Using of fractional factorial design (r^{k-p}) in data envelopment analysis to selection of outputs and inputs

Hulya Bayrak[*], Oday Jarjies[†] and Kubra Durukan[‡§]

Abstract

Data envelopment analysis (DEA) is a linear programming based technique for measuring the relative performance of organisational units where the presence of multiple inputs and outputs makes comparisons difficult. We used, Morita and Avkiran propose after it has been developed an input-output selection method that uses fractional factorial design, which is a statistical approach to find an optimal combination. Energy efficiency and greenhouse gas emissions are closely linked in the last two decades. We demonstrate the proposed method using data that increase energy efficiency and heating gas emissions in the European Union (EU) countries.

---

[*]Gazi University, Science Faculty, Statistics Department, Ankara, TURKEY,
Email: `hbayrak@gazi.edu.tr`
[†]Department Operations Research and Technology Intelligence Mosul University, IRAQ
Email: `oday_alubade@yahoo.com`
[‡]Kirikkale University, Faculty of Arts and Sciences, Department of Statistics, 71100, Kampus, Yahsihan, Kirikkale - TURKEY,
Email: `kubraba@yahoo.com`
[§]Corresponding Author.

1. Introduction

Data envelopment analysis (DEA), introduced by Charnes, Cooper and Rhodes (CCR) [1], is a mathematical programming method for measuring the relative efficiency of decision-making units (DMUs) with multiple inputs and outputs. Most models DEA has the best performance and efficiency to determine the degree of expertise and decision-making units (DMUs). Differentiating efficient DMUs is an interesting research area. The original DEA method evaluates each DMU against a set of efficient DMUs and cannot identify which efficient DMU is a better option with respect to the inefficient DMU. This is because all efficient DMUs have an efficiency score of one. Authors have proposed methods for ranking the best performers, for instance using super-efficiency DEA model.

In this paper, in order to rank DMUs, we use the evaluation contexts that are obtained by partitioning the set of DMUs into several levels of efficiency, and rank all DMUs with two criteria: the high and low performers. The influence of all DMUs, both efficient and inefficient, in ranking is this method's preference.

1.1. Data Envelopment Analysis. Consider n decision making units ($DMUj, j = 1, \ldots, n$) in which each DMU consumes input levels $x_{ij}(i = 1, ..., m)$ to produce output levels $y_{rj}(j = 1, ..., s)$. Suppose that $x_j = (x_{1j}, ..., x_{mj})^T$ and $y_j = (y_{1j}, ..., y_{sj})^T$ are the vectors of inputs and outputs values respectively, the relative efficiency score of the $DMU_O, O \in \{1, ..., n\}$ is obtained from the following model which is called input-oriented CCR envelopment model [7, 8, 9]

$$(1.1) \qquad \theta_O^* = \min \theta$$
$$s.t. \sum_j \lambda_j x_{ij} \leq \theta x_{io}, i = 1, ..., m$$
$$\sum_j \lambda_j y_{ij} x_{ij} \geq y_{ro}, r = 1, ..., s$$
$$\lambda_j \geq 0, j = 1, ..., n$$

This model is an input oriented constant returns to scale (CRS) model. The efficiency of DMUo is determined from efficiency score θ_O^* and its slack values. If and only if $\theta_O^* = 1$ there is no slack, DMUo is said to be efficient. If and only if $\theta_O^* = 1$ there are non-zero slacks, DMUo is inefficient and we can called it a weak-efficient. The weak-efficient DMUs and efficient DMUs comprise the efficient frontier [6].

Morita and Haba in a previous study, select the output of the of preference between the two groups based on public information and previous experience has nothing to do with data where they are exploiting the experience of planning two-level orthogonal and optimal variables can be found statistically. On the other hand, Ediridsinghe and Zhang proposed DEA generalized approach to determine the input and output by maximizing the correlation coefficient between the DEA and the result of external performance indicator. Morita and Avkiran propose an input output selection method that uses diagonal layout experiments and demonstrate the proposed method using financial statement data from NIKKEI 500 index. They utilize a two-step heuristic algorithm that combines random sampling and local search to find an optimal combination of inputs and outputs [5, 4].

In this paper, we show the method of selection of inputs and outputs based on an analysis using the Mahalanobis distance of difference between the two group of data. We use 3-level orthogonal layout experiment to find a suitable combination of inputs and outputs, where trials are independent of each other.

