

STATISTIQUE ET ANALYSE DES DONNÉES

THIERRY FOUCART

Programmation de la méthode STATIS sur micro-ordinateurs

Statistique et analyse des données, tome 7, n° 3 (1982), p. 68-82

http://www.numdam.org/item?id=SAD_1982__7_3_68_0

© Association pour la statistique et ses utilisations, 1982, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*

<http://www.numdam.org/>

STATISTIQUE ASSISTEE PAR ORDINATEUR

PROGRAMMATION DE LA METHODE STATIS SUR MICRO-ORDINATEURS.

Thierry FOUCART

Laboratoire de Statistique et Probabilité
I.U.T. Kercado
56008 VANNES

Résumé : L'utilisation généralisée des micro-ordinateurs conduit évidemment au développement de logiciels conçus pour ces calculateurs. Nous proposons ici l'analyse de la méthode STATIS, légèrement modifiée, en vue de sa programmation sur ce type de matériel. L'implantation du programme a été réalisée sur un TRS 80 Model I fonctionnant avec un BASIC très proche du BASIC de Microsoft.

Abstract : The general spreading use of Micro-computers obviously leads to the development of softwares specially devised for these calculators. We propose here the analysis of the STATIS method, slightly altered, for the purpose of adapting it on this type of hardware. The lay-out of this programme has been achieved on a TRS 80 Model I, working on a BASIC language, very close to the Microsoft BASIC.

Mots clés : Analyse factorielle, étude d'une famille d'opérateurs.

1 - LES DONNEES ET LES RESULTATS

Nous précisons dans ce paragraphe le type de données analysées par le logiciel et la forme des résultats que l'on obtient. Le logiciel utilise

un certain nombre de fichiers et nous supposons que les données sont sauvegardées sur certains d'entre eux, créés par l'utilisateur selon des conventions précises au fur et à mesure. On pourra se reporter [1] pour la programmation en Basic.

1.1 - Caractéristiques des données.

Les données initiales analysées par le logiciel sont constituées de matrices symétriques.

Sur le plan méthodologique, cela signifie :

- i) dans le cas des opérateurs de covariance, la métrique définie sur l'espace euclidien \mathbb{R}^P a une matrice M telle que VM soit symétrique : en général $M = I$. On pourra donc comparer des matrices de covariance ou de corrélation
- ii) dans le cas des opérateurs de produit scalaire, les poids des individus doivent être égaux (on pourrait envisager de traiter les matrices de la forme $P^{1/2} W P^{1/2}$ [2]).

Sur le plan informatique, la symétrie des matrices permet de n'en écrire que la moitié (diagonale comprise) sur les fichiers : on limite ainsi la place occupée sur la disquette ; les matrices sont écrites sur un fichier chacune ; leur nombre maximum est fixé à 50 et leur taille à 10, ces deux nombres pouvant être facilement modifiés.

Les programmes calculant ces matrices sont faciles à rédiger en BASIC : ce sont des programmes classiques de calcul de covariances ou de produits scalaires [2].

Par contre, étant donné la spécificité des fichiers suivants, nous en donnons la définition précise ; il s'agit de :

- PARAM/PAR : paramètres généraux de l'analyse
- IDENTIFI/DAT : identificateurs (représentation du compromis et des infrastructures).

1.2 - Description des fichiers initiaux PARAM/PAR et IDENTIFI/DAT.

Fichier PARAM/PAR

P	nombre de tableaux
L	taille des matrices
IT	type des matrices : 1 covariance 2 produit scalaire

- Q métrique (interstructure : canonique : 1 normée : 2)
 P∅ nombre d'axes (interstructure)
 L∅ nombre d'axes (compromis et intrastructures)
 OP options : 1 interstructure
 2 interstructure + compromis
 3 interstructure + compromis + composantes principales ou axes principaux
 4 interstructure + compromis + variables (ou individus)
 5 interstructure + compromis + composantes ou axes principaux + variables ou individus
- A\$ fichier contenant les noms des tableaux
 B\$ fichier contenant les identificateurs des variables (ou des individus)
- J1,J2,...,0,0 axes des plans représentés (interstructure)
 J1,J2,...,0,0 id. (intrastructures et compromis).

