

VIRTUE U. EKHOSUEHI (Benin City)

On graduation of enrolment size in a multi-echelon educational system

Abstract This paper focuses on an educational system wherein demotion and double promotion are not allowed. The total enrolment in such a system is modelled as a linear model within the context of factor analysis. The goal is to represent the total enrolment in terms of latent factors which generate the flows in the system. The notion of the matrix spectrum and the spectral radius are used to benchmark the specific variances and to approximate the factor loading vector, respectively. Two main theorems are propounded alongside with their proofs. A numerical illustration is given.

2010 Mathematics Subject Classification: Primary 62H25; Secondary 97B10..

Key words and phrases: educational system, enrolment, factor analysis, spectral radius, positive definite matrix.

1. Introduction The analysis of flows (such as promotion, admission, attrition, etc.) in an educational system holds a central position in modelling enrolment stock [2, 3, 6]. Early research was concerned with predicting enrolments using transition models based on the theory of Markov chains. As opposed to the Markov chain approach, this study concentrates on modelling enrolments using a single latent construct which generates the flows in the system. Flows take place between two consecutive sessions and are described in numerical terms. The term ‘educational system’ is used loosely in this paper to denote any institutional arrangement under a common management where individuals study, such as a secondary school, a department in a university or polytechnic, etc. The flows in the multi-echelon educational system (system hereinafter) may be either within or between the system and the outside world. Flows such as repetition, promotion, demotion are internal to the system. Repetition is counted as a part of flows because it occurs between sessions. Demotion is often not a common practice in school administration. For this reason, it is excluded as a flow in the remaining part of this paper. There is a two-way flow between the system and the outside world: admission and attrition. Attrition herein refers to all losses in the system such as graduation and dropouts for whatever reason.

Let $S = \{i : 1 \leq i \leq k, i \text{ is an integer}\}$ denote the levels in the system with $i = 1$ being the lowest level and $i = k$ the final level after which a student graduates. Since the system interacts with the outside world, the notation 0 is introduced to represent the outside world. Transitions therefore occurs within the state space $\{0\} \cup S = \Omega$. Let $n_{ij}(t-1)$ represent the number of students moving from level i in session $t-1$ to level j in session t for all $i, j \in \Omega, i = j \neq 0$. More precisely, $n_{ij}(t-1)$ is the number of students who were previously in level i and are still in the system in level j until the end of session t . In reality, $n_{ij}(t-1) = 0$ for $j \neq i, i+1$ as it is less common to have demotion and double promotion. The enrolment account for each level is expressed as

$$n_i(t) = n_{0i}(t) + n_{ii}(t-1) + n_{(i-1)i}(t-1) + n_{i0}(t), \quad i = 1, 2, \dots, k, \quad (1)$$

where $n_i(t)$ is the total enrolment in level i at session t . This study uses $j = 0$ instead of $j = k+1$ as in [1] to denote the index for attrition because students who leave the system are assumed to be in a state external to the system. Comparing equation (1) to that of the manpower system [1], $n_{i0}(t)$ may be thought of as admission in session t to replace leavers and $n_{0i}(t)$ as the admission in session t to achieve a desired expansion of the system. Relying on the flow variables, $n_{0i}(t)$, $n_{ii}(t-1)$, $n_{(i-1)i}(t-1)$ and $n_{i0}(t)$, this study aims to represent the enrolment size, $E_t = \sum_{i=1}^k n_i(t)$, in terms of latent factors which generate the flows. Total enrolment (or enrolment size) plays a significant role in educational planning. For instance, one of the key factors for programme accreditation in the Nigerian university system is the staff to student ratio, which is computed using the total enrolment for three consecutive sessions. There is a need to know the pattern of the enrolment size because it serves as a guide for resource planning for the system.

This study adopts the conceptual framework of factor analysis [4, 8], wherein the flow variables are expressed as a linear combination of latent constructs with an accompanying error term to account for the part of a flow variable that is not common to the other variables. The notion of the matrix spectrum and spectral radius are employed as a means to benchmark the specific variances and to approximate the factor loading vector, respectively. The flow data considered herein is a point in \mathbb{R}^4 .

