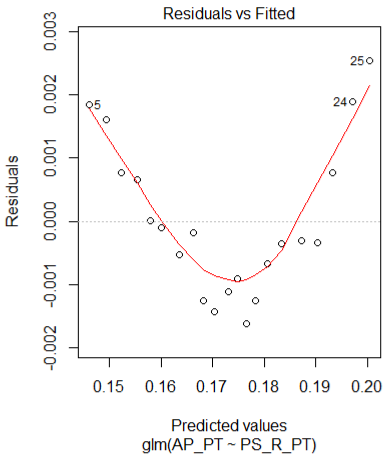
A166: Hierarchical clustering dendrogram for AP – H_ME AT



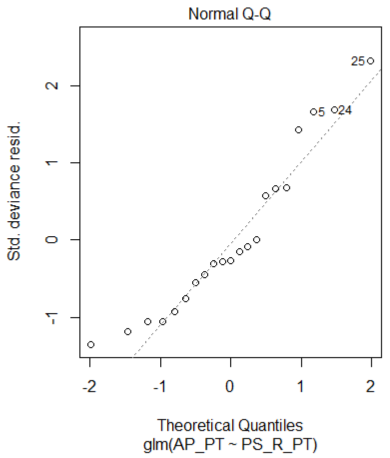
A167: K-means cluster separation in scatter plot for AP – H_ME AT



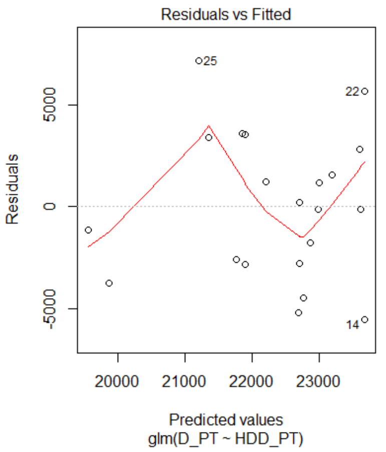
A168: Residual vs Fitted plot for AP – PS_R PT



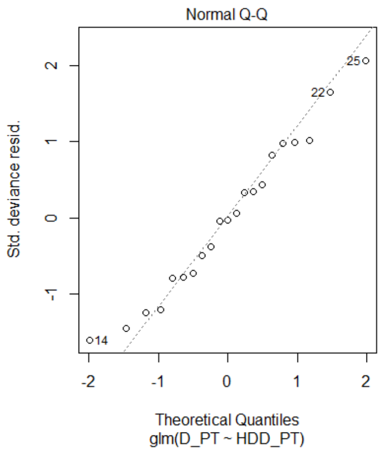
A169: Fitted line of normalized residuals for AP – PS_R PT



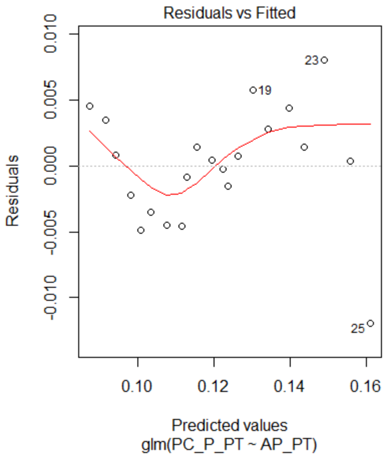
A170: Residual vs Fitted plot for D - HDD PT



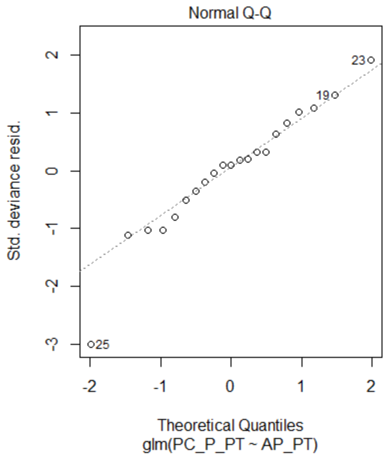
A171: Fitted line of normalized residuals for D - HDD PT



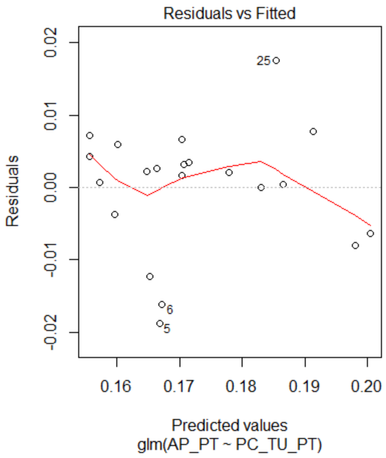
A172: Residual vs Fitted plot for AP - PC_P PT



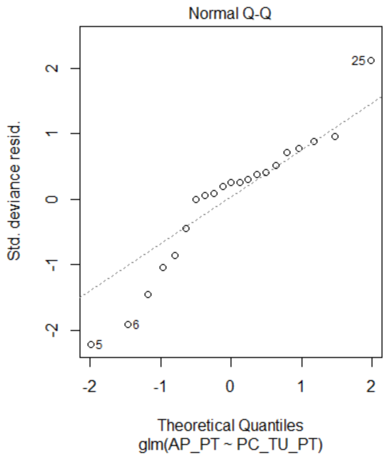
A173: Fitted line of normalized residuals for AP - PC_P PT



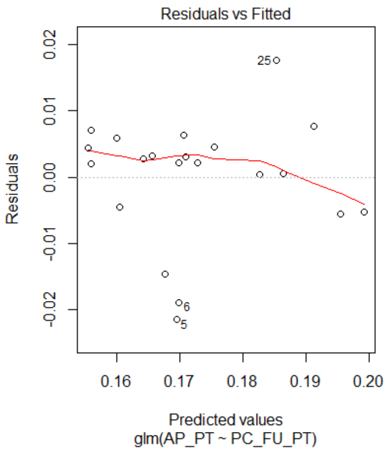
A174: Residual vs Fitted plot for AP - PC_TU PT



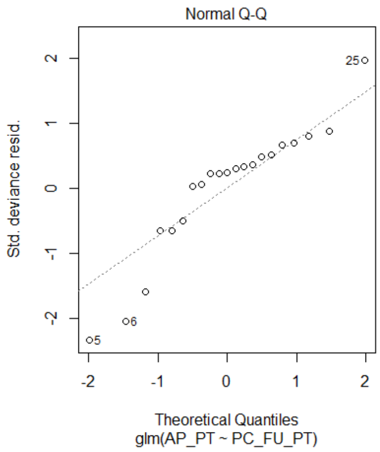
A175: Fitted line of normalized residuals for AP - PC_TU PT



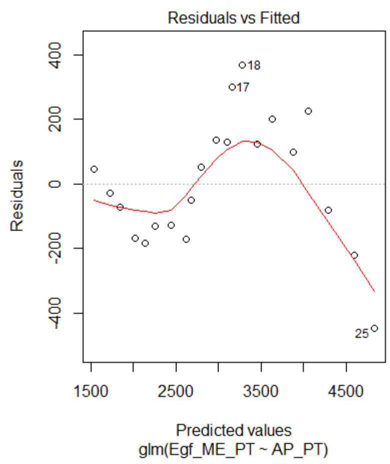
A176: Residual vs Fitted plot for AP - PC_FU PT



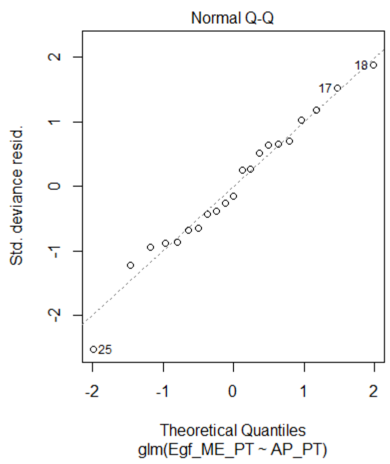
A177: Fitted line of normalized residuals for AP - PC_FU PT



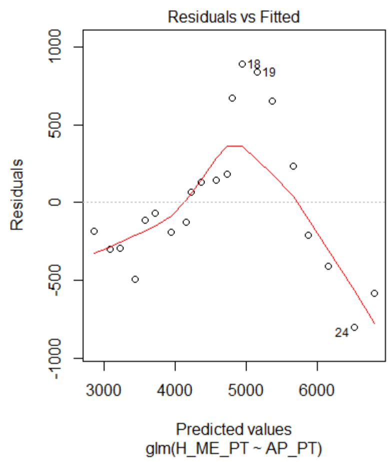
A178: Residual vs Fitted plot for AP – Egf_ME PT



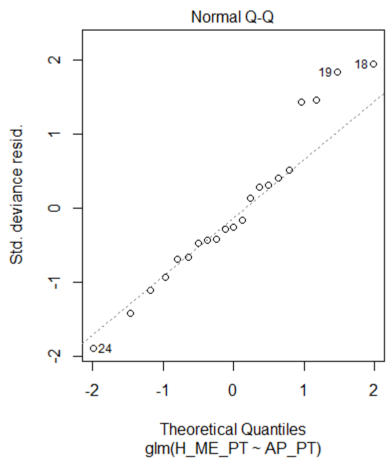
A179: Fitted line of normalized residuals for AP – Egf_ME PT



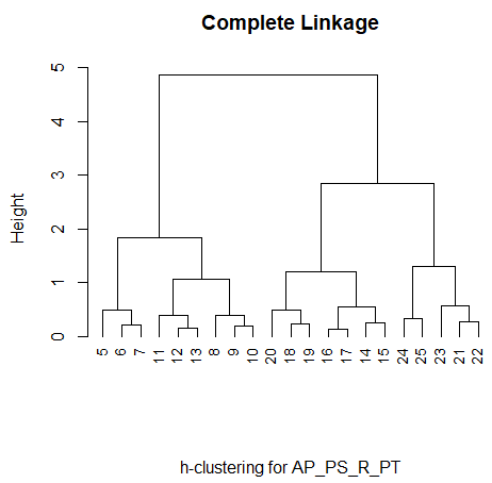
A180: Residual vs Fitted plot for AP – H_ME PT



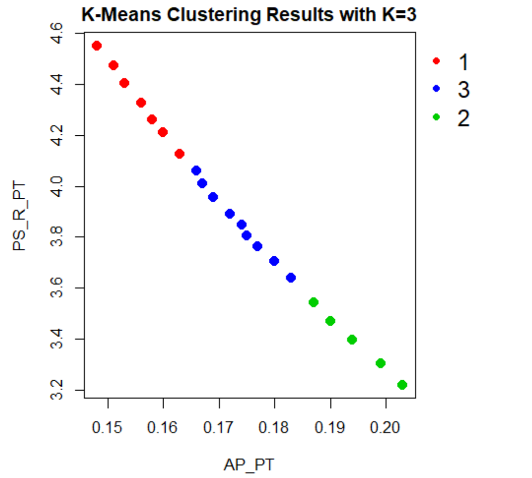
A181: Fitted line of normalized residuals for AP – H_ME PT



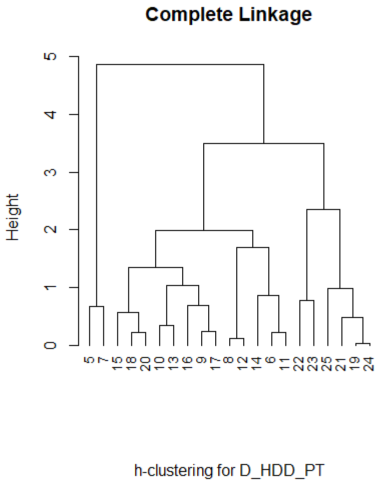
A182: Hierarchical clustering dendrogram for AP – PS_R PT



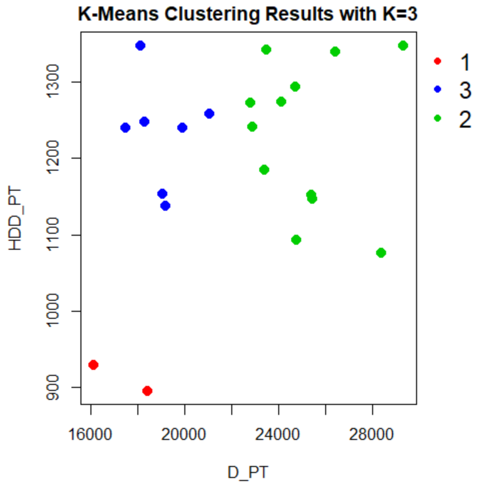
A183: K-means cluster separation in scatter plot for AP – PS_R PT



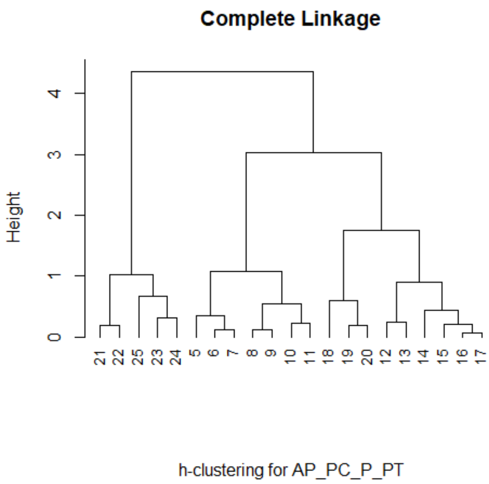
A184: Hierarchical clustering dendrogram for D - HDD PT



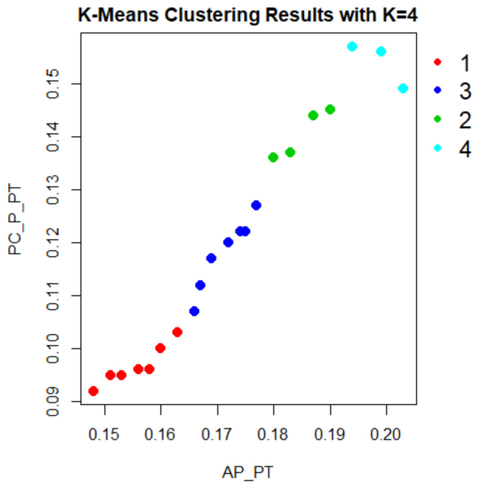
A185: K-means cluster separation in scatter plot for D - HDD PT



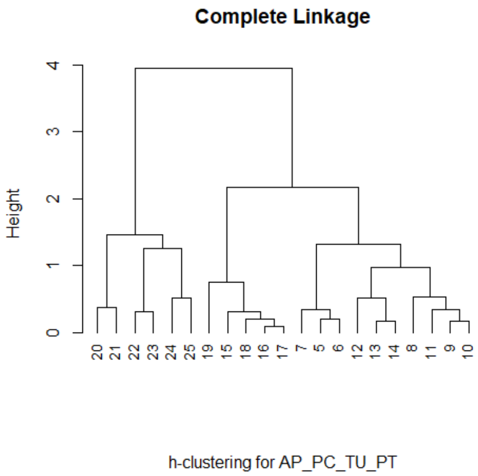
A186: Hierarchical clustering dendrogram for AP – PC_P PT



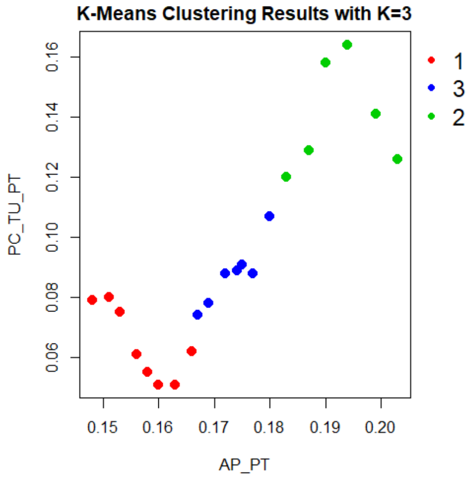
A187: K-means cluster separation in scatter plot for AP – PC_P PT



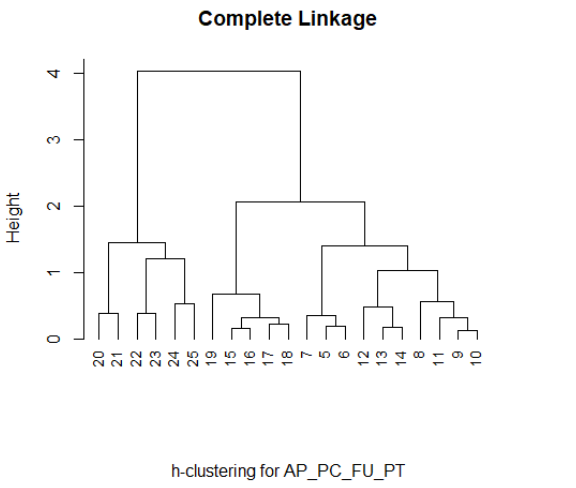
A188: Hierarchical clustering dendrogram for AP – PC_TU PT



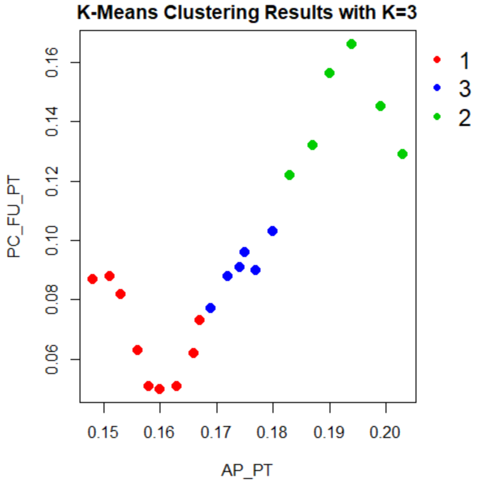
A189: K-means cluster separation in scatter plot for AP – PC_TU PT



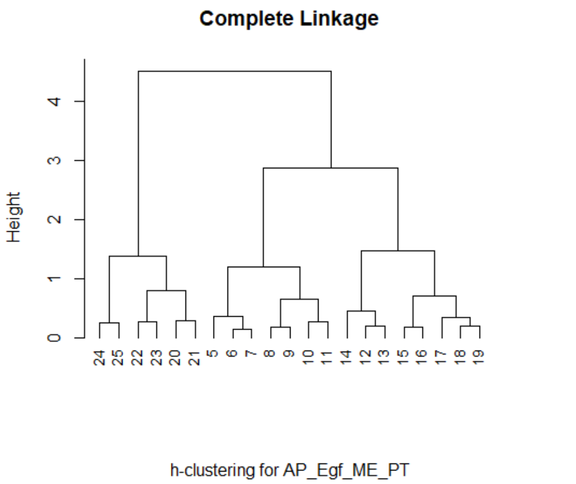
A190: Hierarchical clustering dendrogram for AP – PC_FU PT



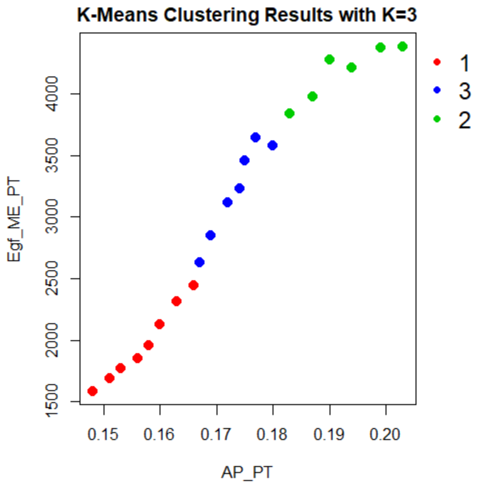
A191: K-means cluster separation in scatter plot for AP – PC_FU PT



