ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

A DEA Analysis of Risk, Cost and Revenues in Insurance

Maria Manuela Segovia Gonzalez
Universidad Pablo Olavide

Ignacio Contreras Rubio
Universidad Pablo Olavide

Cecilio Mar Molinero
Kent Business School

Kent Business School
UNIVERSITY OF KENT

Working Paper No.129
October 2006

A DEA ANALYSIS OF RISK, COST, AND REVENUES IN INSURANCE

By

Maria Manuela Segovia Gonzalez
Universidad Pablo Olavide
Sevilla, Spain.

Ignacio Contreras Rubio
Universidad Pablo Olavide
Sevilla, Spain.

Cecilio Mar Molinero*
Kent Business School
University of Kent
Canterbury, Kent CT2 7PE, UK
C.Mar-Molinero@kent.ac.uk
+44 1227 824554

* Corresponding author

Abstract

Insurance companies have to take risk and cost into account when pricing car insurance policies that cover the risk of private use of cars. In this paper we use data from 80,000 car insurance policies in order to assess, once risk and cost have been taken into account, the policies that generate the highest returns for the company. We use Data Envelopment Analysis and frame the study within an analysis of experiments context. There are consequences for the pricing policy of the company.

Keywords: efficiency, car insurance, data envelopment analysis, risk, statistics.

JEL Classification: C44, D81, G22

Introduction

The market for vehicle insurance in Spain has been mature for a long time, as reflected in its slow growth, excess supply, little product innovation, standard products, little differentiation between companies (except in price), and a strong level of competition. But in Spain, as in the rest of Europe, the situation has changed in recent times. Competition between insurance companies has become fierce. This change can be traced to two causes. First, the publication of injury data has resulted in better tuning of insurance products; second, the popularity of new distribution channels- such as telephone and Internet sales.

Some insurance companies have been following very aggressive strategies, offering products with very low profitability, even at a loss. The only justification for this behaviour is the search for a higher share of the market. There are now many new products, covering a variety of risks at a competitive price.

Customers can now choose amongst many insurance products from many firms that cover the same risks. Companies also offer many insurance products that are tailored to the characteristics of the individual driver. In this situation, the behaviour of the insurance consumer has been the subject of much interest (Cotter and Jensen, 1989).

Price is not the only aspect of the insurance product that the customer values. Pujol and Bolancé (2004) show that there is a fidelity effect: a the customer who is satisfied with the service and the way he/she has been treated, is more inclined not to switch companies. But, when switching companies, price is the determining factor, although there are others. Another important factor is the fear of not being compensated in full for a loss. Mayers and Smith (1983), and Doherty and Schlesinger (1983) were the first to study the possibility that the insurer fails to honour its duty to pay full indemnity in case of accident. They also analysed other secondary risks that may affect the final wealth of the insurance taker. The

perception that the insurer may not pay up can affect the consumer's choice of policy: a consumer who does not fully trust an insurance company may over-insure in order to achieve additional protection. Other authors that have studied similar issues are Schlesinger and Schulenburg (1987), Doherty and Schlesinger (1990), Meyer and Meyer (1998), Doherty and Dionne (1993), Johnson, et al. (1993) and Mahul (2000).

Vehicle insurance products are priced on the grounds of risk, but each insurance product has its pricing peculiarities. Boj, et al. (2004), and Guillén et al. (2005) used multivariate statistical analysis in order to assess how the various factors that contribute to risk are related to policy price. There is a need to explore how the various components of risk are related to the success or otherwise of a particular kind of insurance policy, and this is what we try to assess in this paper.

Insurance companies, when pricing a policy, segment the market into various groups according to the risk and cost of a claim. In Spain they do this using their own experience and the statistical information provided by the association of insurance and reassurance companies (UNESPA). Risk factors are: vehicle insurance group, area where the vehicle resides, type of use, the driver's personal characteristics, and the driver's history of accidents. In this paper, we will also take into account some of these risk factors in order to produce a taxonomy of insurance policies to be studied. The data was provided by a car insurance company with a large share of the Spanish market that desires to remain anonymous.

In this paper a particular insurance policy taken by a group of customers with similar risk factors is treated as a Unit of Assessment (UOA) in DEA. We prefer the term UOA rather than Decision Making Unit (DMU), since we are assessing the efficiency of insurance policies and no decision making is involved. We follow Sarrico et al. (1997) in using form of words. DEA will be used to assess the performance of the various insurance policies sold by a particular company to individuals who insure a car for private use. Three inputs and two

outputs are collected for each UOA. As inputs we choose those factors that result in a cost to the company: total number of claims, number of claims in which the insurance taker is guilty- since they are more expensive from the point of view of the insurance company-, and the cost of a claims. We take as outputs two measures of income to the company: gross premium, and net premium. All the UOA are standardised per hundred policy takers in order to control for the different numbers sold, as we are interested in assessing how the products compare and the reasons why some are more efficient than others from the point of view of the company.

A UOA is efficient if with a low level of inputs is able to generate a high level of output. But the efficiency of a UOA depends on the specification of the model: the particular set of inputs and outputs that are used; Pedraja Chaparro et al. (1999), Parkin and Hollingsworth (1997), Jenkins and Anderson (2003). We are interested in obtaining an overall measure of efficiency over all possible models. The approach followed here is due to Serrano-Cinca et al. (2005), and Serrano-Cinca and Mar-Molinero (2004). In essence, this methodology estimates efficiencies under a series of specifications, and then applies Principal Component Analysis (PCA) to the results.

Having obtained an overall measure of efficiency, we assess the impact of the various risk factors. We find that overall efficiency is affected by gender, age, area, and type of car, but that the relationship between efficiency and these risk factors is not a straight forward one due to the presence of interactions.

After this introduction we discuss the data used. The next section contains the technical aspects of the analysis. This is followed by a discussion of the results obtained. The paper ends with a concluding section.

1. The Data

The data on which this study is based were provided by one of the main Spanish insurance companies. This data set is not generally available due privacy legislation and industrial confidentiality issues. The company also insures industrial vehicles and private vehicles for industrial use, but we only concern ourselves in this paper with policies issued to individuals who drive a car for personal use. In total, 80,000 policies issued in 2002 have been included in the analysis.

The data were stratified taking into account the risk factors usually accepted in the insurance business: gender, driver's age, type of car, and geographical area where the vehicle is domiciled. Gender is important, as the risk of an accident is different for men and women. As far as age is concerned, we consider two cases: drivers who are 30 years old or younger at the time when the policy is issued, and drivers who are older. Apart from the well-known relationship between age and risk, we must point out that in Spain, drivers need to renew their licence at regular intervals, and this requires a medical examination, hence physical condition should not be an issue. Consider, as a counter-example, the case of the UK, where drivers do not require a medical examination until they reach the age of 70; if we had been dealing with UK data, it would have been appropriate to divide the variable "age" into more than two intervals. Cars have been divided by type into: premium, intermediate, and economy. This is not only related to the risk of accident, but to the cost of the accident, should it occur. An accident involving a premium car is, in general, more expensive to the company than an accident involving an economy car. This is taken into account in the policy pricing decision, and the amount of revenue generated by the policy will reflect it. Finally, we segment on geographical area in Spain, as the different areas have different weather conditions and the level of investment in road infrastructure is different. The condition of the road and the

weather conditions are known to affect the risk of an accident. We consider four different areas: North, Centre, Mediterranean, and South.

