

Capitolo V: un'analisi empirica

L'impostazione del lavoro empirico si basa sull'utilizzo della metodologia della statistica multivariata. In particolare viene utilizzata l'analisi fattoriale dinamica, la quale permette la misurazione di un fenomeno complesso attraverso l'impiego di un numero elevato di variabili e la sintesi dei risultati senza perdere la complessità.¹

Il risultato dell'analisi è quello di fornirci un quadro descrittivo attraverso il quale è possibile indagare sulle dinamiche dei paesi circa la loro partecipazione all'età dell'informazione.

Descrizione della metodologia statistica utilizzata

La metodologia statistica alla quale ci si riferisce è l'Analisi Fattoriale a più vie o multivariata, in particolare verrà adottata l'Analisi fattoriale dinamica, nell'ambito della analisi statistica "multiway". Ci si riferisce a matrici di dati a tre o più indici, con 1 o più modi, con 2 o più dimensioni.

Nella nostra analisi tratteremo un array di dati del tipo Unità x Variabili x Occasioni $X = (x_{ijk})$ $i=1, I$ (Unità) $j=1, J$ (Variabili) $k=1, K$ (Occasioni).

Nel caso di un'analisi fattoriale dinamica è possibile trattare le occasioni come un insieme caratterizzato da una struttura, in questo caso temporale.

L'oggetto dell'analisi diviene quindi un array del tipo stesse unità x stesse

¹"L'applicazione dell'analisi multivariata ai problemi di carattere economico non è ancora molto diffusa, mentre invece a nostro avviso può dare un contributo importante al superamento di alcuni dei problemi che si pongono nella determinazione delle variabili e dei fattori che indicano fenomeni dinamici come la crescita e lo sviluppo economico"
Palazzi, 1997

variabili x tempi, le variabili oggetto di analisi sono inoltre di natura quantitativa.

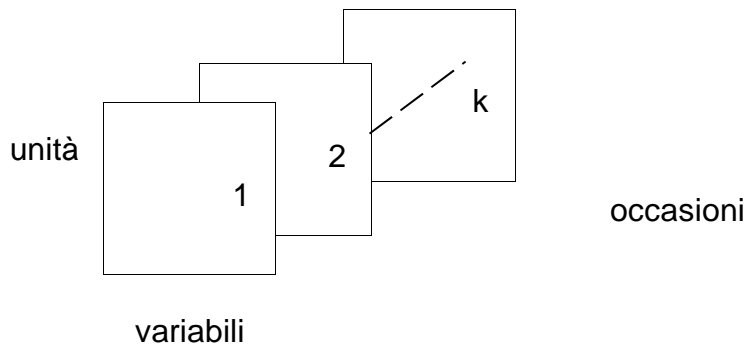


fig. 5.1 Array stesse unità, stesse variabili, occasioni.

L'array può essere visto in diverse maniere a seconda dell'aggregazione o meno degli indici i , j , t in modi "combinati". Qui utilizzeremo la combinazione $X (IT, J)$ la quale può essere studiata in due maniere $X (IT, J) = X_t$ con $t=1, T$ e $X(IT, J) = X_i$ con $i=1, I$, nel primo caso t gioca il ruolo di indice di gruppo, ed i quello di unità elementare (e viceversa).

In base alla forma $X (IT, J)$ si costruisce la metodologia dell'AFD diretta che si fonda essenzialmente su due aspetti: la decomposizione della variabilità totale di $X (IT, J)$ in tre componenti essenziali (struttura statica delle unità, dinamica media del sistema, dinamica differenziale delle singole unità) e la modellizzazione (in senso esplorativo) di tali componenti in modo tale da spiegare la variabilità misurata da ciascuna delle componenti dette. La modellizzazione in questione si basa sull'uso congiunto di due tecniche multivariate: l'Analisi delle Componenti Principali (ACP) ed il Modello di Regressione, diversamente combinate all'interno dei vari modelli.

Due sono gli aspetti preliminari da affrontare prima dell'analisi. Innanzitutto è necessario "omogeneizzare" i dati dell'array X . Infatti spesso le variabili quantitative si riferiscono a fenomeni diversi per dimensione ed unità di misura e la valutazione della variabilità congiunta di tali fenomeni può risultare distorta dall'eterogeneità dei dati grezzi. Al fine di eliminare questa fonte di disturbo è necessaria un'operazione di normalizzazione dei dati x_{ijt} , ma si deve anche evitare di incidere troppo sulle fonti di disturbo. Nella nostra analisi normalizzeremo il dato grezzo elementare x_{ijt} dividendolo per un opportuno valore medio², ovvero la media generale di ciascuna variabile $\underline{x}_{.j}$. (mi scuso ma qui come di seguito la sottolineatura va intesa come se fosse posta sopra la variabile). La normalizzazione proposta porta alla conseguente trasformazione preliminare $z_{ijk} = x_{ijt} / \underline{x}_{.j}$, conseguentemente l'oggetto di analisi diviene l'array $Z(I, J)$.

