Macrothink Institute™

Research in Applied Economics
ISSN 1948-5433
2016, Vol. 8, No. 2

Clustering of OECD Countries Out of Pocket Health Expenditure Time Series Data

Songul Cinaroglu[1]

[1]Hacettepe University, Faculty of Economics and Administrative Sciences (FEAS) Department of Health Care Management, Beytepe, Ankara, Turkey.

*Correspondence: Hacettepe University, Faculty of Economics and Administrative Sciences (FEAS) Department of Health Care Management, Beytepe, Ankara, Turkey. Tel: 90-312-297-6356 (278) E-mail: cinaroglus@hacettepe.edu.tr

Received: April 24, 2016 Accepted: June 20, 2016 Published: June 24, 2016

doi:10.5296/rae.v8i2.9377 URL: http://dx.doi.org/10.5296/rae.v8i2.9377

Abstract

Out of pocket health expenditures points out to the payments made by households at the point they receive health services. Frequently these include doctor consultation fees, purchase of medication and hospital bills. In this study hierarchical clustering method was used for classification of 34 countries which are members of OECD (Organization for Economic Cooperation and Development) in terms of out of pocket health expenditures for the years between 1995-2011. Longest common subsequences (LCS), correlation coefficient and Euclidean distance measure was used as a measure of similarity and distance in hierarchical clustering. At the end of the analysis it was found that LCS and Euclidean distance measures were the best for determining clusters. Furthermore, study results led to understand grouping of OECD countries according to health expenditures.

Keywords: Hierarchical Clustering, Time Series Data, OECD Countries

1. Introduction

Today there is an increasing interest in the cross-country comparisons of the performance of national health care systems (Varabyova & Schreyögg, 2013). Health expenditure is an important indicator for better understanding and measuring performance of health care systems (Lette et al., 2016). OECD countries are currently spending record amounts on health care compared with other parts of the world (Ozcan & Khushalani 2016; Huber & Orosz, 2003). In many OECD countries, the health care system constitutes the largest service industry, with an average health spending reaching 9.5% of GDP in 2010 (Varabyova & Schreyögg, 2013). Health spending has different trend in most of countries. For example, health spending growth slowed in 2010 and 2011, notably in Canada (3.0 % in 2010 and 0.8% in 2011 in real terms) and the United States (2.5 % in 2010 and 1.8% in 2011, in real terms). But this is the opposite for European countries. In the United States, the share of health spending to GDP has reached at 17.7 percent between the years 2009-2011.

Out-of-pocket spending on health care is one of the most dynamic components of private consumption in lots of OECD countries (Huber & Orosz, 2003; Penders et al. 2016). Out-of-pocket health expenditures consists of gratuities, in-kind payments to health practitioners, suppliers of pharmaceuticals, therapeutic appliances and payments for other goods and services whose primary aim is enhancement of the health status of population (WB, 2016). There are differences between countries in the components of goods and services that are paid out of pocket (Huber & Orosz, 2003). Generally private health expenditures consist of doctor's consultation fees, purchases of medication and hospital bills (WHO, 2005). Moreover, pharmaceuticals are one of the big components in all countries (Huber & Orosz, 2003).

It is also important to note that out-of-pocket payments are refers to any insurance reimbursement (WHO, 2005). In most of other OECD countries, private health insurance represents a small proportion of total health revenues, but in the United States private insurance reflects for almost the same percentage of health care revenues as public insurance (Anderson & Frogner, 2008; Moses et al., 2015). Recent reports emphasize that despite health spending grows slowly after 2010 in European countries, health expenditure differences between European countries and US is still continuing (OECD 2015). Moreover, OECD countries are helping developing countries to fight against health inequalities and increasing health costs (Devaux 2015). To improve general health status of people, reduce health disparities and effectively manage capital flows, it become important to understand how developed OECD countries are grouping according to their out of pocket health expenditures.

In this study hierarchical clustering methods was used for classification of countries which are members of OECD in terms of out of pocket health expenditures for the years between 1995-2011. Longest common subsequences (1-LCS), correlation coefficients ($1 - \mathrm{cor}, 1 - \mathrm{cor}^2$) was used as a measure of dissimilarity and Euclidean distance measure was used for hierarchical clustering. The plan of the paper is as follows. Section 2 introduces materials and methods, Section 3 summarizes empirical study results and Section 4 is conclusion part.

2. Materials and Methods

Out of pocket health expenditures (% of private health expenditure) of 34 OECD countries for the years of 1995-2011 derived from World Bank (WB) website which is the broadest source of comparable statistics on diverse health systems across WB countries (WB-World Bank, 2016). Considerable amount of scientific and business data is represented in the form of time series. Similarity detection and clustering are two common methods of time series analysis. In time series similarity detection, it is aimed to detect similarities between different time series. Furthermore, the second one is time series clustering according to definite features (Grabusts and Borisov, 2009).