In our data set there are more men than women: 82.73% men versus 17.27% women. Only 8.33% of insurance policies were issued to people aged 30 and under, the remaining 91.67% being issued to people older than 30. This difference probably reflects the very high premiums attached to policies for young people with little driving experience. If we now turn our attention to the type of car, 12.48% of them are classified as premium, 40.38% are medium, and 47.14% are economy. The insurance company specialises in the South of Spain, and this explains that 52.22% of the insurance takers are located in that region; the Mediterranean area accounts for 18.31% of the policies issued, the North accounts for 17.67% of the policies, and the remaining 11.80% of the policies are sold in the Centre. Each combination of gender, age, type of car, and area is treated as a UOA. A mnemonic coding system has been devised to describe the different UOAs. This can be seen in Table 1.

		Premium (P)		*Intermediate (I)*		*Economy (E)*	
		Female (F)	*Male (M)*	*Female (F)*	*Male (M)*	*Female (F)*	*Male (M)*
North (N)	*>30 (+)*	PFN+	PMN+	IFN+	IMN+	EFC+	EMN+
	=30 (-)	-	PMN-	IFN-	IMN-	EFC-	EMN-
Centre (C)	*>30 (+)*	PFC+	PMC+	IFC+	IMC+	EFC+	EMC+
	=30 (-)	-	PMC-	IFC-	IMC-	EFC-	EMC-
Mediterranean (M)	*>30 (+)*	PFM+	PMM+	IFM+	IMM+	EFM+	EMM+
	=30 (-)	-	PMM-	IFM-	IMM-	EFM-	EMM-
South (S)	*>30 (+)*	PFS+	PMS+	IFS+	IMS+	EFS+	EMS+
	=30 (-)	PFS-	PMS-	IFS-	IMS-	EFS-	EMS-

Table 1 Profiles of clients with mnemonic coding.

The experimental design just described produces 48 possible UOAs, but three were not included in the data because not enough policies were issued that combined these characteristics. They are identified as missing data (-) in Table 1. The final number of UOAs to be analysed by means of DEA is, therefore, 45.

2. DEA input and output selection

DEA has been applied before in the area of insurance, but the aim of the studies has been to compare the performance of the various companies, not to study insurance products; examples are Cummins and Zhi (1998), Cummins and Rubio-Misas (1998), Fecher et al. (1993), Fukuyama (1997), and Tone and Sahoo (2005). All the studies reviewed take the point of view of the company, and not the point of view of the consumer. In general, premium income and number of policies are treated as outputs. We will also treat premium income as an output. We accept that measures of input can have high correlations, and the same thing can happen with measures of output. Rather than study their correlations prior to the estimation of the model, we prefer to estimate a variety of models and check how the presence or absence of a particular input- or output- in the model specification affects the results.

From the point of view of the company, the relevant information needed to assess the performance of an insurance product can be measured by two types of variables: those that, directly or indirectly, generate a cost and should be kept to as low a value as possible; and those that generate an income. We associate the first set of variables with inputs and the second set with outputs.

As inputs we have considered three variables: number of claims, cost of a claims, and number of claims in which the customer is the guilty part. The third input has been added on the grounds that, when the insurance taker is the guilty part, the company has to pay more for the claim. More specifically:

Input A: Number of claims per 100 policy takers.

Input B: Cost of claims per 100 insurance takers.

Input C: Number of claims in which the insurance taker is the guilty part (per 100 insurance takers).

Notice that by standardising the data to 100 customers we are, in the case of inputs A and C, calculating probabilities. Thus, inputs A and C are equivalent to risks. Input B, for the same reason, can be interpreted as the cost associated with the risk.

Outputs are associated with premium income, as done in previous studies. We distinguish between gross and net premium. Net premium is obtained from gross premium after removing concessions and discounts. This provides a link with the past and, in some way, captures the quality of the customer and the fidelity effect, since companies offer no-claim bonuses, and try to retain their best customers by offering them extra discounts. This way of treating premiums is in line with the literature; see Lemaire (1985), Henriet and Rochet (1986), and Dionne and Vanasse (1992). Again, standardising to 100 policies allows us to assess the income obtained for a given level of risk. An "efficient policy" would be one that, despite a high level of risk, generates a high level of revenues.

Output 1: Gross premium per 100 insurance takers.

Output 2: Net premium per 100 insurance takers.

Table 2 shows summary statistics for the values of inputs and outputs.

	Inputs			*Outputs*	
	A	B	C	1	2
Max	76.92	51,009.10	30.77	142,989.31	99,631.13
Min	21.28	1,125.38	2.13	48,812.08	30,016.87
Average	37.49	17,774.16	16.07	88,362.13	58,729.66
Std. dev.	9.70	8,743.68	4.74	26,381.49	18,732.99

Table 2 Values of inputs and ouputs

3. DEA analysis

Efficiency estimates were obtained with the CCR model. We were aware of the fact that there are strong correlations between the inputs and between the outputs. For example, when the gross premium is high the net premium is also high. But, instead of simplifying the model prior to estimation, we chose to estimate a series of different specifications (in terms of inputs and outputs) of the model. Since we have three inputs and two outputs, 21 different models are possible. The notation that has been used to describe the models is summarised in Table 3.

Model	*Inputs*	*Outputs*
A1	A	1
AB1	A,B	1
ABC1	A,B,C	1
AC1	A,C	1
BC1	B,C	1
B1	B	1
C1	C	1
A2	A	2
AB2	A,B	2
ABC2	A,B,C	2
AC2	A,C	2
BC2	B,C	2
B2	B	2
C2	C	2
A12	A	1,2
AB12	A,B	1,2
ABC12	A,B,C	1,2
AC12	A,C	1,2
BC12	B,C	1,2
B12	B	1,2
C12	C	1,2

Table 3 DEA models

Estimating all possible models has, at least two advantages. First, we know that the efficiency of a UOA depends on the particular mix of inputs and outputs that are included in the model, and in this way we can see how, for a particular UOA, efficiency depends on the specification used. Second, the multivariate study of the calculated efficiencies will produce an overall

ranking of UOAs that encompasses all possible specifications. Table 4 shows, in the form of percentages, the efficiency of each of the 45 UOA under each of the 21 models estimated.