Un secondo aspetto da prendere in considerazione riguarda invece la formalizzazione matematica dei dati nell'AFD. In quanto la metodologia considerata ha una natura essenzialmente esplorativa, non probabilistica, gli strumenti matematici utilizzati sono di tipo geometrico-algebrico.

Lo spazio di riferimento più rilevante è lo spazio vettoriale R_{j+1} nel quale gli assi sono legati alle j variabili ed al tempo. In tale spazio ciascuna unità è rappresentabile come un vettore per ognuno dei tempi, se teniamo fisso t la matrice X_t è rappresentata dalla nuvola di punti-vettore $N(t)$, se fissiamo invece i è rappresentata dalla nuvola $N_T(i)$. quest'ultima descrive la traiettoria delle unità al variare del tempo.

In base alla metrica adottata ed alle distanze da questa introdotte è possibile quindi valutare la similarità/dissimilarità tra traiettorie, nello spazio "ambiente" R_{j+1} (i cui assi sono dati dalle j variabili più il tempo). L'applicazione di uno dei modelli ci permetterà di proiettare tali distanze

² Il che elimina simultaneamente il problema dell'eventuale differente unità di misura e della differente intensità dei caratteri legata alla loro dimensionalità.

su opportuni sottospazi e quindi di elaborare specifiche analisi delle traiettorie basate su tali proiezioni.

Riferendoci all'array $Z(IT, J)$ costruiamo S , la matrice di varianze e covarianze associate all'array. Tale matrice è decomponibile nella somma di tre matrici di varianze e covarianze, ciascuna delle quali misura una particolare forma di variabilità; avremo quindi $S = S^*_{I} + S^*_{T} + S_{IT}$, dove S^*_{I} è la matrice delle varianze e covarianze dei centri medi delle unità rispetto ai tempi, S^*_{T} è la matrice delle varianze e covarianze dei centri medi dei tempi rispetto alle unità e S_{IT} è la matrice di varianze e covarianze delle interazioni tra unità e tempi. Seguendo l'ordine possiamo denominare la variabilità delle tre matrici come 'struttura statica' delle unità, 'dinamica media' e 'dinamica differenziale'.

I modelli dell'AFD si propongono di 'spiegare' le diverse componenti della matrice S in termini di componenti principali e di regressione rispetto al tempo. Nella nostra analisi ci riferiremo al modello 1 dell'AFD il quale si fonda sulla decomposizione della variabilità totale in due componenti $S = S_{IT} + S^*_{T}$ in quanto $S_{IT} = S^*_{I} + S_{IT}$. Da tale aggregazione discende che S_{IT} è la matrice di dispersione media all'interno dei tempi mentre la matrice S^*_{T} rappresenta la variabilità tra i tempi.

Le variabili

Uno studio che tenti di dare un quadro generale dell'evoluzione dello sviluppo e del legame tra sviluppo e ICT comporta innanzitutto l'individuazione dei parametri e quindi delle variabili attraverso le quali identificare ed interpretare il ruolo e la posizione dei singoli paesi, o di gruppi di paese, e la loro dinamica temporale.

Nella scelta delle variabili si è cercato di tener conto della necessità di introdurre nell'analisi sia i paesi industrializzati (i cui dati sono

ampiamente disponibili) sia i PVS (che presentano invece carenze nella raccolta dei dati).

Le variabili considerate nell'analisi riguardano quindi differenti aspetti inerenti allo sviluppo ed alle nuove tecnologie. Le variabili "tasso di alfabetizzazione" e "tasso di iscrizione all'istruzione superiore in campo scientifico" sono state scelte per valutare il livello del capitale umano (focalizzandosi su un'ottica orientata al contesto delle nuove tecnologie); vengono poi considerate alcune variabili che sintetizzano il grado di diffusione dei mezzi di comunicazione tradizionali (diffusione dei quotidiani, della televisione, della radio, linee telefoniche ed utenti di cellulari), variabili che caratterizzano il livello di diffusione dell'ICT (pc, internet users, internet host) ed infine variabili economiche.

All'interno di quest'ultimo gruppo viene considerato il PIL pro capite, la percentuale di esportazioni di beni ad alta tecnologia come quota del PIL (ad indicare la presenza di un settore di produzione ICT competitivo), il valore aggiunto nel settore manifatturiero nei servizi come percentuale del PIL (come misura dello sviluppo di tali settori e della loro influenza sull'economia), il tasso di formazione di capitale e la commercializzazione di beni.

I valori sono stati ricavati dal World Development Indicator della Banca Mondiale per gli anni 1997, 1998, 1999, 2000, 2001; dall' Handbook of International trade and development dell'UNCTAD per il 1995 e 1996/7; dal Millennium Indicators (per il periodo 1988-2001) delle Nazioni Unite.