Les fichiers A\$ et B\$ sont créés par l'utilisateur.

Fichiers IDENTIFI/DAT

Ce fichier contient les identificateurs des éléments représentés dans les intrastructures ; il dépend de l'option choisie OP et du type de tableau IT :

Identificateurs	type des opérateurs		option				
	covariances	Escoufier	1	2	3	4	5
Variables (ou individus) du compromis	ZB(I)	id.	non	oui	oui	oui	oui
Composante principale I (tableau J)	CIJ	AXIJ	non	non	oui	non	oui
Variable (ou individu) I (tableau J)	ZB(I)J	id.	non	non	non	oui	oui

Les caractères ZB(I) sont définis par l'utilisateur dans un fichier dont le nom B\$ est donné dans le fichier PARAM/PAR. Le programme IDENTIFI/STA crée les identificateurs du caractère ZB(I) associé à chaque opérateur en lui ajoutant le numéro ; ainsi, soit V3 l'identificateur de la 3ème variable dans l'opérateur de covariance compromis ; V31, V32, etc. seront les identificateurs de la 3ème variable dans les opérateurs $\Gamma_1, \Gamma_2 \dots$. La longueur des identificateurs est limitée à 4.

1.3 - Présentation des résultats.

Le logiciel permet d'effectuer les calculs de la méthode STATIS telle qu'elle est décrite en [3]. Suivant l'option choisie, sont établis les résultats suivants :

- interstructure : matrice des produits scalaires entre les opérateurs, vecteurs propres et valeurs propres, coordonnées et cosinus carrés des opérateurs suivant les axes choisis. Représentation graphique
- compromis : analyse en composantes principales du compromis ; calcul des vecteurs propres et valeurs propres, coordonnées et cosinus carrés des variables (resp. individus) dans le cas des opérateurs de covariance (resp. de produit scalaire)
- intrastructures : calcul des coordonnées, à l'aide des opérateurs associés à chaque tableau, des variables ou des individus
- représentation graphique des plans principaux de l'analyse commune au compromis et aux intra-structures.

Remarque. Dans la représentation de l'interstructure, la prédominance de la 1ère valeur propre est souvent telle que nous avons préféré définir l'axe 2 comme axe horizontal et l'axe 1 comme axe vertical, et programmer une translation éventuelle de l'axe 2 : dans ce cas, le point d'intersection des axes ne correspond pas à l'origine du repère. Cela peut également se produire dans la représentation du compromis si l'origine des axes se trouve à l'extérieur du rectangle défini par les termes extrêmes de chaque coordonnée.

2 - ANALYSE DE LA METHODE.

Le logiciel proposé est lui-même constitué d'un ensemble de programmes s'enchaînant automatiquement les uns les autres et lisant, créant ou modifiant des fichiers intermédiaires.

L'organigramme précis de tous ces enchaînements est donné au paragraphe 2.3, la description des fichiers intermédiaires au paragraphe 2.2. Nous commençons par les fichiers de programmes et paramètres.

2.1 - Fichiers de programmes et de paramètres.

Par convention, les fichiers de programmes sont codés/STA et de paramètres/PAR. Ces derniers ont un but précis : les micro-ordinateurs, en général, disposent d'un langage BASIC qui ne possède pas d'instruction de sous-programmes avec passage de paramètres comme les SUBROUTINES en FORTRAN. Pour exécuter plusieurs fois un même programme dans une même session, on définit les paramètres par l'intermédiaire d'un fichier.

