International comparisons of road safety using Singular Value Decomposition

Siem Oppe

D-2001-9

International comparisons of road safety using Singular Value Decomposition

D-2001-9
Siem Oppe
Leidschendam, 2001
SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number:	D-2001-9
Title:	International comparisons of road safety using Singular Value Decomposition
Author(s):	Siem Oppe
Research theme:	Road safety analysis
Theme leader:	Ingrid van Schagen
Project number SWOV:	37.420
Keywords:	Fatality, safety, development, trend (stat), analysis (math), statistics, Europe, international
Contents of the project:	Road safety developments in different countries are often compared on the basis of single parameters. In this analysis, the similarities and dissimilarities of road safety developments in 24 countries over 28 years are regarded in more detail through use of a Singular Value Decomposition (SVD) technique. The present analysis is restricted to developments in the number of fatalities. SVD analysis might help to understand the major factors that effect safety for countries in general.
Number of pages:	15 pp.
Price:	Dfl. 15.-
Published by:	SWOV, Leidschendam, 2001

SWOV Institute for Road Safety Research
P.O. Box 1090
2260 BB Leidschendam
The Netherlands
Telephone 31703209323
Telefax 31703201261

Abstract

There is a general interest in the comparison of road safety developments in different countries. Comparisons have been made, based on absolute levels of accident or fatality risk or on the rate of change of functions regarding risk, the number of accidents, fatalities or injuries over time. Such comparisons are mostly based on single parameters. The method described here goes into more detail regarding similarities and dissimilarities of such developments, using a Singular Value Decomposition (SVD) technique.

A first analysis was carried out on the number of fatalities in 24 countries (mainly European) over 28 years (1970 through 1997). The data was taken from the International Road Traffic and Accident Database (IRTAD), initiated by the Organisation for Economic Cooperation and Development (OECD). The year 1970 is the starting year; from this year onwards, annual data is available for a large number of countries.

The analysis showed that most of the similarities and dissimilarities are represented by four dimensions. There turned out to be one major common trend describing the general decrease of the number of fatalities over time from the starting year 1970 onwards. It is well known that before 1970 the number of fatalities increased in many countries and decreased shortly afterwards. However, this trend cannot be shown because data from the period before 1970 is missing.
For Greece, Korea, Spain and Portugal the trend is on average increasing instead of decreasing, in particular for Greece and Korea. In all other countries there is a decreasing trend, generally at the same level, except for New Zealand. Here the decrease is less marked than in most Western countries.
On all four dimensions the similarities between the European countries were large, compared with non-European countries, except for Greece, Spain, Portugal, Hungary and the Czech Republic and to a minor extent for Ireland. Within the cluster of European countries sub-clusters may be distinguished. The developments in Germany and the Netherlands are for instance rather similar, and the same goes for Belgium, France, Switzerland and Austria.
This suggests that, apart from economic developments, geographical and cultural factors are also important.

The analysis is only applied to the total number of fatalities. The same procedure could be used to compare the fatality rates for countries for which traffic volume developments are known. Such analyses might help to understand the major factors that affect safety for countries in general.

Contents

1. Introduction

There is a general interest in the comparison of road safety developments in different countries. Comparisons have been made, based on absolute levels of accident or fatality risk or on the rate of change of functions regarding risk, the number of accidents, fatalities or injuries over time.

The method described here is an effort to go into more detail regarding similarities and dissimilarities of such developments, using Singular Value Decomposition (SVD). This technique is related to the better known Factor Analysis and Principal Components Analysis techniques. The latter techniques are used to analyse correlation matrices or variance-covariance matrices, in order to find common trends between variables (in this case countries or states). Singular Value Decomposition can be used to make comparisons between traffic developments (measured in motor vehicle kilometres per year for a country or state), safety developments (primarily developments in fatalities) and corresponding risk developments in different countries.