Per non ridurre eccessivamente il numero delle unità laddove sia risultato mancante il valore di una variabile per uno specifico anno, si è assegnato a tale mancanza assegnando alla variabile il valore relativo all'anno più vicino.

I Paesi

I paesi scelti per l'analisi sono 115; la scelta si è basata sulla disponibilità dei valori per le variabili adottate, cercando in ogni caso di fornire un quadro rappresentativo delle varie regioni del mondo. Sono stati considerati 22 paesi 'avanzati', 19 paesi dell'America latina, 15 paesi dell'area est europea, 34 paesi africani e 25 paesi rappresentativi di Asia e medio oriente. Nella tab 5.2 sono indicati tutti i paesi considerati nell'analisi.

1 Albania	35 Etiopia	69 Mongolia	103 Togo
2 Algeria	36 Finlandia	70 Marocco	104 Tunisia
3 Argentina	37 Francia	71 Mozambico	105 Turchia
4 Armenia	38 Germania	72 Namibia	106 Uganda
5 Australia	39 Ghana	73 Nepal	107 Ucraina
6 Austria	40 Grecia	74 Olanda	108 Inghilterra
7 Bangladesh	41 Guatemala	75 Nuova Zelanda	109 Usa
8 Belgio	42 Guinea	76 Nicaragua	110 Uruguay
9 Benin	43 Honduras	77 Niger	111 Venezuela
10 Bolivia	44 Hong Kong	78 Nigeria	112 Vietnam
11 Botswana	45 Ungheria	79 Norvegia	113 Yemen
12 Brasile	46 India	80 Pakistan	114 Zambia
13 Bulgaria	47 Indonesia	81 Panama	115 Zimbabwe
14 Burkina Faso	48 Iran	82 Paraguay	
15 Cambogia	49 Irlanda	83 Perù	
16 Cameroon	50 Israele	84 Filippine	
17 Canada	51 Italia	85 Polonia	
18 Rep. Cen. Afric.	52 Giamaica	86 Portogallo	
19 Chad	53 Giappone	87 Romania	
20 Cile	54 Giordania	88 Russia Fed.	
21 Cina	55 Kenya	89 Ruanda	
22 Colombia	56 Corea del Sud	90 Arabia Saudita	
23 Congo	57 Kuwait	91 Senegal	
24 Costa Rica	58 Laos	92 Singapore	
25 Costa D'Avorio	59 Lettonia	93 Rep. Slovacchia	
26 Croazia	60 Libano	94 Slovenia	
27 Cuba	61 Lituania	95 Sud Africa	
28 Rep. Ceca	62 Madagascar	96 Spagna	
29 Danimarca	63 Malawi	97 Sri Lanka	
30 Ecuador	64 Malesia	98 Sudan	
31 Egitto	65 Mali	99 Svezia	

tab 5.2 Paesi sui quali si è effettuata l'AFD

Risultati dell'analisi

Come supporto all'analisi è stato utilizzato il programma di calcolo XLISP-STAT³.

Considerato il modello 1 con riferimento all'analisi delle componenti principali (ACP) relativa alla matrice S_t abbiamo ottenuto i seguenti autovalori: 7.42 1.61 1.03 .61 .58 .46 .39 .30 .27 .22 .14 .08 .05 .04 .02 .

Sono stati allora presi in considerazione i primi tre autovettori della matrice S_t la cui quota di variabilità spiegata è pari al 72,42%.

La tav 5.3 sintetizza la qualità di rappresentazione del modello adottato. La serie di Indicatori $\{I(t)\}$ con $t=1,T$, misura il grado di spiegazione fornito, ad ogni tempo, dalla struttura fattoriale in questione rispetto alla variabilità complessiva osservata in ogni tempo. L'andamento di tali indici consente di evidenziare la qualità globale di rappresentazione delle diverse occasioni, in tal caso l'anno meglio rappresentato nel complesso appare il 2000.

Analisi di S_t

$I_t=0.724234$

$I(97)=0.708509$

$I(98)=0.712748$

$I(99)=0.717042$

$I(00)=0.747711$

* $I_i=0.757500$

$I_{it}=0.446899$

³ Programma di analisi fattoriale dinamica a cura di I. Corazzieri

Tab 5.3 Indici di qualità di rappresentazione dell'analisi tramite il Mod 1

L'indice It mostra la valutazione globale della capacità esplicativa del Modello 1 rispetto alla variabilità rappresentata dalla matrice di dispersione media all'interno dei tempi ($\underline{St} = S^*i + Sit$) che misura la variabilità all'interno dei tempi visti come "gruppi".

La qualità della rappresentazione fattoriale della struttura statica delle unità è invece data dall'indicatore I^* , mentre lit indica la qualità rispetto alla dinamica differenziale delle unità.