2. Theoretical framework Let $w_t = \sum_{i=1}^k n_{0i}(t)$, $x_t = \sum_{i=1}^k n_{ii}(t-1)$, $y_t = \sum_{i=1}^k n_{(i-1)i}(t-1)$ and $z_t = \sum_{i=1}^k n_{i0}(t)$ denote the total number of transitions expressed in the respective summands in a session t . Let $\mathbf{N} = [w, x, y, z]^T$ denote the observation vector for a session. Suppose that the realisations of the 4-dimensional random vector, \mathbf{N} , is observed for T sessions. Then a data matrix, say \mathbf{M} , is obtained as a cloud of T points in \mathbb{R}^4

as

$$\mathbf{M} = \begin{bmatrix} \mathbf{N}'_1 \\ \mathbf{N}'_2 \\ \vdots \\ \mathbf{N}'_T \end{bmatrix} = \begin{bmatrix} w_1 & x_1 & y_1 & z_1 \\ w_2 & x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots & \vdots \\ w_T & x_T & y_T & z_T \end{bmatrix}.$$

The rows $\mathbf{N}'_t = [w_t, x_t, y_t, z_t] \in \mathbb{R}^4$ denote the t -th observation of the 4-dimensional random variable $\mathbf{N} \in \mathbb{R}^4$. The mean of the flows is given as

$$\bar{\mathbf{N}} = \frac{1}{T} \mathbf{M}' \mathbf{1}, \tag{2}$$

where $\mathbf{1} = [1, 1, 1, 1]'$, and the covariance of the flows is expressed as

$$\text{cov}(\mathbf{N}) = \frac{1}{T} \mathbf{M}' \mathbf{M} - \bar{\mathbf{N}} \bar{\mathbf{N}}'. \tag{3}$$

The analysis in this section is centred on the covariance matrix, $\text{cov}(\mathbf{N})$. This is because the covariance matrix summarises the disparity among the variables. Without loss of generality, this study assumes that $\text{cov}(\mathbf{N})$ is positive definite and so the flows are not concentrated on their means.

2.1. Justification for an underlying single construct The justification for using a single construct (or latent factor) as a means for generating enrolment follows from Theorem 2.1.

THEOREM 2.1 *The total enrolment in a multi-echelon educational system, where demotion and double promotion are not allowed, is generated by a single, non-unique latent factor.*

PROOF The degrees of freedom, d , of factor analysing the flow variables is given by [4]:

$$d = \frac{1}{2} (p - q)^2 - \frac{1}{2} (p + q),$$

where p is the number of variables and q is the number of factors. With $p = 4$, $q = 1$ is the only number of factors for which $d \geq 0$. Since $d \geq 0$ the factor model is over determined. ■

Consequent upon Theorem 2.1, all flows are associated with one latent factor. The factor may be named as the 'flow status' factor.

2.2. Factor model of flows Let f denote the underlying latent factor. With this notation, the observation vector \mathbf{N} may have been generated by a model of the form

$$\mathbf{N} = \bar{\mathbf{N}} + \boldsymbol{\beta} f + \mathbf{e} \tag{4}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]'$ is the factor loading vector with its elements β_p serving as weights, and $\mathbf{e} = [e_1, e_2, e_3, e_4]'$ as the specific factor vector. It is

common to assume that $\mathbb{E}(f) = 0$, $\text{var}(f) = 1$, $\mathbb{E}(e_p) = 0$, $\text{var}(e_p) = \psi_p$ and $\text{cov}(e_p, f) = 0$. These assumptions are consistent with the factor model [9]. Since $\bar{\mathbf{N}}$ does not affect variances and covariances,

$$\text{cov}(\mathbf{N}) = \text{cov}(\boldsymbol{\beta}f + \mathbf{e}) = \boldsymbol{\beta} \text{var}(f)\boldsymbol{\beta}' + \text{cov}(\mathbf{e}) = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Psi},$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \psi_2, \psi_3, \psi_4)$. The matrices $\boldsymbol{\beta}\boldsymbol{\beta}'$ and $\boldsymbol{\Psi}$ are unknown. The approximation to the matrix $\boldsymbol{\beta}\boldsymbol{\beta}'$ is provided by the following theorem.