PROSCA/STA	calcul des produits scalaires ou des cosinus entre opérateurs
DIAGONAL/STA	diagonalisation d'une matrice symétrique (algorithme de Jacobi)
DIAGONAL/PAR	fichier de paramètres du programme précédent : P taille de la matrice à diagonaliser P0 nombre d'axes factoriels retenus LP impression des résultats (OUI : 1, NON : 2) Z1\$ nom du fichier contenant la matrice à diagonaliser Z2\$ nom du fichier contenant les vecteurs propres Z3\$ nom du fichier contenant les valeurs propres Z4\$ nom du programme à enchaîner
INTERSTR/STA	calcul des coordonnées et cosinus carrés des opérateurs sur les axes principaux
GRAPH/STA	représentation graphique des plans principaux (interstructure et instructure)
GRAPH/PAR	fichier de paramètres du programme précédent :

	P	nombre de points à représenter
	PØ	nombre d'axes factoriels retenus
	Z1\$	nom du fichier contenant les coordonnées des points sur les axes
	Z2\$	nom du programme à enchaîner
	Z3\$	nom du fichier contenant les identificateurs des points
	J1,J2,...,0,0	numéros des axes des plans à représenter
COMPROMIS/STA		calcul du compromis. Création du fichier INTPASTR/PAR
INTRASTR/STA		gestion des fichiers INTRASTR/PAR et DIAGONAL/PAR (programme exécuté si OP = 5 et IT = 2)
INTRASTR/PAR		fichier de paramètres du programme précédent :
	P	taille de la matrice à diagonaliser
	PØ	nombre d'axes factoriels retenus
	IT	type des opérateurs
	ZI\$	nom du fichier contenant la matrice à diagonaliser
	Z\$(1),...,Z\$(I),...	fichier contenant les matrices suivantes.
COORDCOV/STA		calcul des coordonnées des variables dans les intrastuctures (opérateurs de covariance)
ESCOUF1/STA		calcul des coordonnées des individus (1ère méthode ; opérateurs de produit scalaire)
ESCOUF2/STA		coordonnées des individus (2ème méthode ; opérateurs de produit scalaire)
RESULTAT/STA		édition des résultats numériques

2.2 - Fichiers intermédiaires.

Fichiers intermédiaires créés ou modifiés (codés/DAT) :

PROSCA/STA	PROSCA/DAT	produits scalaires
	VAR/DAT	norme des opérateurs
DIAGONAL/STA	VECPRO/DAT	vecteurs propres de PROSCA/DAT
	VALPRO/DAT	valeurs propres id.
	VEC/DAT	vecteurs propres de COMPROMIS/DAT
	VAL/DAT	valeurs propres id.

INTERSTR/STA	VECPRO/DAT	coordonnées des opérateurs sur les axes principaux
COMPROMIS/STA	COMPROMIS/DAT	matrice du compromis
	VARTOT/DAT	normes des individus ou des variables
INTRASTR/STA	VECTOT/DAT	fichiers réunissant tous les vecteurs propres et toutes les valeurs propres
	VALTOT/DAT	
COORDCOV/STA		
ESCOUF1/STA	VECTOT1/DAT	coordonnées de tous les éléments
ESCOUF2/STA		

Fichiers détruits.

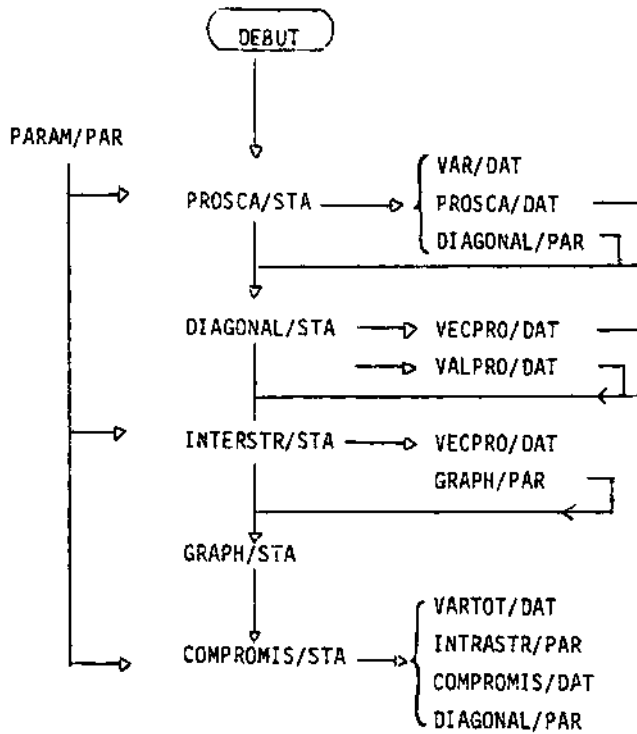
Certains programmes détruisent des fichiers pour augmenter la place disponible sur la disquette :