It is well known that cluster analysis is about finding groups in datasets (Singhal and Seborg, 2005). Data clustering is an important method to analyze a data set according to find its structure. Number of clustering algorithms have been developed and applied in many different research fields (Ozkan & Turkşen, 2013). One of the most common clustering algorithms is hierarchical clustering. This method works by grouping data objects into a tree of clusters (Liao, 2005). Hierarchical clustering makes a hierarchy of clusters that can be shown by a tree called a dendrogram. These algorithms are divided into two groups which are called agglomerative and divisive approaches (Das et al., 2007). Agglomerative Hierarchical Clustering (AHC) is one of semi supervised clustering methods discussed in the literature (Hamasuna et al., 2012). In agglomerative hierarchical clustering, the dissimilarity is used for measuring the closeness of two clusters (Hamasuna et al., 2012). In this study longest common subsequences, correlation coefficients and Euclidean distance algorithms was examined for hierarchical clustering. R data analysis software used for the analysis and "as.dist" function was used to assign the correlation values to be "distances". Details about clustering algorithms are represented below.

2.1 Longest Common Subsequences (LCS)

The longest common subsequence is a method for measuring similarity between subsequences. This method is uses information contained in the Longest Common Subsequences as an indication of similarity (Wang, 2007).

LCS uses dynamic programming algorithm and figure out how well the two flows can match one another. For example, here are two flows having the same last element: (BANANA) and (ATANA). Omit the same last element. Rerun the procedure as far as you find no common last element. The omitted sequence will be (ANA) (Wikipedia). LCSS method allows to match some elements which are unmatched in Euclidean and DTW algorithms. Moreover, LCSS allows more efficient approximate computation (Hamasuna et al., 2012).

2.2 Correlation

Correlation is one of the well-known used similarity measures like Euclidean distances. However, literature suggests that there is a very little work has been done about using correlation as a dissimilarity measure. Correlation has numerous types like normalized correlation, Pearson correlation coefficient and cosine similarity. This is used to describe similarities between two vectors. This is used in pattern recognition, multivariate statistics and data mining (Ma et al., 2007).

For time series data comparison where trends and evolutions are intended to be evaluated, or when the shape formed by the ordered succession of features is relevant, similarity measures based on Pearson's correlation have also been utilized (Iglesias & Kastner, 2013). The dissimilarity forms of correlation coefficients are represented below (Glynn, 2005):

Dissimilarity = 1-Correlation, Dissimilarity = (1-Correlation)/2, Dissimilarity = 1-Abs (Correlation), Dissimilarity = Sqrt (1-Correlation^2)

2.3. *Euclidean Distance*

Euclidean distance is calculated by using Ptyhagonean formula. Literature suggests this as Pythagonean metric (Deza & Deza, 2009). The Euclidean distance between p and q is the length of the line connecting these two points (p.q). In Cartesian coordinates, if p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n) are two points in Euclidean n-space, then the distance from p to q, or from q to p is given below (Deza & Deza, 2009):

$$d(p,q)=d(q,p)=\sqrt{(q1-p1)^2+(q2-p2)^2+\ldots\ldots\ldots\ldots\ldots\ldots+(qn-pn)^2}$$
$$=\sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}$$

3. Results

Table 1 shows 34 OECD country averages about out of pocket health expenditures as a percentage of private health expenditure from 1995 to 2011. In this table it is seen that Iceland has the highest mean value 95.97 (±3.93) and United States has the lowest 24.36 (±1.65) one. According to the OECD statistics in Iceland 80.4% of health spending was funded by public sources in 2010 (OECD, 2012). This table shows that United States spends less than any other country. This can be explained by much more spending on private health insurance and less on public health (Berdahl et al., 2013).

Table 1. Descriptive Statistics

Country	Min.	Median	Mean	Sd.
Australia	47.01	56.44	56.03	±4.08
Austria	56.99	63.94	64.39	±3.55
Belgium	76.88	80.32	81.94	±4.13
Canada	48.61	49.53	51.42	±2.98
Chile	63.04	66.22	66.63	±2.24
Czech Republic	84.29	97.67	94.98	±5.74
Denmark	88.05	90.49	90.44	±1.44
Estonia	77.70	90.08	90.19	±6.40
Finland	74.48	76.00	76.84	±1.86
France	31.00	32.54	33.41	±1.92
Germany	50.53	51.39	51.75	±0.90
Greece	94.15	94.54	94.93	±0.58
Hungary	73.89	88.19	85.72	±8.87
Iceland	91.71	92.92	95.97	±3.93
Ireland	32.81	49.15	48.05	±11.73
Israel	58.67	63.02	65.57	±6.29
Italy	84.88	87.81	88.19	±1.71
Japan	78.85	82.01	81.68	±1.58
Korea Republic	76.85	80.50	80.20	±2.19
Luxembourg	71.85	77.64	80.81	±9.36
Mexico	92.02	94.70	94.49	±1.57
Netherlands	20.83	24.10	28.26	±7.01
New Zealand	62.58	69.43	68.46	±3.43
Norway	94.54	95.46	95.38	±0.47
Poland	79.42	88.20	90.59	±8.60
Portugal	63.99	73.81	73.07	±3.95
Slovak Republic	72.22	89.22	88.55	±10.71
Slovenia	39.18	46.45	45.96	±3.16
Spain	72.60	77.23	79.31	±4.23
Sweden	86.59	88.68	91.17	±4.73
Switzerland	69.18	73.80	73.39	±2.03
Turkey	64.41	69.40	75.65	±14.17
United Kingdom	52.68	54.09	56.21	±4.38
United States	20.86	24.07	24.36	±1.65

