

Predicting Insurance Investment: A Factor Analytic Approach

¹Samuel Obi-Nnamdi Agwuegbo, ²Adetunji Philip Adewole and ³A.N. Maduegbuna

¹Department of Statistics, University of Agriculture, Abeokuta

²Department of Computer Science, University of Agriculture, Abeokuta

³Department of Computer Science, Paul University, Awka

Abstract: Problem statement: In the last decade growing attention has been paid to the pattern of investments by the insurance industry and the question of how to evaluate such investments. In an economy where the capital market is huge and active, mathematical considerations come into play in the selection of investments to ensure yield maximisation. **Approach:** This study examined the use of factor analysis as an emerging technique for the analysis of insurance investment in Nigeria. **Results:** The proposed technique described a number of methods designed to analyze interrelationships within the investment variables in terms of few underlying but unobservable random quantities called factors. The factors were constructed in a way that reduces the overall complexity of the data by taking advantage of inherent interdependencies. **Conclusion:** The result obtained through this approach were promising and shows that two principal components of the factor loadings have a cumulative proportion of variance accounted for 94.5% of the total variations of the investments pattern.

Key words: Analysis, forecasting, regression analysis, multivariate analysis, principal components, eigenvalues

INTRODUCTION

Economic recession and global mettle-down have brought the question of insurance company investment to the forefront. Growing attention has shifted to the pattern of investments by the insurance industry and the question of how to evaluate such investments. The annual published accounts of the insurance industry are a valuable instrument for the assessment of the investment pattern of the insurance industry.

The investment of insurance funds serves as a shield for insurance against predictable underwriting losses which are more prominent than profits. The size of funds held by the Insurance industry represents a reasonable percentage of the Country's total invisible funds generated by the capital market.

Because of the need to invest funds to provide for their future liabilities, insurance companies look for certain desirable characteristics in the asset which they hold. The assets duly are expected to possess good stable yield, stable capital growth, liquidity and marketability and as well as good tax status. Of course, no assets have all of these attributes but by an intelligent mix of different types, the insurer can obtain a portfolio that has broadly desirable characteristics.

In an economy where the capital market is huge and active, mathematical considerations come into play

in the selection of investments to ensure yield maximisation. Regression analysis enjoys very wide use and application in marketing and financial research. Most researchers' in this field are ignorant of the fact that the techniques of regression analysis are critically dependent upon the assumption that the so called independent variables are truly independent. If they are correlated among themselves, problems of multicollinearity and singularity arise and may cause the entire analysis to fail. Factor analysis may be used to reduce this problem. Factor analysis, Johnson and Wichern (2002); Lawley and Maxwell (1971); Kent and Bibby (1980) and Krzanowski (2000) is given to a body of techniques concerned with the study of such interrelationships. A factor analytic study is carried out in this study, to determine linear combinations of the variables that assist insurance investment in Nigeria. The essential purpose of factor analytic approach in this study is to describe a number of methods designed to analyze interrelationships within the investment variables. Factor analysis in this study attempts to explain the interrelationships in terms of few underlying but unobservable random quantities called factors (Johnson and Wichern, 2002; Mardia *et al.*, 1979). The factors are constructed in a way that reduces the overall complexity of the data by taking advantage of inherent interdependencies. Under factor analytic

model, each response variates are represented as a linear function of small number of unobservable common-factor variates and a single latent specific variate. The common factors generate the co-variances among the observable responses, while the specific terms contribute only to the variances of their particular responses. This approach can be thought of as a robust method for reducing the number of explanatory variables in a model in terms of smaller number of variables.

MATERIALS AND METHODS

Data on Insurance assets (both life and non-life) were extracted from the Central Bank of Nigeria (CBN) Statistical bulletin as shown in Table 1. To descry the hidden underlying factors which generated a dependence or variation in Table 1, we adopted a factor analytic approach.