Programmes	Fichiers détruits
DIAGONAL/STA	PROSCA/DAT
INTEPSTR/STA	VAR/DAT
	VALPRO/DAT
COMPROMIS/STA	VECPRO/DAT

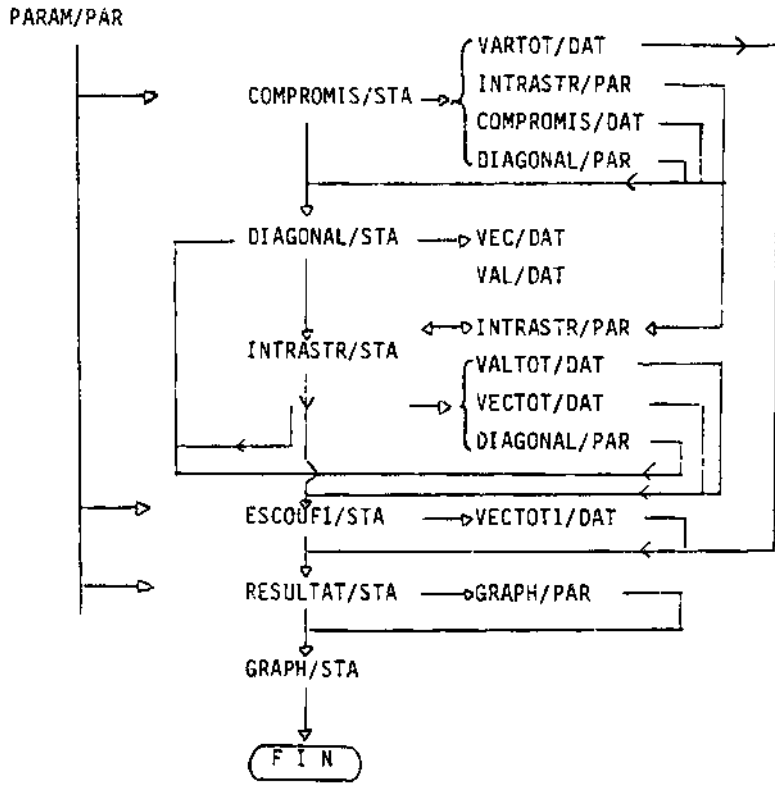
2.3 - Organigramme général.

On distingue 2 stades dans la méthode : le premier est commun aux différents types d'opérateurs : interstructure et compromis ; le deuxième concerne les infrastructures et change suivant l'option choisie dans l'analyse.

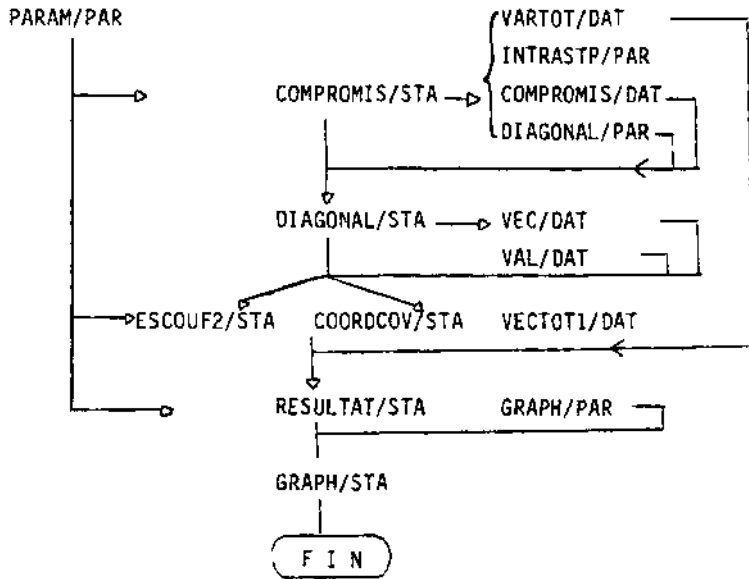
Partie commune.