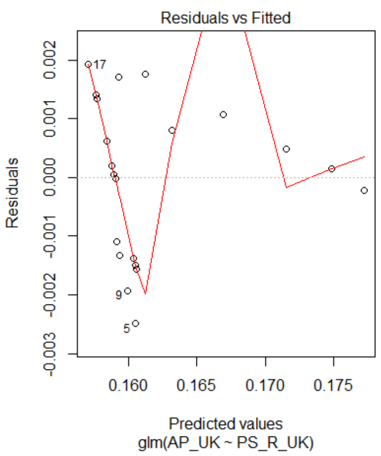
A192: Hierarchical clustering dendrogram for AP – Egf_ME PT



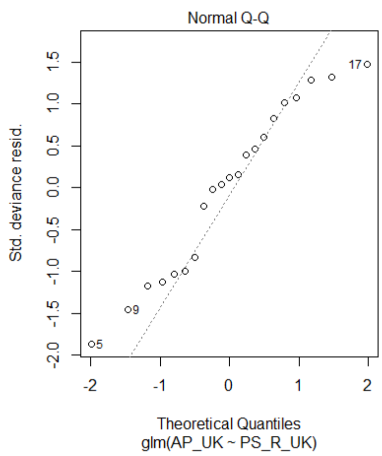
A193: K-means cluster separation in scatter plot for AP – Egf_ME PT



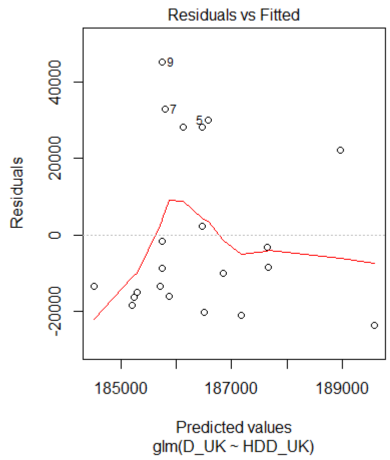
A194: Residual vs Fitted plot for AP – PS_R UK



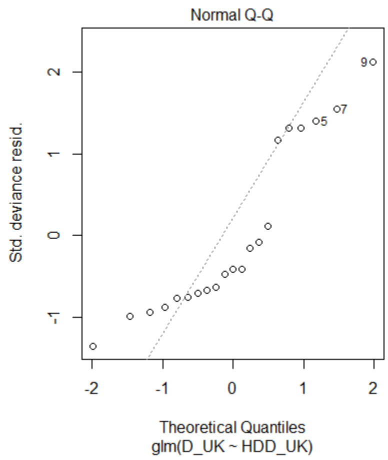
A195: Fitted line of normalized residuals for AP – PS_R UK



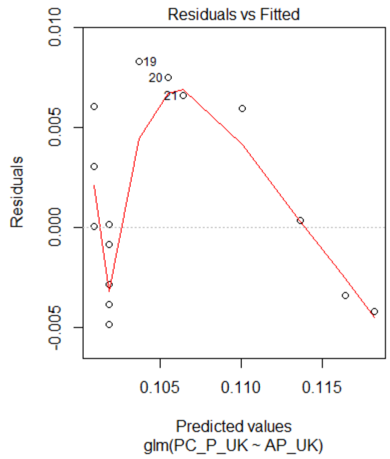
A196: Residual vs Fitted plot for D - HDD UK



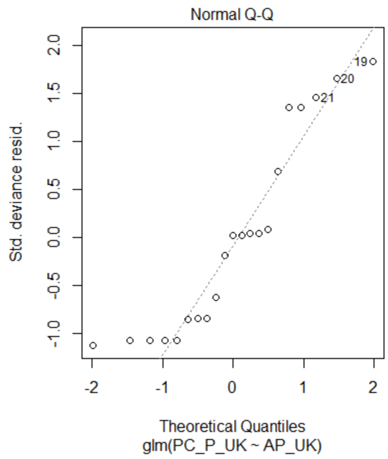
A197: Fitted line of normalized residuals for D - HDD UK



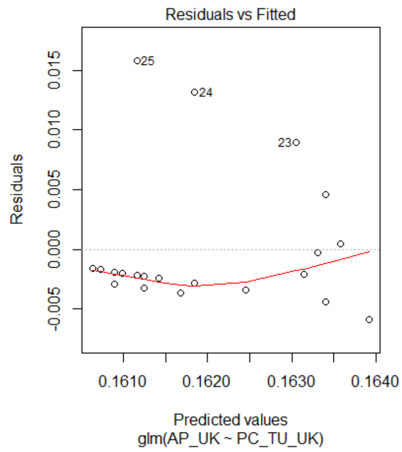
A198: Residual vs Fitted plot for AP – PC_P UK



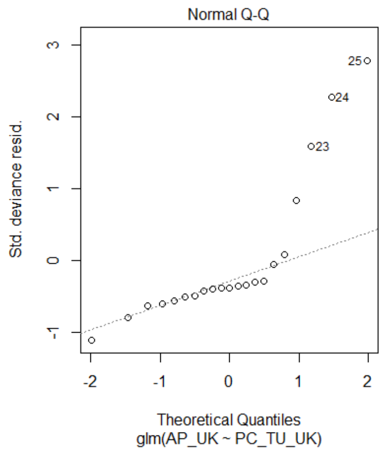
A199: Fitted line of normalized residuals for AP – PC_P UK



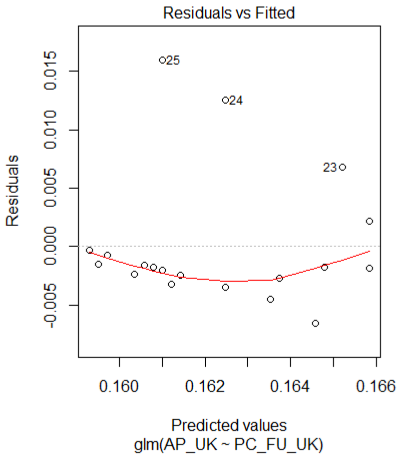
A200: Residual vs Fitted plot for AP – PC_TU UK



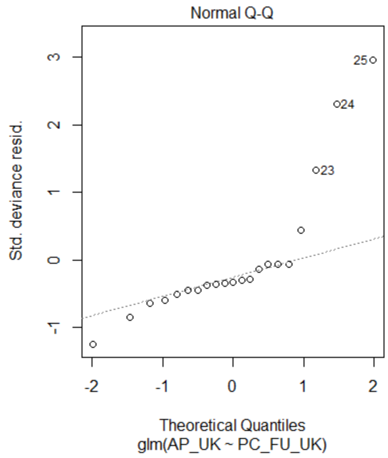
A201: Fitted line of normalized residuals for AP – PC_TU UK



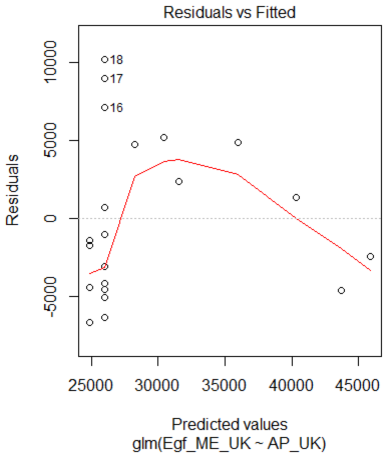
A202: Residual vs Fitted plot for AP – PC_FU UK



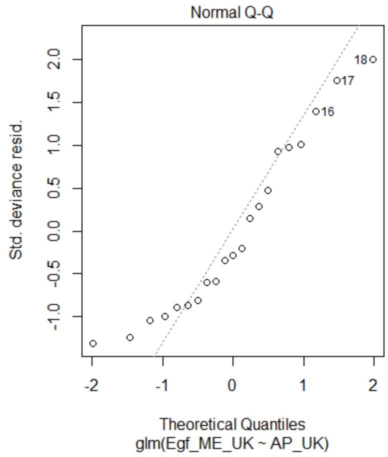
A203: Fitted line of normalized residuals for AP – PC_FU UK



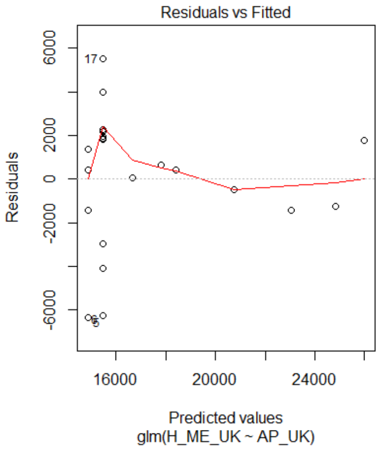
A204: Residual vs Fitted plot for AP – Egf_ME UK



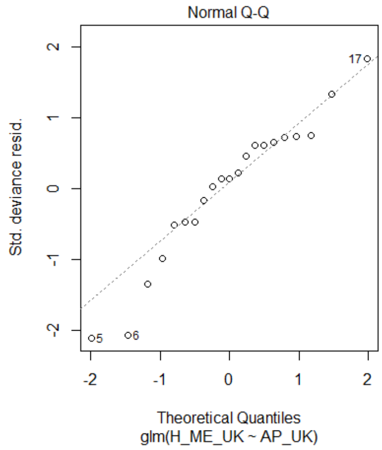
A205: Fitted line of normalized residuals for AP – Egf_ME UK



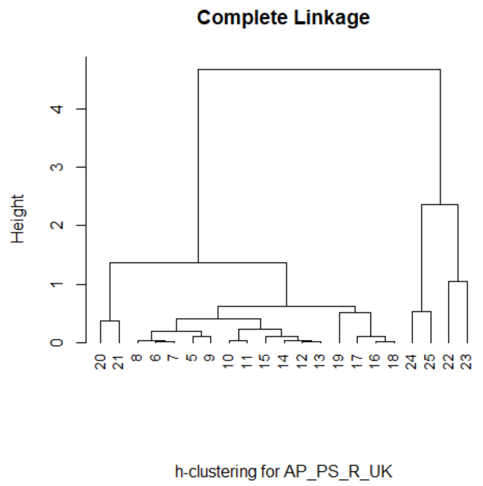
A206: Residual vs Fitted plot for AP – H_ME UK



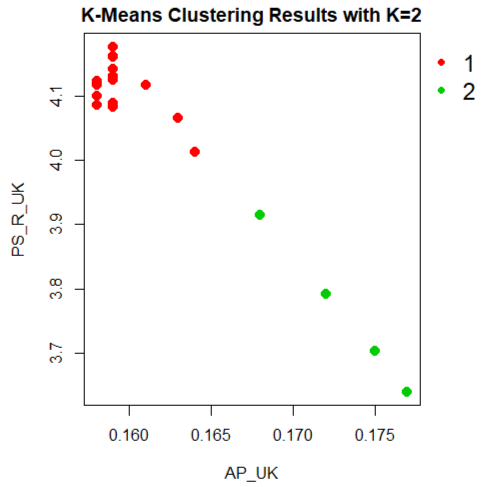
A207: Fitted line of normalized residuals for AP – H_ME UK



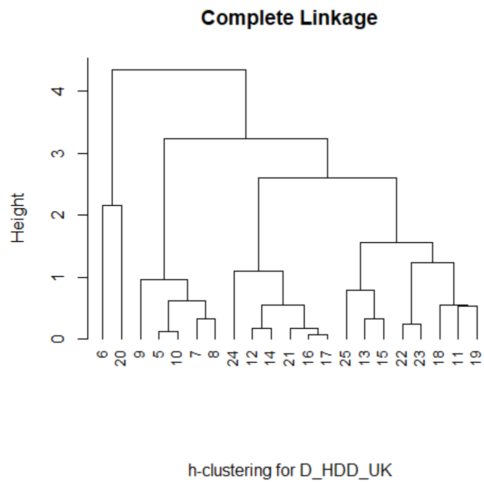
A208: Hierarchical clustering dendrogram for AP – PS_R UK



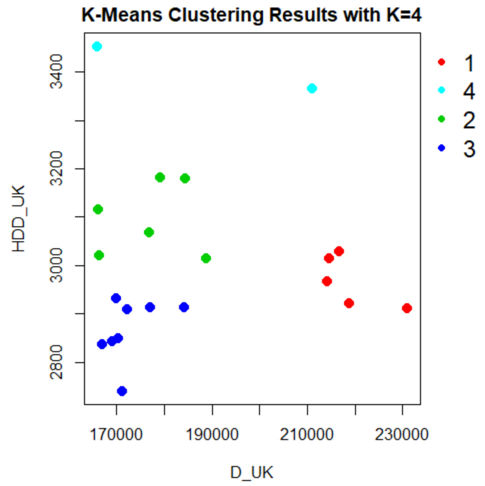
A209: K-means cluster separation in scatter plot for AP – PS_R UK



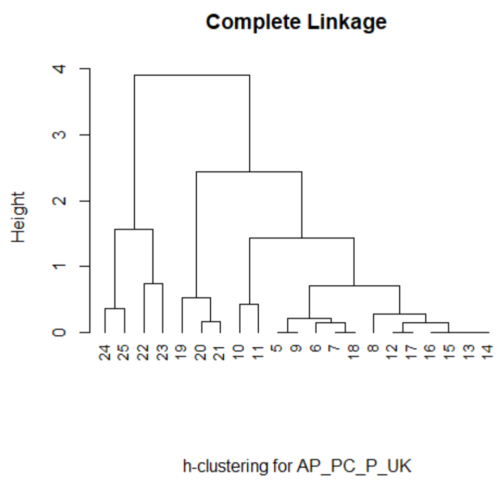
A210: Hierarchical clustering dendrogram for D - HDD UK



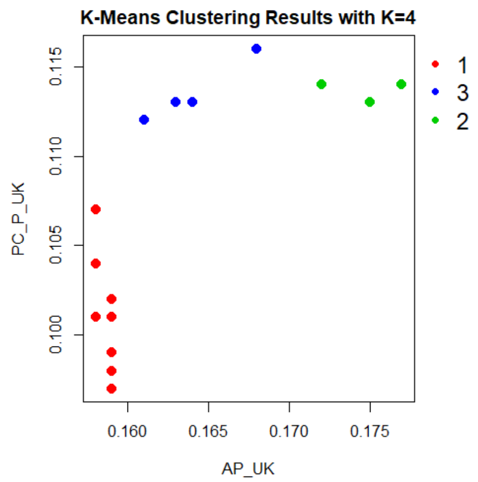
A211: K-means cluster separation in scatter plot for D - HDD UK



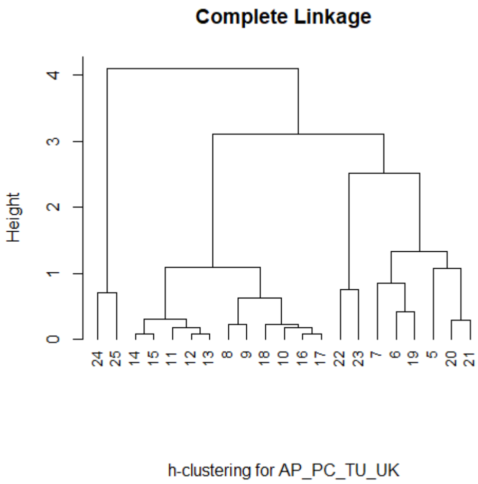
A212: Hierarchical clustering dendrogram for AP – PC_P UK



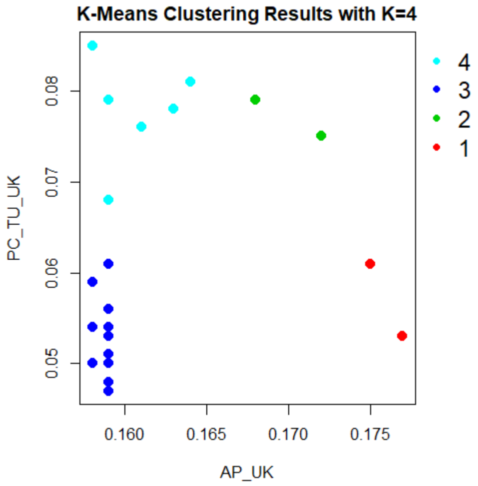
A213: K-means cluster separation in scatter plot for AP – PC_P UK



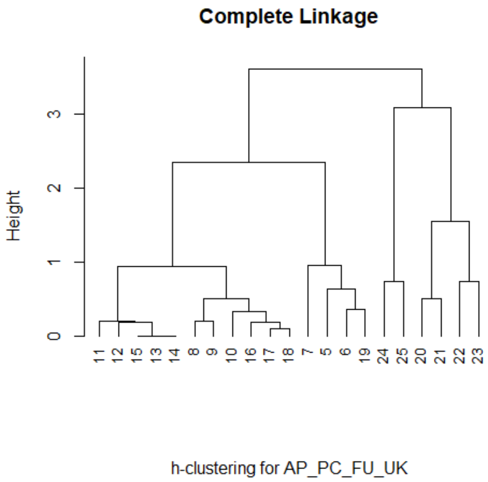
A214: Hierarchical clustering dendrogram for AP – PC_TU UK



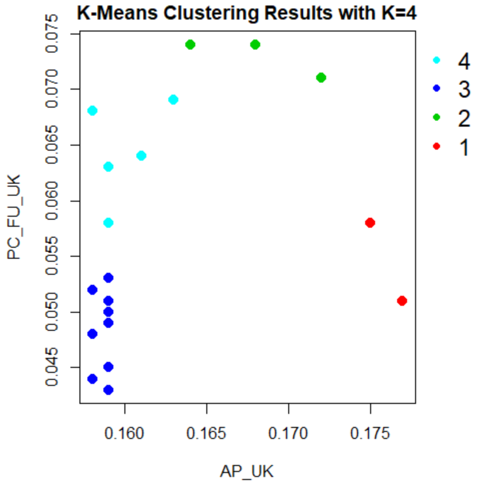
A215: K-means cluster separation in scatter plot for AP – PC_TU UK



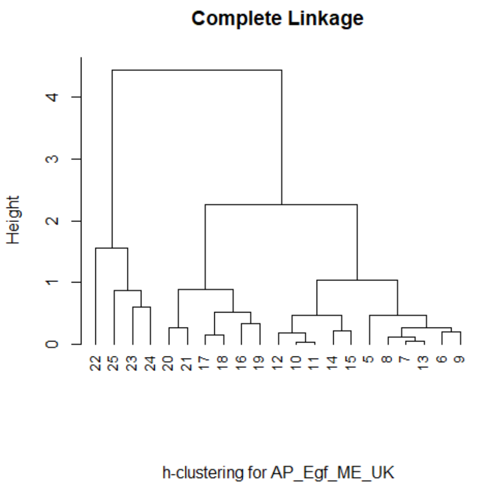
A216: Hierarchical clustering dendrogram for AP – PC_FU UK



