PREDICTION AND SUFFICIENCY IN THE MODEL OF FACTOR ANALYSIS*

Ramses H Abul Naga
University of Lausanne

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6678

Discussion Paper
No.DARP/31
November 1997

---

* This paper was presented at the 1996 conference of the Society for Computation in Economics and Finance. I thank Robin Burgess, Frank Cowell, Ana Lamo, Elvezio Ronchetti, and Christ Schluter for some helpful suggestions. I am especially grateful to Gerard Antille for providing detailed comments on an earlier draft on the paper.

Abstract

We contrast two approaches to the prediction of latent variables in the model of factor analysis. The likelihood static is a sufficient statistic for the unobservables when sampling arises from the exponential family of distributions. Linear predictors on the other hand can be obtained as distribution-free statistics. We provide conditions under which a class of linear predictors is sufficient for the exponential family of distributions. We also examine various predicators in the light of the following criteria: (i) sufficiency, (ii) mean-square error, (iii) unbiasedness and illustrate our results with the help of Chinese date on living standards.

Keywords: Latent variables, factor analysis, sufficiency, prediction, exponential family of distributions, living standards analysis.

JEL Nos.: C1, C4, C8, I3.

Contact address: Dr Ramses H Abul Naga, University of Lausanne, DEEP, CH-1015 Lausanne, Switzerland. Email: AbulNaga@pop-serve.unil.ch

Introduction

The model of factor analysis [FA] is one of several techniques which seek to explain the correlation between a set of variables by a smaller set of random variables. In its standard form the model can be expressed as

$$X = \beta y + U \qquad (1)$$

where X is a p dimensional vector of observed variables known as indicators, and U is a vector of disturbances of the same dimension as X. y is a q x 1 vector of unobserved random variables where $q < p$, and β is p x q matrix of regression coefficients. As the size of the vector y is not critical to the results of the paper, in what follows we take $q=1$ so that y becomes a unique unobserved random variable.

The paper investigates conditions under which a class of distribution-free predictors is *sufficient* for the exponential family of distributions. We also contrast various approaches to prediction in the model of factor analysis according to the following three criteria: (i) sufficiency, (ii) mean-square error, and (iii) unbiasedness. Our first result identifies conditions under which the best linear predictor [BLP] is sufficient. As the best linear unbiased predictor [BLUP] is a scaled version of the BLP, in the model of factor analysis the same results can be used to address the relationship between unbiased predictors and sufficient statistics. In the case of multivariate normality it is a well known result that the best linear predictor of y coincides with the regression function $E(y|X)$. Then the BLP and BLUP are multiples of the sufficient statistic underlying the model. However, while the former two statistics exhibit the sufficiency property, the BLP is efficient (in the mean-square error sense) but biased, while the BLUP is (by definition) unbiased but not efficient. We thus conclude the paper by highlighting the existing trade-off between unbiasedness and mean-square error efficiency in the general problem of predicting the latent variables of the model of factor analysis.

The vector X of equation (1) can comprise cross-section, time-series, or panel data. Applications of the FA model in the behavioural sciences are many. Goldberger (1972), Everitt (1984), and Basilevsky (1994) provide good guides to this literature. Macroeconomics applications of factor analysis mainly fall in the domain of time-series analysis of production and financial aggregates (see for instance Geweke, 1977 and Geweke and Singleton 1981). In time-series factor analysis equation (1) is supplemented by a law of motion for y. X variables are often taken as repeated observations on an indicator of economic activity (say unemployment or GDP). The researcher then wishes to use these data in order to extract information on agents' expectations, the long run rate of interest etc. Garratt and Hall (1996) use the FA model in order to construct an index of *underlying economic activity*, while Mills and Crafts (1996) provide applications in the modelling of historical growth trends. In this literature a time series on X (typically GDP) is decomposed into a trend y and a cyclical component U. The jargon used in this area is borrowed from engineering sciences: predictors are commonly referred to as *filters*, and latent variables are known as *signals*.

Microeconomics applications of factor analysis include applications in the field of income distribution. In their study on the influence of classification and measurement error on the quantification of income inequality, Van Praag et al. (1983) take X to be a vector of reported income at different points in time, and y to be true (unobserved) income status. Likewise, Zimmerman (1992) constructs a longitudinal measurement error model in order to estimate the intergenerational correlation between the (unobserved) permanent incomes of a U.S. sample of fathers and sons. In his study on the identification of the poor, Abul Naga (1994) defines y as permanent income and X contains cross-section data on family income, consumption, and employment. Mercader (1995) also follows a similar approach in her study on living standards in Spain. Using a more complex specification of the FA model, Chamberlain and Griliches (1975) examine the influence of unobserved family background on the earnings of brothers.