A first analysis was carried out on the number of fatalities in 24 countries (mainly European countries) over 28 years (1970 through 1997). The data was taken from the IRTAD, initiated by the OECD. The year 1970 is the starting year; from this year onwards, annual data is available for a large number of countries.
The average level for the total number of accidents is mainly dependent on country size. In order to avoid a trivial solution, regarding the mean difference in the number of fatalities over the years, the numbers are first normalised: for each year the number of fatalities, minus the mean number of fatalities for that country over the years, is divided by the standard deviation of the number of fatalities for that country over the years. Alternative transformations, or no transformation at all, are possible.

A possible extension of this analysis is to add data from the individual states of the USA to this comparison. Furthermore, the same type of analysis can be applied to the fatality rates as well as the motor vehicle kilometres, to find out whether developments in safety and risk can be addressed to developments in the amount of traffic, using similarities and dissimilarities between countries as a basis. Unfortunately, traffic data is not available for many countries.
The method may also be used to look for explanations for the similarities and differences in developments in subgroups of countries, based on trafiic growth in, and infra-structural and economic characteristics of the subgroups.

2. Method

A singular value decomposition (SVD) of a matrix, also referred to as an ‘Eckart-Young analysis’ (Eckart & Young, 1937), is a decomposition of a non-symmetric matrix in eigenvalues and left and right eigenvectors. A given matrix X with n rows and m columns ($m \leq n$), can be written as a product of three matrices:

$$X_{n,m} = L_{n,p} \; E_{p,p} \; R^T_{p,m}$$

with $L^TL = \mathbf{I}_{p,p}$ and $R^TR = \mathbf{I}_{p,p}$ (where T stands for transposition of the matrix, $\mathbf{I}$ is the identity matrix, E a diagonal matrix and $p \leq m$).

In fact L and R are orthonormal matrices, existing of p uncorrelated column vectors with n elements and p uncorrelated row vectors with m elements respectively, with p the rank of matrix X. The vectors of L are called the left eigenvectors of X, the vectors of R the right eigenvectors of X. The p diagonal values of E are called the eigenvalues of X.

The matrix X can be written as the sum of p matrices X_i , where X_i (a matrix of rank 1) is the product of the i’s left eigenvector, the i’s eigenvalue and the i’s right eigenvector. Or:

$$X = X_1 + X_2 + \ldots + X_p, \text{ where } X_i = l_i \, e_i \, r^T_i$$

The eigenvalue of effect i is a measure of the contribution of X_i to the description of X. If the numbering of the X_i matrices is in decreasing order of the eigenvalues, then X_1 ’explains’ most of the variance in the values of X etc.

If p=1, then the matrix X can be decomposed into one r-vector of m values, one l-vector of n values and one single eigenvalue. In fact this means that all columns of X, and also all rows of X are equal up to a scalar.

For the example of our matrix of fatalities for 24 countries over 28 years , this would mean that all countries have the same number of fatalities, except for a multiplicative constant. Such a simple model will of course not hold. However, in this way it is possible to find out to what extent such similarities exist. If X_1 is very similar to X, then the hypothesis of ‘no differences in development over time’ can be maintained. If there are differences in the development of safety over time, such as differences between developed and under-developed countries or European and non-European countries, one might expect to find a decomposition of the matrix in (at least) two submatrices X_1 and X_2, each with a large eigenvalue. All other eigenvalues will be small and more or less equal, if all the other deviations are supposed to be due to error.

The decomposition of X is found from an ordinary Principal Components Analysis (CPA) of the matrix $X^TX = R\,E\,L^T\,L\,E\,R^T = R\,E^2\,R^T$.

The square roots of the eigenvalues of X^TX are equal to those of X. The eigenvectors of X^TX are equal to the right eigenvectors of X. The left-eigenvectors of X can be found from these values as: $L = XRE^{-1}$, with E^{-1} the inverse of the diagonal matrix E, with reciprocal values on the diagonal.

3. Results

From the eigenvalues (see *Table 1*) it can be seen that the eigenvalue (4.09) for the first dimension is dominant. Furthermore, that four eigenvalues are larger than 1. The values for these four dimensions will be represented.