Let the investment portfolio in Table 1, be described by an n -dimensional random variables $X_1, X_2, \dots, X_n$. The vector random variables X_i have a non singular multi-normal distribution. To describe the data, X_i is represented as an $n \times p$ data matrix, observations of

the variables $X' = (x_1, \dots, x_n)$. For convenience $E(X) = 0$ and the covariance matrix of these responses will be denoted by $E(XX') = \Sigma$. The covariance Σ of the standardized variables is nothing more than the correlation matrix. Our interest is to develop the factor model in relation to the data matrix and this can be parsimoniously be written as:

$$Y = \Lambda F + E \quad (1)$$

Where:

Y = An $n \times p$ data matrix

Λ = The $p \times k$ matrix of factor loadings

F = The $k \times n$ matrix of factor scores

E = An $n \times p$ matrix of residuals or error terms

Here k is a scalar denoting the number of factors to be used. It is always less than p the number of variables. Equation 1 is the fundamental model equation for all forms of R-mode factor. It states that each observed variable is weighted sum of factors plus an error term or residual.

Table 1: Assets of all insurance companies

Year	Government securities	Stock, shares and bonds	Mortgages and loans	Cash and bills receivable	Miscellaneous
1970	6,733.0	9,517.0	7,573.0	20,424.0	2,206.0
1971	10,840.0	9,239.0	7,404.0	30,063.0	3,687.0
1972	16,832.0	11,853.0	11,647.0	33,915.0	9,667.0
1973	22,635.0	13,386.0	12,943.0	44,982.0	14,847.0
1974	19,660.0	18,074.0	20,087.0	47,830.0	33,597.0
1975	29,105.0	20,668.0	23,889.0	64,893.0	48,716.0
1976	42,434.0	30,572.0	38,093.0	82,531.0	67,187.0
1977	61,427.0	37,236.0	58,137.0	129,677.0	121,880.0
1978	78,092.0	53,620.0	72,937.0	144,612.0	174,054.0
1979	91,895.0	70,736.0	93,535.0	138,369.0	253,665.0
1980	109,963.0	121,926.0	180,986.0	195,519.0	340,496.0
1981	133,883.0	145,453.0	147,266.0	159,307.0	428,339.0
1982	131,722.0	190,389.0	217,587.0	177,173.0	421,551.0
1983	204,678.0	169,496.0	70,673.0	145,598.0	381,736.0
1984	212,234.0	199,496.0	88,720.0	196,608.0	527,604.0
1985	460,637.0	277,681.0	106,076.0	236,078.0	662,864.0
1986	822,644.0	290,626.0	112,595.0	289,018.0	806,144.0
1987	850,466.0	324,255.0	146,097.0	280,485.0	924,767.0
1988	1,301,149.0	326,049.0	182,337.0	534,138.0	1,049,940.0
1989	1,836,882.0	414,653.0	112,854.0	586,001.0	1,942,475.0
1990	1,852,146.0	498,160.0	176,304.0	518,209.0	2,777,283.0
1991	562,923.0	568,903.0	227,008.0	1,034,552.0	3,258,116.0
1992	628,882.0	987,437.0	909,505.0	1,603,900.0	5,695,873.0
1993	3,856,351.0	5,094,023.0	1,283,881.0	2,189,626.0	6,002,170.0
1994	5,417,112.0	2,078,444.0	628,561.0	2,340,657.0	6,894,799.0
1995	981,338.0	1,817,680.0	666,800.0	3,158,194.0	19,401,775.0
1996	1,662,092.0	4,232,125.0	583,782.0	3,551,990.0	5,537,913.0
1997	2,325,002.0	5,191,335.0	984,860.0	2,846,231.0	8,660,972.0
1998	3,368,788.0	5,350,113.0	1,236,582.0	3,696,894.0	8,723,292.0