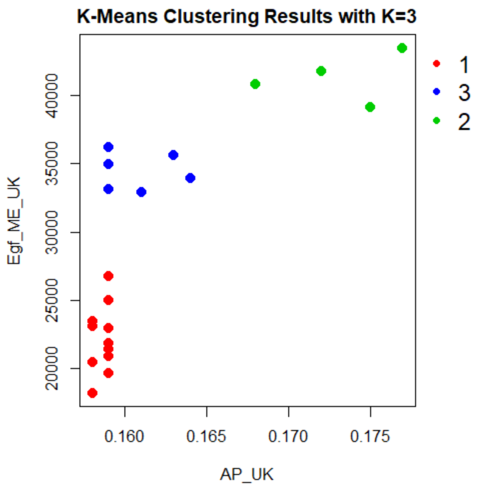
A217: K-means cluster separation in scatter plot for AP – PC_FU UK



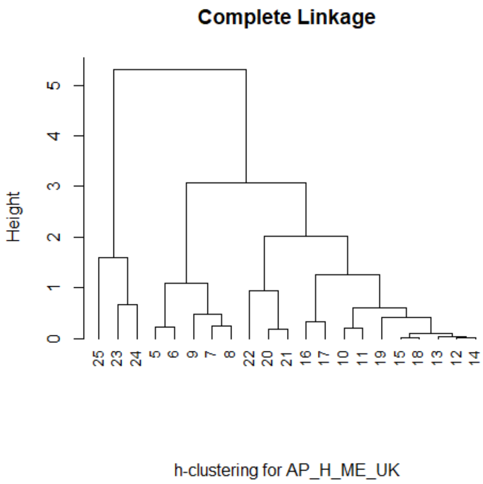
A217: Hierarchical clustering dendrogram for AP – Egf_ME UK



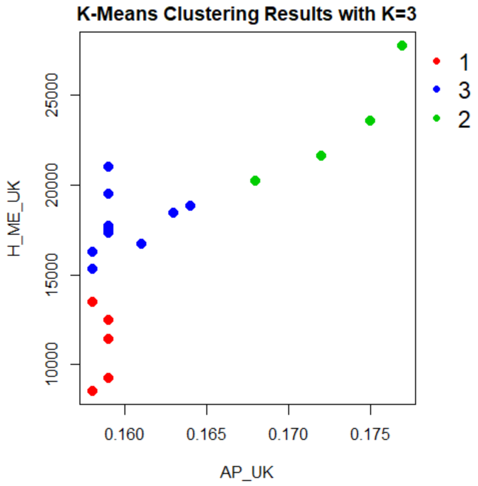
A218: K-means cluster separation in scatter plot for AP – Egf_ME UK



A219: Hierarchical clustering dendrogram for AP – H_ME UK



A220: K-means cluster separation in scatter plot for AP – H_ME UK



A221: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Denmark

Years	AP - PS_R _DK clusters	centroid for AP_DK	centroid for PS_R_DK	AP - PC_TU _DK clusters	centroid for AP_DK	centroid for PC_TU_DK	AP - PC_FU _DK clusters	centroid for AP_DK	centroid for PC_FU_DK	D - HDD _DK clusters	centroid for D_DK	centroid for HDD_DK
1995	1	0.150	4.445	2	0.154	0.063	4	0.159	0.074	3	12,637.750	3,349.282
1996	1	0.150	4.445	2	0.154	0.063	4	0.159	0.074	4	12,992.000	4,012.280
1997	1	0.150	4.445	1	0.149	0.049	1	0.149	0.060	3	12,637.750	3,349.282
1998	1	0.150	4.445	1	0.149	0.049	1	0.149	0.060	3	12,637.750	3,349.282
1999	1	0.150	4.445	1	0.149	0.049	1	0.149	0.060	1	13,301.750	3,028.067
2000	1	0.150	4.445	1	0.149	0.049	3	0.152	0.047	3	12,637.750	3,349.282
2001	1	0.150	4.445	1	0.149	0.049	3	0.152	0.047	3	12,637.750	3,349.282
2002	1	0.150	4.445	1	0.149	0.049	3	0.152	0.047	2	13,866.000	3,258.180
2003	1	0.150	4.445	1	0.149	0.049	1	0.149	0.060	2	13,866.000	3,258.180
2004	1	0.150	4.445	1	0.149	0.049	1	0.149	0.060	3	12,637.750	3,349.282
2005	1	0.150	4.445	1	0.149	0.049	3	0.152	0.047	3	12,637.750	3,349.282
2006	1	0.150	4.445	3	0.154	0.037	3	0.152	0.047	3	12,637.750	3,349.282
2007	1	0.150	4.445	3	0.154	0.037	3	0.152	0.047	1	13,301.750	3,028.067
2008	3	0.162	4.072	3	0.154	0.037	3	0.152	0.047	1	13,301.750	3,028.067
2009	3	0.162	4.072	2	0.154	0.063	3	0.152	0.047	2	13,866.000	3,258.180
2010	3	0.162	4.072	5	0.168	0.075	4	0.159	0.074	4	12,992.000	4,012.280
2011	3	0.162	4.072	5	0.168	0.075	4	0.159	0.074	2	13,866.000	3,258.180
2012	2	0.180	3.594	5	0.168	0.075	2	0.180	0.070	2	13,866.000	3,258.180
2013	2	0.180	3.594	4	0.182	0.066	2	0.180	0.070	2	13,866.000	3,258.180
2014	2	0.180	3.594	4	0.182	0.066	2	0.180	0.070	1	13,301.750	3,028.067
2015	2	0.180	3.594	4	0.182	0.066	2	0.180	0.070	2	13,866.000	3,258.180

A222: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Denmark

Years	AP - Egf_ME _DK clusters	centroid for AP_DK	centroid for Egf_ME_DK	AP - H_ME _DK clusters	centroid for AP_DK	centroid for H_ME_DK	AP - PC_P _DK clusters	centroid for AP_DK	centroid for PC_P_DK
1995	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
1996	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
1997	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
1998	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
1999	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
2000	1	0.150	4,650.033	1	0.149	2,022.322	1	0.150	0.108
2001	3	0.152	6,035.025	1	0.149	2,022.322	1	0.150	0.108
2002	3	0.152	6,035.025	1	0.149	2,022.322	1	0.150	0.108
2003	3	0.152	6,035.025	1	0.149	2,022.322	1	0.150	0.108
2004	3	0.152	6,035.025	3	0.152	2,791.280	1	0.150	0.108
2005	3	0.152	6,035.025	3	0.152	2,791.280	1	0.150	0.108
2006	3	0.152	6,035.025	3	0.152	2,791.280	1	0.150	0.108
2007	3	0.152	6,035.025	3	0.152	2,791.280	1	0.150	0.108
2008	3	0.152	6,035.025	3	0.152	2,791.280	1	0.150	0.108
2009	3	0.152	6,035.025	2	0.163	3,311.000	3	0.166	0.127
2010	2	0.168	7,581.900	2	0.163	3,311.000	3	0.166	0.127
2011	2	0.168	7,581.900	2	0.163	3,311.000	3	0.166	0.127
2012	2	0.168	7,581.900	4	0.180	3,491.575	3	0.166	0.127
2013	4	0.182	7,395.167	4	0.180	3,491.575	2	0.182	0.136
2014	4	0.182	7,395.167	4	0.180	3,491.575	2	0.182	0.136
2015	4	0.182	7,395.167	4	0.180	3,491.575	2	0.182	0.136

A223: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Germany

Years	AP - PS_R _DE clusters	centroid for AP_DE	centroid for PS_R_DE	AP - PC_TU_ DE clusters	centroid for AP_DE	centroid for PC_TU_DE	AP - PC_FU_ DE clusters	centroid for AP_DE	centroid for PC_FU_DE	D - HDD _DE clusters	centroid for D_DE	centroid for HDD_DE
1995	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
1996	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	3	114,802.500	3,722.920
1997	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
1998	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
1999	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
2000	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
2001	4	0.158	4.329	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
2002	1	0.171	3.965	1	0.160	0.086	1	0.160	0.092	1	91,363.847	3,060.554
2003	1	0.171	3.965	3	0.183	0.104	3	0.186	0.098	1	91,363.847	3,060.554
2004	2	0.186	3.596	3	0.183	0.104	3	0.186	0.098	1	91,363.847	3,060.554
2005	2	0.186	3.596	3	0.183	0.104	3	0.186	0.098	1	91,363.847	3,060.554
2006	2	0.186	3.596	3	0.183	0.104	3	0.186	0.098	1	91,363.847	3,060.554
2007	3	0.206	3.212	2	0.202	0.076	3	0.186	0.098	1	91,363.847	3,060.554
2008	3	0.206	3.212	2	0.202	0.076	2	0.206	0.057	1	91,363.847	3,060.554
2009	3	0.206	3.212	2	0.202	0.076	2	0.206	0.057	2	141,017.167	2,990.145
2010	3	0.206	3.212	2	0.202	0.076	2	0.206	0.057	3	114,802.500	3,722.920
2011	3	0.206	3.212	4	0.208	0.052	2	0.206	0.057	2	141,017.167	2,990.145
2012	3	0.206	3.212	4	0.208	0.052	2	0.206	0.057	2	141,017.167	2,990.145
2013	3	0.206	3.212	4	0.208	0.052	2	0.206	0.057	2	141,017.167	2,990.145
2014	3	0.206	3.212	4	0.208	0.052	2	0.206	0.057	2	141,017.167	2,990.145
2015	3	0.206	3.212	4	0.208	0.052	2	0.206	0.057	2	141,017.167	2,990.145

A224: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Germany

Years	AP - Egf_ME_ DE clusters	centroid for AP_DE	centroid for Egf_ME_DE	AP - H_ME_ DE clusters	centroid for AP_DE	centroid for H_ME_DE	AP - PC_P_ DE clusters	centroid for AP_DE	centroid for PC_P_DE
1995	1	0.158	38,815.467	1	0.159	41,038.701	3	0.159	0.125
1996	1	0.158	38,815.467	1	0.159	41,038.701	3	0.159	0.125
1997	1	0.158	38,815.467	1	0.159	41,038.701	3	0.159	0.125
1998	1	0.158	38,815.467	1	0.159	41,038.701	3	0.159	0.125
1999	1	0.158	38,815.467	1	0.159	41,038.701	3	0.159	0.125
2000	3	0.173	47,123.000	1	0.159	41,038.701	3	0.159	0.125
2001	3	0.173	47,123.000	1	0.159	41,038.701	3	0.159	0.125
2002	3	0.173	47,123.000	3	0.178	52,484.000	4	0.178	0.130
2003	3	0.173	47,123.000	3	0.178	52,484.000	4	0.178	0.130
2004	2	0.192	56,715.333	3	0.178	52,484.000	4	0.178	0.130
2005	2	0.192	56,715.333	3	0.178	52,484.000	4	0.178	0.130
2006	2	0.192	56,715.333	2	0.199	59,954.500	2	0.201	0.126
2007	4	0.206	67,577.625	2	0.199	59,954.500	1	0.206	0.119
2008	4	0.206	67,577.625	2	0.199	59,954.500	1	0.206	0.119
2009	4	0.206	67,577.625	2	0.199	59,954.500	2	0.201	0.126
2010	4	0.206	67,577.625	4	0.208	75,306.833	2	0.201	0.126
2011	4	0.206	67,577.625	4	0.208	75,306.833	1	0.206	0.119
2012	4	0.206	67,577.625	4	0.208	75,306.833	1	0.206	0.119
2013	4	0.206	67,577.625	4	0.208	75,306.833	1	0.206	0.119
2014	4	0.206	67,577.625	4	0.208	75,306.833	1	0.206	0.119
2015	4	0.206	67,577.625	4	0.208	75,306.833	1	0.206	0.119

A225: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Ireland

Years	AP - PS_R _IE clusters	centroid for AP_IE	centroid for PS_R_IE	AP - PC_TU_ IE clusters	centroid for AP_IE	centroid for PC_TU_IE	AP - PC_FU_ IE clusters	centroid for AP_IE	centroid for PC_FU_IE	D - HDD _IE clusters	centroid for D_IE	centroid for HDD_IE
1995	3	0.115	5.736	2	0.114	0.131	4	0.113	0.113	3	8,397.667	2,622.335
1996	3	0.115	5.736	2	0.114	0.131	4	0.113	0.113	2	8,126.429	2,840.959
1997	3	0.115	5.736	2	0.114	0.131	4	0.113	0.113	1	11,172.000	2,601.520
1998	3	0.115	5.736	3	0.111	0.052	3	0.111	0.051	1	11,172.000	2,601.520
1999	3	0.115	5.736	3	0.111	0.052	3	0.111	0.051	1	11,172.000	2,601.520
2000	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	1	11,172.000	2,601.520
2001	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	5	9,761.250	2,729.887
2002	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	5	9,761.250	2,729.887
2003	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	5	9,761.250	2,729.887
2004	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	3	8,397.667	2,622.335
2005	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	3	8,397.667	2,622.335
2006	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	3	8,397.667	2,622.335
2007	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	3	8,397.667	2,622.335
2008	2	0.110	6.157	3	0.111	0.052	3	0.111	0.051	2	8,126.429	2,840.959
2009	2	0.110	6.157	2	0.114	0.131	4	0.113	0.113	2	8,126.429	2,840.959
2010	2	0.110	6.157	2	0.114	0.131	4	0.113	0.113	4	7,242.000	3,168.560
2011	3	0.115	5.736	2	0.114	0.131	4	0.113	0.113	2	8,126.429	2,840.959
2012	3	0.115	5.736	2	0.114	0.131	2	0.121	0.126	2	8,126.429	2,840.959
2013	1	0.126	5.241	1	0.126	0.119	2	0.121	0.126	2	8,126.429	2,840.959
2014	1	0.126	5.241	1	0.126	0.119	1	0.128	0.099	3	8,397.667	2,622.335
2015	1	0.126	5.241	1	0.126	0.119	1	0.128	0.099	2	8,126.429	2,840.959

A226: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Ireland

Years	AP - Egf_ME_ IE clusters	centroid for AP_IE	centroid for Egf_ME_IE	AP - H_ME_ IE clusters	centroid for AP_IE	centroid for H_ME_IE	AP - PC_P_IE clusters	centroid for AP_IE	centroid for PC_P_IE
1995	1	0.113	1,447.980	1	0.159	41,038.701	1	0.114	0.043
1996	1	0.113	1,447.980	1	0.159	41,038.701	1	0.114	0.043
1997	1	0.113	1,447.980	1	0.159	41,038.701	1	0.114	0.043
1998	1	0.113	1,447.980	1	0.159	41,038.701	1	0.114	0.043
1999	1	0.113	1,447.980	1	0.159	41,038.701	1	0.114	0.043
2000	1	0.113	1,447.980	1	0.159	41,038.701	3	0.110	0.054
2001	1	0.113	1,447.980	1	0.159	41,038.701	3	0.110	0.054
2002	1	0.113	1,447.980	1	0.159	41,038.701	3	0.110	0.054
2003	1	0.113	1,447.980	1	0.159	41,038.701	3	0.110	0.054
2004	1	0.113	1,447.980	1	0.159	41,038.701	3	0.110	0.054
2005	2	0.110	2,802.771	2	0.199	59,954.500	3	0.110	0.054
2006	2	0.110	2,802.771	2	0.199	59,954.500	3	0.110	0.054
2007	2	0.110	2,802.771	2	0.199	59,954.500	3	0.110	0.054
2008	2	0.110	2,802.771	2	0.199	59,954.500	3	0.110	0.054
2009	2	0.110	2,802.771	2	0.199	59,954.500	2	0.114	0.076
2010	2	0.110	2,802.771	2	0.199	59,954.500	2	0.114	0.076
2011	2	0.110	2,802.771	2	0.199	59,954.500	2	0.114	0.076
2012	3	0.124	3,225.100	3	0.178	52,484.000	2	0.114	0.076
2013	3	0.124	3,225.100	3	0.178	52,484.000	4	0.126	0.068
2014	3	0.124	3,225.100	3	0.178	52,484.000	4	0.126	0.068
2015	3	0.124	3,225.100	3	0.178	52,484.000	4	0.126	0.068

A227: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Greece

Years	AP - PC_P_EL clusters	centroid for AP_EL	centroid for PC_P_EL	AP - PC_TU_EL clusters	centroid for AP_EL	centroid for PC_TU_EL	AP - PC_FU_EL clusters	centroid for AP_EL	centroid for PC_FU_EL	D - HDD _EL clusters	centroid for D_EL	centroid for HDD_EL
1995	1	0.162	0.101	1	0.160	0.104	3	0.164	0.167	3	12,002.500	1,793.168
1996	1	0.162	0.101	1	0.160	0.104	3	0.164	0.167	3	12,002.500	1,793.168
1997	1	0.162	0.101	1	0.160	0.104	3	0.164	0.167	3	12,002.500	1,793.168
1998	1	0.162	0.101	1	0.160	0.104	3	0.164	0.167	1	14,734.143	1,649.234
1999	1	0.162	0.101	3	0.175	0.107	3	0.164	0.167	1	14,734.143	1,649.234
2000	3	0.180	0.115	3	0.175	0.107	3	0.164	0.167	1	14,734.143	1,649.234
2001	3	0.180	0.115	3	0.175	0.107	4	0.183	0.146	1	14,734.143	1,649.234
2002	3	0.180	0.115	3	0.175	0.107	4	0.183	0.146	1	14,734.143	1,649.234
2003	3	0.180	0.115	3	0.175	0.107	4	0.183	0.146	3	12,002.500	1,793.168
2004	3	0.180	0.115	3	0.175	0.107	4	0.183	0.146	1	14,734.143	1,649.234
2005	3	0.180	0.115	2	0.186	0.090	4	0.183	0.146	1	14,734.143	1,649.234
2006	3	0.180	0.115	2	0.186	0.090	4	0.183	0.146	5	21,020.333	1,749.340
2007	3	0.180	0.115	2	0.186	0.090	4	0.183	0.146	2	22,388.167	1,478.378
2008	2	0.188	0.141	2	0.186	0.090	4	0.183	0.146	2	22,388.167	1,478.378
2009	2	0.188	0.141	2	0.186	0.090	4	0.183	0.146	2	22,388.167	1,478.378
2010	2	0.188	0.141	5	0.191	0.153	4	0.183	0.146	2	22,388.167	1,478.378
2011	4	0.201	0.172	5	0.191	0.153	2	0.193	0.215	5	21,020.333	1,749.340
2012	4	0.201	0.172	4	0.203	0.259	1	0.203	0.297	5	21,020.333	1,749.340
2013	4	0.201	0.172	4	0.203	0.259	1	0.203	0.297	2	22,388.167	1,478.378
2014	4	0.201	0.172	4	0.203	0.259	1	0.203	0.297	2	22,388.167	1,478.378
2015	4	0.201	0.172	4	0.203	0.259	1	0.203	0.297	4	31,561.000	1,577.850