THEOREM 2.2 *Given the decomposition of the covariance matrix of total flows*

$$\text{cov}(\mathbf{N}) = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Psi}. \quad (5)$$

Under the assumption that $\boldsymbol{\beta}\boldsymbol{\beta}'$ is positive definite, the factor loading vector $\boldsymbol{\beta}$ is approximated by

$$\tilde{\boldsymbol{\beta}} = \rho^{1/2}(\mathbf{A}^0)\mathbf{c}^*, \quad (6)$$

where \mathbf{A}^0 is a reduced covariance matrix of the form $\mathbf{A}^0 = \text{cov}(\mathbf{N}) - \boldsymbol{\Psi}^0$ with $\boldsymbol{\Psi}^0 = \text{diag}(\psi_1^0, \psi_2^0, \psi_3^0, \psi_4^0)$ being a preliminary guess on the specific variance such that $0 \leq \psi_p^0 < \gamma_p$ and γ_p belongs to the spectrum of $\text{cov}(\mathbf{N})$, $\rho(\mathbf{A}^0)$ is the spectral radius of \mathbf{A}^0 and \mathbf{c}^ is the normalised eigenvector corresponding to the spectral radius $\rho(\mathbf{A}^0)$.*

PROOF Following Nwosu *et al.* [7],

$$[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi} [\text{cov}(\mathbf{N})]^{-1} = [\text{cov}(\mathbf{N})]^{-1} \boldsymbol{\Psi} [\boldsymbol{\beta}\boldsymbol{\beta}']^{-1}. \quad (7)$$

The commutativity of matrices $[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi}$ and $[\text{cov}(\mathbf{N})]^{-1} \boldsymbol{\Psi}$ as expressed in equation (7) implies that they have the same eigenvectors [10]. Let λ_p belong to the spectrum of $[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi}$ and $\boldsymbol{\Phi} = \text{diag}(\phi_1, \phi_2, \phi_3, \phi_4)$ with each eigenvector ϕ_p corresponding to λ_p . Then

$$[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi} \boldsymbol{\Phi} = \lambda_p \boldsymbol{\Phi} \text{ and } [\text{cov}(\mathbf{N})]^{-1} \boldsymbol{\Psi} \boldsymbol{\Phi} = \frac{\psi_p}{\gamma_p} \boldsymbol{\Phi}.$$

From the decomposition of the covariance matrix, $\text{cov}(\mathbf{N})$, expressed in equation (5), it follows that

$$[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} = [\mathbf{I} + [\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi}] [\text{cov}(\mathbf{N})]^{-1}.$$

Thus

$$[\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi} \boldsymbol{\Phi} = [\mathbf{I} + [\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi}] [\text{cov}(\mathbf{N})]^{-1} \boldsymbol{\Psi} \boldsymbol{\Phi},$$

so that

$$\lambda_p \boldsymbol{\Phi} = [\mathbf{I} + [\boldsymbol{\beta}\boldsymbol{\beta}']^{-1} \boldsymbol{\Psi}] \frac{\psi_p}{\gamma_p} \boldsymbol{\Phi} = \left(\frac{\psi_p}{\gamma_p} + \frac{\psi_p \lambda_p}{\gamma_p} \right) \boldsymbol{\Phi}.$$