2. 3^{k-p} Fractional Factorial Designs and Selecting Input and Output Variables:

The whole point of looking at this structure is because sometimes we want to only conduct a fractional factorial. We sometimes can't afford 27 runs. Often we can only afford a fraction of the design. So, let's construct a 3^{3-1} design which is a 1/3 fraction of a 3^3 design. In this case, $N = 3^{3-1} = 3^2 = 9$, the total number of runs. This is a small, compact design [2].

We again start out with a 3^3 design which has 27 treatment combinations and assign them to 3 blocks. What we want to do in this part, going beyond the 3^2 design, is to describe the ANOVA for this 3^3 design. Then we also want to look at the connection between confounding in blocks and 3^{k-p} fractional factorials, See Appendix 1.

2.1. 3-level Full Factorial Designs and Other Factorials. The 3-level design is written as a 3^k factorial design. It means that k factors are considered, each at 3-levels. These are (usually) referred to as low, intermediate and high levels. These levels are numerically expressed as $0, 1$ and 2. One could have considered the digits $-1, 0$ and $+1$, but this may be confusing with respect to the 2-level designs since 0 is reserved for center points. Therefore, we will use the $0, 1, 2$ scheme. The reason that the 3-level designs were proposed is to model possible curvature in the response function and to handle the case of nominal factors at 3-levels. A third level for a continuous factor facilitates investigation of a quadratic relationship between the response and each of the factors [2].

Unfortunately, the 3-level design is prohibitive in terms of the number of runs, and thus in terms of cost and effort. For example a 2-level design with center points is much less expensive while it still is a very good (and simple) way to establish the presence or absence of curvature. Table 1 shows us the difference between full factorial designs and other factorials.

Table 1. 3-level designs

Factors	3	4	**5**	6	7
Full	27	81	**243**	729	2187
1/3	9	27	**81**	243	729
1/9	3	9	**27**	81	243
1/27	NA	NA	**9**	27	81

3. Mahalanobis Distance

In statistics, Mahalanobis distance (MD) is a distance measure introduced by P.C. Mahalanobis in 1936. It is based on correlations between variables by which different patterns can be identified and analyzed. It gauges similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant. In other words, it is a multivariate effect size [3].

Formally, the MD of a multivariate vector $x = (x_1, x_2, ..., x_N)^T$ from a group of values with mean vector $\mu = (\mu_1, \mu_2, ..., \mu_N)^T$ and covariance matrix Σ is defined as [4, 12]:

$$(3.1) \qquad D_M(x) = \sqrt{(x - \mu)^T \, \Sigma^{-1} \, (x - \mu)}$$

Equation (3.1) is rewritten for the sample following as

$$\hat{D}_M(x) = \sqrt{(x - \bar{x})^T S^{-1} (x - \bar{x})}$$

where the mean vector and covariance matrix of the sample are given as $\bar{x}$ and S respectively.

MD is widely used in cluster analysis and classification techniques. It is closely related to Hotelling's T-square distribution used for multivariate statistical testing and Fisher's Linear Discriminant Analysis that is used for supervised classification [12].

In order to use the MD to classify a test point as belonging to one of N classes, one first estimates the covariance matrix of each class, usually based on samples known to belong to each class. Then, given a test sample, one computes the MD to each class, and classifies the test point as belonging to that class for which the MD is minimal [3].

MD and leverage are often used to detect outliers, especially in the development of linear regression models. A point that has a greater the MD from the rest of the sample population of points is said to have higher leverage since it has a greater influence on the slope or coefficients of the regression equation. MD is also used to determine multivariate outliers. Regression techniques can be used to determine if a specific case within a sample population is an outlier via the combination of two or more variable scores. A point can be a multivariate outlier even if it is not a univariate outlier on any variable [7, 8].

3.1. MD Threshold Selection. The MD threshold is another important element of prognostics analysis. An MD threshold value which is either too large or too small leads to false negatives or false positives, respectively. In this study, we consider the distance of one-dimensional variables, where MD coincides with the Welch statistics [5]. The Welch statistics is given as

$$(3.2) \qquad \hat{d} = \frac{\bar{x}_h - \bar{x}_l}{\sqrt{\frac{S_h^2}{n_h} + \frac{S_l^2}{n_l}}}$$

where $\bar{x}_h$, S_h^2 and n_h are the sample mean, sample variance and sample size of high group, respectively. Also $\bar{x}_l$, S_l^2 and n_l are the sample mean, sample variance and sample size of low group, respectively.