Opérateurs de produit scalaire (1ère méthode)



Opérateurs de covariance et opérateurs de produit scalaire (2ème méthode).



BIBLIOGRAPHIE.

- [1] J. BOISGONTIER - "Le BASIC est ses fichiers. Méthodes pratiques".
Editions du PSI, Lagny/Marne, 1980.
- [2] T. FOUCART - "Analyse factorielle. Programmation sur micro-
ordinateurs",
Masson, Paris, 1982.
- [3] T. FOUCART - "STATIS : une nouvelle approche".
(à paraître dans la RSA).

CORRELATIONS SUIVANT ECHEC OU REUSSITE .79.

UTILISATEUR T.FOUCART DATE 3 DECEMBRE 1982

NOMBRE DE TABLEAUX	6
TAILLE DES TABLEAUX	8
TYPE DES OPERATEURS	1
COVARIANCES:1 ESCOUFIER:2	1
METRIQUE CHOISIE CANONIQUE:1 NORMEE:2	2
NOMBRE D'AXES (INTERSTRUCTURE)	4
NOMBRE D'AXES (COMPROMIS)	4
INTRASTRUCTURE: TOUT	
FICHER CONTENANT LES NOMS DES TABLEAUX	
FICHER CONTENANT LES IDENTIFICATEURS DES VARIABLES (OU DES INDIVIDUS)	CORREL/STR
RATIOS/STR	
AXES DES PLANS REPRESENTES (INTERSTRUCTURE)	
1 2	
AXES DES PLANS REPRESENTES (COMPROMIS ET INTRASTRUCTURE)	
1 2	

MATRICE DES PRODUITS SCALAIRES NORMES

1.000000							
0.920635	1.000000						
0.876092	0.832593	1.000000					
0.827361	0.876476	0.922650	1.000000				
0.947635	0.843751	0.881074	0.847291	1.000000			
0.890501	0.889823	0.912287	0.940481	0.855746	1.000000		

DIAGONALISATION

1E LIGNE:VALEURS PROPRES
2E LIGNE:CONTRIBUTION A L'INERTIE TOTALE

117.2300 5.2704

90.5% 4.1%
VECTEURS PROPRES (EN COLONNES)

0.4680	0.5415
0.4441	0.1394
0.3661	-0.3233
0.3597	-0.5413
0.3973	0.3806
0.4032	-0.3807

OPERATEURS

COORDONNEES DES OPERATEURS (1E COLONNE)
RECONSTRUCTION DE LA NORME (2E COLONNE)

OPERATEURS *	AXE 1	AXE 2
COR11 *	5.0670 0.9332 *	1.2430 0.0562 *
COR12 *	4.8080 0.8977 *	0.3101 0.0040 *
COR21 *	3.9634 0.8894 *	-0.7422 0.0312 *
COR22 *	3.8940 0.8833 *	-1.2427 0.0900 *
COR31 *	4.3018 0.8922 *	0.8738 0.0368 *
COR32 *	4.3659 0.9193 *	-0.8740 0.0368 *

COMPROMIS

1.000000							
0.003893	1.000000						
0.369315	-0.389999	1.000000					
0.781495	-0.286489	0.353701	1.000000				
0.509732	-0.243007	0.346035	0.414280	1.000000			
-0.645204	0.063195	-0.278404	-0.670950	-0.232650	1.000000		
0.367915	-0.226606	0.656476	0.454360	0.517940	-0.232552	1.000000	
0.340773	-0.163669	0.695700	0.321116	0.490750	-0.203409	0.967100	1.000000