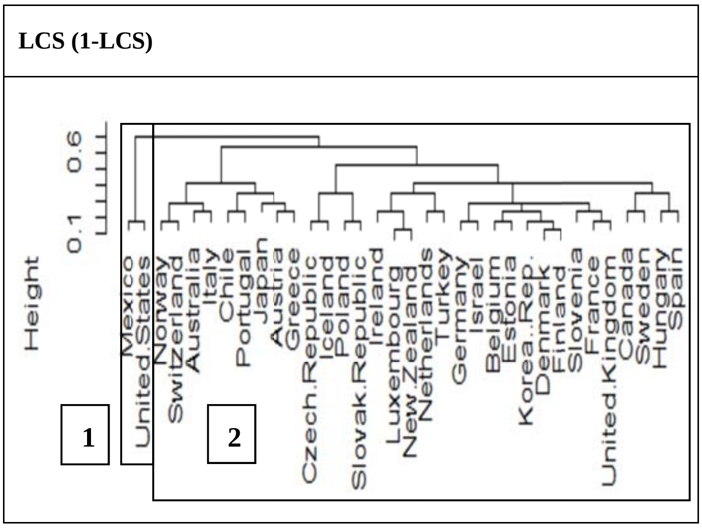


Figure 1. Hierarchical Clustering Using LCS (1-LCS) on Raw Data

Figure 1 shows hierarchical clustering using LCS (1-LCS), on raw data. In this figure the dendrogram illustrates that there are two clusters. In the first cluster there are two countries which are Mexico and United States. All other 32 OECD countries are in the second cluster. The dynamic programming algorithm of the longest common subsequence-LCS is used to determine QSI as index of similarity of the patterns (Boogaart et al., 2013). On the other hand, according to OECD reports public sector is the main source of health funding in all OECD countries except United States and Mexico (OECD 2016). Quality Similarity Index (QSI) for Mexico and United States is 0.76. This means that these two countries have similar out of pocket health expenditures and they have similar country policy about health funding (see Appendix).

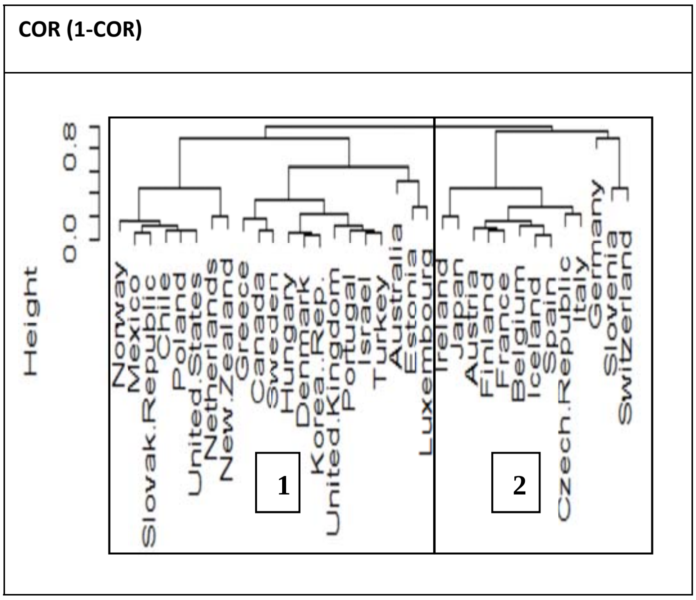


Figure 2. Hierarchical Clustering Using Cor (1-Cor) on Raw Data

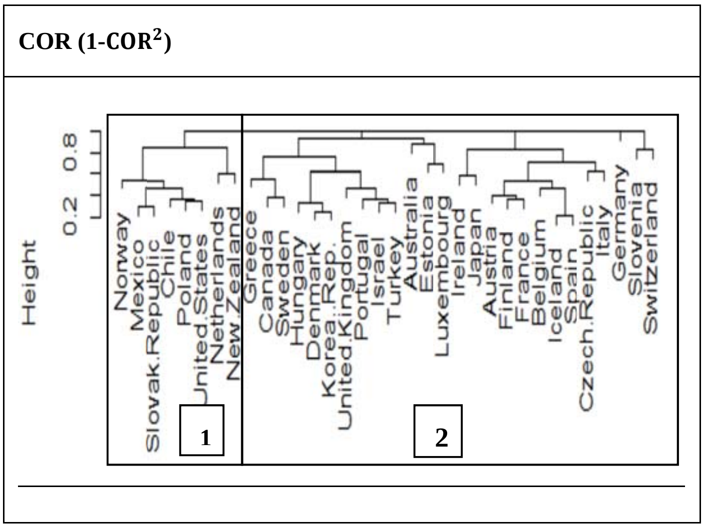


Figure 3. Hierarchical Clustering Using Cor (1-COR^2) on Raw Data

Figure 2 and Figure 3 shows hierarchical clustering results based on correlation coefficients (1-cor, 1-cor^2). Unfortunately, these two dendrograms are not very effective to determine different groups in the clusters.