For example, in run No. 5, x_1, x_{11}, x_{12} variables are selected as an input, variables $x_2, x_3, x_4, x_5, x_6, x_7$ are selected as an output; and variables x_8, x_9, x_{10} are not selected as an input or an output. Based on the fractional factorial design in Appendix 1, we calculate the efficiency scores and MD between the two groups using selected inputs and outputs. Where "1" means that the variable is selected as an input, "2" means that the variable is selected as an output, and "3" means that the variable is not selected.

The ANOVA table for the fractional factorial design appears in Table 2. The sum of squares and the degrees of freedom are given as

$$(3.3) \qquad S_T = \sum_{i=1}^{27} \left(\hat{d}_i - \overset{\bar{\wedge}}{d} \right)^2, \qquad df_T = 26$$

$$(3.4) \qquad S_i = 3 \left[\overset{\bar{\wedge}}{d}^2 (x_i = 1) + \overset{\bar{\wedge}}{d}^2 (x_i = 2) + \overset{\bar{\wedge}}{d}^2 (x_i = 3) \right] - 27 \overset{\bar{\wedge}}{d}^2, \qquad df_i = 2, \quad i = 1, ..., 12$$

$$(3.5) \qquad S_E = S_T - (S_1 + S_2 + ... + S_{12})$$

Table 2. ANOVA table for fractional factorial design of 3^{12-9}

Variables	SS	Df	MS	F Statistics
X_1	S_1	2	$V_1 = S_1/2$	V_1/V_E
X_2	S_2	2	$V_2 = S_2/2$	V_2/V_E
X_3	S_3	2	$V_3 = S_3/2$	V_3/V_E
X_4	S_4	2	$V_4 = S_4/2$	V_4/V_E
X_5	S_5	2	$V_5 = S_5/2$	V_5/V_E
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
X_{12}	S_{12}	2	$V_{12} = S_{12}/2$	V_{12}/V_E
Error	S_E	2	$V_E = S_E/2$	
Total	S_T	26		

where $\overline{\hat{d}}(x_i = 1)$ is the mean of the Mahalanobis distances observed when $(x_i = 1)$. The null hypothesis that the candidate has no effect as an input or output is tested by using the F statistics

$$(3.6) \qquad F = \frac{S_i/df_i}{S_E/df_E}$$

and hypotesis tests is as following:

H0: The variable candidate has no effect on output and input.

H1: The variable candidate has effect on output and input.

This results in the optimal combination of input and output variables. The following is a summary procedure for the selection of variables.

Step 1. Choose a list of data envelope (DEA), which contains the input and output variables are possible.

Step 2. The use of external standards to distinguish between the performance of the two groups. For example, the high and low performance.

Step 3. To create a table perpendicular to try to set the input and output variables that are not determined.

Step 4. Calculate MD between the two groups using Welch statistics.

Step 5. Determine the optimal mix of input and output variables based on the results Analysis of variance.

Step 6. Determine the optimal variables are statistically significant either input or output using sum of MD.

Step 7. We use DEA model (1) with the data that have been selected from the output and input.

4. A Case Study Using Greenhouse Gas Emissions Intensity of Energy Consumption Data

We used the greenhouse gas intensity of energy consumption that is the ratio between energy-related greenhouse gas emissions (carbon dioxide, methane and nitrous oxide) and gross inland energy consumption for EU countries.

There are key factors leading to greenhouse emissions: Electricity production, Transportation, Industry, Commercial and Residential, Agriculture and Land Use and Forestry [10, 11].

The table in the Appendix 2 shows a part of the data set all variables have large rangs. In Step 1 the following twelve variables are collected to evaluate the managerial performance, that is, the standard deviation is greater than the mean.

    A. Total emissions.
    B. Total net emissions.
    C. Energy.
    D. Energy industries.
    E. Manufacturing industries and construction.
    F. Transport.
    G. Road transportation.
    H. Other sectors.
    I. Industrial processes.
    J. Solvent and other product use.
    K. Agriculture.
    L. Waste.

In Step 2, we construct two groups, high-performers and low-performers, the table in the Appendix 3 shows the mean and standard deviation for each variable. When we select the variables to capture the difference between high-performers and low-performers, we choose a variable with a large difference between these two groups. MD between the 15 high cuntery and 15 low cuntery for each variable is also shown in the Appendix 3, where we find that (A),(B),(C),(E), (I) and (L) have a large d and may be intuitively selected as inputs or outputs.