Without loss of generality let y have zero mean and unit variance. Also let E(U)=[0] where [0] denotes a vector of zeros, cov(y,U)=[0] and var(U)=Ω, a diagonal matrix. An important feature of the FA model is the additive structure of the covariance matrix of the observables:

$$\Sigma = \mathrm{var}(X) = \beta\beta' + \Omega \qquad (2)$$

The variance of X thus consists of a component $\beta\beta'$ of rank one originating from the joint dependence of the indicators on y, together with a full rank diagonal matrix Ω arising from the presence of the disturbance terms. The model (1) can be estimated by standard procedures such as maximum likelihood or least squares, discussed in for example Basilevsky (1994) ch. 6, and Wegge (1996).

The problem we are dealing with in this paper is concerned with the question of predicting the unobservable y once we have observed X. We therefore treat the structural parameters of the model (the vector β and the matrix Ω) as given and instead we focus our attention on the various approaches to the prediction of y, their similarities, and their specific strengths and disadvantages.

Section 1 of the paper contains the background theory on sufficiency and the related results of Bartholomew (1984) within the context of factor analysis. Section 2 discusses conditions under which the best linear predictor of y can be expected to be a sufficient statistic when the joint conditional distribution of the indicators, $g(X|y)$, is a member of the exponential family of distributions. Section 3 of the paper reviews the results on prediction and sufficiency under the general normality assumption, whereas section 4 examines the relation between sufficiency and unbiased linear prediction. In section 5 we construct and compare various predictors using living standards data from rural China, which were the subject of an earlier study by Burgess and Murthi (1996). The final section of the paper contains concluding comments.

1 Sufficiency results

The aspect of factor analysis we are dealing with in this paper is the problem of locating observations on the space of the latent variable once we have collected related data in the form of a vector X of indicators. The task of predicting y is to construct a function $\tilde{y}=T(X)$ which satisfies some desirable statistical criteria. An on-going debate in the factor analysis literature has arisen from the failure of researchers in the area to agree on the set of properties $\tilde{y}$ should satisfy. Since y is a random variable, Bartholomew (1981,1984) has argued that the appropriate criterion to require on $\tilde{y}$ to possess is sufficiency, rather than unbiasedness which he argues is more relevant within the context of estimating fixed (i.e. non-random) parameters, such as the vector β of regression coefficients in (1). Bartlett's (1937) work places considerable emphasis on the choice of $\tilde{y}$ as an unbiased predictor.

By an unbiased predictor of a random variable y it is generally meant that $E(\tilde{y})=E(y)$. In the factor analysis literature however the unbiasedness condition takes a different meaning: a predictor $\tilde{y}$ of y is taken to be unbiased if for a given value y_o, $E[\ \tilde{y} \mid y= y_o\ ] = y_o$. We note that in general this requirement is not equivalent to the usual definition that $E(\tilde{y}) = E(y)$, and through out the paper we refer it as the Unbiasedness Restriction in Factor Analysis [URFA]. However, as the model of (1) is linear in y, we shall see in section 4 that the two definitions of unbiasedness can be regarded as equivalent in factor analysis.

By a sufficient statistic it is meant that the predictor exhausts all the sample information in X contained about the latent variable y:

Definition: A statistic T(X) is sufficient for y if and only if the conditional distribution of X given T(X) does not depend upon y.

That is, if one knows the value of a sufficient statistic, the sample can tell nothing more about y. The following theorem is often used to characterize sufficient statistics (Mood et al. (1974) p. 307:

Theorem 1 (Factorisation): Let g(X|y) denote the probability density function of X. A necessary and sufficient condition for a statistic T(X) to be sufficient for y is that there exists a factorisation

$$g(X \mid y) = \pi_1\ [\ T(x_1, \ldots, x_p)\ ;\ y]\cdot\ \pi_2(x_1, \ldots, x_p)$$

where $\pi_1\ [T(x_1, \ldots, x_p)\ ;\ y]$ is a non-negative function that depends on X only through T(X), and the second factor $\pi_2(x_1, \ldots, x_p)$ is also non-negative, and is independent of y.