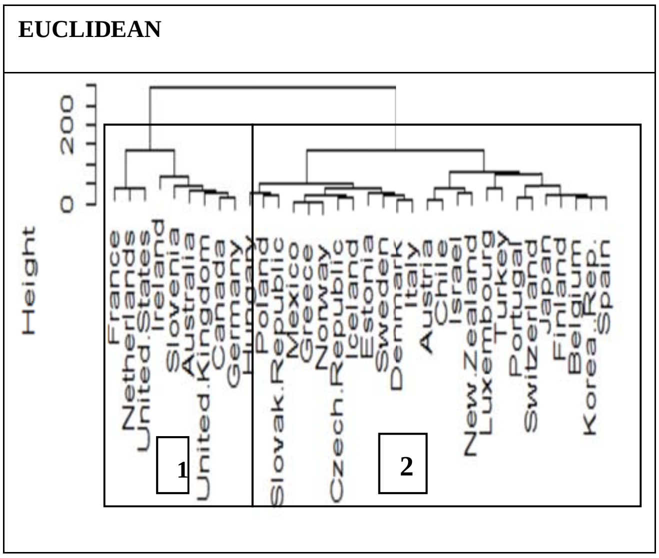


Figure 4. Hierarchical Clustering Using Euclidean Distance on Raw Data

Figure 4 illustrates hierarchical clustering results of OECD countries based on Euclidean distance measure. It can be seen that first cluster consists of 10 countries which are have high out of pocket health expenditure averages includes US. Moreover, the second one includes countries which are have low level of out of pocket health expenditures. Countries which are in the first cluster are France, Netherlands, United States, Ireland, Slovenia, Australia, United Kingdom, Canada, and Germany. These countries have different out of pocket health expenditure patterns (see Appendix) but their out of pocket health expenditure averages for the years between 1995 and 2011 is similar. Countries in the second cluster have low level of expenditures. Iceland is in the second cluster and this has the highest 95.97 (±3.93) mean value of out of pocket health expenditure (see Table 1). To conclude mean values of OECD countries health expenditures become a determinator factor of this hierarchical clustering method.

4. Conclusion

Health systems improve life-enhancing interventions and provide care for the people who demand them. If health systems are powerless, the power of care interventions is weakened. Resource allocation is a central part of the decision making process but there is a scarcity of resources allocation in health care all over the world (WHO, 2000). To improve performance of health care systems to determine better health system financing policy is essential. Not only improving accessibility but also protecting households from financial catastrophe, by reducing out of pocket health spending is necessary (Yardim et al., 2010). According to reports prepared by intergovernmental organizations like WB, OECD and WHO the share of out of pocket health expenditures among health spending is higher in most of OECD countries. To bear this in mind that this study is focused on grouping OECD countries according to their out of pocket health expenditures.

The results of this study which was aimed clustering OECD countries according to their out of pocket health expenditures for the years between 1995 and 2011 shows that; determinator factors for identifying clusters were; countries out of pocket health expenditure trends, average out of pocket health expenditure ratios and countries health funding policies.

In this study longest common subsequence (LCS), correlation (Cor) and Euclidean distance (EU) methods were used for hierarchical clustering. Previous studies will be use different clustering algorithms and they can compare them. In the future these applications can be extract knowledge from data through rule extraction. The direction of further research activities will help health policy makers to discover health expenditure patterns and to explore similarities between developed countries.

Acknowledgements

This study was supported by The Turkish Scientific and Technological Research Council of Turkey (TUBITAK).

References

Anderson, G.F., & Frogner, B.K. (2008). Health spending in OECD countries: obtaining value per dollar. *Health Affairs*. *27*(6), 1718-1727. http://dx.doi.org/10.1377/hlthaff.27.6.1718

Baltagi, B.H., & Moscone, F. (2010). Health care expenditure and income in the OECD reconsidered: Evidence from panel data. *Economic Modelling*, *27*(4), 804-811. http://dx.doi.org/10.1016/j.econmod.2009.12.001.

