

2

Noções básicas de Análise de Componentes Principais e Análise de Fatores

Diversos estudos usaram a análise de componentes principais – ACP para a decomposição da Estrutura a Termo da Taxa de Juros. O emprego dessa técnica por diversos autores foi devido a característica de possibilitar a redução do número de variáveis a serem avaliadas e auxiliar no processo de interpretação dos vínculos existentes entre as variáveis, ou seja, ocorre uma redução na dimensão do problema.

De acordo com Richard Johnson e Dean Wichern, a análise de componentes principais (doravante denominada ACP):

“... is concerned with explaining the variance-covariance structure of a set of variables through a few linear combination of these variables. Its general objectives are (1) data reduction and (2) interpretation”. Ainda de acordo com os mesmos autores, a técnica se baseia no fato de que “... although the original p components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number k of principal components. If so, there is almost as much information in the k components as there is in the original p variables. The k principal components can then replace the initial p variables, and the original data set, consisting of n measurements on p variables, is reduced to a data set consisting of n measurements on k principal components”.

Trata-se, portanto, de uma representação mais parcimoniosa dos dados, que permite uma interpretação facilitada dos vínculos existentes entre as variáveis.

Para formalizar matematicamente as idéias contidas no parágrafo anterior, considere uma coleção de p variáveis aleatórias x_1, x_2, \dots, x_p com matriz de variância-covariância Σ e construa a partir delas um conjunto de p variáveis aleatórias z_1, z_2, \dots, z_p tal que:

$$z_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = a_1^T X$$

$$\begin{aligned} z_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p = a_2^r X \\ &\vdots \\ z_p &= a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p = a_p^r X \end{aligned}$$

onde $a_i^r = [a_{i1} \ a_{i2} \ \dots \ a_{ip}]$ e $X^r = [x_1 \ x_2 \ \dots \ x_p]$. Por definição, as p componentes principais das variáveis aleatórias x_1, x_2, \dots, x_p são as combinações lineares z_1, z_2, \dots, z_p que têm variância máxima e covariância cruzada zero.

O procedimento utilizado para obter as p componentes principais de x_1, x_2, \dots, x_p começa com o cálculo das variâncias e covariâncias cruzadas das combinações lineares z_1, z_2, \dots, z_p :

$$Var(z_i) = Var(a_i^r X) = a_i^r \Sigma a_i \quad ; \quad i = 1, 2, \dots, p \quad (1)$$

$$Cov(z_i, z_j) = Cov(a_i^r X, a_j^r X) = a_i^r \Sigma a_j \quad ; \quad i, j = 1, 2, \dots, p \quad (2)$$

Para calcular a primeira componente principal de x_1, x_2, \dots, x_p é necessário encontrar o vetor a_1 de norma unitária e dimensão $p \times 1$ que faz com que a expressão $a_1^r \Sigma a_1$ assumo o maior valor possível. A condição de 1ª ordem desse problema exige que o vetor a_1 satisfaça $\Sigma a_1 = \alpha_1 a_1$, onde α_1 é o multiplicador de Lagrange associado à restrição de normalização. Essa condição força a_1 a ser um autovetor da matriz Σ , com α_1 sendo o seu respectivo autovalor. Para a variância de z_1 ser, de fato, a maior possível, é necessário que α_1 seja o autovalor de Σ com a maior magnitude.

Para calcular a segunda componente principal do conjunto x_1, x_2, \dots, x_p é necessário encontrar o vetor a_2 de norma unitária, dimensão $p \times 1$ e perpendicular

a a_1 ⁶ que faz com que a expressão $a_2^T \Sigma a_2$ assumo o maior valor possível. A condição de 1ª ordem desse problema exige novamente que $\Sigma a_2 = \alpha_2 a_2$, ou seja, o par (a_2, α_2) deve corresponder a um autovetor/autovalor de Σ . Mais especificamente, α_2 é o autovalor com a segunda maior magnitude, pois z_2 é a componente principal com a segunda maior variância. A determinação de a_3, \dots, a_p segue um raciocínio análogo.

