BIJLAGE II: DE TECHNIEK VAN DE HOOFDCOMPONENTENANALYSE: een beknopte uiteenzetting

Een principale componenten analyse is een bijzondere toepassingsvorm van de zgn. factoranalyse, welke laatste methode veelvuldig toepassing vindt in psychologisch onderzoek. De methode van de principale componenten beoogt de variatie van een aantal afzonderlijke variabelen zo goed mogelijk te beschrijven door een bij voorkeur geringer aantal geconstrueerde variabelen. Deze worden principale componenten genoemd. Hiertoe wordt verondersteld dat de variabele z_i geschreven kan worden als een lineaire combinatie van de principale componenten f_j, zodat voor $i = 1, \ldots\ldots, p$:

$$z_i = \sum_{j=1}^{q} \alpha_{ij} f_j$$

met p het aantal variabelen en q het aantal principale componenten. De beschouwde variabelen worden gemeten in afwijking van hun gemiddelde. Dit betekent dat het gemiddelde van de z_i steeds nul is. In sommige toepassingen, zoals in die uit dit artikel, worden de z_i bovendien gestandaardiseerd zodat ze een variantie van één bezitten. De α_{ij} en f_j zijn niet waarneembaar en moeten worden berekend; de α_{ij} heten factorladingen. In de rest van deze appendix wordt de werkwijze geschetst om de principale componenten en factorladingen te berekenen in zoverre dit noodzakelijk is voor een goed begrip van de uitkomsten van ons onderzoek. Voor een uitvoerige behandeling zij verwezen naar de desbetreffende literatuur*.

Zij z_{it} de gestandaardiseerde i^e^ koersindex op tijdstip t. Indien alle koersindices in een bepaalde vaste verhouding tot elkaar zouden staan dan geldt:

$$z_{it} = \alpha_{i1} f_{1t} \tag{1}$$

voor alle i en t, waarbij α_{i1} een rij nader te bepalen constanten is en f_1 de niet waarneembare eerste principale component is. In het algemeen geldt (1) slechts bij benadering. We zoeken daarom naar de α_{i1} en f_{1t} die een zo groot mogelijk deel van de residuele variantie S_1, met

$$S_1 = \sum_t \sum_i (z_{it} - \alpha_{i1} f_{1t})^2 \tag{2}$$

verklaart. Vanwege de onbepaaldheid van (2) voeren we de normalisatie $\sum_t f_{1t}^2 = 1$ in. Aangetoond kan worden dat (2) minimaal is bij

$$f_{1t} = \frac{1}{\lambda_1} \sum_i z_{it} \alpha_{i1} \tag{3}$$

* Zie b.v. Lawley en Maxwell (1971), of de betrekkelijk eenvoudige inleiding door Van de Geer (1967). Voor een kritische evaluatie van de factoranalyse en een vergelijking met de hoofdcomponentenmethode raadplege men b.v. Bethlehem *et. al.* (1977).

waarbij λ_1 de grootste eigenwaarde is van de (p×p) matrix (m_{ih}) met

$$m_{ih} = \sum_t z_{it} z_{ht}$$

met i, h = 1, 2, ..., p terwijl de α_{i1} worden berekend uit de elementen van de overeenkomstige eigenvector door vermenigvuldiging met $\sqrt{\lambda_1}$. Voorts kan worden afgeleid dat

$$\lambda_1 = \sum_i \alpha_{i1}^2$$

De uitdrukking (3) betekent dat f_1 een lineaire functie is van de waargenomen variabelen met coëfficiënten die evenredig zijn aan de elementen uit de eigenvector welke behoort bij de grootste eigenwaarde λ_1.

Bovenstaande werkwijze kan worden herhaald voor de tweede en volgende principale componenten. De k^e principale component wordt verkregen als

$$f_{kt} = \frac{1}{\lambda_k} \sum_i z_{it} \alpha_{ik}$$

met λ_k de naar aflopende grootte gerangschikte k^e eigenwaarde van de matrix (m_{ih}). De α_{ik} wordt berekend uit de corresponderende eigenvector. Verder geldt

$$\lambda_k = \sum_i \alpha_{ik}^2$$

Zoals gemakkelijk valt in te zien, is voor gestandaardiseerde variabelen de matrix (m_{ih}) gelijk aan de correlatiematrix. In dat geval kunnen de factorladingen α_{ik} worden geïnterpreteerd als enkelvoudige correlatiecoëfficiënten tussen de i^e variabele z_i en de k^e principale component f_k. Immers per definitie geldt

$$R(z_i, f_k) = \frac{\sum_t z_{it} f_{kt}}{\sqrt{\sum_t z_{it}^2 \sum_t f_{kt}^2}}$$