DMU	*A1*	*AB1*	*ABC1*	*AC1*	*BC1*	*B1*	*C1*	*A2*	*AB2*	*ABC2*	*AC2*	*BC2*	*B2*	*C2*	*A12*	*AB12*	*ABC12*	*AC12*	*BC12*	*B12*	*C12*
PFN+	60	60	60	60	53	31	23	54	54	54	54	48	25	21	60	60	60	60	53	31	23
PFC+	34	37	37	34	37	24	14	33	35	36	33	36	21	14	34	37	37	34	37	24	14
PFM+	43	43	44	44	33	15	20	42	42	42	42	31	13	19	43	43	44	44	33	15	20
PFS+	34	34	34	34	29	17	13	34	34	34	34	29	15	13	34	34	34	34	29	17	13
PFS-	67	100	100	67	100	100	22	75	100	100	75	100	100	24	75	100	100	75	100	100	24
PMN+	58	58	58	58	45	24	23	51	51	51	51	40	19	20	58	58	58	58	45	24	23
PMN-	86	86	86	86	68	39	30	85	85	85	85	67	35	30	86	86	86	86	68	39	30
PMC+	44	44	44	44	41	25	18	41	41	41	41	39	21	16	44	44	44	44	41	25	18
PMC-	34	34	37	37	23	10	17	36	36	38	38	24	9	18	36	36	38	38	24	10	18
PMM+	47	47	49	49	41	20	22	44	44	46	46	38	17	21	47	47	49	49	41	20	22
PMM-	47	60	67	49	67	45	22	51	61	70	53	70	44	24	51	61	70	53	70	45	24
PMS+	40	41	41	40	40	25	15	39	40	40	39	38	22	14	40	41	41	40	40	25	15
PMS-	53	53	55	55	48	25	25	57	57	59	59	51	24	26	57	57	59	59	51	25	26
IFN+	48	48	48	48	40	24	18	42	42	42	42	36	19	16	48	48	48	48	40	24	18
IFN-	100	100	100	100	100	59	44	100	100	100	100	100	53	43	100	100	100	100	100	59	44
IFC+	36	45	46	36	46	34	14	32	38	39	32	39	28	12	36	45	46	36	46	34	14
IFC-	61	63	100	100	100	39	100	61	64	100	100	100	35	100	61	64	100	100	100	39	100
IFM+	38	38	38	38	31	16	16	35	35	35	35	28	13	15	38	38	38	38	31	16	16
IFM-	51	51	52	52	43	21	24	52	52	54	54	43	19	24	52	52	54	54	43	21	24
IFS+	27	31	31	27	29	21	9	26	28	28	26	27	18	9	27	31	31	27	29	21	9
IFS-	39	48	54	41	54	35	19	42	49	56	44	56	34	20	42	49	56	44	56	35	20
IMN+	51	51	51	51	41	25	18	42	42	42	42	35	18	15	51	51	51	51	41	25	18
IMN-	55	55	55	55	42	25	18	55	55	55	55	41	22	18	55	55	55	55	42	25	18
IMC+	40	46	47	40	47	31	17	36	40	41	36	41	24	15	40	46	47	40	47	31	17
IMC-	35	35	35	35	29	15	16	37	37	37	37	30	14	16	37	37	37	37	30	15	16
IMM+	44	44	44	44	41	24	17	38	38	38	38	35	19	15	44	44	44	44	41	24	17
IMM-	34	34	34	34	29	16	13	36	36	36	36	31	15	14	36	36	36	36	31	16	14
IMS+	27	31	31	27	30	21	10	25	27	27	25	26	17	9	27	31	31	27	30	21	10
IMS-	31	31	31	31	29	17	12	34	34	34	34	31	17	13	34	34	34	34	31	17	13
EFC+	39	43	43	39	41	28	13	34	36	36	34	34	21	11	39	43	43	39	41	28	13
EFC-	51	63	63	51	63	46	19	49	58	58	49	58	40	18	51	63	63	51	63	46	19
EFC+	29	38	40	29	40	30	12	26	32	35	26	35	24	11	29	38	40	29	40	30	12
EFC-	47	59	62	47	62	44	19	51	60	63	51	63	43	20	51	60	63	51	63	44	20
EFM+	37	41	41	37	38	26	13	33	35	35	33	33	21	11	37	41	41	37	38	26	13
EFM-	35	38	38	35	36	24	12	37	39	39	37	36	23	12	37	39	39	37	36	24	12
EFS+	24	29	29	24	26	20	8	22	25	25	22	23	17	7	24	29	29	24	26	20	8
EFS-	25	33	33	25	33	26	9	28	33	34	28	34	25	10	28	33	34	28	34	26	10
EMN+	48	48	48	48	43	26	17	40	40	40	40	35	19	14	48	48	48	48	43	26	17
EMN-	40	40	40	40	37	23	14	40	40	40	40	36	20	14	40	40	40	40	37	23	14
EMC+	38	44	44	38	44	29	15	33	37	37	33	37	23	13	38	44	44	38	44	29	15
EMC-	28	37	39	28	39	30	11	28	35	37	28	37	27	11	28	37	39	28	39	30	11
EMM+	40	42	42	40	41	26	16	35	36	36	35	35	20	13	40	42	42	40	41	26	16
EMM-	25	26	26	25	26	16	10	28	29	29	28	28	16	11	28	29	29	28	28	16	11
EMS+	27	31	31	27	29	21	9	24	27	27	24	25	17	8	27	31	31	27	29	21	9
EMS-	26	29	29	26	28	19	9	28	31	31	28	29	18	10	28	31	31	28	29	19	10

Table 4 DEA efficiencies under the 21 DEA models

An examination of Table 4 shows the influence on efficiency of model specification. If we consider the saturated model, ABC12, we observe that only three units are 100% efficient: IFN-, IFC- and PFS-. They have in common that they involve females, aged 30 or less, and cars in the premium and intermediate ranges. This can be interpreted to mean that, having taken into account the number of accidents, the cost of accidents, and the number of accidents in which the insurance taker was the guilty part, the value of premiums to the company is high. In other words, having taken into account the risk factors associated with young women in relatively expensive cars, we conclude that they are good business. To further explore this subject we examine each one of these three efficient UAOs.

Take first PFS-, we notice that this UOA achieves very low efficiency values under specifications C1, C2, and C12. These three specifications use as an input only the accidents in which the insurance taker was guilty. There are two possible explanations for this low level of efficiency, either these young women in expensive cars in the South are guilty relatively often when involved in an accident; or, they are charged a low premium given the risk involved. The high efficiency value in model ABC12 for this insurance product may be explained by the fact the cost of claims is relatively low in this group (input B); or, perhaps, these policies are associated with relatively few claims (input A).