DIAGONALISATION

1E LIGNE: VALEURS PROPRES
 2E LIGNE: CONTRIBUTION A L'INERTIE TOTALE

3.9564 1.5338 0.9800 0.7017

49.5% 19.2% 12.3% 8.8%
 VECTEURS PROPRES (EN COLONNES)

0.3712	-0.4423	-0.1218	-0.1219
-0.1751	-0.2391	-0.0870	0.0575
0.3715	0.2951	0.1015	0.3882
0.3910	-0.3726	0.1760	0.0369
0.3402	0.0435	-0.0293	-0.2315
-0.2961	0.5105	-0.0606	-0.3425
0.4181	0.3431	-0.2416	0.2593
0.4029	0.3755	-0.3067	0.1323

RESULTATS NUMERIQUES DE L'ANALYSE DEMANDEE

COMPROMIS

POUR CHAQUE AXE LA 1ERE COLONNE DONNE LA COORDONNEE
 LA 2E LE COSINUS CARRE

	AXE 1	AXE 2	AXE 3	AXE 4
ATF *	0.73840	0.54524	-0.54780	0.30008
FIN *	-0.34823	0.12126	-0.29609	0.08767
IRL *	0.73902	0.54615	0.36551	0.13360
REN *	0.77781	0.60499	-0.46149	0.21297
SOL *	0.67661	0.45781	-0.05399	0.00290
RVA *	-0.58901	0.34694	0.63224	0.39972
RFR *	0.83166	0.69165	0.42498	0.19061
LOG *	0.80150	0.64240	0.46504	0.21626

COMPOSANTES PRINCIPALES DES TABLEAUX

COORDONNEES SUR LES COMPOSANTES PRINCIPALES DU COMPROMIS

	AXE 1	AXE 2	AXE 3	AXE 4
C11 *	2.48774	-0.21705	0.30060	0.06930
C21 *	-0.21061	0.78137	0.20999	0.30739
C31 *	0.14961	0.02309	0.79509	-0.26688
C41 *	0.02918	0.20790	-0.22583	0.77837
C12 *	2.29835	-0.39508	-0.45878	-0.32250
C22 *	-0.24599	0.74510	-0.04330	0.10495
C32 *	-0.22834	-0.03461	1.39059	0.33841
C42 *	-0.13582	0.07098	0.28635	0.41373
C13 *	1.56741	0.08482	-0.12645	-0.01638
C23 *	0.05281	1.87127	0.04044	0.13390
C33 *	-0.06393	0.03233	0.06447	-0.35110
C43 *	-0.00690	0.09056	-0.29709	1.04950
C14 *	1.54406	-0.01777	-0.20921	0.04643
C24 *	-0.01107	1.67478	0.01990	-0.47642
C34 *	-0.10412	0.01590	1.11559	0.31946
C44 *	0.01955	-0.32224	0.27031	1.00098
C15 *	1.93273	0.07858	0.48150	0.64687
C25 *	0.04893	0.94030	-0.27846	0.26573
C35 *	0.23964	-0.22258	0.87283	-0.18413
C45 *	0.27242	0.17973	-0.15580	1.36913
C16 *	1.88524	0.31612	-0.01489	-0.38891
C26 *	0.19683	1.63089	0.23409	-0.43168
C36 *	-0.00741	0.18712	0.89218	0.15221
C46 *	-0.16378	-0.29197	0.12879	0.51170