Since X and y are random variables, we can decompose their joint density as follows:

$$f(X, y) = h(y|X) \cdot g(X) \qquad (3)$$

Thus all the information available to us about y is contained in the conditional distribution h(y|X). This result is due to Bartholomew (1984) where he shows that the following two conditions suffice to guarantee the existence of sufficient statistics for y within the context of the factor analysis model:

<u>Assumption 1</u>. The x's are independent when y is held fixed, i.e. ,

$$g(x_1, \ldots, x_p \mid y) = g(x_1|y) \cdot \ldots \cdot g(x_p|y) \qquad (4)$$

<u>Assumption 2</u>. The distribution of x_i when y is held fixed is a member of the exponential family:

$$g_i(x_i| y) = a_i(x_i)\, \phi_i(y)\, \exp\left[c_i(x_i)\, d(y) \right] \quad i = 1, \ldots, p \qquad (5)$$

<u>Theorem 2</u> (Bartholomew, 1984): Assumptions 1 and 2 above constitute a set of sufficient conditions for the existence of a sufficient statistic for y [(1)] .

Assumption 1 is known as the *axiom of conditional independence.* Stated differently, the axiom postulates that the observed correlation between the x's is solely induced by their joint dependence on y. As Bartholomew points out, assumption 2 is a fairly mild requirement since most distributions used in practice (normal, multinomial, Poisson, Beta etc.) are members of the exponential family. Note also that each distribution $g_i(x_i| y)$ could be chosen as a different member of the exponential family. To derive the sufficient statistic for y, substitute (5) into (4) to obtain

$$\Pi_i g_i(x_i|y) = \Pi_i a_i(x_i)\,\phi_i(y)\,\exp[\Sigma_i c_i(x_i)\,d(y)] \qquad (6)$$

The factorisation theorem can be used to decompose (6) as follows:

$$\pi_1[T(x_1,\ldots,x_p);y] = \Pi_i \phi_i(y)\,\exp[\Sigma_i c_i(x_i)\,d(y)]$$

and

$$\pi_2(x_1,\ldots,x_p) = \Pi_i a_i(x_i)$$

From theorem 1 it follows that the required predictor for y which exhibits the sufficiency property is given by

$$C(X) = \Sigma_i c_i(x_i) \qquad (7)$$

In the terminology of Bartholomew C(X) is referred to as the component function. The term "component " is used to illustrate the similarity between (7) and the related model of principal component analysis (cf. for eg. Abul Naga and Antille, 1990). In more general terms, the function C(X) constitutes the likelihood statistic for the exponential model (5).

2 Best linear prediction

An alternative route to the prediction problem is to follow a distribution-free approach. The advantage of pursuing such a line of thought resides in the fact that the statistical properties we postulate about the index of y will hold regardless of the exact distribution of X. As can be seen from (6) the sufficient statistic for y, $\Sigma_i c_i(x_i)$, is likely to differ according to the distributional assumptions we retain for each of the conditional distributions $g_i(x_i|y$.

The linear relation between X and y implied by the FA model (1) will in fact turn out to be the property which will enable us to derive a distribution-free predictor for y. Let y^* be a linear predictor of y. We can then write

$$y^* = b'X$$

The *best linear predictor* of y (Amemiya (1985), p. 3) chooses b in a way as to minimize the mean square error between y and a linear function b'X of the indicators. The problem can thus be written as follows

$$\min_{b} \; E[\,y - b'X\,]^2 \qquad (8)$$

The optimal choice of b is given by

$$b^o = [\,var(\,X\,)\,]^{-1} cov(X,y)$$

Noting from (2) that $var(X)=\Sigma$, and that $cov(X,y)=\beta$, we obtain the following expression for the best linear predictor [BLP] of y:

$$y^* = \beta'\Sigma^{-1}X \qquad (9)$$

Under what type of distributional assumptions would the BLP constitute a sufficient statistic for y? One first answer to the above question is that this scenario would occur when the likelihood statistic induced by a set of distributional assumptions is identical with (9). However it turns out that one can be more specific by making use of an important property of sufficient statistics, namely that any one-to-one transformation of a sufficient statistic is also sufficient (Mood et al. (1974), theorem 3, p. 307). Since the BLP is by definition linear, we can attempt to draw similarities in situations where the likelihood statistic is linear:

$$C(X) = \Sigma_i c_i \, x_i \qquad (10)$$

The proposition below provides sufficient conditions for the best linear predictor to exhibit the sufficiency property.