Berdahl, T.A., Friedman, B.S., McCormick, M.C., & Simpson, L. (2013). Annual report on health care for children and youth in the United States: Trends in racial/ethnic, income, and insurance disparities over time, 2002–2009. *Academic Pediatrics*, *13*(3), 191-203. http://dx.doi.org/10.1016/j.acap.2013.02.003

Boogaart, K.G., Rost, S., & Petzoldt, T. (2013). Package "qualV". P.15 http://cran.rproject.org/web/ packages/qualV/qualV.pdf (20 March 2016).

Das, R., Bhattacharyya, D.K., & Kalita, J.K. (2007). An effective dissimilarity measure for clustering gene expression time series data. *Biotechnology and Bioinformatics Symposium*, *Paper Identification Number* 56, 1-6.

Devaux M. (2015). Income-related inequalities and inequities in health care services utilisation in 18 selected OECD countries. *European Journal of Health Economics*, *16*(1), 21-33. http://dx.doi.org/10.1007/s10198-013-0546-4

Deza, M.M., & Deza, E. (2009). *Encyclopedia of Distances*. (1st Edition), Springer-Verlag Berlin Heidelberg, pp.291-305.

Glynn, E.F. (2005). Correlation "Distance" and hierarchical clustering. R Teach Notes and Gallery. Retrieved 23 April 2016 from http://research.stowers-institute .org/efg/R/ Visualization /cor-cluster/

Grabusts, P., & Borisov, A. (2009). Clustering methodology for time series mining. *Scientific Journal of Riga Technical University*, *40*(1), 81-86.

Hamasuna, Y., Endo, Y., & Miyamoto, S. (2012). On agglomerative hierarchical clustering using clusterwise tolerance based pairwise constraints. *Journal of Advanced Computational Intelligence Informatics*, *16*(1), 174-175. http://dx.doi.org/10.20965/jacii.2012.p0174

Huber, M., & Orosz, E. (2003). Health expenditure trends in OECD countries, 1990-2001. *Health Care Financing Review*, *25*(1), 1-22.

Iglesias, F., & Kastner, W. (2013). Analysis of similarity measures in time series clustering for the discovery of building energy patterns. *Energies*, *6*(2), 579-597. http://dx.doi.org/10.3390/en6020579.

Lette, M., Bemelmans, W.J.E., Breda, J., Slobbe, L.C.J., & Boshuizen, H.C. (2016). Health care costs attributable to overweight calculated in a standardized way for three European countries. *The European Journal of Health Economics, 17*(1), 61-69. http://dx.doi.org/10.1007/s10198-014-0655-8.

Liao, T.W. (2005). Clustering of time series data-a survey. *Pattern Recognition*, *38*, 1857-1874. http://dx.doi.org/10.1016/j.patcog.2005.01.025

Ma, Y., Lao, S., Takikawa, E., & Kawade, M. (2007). Discriminant analysis in correlation similarity measure space. *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR.

Moses H., Matheson D.H.M., Cairns-Smith S., George B.P., Pallisch C., & Dorsey E.R. (2015). The anatomy of medical research: US and international comparisons. *JAMA, 313*(2), 174-189. http://dx.doi.org/10.1001/jama.2014.15939.

OECD (2012) How does Iceland compare Retrieved 19 February 2016 from http://www.oecd.org/iceland/BriefingNoteICELAND2012.pdf

OECD (2015) Focus on Health Spending Retrieved 19 June 2016 from https://www.oecd.org/health/health-systems/Focus-Health-Spending-2015.pdf

OECD (2016) Health Data, Retrieved 25 April 2016 from http://www.oecd.org/els/health-systems/health-spending-continues-to-stagnate-says-oecd.htm

Ozcan, Y.A., & Khushalani, J. (2016). Assessing efficiency of public health and medical care provision in OECD countries after a decade of reform. *Central European Journal of Operations Research*, 1-19. http://dx.doi.org/10.1007/s10100-016-0440-0

Ozkan, I., & Türkşen, B.I. (2013). MiniMax e-stable cluster validity index for Type-2 fuzziness. *Information Sciences*, *184*(1), 64-74. http://dx.doi.org/10.1016/j.ins.2011.07.036.

Penders, Y.W.H. Rietjens, J. Albers, G. Croezen, S. & van den Block, L. (2016). Differences in out-of-pocket costs of healthcare in the last year of life of older people in 13 European countries. *Palliative Medicine*, 1-11. http://dx.doi.org/10.1177/0269216316647206.