VARIABLES OU INDIVIDUS DES TABLEAUX

COORDONNEES SUR LES COMPOSANTES PRINCIPALES DU COMPROMIS

	*	AXE 1	AXE 2	AXE 3	AXE 4	*
ATF1	*	0.96366	* -0.29698	* -0.02825	* -0.17455	*
FIN1	*	-0.57975	* -0.20350	* -0.79458	* 0.17234	*
TRL1	*	0.95335	* 0.29837	* -0.04957	* 0.32514	*
REN1	*	0.95016	* -0.17880	* 0.43798	* -0.09231	*
SOL1	*	0.80467	* -0.22952	* 0.18537	* -0.62295	*
RVA1	*	-0.75854	* 0.50528	* 0.04745	* -0.11780	*
RFR1	*	1.02519	* 0.24737	* 0.00409	* 0.23351	*
LQG1	*	0.93252	* 0.34786	* -0.07839	* 0.39448	*
ATF2	*	0.91810	* -0.41025	* -0.40964	* -0.19856	*
FIN2	*	-0.5721	* -0.04651	* -1.08967	* -0.23636	*
TRL2	*	0.65378	* 0.23740	* 0.18117	* 0.18419	*
REN2	*	0.91006	* -0.49853	* 0.00940	* -0.14877	*
SOL2	*	0.92949	* -0.13082	* -0.34899	* -0.44690	*
RVA2	*	-0.85700	* 0.50447	* -0.16278	* 0.00010	*
RFR2	*	0.94530	* 0.03139	* -0.58838	* -0.18244	*
LQG2	*	0.90922	* 0.09077	* -0.62526	* -0.17892	*
ATF3	*	0.63576	* -0.69048	* -0.18714	* -0.19238	*
FIN3	*	-0.20362	* -0.58305	* -0.78623	* 0.33108	*
TRL3	*	0.71129	* 0.45361	* -0.11900	* 0.36612	*
REN3	*	0.55697	* -0.75944	* 0.05468	* -0.11334	*
SOL3	*	0.55717	* -0.05178	* 0.19639	* -0.84466	*
RVA3	*	-0.37436	* 0.87759	* 0.03630	* -0.33462	*
RFR3	*	0.61295	* 0.80038	* -0.26504	* 0.21373	*
LQG3	*	0.63782	* 0.77021	* -0.29541	* 0.29812	*
ATF4	*	0.56314	* -0.82678	* -0.07407	* 0.21453	*
FIN4	*	-0.16928	* -0.43982	* -0.92250	* -0.09196	*
TRL4	*	0.55399	* 0.30439	* 0.39206	* 0.48269	*
REN4	*	0.62420	* -0.70399	* -0.11419	* 0.20269	*
SOL4	*	0.50997	* 0.39048	* -0.27076	* -0.83780	*
RVA4	*	-0.45765	* 0.81003	* -0.09177	* -0.53729	*
RFR4	*	0.67005	* 0.58797	* -0.43919	* -0.21329	*
LQG4	*	0.66089	* 0.58457	* -0.48559	* -0.23041	*
ATF5	*	0.71201	* -0.44055	* 0.25659	* -0.11978	*
FIN5	*	-0.55842	* -0.01721	* -0.85097	* 0.03538	*
TRL5	*	0.85401	* 0.31412	* 0.01855	* 0.62021	*
REN5	*	0.85908	* -0.24282	* 0.40009	* 0.35662	*
SOL5	*	0.37260	* -0.09123	* 0.20873	* -0.97453	*
RVA5	*	-0.55326	* 0.44615	* -0.30165	* -0.47563	*
RFR5	*	0.81994	* 0.43538	* -0.01344	* 0.57500	*
LQG5	*	0.75446	* 0.49254	* -0.20048	* 0.68962	*
ATF6	*	0.55526	* -0.72940	* -0.26251	* -0.09444	*
FIN6	*	-0.37360	* -0.56436	* -0.81276	* 0.09829	*
TRL6	*	0.66175	* 0.61884	* 0.20574	* 0.01218	*
REN6	*	0.68881	* -0.47746	* 0.19333	* 0.02901	*
SOL6	*	0.80725	* 0.52537	* -0.14848	* -0.52248	*
RVA6	*	-0.44447	* 0.72266	* 0.10731	* -0.34706	*
RFR6	*	0.83571	* 0.56860	* -0.15770	* -0.34013	*
LQG6	*	0.85085	* 0.60161	* -0.15770	* -0.33566	*