Dimension	Eigenvalue	CPEV
1	4.09	0.70
2	1.61	0.80
3	1.31	0.88
4	1.08	0.92
5	0.76	0.95
6	0.58	0.96
7	0.49	0.97
8	0.47	0.98
9	0.33	0.99
10	0.30	0.99

Table 1. *Eigenvalues and cumulative proportion of explained variance (CPEV).*

From the plot of the eigenvectors for years (see *Figure 1*), it can be seen that the first common dimension (dim1) describes the general decrease of the number of fatalities over time. Note that the starting year is 1970. Before 1970 the number of fatalities went up for many countries to decrease shortly afterwards. Data regarding the trend before 1970 is missing. Dim2 through dim4 describe deviations from this trend (see below). Dim 4 seems to be similar to dim2, but shifted in time.

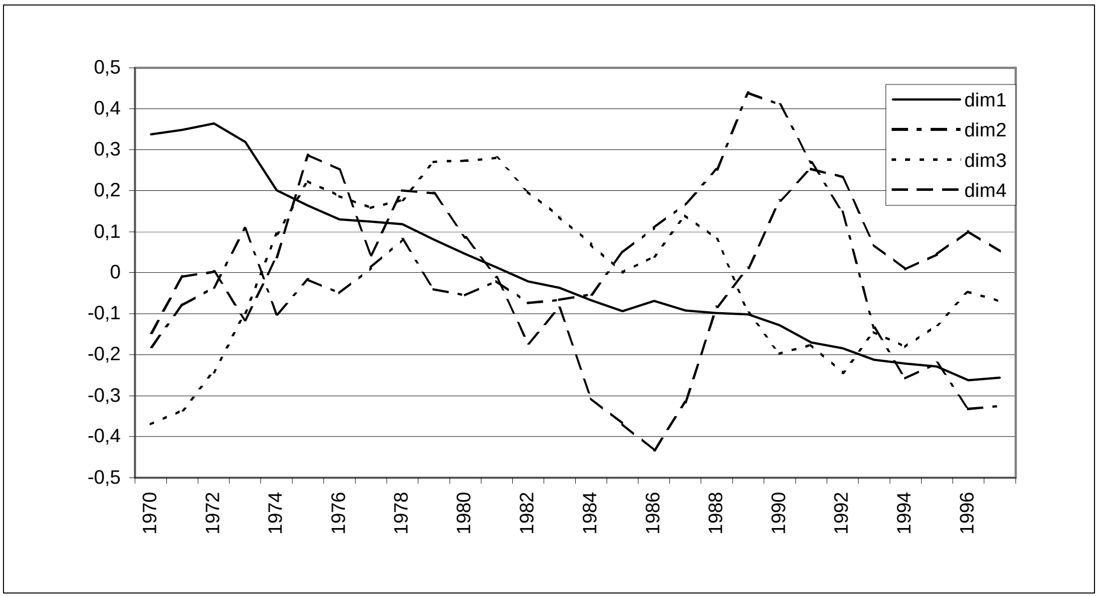

Figure 1. *Plot of the eigenvector values for years, from 1970 to 1997.*

Figure 2 shows that Greece, Korea, Spain and Portugal are low and even negative on (and therefore most deviant from) this trend. This means that for these countries, on average, the trend is increasing instead of decreasing, particularly for Greece and Korea. The other countries are all positive and generally at the same level, except for New Zealand. Here the decrease is a slightly less marked than in most other Western countries.