A similar analysis can be performed with UOA IFN- , involving young females in cars of intermediate value (in the North of the country). Here we also observe low efficiency values under specifications whose only input is the number of claims in which the insurance taker was guilty. Low efficiency values are also observed under models B1, B2, and B12, all of them containing input B, cost of accidents. We conclude that what makes this UOA efficient is a good performance under models that include input A, the number of claims. This could mean that young females in intermediate range cars in the North, if one takes into

account that they make relative few claims, or pay high premiums to the company. This is a segment of the market that is worth keeping.

Finally, for completeness, we will look at IFC- the remaining UOA that is efficient under the full model- ABC12. This is, again, young females driving intermediate range cars, this time in the Centre of the country. The specifications under which this UOA is inefficient are B1, B2, and B12, all of them associated with input B, the cost of the claims. We deduce that the individuals in this segment of the market may be involved in relatively expensive accidents, but that the premium they pay fully compensates for this, perhaps because they are involved in relatively few accidents, and are not often the guilty party.

We have seen that, by calculating efficiencies under a variety of models, we can obtain important information about the DEA performance of a UOA. However, the verbal analysis that we have just performed has its limitations. As the number of UOAs increases, and the number of possible specifications also increases, the details of why a UOA performs well or badly is buried under a mass of data. In what follows we show that it is possible to use multivariate analysis in order to visualise the reasons why a UOA achieves good performance under some input/output specifications but not under others.

4. Factor Analysis of DEA efficiencies

Table 4 can be treated as a multivariate data set: model specifications are the variables, and UOAs are the observations. This produces a table of 21 variables and 45 observations. There are clear correlations between the various variables, and a data reduction procedure is possible. Various data reduction techniques are available, and we choose to use Factor Analysis (FA). Prior to the FA study, a Principal Components Analysis (PCA) was performed in order to study the number of factors to be estimated; see Dunteman (1989).

Eigenvalues were considered important in PCA if they were greater than 0.8, as recommended by Joliffe (1972). Table 5 shows the eigenvalues and the variances explained. It is clear that, rather than operating with 21 variables, it is sufficient to concentrate on three principal components. The first component explains 82.43% of total variability, the addition of the second component increases total variance explained to 93.45%, and the third component brings the total variance explained to 99.03%. This was interpreted to mean that the data set can be described by means of three factors. In what follows we will concentrate on a discussion of the first two factors. The third factor was also studied, but no clear interpretation was found and no recommendations followed, and it will not be discussed in what follows.

Component	*Eigenvalue*	*% of variance*	*Cumulative %*
PC1	17.31	82.43	82.43
PC2	2.31	11.02	93.45
PC3	1.17	5.58	99.03
PC4	0.12	0.57	99.60

Table 5 Principal Component Analysis results

FA was performed on the data contained in Table 4 using the PCA approach. Factors were not rotated. As it is common practice in FA, meaning is attached to each factor after studying the correlations between the factors and the original variables. These can be seen in Table 6. The results have been ordered in terms of the correlations of the models with the first factor.

	Component		
	PC1	*PC2*	*PC3*
ABC12	0.996	0.002	0.024
ABC1	0.995	-0.001	0.022
ABC2	0.989	0.013	0.038
BC1	0.968	-0.078	0.204
BC12	0.968	-0.077	0.209
BC2	0.966	-0.061	0.210
AC2	0.962	0.222	-0.111
AC12	0.960	0.229	-0.147
AB2	0.959	-0.187	-0.164
AB12	0.959	-0.207	-0.186
AB1	0.956	-0.213	-0.183
AC1	0.941	0.274	-0.174
A2	0.934	-0.002	-0.332
A12	0.925	0.001	-0.375
A1	0.904	0.044	-0.413
B12	0.789	-0.551	0.261
B1	0.789	-0.551	0.261
B2	0.772	-0.556	0.287
C12	0.750	0.598	0.279
C2	0.748	0.590	0.299
C1	0.741	0.613	0.269

Table 6 Factor analysis: correlations between factors and variables

It can be seen in Table 6 that the correlations between the models and the first factor are positive and high, the lowest one being 0.741, and the highest one 0.996, this last correlation is associated with the saturated model ABC12. The first factor is taken as a measure of "global efficiency" over all possible specifications, as suggested by Serrano-Cinca and Mar-Molinero (2004).

To attach a meaning to the second factor we need to take into account the value of the correlations between the model and the second factor, as well as the sign of the correlations. We find that the second component has high positive correlations in the case models C1, C2, and C12, and has negative correlations with models B1, B2, and B12. Positive correlations are associated with input C (number of guilty claims), and negative correlations are associated with input B (cost of claims). The second factor can, therefore, be interpreted as a contrast between responsibility for the accident and cost of the accident.

The correlations between the models and the third factor are low, and no attempt was made to interpret the third factor.

Figure 1 plots the UOAs using as coordinates factor loadings in the first and second factor. On the right hand side of Figure 1 we find the UOAs that have high global efficiency, and on the left hand side those whose global efficiency level is low. We can observe that IFN-, PFS-, and IFC- have the highest global efficiencies. All of them are associated with young female drivers. In fact, all but one (EFM-) UOAs that involve young female drivers appear on the right hand side of the figure, indicating that, having taken into account all the elements of risk, and considering the premium they pay, young female drivers are good business for the insurance company, perhaps because premiums are too high in relation to risk. If we look at the other side, we find that older females are concentrated on the left hand side of the figure, indicating a low level of global efficiency. In other words, considering the risk involved, older females are bad business for the company, perhaps because they accumulate high no-claim bonuses. In Figure 2 we have replaced the specification of the UOS for the sex of the driver, and no clear sex effect appears, although it is clear that the points that show the highest global efficiency are all associated with females but, on the other hand, the point that has the lowest global efficiency is also associated with females (IFS+). No clear sex effect appears, although there appears to be an interaction between sex and age. We will further study this issue in what follows.