In Step 3, we assign 12 factors into a 3-level orthogonal layout, where at least 27 runs are required. That is, we utilize the fractional factorial design 3^{12-9}. Table 3 shows the selected variable combinations for efficiency score calculation. The MD for each experiment is calculated in Step 4, which is also shown in the last column of Table 3.

Table 3. Selected inputs and outputs and MD

Runs	A	B	C	D	E	F	G	H	I	J	K	L	Selected Input	Selected Output	Not Selected	$\hat{d}$
1	1	1	1	1	1	1	1	1	1	1	1	1	A,B,C,D,E,F,G,H,I,J,K,L	None	None	2.44
2	1	1	1	1	2	2	2	2	2	2	2	2	A,B,C,D	E,F,G,H,I,J,K,L	None	1.19
3	1	1	1	1	3	3	3	3	3	3	3	3	A,B,C,D	None	E,F,G,H,I,J,K,L	1.4
4	1	2	2	2	1	1	1	2	2	2	3	3	A,E,F,G	B,C,D,H,I,J	K,L	0
5	1	2	2	2	2	2	2	3	3	3	1	1	A,K,L	B,C,D,E,F,G	H,I,J	-0.13
6	1	2	2	2	3	3	3	1	1	1	2	2	A,H,I,J	B,C,D,K,L	E,F,G	-0.35
7	1	3	3	3	1	1	1	3	3	3	2	2	A,E,F,G	K,L	B,C,D,H,I,J	1.18
8	1	3	3	3	2	2	2	1	1	1	3	3	A,H,I,J	E,F,G	B,C,D,K,L	0.55
9	1	3	3	3	3	3	3	2	2	2	1	1	A,K,L	H,I,J	B,C,D,E,F,G	0.98
10	2	1	2	3	1	2	3	1	2	3	1	2	B,E,H,K	A,C,F,I,L	D,G,J	-0.26
11	2	1	2	3	2	3	1	2	3	1	2	3	B,G,J	A,C,E,H,K	D,F,I,L	-0.15
12	2	1	2	3	3	1	2	3	1	2	3	1	B,F,I,L	A,C,G,J	D,E,H,K	-0.46
13	2	2	3	1	1	2	3	2	3	1	3	1	D,E,J,L	A,B,F,H	C,G,I,K	-1.11
14	2	2	3	1	2	3	1	3	1	2	1	2	D,G,I,K	A,B,E,J,L	C,F,H	-0.87
15	2	2	3	1	3	1	2	1	2	3	2	3	D,F,H	A,B,G,I,K	C,E,J,L	-0.7
16	2	3	1	2	1	2	3	3	1	2	2	3	C,E,I	A,D,F,J,K	B,G,H,L	0.03
17	2	3	1	2	2	3	1	1	2	3	3	1	C,G,H,L	A,D,E,I	B,F,J,K	-0.24
18	2	3	1	2	3	1	2	2	3	1	1	2	C,F,J,K	A,D,G,H,L	B,E,I	-0.21
19	3	1	3	2	1	3	2	1	3	2	1	3	B,E,H,K	D,G,J	A,C,F,I,L	0.59
20	3	1	3	2	2	1	3	2	1	3	2	1	B,F,I,L	D,E,H,K	A,C,G,J	0.55
21	3	1	3	2	3	2	1	3	2	1	3	2	B,G,J	D,F,I,L	A,C,E,H,K	0.67
22	3	2	1	3	1	3	2	2	1	3	3	2	C,E,I	B,G,H,L	A,D,F,J,K	0.02
23	3	2	1	3	2	1	3	3	2	1	1	3	C,F,J,K	B,E,I	A,D,G,H,L	-0.28
24	3	2	1	3	3	2	1	1	3	2	2	1	C,G,H,L	B,F,J,K	A,D,E,I	-0.05
25	3	3	2	1	1	3	2	3	2	1	2	1	D,E,J,L	C,G,I,K	A,B,F,H	-0.8
26	3	3	2	1	2	1	3	1	3	2	3	2	D,F,H	C,E,J,L	A,B,G,I,K	-0.08
27	3	3	2	1	3	2	1	2	1	3	1	3	D,G,I,K	C,F,H	A,B,E,J,L	-0.65

Table 4 shows the analysis of variance for the data in Table 3, where we have pooled the negligible variables into the residual (Step 5). The level of significance is shown as the p value, where we find four variables (A,B,C, F, H and K) significant at the 5% level and their p values are very low, we leave them in the analysis for illustrative purposes.