A228: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Greece

Years	AP - Egf_ME_EL clusters	centroid for AP_EL	centroid for Egf_ME_EL	AP - H_ME_EL clusters	centroid for AP_EL	centroid for H_ME_EL	AP - PC_P_EL clusters	centroid for AP_EL	centroid for PC_P_EL
1995	1	0.166	1,807.543	4	0.158	3,373.033	1	0.162	0.101
1996	1	0.166	1,807.543	4	0.158	3,373.033	1	0.162	0.101
1997	1	0.166	1,807.543	4	0.158	3,373.033	1	0.162	0.101
1998	1	0.166	1,807.543	2	0.172	4,453.160	1	0.162	0.101
1999	1	0.166	1,807.543	2	0.172	4,453.160	1	0.162	0.101
2000	1	0.166	1,807.543	2	0.172	4,453.160	3	0.180	0.115
2001	1	0.166	1,807.543	2	0.172	4,453.160	3	0.180	0.115
2002	3	0.181	3,558.300	2	0.172	4,453.160	3	0.180	0.115
2003	3	0.181	3,558.300	3	0.185	7,265.925	3	0.180	0.115
2004	3	0.181	3,558.300	3	0.185	7,265.925	3	0.180	0.115
2005	3	0.181	3,558.300	3	0.185	7,265.925	3	0.180	0.115
2006	3	0.181	3,558.300	3	0.185	7,265.925	3	0.180	0.115
2007	3	0.181	3,558.300	3	0.185	7,265.925	3	0.180	0.115
2008	2	0.191	5,700.060	3	0.185	7,265.925	2	0.188	0.141
2009	2	0.191	5,700.060	3	0.185	7,265.925	2	0.188	0.141
2010	2	0.191	5,700.060	3	0.185	7,265.925	2	0.188	0.141
2011	2	0.191	5,700.060	1	0.201	5,272.020	4	0.201	0.172
2012	2	0.191	5,700.060	1	0.201	5,272.020	4	0.201	0.172
2013	4	0.205	3,601.263	1	0.201	5,272.020	4	0.201	0.172
2014	4	0.205	3,601.263	1	0.201	5,272.020	4	0.201	0.172
2015	4	0.205	3,601.263	1	0.201	5,272.020	4	0.201	0.172

A229: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Spain

Years	AP - PS_R _ ES clusters	centroid for AP_ES	centroid for PS_R_ES	AP - PC_TU_ES clusters	centroid for AP_ES	centroid for PC_TU_ES	AP - PC_FU_ ES clusters	centroid for AP_ES	centroid for PC_FU_ES	D - HDD _ ES clusters	centroid for D_ES	centroid for HDD_ES
1995	1	0.153	4.465	4	0.155	0.189	1	0.155	0.258	5	77,231.000	1,567.725
1996	1	0.153	4.465	4	0.155	0.189	1	0.155	0.258	3	86,436.000	1,870.538
1997	1	0.153	4.465	4	0.155	0.189	1	0.155	0.258	5	77,231.000	1,567.725
1998	3	0.164	4.184	4	0.155	0.189	1	0.155	0.258	1	99,461.286	1,876.921
1999	3	0.164	4.184	3	0.167	0.107	2	0.168	0.185	1	99,461.286	1,876.921
2000	3	0.164	4.184	3	0.167	0.107	2	0.168	0.185	3	86,436.000	1,870.538
2001	2	0.169	4.046	3	0.167	0.107	3	0.166	0.128	3	86,436.000	1,870.538
2002	2	0.169	4.046	3	0.167	0.107	2	0.168	0.185	4	93,192.200	1,656.320
2003	2	0.169	4.046	3	0.167	0.107	2	0.168	0.185	1	99,461.286	1,876.921
2004	2	0.169	4.046	3	0.167	0.107	3	0.166	0.128	3	86,436.000	1,870.538
2005	3	0.164	4.184	3	0.167	0.107	3	0.166	0.128	1	99,461.286	1,876.921
2006	3	0.164	4.184	3	0.167	0.107	3	0.166	0.128	4	93,192.200	1,656.320
2007	3	0.164	4.184	3	0.167	0.107	3	0.166	0.128	1	99,461.286	1,876.921
2008	3	0.164	4.184	3	0.167	0.107	3	0.166	0.128	1	99,461.286	1,876.921
2009	3	0.164	4.184	1	0.168	0.197	2	0.168	0.185	4	93,192.200	1,656.320
2010	2	0.169	4.046	1	0.168	0.197	2	0.168	0.185	3	86,436.000	1,870.538
2011	2	0.169	4.046	1	0.168	0.197	2	0.168	0.185	4	93,192.200	1,656.320
2012	5	0.175	3.843	2	0.179	0.244	4	0.179	0.252	1	99,461.286	1,876.921
2013	5	0.175	3.843	2	0.179	0.244	4	0.179	0.252	1	99,461.286	1,876.921
2014	4	0.183	3.630	2	0.179	0.244	4	0.179	0.252	4	93,192.200	1,656.320
2015	4	0.183	3.630	2	0.179	0.244	4	0.179	0.252	2	113,792.000	1,639.980

A230: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Spain

Years	AP - Egf_ME_ ES clusters	centroid for AP_ES	centroid for Egf_ME_ES	AP - H_ME_ ES clusters	centroid for AP_ES	centroid for H_ME_ES	AP - PC_P_ ES clusters	centroid for AP_ES	centroid for PC_P_ES
1995	1	0.155	7,558.250	2	0.153	9,743.600	3	0.155	0.099
1996	1	0.155	7,558.250	2	0.153	9,743.600	3	0.155	0.099
1997	1	0.155	7,558.250	2	0.153	9,743.600	3	0.155	0.099
1998	1	0.155	7,558.250	4	0.163	11,881.700	3	0.155	0.099
1999	3	0.167	10,136.143	4	0.163	11,881.700	4	0.166	0.091
2000	3	0.167	10,136.143	4	0.163	11,881.700	4	0.166	0.091
2001	3	0.167	10,136.143	5	0.168	15,860.600	4	0.166	0.091
2002	3	0.167	10,136.143	5	0.168	15,860.600	4	0.166	0.091
2003	3	0.167	10,136.143	5	0.168	15,860.600	4	0.166	0.091
2004	3	0.167	10,136.143	5	0.168	15,860.600	4	0.166	0.091
2005	3	0.167	10,136.143	5	0.168	15,860.600	4	0.166	0.091
2006	2	0.165	16,411.000	3	0.167	21,754.603	4	0.166	0.091
2007	2	0.165	16,411.000	3	0.167	21,754.603	4	0.166	0.091
2008	2	0.165	16,411.000	3	0.167	21,754.603	4	0.166	0.091
2009	2	0.165	16,411.000	3	0.167	21,754.603	4	0.166	0.091
2010	5	0.171	23,848.000	3	0.167	21,754.603	2	0.171	0.111
2011	5	0.171	23,848.000	3	0.167	21,754.603	2	0.171	0.111
2012	5	0.171	23,848.000	1	0.179	25,301.000	2	0.171	0.111
2013	4	0.181	24,898.000	1	0.179	25,301.000	1	0.181	0.127
2014	4	0.181	24,898.000	1	0.179	25,301.000	1	0.181	0.127
2015	4	0.181	24,898.000	1	0.179	25,301.000	1	0.181	0.127

A231: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Netherlands

Years	AP - PS_R _ NL clusters	centroid for AP_NL	centroid for PS_R_NL	AP - PC_TU_NL clusters	centroid for AP_NL	centroid for PC_TU_NL	AP - PC_FU_NL clusters	centroid for AP_NL	centroid for PC_FU_NL	D - HDD _ NL clusters	centroid for D_NL	centroid for HDD_NL
1995	1	0.136	5.007	2	0.132	0.080	1	0.133	0.103	1	24,367.583	2,652.022
1996	1	0.136	5.007	2	0.132	0.080	1	0.133	0.103	3	27,268.473	3,257.033
1997	1	0.136	5.007	5	0.138	0.055	1	0.133	0.103	1	24,367.583	2,652.022
1998	1	0.136	5.007	5	0.138	0.055	3	0.137	0.058	1	24,367.583	2,652.022
1999	1	0.136	5.007	4	0.139	0.038	3	0.137	0.058	1	24,367.583	2,652.022
2000	1	0.136	5.007	4	0.139	0.038	3	0.137	0.058	1	24,367.583	2,652.022
2001	1	0.136	5.007	4	0.139	0.038	3	0.137	0.058	1	24,367.583	2,652.022
2002	1	0.136	5.007	4	0.139	0.038	3	0.137	0.058	1	24,367.583	2,652.022
2003	1	0.136	5.007	5	0.138	0.055	3	0.137	0.058	1	24,367.583	2,652.022
2004	1	0.136	5.007	5	0.138	0.055	3	0.137	0.058	1	24,367.583	2,652.022
2005	1	0.136	5.007	5	0.138	0.055	3	0.137	0.058	1	24,367.583	2,652.022
2006	3	0.149	4.521	5	0.138	0.055	3	0.137	0.058	1	24,367.583	2,652.022
2007	3	0.149	4.521	4	0.139	0.038	2	0.150	0.051	2	31,001.500	2,634.457
2008	3	0.149	4.521	4	0.139	0.038	2	0.150	0.051	2	31,001.500	2,634.457
2009	3	0.149	4.521	3	0.155	0.050	2	0.150	0.051	2	31,001.500	2,634.457
2010	3	0.149	4.521	3	0.155	0.050	2	0.150	0.051	3	27,268.473	3,257.033
2011	3	0.149	4.521	3	0.155	0.050	2	0.150	0.051	2	31,001.500	2,634.457
2012	2	0.170	3.871	3	0.155	0.050	4	0.170	0.071	2	31,001.500	2,634.457
2013	2	0.170	3.871	1	0.173	0.072	4	0.170	0.071	3	27,268.473	3,257.033
2014	2	0.170	3.871	1	0.173	0.072	4	0.170	0.071	1	24,367.583	2,652.022
2015	2	0.170	3.871	1	0.173	0.072	4	0.170	0.071	2	31,001.500	2,634.457

A232: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Netherlands

Years	AP - Egf_ME_ NL clusters	centroid for AP_NL	centroid for Egf_ME_NL	AP - H_ME_ NL clusters	centroid for AP_NL	centroid for H_ME_NL	AP - PC_P_NL clusters	centroid for AP_NL	centroid for PC_P_NL
1995	1	0.135	6,683.360	1	0.134	8,328.771	1	0.133	0.130
1996	1	0.135	6,683.360	1	0.134	8,328.771	1	0.133	0.130
1997	1	0.135	6,683.360	1	0.134	8,328.771	1	0.133	0.130
1998	1	0.135	6,683.360	1	0.134	8,328.771	3	0.139	0.116
1999	1	0.135	6,683.360	1	0.134	8,328.771	3	0.139	0.116
2000	1	0.135	6,683.360	1	0.134	8,328.771	3	0.139	0.116
2001	1	0.135	6,683.360	1	0.134	8,328.771	3	0.139	0.116
2002	1	0.135	6,683.360	2	0.138	11,950.000	3	0.139	0.116
2003	1	0.135	6,683.360	2	0.138	11,950.000	3	0.139	0.116
2004	1	0.135	6,683.360	2	0.138	11,950.000	3	0.139	0.116
2005	3	0.148	10,355.950	2	0.138	11,950.000	3	0.139	0.116
2006	3	0.148	10,355.950	4	0.149	7,929.167	3	0.139	0.116
2007	3	0.148	10,355.950	4	0.149	7,929.167	3	0.139	0.116
2008	3	0.148	10,355.950	4	0.149	7,929.167	3	0.139	0.116
2009	3	0.148	10,355.950	4	0.149	7,929.167	2	0.153	0.123
2010	3	0.148	10,355.950	4	0.149	7,929.167	2	0.153	0.123
2011	3	0.148	10,355.950	4	0.149	7,929.167	2	0.153	0.123
2012	2	0.170	11,113.500	3	0.170	10,113.750	4	0.170	0.130
2013	2	0.170	11,113.500	3	0.170	10,113.750	4	0.170	0.130
2014	2	0.170	11,113.500	3	0.170	10,113.750	4	0.170	0.130
2015	2	0.170	11,113.500	3	0.170	10,113.750	4	0.170	0.130

A233: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Austria

Years	AP - PS_R _ AT clusters	centroid for AP_AT	centroid for PS_R_AT	AP - PC_TU_AT clusters	centroid for AP_AT	centroid for PC_TU_AT	AP - PC_FU_AT clusters	centroid for AP_AT	centroid for PC_FU_AT	D - HDD _ AT clusters	centroid for D_AT	centroid for HDD_AT
1995	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
1996	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
1997	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
1998	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
1999	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
2000	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	2	9,821.100	3,423.588
2001	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	1	8,355.833	3,782.585
2002	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	2	9,821.100	3,423.588
2003	3	0.154	4.382	1	0.153	0.044	3	0.153	0.045	3	10,825.000	3,760.024
2004	3	0.154	4.382	3	0.159	0.055	2	0.162	0.058	3	10,825.000	3,760.024
2005	3	0.154	4.382	3	0.159	0.055	2	0.162	0.058	3	10,825.000	3,760.024
2006	4	0.168	4.015	3	0.159	0.055	2	0.162	0.058	3	10,825.000	3,760.024
2007	4	0.168	4.015	2	0.174	0.047	1	0.178	0.053	2	9,821.100	3,423.588
2008	4	0.168	4.015	2	0.174	0.047	4	0.175	0.046	2	9,821.100	3,423.588
2009	2	0.176	3.835	4	0.181	0.055	1	0.178	0.053	2	9,821.100	3,423.588
2010	2	0.176	3.835	2	0.174	0.047	4	0.175	0.046	3	10,825.000	3,760.024
2011	2	0.176	3.835	2	0.174	0.047	4	0.175	0.046	2	9,821.100	3,423.588
2012	2	0.176	3.835	2	0.174	0.047	4	0.175	0.046	2	9,821.100	3,423.588
2013	1	0.183	3.688	4	0.181	0.055	1	0.178	0.053	2	9,821.100	3,423.588
2014	1	0.183	3.688	4	0.181	0.055	1	0.178	0.053	2	9,821.100	3,423.588
2015	1	0.183	3.688	4	0.181	0.055	1	0.178	0.053	2	9,821.100	3,423.588

A234: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Austria

Years	AP - Egf_ME_ AT clusters	centroid for AP_AT	centroid for Egf_ME_AT	AP - H_ME_AT clusters	centroid for AP_AT	centroid for H_ME_AT	AP - PC_P_ AT clusters	centroid for AP_AT	centroid for PC_P_AT
1995	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
1996	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
1997	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
1998	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
1999	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
2000	1	0.154	4,575.680	1	0.153	3,894.200	1	0.154	0.140
2001	1	0.154	4,575.680	3	0.157	5,002.883	1	0.154	0.140
2002	1	0.154	4,575.680	3	0.157	5,002.883	1	0.154	0.140
2003	1	0.154	4,575.680	3	0.157	5,002.883	1	0.154	0.140
2004	1	0.154	4,575.680	3	0.157	5,002.883	1	0.154	0.140
2005	3	0.164	6,006.200	3	0.157	5,002.883	3	0.166	0.134
2006	3	0.164	6,006.200	3	0.157	5,002.883	3	0.166	0.134
2007	3	0.164	6,006.200	2	0.173	5,816.840	3	0.166	0.134
2008	2	0.178	6,849.425	2	0.173	5,816.840	3	0.166	0.134
2009	2	0.178	6,849.425	2	0.173	5,816.840	2	0.179	0.145
2010	2	0.178	6,849.425	2	0.173	5,816.840	2	0.179	0.145
2011	2	0.178	6,849.425	2	0.173	5,816.840	2	0.179	0.145
2012	2	0.178	6,849.425	4	0.182	6,734.725	2	0.179	0.145
2013	2	0.178	6,849.425	4	0.182	6,734.725	2	0.179	0.145
2014	2	0.178	6,849.425	4	0.182	6,734.725	2	0.179	0.145
2015	2	0.178	6,849.425	4	0.182	6,734.725	2	0.179	0.145

A235: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in Portugal

Years	AP - PS_R _ PT clusters	centroid for AP_PT	centroid for PS_R_PT	AP - PC_TU_ PT clusters	centroid for AP_PT	centroid for PC_TU_PT	AP - PC_FU_ PT clusters	centroid for AP_PT	centroid for PC_FU_PT	D - HDD _ PT clusters	centroid for D_PT	centroid for HDD_PT
1995	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	1	17,275.000	911.760
1996	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	3	19,022.428	1,232.054
1997	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	1	17,275.000	911.760
1998	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	3	19,022.428	1,232.054
1999	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	2	25,121.167	1,230.133
2000	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	3	19,022.428	1,232.054
2001	1	0.156	4.335	1	0.157	0.064	1	0.158	0.067	3	19,022.428	1,232.054
2002	3	0.174	3.853	1	0.157	0.064	1	0.158	0.067	3	19,022.428	1,232.054
2003	3	0.174	3.853	3	0.173	0.088	1	0.158	0.067	3	19,022.428	1,232.054
2004	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	3	19,022.428	1,232.054
2005	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	2	25,121.167	1,230.133
2006	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	2	25,121.167	1,230.133
2007	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	2	25,121.167	1,230.133
2008	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	2	25,121.167	1,230.133
2009	3	0.174	3.853	3	0.173	0.088	3	0.175	0.091	2	25,121.167	1,230.133
2010	3	0.174	3.853	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133
2011	2	0.195	3.386	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133
2012	2	0.195	3.386	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133
2013	2	0.195	3.386	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133
2014	2	0.195	3.386	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133
2015	2	0.195	3.386	2	0.193	0.140	2	0.193	0.142	2	25,121.167	1,230.133

A236: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in Portugal

Years	AP - Egf_ME_ PT clusters	centroid for AP_PT	centroid for Egf_ME_PT	AP - H_ME_PT clusters	centroid for AP_PT	centroid for H_ME_PT	AP - PC_P_ PT clusters	centroid for AP_PT	centroid for PC_P_PT
1995	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
1996	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
1997	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
1998	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
1999	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
2000	1	0.157	1,968.062	1	0.154	3,073.183	1	0.156	0.097
2001	1	0.157	1,968.062	3	0.169	4,367.183	1	0.156	0.097
2002	1	0.157	1,968.062	3	0.169	4,367.183	3	0.171	0.118
2003	3	0.173	3,214.729	3	0.169	4,367.183	3	0.171	0.118
2004	3	0.173	3,214.729	3	0.169	4,367.183	3	0.171	0.118
2005	3	0.173	3,214.729	3	0.169	4,367.183	3	0.171	0.118
2006	3	0.173	3,214.729	3	0.169	4,367.183	3	0.171	0.118
2007	3	0.173	3,214.729	2	0.180	5,844.540	3	0.171	0.118
2008	3	0.173	3,214.729	2	0.180	5,844.540	3	0.171	0.118
2009	3	0.173	3,214.729	2	0.180	5,844.540	2	0.185	0.140
2010	2	0.193	4,176.083	2	0.180	5,844.540	2	0.185	0.140
2011	2	0.193	4,176.083	2	0.180	5,844.540	2	0.185	0.140
2012	2	0.193	4,176.083	4	0.196	5,842.600	2	0.185	0.140
2013	2	0.193	4,176.083	4	0.196	5,842.600	4	0.199	0.154
2014	2	0.193	4,176.083	4	0.196	5,842.600	4	0.199	0.154
2015	2	0.193	4,176.083	4	0.196	5,842.600	4	0.199	0.154