Hence

$$\lambda_p = \frac{\psi_p}{\gamma_p - \psi_p}, \quad p = 1, 2, 3, 4,$$

provided that $\psi_p \neq \gamma_p$. Since it is common that the preliminary guess on the specific variances should be as small as possible [9], $\psi_p^0 < \gamma_p$. As variance cannot be negative, $0 \leq \psi_p^0 < \gamma_p$. With $\Psi^0 = \text{diag} [\psi_1^0, \psi_2^0, \psi_3^0, \psi_4^0]$,

$$\beta\beta' = \text{cov}(\mathbf{N}) - \Psi^0.$$

Let $\mathbf{A} = \text{cov}(\mathbf{N}) - \Psi^0$ and $\sigma(\mathbf{A}^0)$ be the spectrum of \mathbf{A}^0 . The the spectral decomposition of \mathbf{A}^0 may be expressed as

$$\mathbf{A}^0 = \mathbf{C}\mathbf{V}^{1/2}\mathbf{V}^{1/2}\mathbf{C}' = (\mathbf{C}\mathbf{V}^{1/2}) (\mathbf{C}\mathbf{V}^{1/2})',$$

where $\mathbf{V}^{1/2} = \text{diag} (\sqrt{v_1}, \sqrt{v_2}, \sqrt{v_3}, \sqrt{v_4})$ with $v_p \in \sigma(\mathbf{A}^0)$ and \mathbf{C} is a matrix whose columns are normalised eigenvectors and each eigenvector corresponds to each eigenvalue in $\sigma(\mathbf{A}^0)$. Setting $\beta = \mathbf{C}\mathbf{V}^{1/2}$ is a contradiction as $\mathbf{C}\mathbf{V}^{1/2}$ is a square matrix of order 4. To circumvent this, the spectral radius of \mathbf{A}^0 , which is defined as $\rho(\mathbf{A}^0) = \max \{|v_p| : v_p \in \sigma(\mathbf{A}^0)\}$, and its corresponding normalised eigenvector \mathbf{c}^* are employed. Since \mathbf{A}^0 is symmetric and positive definite, its spectral radius is an eigenvalue which is a positive real number. In the context of principal component extraction, $\tilde{\beta}$ approximates β as $\tilde{\beta} = \rho^{1/2}(\mathbf{A}^0)\mathbf{c}^*$. ■

2.3. Enrolment size graduation The enrolment size is graduated using factor scores as the predictor. This study adopts the centred regression method [9] to estimate the sessional factor scores as

$$\hat{\mathbf{F}} = \mathbf{N}_c [\text{cov}(\mathbf{N})]^{-1} \tilde{\beta}, \tag{8}$$

where $\hat{\mathbf{F}} = [\hat{f}_1, \hat{f}_2, \dots, \hat{f}_T]'$ with \hat{f}_t being the estimated factor score for session t , $\mathbf{N}_c = [(\mathbf{N}_1 - \bar{N})', (\mathbf{N}_2 - \bar{N})', \dots, (\mathbf{N}_T - \bar{N})']'$, and the hat is used to denote an estimate. The enrolment size is regressed on the estimated factor scores using a linear model of the form

$$\mathbf{E} = \mathbf{H}\mathbf{\Delta} + \epsilon, \tag{9}$$

where \mathbf{E} is a $T \times 1$ vector of enrolment size at each session, \mathbf{H} is a $T \times s$ design matrix with the first column containing ones and the remaining columns constructed according to the regression function which is expressed as a polynomial of the estimated factor scores up to degree $s - 1$, $\mathbf{\Delta}$ is an $s \times 1$ vector of the regression coefficients and ϵ is a $T \times 1$ vector of the error term. The specification of the regression function is made after examining a scatter plot

of the enrolment size against the factor scores. The linear model (9) may be estimated using the method of least squares [5].