Wegens $\sum_t z_{it} f_{kt} = \alpha_{ik}$ * en de standaardisatie van z_i en f_k volgt onmiddellijk dat

$$R(z_i, f_k) = \alpha_{ik}$$

We merken op dat de som van de eigenwaarden λ_k gelijk is aan het spoor van de matrix (m_{ih}), gewoonlijk genoteerd als $sp(m_{ih})$. Zoals bekend is deze grootheid gelijk aan p indien, zoals in dit artikel het geval is, wordt uitgegaan van gestandaardiseerde variabelen. Elke factor f_k verklaart daarom de fractie ϕ_k van de totale variatie in z_i, met

$$\phi_k = \frac{\lambda_k}{sp(m_{ih})}$$

De grootheid ϕ_k is enigszins vergelijkbaar met de correlatiecoëfficiënt uit de regressie-analyse wanneer deze wordt gebruikt om de kwaliteit van de aanpassing te meten.

* Voor k = 1 volgt dat uit (2) door S_1 te minimeren met betrekking tot α_{i1}.

Tot slot een enkel woord over de varimax-rotatie. In sommige situaties is het verhelderend voor de interpretatie van de principale componenten, de kolommen van de in eerste aanleg berekende matrix met factorladingen te transformeren. De varimax-rotatie biedt hiervoor een veelvuldig gebruikte werkwijze. Het probleem is daarbij een orthogonale transformatiematrix te construeren, zodanig dat de functie

$$V = p \sum_j \left[\sum_i \beta_{ij}^4 - (\sum_i \beta_{ij}^2)^2 \right] \tag{4}$$

$$\text{met } \beta_{ij} = \frac{\alpha_{ij}}{\sqrt{\sum_j \alpha_{ij}^2}}$$

en $i = 1, \ldots, p$ en $j = 1, \ldots, q$ wordt gemaximeerd.

De gezochte factorladingen zijn functies van de rotatiehoek, die iteratief wordt berekend uit de telkens beschikbaar komende factorladingen β_{ij} tot criterium (4) maximaal is.

Geraadpleegde literatuur

Agmon, T., 1972, The relations among equity markets: a study of share price co-movements in the United States, United Kingdom, Germany and Japan, *Journal of Finance 27,* 839-855.

Bethlehem, J. G., Elffers, H., Gill, R. D. en Rijvordt, J., 1977, *Methoden, voetangels en klemmen in de factoranalyse* (gestencild rapport Mathematisch Centrum, Amsterdam).

Eijgenhuijsen, H. G., 1977, *Risico en rendement van aandelenportefeuilles* (Stenfert Kroese, Leiden).

Fase, M. M. G., 1973, A principal components analysis of market interest rates in the Netherlands, *European Economic Review 4,* 107-134.

Fase, M. M. G., 1976, The interdependence of short-term interest rates in the major financial centres of the world, *KYKLOS 29,* 63-96.

Fase, M. M. G., 1976, Een index voor het koersverloop van aandelen: een hoofdcomponenten-analyse, *Bank- en Effectenbedrijf 25,* 409-415.

Feeney, G. J. en Hester, D. D. 1967, Stock market indices: a principal component analysis in: D. D. Hester en J. Tobin (eds): *Risk aversion and portfolio choice* (Wiley, New York), 110-138.

Geer, J. P. van de, 1967, *Inleiding in de multivariate analyse* (Van Loghem Slaterus, Arnhem).

Generale Bankmaatschappij, 1973, *De beurs in België en portefeuillebeheer* (Brussel).

Groot, A. D. de, 1971, *Een minimale methodologie* (Mouton, Den Haag).

King, B. F., 1966, Market and industry factors in stock price behavior, *Journal of Business 39,* 139-190.

Kredietbank, 1978, De internationalisatie van de aandelenmarkten, *Weekberichten 33,* 1-6.

Lawley, D. N. en Maxwell, A. E. 1971, *Factor analysis as a statistical method* (Butterworths, London, 2nd edition).

Pais, A., 1967. Het Damrak en het buitenland, *Economisch Statistische Berichten 52,* 818-820.

Pais, A. en Van Dorsser, A. M., 1963, Tussen Wall Street en Damrak, *Economisch Statistische Berichten 48,* 967-972.

Ripley, D. M., 1973, Systematic elements in the linkage of national stock market indices, *Review of Economics and Statistics 55,* 356-361.

Solnik, B. H., 1973, Note on the validity of the random walk for European stock prices, *The Journal of Finance 5,* 1151-1159.

Wolf, H., 1978, *Koersfluctuaties op de Amsterdamse effectenbeurs en de groepsindeling van het CBS. Een factoranalytisch onderzoek* (gestencilde doctoraalscriptie Vrije Universiteit te Amsterdam, jan. 1978).