A237: Results from K-means clustering for AP – PS_R, AP – PC_TU, AP – PC_FU and D – HDD in United Kingdom

Years	AP - PS_R _ UK clusters	centroid for AP_UK	centroid for PS_R_UK	AP - PC_TU_ UK clusters	centroid for AP_UK	centroid for PC_TU_UK	AP - PC_FU_ UK clusters	centroid for AP_UK	centroid for PC_FU_UK	D - HDD _ UK clusters	centroid for D_UK	centroid for HDD_UK
1995	1	0.159	4.111	4	0.161	0.078	4	0.160	0.064	1	219,115.800	2,968.158
1996	1	0.159	4.111	4	0.161	0.078	4	0.160	0.064	4	188,523.500	3,409.165
1997	1	0.159	4.111	4	0.161	0.078	4	0.160	0.064	1	219,115.800	2,968.158
1998	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	1	219,115.800	2,968.158
1999	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	1	219,115.800	2,968.158
2000	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	1	219,115.800	2,968.158
2001	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	2	176,921.000	3,096.373
2002	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2003	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2004	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2005	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2006	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2007	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	3	172,602.500	2,866.421
2008	1	0.159	4.111	3	0.159	0.053	3	0.159	0.047	2	176,921.000	3,096.373
2009	1	0.159	4.111	4	0.161	0.078	4	0.160	0.064	2	176,921.000	3,096.373
2010	1	0.159	4.111	4	0.161	0.078	4	0.160	0.064	4	188,523.500	3,409.165
2011	1	0.159	4.111	4	0.161	0.078	2	0.168	0.073	3	172,602.500	2,866.421
2012	2	0.173	3.762	2	0.170	0.077	2	0.168	0.073	2	176,921.000	3,096.373
2013	2	0.173	3.762	2	0.170	0.077	2	0.168	0.073	2	176,921.000	3,096.373
2014	2	0.173	3.762	1	0.176	0.057	1	0.176	0.054	3	172,602.500	2,866.421
2015	2	0.173	3.762	1	0.176	0.057	1	0.176	0.054	2	176,921.000	3,096.373

A238: Results from K-means clustering for AP – Egf_ME, AP – H_ME and AP – PC_P in United Kingdom

Years	AP - Egf_ME_ UK clusters	centroid for AP_UK	centroid for Egf_ME_UK	AP - H_ME_ UK clusters	centroid for AP_UK	centroid for H_ME_UK	AP - PC_P_ UK clusters	centroid for AP_UK	centroid for PC_P_UK
1995	1	0.159	22,162.691	1	0.159	11,018.260	1	0.159	0.100
1996	1	0.159	22,162.691	1	0.159	11,018.260	1	0.159	0.100
1997	1	0.159	22,162.691	1	0.159	11,018.260	1	0.159	0.100
1998	1	0.159	22,162.691	1	0.159	11,018.260	1	0.159	0.100
1999	1	0.159	22,162.691	1	0.159	11,018.260	1	0.159	0.100
2000	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2001	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2002	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2003	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2004	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2005	1	0.159	22,162.691	3	0.160	17,792.108	1	0.159	0.100
2006	3	0.161	34,445.550	3	0.160	17,792.108	1	0.159	0.100
2007	3	0.161	34,445.550	3	0.160	17,792.108	1	0.159	0.100
2008	3	0.161	34,445.550	3	0.160	17,792.108	1	0.159	0.100
2009	3	0.161	34,445.550	3	0.160	17,792.108	3	0.164	0.113
2010	3	0.161	34,445.550	3	0.160	17,792.108	3	0.164	0.113
2011	3	0.161	34,445.550	3	0.160	17,792.108	3	0.164	0.113
2012	2	0.173	41,261.900	2	0.173	23,297.850	3	0.164	0.113
2013	2	0.173	41,261.900	2	0.173	23,297.850	2	0.175	0.114
2014	2	0.173	41,261.900	2	0.173	23,297.850	2	0.175	0.114
2015	2	0.173	41,261.900	2	0.173	23,297.850	2	0.175	0.114

APPENDIX B

B1: R script code for regression analysis

```
setwd("C:/Users/Leandros-Maria/Desktop/Msc BIS Analytics/c semester/R programming/wdir")
rm(list = ls())
library(lmtest)
AP<- read.csv("AP.csv")
PS_R<- read.csv("PS_R.CSV")
D<- read.csv("D.CSV")
HDD<- read.csv("HDD.csv")
PC_P<- read.csv("PC_P.csv")
PC_TU<- read.csv("PC_TU.csv")
PC_FU<- read.csv("PC_FU.csv")
Egf_ME<- read.csv("Egf_ME.csv")
H_ME<-read.csv("H_ME.csv")

DKvars<-data.frame(AP[5:25,1:2], PC_P[1:21,2], PS_R[5:25,2], PC_TU[1:21,2], PC_FU[1:21,2], Egf_ME[1:21,2],H_ME[1:21,2], D[1:21,2], HDD[1:21,2])
names(DKvars)<- c("TIME", "AP_DK", "PC_P_DK", "PS_R_DK", "PC_TU_DK", "PC_FU_DK", "Egf_ME_DK", "H_ME_DK", "D_DK", "HDD_DK")

DEvars<-data.frame(AP[5:25, c(1,3)], PC_P[1:21,3], PS_R[5:25,3], PC_TU[1:21,3], PC_FU[1:21,3], Egf_ME[1:21,3], H_ME[1:21,3], D[1:21,3], HDD[1:21,3])
names(DEvars)<- c("TIME", "AP_DE", "PC_P_DE", "PS_R_DE", "PC_TU_DE", "PC_FU_DE", "Egf_ME_DE", "H_ME_DE", "D_DE", "HDD_DE")

IEvars<-data.frame(AP[5:25, c(1,4)], PC_P[1:21,4], PS_R[5:25,4], PC_TU[1:21,4], PC_FU[1:21,4], Egf_ME[1:21,4], H_ME[1:21,4], D[1:21,4], HDD[1:21,4])
names(IEvars)<- c("TIME", "AP_IE", "PC_P_IE", "PS_R_IE", "PC_TU_IE", "PC_FU_IE", "Egf_ME_IE", "H_ME_IE", "D_IE", "HDD_IE")

ELvars<-data.frame(AP[5:25,c(1,5)], PC_P[1:21,5], PS_R[5:25,5], PC_TU[1:21,5], PC_FU[1:21,5], Egf_ME[1:21,5],H_ME[1:21,5], D[1:21,5], HDD[1:21,5])
names(ELvars)<- c("TIME", "AP_EL", "PC_P_EL", "PS_R_EL", "PC_TU_EL", "PC_FU_EL", "Egf_ME_EL", "H_ME_EL", "D_EL", "HDD_EL")

ESvars<-data.frame(AP[5:25,c(1,6)], PC_P[1:21,6], PS_R[5:25,6], PC_TU[1:21,6], PC_FU[1:21,6], Egf_ME[1:21,6],H_ME[1:21,6], D[1:21,6], HDD[1:21,6])
names(ESvars)<- c("TIME", "AP_ES", "PC_P_ES", "PS_R_ES", "PC_TU_ES", "PC_FU_ES", "Egf_ME_ES", "H_ME_ES", "D_ES", "HDD_ES")

NLvars<-data.frame(AP[5:25,c(1,7)], PC_P[1:21,7], PS_R[5:25,7], PC_TU[1:21,7], PC_FU[1:21,7], Egf_ME[1:21,7],H_ME[1:21,7], D[1:21,7], HDD[1:21,7])
names(NLvars)<- c("TIME", "AP_NL", "PC_P_NL", "PS_R_NL", "PC_TU_NL", "PC_FU_NL", "Egf_ME_NL", "H_ME_NL", "D_NL", "HDD_NL")

ATvars<-data.frame(AP[5:25,c(1,8)], PC_P[1:21,8], PS_R[5:25,8], PC_TU[1:21,8], PC_FU[1:21,8], Egf_ME[1:21,8],H_ME[1:21,8], D[1:21,8], HDD[1:21,8])
names(ATvars)<- c("TIME", "AP_AT", "PC_P_AT", "PS_R_AT", "PC_TU_AT", "PC_FU_AT", "Egf_ME_AT", "H_ME_AT", "D_AT", "HDD_AT")

PTvars<-data.frame(AP[5:25,c(1,9)], PC_P[1:21,9], PS_R[5:25,9], PC_TU[1:21,9], PC_FU[1:21,9], Egf_ME[1:21,9],H_ME[1:21,9], D[1:21,9], HDD[1:21,9])
names(PTvars)<- c("TIME", "AP_PT", "PC_P_PT", "PS_R_PT", "PC_TU_PT", "PC_FU_PT", "Egf_ME_PT", "H_ME_PT", "D_PT", "HDD_PT")

UKvars<-data.frame(AP[5:25,c(1,10)], PC_P[1:21,10], PS_R[5:25,10], PC_TU[1:21,10], PC_FU[1:21,10], Egf_ME[1:21,10],H_ME[1:21,10], D[1:21,10], HDD[1:21,10])
names(UKvars)<- c("TIME", "AP_UK", "PC_P_UK", "PS_R_UK", "PC_TU_UK", "PC_FU_UK", "Egf_ME_UK", "H_ME_UK", "D_UK", "HDD_UK")
```

```
AP_PSR_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PS_R_DK"))
cor(AP_PSR_DK)
ap_psr_dk_glm <- glm(AP_DK ~ PS_R_DK , data = AP_PSR_DK)
summary(ap_psr_dk_glm)
dwtest(ap_psr_dk_glm)
plot(ap_psr_dk_glm)

AP_PSR_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PS_R_DE"))
cor(AP_PSR_DE)
ap_psr_de_glm <- glm(AP_DE ~ PS_R_DE , data = AP_PSR_DE)
summary(ap_psr_de_glm)
dwtest(ap_psr_de_glm)
plot(ap_psr_de_glm)

AP_PSR_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PS_R_IE"))
cor(AP_PSR_IE)
ap_psr_ie_glm <- glm(AP_IE ~ PS_R_IE , data = AP_PSR_IE)
summary(ap_psr_ie_glm)
dwtest(ap_psr_ie_glm)
plot(ap_psr_ie_glm)

AP_PSR_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PS_R_EL"))
cor(AP_PSR_EL)
ap_psr_el_glm <- glm(AP_EL ~ PS_R_EL , data = AP_PSR_EL)
summary(ap_psr_el_glm)
dwtest(ap_psr_el_glm)
plot(ap_psr_el_glm)

AP_PSR_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PS_R_ES"))
cor(AP_PSR_ES)
ap_psr_es_glm <- glm(AP_ES ~ PS_R_ES , data = AP_PSR_ES)
summary(ap_psr_es_glm)
dwtest(ap_psr_es_glm)
plot(ap_psr_es_glm)

AP_PSR_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PS_R_NL"))
cor(AP_PSR_NL)
ap_psr_nl_glm <- glm(AP_NL ~ PS_R_NL , data = AP_PSR_NL)
summary(ap_psr_nl_glm)
dwtest(ap_psr_nl_glm)
plot(ap_psr_nl_glm)

AP_PSR_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PS_R_AT"))
cor(AP_PSR_AT)
ap_psr_at_glm <- glm(AP_AT ~ PS_R_AT , data = AP_PSR_AT)
summary(ap_psr_at_glm)
dwtest(ap_psr_at_glm)
plot(ap_psr_at_glm)

AP_PSR_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PS_R_PT"))
cor(AP_PSR_PT)
ap_psr_pt_glm <- glm(AP_PT ~ PS_R_PT , data = AP_PSR_PT)
summary(ap_psr_pt_glm)
dwtest(ap_psr_pt_glm)
plot(ap_psr_pt_glm)

AP_PSR_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PS_R_UK"))
cor(AP_PSR_UK)
ap_psr_uk_glm <- glm(AP_UK ~ PS_R_UK , data = AP_PSR_UK)
```

```
summary(ap_psr_uk_glm)
dwtest(ap_psr_uk_glm)
plot(ap_psr_uk_glm)

D_HDD_DK<- subset.data.frame(DKvars, select = c("D_DK", "HDD_DK"))
cor(D_HDD_DK)
d_hdd_dk_glm<-glm(D_DK ~ HDD_DK, data = D_HDD_DK)
summary(d_hdd_dk_glm)
dwtest(d_hdd_dk_glm)
plot(d_hdd_dk_glm)

D_HDD_DE<- subset.data.frame(DEvars, select = c("D_DE", "HDD_DE"))
cor(D_HDD_DE)
d_hdd_de_glm<-glm(D_DE ~ HDD_DE, data = D_HDD_DE)
summary(d_hdd_de_glm)
dwtest(d_hdd_de_glm)
plot(d_hdd_de_glm)

D_HDD_IE<- subset.data.frame(IEvars, select = c("D_IE", "HDD_IE"))
cor(D_HDD_IE)
d_hdd_ie_glm<-glm(D_IE ~ HDD_IE, data = D_HDD_IE)
summary(d_hdd_ie_glm)
dwtest(d_hdd_ie_glm)
plot(d_hdd_ie_glm)

D_HDD_EL<- subset.data.frame(ELvars, select = c("D_EL", "HDD_EL"))
cor(D_HDD_EL)
d_hdd_el_glm<-glm(D_EL ~ HDD_EL, data = D_HDD_EL)
summary(d_hdd_el_glm)
dwtest(d_hdd_el_glm)
plot(d_hdd_el_glm)

D_HDD_ES<- subset.data.frame(ESvars, select = c("D_ES", "HDD_ES"))
cor(D_HDD_ES)
d_hdd_es_glm<-glm(D_ES ~ HDD_ES, data = D_HDD_ES)
summary(d_hdd_es_glm)
dwtest(d_hdd_es_glm)
plot(d_hdd_es_glm)

D_HDD_NL<- subset.data.frame(NLvars, select = c("D_NL", "HDD_NL"))
cor(D_HDD_NL)
d_hdd_nl_glm<-glm(D_NL ~ HDD_NL, data = D_HDD_NL)
summary(d_hdd_nl_glm)
dwtest(d_hdd_nl_glm)
plot(d_hdd_nl_glm)

D_HDD_AT<- subset.data.frame(ATvars, select = c("D_AT", "HDD_AT"))
cor(D_HDD_AT)
d_hdd_at_glm<-glm(D_AT ~ HDD_AT, data = D_HDD_AT)
summary(d_hdd_at_glm)
dwtest(d_hdd_at_glm)
plot(d_hdd_at_glm)

D_HDD_PT<- subset.data.frame(PTvars, select = c("D_PT", "HDD_PT"))
cor(D_HDD_PT)
d_hdd_pt_glm<-glm(D_PT ~ HDD_PT, data = D_HDD_PT)
summary(d_hdd_pt_glm)
dwtest(d_hdd_pt_glm)
plot(d_hdd_pt_glm)
```

```
D_HDD_UK<- subset.data.frame(UKvars, select = c("D_UK", "HDD_UK"))