Singhal, A., & Seborg, D.E. (2005). Clustering multivariate time-series data. *Journal of Chemometrics*, *19*(8), 427-438. http://dx.doi.org/10.1002/cem.945

The World Bank (WB) (2016), Out of Pocket Health Expenditure. Retrieved 15 March 2016 from http://data.worldbank.org/indicator/ SH.XPD.OOPC.ZS

Varabyova, Y., & Schreyögg, J. (2013). International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy*, *112*(1-2), 70-79. http://dx.doi.org/10.1016/j.healthpol.2013.03.003

Wang, H. (2007). Subsequence counting as a measure of similarity for sequences. *International Journal of Pattern Recognition Artificial Intelligence*, *21*(4), 745-758. http://dx.doi.org/10.1142/S021800140700565X

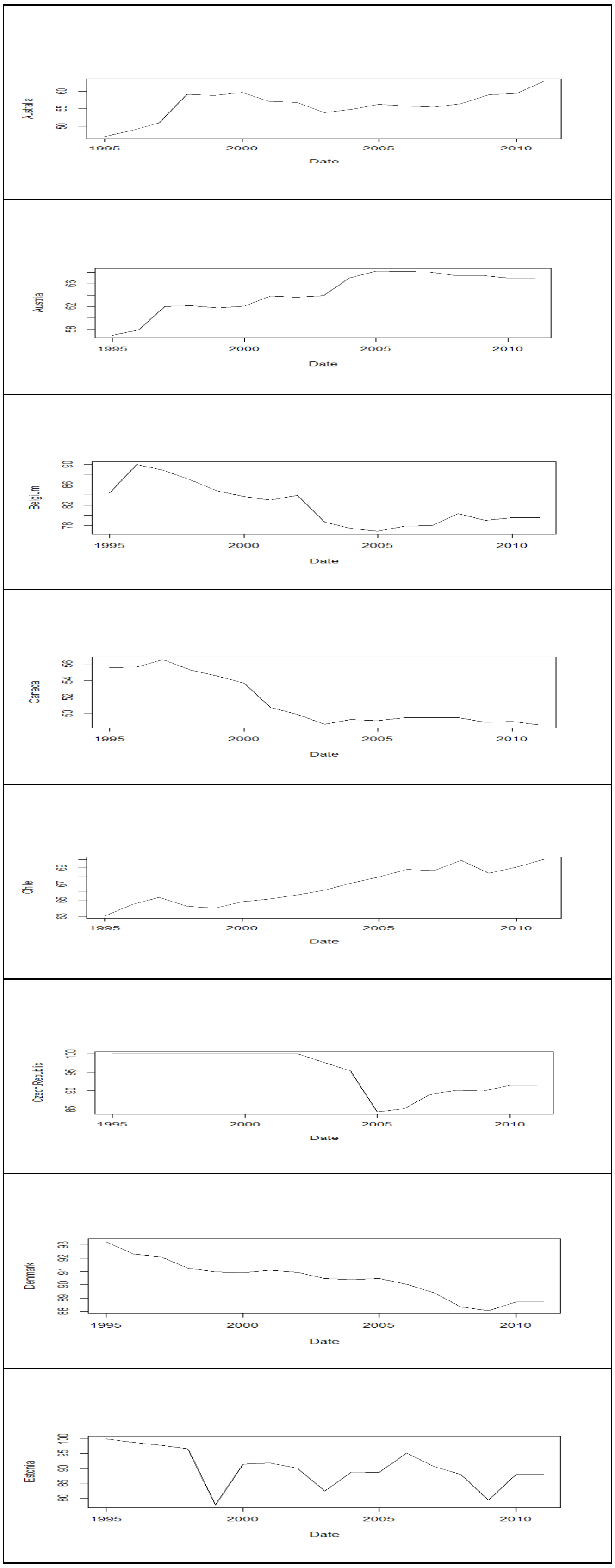
World Bank (WB). Retrieved 25 April 2016 from http://data.worldbank.org/indicator/ SH.XPD.OOPC.ZS?page=1

World Health Organization (WHO) (2000). The World Health Report 2000. Retrieved 20 March 2016 from http://www.who.int/whr/2000/en/whr00_en.pdf?ua