<u>Proposition 1</u>: Make the assumptions of theorem 2. Then for any joint conditional distribution $g(X|y)$ such that

(i) the likelihood statistic is linear

(ii) for each i $c_i = k\, b_i$

where k is an arbitrary constant, and c_i and b_i are respectively the coefficients on x_i in the likelihood statistic and the BLP, the best linear predictor is a sufficient statistic.

Proof: If the likelihood statistic is linear, we can write C(X) as in (10). The second requirement for the BLP to be sufficient is that for every i $c_i = k\, b_i$. Therefore

$$C(X) = k\, \Sigma_i b_i \, x_i \qquad (11)$$

and substituting (11) into (6) we get

$$\Pi_i \, g_i(x_i | y) = \Pi_i \, a_i(x_i) \, \phi_i(y) \, \exp[\Sigma_i b_i(x_i) \, \delta(y)]$$

where $\delta(y)= k.d(y)$. By the factorisation theorem it then follows that $y^* = \Sigma_i b_i \, x_i$ is a sufficient statistic for y $\square$.

The conditions stated in proposition 1 are not necessary since the class of one-to-one transformations need not be only linear. However as will become apparent when we come to discuss the multivariate normal case, as well as other cases of linear predictors, they turn out to be relevant in many practical cases. Proposition 2 below is more general in that it extends the conclusion for all one-to-one transformations of the BLP.

Proposition 2: Make the assumptions of theorem 2. Then for any joint conditional distribution $g(X|y)$ such that the likelihood statistic is a one-to-one mapping of the BLP, the best linear predictor is a sufficient statistic for y.

Proof. If the likelihood statistic is a 1-1 mapping of the BLP, we have

$$C(X) = \Phi(\Sigma_i b_i \, x_i) \qquad (12)$$

and substituting (12) into (6) we get

$$\Pi_i g_i(x_i | y) = \Pi_i a_i(x_i) \phi_i(y) \exp [\Phi (\Sigma_i b_i x_i) d (y)]$$

Using the factorisation theorem we can decompose $g(X|y)$ as a product of two functions, $g(X|y)= \pi_1 [\Sigma_i b_i(x_i) \; ; y]. \pi_2(x_1 ,. . . ,x_p)$ where

$$\pi_1 = \Pi_i \phi_i(y) . [\Delta(\Sigma_i b_i(x_i))]^{d (y)}$$

where $\Delta(.)= \exp[\Phi(.)]$, and

$$\pi_2(x_1 ,. . . ,x_p)= \Pi_i a_i(x_i)$$

As required, π_1 depends on X only through $\Sigma_i b_i x_i$. Once again it follows from the factorisation theorem that the best linear predictor is a sufficient statistic for y □.

3 Prediction and sufficiency under normality

In many practical applications of the FA model normality of the random variables is assumed. Though the structural parameters of the model can be estimated using least squares procedures, exact tests of hypotheses and maximum likelihood estimation will usually be carried out using the general normality assumption because of its tractability. In order to contrast Bartholomew's component function (the likelihood statistic) with the BLP, we construct the sufficient statistic for the latent variable model under normality (Bartholomew 1981, 1984). As this statistic turns out to be a constant multiple of the BLP, proposition 1 applies, and we may therefore readily conclude that the BLP will also be a sufficient statistic under the general normality assumption.

Under normality we have that $y \sim N(0,1)$ and $U \sim N(0, \Omega)$. From this it follows that

$$X \mid y \sim N\ [\beta y,\ \Omega] \qquad (13)$$

and

$$y \mid X \sim N\,[\,\beta' \Sigma^{-1} X\,;\ (\beta' \Omega^{-1} \beta + 1\,)^{-1}\,] \quad (14)$$

The distribution of X when y is held fixed can be written as

$$g\ (X|\ y) = (2\pi)^{-p/2}\ [\det\ (\Omega)\ ]^{-1/2} \exp\{-\frac{1}{2}(X'\Omega^{-1}X)\}$$

$$.\exp\{-\frac{1}{2}y'\beta'\Omega^{-1}\beta\ y\ \}\ .\exp\{(\ X'\ \Omega^{-1}\beta\ y\ )\}$$