Table 4. Table of ANOVA

Variables	Sum of Squares	Degrees of Freedom	Mean Squares	F Statistics	p value
A	4.32	2	2.16	11.37**	0.0837
B	3.78	2	1.39	7.32**	0.1265
C	1.58	2	0.79	4.16**	0.2109
F	1.90	2	0.95	5.00**	0.1790
H	1.20	2	0.60	3.16**	0.2662
K	1.40	2	0.70	3.68**	0.2340
Error	2.68	14	0.19		
Total	16.86	26			

Step 6, the final step in our procedure, generates Table 5 which shows the sum of MD for each variable at each level in Table 3. For example, when variable A is selected as an input, the sum of MD is 7.26, and when variable A is selected as an output, $\hat{d}$ is -3.97, it should be selected as an input. Maxima are indicated in bold font in Table 5. Thus we select four input (A) Total emissions, (B) Total net emissions, (C) Energy and (F) Transport and two outputs, namely, (H) Other sectors and (K) Agriculture.

Step 7, we run the DEA model (1) using this inputs and outputs combination.

Table 5. The sum of MD

Variables	Selected as Input	Selected as Output
A	7.26	-3.97
B	5.97	-3.47
C	4.30	-2.88
F	2.44	0.24
H	1.90	2.62
K	1.61	1.90

Note, we got 80% of the major factors leading to emissions of greenhouse gases That was previously displayed.

5. Conclusion

It is possible to attempt more than fractional factorial design at level 3 for example Latin square design or partial design. The MD and ANOVA was used to distinguish between the two groups after selecting the input and output from ANOVA results note it was investigating maximum MD between the two groups we demonstrate the effectiveness of this new approach using a case study with any DEA can set inputs and outputs and measuring the efficiency of performance that can effectively distinguish between groups of high and low performance.

Situation as you know it can not always be perfect, but is close to ideal combination that have been obtained are a limited number of 27 trials experience It can experimentation on a larger number of factors and a larger number of experiments.

References

[1] Charnes, A., Cooper, W.W. and Rhodes, E. Measuring the efficiency of decision making units. European Journal of Operational Research Vol. 2, 429-441, 1978.

[2] Box, G.E., Hunter, J.S., Hunter, W.G. Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition. Wiley, 2005.

[3] Jahanshahloo, G.R. and Hosseinzadeh, L.F. Introduction to Data Envelopment Analysis, (forth coming) 2004.

[4] Taguchi, G., Wu, Y. and Chowdhury, S. Mahalanobis-Taguchi System. McGraw-Hill Professional, 2000.

[5] Morita, H. and Haba, Y. Variable selection in data envelopment analysis based on external information. Proceedings of the eighth Czech-Japan Seminar 2005 on Data Analysis and Decision Making under Uncertainty, 181–187, 2005.

[6] Morita, H. and Avrikan, N.K. Selecting Input and Output Variables in Data envelopment analysis by designing statictcal experiments. Journal of the Operations Research Society of Japan, 52(2), 163-173, 2009.

[7] Edirisinghe, N.C.P. and Zhang, X. Generalized DEA model of fundamental analysis and its application to portfolio optimization. Journal of Banking & Finance, 31, 3311–3335, 2007.

[8] Seiford, L.M. and Zhu, J. Context-dependent data envelopment analysis: measuring attractiveness and progress. Omega, 31, 397–408, 2003.

[9] Cooper, W.W, Seiford, L.M. and Tone, K. Data Envelopment Analysis. Kluwer Academic Publishers, 2000.

[10] http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes.

[11] http://www.eia.gov/countries/data.cfm.

[12] http://en.wikipedia.org/wiki/Mahalanobis_distance.