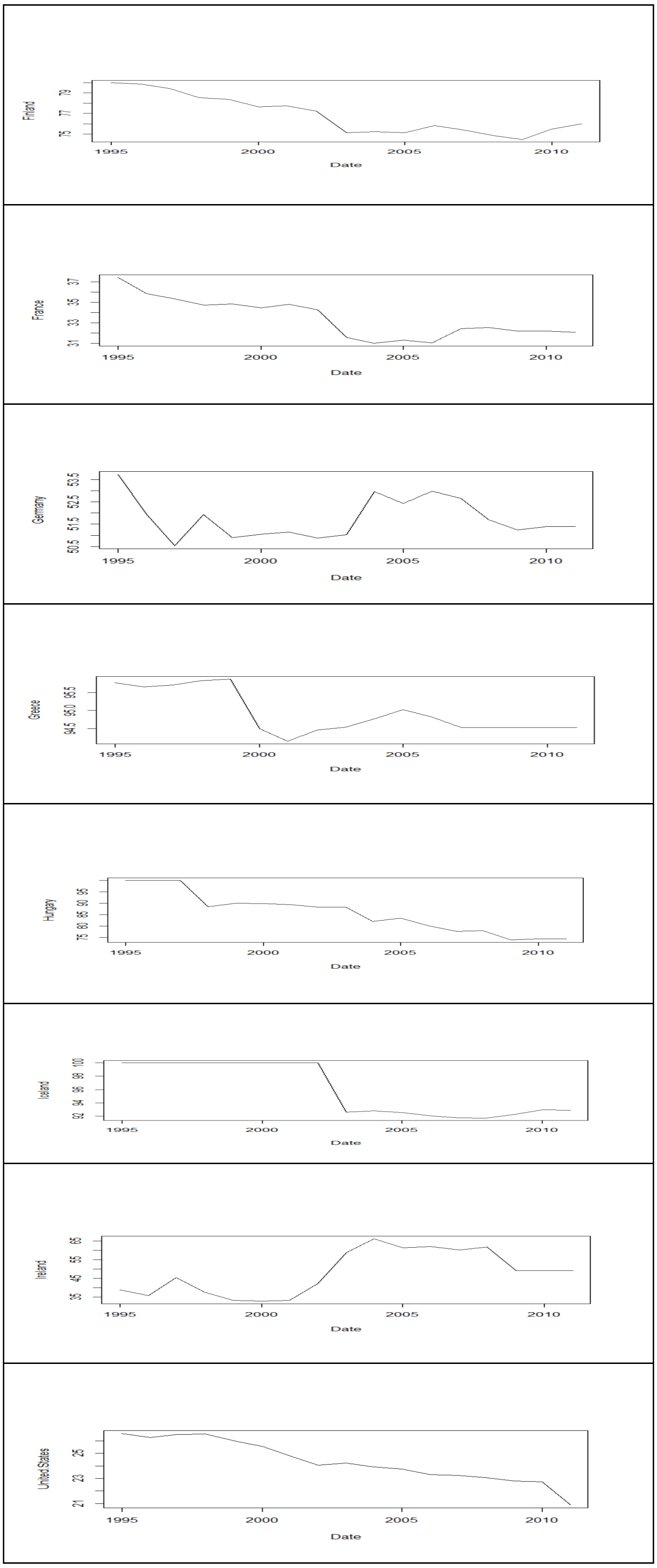
World Health Organization (WHO) (2005), Distribution of health payments and catastrophic expenditures methodology, Discussion Paper. Retrieved 20 March 2016 from http://www.who.int/health_financing/documents/dp_e_05_2-distribution_of_health_payments.pdf

Yardım, M.S., Cilingiroglu, N., & Yardım, N. (2010). Catastrophic health expenditure and impoverishment in Turkey. *Health Policy*, *94*(1), 26-33. http://dx.doi.org/10.1016/j.healthpol.2009.08.006.

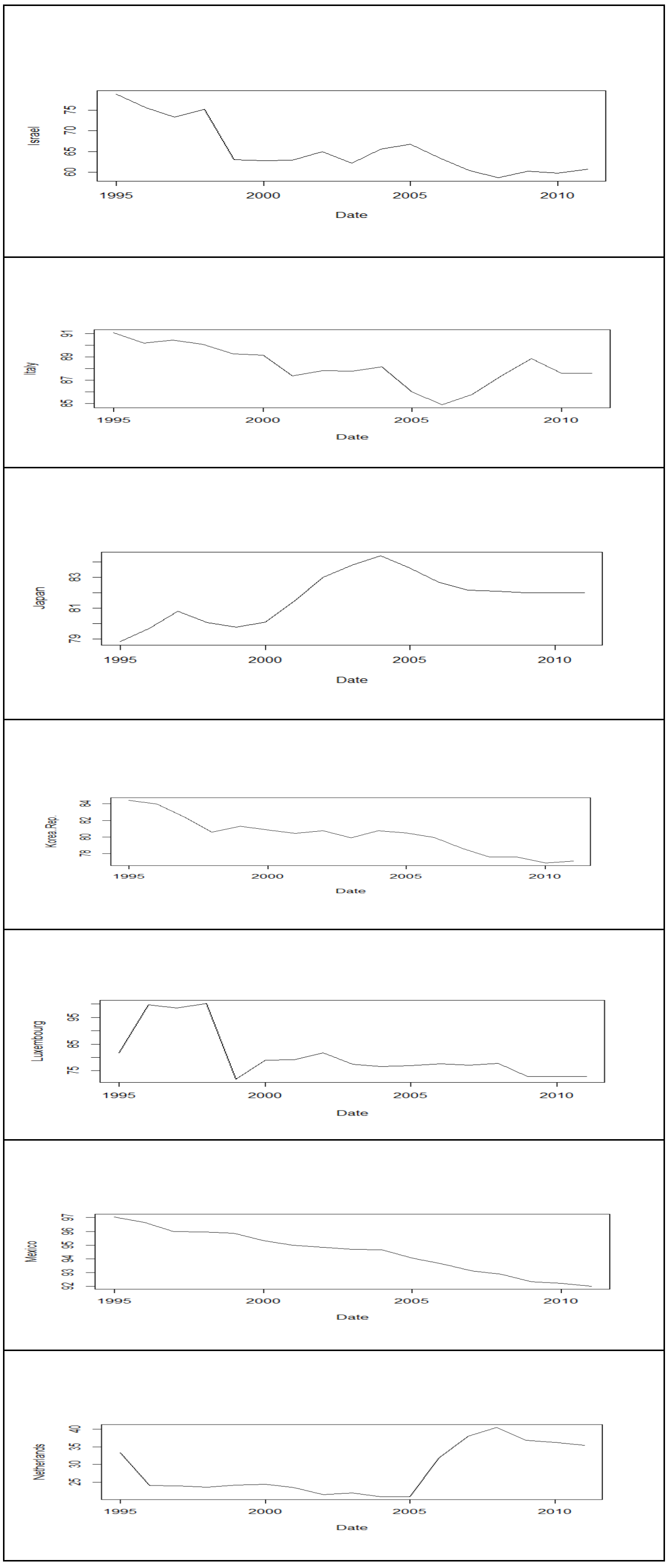
Appendix: Time Plots of 34 OECD Countries Out of Pocket Health Expenditure