In base alla matrice di correlazione tra le componenti e le variabili (tab 5.4) è possibile costruire un grafico attraverso il quale interpretare la struttura fattoriale individuata. Nel grafico 5.5 sono stati quindi riportati i valori in riferimento al piano formato dalla prima e dalla seconda componente principale e dalla seconda e la terza.

Dalla posizione delle variabili su tali piani si nota come il I asse fattoriale sia caratterizzato dai valori delle variabili relative all'educazione terziaria in campo scientifico ($grter$), alla diffusione dei mezzi di comunicazione tradizionali ($quot$, tv , $radio$, $tecel$), alla diffusione dell'ICT (pc , $intus$, $intho$), dal PIL. Il II asse fattoriale è caratterizzato dai valori delle variabili relative a $va1$, $va2$ e grf .

MATRICE DI CORRELAZIONE TRA LE COMPONENTI E LE VARIABILI

vars	comp1	comp2	comp3
Edu	0.57	0.39	0.14
Grter	0.88	0.03	0.21
Quot	0.74	0.08	-0.07
Tv	0.89	0.08	0.22
Radio	0.89	-0.04	0.15
Tecel	0.95	0.02	0.01
Pc	0.95	-0.10	-0.07

Intus	0.90	-0.15	-0.06
Intho	0.74	-0.29	0.03
Exht	0.55	0.15	-0.52
Pil	0.91	-0.08	0.03
Va1	-0.11	0.75	0.21
Va2	0.04	0.66	0.46
Gcf	-0.00	0.71	-0.17
Trg	0.28	0.47	-0.70

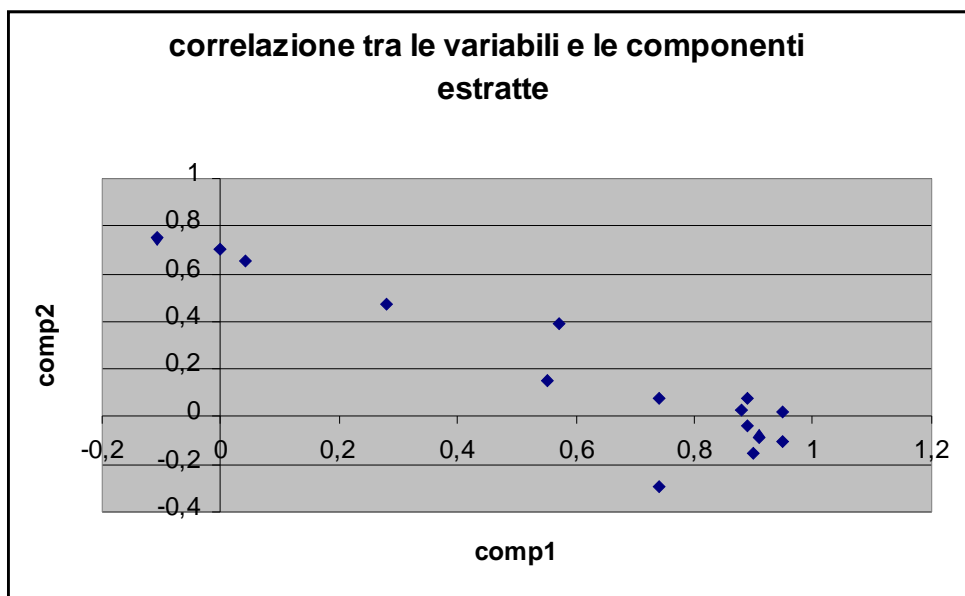
Tab. 5.4 Valori delle variabili sugli assi

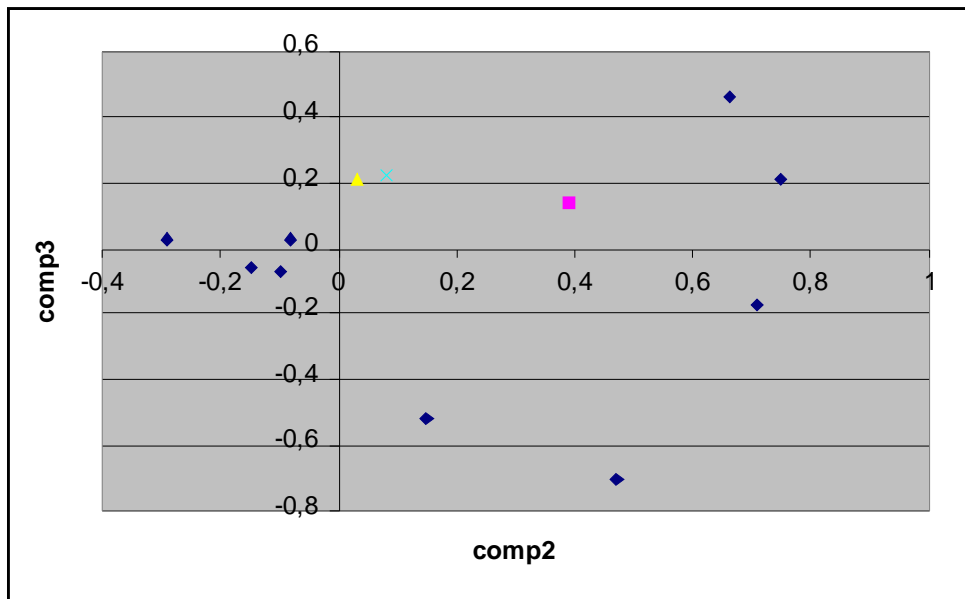
Sul piano fattoriale della I e II componente estratte è possibile allora distinguere tre situazioni: la prima caratterizzata da un alto livello di istruzione, da una buona diffusione dei vecchi mezzi di comunicazione, da un'ampia quota di esportazioni di beni ad alta tecnologia, e da un elevato scambio di beni manifatturieri (nel I quadrante); la seconda caratterizzata dal valore aggiunto nell'industria (nel II quadrante) ed infine da un livello di reddito pro capite particolarmente elevato e da un'ampia diffusione delle nuove tecnologie e delle telecomunicazioni nel IV quadrante.