cor(D_HDD_UK)
d_hdd_uk_glm<-glm(D_UK ~ HDD_UK, data = D_HDD_UK)
summary(d_hdd_uk_glm)
dwtest(d_hdd_uk_glm)
plot(d_hdd_uk_glm)

AP_PCP_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_P_DK"))
cor(AP_PCP_DK)
ap_pcp_dk_glm<-glm(PC_P_DK ~ AP_DK, data = AP_PCP_DK)
summary(ap_pcp_dk_glm)
dwtest(ap_pcp_dk_glm)
plot(ap_pcp_dk_glm)

AP_PCP_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_P_DE"))
cor(AP_PCP_DE)
ap_pcp_de_glm<-glm(PC_P_DE ~ AP_DE, data = AP_PCP_DE)
summary(ap_pcp_de_glm)
dwtest(ap_pcp_de_glm)
plot(ap_pcp_de_glm)

AP_PCP_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_P_IE"))
cor(AP_PCP_IE)
ap_pcp_ie_glm<-glm(PC_P_IE ~ AP_IE, data = AP_PCP_IE)
summary(ap_pcp_ie_glm)
dwtest(ap_pcp_ie_glm)
plot(ap_pcp_ie_glm)

AP_PCP_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_P_EL"))
cor(AP_PCP_EL)
ap_pcp_el_glm<-glm(PC_P_EL ~ AP_EL, data = AP_PCP_EL)
summary(ap_pcp_el_glm)
dwtest(ap_pcp_el_glm)
plot(ap_pcp_el_glm)

AP_PCP_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_P_ES"))
cor(AP_PCP_ES)
ap_pcp_es_glm<-glm(PC_P_ES ~ AP_ES, data = AP_PCP_ES)
summary(ap_pcp_es_glm)
dwtest(ap_pcp_es_glm)
plot(ap_pcp_es_glm)

AP_PCP_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_P_NL"))
cor(AP_PCP_NL)
ap_pcp_nl_glm<-glm(PC_P_NL ~ AP_NL, data = AP_PCP_NL)
summary(ap_pcp_nl_glm)
dwtest(ap_pcp_nl_glm)
plot(ap_pcp_nl_glm)

AP_PCP_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_P_AT"))
cor(AP_PCP_AT)
ap_pcp_at_glm<-glm(PC_P_AT ~ AP_AT, data = AP_PCP_AT)
summary(ap_pcp_at_glm)
dwtest(ap_pcp_at_glm)
plot(ap_pcp_at_glm)

AP_PCP_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_P_PT"))
cor(AP_PCP_PT)
ap_pcp_pt_glm<-glm(PC_P_PT ~ AP_PT, data = AP_PCP_PT)
summary(ap_pcp_pt_glm)
```

```
dwtest(ap_pcp_pt_glm)
plot(ap_pcp_pt_glm)

AP_PCP_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_P_UK"))
cor(AP_PCP_UK)
ap_pcp_uk_glm<-glm(PC_P_UK ~ AP_UK, data = AP_PCP_UK)
summary(ap_pcp_uk_glm)
dwtest(ap_pcp_uk_glm)
plot(ap_pcp_uk_glm)

AP_PCTU_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_TU_DK"))
cor(AP_PCTU_DK)
ap_pctu_dk_glm<-glm( AP_DK ~ PC_TU_DK, data = AP_PCTU_DK)
summary(ap_pctu_dk_glm)
dwtest(ap_pctu_dk_glm)
plot(ap_pctu_dk_glm)

AP_PCTU_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_TU_DE"))
cor(AP_PCTU_DE)
ap_pctu_de_glm<-glm( AP_DE ~ PC_TU_DE, data = AP_PCTU_DE)
summary(ap_pctu_de_glm)
dwtest(ap_pctu_de_glm)
plot(ap_pctu_de_glm)

AP_PCTU_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_TU_IE"))
cor(AP_PCTU_IE)
ap_pctu_ie_glm<-glm( AP_IE ~ PC_TU_IE, data = AP_PCTU_IE)
summary(ap_pctu_ie_glm)
dwtest(ap_pctu_ie_glm)
plot(ap_pctu_ie_glm)

AP_PCTU_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_TU_EL"))
cor(AP_PCTU_EL)
ap_pctu_el_glm<-glm( AP_EL ~ PC_TU_EL, data = AP_PCTU_EL)
summary(ap_pctu_el_glm)
dwtest(ap_pctu_el_glm)
plot(ap_pctu_el_glm)

AP_PCTU_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_TU_ES"))
cor(AP_PCTU_ES)
ap_pctu_es_glm<-glm( AP_ES ~ PC_TU_ES, data = AP_PCTU_ES)
summary(ap_pctu_es_glm)
dwtest(ap_pctu_es_glm)
plot(ap_pctu_es_glm)

AP_PCTU_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_TU_NL"))
cor(AP_PCTU_NL)
ap_pctu_nl_glm<-glm( AP_NL ~ PC_TU_NL, data = AP_PCTU_NL)
summary(ap_pctu_nl_glm)
dwtest(ap_pctu_nl_glm)
plot(ap_pctu_nl_glm)

AP_PCTU_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_TU_AT"))
cor(AP_PCTU_AT)
ap_pctu_at_glm<-glm( AP_AT ~ PC_TU_AT, data = AP_PCTU_AT)
summary(ap_pctu_at_glm)
dwtest(ap_pctu_at_glm)
plot(ap_pctu_at_glm)

AP_PCTU_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_TU_PT"))
```

```
cor(AP_PCTU_PT)
ap_pctu_pt_glm<-glm( AP_PT ~ PC_TU_PT, data = AP_PCTU_PT)
summary(ap_pctu_pt_glm)
dwtest(ap_pctu_pt_glm)
plot(ap_pctu_pt_glm)

AP_PCTU_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_TU_UK"))
cor(AP_PCTU_UK)
ap_pctu_uk_glm<-glm( AP_UK ~ PC_TU_UK, data = AP_PCTU_UK)
summary(ap_pctu_uk_glm)
dwtest(ap_pctu_uk_glm)
plot(ap_pctu_uk_glm)

AP_PCFU_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_FU_DK"))
cor(AP_PCFU_DK)
ap_pcfu_dk_glm<-glm( AP_DK ~ PC_FU_DK, data = AP_PCFU_DK)
summary(ap_pcfu_dk_glm)
dwtest(ap_pcfu_dk_glm)
plot(ap_pcfu_dk_glm)

AP_PCFU_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_FU_DE"))
cor(AP_PCFU_DE)
ap_pcfu_de_glm<-glm( AP_DE ~ PC_FU_DE, data = AP_PCFU_DE)
summary(ap_pcfu_de_glm)
dwtest(ap_pcfu_de_glm)
plot(ap_pcfu_de_glm)

AP_PCFU_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_FU_IE"))
cor(AP_PCFU_IE)
ap_pcfu_ie_glm<-glm( AP_IE ~ PC_FU_IE, data = AP_PCFU_IE)
summary(ap_pcfu_ie_glm)
dwtest(ap_pcfu_ie_glm)
plot(ap_pcfu_ie_glm)

AP_PCFU_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_FU_EL"))
cor(AP_PCFU_EL)
ap_pcfu_el_glm<-glm( AP_EL ~ PC_FU_EL, data = AP_PCFU_EL)
summary(ap_pcfu_el_glm)
dwtest(ap_pcfu_el_glm)
plot(ap_pcfu_el_glm)

AP_PCFU_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_FU_ES"))
cor(AP_PCFU_ES)
ap_pcfu_es_glm<-glm( AP_ES ~ PC_FU_ES, data = AP_PCFU_ES)
summary(ap_pcfu_es_glm)
dwtest(ap_pcfu_es_glm)
plot(ap_pcfu_es_glm)

AP_PCFU_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_FU_NL"))
cor(AP_PCFU_NL)
ap_pcfu_nl_glm<-glm( AP_NL ~ PC_FU_NL, data = AP_PCFU_NL)
summary(ap_pcfu_nl_glm)
dwtest(ap_pcfu_nl_glm)
plot(ap_pcfu_nl_glm)

AP_PCFU_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_FU_AT"))
cor(AP_PCFU_AT)
ap_pcfu_at_glm<-glm( AP_AT ~ PC_FU_AT, data = AP_PCFU_AT)
summary(ap_pcfu_at_glm)
dwtest(ap_pcfu_at_glm)
```

```
plot(ap_pcfu_at_glm)

AP_PCFU_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_FU_PT"))
cor(AP_PCFU_PT)
ap_pcfu_pt_glm<-glm( AP_PT ~ PC_FU_PT, data = AP_PCFU_PT)
summary(ap_pcfu_pt_glm)
dwtest(ap_pcfu_pt_glm)
plot(ap_pcfu_pt_glm)

AP_PCFU_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_FU_UK"))
cor(AP_PCFU_UK)
ap_pcfu_uk_glm<-glm( AP_UK ~ PC_FU_UK, data = AP_PCFU_UK)
summary(ap_pcfu_uk_glm)
dwtest(ap_pcfu_uk_glm)
plot(ap_pcfu_uk_glm)

AP_EgfME_DK<-subset.data.frame(DKvars, select = c("AP_DK", "Egf_ME_DK"))
cor(AP_EgfME_DK)
ap_egfme_dk_glm<-glm(Egf_ME_DK ~ AP_DK, data = AP_EgfME_DK)
summary(ap_egfme_dk_glm)
dwtest(ap_egfme_dk_glm)
plot(ap_egfme_dk_glm)

AP_EgfME_DE<-subset.data.frame(DEvars, select = c("AP_DE", "Egf_ME_DE"))
cor(AP_EgfME_DE)
ap_egfme_de_glm<-glm(Egf_ME_DE ~ AP_DE, data = AP_EgfME_DE)
summary(ap_egfme_de_glm)
dwtest(ap_egfme_de_glm)
plot(ap_egfme_de_glm)

AP_EgfME_IE<-subset.data.frame(IEvars, select = c("AP_IE", "Egf_ME_IE"))
cor(AP_EgfME_IE)
ap_egfme_ie_glm<-glm(Egf_ME_IE ~ AP_IE, data = AP_EgfME_IE)
summary(ap_egfme_ie_glm)
dwtest(ap_egfme_ie_glm)
plot(ap_egfme_ie_glm)

AP_EgfME_EL<-subset.data.frame(ELvars, select = c("AP_EL", "Egf_ME_EL"))
cor(AP_EgfME_EL)
ap_egfme_el_glm<-glm(Egf_ME_EL ~ AP_EL, data = AP_EgfME_EL)
summary(ap_egfme_el_glm)
dwtest(ap_egfme_el_glm)
plot(ap_egfme_el_glm)

AP_EgfME_ES<-subset.data.frame(ESvars, select = c("AP_ES", "Egf_ME_ES"))
cor(AP_EgfME_ES)
ap_egfme_es_glm<-glm(Egf_ME_ES ~ AP_ES, data = AP_EgfME_ES)
summary(ap_egfme_es_glm)
dwtest(ap_egfme_es_glm)
plot(ap_egfme_es_glm)

AP_EgfME_NL<-subset.data.frame(NLvars, select = c("AP_NL", "Egf_ME_NL"))
cor(AP_EgfME_NL)
ap_egfme_nl_glm<-glm(Egf_ME_NL ~ AP_NL, data = AP_EgfME_NL)
summary(ap_egfme_nl_glm)
dwtest(ap_egfme_nl_glm)
plot(ap_egfme_nl_glm)

AP_EgfME_AT<-subset.data.frame(ATvars, select = c("AP_AT", "Egf_ME_AT"))
cor(AP_EgfME_AT)
```

```
ap_egfme_at_glm<-glm(Egf_ME_AT ~ AP_AT, data = AP_EgfME_AT)
summary(ap_egfme_at_glm)
dwtest(ap_egfme_at_glm)
plot(ap_egfme_at_glm)

AP_EgfME_PT<-subset.data.frame(PTvars, select = c("AP_PT", "Egf_ME_PT"))
cor(AP_EgfME_PT)
ap_egfme_pt_glm<-glm(Egf_ME_PT ~ AP_PT, data = AP_EgfME_PT)
summary(ap_egfme_pt_glm)
dwtest(ap_egfme_pt_glm)
plot(ap_egfme_pt_glm)

AP_EgfME_UK<-subset.data.frame(UKvars, select = c("AP_UK", "Egf_ME_UK"))
cor(AP_EgfME_UK)
ap_egfme_uk_glm<-glm(Egf_ME_UK ~ AP_UK, data = AP_EgfME_UK)
summary(ap_egfme_uk_glm)
dwtest(ap_egfme_uk_glm)
plot(ap_egfme_uk_glm)

AP_HME_DK<-subset.data.frame(DKvars, select = c("AP_DK", "H_ME_DK"))
cor(AP_HME_DK)
ap_hme_dk_glm<-glm(H_ME_DK ~ AP_DK, data = AP_HME_DK)
summary(ap_hme_dk_glm)
dwtest(ap_hme_dk_glm)
plot(ap_hme_dk_glm)

AP_HME_DE<-subset.data.frame(DEvars, select = c("AP_DE", "H_ME_DE"))
cor(AP_HME_DE)
ap_hme_de_glm<-glm(H_ME_DE ~ AP_DE, data = AP_HME_DE)
summary(ap_hme_de_glm)
dwtest(ap_hme_de_glm)
plot(ap_hme_de_glm)

AP_HME_IE<-subset.data.frame(IEvars, select = c("AP_IE", "H_ME_IE"))
cor(AP_HME_IE)
ap_hme_ie_glm<-glm(H_ME_IE ~ AP_IE, data = AP_HME_IE)
summary(ap_hme_ie_glm)
dwtest(ap_hme_ie_glm)
plot(ap_hme_ie_glm)

AP_HME_EL<-subset.data.frame(ELvars, select = c("AP_EL", "H_ME_EL"))
cor(AP_HME_EL)
ap_hme_el_glm<-glm(H_ME_EL ~ AP_EL, data = AP_HME_EL)
summary(ap_hme_el_glm)
dwtest(ap_hme_el_glm)
plot(ap_hme_el_glm)

AP_HME_ES<-subset.data.frame(ESvars, select = c("AP_ES", "H_ME_ES"))
cor(AP_HME_ES)
ap_hme_es_glm<-glm(H_ME_ES ~ AP_ES, data = AP_HME_ES)
summary(ap_hme_es_glm)
dwtest(ap_hme_es_glm)
plot(ap_hme_es_glm)

AP_HME_NL<-subset.data.frame(NLvars, select = c("AP_NL", "H_ME_NL"))
cor(AP_HME_NL)
ap_hme_nl_glm<-glm(H_ME_NL ~ AP_NL, data = AP_HME_NL)
summary(ap_hme_nl_glm)
dwtest(ap_hme_nl_glm)
plot(ap_hme_nl_glm)
```

```
AP_HME_AT<-subset.data.frame(ATvars, select = c("AP_AT", "H_ME_AT"))
cor(AP_HME_AT)
ap_hme_at_glm<-glm(H_ME_AT ~ AP_AT, data = AP_HME_AT)
summary(ap_hme_at_glm)
dwtest(ap_hme_at_glm)
plot(ap_hme_at_glm)

AP_HME_PT<-subset.data.frame(PTvars, select = c("AP_PT", "H_ME_PT"))
cor(AP_HME_PT)
ap_hme_pt_glm<-glm(H_ME_PT ~ AP_PT, data = AP_HME_PT)
summary(ap_hme_pt_glm)
dwtest(ap_hme_pt_glm)
plot(ap_hme_pt_glm)

AP_HME_UK<-subset.data.frame(UKvars, select = c("AP_UK", "H_ME_UK"))
cor(AP_HME_UK)
ap_hme_uk_glm<-glm(H_ME_UK ~ AP_UK, data = AP_HME_UK)
summary(ap_hme_uk_glm)
dwtest(ap_hme_uk_glm)
plot(ap_hme_uk_glm)
```