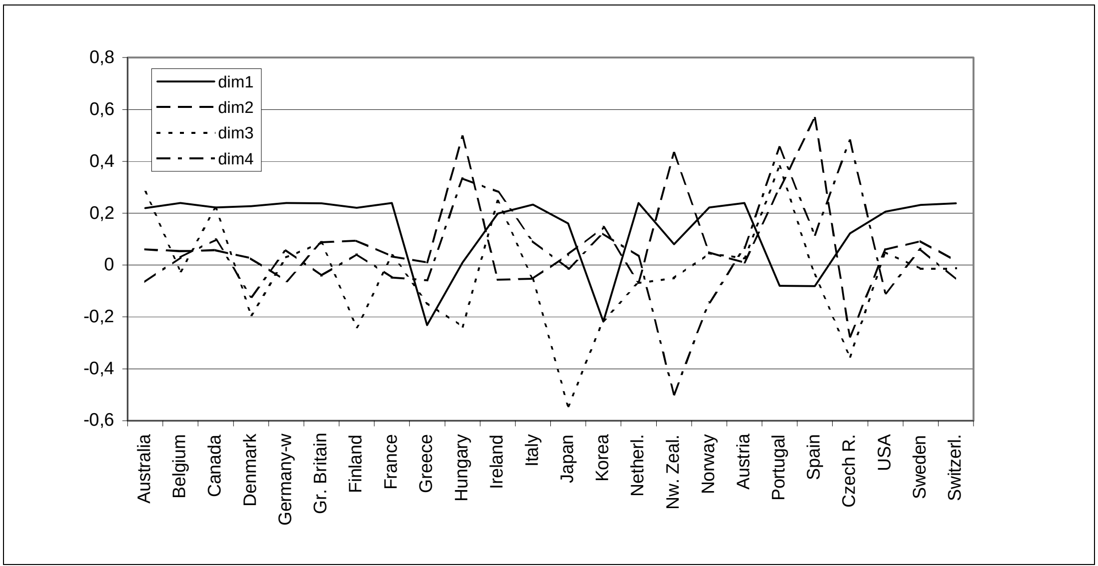

Figure 2. *Plot of the eigenvector values for countries.*

The second dimension expresses a considerate increase in the number of fatalities during the end of the eighties and beginning of the nineties and a further decrease in the late nineties. Spain, Hungary and New Zealand and to a minor degree Korea (subgroup 1) have the highest 'loading' on this dimension, while primarily the Czech Republic and to a lower extent the Netherlands, West-Germany, Ireland and Italy (subgroup 2) have a reversed trend. For the first subgroup this means a smaller decrease or even a rise in fatalities in the second half of the eighties and a higher decrease in the nineties. For the second subgroup a steeper decrease from 1984 onwards, was followed by a retarded decrease in the nineties.

The third dimension shows an increase in the late seventies, and late eighties for Australia, Canada, Ireland and Portugal and (because of the negative sign) a decrease for Japan and to a lesser extent for the Czech Republic.

The fourth dimension shows a steep dip for the middle of the eighties as compared to the middle seventies and the beginning of the nineties. This trend is present for the Czech Republic, Portugal, Hungary and Ireland and reversed for New Zealand and to a lesser extent for Denmark and the USA.