Going back to (6), we can express g(X|y) as a member of the exponential family with:

$$a(X) = (2\pi)^{-p/2}\,[\det(\Omega)]^{-1/2}\exp\{-\frac{1}{2}(X'\Omega^{-1}X)\}$$

$$\phi(y) = \exp\{-\frac{1}{2}y'\,\beta'\Omega^{-1}\beta\,y\}$$

and

$$C(X) = \beta'\Omega^{-1}X \qquad (15)$$

Bartholomew's component function is this latter C(X) function. It is important to note that as in the case of the BLP, under normality C(X) is linear. Furthermore, from Basilevsky (1994) lemma 6.5, p. 375 we have that

$$[1+\beta'\Omega^{-1}\beta]\beta'\Sigma^{-1} = \beta'\Omega^{-1}$$

In other words $C(X) = k_1\, y^*$, where y^* is the best linear predictor of y, and

$$k_1 = [1+\beta'\Omega^{-1}\beta] \qquad (16)$$

It follows from proposition 1 that under normality the BLP y^* is a sufficient statistic for y. Another well known feature of the normal distribution is that the regression function, $E(y|X)=\beta'\Sigma^{-1}X$, is linear and coincides with the expression of the BLP. Therefore under normality y^* is also the *Best Predictor* of y (cf. Amemiya (1985), p. 3).

The above example is an illustration of a result well known to Bayesian statisticians, namely that knowledge of the statistic C(X) is equivalent to working with the conditional distribution h(y|X) [(2)].

4 Unbiased linear prediction and sufficiency

The purpose of this section is to extend our discussion in order to explore the common ground between unbiased linear prediction and sufficiency. The Rao-Blackwell and Lehmann-Scheffé theorems (Mood et al., pp. 321-26) constitute the bulk of the results in parametric statistics linking unbiasedness and sufficiency. Our approach here is somewhat different: we wish to inquire when would a distribution-free linear unbiased predictor exhibit the sufficiency property for the exponential class of distributions. As the best linear unbiased predictor [BLUP] of y happens to be a scaled version of the BLP, the same conditions of proposition 1 can be used to answer the above question.

Our starting point is to observe that the BLP is generally biased. This can be seen by noting that the inverse of Σ has the form

$$\Sigma^{-1} = \Omega^{-1} - \Omega^{-1}\beta\beta'\Omega^{-1} / (1+\beta'\Omega^{-1}\beta) \qquad (17)$$

and from substituting (1) into y^* for a given realisation y_o of y:

$$E(y^* | \; y = y_o) = E[\beta'\Sigma^{-1}(\beta y_o + U)] \neq 0$$

Bartlett (1937) suggested treating y as a fixed effect, specific to each observation, as opposed to its earlier treatment as a random variable. If through prior estimation of the FA model one possesses knowledge of β and Ω, one can use this information to write down the following least squares estimation problem for y:

$$\min_y [\; X - \beta y]'\Omega^{-1}[X - \beta y] \qquad (18)$$

From Generalized Least Squares theory, it follows that the Bartlett estimator

$$\hat{y} = [\beta' \Omega^{-1} \beta]^{-1} \beta' \Omega^{-1} X \qquad (19)$$

is the best linear unbiased estimator [BLUE] for y in the minimisation of the sum of squares of standardized residuals in problem (18). It is important to note that $\hat{y}$ is derived as an estimator rather than a predictor , since it treats y as an unknown parameter rather than a random variable. Note also that under the normality assumption the Bartlett estimator is a sufficient statistic when the axiom of conditional independence holds. Define the following constant:

$$k_2 = [\beta' \Omega^{-1} \beta]^{-1} \qquad (20)$$

We can thus write $\hat{y}$ as a multiple of the likelihood statistic (7) of the normality model:

$$\hat{y} = k_2 \ C(X) \qquad (21)$$

On the basis of proposition 1 the sufficiency of the Bartlett estimator follows from (21).

There is a conceptual problem however in treating y's in the sample as fixed parameters since as the sample size goes to infinity, the number of parameters in the model which require estimation increases at the same rate. Thus, Anderson and Rubin (1956) have shown that when β and Ω have to be estimated jointly with the y's, the maximum likelihood estimator is undefined.

As the Bartlett approach appears to be conceptually problematic, it is worth investigating whether one may be able to derive a best linear unbiased predictor of y under the assumption that the latent variable is random. The problem is a straightforward extension of the derivation of the best linear predictor considered in section 2. Lawley and Maxwell (1971) ch. 8, have studied this problem, where they have shown that the Bartlett statistic can be chosen as the minimum mean-square error predictor of y constrained to satisfy the unbiasedness condition [URFA].