O procedimento acima descrito conduz aos resultados de que (i) as p componentes principais de um dado conjunto de p variáveis aleatórias correspondem às p combinações lineares das variáveis pertencentes ao conjunto original; e (ii) os componentes dos p autovetores da matriz de variância-covariância Σ definem os coeficientes de cada combinação. É possível demonstrar também os seguintes resultados⁷:

$$\bullet \sum_{i=1}^p \text{Var}(x_i) = \sum_{i=1}^p \alpha_i, \text{ ou seja, a variância total do sistema formado pelas}$$

variáveis aleatórias originais é igual à soma dos autovalores de Σ . Em consequência disso, a proporção da variância total que é explicada pela l -ésima

componente principal é dada por $\alpha_l / \sum_{i=1}^p \alpha_i$.

• O coeficiente de correlação entre a l -ésima componente principal z_l e a m -ésima variável aleatória x_m é dado por $e_{lm} \alpha_l^{1/2} / (\text{Var}(x_m))^{1/2}$, onde e_{lm} denota a m -ésima componente do l -ésimo autovetor de Σ . Dessa maneira, a magnitude de e_{lm}

⁶ A restrição de que as duas componentes principais z_1 e z_2 devem ser descorrelatadas exige que $a_1^T \Sigma a_2 = 0$ (ver [2]). Substituindo a condição $\Sigma a_2 = \alpha_2 a_2$ nesse resultado chega-se a $a_1^T \alpha_2 a_2 = \alpha_2 a_1^T a_2 = 0$; isso implica que os vetores a_1 e a_2 devem ser ortogonais.

⁷ Para maiores detalhes ver, por exemplo, Applied multivariate statistical analysis, de Johnson e Wichern, Englewood Cliffs: Prentice-Hall, 2002.

pode ser interpretada como uma medida da importância da m -ésima variável para a l -ésima componente principal.

A análise de componentes principais é um dos insumos mais importantes da análise de fatores. Novamente citando Richard Johnson e Dean Wichern, “... *the essential purpose of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors*”. O modelo matemático formal pode ser descrito da seguinte maneira: seja X um vetor aleatório de dimensão $p \times 1$ (ou seja, suas componentes são as variáveis aleatórias x_1, x_2, \dots, x_p) tal que $E[X] = \mu$ e $Var(X) = \Sigma$. O modelo de fatores postula que cada variável aleatória em X depende linearmente de m variáveis aleatórias não-observáveis F_1, F_2, \dots, F_m , chamadas fatores comuns, e p fontes adicionais de variação $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas erros ou fatores específicos:

$$\begin{aligned}x_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\x_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\&\vdots \\x_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p\end{aligned}$$

Em notação matricial, $X - \mu = LF + \varepsilon$, onde L , F e ε têm dimensões $p \times m$, $m \times 1$ e $p \times 1$, respectivamente. Para que o modelo seja uma representação parcimoniosa do comportamento de x_1, x_2, \dots, x_p é necessário ter $m < p$. O elemento (i, j) da matriz L corresponde ao *factor loading* l_{ij} , que mede a influência que o j -ésimo fator exerce sobre a i -ésima variável, enquanto que a i -ésima componente do vetor de fatores específicos ε (ε_i) mede o quanto o comportamento da variável aleatória x_i não pode ser explicado pelo conjunto de fatores comuns F_1, F_2, \dots, F_m . Assume-se também que:

$$\bullet E[F] = 0_{m \times 1}^8. \quad (3)$$

$$\bullet E[\varepsilon] = 0_{p \times 1}. \quad (4)$$

$$\bullet \text{Var}(F) = E[FF^tr] = I_{m \times m}. \quad (5)$$

$$\bullet \text{Var}(\varepsilon) = E[\varepsilon\varepsilon^tr] = \Psi_{p \times p} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}. \quad (6)$$

$$\bullet \varepsilon \text{ e } F \text{ são descorrelatados, de maneira que } E[\varepsilon F^tr] = 0_{p \times m}. \quad (7)$$