Sul piano fattoriale relativo alla II e III componente principale invece il primo quadrante è caratterizzato dalle variabili relative al valore aggiunto nell'industria e nei servizi e dalle variabili relative ad istruzione e tv (asse3); il II quadrante dal numero degli host e dal PIL, il III quadrante dalle variabili relative alla diffusione dell'ICT (pc, intus); infine il IV quadrante è caratterizzato dalle variabili relative agli scambi, ovvero un elevata quota di esportazioni di beni ad alta tecnologia e un alto valore di scambio di beni come percentuale del PIL, mentre vicino all'asse della seconda componente principale si posizione la variabile relativa alla formazione del capitale.

La visualizzazione dei centri medi \bar{z}_{ij} sugli assi fattoriali mette in evidenza le caratteristiche dei paesi considerati. Riguardo al piano fattoriale

formato dalla I e II componente principale, tra i paesi che hanno caratteristiche più marcate rispetto al complesso degli altri, si evidenziano gli Stati Uniti che si proietta sui quadranti caratterizzati da un numero elevato di host e PIL elevato, Singapore tra quelli caratterizzati da un ampio settore di esportazione di beni tecnologici. Nello stesso quadrante degli Usa, ma con caratteristiche meno accentuate si presentano Nuova Zelanda, Svizzera, Canada ed Australia; in quello di Singapore la Corea del Sud, Hong Kong, Norvegia (paesi caratterizzati dalla presenza di industrie dedite alla produzione di ICT). I paesi europei appaiono fortemente caratterizzati dalla distribuzione dei mass media: Danimarca, Irlanda, Italia maggiormente da quelli tradizionali (I quadrante) mentre Norvegia, Finlandia, Inghilterra, Olanda e Austria dalle nuove tecnologie (IV quadrante). Infine i paesi nel II quadrante condividono con quelli del III la poca diffusione dei mezzi di comunicazione (tradizionali e nuovi), un minor livello di istruzione e di PIL, ma si differenziano in seguito ad un elevato livello di valore aggiunto nell'industria.





Tab 5.5 Proiezione delle variabili sugli assi delle componenti estratte

Sul III quadrante i valori più negativi sono assunti da paesi quali Niger, Etiopia, Rep. Centro Africana, Uganda, Benin, Sudan, i quali presentano un basso livello di reddito pro capite unito a scarsa diffusione dei mezzi di comunicazione in generale, basso livello di istruzione (soprattutto scientifica) e basso valore aggiunto nell'industria e nei servizi. Si tratta di paesi che presentano una carenza nel livello delle infrastrutture e di capitale umano che si rispecchia in una scarsa capacità di partecipare alla 'rivoluzione digitale' o di coglierne i benefici.

Sempre all'interno del IV quadrante, vicino all'asse delle ascisse, troviamo Zambia, Sud Africa, Zimbabwe, Colombia, Cuba, i quali presentano un grado di diffusione delle tecnologie ICT maggiore (si ricordi come il Sud Africa sia lo Stato con il maggior numero di Internet host nell'area africana). Sul II quadrante troviamo quei paesi che sono maggiormente caratterizzati dalla variabile valore aggiunto nell'industria. Questo settore potrebbe essere un profittevole canale di impiego delle ICT al fine di

ampliare i mercati e renderli più spessi nonché di stimolo per la produttività. Di contro il livello di istruzione, soprattutto scientifica, è ancora basso, quindi l'assenza di capitale umano 'adeguato' può rendere tali politiche di adozione meno efficaci.

Anche sul piano composto dalla II e III componente i paesi maggiormente caratterizzati sono gli Usa, che giacciono lungo l'asse delle x nel II quadrante, e Singapore, nel IV quadrante assieme a Tailandia, Filippine (mentre vicino agli asse delle x giacciono i paesi dell'est europeo).