The prediction problem may be stated as one of choosing a parameter vector α in a way as to minimize the mean square error

$$E[y-\alpha'X]^2$$

subject to the unbiasedness restriction [URFA]

$$E[\alpha'\beta y - y \mid y=y_o] = 0$$

The Lagrangean of the above problem can be written as:

$$L(\alpha, \lambda; \beta, \Omega) = E[y-\alpha'X]^2 + \lambda(1-\alpha'\beta) \qquad (22)$$

which yields for solution

$$y^*_u = [\beta'\Sigma^{-1}\beta]^{-1}\beta'\Sigma^{-1}X \qquad (23)$$

See Lawley and Maxwell (1971) pp. 109-111 for a derivation.

Thus, the effect of restricting the choice of predictors to the class of unbiased statistics in the URFA sense has the consequence of scaling the BLP of problem (8) by the constant $(\beta'\Sigma^{-1}\beta)^{-1}$. Whereas the statistic

$\beta'\Sigma^{-1}X$ is the minimum MSE linear predictor, $(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}X$ constitutes the minimum mean square error unbiased predictor of y under the unbiasedness restriction adopted in the factor analysis literature.

It is worth noting that the BLUP is the GLS estimator of y, viz the Bartlett Statistic (19). To see that this is the case, note from (17) that

$$\beta'\Sigma^{-1} = \beta'\Omega^{-1} - \beta'\Omega^{-1}\beta\beta'\Omega^{-1} / (1 + \beta'\Omega^{-1}\beta)$$

and that

$$\begin{aligned}\beta'\Sigma^{-1}\beta &= \left[1 - \beta'\Omega^{-1}\beta / (1 + \beta'\Omega^{-1}\beta)\right]\beta'\Omega^{-1}\beta \\ &= \beta'\Omega^{-1}\beta / (1 + \beta'\Omega^{-1}\beta)\end{aligned}$$

The argument generalizes to the case where y is a q-dimensional vector (q>1), through the inversion of an additive form $[BB' + \Omega]$, where B is a p x q matrix.

Note that the predictor y^*_u is also unbiased in the conventional sense that $E(y^*_u)=E(y)$. Since the FA model is linear, for any unbiased linear predictor $\hat{y}=b'X$ such that $E(\hat{y})= E(y)$ we have $E(\hat{y}) = E(b'\beta y)$. Therefore the unbiasedness condition in the conventional sense takes the simple form

$$b'\beta = 1 \qquad (24)$$

This condition is identical to the way unbiasedness is enforced in the [URFA] sense (see the Lagrangean (22)). Thus in the linear FA model the distinction between the two concepts of unbiasedness is not fundamental.

The condition (24) can in fact be used to generate unbiased predictors for y. For instance, for $b' = (\beta'\beta)^{-1}\beta'$, we can define the predictor

$$y^*_o = \sum_i \beta_i x_i \Big/ \sum_i \beta_i^2 \qquad (25)$$

which provides an example of an unbiased, but not sufficient predictor for y, since it cannot be expressed as a one-to-one mapping of the likelihood statistic (15). Only in the specific case where $\Omega = \kappa.\, I$ (where κ is a positive scalar), will y^*_o be sufficient and best linear unbiased. By analogy with least squares estimation theory we refer to (25) as the OLS predictor.

5 A numerical example

The preceding discussion can be illustrated by means of an empirical example which has been the subject of a more detailed study by Burgess and Murthi (1996). We consider a factor analysis model with three indicators:

x_1 : log per capita income in the household

x_2 : log per capita calorie intake in household

x_3 : log household size.

The data pertain to a sample of 5380 rural households in the Chinese province of Sichuan for the year 1990. The unobserved variable is interpreted as the family's long run income. The purpose of the example is to contrast predictors of the latent variable. The sample correlation matrix is the following:

$$S = \begin{bmatrix} 1 & & \\ 0.477 & 1 & \\ -0.194 & -0.266 & 1 \end{bmatrix}$$

Parameter estimates of the model obtained by the method of moments, are reported in table 1:

Table 1: Parameters estimates and standard errors.

parameter	β_1	β_2	β_3	ω_{11}	ω_{22}	ω_{33}
estimate	0.593	0.807	-0.327	0.649	0.350	0.893
s.e.	0.022	0.027	0.017	0.025	0.041	0.018

Table 2: Alternative predictors of the latent variable using Chinese data on living standards.

statistic	x1	x2	x3
C(X)	0.914	2.306	-0.366
BLP	0.259	0.656	-0.104
BLUP	0.363	0.915	-0.145
OLS	0.534	0.726	-0.294
$y^*(\theta)$ [a]	0.441	0.786	-0.125

Note: (a) Computed at the value θ=0.5.