Source: Central bank statistical bulletin

The product ΛF produces a vector of estimates of X the vector E represents the difference between this estimate and the observed vector. These residuals are assumed to be uncorrelated with the factors. From the properties of the latent variates it follows that the covariance matrix of the observable and common-factor variates can be written as:

$$E[(X - \Lambda F)(X - \Lambda F)'] = \Psi \quad (2)$$

Evaluation of the expectations in Eq. 2 shows that:

$$\Sigma = \Psi + \Lambda \Lambda' \quad (3)$$

Where:

Σ = The $p \times p$ population covariance matrix of the observed variables

Λ = The $p \times k$ matrix of factor loadings

Ψ = The $p \times p$ residual covariance matrix

The factors are designed to account maximally for the inter-correlations of the variables. The diagonal elements $\Lambda \Lambda'$ are called the communalities of the responses. By choosing different orthogonal transformations, an infinity of loading matrices can be computed from Λ which would lead to the same covariance.

Equation 3 presents, in matrix form, the complete factor model for the variance covariance of the observed variables. The model holds for an R-mode factor model; since the relationship within the set of p variables is regarded as reflecting the correlations of each of the variables with k mutually uncorrelated underlying factors.

In practice the parameters of the factor models are never known and must be estimated from sample observations. The information in the sample covariance matrix S is sufficient for the estimation of the factor parameters. With uncorrelated factors as in Eq. 3, a reasonable criterion for fitting the model to the data is to make $S - \Lambda \Lambda' - \Psi$ as small as possible, by choosing Λ so that the sum of squares of all the elements of $S - \Lambda \Lambda' - \Psi$ is minimized. But the diagonal elements of Ψ are unique variance that must be estimated from the data, together with the factor loading matrix Λ . Of course, the magnitude of the communalities $\Lambda \Lambda'$ is dependent upon the number of factors k , that are retained. If $k = p$, Ψ will vanish in Eq. 3 and our problem is equivalent to Principal Component Analysis (PCA). In cases where $k < p$, the matrix of parameters Λ which are the loadings

on the factors and the unique variance ψ is estimated. The usual assumption is that $k < p$ factors are retained. There are mathematical criteria as well as some subjective decisions involved in determining the number of factors k to be extracted, in locating these factors prior to rotation and in rotating the factor structure. By using Kaiser's varimax criterion for rotation to a simple structure, the best fit or most useful fit of the factors to the data is provided.

Most factor analyses are implemented by using standardized variables. Standardization is in turn effected because in many problems the raw variables reflect widely differing units of measurement. By standardizing the variables to mean zero and unit standard deviation, the impact of units of measurement on the final solution is removed.

RESULTS

A correlation analysis of the Assets is as shown in Table 2.

Table 2 contains the simple pair wise correlations of the investment variables.

Each of this investment mix in Table 2 is measured on an annual basis. For simplicity, the assets have been transformed to standardized form.

Given these result in Table 2, it is clear that variables X_1 and X_2 , X_2 and X_3 , X_3 and X_4 and X_4 and X_5 form groups. Variable 4 is closer to the (3, 4) and (4, 5) groups. The factor-loading matrix is the key outputs of a factor-analytic solution. Specifically, the estimated factor loadings are the sample principal component coefficients (eigenvectors of R), scaled by the square root of the corresponding eigenvalues. A rotation of the factors revealed a simple structure and aids interpretations. Rotated principal-component estimates of factor loadings based upon the sample correlation matrix R are presented in Table 3, with the estimated factor loadings, communalities, eigenvalues and proportion of total (standardized) sample variance explained by each factor for $k = 1$ and $k = 2$ factor solutions.