B2: R script code for cluster analysis

```
setwd("C:/Users/Leandros-Maria/Desktop/Msc BIS Analytics/c semester/R programming/wdir")
rm(list = ls())
AP<- read.csv("AP.csv")
PS_R<- read.csv("PS_R.CSV")
D<- read.csv("D.CSV")
HDD<- read.csv("HDD.csv")
PC_P<- read.csv("PC_P.csv")
PC_TU<- read.csv("PC_TU.csv")
PC_FU<- read.csv("PC_FU.csv")
Egf_ME<- read.csv("Egf_ME.csv")
H_ME<-read.csv("H_ME.csv")

DKvars<-data.frame(AP[5:25,1:2], PC_P[1:21,2], PS_R[5:25,2], PC_TU[1:21,2], PC_FU[1:21,2], Egf_ME[1:21,2],H_ME[1:21,2], D[1:21,2], HDD[1:21,2])
names(DKvars)<- c("TIME", "AP_DK", "PC_P_DK", "PS_R_DK", "PC_TU_DK", "PC_FU_DK", "Egf_ME_DK", "H_ME_DK", "D_DK", "HDD_DK")

DEvars<-data.frame(AP[5:25, c(1,3)], PC_P[1:21,3], PS_R[5:25,3], PC_TU[1:21,3], PC_FU[1:21,3], Egf_ME[1:21,3], H_ME[1:21,3], D[1:21,3], HDD[1:21,3])
names(DEvars)<- c("TIME", "AP_DE", "PC_P_DE", "PS_R_DE", "PC_TU_DE", "PC_FU_DE", "Egf_ME_DE", "H_ME_DE", "D_DE", "HDD_DE")

IEvars<-data.frame(AP[5:25, c(1,4)], PC_P[1:21,4], PS_R[5:25,4], PC_TU[1:21,4], PC_FU[1:21,4], Egf_ME[1:21,4], H_ME[1:21,4], D[1:21,4], HDD[1:21,4])
names(IEvars)<- c("TIME", "AP_IE", "PC_P_IE", "PS_R_IE", "PC_TU_IE", "PC_FU_IE", "Egf_ME_IE", "H_ME_IE", "D_IE", "HDD_IE")

ELvars<-data.frame(AP[5:25,c(1,5)], PC_P[1:21,5], PS_R[5:25,5], PC_TU[1:21,5], PC_FU[1:21,5], Egf_ME[1:21,5],H_ME[1:21,5], D[1:21,5], HDD[1:21,5])
names(ELvars)<- c("TIME", "AP_EL", "PC_P_EL", "PS_R_EL", "PC_TU_EL", "PC_FU_EL", "Egf_ME_EL", "H_ME_EL", "D_EL", "HDD_EL")

ESvars<-data.frame(AP[5:25,c(1,6)], PC_P[1:21,6], PS_R[5:25,6], PC_TU[1:21,6], PC_FU[1:21,6], Egf_ME[1:21,6],H_ME[1:21,6], D[1:21,6], HDD[1:21,6])
names(ESvars)<- c("TIME", "AP_ES", "PC_P_ES", "PS_R_ES", "PC_TU_ES", "PC_FU_ES", "Egf_ME_ES", "H_ME_ES", "D_ES", "HDD_ES")

NLvars<-data.frame(AP[5:25,c(1,7)], PC_P[1:21,7], PS_R[5:25,7], PC_TU[1:21,7], PC_FU[1:21,7], Egf_ME[1:21,7],H_ME[1:21,7], D[1:21,7], HDD[1:21,7])
names(NLvars)<- c("TIME", "AP_NL", "PC_P_NL", "PS_R_NL", "PC_TU_NL", "PC_FU_NL", "Egf_ME_NL", "H_ME_NL", "D_NL", "HDD_NL")

ATvars<-data.frame(AP[5:25,c(1,8)], PC_P[1:21,8], PS_R[5:25,8], PC_TU[1:21,8], PC_FU[1:21,8], Egf_ME[1:21,8],H_ME[1:21,8], D[1:21,8], HDD[1:21,8])
names(ATvars)<- c("TIME", "AP_AT", "PC_P_AT", "PS_R_AT", "PC_TU_AT", "PC_FU_AT", "Egf_ME_AT", "H_ME_AT", "D_AT", "HDD_AT")

PTvars<-data.frame(AP[5:25,c(1,9)], PC_P[1:21,9], PS_R[5:25,9], PC_TU[1:21,9], PC_FU[1:21,9], Egf_ME[1:21,9],H_ME[1:21,9], D[1:21,9], HDD[1:21,9])
names(PTvars)<- c("TIME", "AP_PT", "PC_P_PT", "PS_R_PT", "PC_TU_PT", "PC_FU_PT", "Egf_ME_PT", "H_ME_PT", "D_PT", "HDD_PT")

UKvars<-data.frame(AP[5:25,c(1,10)], PC_P[1:21,10], PS_R[5:25,10], PC_TU[1:21,10], PC_FU[1:21,10], Egf_ME[1:21,10],H_ME[1:21,10], D[1:21,10], HDD[1:21,10])
names(UKvars)<- c("TIME", "AP_UK", "PC_P_UK", "PS_R_UK", "PC_TU_UK", "PC_FU_UK", "Egf_ME_UK", "H_ME_UK", "D_UK", "HDD_UK")

set.seed(2)
AP_PSR_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PS_R_DK"))
```

```
ap_psr_dk_hcomp<-hclust(dist(scale(AP_PSR_DK)), method = "complete")
plot(ap_psr_dk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_DK", cex=.9, hang = -1)
ap_psr_dk_kmout<-kmeans(scale(AP_PSR_DK),3, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_DK, col = (ap_psr_dk_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_psr_dk_kmout$cluster) , col = unique(ap_psr_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_dk_kmout$cluster, DKvars$TIME)
ap_psr_dk_kmout$centers

set.seed(2)
D_HDD_DK<- subset.data.frame(DKvars, select = c("D_DK", "HDD_DK"))
d_hdd_dk_hcomp<-hclust(dist(scale(D_HDD_DK)), method = "complete")
plot(d_hdd_dk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_DK", cex=.9, hang = -1)
d_hdd_dk_kmout<-kmeans(scale(D_HDD_DK),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_DK, col = (d_hdd_dk_kmout$cluster+1),  main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_dk_kmout$cluster) , col = unique(d_hdd_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_dk_kmout$cluster, DKvars$TIME)
d_hdd_dk_kmout$centers

set.seed(2)
AP_PCP_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_P_DK"))
ap_pcp_dk_hcomp<-hclust(dist(scale(AP_PCP_DK)), method = "complete")
plot(ap_pcp_dk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_DK", cex=.9, hang = -1)
ap_pcp_dk_kmout<-kmeans(scale(AP_PCP_DK),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_DK, col = (ap_pcp_dk_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_dk_kmout$cluster) , col = unique(ap_pcp_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_dk_kmout$cluster, DKvars$TIME)
ap_pcp_dk_kmout$centers

set.seed(2)
AP_PCTU_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_TU_DK"))
ap_pctu_dk_hcomp<-hclust(dist(scale(AP_PCTU_DK)), method = "complete")
plot(ap_pctu_dk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_DK", cex=.9, hang = -1)
ap_pctu_dk_kmout<-kmeans(scale(AP_PCTU_DK),5, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_DK, col = (ap_pctu_dk_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_dk_kmout$cluster) , col = unique(ap_pctu_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_dk_kmout$cluster, DKvars$TIME)
ap_pctu_dk_kmout$centers

set.seed(2)
AP_PCFU_DK<-subset.data.frame(DKvars, select = c("AP_DK", "PC_FU_DK"))
ap_pcfu_dk_hcomp<-hclust(dist(scale(AP_PCFU_DK)), method="complete")
plot(ap_pcfu_dk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_DK", cex=.9, hang = -1)
ap_pcfu_dk_kmout<-kmeans(scale(AP_PCFU_DK),4,nstart=15)
```

```
par(mar=c(5,4,2,4))
plot(AP_PCFU_DK, col=(ap_pcfu_dk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_dk_kmout$cluster) , col = unique(ap_pcfu_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_dk_kmout$cluster,DKvars$TIME)
ap_pcfu_dk_kmout$centers

set.seed(2)
AP_EgfME_DK<-subset.data.frame(DKvars, select = c("AP_DK", "Egf_ME_DK"))
ap_egfme_dk_hcomp<-hclust(dist(scale(AP_EgfME_DK)), method = "complete")
plot(ap_egfme_dk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_DK", cex=.9, hang = -1)
ap_egfme_dk_kmout<-kmeans(scale(AP_EgfME_DK),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_DK, col= (ap_egfme_dk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_dk_kmout$cluster) , col = unique(ap_egfme_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_dk_kmout$cluster, DKvars$TIME)
ap_egfme_dk_kmout$centers

set.seed(2)
AP_HME_DK<-subset.data.frame(DKvars, select = c("AP_DK", "H_ME_DK"))
ap_hme_dk_hcomp<-hclust(dist(scale(AP_HME_DK)), method = "complete")
plot(ap_hme_dk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_DK", cex=.9, hang = -1)
ap_hme_dk_kmout<- kmeans(scale(AP_HME_DK),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_DK, col = (ap_hme_dk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_dk_kmout$cluster) , col = unique(ap_hme_dk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
ap_hme_dk_kmout$centers

set.seed(2)
DKvars_hcomp<-hclust(dist(scale(DKvars)), method = "complete")
plot(DKvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of DK", cex=.9, hang = -1)
DKvars_kmout<- kmeans(scale(DKvars), 4 , nstart = 20)
table(DKvars_kmout$cluster, DKvars$TIME)

set.seed(2)
AP_PSR_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PS_R_DE"))
ap_psr_de_hcomp<-hclust(dist(scale(AP_PSR_DE)), method = "complete")
plot(ap_psr_de_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_DE", cex=.9, hang = -1)
ap_psr_de_kmout<-kmeans(scale(AP_PSR_DE),4, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_DE, col = (ap_psr_de_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_de_kmout$cluster) , col = unique(ap_psr_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_de_kmout$cluster, DEvars$TIME)
ap_psr_de_kmout$centers

set.seed(2)
D_HDD_DE<- subset.data.frame(DEvars, select = c("D_DE", "HDD_DE"))
d_hdd_de_hcomp<-hclust(dist(scale(D_HDD_DE)), method = "complete")
```

```
plot(d_hdd_de_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_DE", cex=.9, hang = -1)
d_hdd_de_kmout<-kmeans(scale(D_HDD_DE),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_DE, col = (d_hdd_de_kmout$cluster+1),  main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_de_kmout$cluster) , col = unique(d_hdd_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_de_kmout$cluster, DEvars$TIME)
d_hdd_de_kmout$centers

set.seed(2)
AP_PCP_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_P_DE"))
ap_pcp_de_hcomp<-hclust(dist(scale(AP_PCP_DE)), method = "complete")
plot(ap_pcp_de_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_DE", cex=.9, hang = -1)
ap_pcp_de_kmout<-kmeans(scale(AP_PCP_DE),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_DE, col = (ap_pcp_de_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_de_kmout$cluster) , col = unique(ap_pcp_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_de_kmout$cluster, DEvars$TIME)
ap_pcp_de_kmout$centers

set.seed(2)
AP_PCTU_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_TU_DE"))
ap_pctu_de_hcomp<-hclust(dist(scale(AP_PCTU_DE)), method = "complete")
plot(ap_pctu_de_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_DE", cex=.9, hang = -1)
ap_pctu_de_kmout<-kmeans(scale(AP_PCTU_DE),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_DE, col = (ap_pctu_de_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_de_kmout$cluster) , col = unique(ap_pctu_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_de_kmout$cluster, DEvars$TIME)
ap_pctu_de_kmout$centers

set.seed(2)
AP_PCFU_DE<-subset.data.frame(DEvars, select = c("AP_DE", "PC_FU_DE"))
ap_pcfu_de_hcomp<-hclust(dist(scale(AP_PCFU_DE)), method="complete")
plot(ap_pcfu_de_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_DE", cex=.9, hang = -1)
ap_pcfu_de_kmout<-kmeans(scale(AP_PCFU_DE),3,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_DE, col=(ap_pcfu_de_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_de_kmout$cluster) , col = unique(ap_pcfu_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_de_kmout$cluster,DEvars$TIME)
ap_pcfu_de_kmout$centers

set.seed(2)
AP_EgfME_DE<-subset.data.frame(DEvars, select = c("AP_DE", "Egf_ME_DE"))
ap_egfme_de_hcomp<-hclust(dist(scale(AP_EgfME_DE)), method = "complete")
plot(ap_egfme_de_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_DE", cex=.9, hang = -1)
ap_egfme_de_kmout<-kmeans(scale(AP_EgfME_DE),4,nstart = 15)
par(mar=c(5,4,2,4))
```

```
plot(AP_EgfME_DE, col= (ap_egfme_de_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_de_kmout$cluster) , col = unique(ap_egfme_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_de_kmout$cluster, DEvars$TIME)
ap_egfme_de_kmout$centers

AP_HME_DE<-subset.data.frame(DEvars, select = c("AP_DE", "H_ME_DE"))
ap_hme_de_hcomp<-hclust(dist(scale(AP_HME_DE)), method = "complete")
plot(ap_hme_de_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_DE", cex=.9, hang = -1)
ap_hme_de_kmout<- kmeans(scale(AP_HME_DE),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_DE, col = (ap_hme_de_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_de_kmout$cluster) , col = unique(ap_hme_de_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_de_kmout$cluster, DEvars$TIME)
ap_hme_de_kmout$centers

set.seed(2)
DEvars_hcomp<-hclust(dist(scale(DEvars)), method = "complete")
plot(DEvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of DE", cex=.9, hang = -1)
DEvars_kmout<- kmeans(scale(DEvars), 4 , nstart = 20)
table(DEvars_kmout$cluster, DEvars$TIME)

set.seed(2)
AP_PSR_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PS_R_IE"))
ap_psr_ie_hcomp<-hclust(dist(scale(AP_PSR_IE)), method = "complete")
plot(ap_psr_ie_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_IE", cex=.9, hang = -1)
ap_psr_ie_kmout<-kmeans(scale(AP_PSR_IE),3, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_IE, col = (ap_psr_ie_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_ie_kmout$cluster) , col = unique(ap_psr_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_ie_kmout$cluster, IEvars$TIME)
ap_psr_ie_kmout$centers

set.seed(2)
D_HDD_IE<- subset.data.frame(IEvars, select = c("D_IE", "HDD_IE"))
d_hdd_ie_hcomp<-hclust(dist(scale(D_HDD_IE)), method = "complete")
plot(d_hdd_ie_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_IE", cex=.9, hang = -1)
d_hdd_ie_kmout<-kmeans(scale(D_HDD_IE),5, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_IE, col = (d_hdd_ie_kmout$cluster+1),  main="K-Means Clustering Results with K=5", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_ie_kmout$cluster) , col = unique(d_hdd_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_ie_kmout$cluster, IEvars$TIME)
d_hdd_ie_kmout$centers

set.seed(2)
AP_PCP_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_P_IE"))
ap_pcp_ie_hcomp<-hclust(dist(scale(AP_PCP_IE)), method = "complete")
plot(ap_pcp_ie_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_IE", cex=.9, hang = -1)
```

```
ap_pcp_ie_kmout<-kmeans(scale(AP_PCP_IE),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_IE, col = (ap_pcp_ie_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_ie_kmout$cluster) , col = unique(ap_pcp_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_ie_kmout$cluster, IEvars$TIME)
ap_pcp_ie_kmout$centers

set.seed(2)
AP_PCTU_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_TU_IE"))
ap_pctu_ie_hcomp<-hclust(dist(scale(AP_PCTU_IE)), method = "complete")
plot(ap_pctu_ie_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_IE", cex=.9, hang = -1)
ap_pctu_ie_kmout<-kmeans(scale(AP_PCTU_IE),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_IE, col = (ap_pctu_ie_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_ie_kmout$cluster) , col = unique(ap_pctu_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_ie_kmout$cluster, IEvars$TIME)
ap_pctu_ie_kmout$centers

set.seed(2)
AP_PCFU_IE<-subset.data.frame(IEvars, select = c("AP_IE", "PC_FU_IE"))
ap_pcfu_ie_hcomp<-hclust(dist(scale(AP_PCFU_IE)), method="complete")
plot(ap_pcfu_ie_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_IE", cex=.9, hang = -1)
ap_pcfu_ie_kmout<-kmeans(scale(AP_PCFU_IE),4,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_IE, col=(ap_pcfu_ie_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_ie_kmout$cluster) , col = unique(ap_pcfu_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_ie_kmout$cluster,IEvars$TIME)
ap_pcfu_ie_kmout$centers

set.seed(2)
AP_EgfME_IE<-subset.data.frame(IEvars, select = c("AP_IE", "Egf_ME_IE"))
ap_egfme_ie_hcomp<-hclust(dist(scale(AP_EgfME_IE)), method = "complete")
plot(ap_egfme_ie_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_IE", cex=.9, hang = -1)
ap_egfme_ie_kmout<-kmeans(scale(AP_EgfME_IE),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_IE, col= (ap_egfme_ie_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_ie_kmout$cluster) , col = unique(ap_egfme_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_ie_kmout$cluster, IEvars$TIME)
ap_egfme_ie_kmout$centers

set.seed(2)
AP_HME_IE<-subset.data.frame(IEvars, select = c("AP_IE", "H_ME_IE"))
ap_hme_ie_hcomp<-hclust(dist(scale(AP_HME_IE)), method = "complete")
plot(ap_hme_ie_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_IE", cex=.9, hang = -1)
ap_hme_ie_kmout<- kmeans(scale(AP_HME_IE),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_IE, col = (ap_hme_ie_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
```

```
legend("topleft", legend = unique(ap_hme_ie_kmout$cluster) , col = unique(ap_hme_ie_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_ie_kmout$cluster, IEvars$TIME)
ap_hme_ie_kmout$centers

set.seed(2)
IEvars_hcomp<-hclust(dist(scale(IEvars)), method = "complete")
plot(IEvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of IE", cex=.9, hang = -1)
IEvars_kmout<- kmeans(scale(IEvars), 4 , nstart = 20)
table(IEvars_kmout$cluster, IEvars$TIME)

set.seed(2)
AP_PSR_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PS_R_EL"))
ap_psr_el_hcomp<-hclust(dist(scale(AP_PSR_EL)), method = "complete")
plot(ap_psr_el_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_EL", cex=.9, hang = -1)
ap_psr_el_kmout<-kmeans(scale(AP_PSR_EL),4, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_EL, col = (ap_psr_el_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_el_kmout$cluster) , col = unique(ap_psr_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_el_kmout$cluster, ELvars$TIME)
ap_psr_el_kmout$centers

set.seed(2)
D_HDD_EL<- subset.data.frame(ELvars, select = c("D_EL", "HDD_EL"))
d_hdd_el_hcomp<-hclust(dist(scale(D_HDD_EL)), method = "complete")
plot(d_hdd_el_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_EL", cex=.9, hang = -1)
d_hdd_el_kmout<-kmeans(scale(D_HDD_EL),5, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_EL, col = (d_hdd_el_kmout$cluster+1),  main="K-Means Clustering Results with K=5", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_el_kmout$cluster) , col = unique(d_hdd_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_el_kmout$cluster, ELvars$TIME)
d_hdd_el_kmout$centers

set.seed(2)
AP_PCP_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_P_EL"))
ap_pcp_el_hcomp<-hclust(dist(scale(AP_PCP_EL)), method = "complete")
plot(ap_pcp_el_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_EL", cex=.9, hang = -1)
ap_pcp_el_kmout<-kmeans(scale(AP_PCP_EL),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_EL, col = (ap_pcp_el_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_el_kmout$cluster) , col = unique(ap_pcp_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_el_kmout$cluster, ELvars$TIME)
ap_pcp_el_kmout$centers

set.seed(2)
AP_PCTU_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_TU_EL"))
ap_pctu_el_hcomp<-hclust(dist(scale(AP_PCTU_EL)), method = "complete")
plot(ap_pctu_el_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_EL", cex=.9, hang = -1)
ap_pctu_el_kmout<-kmeans(scale(AP_PCTU_EL),5, nstart = 15)
```

```
par(mar=c(5,4,2,4))
plot(AP_PCTU_EL, col = (ap_pctu_el_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_el_kmout$cluster) , col = unique(ap_pctu_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_el_kmout$cluster, ELvars$TIME)
ap_pctu_el_kmout$centers

set.seed(2)
AP_PCFU_EL<-subset.data.frame(ELvars, select = c("AP_EL", "PC_FU_EL"))
ap_pcfu_el_hcomp<-hclust(dist(scale(AP_PCFU_EL)), method="complete")
plot(ap_pcfu_el_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_EL", cex=.9, hang = -1)
ap_pcfu_el_kmout<-kmeans(scale(AP_PCFU_EL),4,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_EL, col=(ap_pcfu_el_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_el_kmout$cluster) , col = unique(ap_pcfu_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_el_kmout$cluster,ELvars$TIME)
ap_pcfu_el_kmout$centers

set.seed(2)
AP_EgfME_EL<-subset.data.frame(ELvars, select = c("AP_EL", "Egf_ME_EL"))
ap_egfme_el_hcomp<-hclust(dist(scale(AP_EgfME_EL)), method = "complete")
plot(ap_egfme_el_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_EL", cex=.9, hang = -1)
ap_egfme_el_kmout<-kmeans(scale(AP_EgfME_EL),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_EL, col= (ap_egfme_el_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_el_kmout$cluster) , col = unique(ap_egfme_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_el_kmout$cluster, ELvars$TIME)
ap_egfme_el_kmout$centers

set.seed(2)
AP_HME_EL<-subset.data.frame(ELvars, select = c("AP_EL", "H_ME_EL"))
ap_hme_el_hcomp<-hclust(dist(scale(AP_HME_EL)), method = "complete")
plot(ap_hme_el_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_EL", cex=.9, hang = -1)
ap_hme_el_kmout<- kmeans(scale(AP_HME_EL),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_EL, col = (ap_hme_el_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_el_kmout$cluster) , col = unique(ap_hme_el_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_el_kmout$cluster, ELvars$TIME)
ap_hme_el_kmout$centers

set.seed(2)
ELvars_hcomp<-hclust(dist(scale(ELvars)), method = "complete")
plot(ELvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of EL", cex=.9, hang = -1)
ELvars_kmout<- kmeans(scale(ELvars), 3 , nstart = 20)
table(ELvars_kmout$cluster, ELvars$TIME)

set.seed(2)
AP_PSR_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PS_R_ES"))
ap_psr_es_hcomp<-hclust(dist(scale(AP_PSR_ES)), method = "complete")
```

```
plot(ap_psr_es_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_ES", cex=.9, hang = -1)
ap_psr_es_kmout<-kmeans(scale(AP_PSR_ES),5, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_ES, col = (ap_psr_es_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_es_kmout$cluster) , col = unique(ap_psr_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_es_kmout$cluster, ESvars$TIME)
ap_psr_es_kmout$centers

set.seed(2)
D_HDD_ES<- subset.data.frame(ESvars, select = c("D_ES", "HDD_ES"))
d_hdd_es_hcomp<-hclust(dist(scale(D_HDD_ES)), method = "complete")
plot(d_hdd_es_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_ES", cex=.9, hang = -1)