O *set-up* acima descrito implica que (i) $\text{Var}(X - \mu) = \Sigma = LL^tr + \Psi$ e (ii) $\text{Cov}(X - \mu, F) = L$. A prova de (i) parte da definição da matriz de variância-covariância de um vetor aleatório e usa as hipóteses das eqs.(5) a (7), conforme pode ser visto a seguir:

$$\begin{aligned} (X - \mu)(X - \mu)^tr &= (LF + \varepsilon)(LF + \varepsilon)^tr = (LF + \varepsilon)(F^tr L^tr + \varepsilon^tr) \\ &= LFF^tr L^tr + LF\varepsilon^tr + \varepsilon F^tr L^tr + \varepsilon\varepsilon^tr \Rightarrow E[(X - \mu)(X - \mu)^tr] = E[LFF^tr L^tr] + \\ &E[LF\varepsilon^tr] + E[\varepsilon F^tr L^tr] + E[\varepsilon\varepsilon^tr] = LE[FF^tr]L^tr + LE[F\varepsilon^tr] + E[\varepsilon F^tr]L^tr + \\ &E[\varepsilon\varepsilon^tr] = LL^tr + \Psi \end{aligned} \quad (8)$$

enquanto que a prova de (ii) parte da definição da covariância entre dois vetores aleatórios quaisquer e usa as mesmas hipóteses:

$$\begin{aligned} (X - \mu)F^tr &= (LF + \varepsilon)F^tr = LFF^tr + \varepsilon F^tr \Rightarrow E[(X - \mu)F^tr] \\ &= E[LFF^tr] + E[\varepsilon F^tr] = LE[FF^tr] + E[\varepsilon F^tr] = L \end{aligned} \quad (9)$$

⁸ Para denotar as dimensões de uma matriz M qualquer utilizamos a notação $M_{I,J}$, que indica que a referida matriz possui I linhas e J colunas.

Defina agora a variável $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$ (chamada comunalidade). De acordo com eq.(8), $Var(x_i) = h_i^2 + \psi_i$, ou seja, a parcela da variância da i -ésima variável aleatória x_i que é explicada pelos fatores comuns F_1, F_2, \dots, F_m está associada a h_i^2 . Repare que o modelo de fatores dá margem a uma representação mais econômica do conjunto x_1, x_2, \dots, x_p , pois permite que Σ (que possui $p(p+1)/2$ elementos) seja escrita como $\Sigma = LL^T + \Psi$, que depende de apenas $p(m+1)$ elementos (os mp factor loadings e os p elementos da matriz de variância-covariância de \mathcal{E}).

Em um problema prático o analista dispõe de uma amostra com n observações das variáveis aleatórias x_1, x_2, \dots, x_p e, a partir delas, deseja encontrar um modelo com $m < p$ fatores comuns que possa representar satisfatoriamente a variabilidade dos dados. A amostra permite que o analista calcule uma estimativa para a matriz de variância-covariância “verdadeira” Σ (denotada por S) e, a partir de S , decida se tal representação é razoável ou não. Nas palavras de Richard Johnson e Dean Wichern,

“... if the off-diagonal elements of S are small or those of the sample correlation are essentially zero, the variables are not related and a factor analysis will not prove useful. In these circumstances, the specific factors play the dominant role.”

Mas se S é tal que *“... it deviates from a diagonal matrix, then a factor model can be entertained and the initial problem is one of estimating the factor loadings in L and the specific variances in Ψ .”*

Há inúmeros métodos para estimar os *factor loadings* l_{ij} e as variâncias dos fatores específicos ϵ_i . O mais popular é o que emana da análise de componentes principais, que propõe o seguinte estimador para L :

$$\hat{L} = \begin{bmatrix} \sqrt{\hat{\alpha}_1} \hat{a}_1 & \sqrt{\hat{\alpha}_2} \hat{a}_2 & \dots & \sqrt{\hat{\alpha}_m} \hat{a}_m \end{bmatrix}_{p \times m} \quad (10)$$

onde supõe-se que m fatores comuns são suficientes para construir o modelo. Os números $\sqrt{\hat{\alpha}_1}$, $\sqrt{\hat{\alpha}_2}$, ..., $\sqrt{\hat{\alpha}_m}$ são os m primeiros autovalores de S , enquanto que os vetores \hat{a}_1 , \hat{a}_2 , ..., \hat{a}_m (de dimensão $p \times 1$) são os seus autovetores. As variâncias dos fatores específicos são os elementos da diagonal principal de $S - \hat{L}\hat{L}^r$.