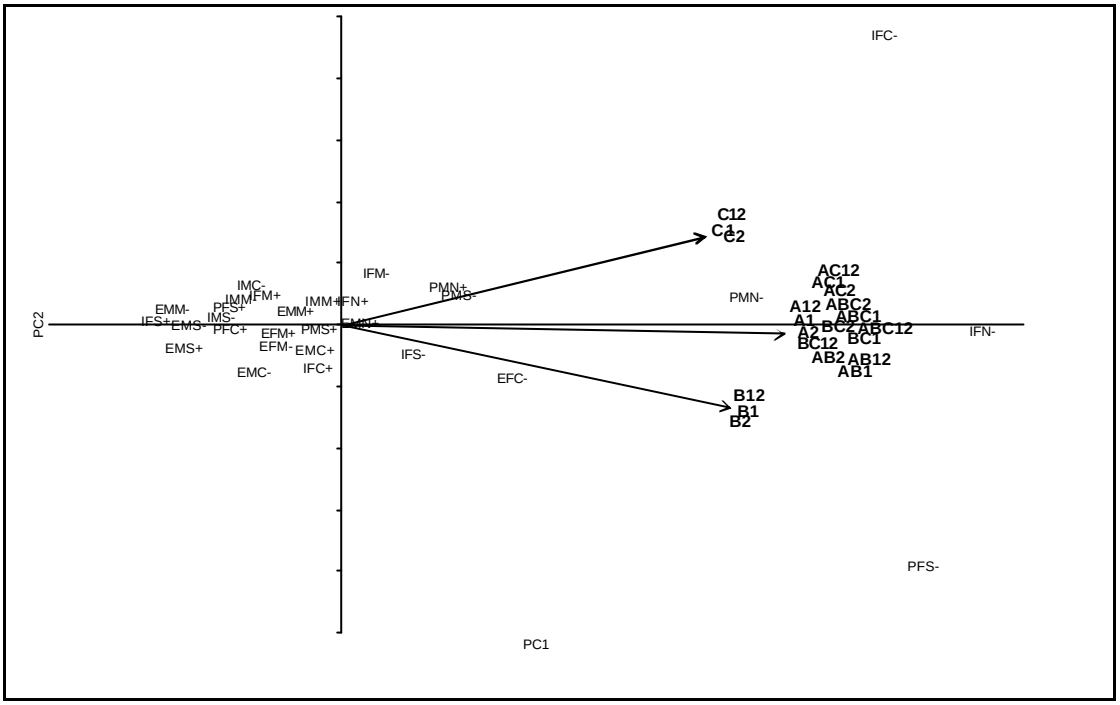

Figure 1 Plot of component loading, PC1 vs PC2

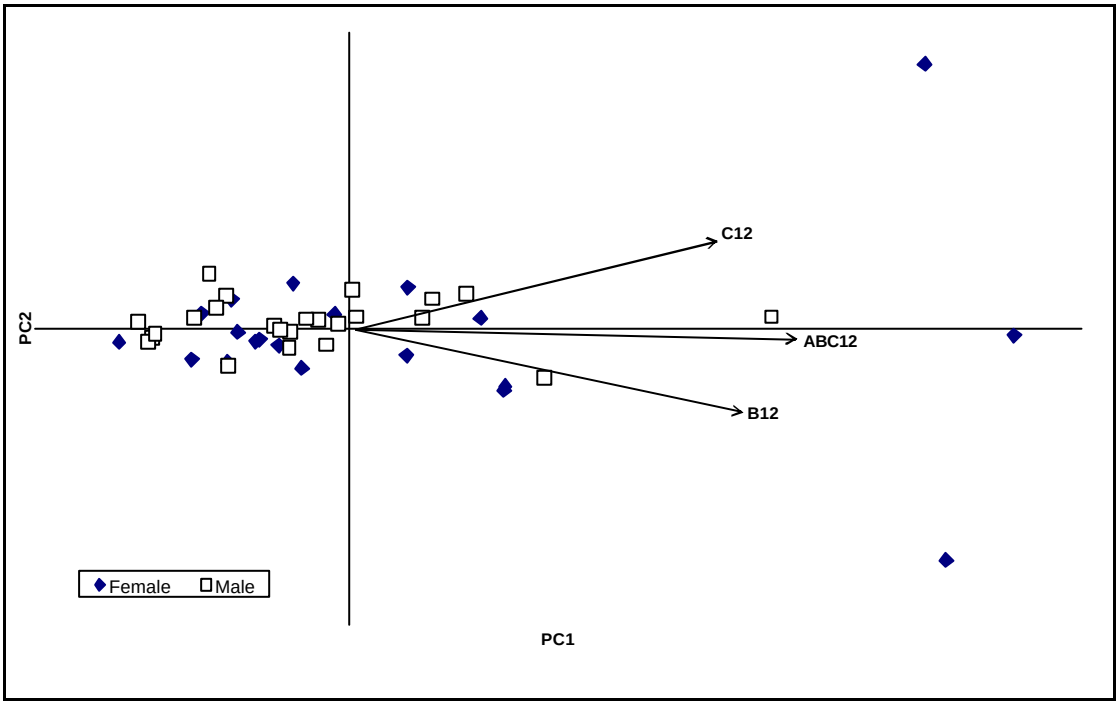

Figure 2 PC1 vs PC2 component loadings. Clients have been identified by means of their gender

Turning now to the second factor, UOAs that achieve high efficiency taking into account the risk of being involved in a guilty accident are situated towards the top of the figure, while UOAs that achieve high efficiency once the cost of the claim has been considered, are situated at the bottom of the figure. This has been found to be related with the type of car. In Figure 3, the type of car replaces the name of the UOA. We see that most Premium cars are on the top of the figure, whilst most economy cars are at the bottom of the figure. It is also apparent that premium cars tend to be located on the right hand side of the figure, while economy cars tend to be located on the left hand side of the figure. In other words, as a general rule, having taken into account the level of risk involved, premium cars are better business for the company than economy cars, perhaps because they are overcharged for the insurance.

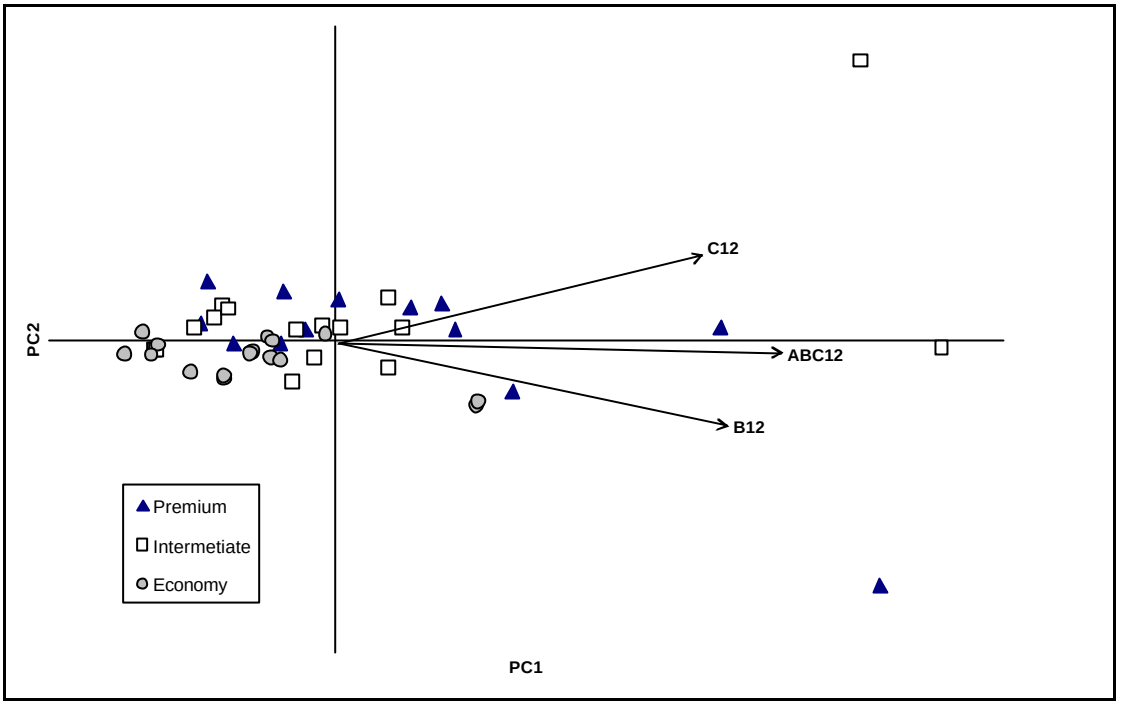

Figure 3 PC1 vs PC2 component loadings. Clients have been identified by means of their vehicle type