3. Numerical illustration Suppose the flows in a system has been observed for five sessions and arising thereof is a data matrix \mathbf{M} which has been organised according to the four flow variables — admission, repetition (probation), promotion and attrition — as follows:

$$\mathbf{M} = \begin{bmatrix} 119 & 21 & 70 & 18 \\ 66 & 27 & 164 & 19 \\ 49 & 8 & 189 & 65 \\ 40 & 34 & 138 & 75 \\ 128 & 68 & 91 & 54 \end{bmatrix}.$$

The actual enrolment size is $\mathbf{E} = [228, 276, 311, 287, 241]'$. This enrolment can be graduated as follows. Firstly, the covariance matrix, $\text{cov}(\mathbf{N})$, is computed as

$$\text{cov}(\mathbf{N}) = 10^3 \times \begin{bmatrix} 1.6453 & 0.5084 & -1.7095 & -0.5199 \\ 0.5084 & 0.5053 & -0.5783 & 0.0834 \\ -1.7095 & -0.5783 & 2.4553 & 0.4507 \\ -0.5199 & 0.0834 & 0.4507 & 0.6947 \end{bmatrix}.$$

The spectrum of this covariance matrix is

$$\sigma(\text{cov}(\mathbf{N})) = \{1.347 \times 10^2, 3.41 \times 10^2, 7.256 \times 10^2, 4.0992 \times 10^3\}.$$

Secondly, a preliminary specific variance matrix is assumed on the basis of the spectrum, $\sigma(\text{cov}(\mathbf{N}))$. For instance,

$$\Psi^0 = \text{diag}(13.47, 34.1, 72.56, 409.92).$$

This would lead to a reduced covariance matrix

$$\mathbf{A}^0 = 10^3 \times \begin{bmatrix} 1.6318 & 0.5084 & -1.7095 & -0.5199 \\ 0.5084 & 0.4712 & -0.5783 & 0.0834 \\ -1.7095 & -0.5783 & 2.3827 & 0.4507 \\ -0.5199 & 0.0834 & 0.4507 & 0.2848 \end{bmatrix}$$

with a spectral radius of $\rho(\mathbf{A}^0) = 4.0393 \times 10^3$. The proportion of explained variance is 76.2%. The eigenvector corresponding to this spectral radius is

$$\mathbf{c}^* = [-0.6099, -0.2040, 0.7467, 0.1696]'$$

Thirdly, the factor loading vector is approximated as

$$\tilde{\beta} = [-38.7643, -12.9642, 47.4597, 10.7761]'$$

The total sum of squared error is 1.6451×10^5 . This total sum of squared error is large because the spectrum of \mathbf{A}^0 is a collection of large eigenvalues. Nonetheless, there is no cause for alarm as it is bounded by the sum of squares of the deleted eigenvalues contained in the set $\{\sigma(\mathbf{A}^0) - \{\rho(\mathbf{A}^0)\}\}$, which is 3.1851×10^5 . Fourthly, the factor scores are estimated for each session as

$$\hat{\mathbf{F}} = [-1.0911, 0.5528, 1.0181, 0.5701, -1.0500]'$$

This factor score vector is used as an explanatory variable in the linear model by examining the nature of the observed enrolment size. The enrolment size in this example may be graduated using

$$\hat{\mathbf{E}}_t = 758.7769 - 892.5008\hat{f}_t - 379.6040\hat{f}_t^2 + 809.7748\hat{f}_t^3, \quad t = 1, 2, 3, 4, 5.$$

4. Concluding remarks This study emphasises the role of latent factors in the graduation of enrolment size. The method of graduation described in this paper has given a 'snapshot' into the use of factor analysis in fitting enrolment size. The main contribution of the paper is the factor analysis of flow variables with bounds on the preliminary assumption on the specific variances and its use for fitting the enrolment size. This methodological contribution is useful to researchers and educational planners in terms of modelling the trend of enrolment size. In practice, before factor analysis is conducted on the four flow variables, the appropriateness for its use is checked using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett's test [4]. The significance of these tests would suggest that factor extraction can be performed. The linear model may as well be subjected to statistical tests so as to ascertain its overall significance. From a methodology standpoint, the approach in this paper requires a methodological shift when the KMO measure and Bartlett's test are not significant. Thus the method described herein should be considered as a guide (at best) as to the possible trend in the enrolment size in a system.