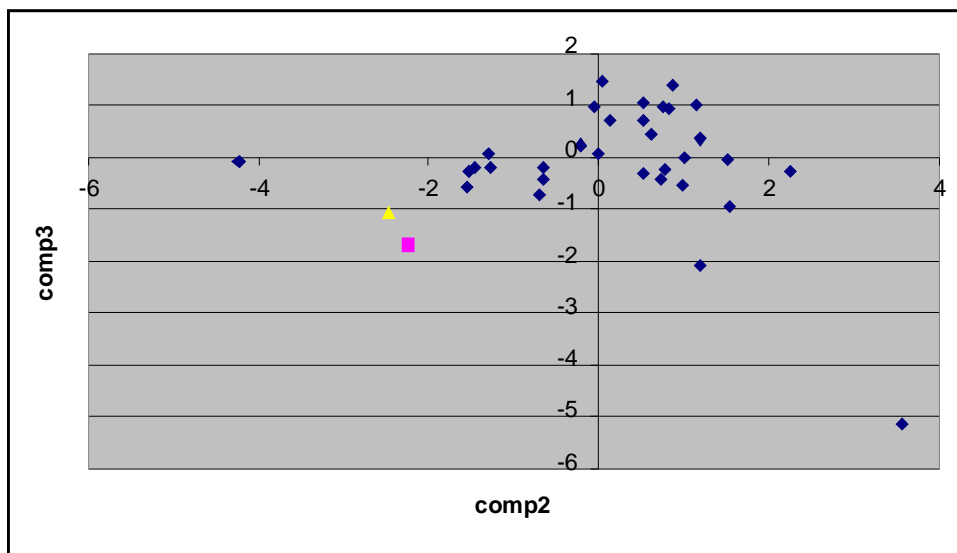
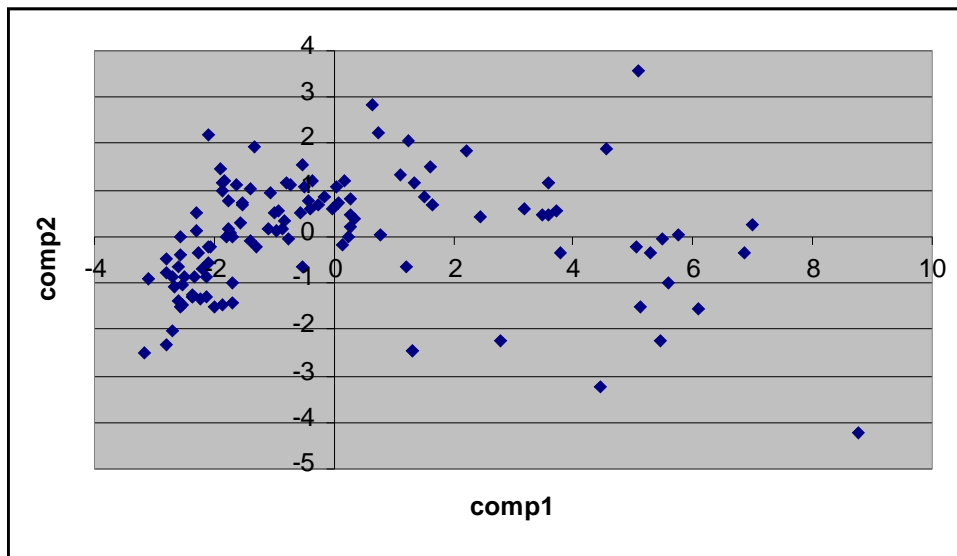
Nel terzo quadrante si posizionano anche il Kuwait ed Israele, assieme alla Svezia e la Svizzera che si caratterizzano però più vicino all'asse delle ascisse.

Passando quindi ad effettuare un'analisi della dinamica temporale delle singole unità rispetto ai tempi seguendo le traiettorie determinate mediante le componenti principali di ciascuna unità ai singoli tempi t.

Tra i paesi che presentano una dinamica temporale più intensa troviamo gli USA, che tendono ad accentuare la loro posizione nel IV quadrante.

Australia, Svizzera, Olanda, Giappone, Canada, Austria, in seguito a miglioramenti nella diffusione dell'ICT, dal I quadrante si dirigono verso il centro destro del IV dove sono posizionati gli Stati Uniti.

Il Mozambico, la Guinea, l'Iran, Egitto e l'India dal III quadrante si spostano verso il II, caratterizzato dal valore aggiunto nell'industria, quadrante all'interno del quale si accentua la dinamica dell'Algeria e del Sud Africa.