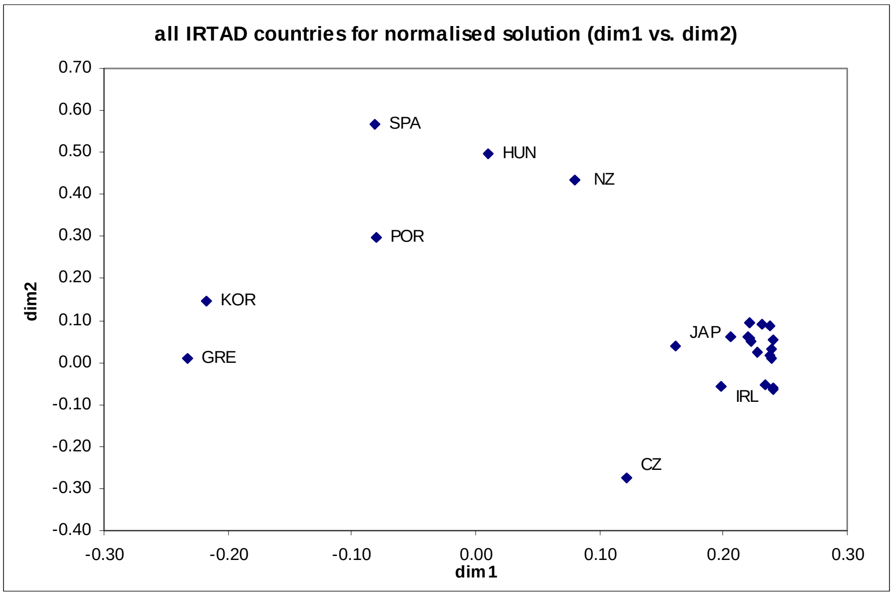

Figure 3a. *Representation of countries on dimension 1 and 2.*

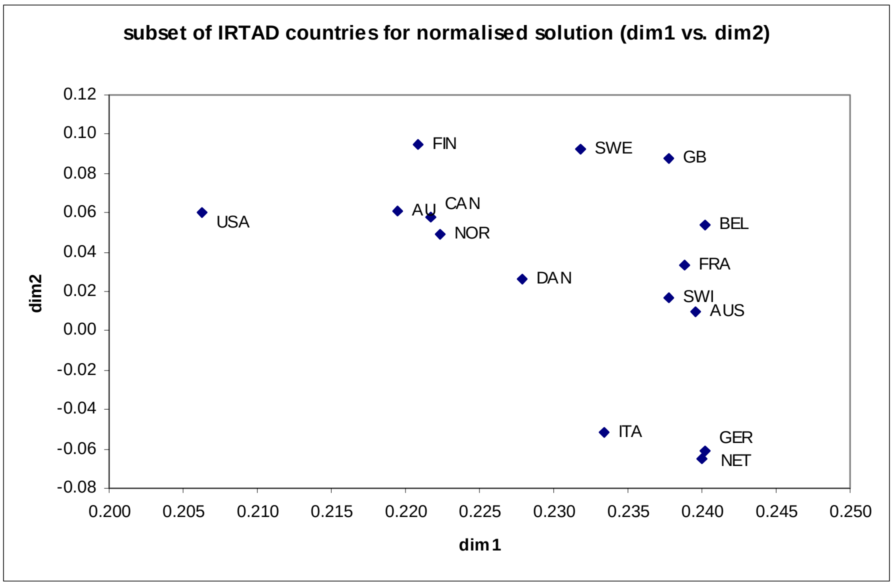

Figure 3b. *Representation of a subset of countries on dimension 1 and 2.*

Figure 3a. represents the values for the countries on dimension 1 and 2 and *Figure 4a* on dimension 3 and 4. A separate plot is made for those countries that cluster on the right side of *Figure 3a* (see *Figure 3b*) and for the countries in the middle of *Figure 4a* (see *Figure 4b*).