d_hdd_es_kmout<-kmeans(scale(D_HDD_ES),5, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_ES, col = (d_hdd_es_kmout$cluster+1),  main="K-Means Clustering Results with K=5", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_es_kmout$cluster) , col = unique(d_hdd_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_es_kmout$cluster, ESvars$TIME)
d_hdd_es_kmout$centers

set.seed(2)
AP_PCP_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_P_ES"))
ap_pcp_es_hcomp<-hclust(dist(scale(AP_PCP_ES)), method = "complete")
plot(ap_pcp_es_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_ES", cex=.9, hang = -1)
ap_pcp_es_kmout<-kmeans(scale(AP_PCP_ES),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_ES, col = (ap_pcp_es_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_es_kmout$cluster) , col = unique(ap_pcp_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_es_kmout$cluster, ESvars$TIME)
ap_pcp_es_kmout$centers

set.seed(2)
AP_PCTU_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_TU_ES"))
ap_pctu_es_hcomp<-hclust(dist(scale(AP_PCTU_ES)), method = "complete")
plot(ap_pctu_es_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_ES", cex=.9, hang = -1)
ap_pctu_es_kmout<-kmeans(scale(AP_PCTU_ES),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_ES, col = (ap_pctu_es_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_es_kmout$cluster) , col = unique(ap_pctu_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_es_kmout$cluster, ESvars$TIME)
ap_pctu_es_kmout$centers

set.seed(2)
AP_PCFU_ES<-subset.data.frame(ESvars, select = c("AP_ES", "PC_FU_ES"))
ap_pcfu_es_hcomp<-hclust(dist(scale(AP_PCFU_ES)), method="complete")
plot(ap_pcfu_es_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_ES", cex=.9, hang = -1)
ap_pcfu_es_kmout<-kmeans(scale(AP_PCFU_ES),4,nstart=15)
par(mar=c(5,4,2,4))
```

```
plot(AP_PCFU_ES, col=(ap_pcfu_es_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_es_kmout$cluster) , col = unique(ap_pcfu_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_es_kmout$cluster,ESvars$TIME)
ap_pcfu_es_kmout$centers

set.seed(2)
AP_EgfME_ES<-subset.data.frame(ESvars, select = c("AP_ES", "Egf_ME_ES"))
ap_egfme_es_hcomp<-hclust(dist(scale(AP_EgfME_ES)), method = "complete")
plot(ap_egfme_es_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_ES", cex=.9, hang = -1)
ap_egfme_es_kmout<-kmeans(scale(AP_EgfME_ES),5,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_ES, col= (ap_egfme_es_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_es_kmout$cluster) , col = unique(ap_egfme_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_es_kmout$cluster, ESvars$TIME)
ap_egfme_es_kmout$centers

set.seed(2)
AP_HME_ES<-subset.data.frame(ESvars, select = c("AP_ES", "H_ME_ES"))
ap_hme_es_hcomp<-hclust(dist(scale(AP_HME_ES)), method = "complete")
plot(ap_hme_es_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_ES", cex=.9, hang = -1)
ap_hme_es_kmout<- kmeans(scale(AP_HME_ES),5,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_ES, col = (ap_hme_es_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_es_kmout$cluster) , col = unique(ap_hme_es_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_es_kmout$cluster, ESvars$TIME)
ap_hme_es_kmout$centers

set.seed(2)
ESvars_hcomp<-hclust(dist(scale(ESvars)), method = "complete")
plot(ESvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of ES", cex=.9, hang = -1)
ESvars_kmout<- kmeans(scale(ESvars), 3 , nstart = 20)
table(ESvars_kmout$cluster, ESvars$TIME)

set.seed(2)
AP_PSR_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PS_R_NL"))
ap_psr_nl_hcomp<-hclust(dist(scale(AP_PSR_NL)), method = "complete")
plot(ap_psr_nl_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_NL", cex=.9, hang = -1)
ap_psr_nl_kmout<-kmeans(scale(AP_PSR_NL),3, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_NL, col = (ap_psr_nl_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_nl_kmout$cluster) , col = unique(ap_psr_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_nl_kmout$cluster, NLvars$TIME)
ap_psr_nl_kmout$centers

set.seed(2)
D_HDD_NL<- subset.data.frame(NLvars, select = c("D_NL", "HDD_NL"))
d_hdd_nl_hcomp<-hclust(dist(scale(D_HDD_NL)), method = "complete")
```

```
plot(d_hdd_nl_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_NL", cex=.9, hang = -1)
d_hdd_nl_kmout<-kmeans(scale(D_HDD_NL),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_NL, col = (d_hdd_nl_kmout$cluster+1),  main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_nl_kmout$cluster) , col = unique(d_hdd_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_nl_kmout$cluster, NLvars$TIME)
d_hdd_nl_kmout$centers

set.seed(2)
AP_PCP_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_P_NL"))
ap_pcp_nl_hcomp<-hclust(dist(scale(AP_PCP_NL)), method = "complete")
plot(ap_pcp_nl_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_NL", cex=.9, hang = -1)
ap_pcp_nl_kmout<-kmeans(scale(AP_PCP_NL),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_NL, col = (ap_pcp_nl_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_nl_kmout$cluster) , col = unique(ap_pcp_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_nl_kmout$cluster, NLvars$TIME)
ap_pcp_nl_kmout$centers

set.seed(2)
AP_PCTU_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_TU_NL"))
ap_pctu_nl_hcomp<-hclust(dist(scale(AP_PCTU_NL)), method = "complete")
plot(ap_pctu_nl_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_NL", cex=.9, hang = -1)
ap_pctu_nl_kmout<-kmeans(scale(AP_PCTU_NL),5, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_NL, col = (ap_pctu_nl_kmout$cluster+1), main="K-Means Clustering Results with K=5", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_nl_kmout$cluster) , col = unique(ap_pctu_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_nl_kmout$cluster, NLvars$TIME)
ap_pctu_nl_kmout$centers

set.seed(2)
AP_PCFU_NL<-subset.data.frame(NLvars, select = c("AP_NL", "PC_FU_NL"))
ap_pcfu_nl_hcomp<-hclust(dist(scale(AP_PCFU_NL)), method="complete")
plot(ap_pcfu_nl_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_NL", cex=.9, hang = -1)
ap_pcfu_nl_kmout<-kmeans(scale(AP_PCFU_NL),4,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_NL, col=(ap_pcfu_nl_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_nl_kmout$cluster) , col = unique(ap_pcfu_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_nl_kmout$cluster,NLvars$TIME)
ap_pcfu_nl_kmout$centers

set.seed(2)
AP_EgfME_NL<-subset.data.frame(NLvars, select = c("AP_NL", "Egf_ME_NL"))
ap_egfme_nl_hcomp<-hclust(dist(scale(AP_EgfME_NL)), method = "complete")
plot(ap_egfme_nl_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_NL", cex=.9, hang = -1)
ap_egfme_nl_kmout<-kmeans(scale(AP_EgfME_NL),3,nstart = 15)
par(mar=c(5,4,2,4))
```

```
plot(AP_EgfME_NL, col= (ap_egfme_nl_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_nl_kmout$cluster) , col = unique(ap_egfme_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_nl_kmout$cluster, NLvars$TIME)
ap_egfme_nl_kmout$centers

set.seed(2)
AP_HME_NL<-subset.data.frame(NLvars, select = c("AP_NL", "H_ME_NL"))
ap_hme_nl_hcomp<-hclust(dist(scale(AP_HME_NL)), method = "complete")
plot(ap_hme_nl_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_NL", cex=.9, hang = -1)
ap_hme_nl_kmout<- kmeans(scale(AP_HME_NL),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_NL, col = (ap_hme_nl_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_nl_kmout$cluster) , col = unique(ap_hme_nl_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_nl_kmout$cluster, NLvars$TIME)
ap_hme_nl_kmout$centers

set.seed(2)
NLvars_hcomp<-hclust(dist(scale(NLvars)), method = "complete")
plot(NLvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of NL", cex=.9, hang = -1)
NLvars_kmout<- kmeans(scale(NLvars), 4 , nstart = 20)
table(NLvars_kmout$cluster, NLvars$TIME)

set.seed(2)
AP_PSR_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PS_R_AT"))
ap_psr_at_hcomp<-hclust(dist(scale(AP_PSR_AT)), method = "complete")
plot(ap_psr_at_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_AT", cex=.9, hang = -1)
ap_psr_at_kmout<-kmeans(scale(AP_PSR_AT),4, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_AT, col = (ap_psr_at_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_at_kmout$cluster) , col = unique(ap_psr_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_at_kmout$cluster, ATvars$TIME)
ap_psr_at_kmout$centers

set.seed(2)
D_HDD_AT<- subset.data.frame(ATvars, select = c("D_AT", "HDD_AT"))
d_hdd_at_hcomp<-hclust(dist(scale(D_HDD_AT)), method = "complete")
plot(d_hdd_at_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_AT", cex=.9, hang = -1)
d_hdd_at_kmout<-kmeans(scale(D_HDD_AT),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_AT, col = (d_hdd_at_kmout$cluster+1),  main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_at_kmout$cluster) , col = unique(d_hdd_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_at_kmout$cluster, ATvars$TIME)
d_hdd_at_kmout$centers

set.seed(2)
AP_PCP_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_P_AT"))
ap_pcp_at_hcomp<-hclust(dist(scale(AP_PCP_AT)), method = "complete")
```

```
plot(ap_pcp_at_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_AT", cex=.9, hang = -1)
ap_pcp_at_kmout<-kmeans(scale(AP_PCP_AT),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_AT, col = (ap_pcp_at_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_at_kmout$cluster) , col = unique(ap_pcp_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_at_kmout$cluster, ATvars$TIME)
ap_pcp_at_kmout$centers

set.seed(2)
AP_PCTU_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_TU_AT"))
ap_pctu_at_hcomp<-hclust(dist(scale(AP_PCTU_AT)), method = "complete")
plot(ap_pctu_at_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_AT", cex=.9, hang = -1)
ap_pctu_at_kmout<-kmeans(scale(AP_PCTU_AT),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_AT, col = (ap_pctu_at_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_at_kmout$cluster) , col = unique(ap_pctu_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_at_kmout$cluster, ATvars$TIME)
ap_pctu_at_kmout$centers

set.seed(2)
AP_PCFU_AT<-subset.data.frame(ATvars, select = c("AP_AT", "PC_FU_AT"))
ap_pcfu_at_hcomp<-hclust(dist(scale(AP_PCFU_AT)), method="complete")
plot(ap_pcfu_at_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_AT", cex=.9, hang = -1)
ap_pcfu_at_kmout<-kmeans(scale(AP_PCFU_AT),4,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_AT, col=(ap_pcfu_at_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_at_kmout$cluster) , col = unique(ap_pcfu_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_at_kmout$cluster,ATvars$TIME)
ap_pcfu_at_kmout$centers

set.seed(2)
AP_EgfME_AT<-subset.data.frame(ATvars, select = c("AP_AT", "Egf_ME_AT"))
ap_egfme_at_hcomp<-hclust(dist(scale(AP_EgfME_AT)), method = "complete")
plot(ap_egfme_at_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_AT", cex=.9, hang = -1)
ap_egfme_at_kmout<-kmeans(scale(AP_EgfME_AT),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_AT, col= (ap_egfme_at_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_at_kmout$cluster) , col = unique(ap_egfme_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_at_kmout$cluster, ATvars$TIME)
ap_egfme_at_kmout$centers

set.seed(2)
AP_HME_AT<-subset.data.frame(ATvars, select = c("AP_AT", "H_ME_AT"))
ap_hme_at_hcomp<-hclust(dist(scale(AP_HME_AT)), method = "complete")
plot(ap_hme_at_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_AT", cex=.9, hang = -1)
ap_hme_at_kmout<- kmeans(scale(AP_HME_AT),4,nstart = 15)
par(mar=c(5,4,2,4))
```

```
plot(AP_HME_AT, col = (ap_hme_at_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_at_kmout$cluster) , col = unique(ap_hme_at_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_at_kmout$cluster, ATvars$TIME)
ap_hme_at_kmout$centers

set.seed(2)
ATvars_hcomp<-hclust(dist(scale(ATvars)), method = "complete")
plot(ATvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of AT", cex=.9, hang = -1)
ATvars_kmout<- kmeans(scale(ATvars),3, nstart = 20)
table(ATvars_kmout$cluster, ATvars$TIME)

set.seed(2)
AP_PSR_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PS_R_PT"))
ap_psr_pt_hcomp<-hclust(dist(scale(AP_PSR_PT)), method = "complete")
plot(ap_psr_pt_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_PT", cex=.9, hang = -1)
ap_psr_pt_kmout<-kmeans(scale(AP_PSR_PT),3, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_PT, col = (ap_psr_pt_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_pt_kmout$cluster) , col = unique(ap_psr_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_pt_kmout$cluster, PTvars$TIME)
ap_psr_pt_kmout$centers

set.seed(2)
D_HDD_PT<- subset.data.frame(PTvars, select = c("D_PT", "HDD_PT"))
d_hdd_pt_hcomp<-hclust(dist(scale(D_HDD_PT)), method = "complete")
plot(d_hdd_pt_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_PT", cex=.9, hang = -1)
d_hdd_pt_kmout<-kmeans(scale(D_HDD_PT),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_PT, col = (d_hdd_pt_kmout$cluster+1),  main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_psr_pt_kmout$cluster) , col = unique(ap_psr_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_pt_kmout$cluster, PTvars$TIME)
d_hdd_pt_kmout$centers

set.seed(2)
AP_PCP_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_P_PT"))
ap_pcp_pt_hcomp<-hclust(dist(scale(AP_PCP_PT)), method = "complete")
plot(ap_pcp_pt_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_PT", cex=.9, hang = -1)
ap_pcp_pt_kmout<-kmeans(scale(AP_PCP_PT),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_PT, col = (ap_pcp_pt_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_pt_kmout$cluster) , col = unique(ap_pcp_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_pt_kmout$cluster, PTvars$TIME)
ap_pcp_pt_kmout$centers

set.seed(2)
AP_PCTU_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_TU_PT"))
ap_pctu_pt_hcomp<-hclust(dist(scale(AP_PCTU_PT)), method = "complete")
```

```
plot(ap_pctu_pt_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_PT", cex=.9, hang = -1)
ap_pctu_pt_kmout<-kmeans(scale(AP_PCTU_PT),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_PT, col = (ap_pctu_pt_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_pt_kmout$cluster) , col = unique(ap_pctu_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_pt_kmout$cluster, PTvars$TIME)
ap_pctu_pt_kmout$centers

set.seed(2)
AP_PCFU_PT<-subset.data.frame(PTvars, select = c("AP_PT", "PC_FU_PT"))
ap_pcfu_pt_hcomp<-hclust(dist(scale(AP_PCFU_PT)), method="complete")
plot(ap_pcfu_pt_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_PT", cex=.9, hang = -1)
ap_pcfu_pt_kmout<-kmeans(scale(AP_PCFU_PT),3,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_PT, col=(ap_pcfu_pt_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_pt_kmout$cluster) , col = unique(ap_pcfu_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_pt_kmout$cluster,PTvars$TIME)
ap_pcfu_pt_kmout$centers

set.seed(2)
AP_EgfME_PT<-subset.data.frame(PTvars, select = c("AP_PT", "Egf_ME_PT"))
ap_egfme_pt_hcomp<-hclust(dist(scale(AP_EgfME_PT)), method = "complete")
plot(ap_egfme_pt_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_PT", cex=.9, hang = -1)
ap_egfme_pt_kmout<-kmeans(scale(AP_EgfME_PT),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_PT, col= (ap_egfme_pt_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_pt_kmout$cluster) , col = unique(ap_egfme_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_pt_kmout$cluster, PTvars$TIME)
ap_egfme_pt_kmout$centers

set.seed(2)
AP_HME_PT<-subset.data.frame(PTvars, select = c("AP_PT", "H_ME_PT"))
ap_hme_pt_hcomp<-hclust(dist(scale(AP_HME_PT)), method = "complete")
plot(ap_hme_pt_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_PT", cex=.9, hang = -1)
ap_hme_pt_kmout<- kmeans(scale(AP_HME_PT),4,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_PT, col = (ap_hme_pt_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_pt_kmout$cluster) , col = unique(ap_hme_pt_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_pt_kmout$cluster, PTvars$TIME)
ap_hme_pt_kmout$centers

set.seed(2)
PTvars_hcomp<-hclust(dist(scale(PTvars)), method = "complete")
plot(PTvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of PT", cex=.9, hang = -1)
PTvars_kmout<- kmeans(scale(PTvars),4, nstart = 20)
table(PTvars_kmout$cluster, PTvars$TIME)
```

```
set.seed(2)
AP_PSR_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PS_R_UK"))
ap_psr_uk_hcomp<-hclust(dist(scale(AP_PSR_UK)), method = "complete")
plot(ap_psr_uk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PS_R_UK", cex=.9, hang = -1)
ap_psr_uk_kmout<-kmeans(scale(AP_PSR_UK),2, nstart = 20)
par(mar=c(5,4,2,4))
plot(AP_PSR_UK, col = (ap_psr_uk_kmout$cluster+1), main="K-Means Clustering Results with K=2", pch=20, cex=2 )
legend("topleft", legend = unique(ap_psr_uk_kmout$cluster) , col = unique(ap_psr_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_psr_uk_kmout$cluster, UKvars$TIME)
ap_psr_uk_kmout$centers

set.seed(2)
D_HDD_UK<- subset.data.frame(UKvars, select = c("D_UK", "HDD_UK"))
d_hdd_uk_hcomp<-hclust(dist(scale(D_HDD_UK)), method = "complete")
plot(d_hdd_uk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for D_HDD_UK", cex=.9, hang = -1)
d_hdd_uk_kmout<-kmeans(scale(D_HDD_UK),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(D_HDD_UK, col = (d_hdd_uk_kmout$cluster+1),  main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(d_hdd_uk_kmout$cluster) , col = unique(d_hdd_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(d_hdd_uk_kmout$cluster, UKvars$TIME)
d_hdd_uk_kmout$centers

set.seed(2)
AP_PCP_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_P_UK"))
ap_pcp_uk_hcomp<-hclust(dist(scale(AP_PCP_UK)), method = "complete")
plot(ap_pcp_uk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_P_UK", cex=.9, hang = -1)
ap_pcp_uk_kmout<-kmeans(scale(AP_PCP_UK),3, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCP_UK, col = (ap_pcp_uk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pcp_uk_kmout$cluster) , col = unique(ap_pcp_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcp_uk_kmout$cluster, UKvars$TIME)
ap_pcp_uk_kmout$centers

set.seed(2)
AP_PCTU_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_TU_UK"))
ap_pctu_uk_hcomp<-hclust(dist(scale(AP_PCTU_UK)), method = "complete")
plot(ap_pctu_uk_hcomp, main = "Complete Linkage", xlab="", sub="h-clustering for AP_PC_TU_UK", cex=.9, hang = -1)
ap_pctu_uk_kmout<-kmeans(scale(AP_PCTU_UK),4, nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_PCTU_UK, col = (ap_pctu_uk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2 )
legend("topleft", legend = unique(ap_pctu_uk_kmout$cluster) , col = unique(ap_pctu_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pctu_uk_kmout$cluster, UKvars$TIME)
ap_pctu_uk_kmout$centers

set.seed(2)
AP_PCFU_UK<-subset.data.frame(UKvars, select = c("AP_UK", "PC_FU_UK"))
ap_pcfu_uk_hcomp<-hclust(dist(scale(AP_PCFU_UK)), method="complete")
```

```
plot(ap_pcfu_uk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_PC_FU_UK", cex=.9, hang = -1)
ap_pcfu_uk_kmout<-kmeans(scale(AP_PCFU_UK),4,nstart=15)
par(mar=c(5,4,2,4))
plot(AP_PCFU_UK, col=(ap_pcfu_uk_kmout$cluster+1), main="K-Means Clustering Results with K=4", pch=20, cex=2)
legend("topleft", legend = unique(ap_pcfu_uk_kmout$cluster) , col = unique(ap_pcfu_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_pcfu_uk_kmout$cluster,UKvars$TIME)
ap_pcfu_uk_kmout$centers

set.seed(2)
AP_EgfME_UK<-subset.data.frame(UKvars, select = c("AP_UK", "Egf_ME_UK"))
ap_egfme_uk_hcomp<-hclust(dist(scale(AP_EgfME_UK)), method = "complete")
plot(ap_egfme_uk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_Egf_ME_UK", cex=.9, hang = -1)
ap_egfme_uk_kmout<-kmeans(scale(AP_EgfME_UK),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_EgfME_UK, col= (ap_egfme_uk_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_egfme_uk_kmout$cluster) , col = unique(ap_egfme_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_egfme_uk_kmout$cluster, UKvars$TIME)
ap_egfme_uk_kmout$centers

set.seed(2)
AP_HME_UK<-subset.data.frame(UKvars, select = c("AP_UK", "H_ME_UK"))
ap_hme_uk_hcomp<-hclust(dist(scale(AP_HME_UK)), method = "complete")
plot(ap_hme_uk_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for AP_H_ME_UK", cex=.9, hang = -1)
ap_hme_uk_kmout<- kmeans(scale(AP_HME_UK),3,nstart = 15)
par(mar=c(5,4,2,4))
plot(AP_HME_UK, col = (ap_hme_uk_kmout$cluster+1), main="K-Means Clustering Results with K=3", pch=20, cex=2)
legend("topleft", legend = unique(ap_hme_uk_kmout$cluster) , col = unique(ap_hme_uk_kmout$cluster)+1, pch=20, cex=1.5,inset=c(1,0), xpd=TRUE, bty="n")
table(ap_hme_uk_kmout$cluster, UKvars$TIME)
ap_hme_uk_kmout$centers

set.seed(2)
UKvars_hcomp<-hclust(dist(scale(UKvars)), method = "complete")
plot(UKvars_hcomp, main="Complete Linkage", xlab="", sub="h-clustering for all variables of UK", cex=.9, hang = -1)
UKvars_kmout<- kmeans(scale(UKvars),4, nstart = 20)
table(UKvars_kmout$cluster, UKvars$TIME)
```