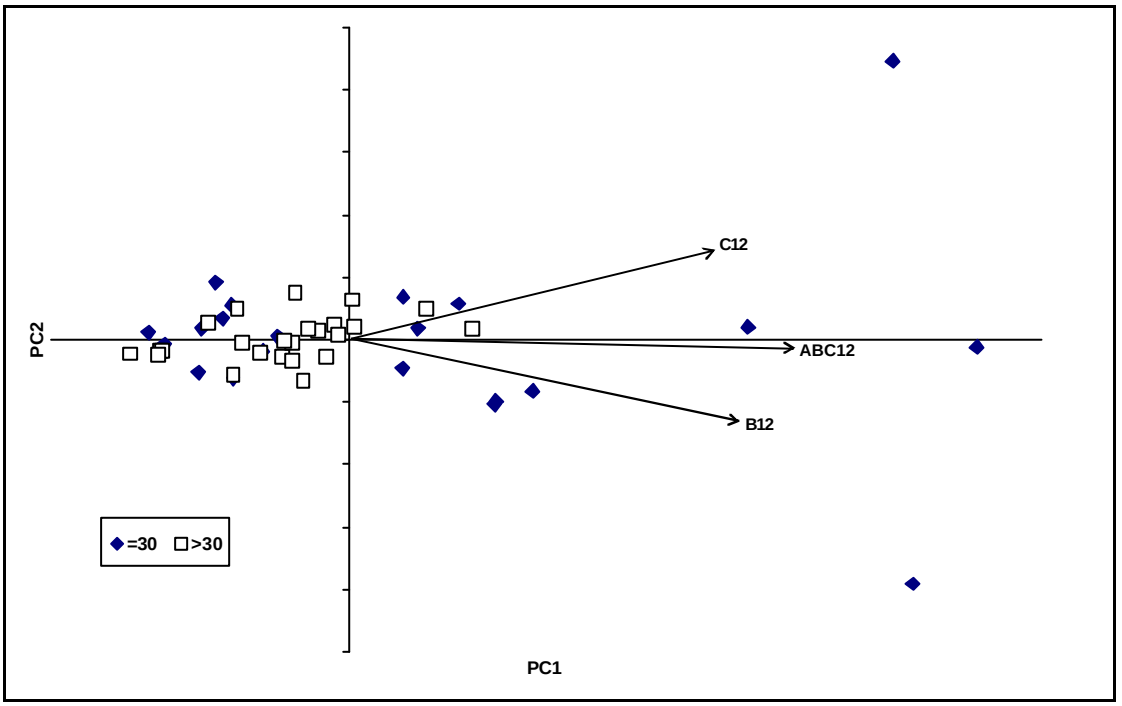

Figure 4 PC1 vs PC2 component loadings. Clients have been identified by means of their age

For completeness, the points in figure 1 have been replaced by an indicator of age in Figure 4, and by indicator of area in Figure 5.

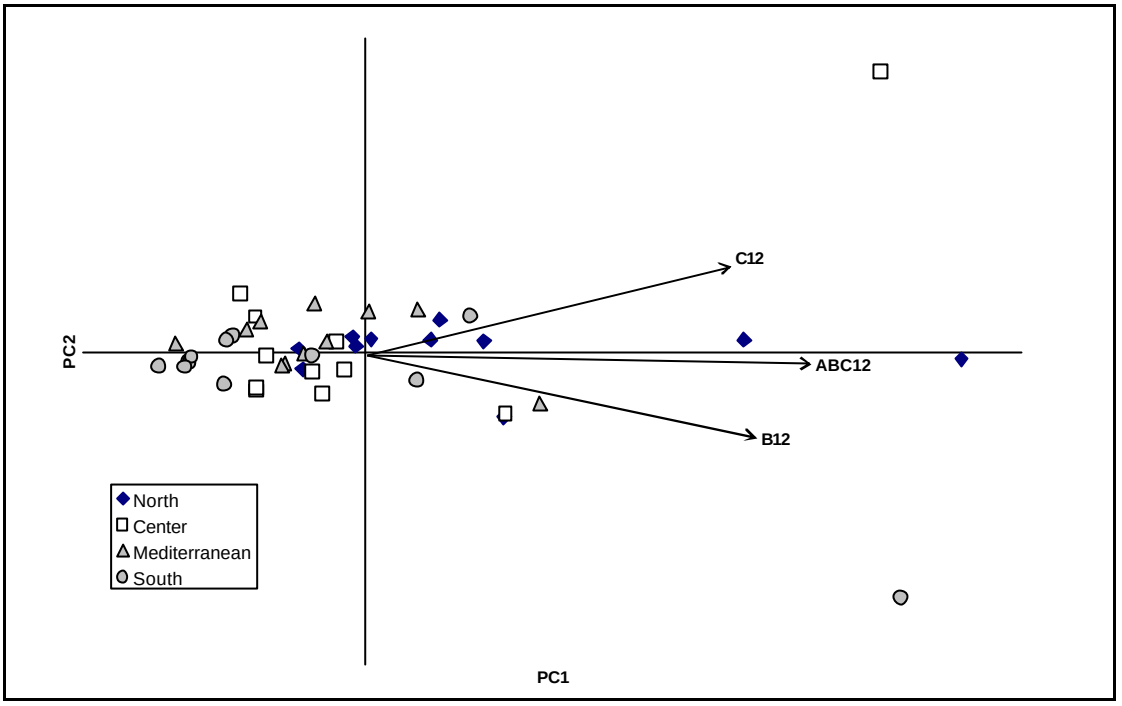

Figure 5 PC1 vs PC2 component loadings. Clients have been identified by means of their geographical area

5. Property fitting results

In order to graphically reveal how the efficiency of the UOAs depends on the various combinations of inputs and outputs that enter into the DEA specification we have used the Property Fitting (Pro-Fit) technique. Pro-Fit draws lines through Figure 1 in the same way that North-South directions are drawn in a geographical map; i.e., the line points in the direction in which the efficiencies under the model under exploration grow. For a description of how the technique works see Schiffman et al. (1981), and for its mathematical basis see Mar-Molinero and Mingers (2006). For each specification, say C12, there is a line in Figure 1 pointing out in the direction in which efficiency under that model grows. These lines are shown in Figure 1, although most of the lines coincide with the horizontal axis and it has been necessary hide some of them in order not to clutter the representation. A property of Pro-Fit is that the angle between any two lines is associated with the correlation of the efficiencies produced by the models generated by the lines. The procedure followed to estimate the

location of the lines requires the calculation of a regression: each UOA is taken as an observation, the dependent variable is the efficiency under a model, and the independent variables are the factor loadings; three factor loadings have been included in the regression. The coefficients of the regression have been normalised so that their squares add up to one. We refer to normalised regression coefficients as "directional cosines". Table 7 shows the regression results, including the coefficient of determination, R^2, for every regression. The results are excellent; the lowest R^2 value was 0.978.