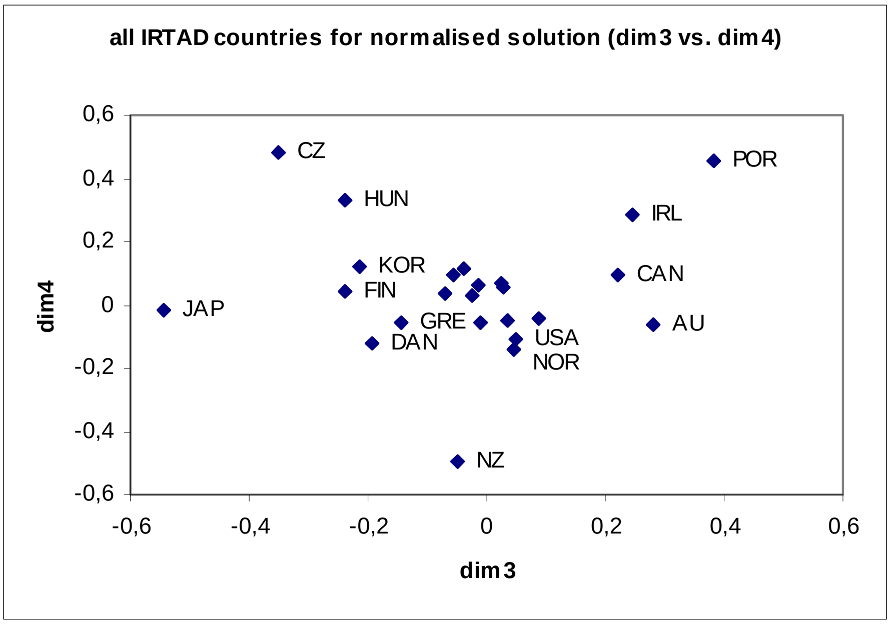

Figure 4a. *Representation of countries on dimension 3 and 4.*

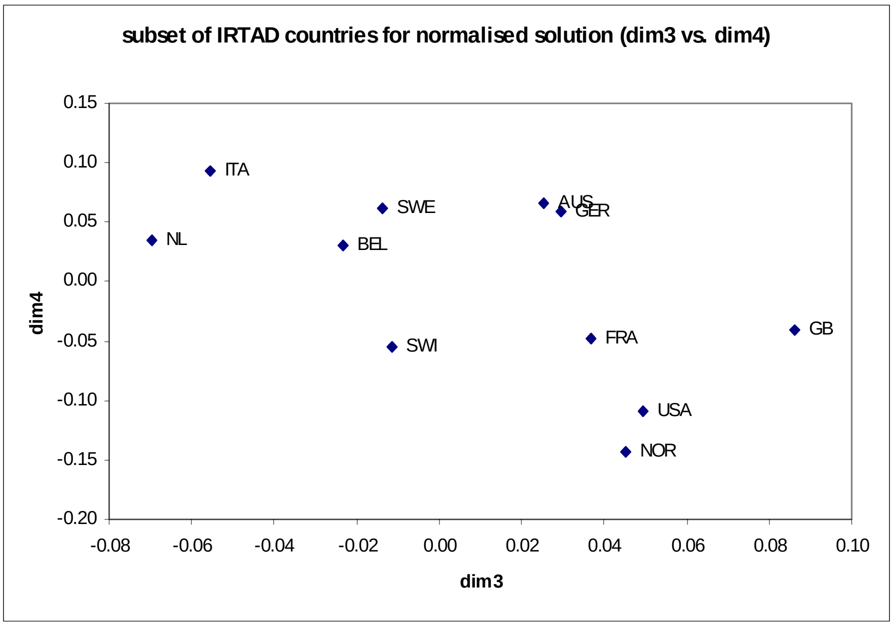

Figure 4b. *Representation of a subset of countries on dimension 3 and 4.*

In general these plots show that the developed Western European countries cluster together. This means that the developments in these countries are similar, and dissimilar to the developments in the non-European countries. The European countries that deviate most on dim1 and dim2 are Greece, Spain, Portugal, Hungary, the Czech Republic, and to a lesser extent Ireland.

When regarding the European countries, we see that Greece is most extreme on dim1, but not on the other dimensions. This means that there is a steady, uninterrupted increase in the number of accidents over the years. The developments in Portugal and Spain on average take place at the same speed (dim1), however, Spain is high on dim 2 and Portugal on dim3 and 4. These suggested differences between Portugal and Spain are in fact only differences in time. The development in Portugal is later (with a peak in the early nineties) than in Spain (with a peak in the late eighties). The major difference between Hungary and the Czech Republic is primarily represented by dim2: the high peak at the late eighties is present in Hungary, but reversed in the Czech Republic.

For the clustering subset in *Figure 3b* We see that Germany and the Netherlands are very similar, and Italy to a lesser extent. Belgium, France, Switzerland and Austria are also close together, with a medium value on dim 2, in contrast with Germany, the Netherlands and Italy that have a negative value.
The Nordic countries, Australia and Canada are also clustering together, with a further similarity between Sweden and Great Britain. Except for Sweden, the annual decrease (dim1) is relatively low and the value on dim 2 relatively high.
In *Figure 4b* we see that Germany and Austria are rather similar and together with Italy and Sweden opposed to Norway and the USA on dim4, while the Netherlands and Italy are opposed to Great Britain on dim3.

4. Conclusion

The development in non-European countries is generally different from the development in the European countries, suggesting that not only economic trends but also geographical trends are important. It would be interesting to see what would happen, if the USA were added. Two main clusters might result, but it is also possible that European countries and the USA. with common structures in traffic or economic development or in infrastructure will show common developments in safety. Such similarities and dis-similarities may help to get a better understanding of the major factors that determine safety developments.

References

Eckart, C. & Young, G. (1937). *The approximation of one matrix by another one of lower rank.* In: Psychometrika 1, pp.211-218.