Table 2: Correlation Matrix of Insurance Asset

Variables	(X_1)	(X_2)	(X_3)	(X_4)	(X_5)
Mortgage loans (X_1)	1.00	0.89	0.50	0.58	0.55
Government securities (X_2)	0.89	1.00	0.71	0.69	0.53
Stock, shares and bonds (X_3)	0.50	0.71	1.00	0.90	0.65
Cash and bills Receivable (X_4)	0.58	0.69	0.90	1.00	0.86
Miscellaneous (X_5)	0.55	0.53	0.65	0.86	1.00

Table 3: Rotated principal component estimates of factor loadings

Variables	Estimated factor loadings		Communalities
	F ₁	F ₂	
Mortgage loans (X ₁)	0.415	-0.628	0.567
Government Securities (X ₂)	0.455	-0.480	0.437
Stock, shares and bonds (X ₃)	0.452	0.300	0.294
Cash and bills receivable (X ₄)	0.483	0.359	0.362
Miscellaneous (X ₅)	0.428	0.396	0.340
Eigen values	3.750	0.760	
Cumulative proportion of total (standardized) sample variance	0.750	0.190	

The eigenvalues in Table 3 represent the lengths of these axes and the eigenvectors of the matrix yield the principal axes of the ellipsoid. The first eigenvalues is 3.75, which is the length of the first vector while the second eigenvalues is 0.76. These eigenvalues represent the lengths of the two principal axes. The principal axes also represent the total variance of the dataset and each account for an amount of the total variance equal to the eigenvalues divided by the trace. The first principal axis contained about 75% of the total variance, while the second axis represents only 19%. The elements of the eigenvectors which are used to compute the scores of observations are called loadings. They are simply coefficients of the linear equation which the eigenvector defines. The first two eigenvectors accounted for 94% of the variance in the dataset. The communalities correspond to the portion of the variance of each variable retained in the factors. The communalities are the sum of squares of the loadings in the factor matrix and are equal to the original variances. The magnitude of the communalities is dependent upon the number of factors that are retained. The factors are defined to account maximally for the intercorrelations of the variables.

DISCUSSION

This study condenses the information contained in the asset portfolio by providing reasonable explanations in terms of a few interpretable factors. The factors explain the correlations that have been discovered among the investment variables. The factor loading matrix Λ , is regarded as containing coefficients that describe the composition of the factor in terms of the original variables.

CONCLUSION

Factor analysis is used in this study to provide a way of explaining the observed variability in the investment (assets) pattern of the Nigerian insurance

industry. The factor analytic solution outlined above resulted from the application of the principal-components procedure to the data. The objective of the principal-components analysis was to transform the set of interrelated investment variables into a set of unrelated linear combinations of these variables. Under the assumption of the classical-factor model, the correlations between investment variables in Table 2 are reproduced by means of the common-factor coefficients alone.

The entries in Table 3 simply are correlations between the variables and the factors. These correlations are called factor loadings. Since the entries are variable-factor correlations, their square indicates the proportion of variation in the variable that is accounted for by the factor. The proportion of variance in the study accounted for two factors. The two factors derived from the study adequately captured the information contained in the five investment variables. The first rotated factor is largely concerned with all the variables. All the five variables are positive in the first loadings. Variables 3, 4 and 5 have positive loadings on the second factor, whereas variables 1 and 2 have negative loadings on the second factor. The second factor weighs mortgage loans against government securities. The first two eigenvalues of R are the only eigenvalues greater than unity. Moreover, $k = 2$ common factors accounted for 94% cumulative proportion of total (standardized) sample variance.

REFERENCES

- Johnson, R.A. and D.W. Wichern, 2002. Applied Multivariate Statistical Analysis. 5th Edn., Prentice-Hall, Englewood Cliffs, ISBN: 0131219731, pp: 767.
- Krzanowski, W.J., 2000. Principles of Multivariate Analysis. A User's Perspective. 1st Edn., Oxford University Press, USA., ISBN: 10: 0198507089, pp: 608.
- Kent, J.T. and J.M. Bibby, 1980. Multivariate Analysis. 2nd Edn., Academic Press, New York, ISBN: 0124712509, pp: 521.
- Lawley, D.N. and A.E. Maxwell, 1971. Factor Analysis as a Statistical Method. 2nd Edn., American Elsevier Pub. Co., London, ISBN: 0444196161, pp: 153.