Appendix 1. Fractional factorial design for 3^{12-9} twelve factors at three levels (27 Runs)

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	Selected Input	Selected Output	Not selected	D
1	1	1	1	1	1	1	1	1	1	1	1	1	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},x_{12}$	None	None	D_1
2	1	1	1	1	2	2	2	2	2	2	2	2	x_1,x_2,x_3,x_4	$x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},x_{12}$	None	D_2
3	1	1	1	1	3	3	3	3	3	3	3	3	x_1,x_2,x_3,x_4	None	$x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},x_{12}$	D_3
4	1	2	2	2	1	1	1	2	2	2	3	3	x_1,x_5,x_6,x_7	$x_2,x_3,x_4,x_8,x_9,x_{10}$	x_{11},x_{12}	D_4
5	1	2	2	2	2	2	3	3	3	1	1	1	x_1,x_{11},x_{12}	x_2,x_3,x_4,x_5,x_6,x_7	x_8,x_9,x_{10}	D_5
6	1	2	2	2	3	3	3	1	1	1	2	2	x_1,x_8,x_9,x_{10}	$x_2,x_3,x_4,x_{11},x_{12}$	x_5,x_6,x_7	D_6
7	1	3	3	3	1	1	1	3	3	3	2	2	x_1,x_5,x_6,x_7	x_{11},x_{12}	$x_2,x_3,x_4,x_8,x_9,x_{10}$	D_7
8	1	3	3	3	2	2	2	1	1	1	3	3	x_1,x_8,x_9,x_{10}	x_5,x_6,x_7	$x_2,x_3,x_4,x_{11},x_{12}$	D_8
9	1	3	3	3	3	3	3	2	2	2	1	1	x_1,x_{11},x_{12}	x_8,x_9,x_{10}	x_2,x_3,x_4,x_5,x_6,x_7	D_9
10	2	1	2	3	1	2	3	1	2	3	1	2	x_2,x_5,x_8,x_{11}	x_1,x_3,x_6,x_9,x_{12}	x_4,x_7,x_{10}	D_{10}
11	2	1	2	3	2	3	1	2	3	1	2	3	x_2,x_7,x_{10}	x_1,x_3,x_5,x_8,x_{11}	x_4,x_6,x_9,x_{12}	D_{11}
12	2	1	2	3	3	1	2	3	1	2	3	1	x_2,x_6,x_9,x_{12}	x_1,x_3,x_7,x_{10}	x_4,x_5,x_8,x_{11}	D_{12}
13	2	2	3	1	1	2	3	2	3	1	3	1	x_4,x_5,x_{10},x_{12}	x_1,x_2,x_6,x_8	x_3,x_7,x_9,x_{11}	D_{13}
14	2	2	3	1	2	3	1	3	1	2	1	2	x_4,x_7,x_9,x_{11}	$x_1,x_2,x_5,x_{10},x_{12}$	x_3,x_6,x_8	D_{14}
15	2	2	3	1	3	1	2	1	2	3	2	3	x_4,x_6,x_8	x_1,x_2,x_7,x_9,x_{11}	x_3,x_5,x_{10},x_{12}	D_{15}
16	2	3	1	2	1	2	3	3	1	2	2	3	x_3,x_5,x_9	$x_1,x_4,x_6,x_{10},x_{11}$	x_2,x_7,x_8,x_{12}	D_{16}
17	2	3	1	2	2	3	1	1	2	3	3	1	x_3,x_7,x_8,x_{12}	x_1,x_4,x_5,x_9	x_2,x_6,x_{10},x_{11}	D_{17}
18	2	3	1	2	3	1	2	2	3	1	1	2	x_3,x_6,x_{10},x_{11}	x_1,x_4,x_7,x_8,x_{12}	x_2,x_5,x_9	D_{18}
19	3	1	3	2	1	3	2	1	3	2	1	3	x_2,x_5,x_8,x_{11}	x_4,x_7,x_{10}	x_1,x_3,x_6,x_9,x_{12}	D_{19}
20	3	1	3	2	2	1	3	2	1	3	2	1	x_2,x_6,x_9,x_{12}	x_4,x_5,x_8,x_{11}	x_1,x_3,x_7,x_{10}	D_{20}
21	3	1	3	2	3	2	1	3	2	1	3	2	x_2,x_7,x_{10}	x_4,x_6,x_9,x_{12}	x_1,x_3,x_5,x_8,x_{11}	D_{21}
22	3	2	1	3	1	3	2	2	1	3	3	2	x_3,x_5,x_9	x_2,x_7,x_8,x_{12}	$x_1,x_4,x_6,x_{10},x_{11}$	D_{22}
23	3	2	1	3	2	1	3	3	2	1	1	3	x_3,x_6,x_{10},x_{11}	x_2,x_5,x_9	x_1,x_4,x_7,x_8,x_{12}	D_{23}
24	3	2	1	3	3	2	1	1	3	2	2	1	x_3,x_7,x_8,x_{12}	x_2,x_6,x_{10},x_{11}	x_1,x_4,x_5,x_9	D_{24}
25	3	3	2	1	1	3	2	3	2	1	2	1	x_4,x_5,x_{10},x_{12}	x_3,x_7,x_9,x_{11}	x_1,x_2,x_6,x_8	D_{25}
26	3	3	2	1	2	1	3	1	3	2	3	2	x_4,x_6,x_8	x_3,x_5,x_{10},x_{12}	x_1,x_2,x_7,x_9,x_{11}	D_{26}
27	3	3	2	1	3	2	1	2	1	3	1	3	x_4,x_7,x_9,x_{11}	x_3,x_6,x_8	$x_1,x_2,x_5,x_{10},x_{12}$	D_{27}