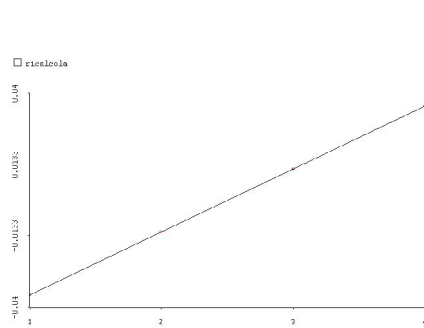
tab 5.6 Coordinate dei centri z_{ij} . sul piano fattoriale

Riguardo alla dinamica delle variabili, le variabili relative all'istruzione, alla diffusione delle passate tecnologie ed al PIL si presentano essenzialmente stabili. Questa scarsa dinamica è molto probabilmente dipendente dal

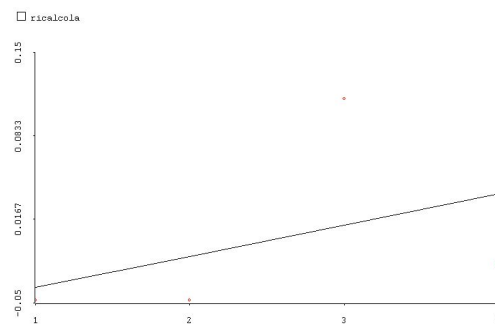
breve arco temporale dell'analisi e dal fatto che si tratta di variabili relative a dimensioni in qualche modo già consolidate nel passato.

Le variabili descritte in modo meno soddisfacente dal modello di regressione temporale lineare, in base a R^2 , sono quelle relative all'istruzione superiore scientifica, alla diffusione dei quotidiani, alle esportazioni di beni altamente tecnologici, al PIL e allo scambio dei beni.

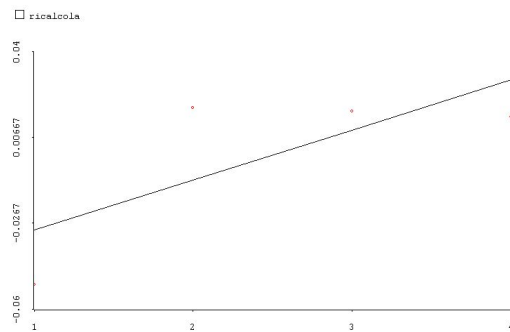
Valori di regressione dei centri medi \underline{z}_{jt} per le variabili considerate:



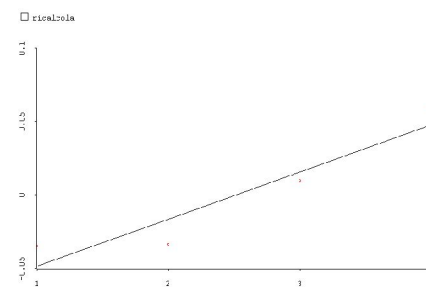
var edu



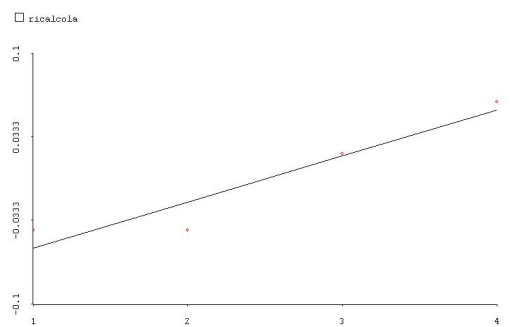
var grter



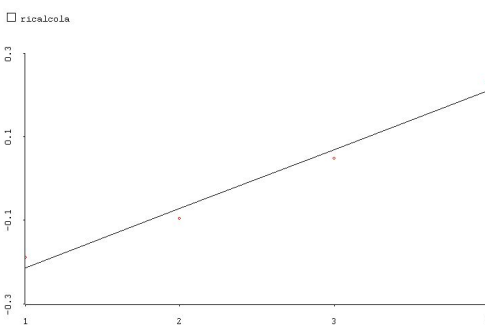
var quot



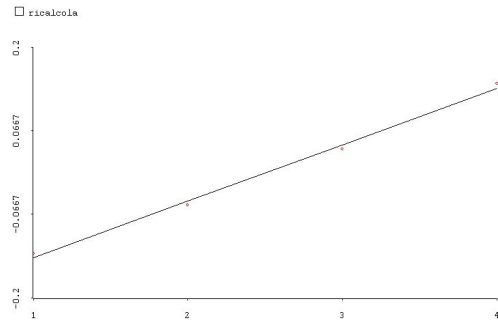
var tv



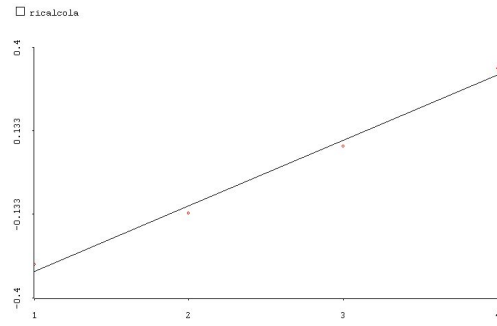
var radio



var tecel



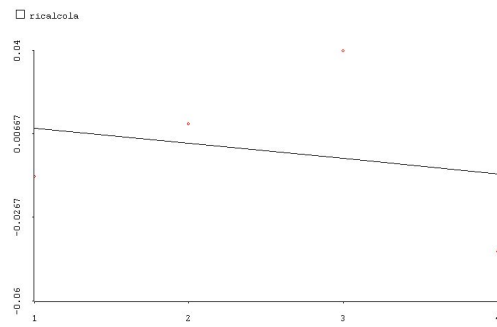
var pc



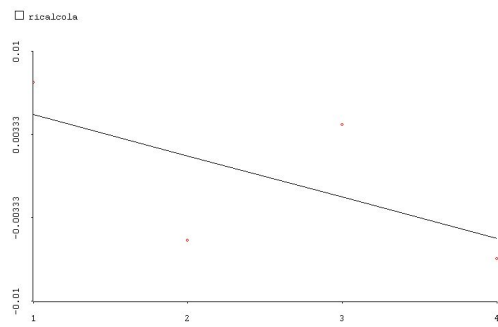
var intus



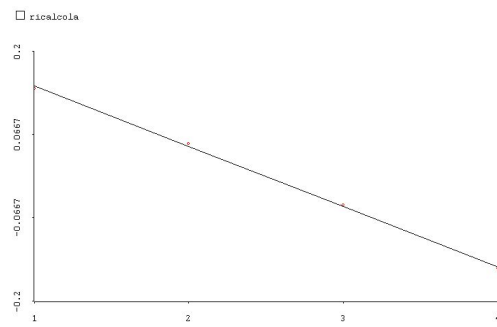
var intho



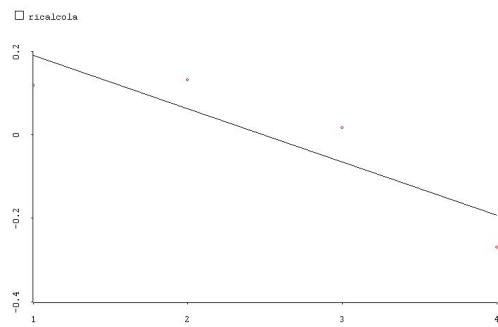
var exht



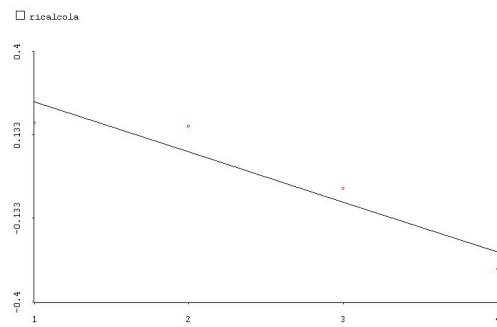
var PIL



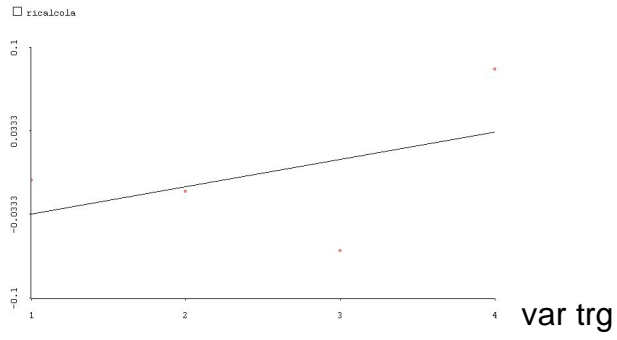
var va1



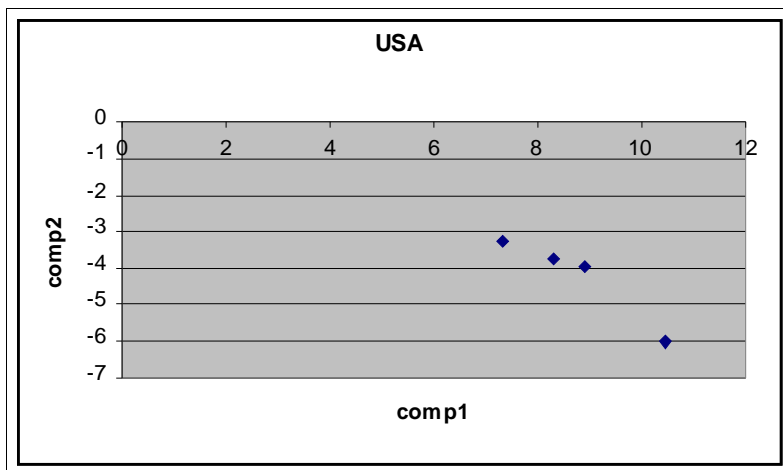
var va2



var gcf



Rappresentazione della dinamica di alcune unità sul primo piano



fattoriale.