	Directional cosines				
	f 1	*f 2*	*f 3*	*F*	*Adj R^2*
A1	0.909	0.045	-0.415	1261.181	0.988
	55.883**	2.738**	-25.556**		
AB1	0.960	-0.214	-0.184	1736.591	0.992
	69.258**	-15.412**	-13.250**		
ABC1	1.000	-0.001	0.022	1363.040	0.989
	63.931**	-0.046	1.394		
AC1	0.945	0.275	-0.175**	1484.181	0.990
	63.077**	18.378**	-11.668**		
BC1	0.975	-0.078	0.206	891.573	0.984
	50.446**	-4.039**	10.657**		
B1	0.791	-0.552	0.262	2410.832	0.994
	67.306**	-46.958**	22.301**		
C1	0.742	0.614	0.270	4509.064	0.997
	86.269**	71.421**	31.368**		
A2	0.942	-0.002	-0.335	805.665	0.982
	46.331**	-0.116	-16.446**		
AB2	0.968	-0.189	-0.165	756.187	0.981
	46.109**	-8.989**	-7.859**		
ABC2	0.999	0.013	0.039	649.945	0.978
	44.120**	0.572	1.712		
AC2	0.968	0.223	-0.112	1084.331	0.987
	55.227**	12.737**	-6.380**		
BC2	0.975	-0.062	0.212	719.621	0.980
	45.314**	-2.875**	9.862**		
B2	0.777	-0.560	0.289	1074.562	0.987
	44.090**	-31.790**	16.406**		
C2	0.749	0.590	0.300	4223.863	0.997
	84.360**	66.470	33.714**		
A12	0.927	0.001	-0.375	4124.636	0.996
	103.101**	0.151	-41.763**		
AB12	0.960	-0.207	-0.187	3676.337	0.996
	100.854**	-21.752**	-19.606**		
ABC12	1.000	0.002	0.024	1749.929	0.992
	72.435**	0.135	1.720		
AC12	0.962	0.230	-0.148	3758.182	0.996
	102.141**	24.407**	-15.688**		
BC12	0.975	-0.078	0.210	951.257	0.985
	52.062**	-4.168**	11.221**		
B12	0.791	-0.552	0.262	2410.832	0.994
	67.306**	-46.958**	22.301**		
C12	0.751	0.598	0.279	8053.763	0.998
	116.713**	93.024**	43.428**		

**Significant al the 0.01 level.

Table 7 Pro-Fit analysis: linear regression results

Normalised directional vectors fall neatly into three groups. On the upper right hand side quadrant of Figure 1 we see the vectors associated with C1, C2, and C12. All these models share a single input, the number of claims in which the insured person is guilty. On the bottom right hand side we find B1, B2, and B12; all of them with the average cost of a claim as the only input. All other models bunch along the right hand side of the horizontal axis.

Figure 1, augmented with the Pro-Fit vectors makes it possible to interpret how the different risk factors contribute to efficiency. Take, for example, PFS-, a case studied above. This UOA is located in an extreme position in the direction of the vectors B12, B1, and B2, indicating that it is efficient under these three models; it is very far to the right, in the direction of the vector associated with the saturated model ABC12; and it no longer projects as an extreme case under directional vectors C12, C1, C2. This is in line with the results in Table 4 that show that PFS- is not efficient under models C12, C1, and C2, and is consistent with the earlier discussion. The advantage of the visualisation over examination of Table 4 is that one can make an intuitive assessment of why a UOA is or is not efficient under the various models.

A question that we may like to answer is: how are the different risk factors associated with the efficiency of the policy under assessment? To answer this question we will only take into account the efficiencies under the saturated model, as the directional line associated with this model is almost undistinguishable from the horizontal axis.

6. Risk factors and efficiencies

In order to assess how the risk factors are associated with efficiency we have conducted a regression in which the dependent variable is the efficiency under model ABC12,

while independent variables are dummy variables: gender takes the value 1 for women and 0 for men; age takes the value 1 for the under 30 year olds and 0 for over 30 year olds; a set if four dummy variables indicate the area; and a set of three dummy variables indicate the type of car insured. The model also contained interactions between age and gender; between age and type of car; between gender and type of car; and between type of car and area. This way of proceeding is equivalent to conducting an analysis of variance (ANOVA); Neter et al. (1996). We take as the basis for comparison mature males in the South who drive an economy car (EMS+), one of the combinations with the lowest efficiency. The aim is to find how, variations in the various risk factors, change the efficiency of this model. We also want to explore any interactions that may exist between the factors. The results are shown in Table 8.

Using standard methodology, we found that there are statistically significant interactions between gender and age, between age and premium cars, between age and intermediate cars, and between premium cars and location in the Centre of the county. There are significant regional effects. We now proceed to describe these.

The largest increase in efficiency is obtained when young people drive premium cars. There are two possible explanations for this, the first one is that insurance premiums are too high for this group; the second explanation is that this group does not qualify for no-claims discounts and the net income is relatively large. Young people who drive intermediate cars are also good business for the company, although the impact on efficiency is lower.

The second largest effect, in terms of augmenting efficiency, is related to the interaction between age and gender: young women are good business for the company. The explanation may be that, having taken risk into account, young women pay high insurance premiums and do not qualify for no-claims bonuses. Notice that this is not the case for young

men: they are not so efficient for the company indicating that, when taking into account risk, they pay relatively low insurance premiums.

The last significant interaction is found between premium cars and location in the Centre of the country. The effect is negative: premium cars in the Centre of the country generate less revenue to the insurance company than premium cars anywhere else. This may be related to an attempt to obtain this segment of the market by offering discounts, as the Centre region is where the capital of the country is located, and the number of premium cars may be very high.

Considering interactions and principal effects one can observe that, in general, efficiency- or income for the company- increases when policies are issued to young people, to females, involve premium cars, and are sold in the North of the country. How are we to interpret these findings? We have said at the beginning that the company specialises in drivers in the South. This is where its market is, and within this market the largest segment is mature men driving economy cars. Given the level of the competition in the insurance business, this company may be working with its main source of income under very low margins and, compensating by charging more to those drivers who are more peripheral from the point of view of the company.