Appendix 2. 30 Curenty Data Extract

	Country	A	B	C	D	E	F	G	H	I	J	K	L
1	Belgium	132459	131417	707608	26434	23708	24257	23459	32606	13460	214	10042	1135
2	Bulgaria	61427	52796	46438	31464	3798	7949	7508	2013	3852	46	6406	4686
3	Czech Republic	139158	133639	115205	56251	23807	17448	16979	12340	12061	503	7777	3612
4	Denmark	61065	58895	48789	23915	4453	13248	12241	6600	1685	76	9520	995
5	Germany	936544	953827	782313	354506	115007	154730	146844	145928	72569	1944	67479	12239
6	Estonia	20517	16759	18185	14644	510	2260	2078	646	498	18	1344	478
7	Ireland	61314	60284	40510	13328	4549	11606	11061	10996	1934	72	17910	889
8	Greece	118287	115645	93213	52522	6764	22996	19184	9839	10542	316	9282	4934
9	Spain	355898	326944	269835	72418	63434	91423	83872	39252	2817	2938	40014	15094
10	France	522373	490148	370022	61564	68440	132154	125273	103199	37509	1224	93876	19741
11	Italy	501318	444787	415727	133255	61375	118849	110018	90453	31963	1658	33741	18229
12	Cyprus	10838	10673	7505	3880	716	2313	2313	580	807	1	670	1857
13	Latvia	12077	−5070	8401	2261	1087	3222	2963	1729	639	42	2330	666
14	Lithuania	20810	9095	12848	5446	1120	4565	4125	1431	2249	93	4458	1161
15	Luxembourg	12075	11780	10647	1271	1401	6288	6273	1643	660	14	690	64
16	Hungary	67679	64307	49070	16668	3903	11870	11598	14270	6386	269	8267	3687
17	Malta	3035	2973	2657	1893	46	577	529	140	100	1	78	199
18	Netherlands	210053	213054	177819	66613	27326	34988	34242	45820	10432	170	16624	5008
19	Austria	84594	80983	64328	14293	15618	22452	21890	11401	10680	327	7453	1806
20	Poland	400865	357985	327117	173536	30969	48766	47151	61954	29769	779	34624	8576
21	Portugal	70599	60719	49654	14586	9487	18936	18255	5242	5756	228	7515	7446
22	Romania	121355	95545	86038	33353	18577	15133	14129	10192	12732	125	16777	5683
23	Slovenia	19522	11031	15980	6219	1900	5272	5228	2228	971	30	1963	577
24	Slovakia	45982	39893	32008	9507	9316	6654	6547	4545	8522	164	3065	2222
25	Finland	74553	52474	60649	30547	9905	13570	12001	5083	5766	73	5882	2186
26	Sweden	66232	32177	49359	13091	10122	20744	19122	4239	6841	311	7873	1848
27	United Kingdom	590247	586493	501113	192184	67546	118455	111735	108315	26731	1	45908	16495
28	Iceland	4542	5276	1866	5	213	900	844	556	1810	6	646	214
29	Norway	53896	20951	40726	14899	3612	15142	10097	3592	7479	170	4273	1248
30	Switzerland	54247	53367	44017	4190	5985	16422	16092	17050	3689	215	5688	612
	Range	933509	958897	780447	354501	114961	154153	146315	145788	72469	2937	93798	19677
	Mean	161119	149628.2	128321.6	48148.1	19823.13	32106.3	30121.7	25252.6	11870.3	400.93	15739.17	4786.03
	Standard Deviation	220938	219088.3	181771.5	75728.93	27823.22	43507.51	41002.2	23603.62	15541.64	678.39	21641.52	5786.77