REFERENCES

- [1] D. J. Bartholomew, A. F. Forbes and S. I. McClean, *Statistical Techniques for Manpower Planning* (2nd ed.), John Wiley & Sons, Chichester, 1991.
- [2] V. U. Ekhosuehi and A. A. Osagiede, *A proposed Markov model for predicting the structure of a multi-echelon educational system in Nigeria*, *Monografias Matematicas Garcia de Galdeano* 38 (2013), 65–74. [MR 3203093](#)
- [3] J. Gani, *Formulae for projecting enrolments and degrees awarded in universities*, *Journal of the Royal Statistical Society Series A*, 126 (1963), no. 3, 400–409. [doi: 10.2307/2982224](#)
- [4] W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis* (2nd ed.), Springer-Verlag, Berlin, 2007. [doi: 10.1007/978-3-540-72244-1](#)
- [5] B. W. Lindgren, *Statistical Theory* (4th ed.), Fourth edition. Chapman & Hall, New York, 1993. xii+633 pp. ISBN: 0-412-04181-2. [MR 1251284](#)

- [6] M. G. Nicholls, *The use of Markov models as an aid to the evaluation, planning and benchmarking of doctoral programs*, Journal of the Operational Research Society, 60 (2009), no. 9, 1183–1190. doi: [10.1057/palgrave.jors.2602639](https://doi.org/10.1057/palgrave.jors.2602639)
- [7] D. F. Nwosu, S. I. Onyeagu, J. I. Mbegbu and V. U. Ekhosuehi, *On refined principal component method for factor analysis*, Journal of the Nigerian Statistical Association, 28 (2016), 16–29.
- [8] T. Raykov and G. A. Marcoulides, *An Introduction to Applied Multivariate Analysis*, Routledge: Taylor & Francis Group, New York, 2008.
- [9] A. C. Rencher, *Methods of Multivariate Analysis* (2nd ed.), John Wiley & Sons Inc., New York, 2002.
- [10] Z. Woznicki, *Matrix splitting principles*, International Journal of Mathematics and Mathematical Sciences, 28 (2001), no. 5, 251–284. doi: <https://doi.org/10.1155/S0161171201007062> MR 1886844 Zbl 1005.65029

Wielkość rekrutacji w wieloetapowych systemach edukacji Virtue U. Ekhosuehi

Streszczenie Artykuł koncentruje się na systemie edukacyjnym, w którym nie jest możliwe powtarzanie i podwójna promocja są niedozwolone. Całkowita liczba uczestników w takim systemie może być badana w ramach modelu liniowego z zastosowaniem analizy czynnikowej. Celem jest przedstawienie całkowitej liczby rekrutowanych w zależności od ukrytych czynników, które generują przepływy w systemie. Pojęcie spektrum macierzy i promienia spektralnego są używane dla porównania wariancji i do przybliżenia odpowiednio wektora wag czynników. Sformułowano dwa twierdzenia wraz z dowodami. Podano numeryczną ilustrację.

2010 Klasyfikacja tematyczna AMS (2010): Primary 62H25; Secondary 97B10..

Słowa kluczowe: system edukacyjny • rekrutacja • analiza czynnikowa • promień spektralny • macierz dodatnia.



Virtue U. Ekhosuehi is a lecturer in the Department of Mathematics, University of Benin, Nigeria. He received his PhD in Industrial Mathematics in 2012 from the same university. He has published quite a number of scientific papers in areas related to Operations Research and Mathematical Economics. He also works on the mathematical aspects of manpower planning. He has been enlisted as a reviewer for Operations Research and Decisions, a journal published by Wrocław University of Technology, Poland. References to his research papers are found in MathSciNet under ID: 998379.

VIRTUE U. EKHOSUEHI
UNIVERSITY OF BENIN
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BENIN, BENIN CITY, NIGERIA
E-mail: virtue.ekhosuehi@uniben.edu

Communicated by: Antoni Leon Dawidowicz

(Received: February 22, 2018; revised: February 22, 2018)