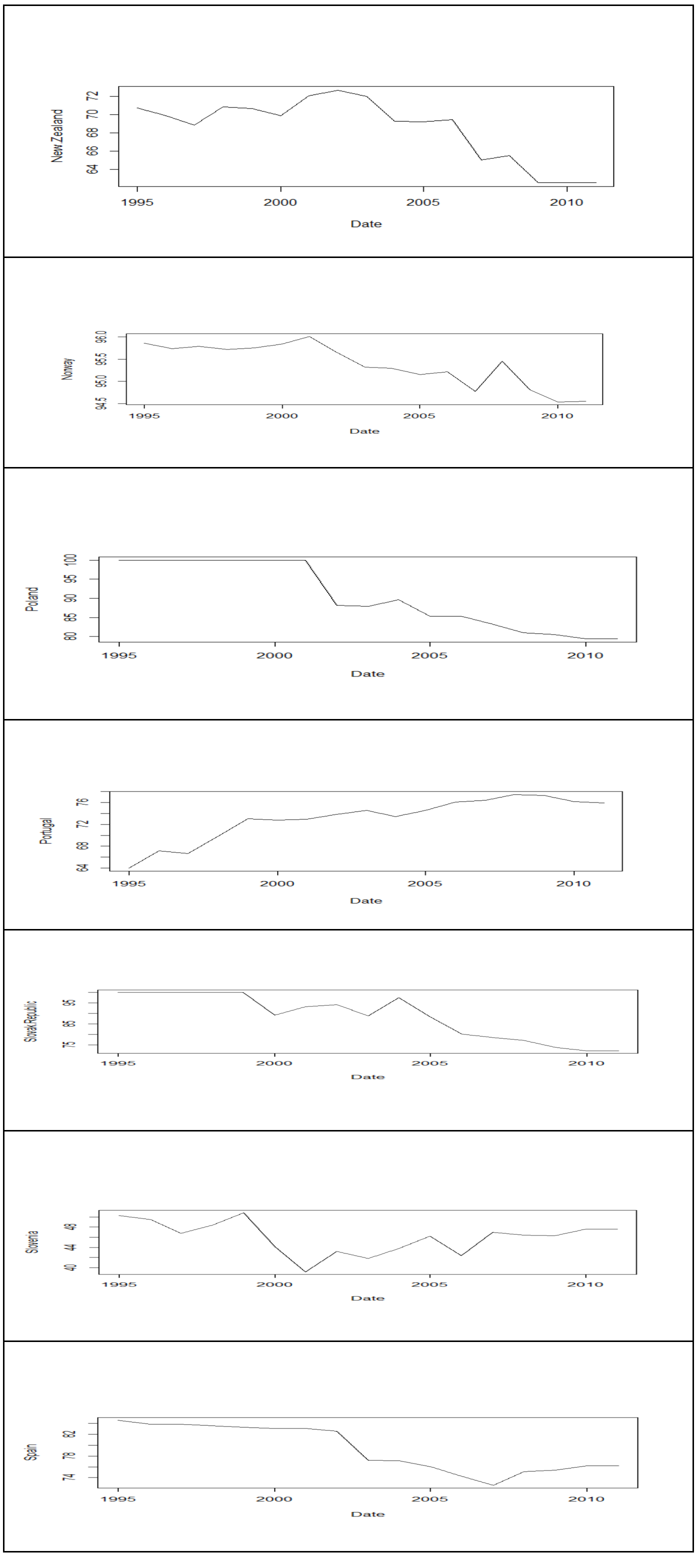
Appendix: Time Plots of 34 OECD Countries Out of Pocket Health Expenditure (Continues)



Appendix: Time Plots of 34 OECD Countries Out of Pocket Health Expenditure (Continues)



Appendix: Time Plots of 34 OECD Countries Out of Pocket Health Expenditure (Continues)



Appendix: Time Plots of 34 OECD Countries Out of Pocket Health Expenditure (Continues)