Appendix 3. High and Low Curenty Data Extract

	Country	A	B	C	D	E	F	G	H	I	J	K	L
1	Belgium	132459	131417	107608	26434	23708	24257	23459	32602	13460	214	10042	1135
2	Czech Republic	139158	133639	115205	56251	23807	17448	16979	12340	12061	503	7777	3612
3	Germany	936544	953827	782313	354506	115007	154730	146844	145928	72569	1944	67479	12239
4	Greece	118287	115645	93213	52222	6764	22996	19184	9839	10542	316	9282	4934
5	Spain	355898	326944	269835	72418	63434	91423	83872	39252	2817	2938	40014	15094
6	France	522373	490148	370022	61564	68440	132154	125273	103199	37509	1224	93876	19741
7	Italy	501318	444787	415727	133255	61375	118849	110018	94153	31963	1658	33741	18229
8	Hungary	67679	64307	49070	16668	3903	11870	11598	14270	6386	269	8267	3687
9	Netherlands	210053	213054	177819	66613	27326	34988	34242	45820	10432	170	16624	5008
10	Austria	84594	80983	64328	14293	15618	22452	21890	11401	10680	327	7453	1806
11	Poland	400865	357985	327117	173536	30969	48766	47151	61954	29769	779	34624	8576
12	Portugal	70599	60719	49654	14586	9487	18936	18255	5242	5756	228	7515	7446
13	Romania	121355	95545	86038	33353	18577	15133	14129	10192	12732	125	16777	5683
14	Finland	74556	52474	60649	30547	9905	13570	12001	5083	5766	73	5882	2186
15	United Kingdom	590247	586493	501113	192184	67546	118455	111735	108315	26731	0	45908	16495
High 15	Range	868865	901353	733243	340213	111104	142860	135246	140845	66813	2938	87994	18606
	Mean	288399	273864.5	231314.1	86562	36391.07	56401.8	53108.67	46639.33	20958.2	717.87	27017.4	8391.4
	Standard Deviation	256553.6	256629	213030.7	92924.08	31703.96	51160.87	48155	45665.55	17771.85	854.18	26026.4	6336.72
1	Bulgaria	61427	52796	46438	31464	3798	7949	7508	2013	3852	46	6406	4686
2	Denmark	61065	58895	48789	23915	4453	13248	12241	6600	1685	76	9520	995
3	Estonia	20517	16759	18185	14644	510	2260	2078	646	498	18	1344	472
4	Ireland	61314	60284	40510	13328	4549	11606	11061	10996	1934	72	17910	889
5	Cyprus	10838	10673	7505	3880	716	2313	2313	580	807	0	670	1857
6	Latvia	12077	−5070	8401	2261	1087	3222	2963	1729	639	42	2330	666
7	Lithuania	20810	9095	12848	5446	1120	4565	4125	1431	2249	93	4458	1161
8	Luxembourg	12075	11780	10647	1271	1401	6288	6273	1643	660	14	690	64
9	Malta	3035	2973	2657	1893	46	577	529	140	100	1	78	199
10	Slovenia	19522	11031	15980	6219	1900	5272	5228	2228	971	30	1963	577
11	Slovakia	45982	39893	32008	9507	9316	6654	6547	4545	8522	164	3065	2222
12	Sweeden	66232	32177	49359	13091	10122	20744	19122	4239	6841	311	7873	1848
13	Iceland	4542	5276	1866	5	213	900	844	556	1810	6	646	214
14	Norway	53896	20951	40726	14899	3612	15142	10097	3592	7479	170	4273	1248
15	Switzerland	54247	53367	44017	4190	5985	16422	16092	17050	3689	215	5688	612
Low 15	Range	63197	65354	47493	31459	10076	20167	18593	16910	8422	311	17832	4622
	Mean	33838.6	25392	25329.07	9734.2	3255.2	7810.8	7134.73	3865.87	2782.4	83.87	4460.93	1180.67
	Standard Deviation	24066.08	22253.43	18152,81	8974.83	3195.32	6221.92	5590.46	4631.91	2742.38	91.88	4692.22	1164.93