7. Conclusions

In this paper we have compared insurance policies under various risk factors, taking into account the revenues that they generate. We have applied a factorial experimental design to the customers of the insurance company, taking into account the various risk factors and the levels at which they present themselves. Having defined each policy, we have standardised for the number of customers that take such policies, in order to control for the

level of risk. Each of these standardised policies has been treated as a unit of assessment in a data envelopment analysis. We were interested in assessing how the risk factors affect the level of efficiency achieved from each policy, and this was found through a procedure involving estimating a variety of models and doing factor analysis on the results. Finally, an analysis of variance has revealed which risk factors influence efficiency, and how they interact.

The analysis has revealed that certain combinations of factors generate more revenue to the insurance company than others. For example, young women are good business, as are young people driving expensive cars. It has been suggested that this is the result of the changing competitive environment in which the company operates, which forces it to give high discounts to the main segment of its market in order to maintain its flow of funds.

This analysis has shown how to assess insurance products under the various factors that influence risk and revenue. This is, of course, not the full story, since one has to take into account, not only the value of an insurance contract to the company, but also the total number of contracts and the relationship between price and demand. There is still much more work to be done.

References

Boj, Eva, Claramunt, M. Mercedes, and Fortiana, Josep. (2004). *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación*, Fundación Mapfre Estudios. Instituto de Ciencias del Seguro.

Cotter, Kevin D., and Jensen, Gail A. (1989). "Choice of purchasing arrangements in insurance markets," *Journal of Risk and Uncertainty* 2(4), 405-414.

Cummins, J.David, and Rubio-Misas, Maria. (1998). "Efficiency in the Spanish insurance industry," *Working paper, The Warton School*, University of Pennsylvania.

Cummins, J.David, and Zi, Hongmin. (1998). "Comparison of frontier efficiency methods: and application to the US life insurance industry," *Journal of Productivity Analysis* 10, 131-152.

Dionne, Georges, and Vanasse, Charles. (1992). "Automobile insurance ratemaking in the presence of asymmetrical information," *Journal of Applied Econometrics* 7(2), 149-166.

Doherty, Neil A., and Schlesinger, Harris. (1983). "Optimal insurance in incomplete markets," *Journal of Political Economy* 91(6), 1045-1054.

Doherty, Neil A., and Schlesinger, Harris. (1990). "Rational insurance purchasing: consideration of contract non-performance," *Quarterly Journal of Economics* 105, 143-153.

Doherty, Neil A. and Dionne, Georges. (1993). "Insurance with undiversifiable risk: Contract structure and organizational form of insurance firms," *Journal of Risk and Uncertainty* 6 (2), 187-203.

Dunteman, George H. (1989). *Principal components analysis*, Series: Quantitative Applications in the Social Sciences, nº 69. Series Editor: Michael S. Lewis-Beck, University of Towa.

Fecher, Fabienne, et al. (1993). "Productive performance of the French insurance industry," *Journal of Productivity Analysis* 4, 77-93.

Fukuyama, Hirofumi. (1997). "Investigating productive efficiency and productivity changes of Japanese life insurance companies," *Pacific-Basin Finance Journal* 5, 481-509.

Guillén, Monserrat, et al. (2005). *El Seguro de automóviles: estado actual y perspectiva de la técnica actuarial*, Fundación Mapfre estudios. Instituto de Ciencias del Seguro.

Henriet, Dominique, and Rochet, Jean-Charles. (1986). "La logique des systèmes Bonus-Malus en assurance automobile: une approche thèorique, " *Annales d'Èconomie et the Statistique* 1, 133-152.

Jenkins, Larry, and Anderson, Murray. (2003). "A multivariate statistical approach to reducing the number of variables in data envelopment analysis," *European Journal of Operational Research* 147, 51-61.

Johnson, Eric J., et al. (1993). "Framing, probability distortions, and insurance decisions," *Journal of Risk and Uncertainty* 7(1), 35-51.

Joliffe, Ian T. (1972). "Discarding variables in Principal Components Analysis," *Applied Statistics* 21, 160-173.

Lemaire, Jean. (1985). *Bonus-Malus systems in automobile insurance actuarial models*: Kluwer Academic Publishers.

Mahul, Olivier. (2000). "Optimal insurance design with random initial wealth," *Economics Letters* 69, 353-358.

Mar-Molinero, Cecilio, and Mingers, John. (2006). "Mapping MBA Programmes: an alternative analysis," Forthcoming in the *Journal of the Operational Research Society.* Available in electronic form.

Mayers, David, and Smith, Clifford W. (1983). "The interdependence of individual portfolio decisions an the demand for insurance," *Journal of Political Economy* 91, 304-311.

Meyer, Donald J., and Meyer, Jack. (1998). "Changes in background risk and the demand for insurance," *The Geneva Papers on Risk and Insurance Theory* 23, 29-40.

Neter, John, et al. (1996). *Applied linear statistical models*: Irwin, Chicago, USA.

Parkin, David, and Hollingsworth, Bruce. (1997). "Measuring production efficiency of acute hospitals in Scotland, 1991-94: validity issues in data envelopment analysis," *Applied Economics* 29, 1425-1433.

Pedraja-Chaparro, Francisco, Salinas Jimenez, Javier, and Smith, Peter. (1999). "On the quality of the Data Envelopment Analysis model," *Journal of the Operational Research Society* 50, 636-645.

Pujol, Maria, and Bolancé, Catalina. (2004). *La matriz valor fidelidad en el análisis de los asegurados en el ramo del automóvil*, Fundación Mapfre estudios. Instituto de Ciencias del Seguro.

Sarrico, Cla'udia S., et al. (1997). "Data envelopment analysis and university selection," *Journal of the Operational Research Society* 48, 1163-1177.

Schiffman Susan S., Reynolds M. Lance, and Young Forrest W. (1981): *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press: London.

Schlesinger, Harris, and Schulenburg, J.Matthias. (1987). "Risk aversion and the purchase of risky insurance," *Journal of Economics* 47, 309-314.

Serrano-Cinca, Carlos, Fuertes-Callen, Yolanda, and Mar-Molinero, Cecilio. (2005). "Measuring DEA efficiency in Internet companies," *Decision Support Systems* 38, 557-573.

Serrano-Cinca, Carlos, and Mar-Molinero, Cecilio. (2004). "Selecting DEA specifications and ranking units via PCA," *Journal of the Operational Research Society* 55, 521-528.

Tone, Kaoru, and Sahoo, Biresh K. (2005). "Evaluating cost efficiency and returns to scale in the Life Insurance Corporation of India using data envelopment analysis," *Socio-Economic Planning Sciences* 39, 261-285.

UNIVERSITY OF KENT

http://www.kent.ac.uk/kbs/research-information/index.htm