The goodness of fit index defined as $1 - \det(\hat{\Omega}) / \det(S)$ takes the value 0.716. Replacing β and Ω by their parameter estimates, we calculate Bartholomew's component function using (14), the BLP using (9), the BLUP (the Bartlett statistic) using (19), and the OLS predictor using (25). The coefficients of the various predictors are given in table 2.

The coefficients on the BLP are the coefficients of the component function scaled down by a factor approximately equal to 3.5. Likewise, the coefficients on the BLUP are those of the component function scaled down by a factor approximately equal to 2.5. For the sake of ranking families in the y space the three approaches will produce identical conclusions. However, if one wishes to obtain *absolute* rankings, i.e. distances between observations, as opposed to just *ordinal* rankings, the data analyst will be confronted with the usual problem of trading off prediction mean square error against unbiasedness. If unbiased predictions are deemed to be the most important requirement, the choice of the BLUP is to be recommended. If this is not the case, one can opt for the minimum MSE linear predictor, viz the BLP.

The OLS predictor while being unbiased will in general not rank households identically with the component function, the BLP, or BLUP. As it is not a sufficient statistic, it coefficients in general cannot be expressed as multiples of those of the component function. Compared to the BLUP for instance, OLS assigns a higher weight on income but a smaller weight on calorie intake. The last line of table 2 will be discussed in the next section.

6 Concluding comments

The purpose of this paper was to contrast several approaches to the problem of prediction in the model of factor analysis. Throughout the discussion we have focussed on three properties of predictors, namely sufficiency, MSE optimality, and unbiasedness. None of the predictors considered here meets simultaneously all three requirements - see table 3. Unbiasedness is only met by the BLUP and OLS, the first of these only being sufficient. The minimum MSE predictor, the BLP, is sufficient, but on the other hand biased. The situation in the FA model is therefore less clear-cut than in other statistical models such as the Gauss-Markov regression model where all three properties can be met by a single statistic.

Table 3: Statistical properties of various predictors in the model of factor analysis.

predictor	unbiased	minimum MSE	sufficient
C(X)	no	no	yes
BLP	no	yes	yes
BLUP	yes	no	yes
OLS	yes	no	no

As the mean-square error of a predictor can be decomposed into the sum of its variance and its square bias, the MSE criterion can be put into good use in ranking biased and unbiased predictors. From the definitions of the BLP and BLUP (and also by noting that the Lagrange multiplier λ in (24) is in general not equal to zero), we note that the BLP will dominate any other linear predictor according to the MSE criterion.

It is possible however to explicitly take into account the existing trade-off between prediction bias and mean-square error. Consider the following class of linear predictors:

$$\tilde{y}(\theta) = \theta\, y^{*} + (1-\theta)\, y^{*}_{u} \qquad 0 \leq \theta \leq 1$$

Then at one end, when θ=1, we obtain the BLP, and at the other end, for θ=0 we have the BLUP. Intermediary values of θ in the range [0;1] allow us to trade off prediction bias against accuracy, with increasing values of θ reducing the

MSE and increasing the bias of the predictor. Furthermore, by writing $\tilde{y}(\theta)$ in the form

$$\tilde{y}(\theta) = [\,\theta\,(1+\beta'\Omega^{-1}\beta)^{-1} + (1-\theta)\,(\beta'\,\Omega^{-1}\beta\,)^{-1}\,]\,\beta'\Omega^{-1}X$$

we note that when the axiom of conditional independence holds such a statistic will be sufficient for the normal distribution.

When $\theta=1/2$, $\tilde{y}(\theta)$ is simply an average of the BLP and BLUP. This is the value computed in the last line of table 2 of our numerical example. There we have a predictor which is sufficient, and which possesses a smaller bias than the BLP, as well as a lower mean square error than the BLUP.

Footnotes

(1) The exact set of sufficient conditions given in Bartholomew (1984) is weaker: in assumption 2 it is only required that p - 1 of the x's be chosen as members of the exponential family.

(2) See for instance the discussion around property 4.2 in Gourieroux and Monfort(1995), pp. 103-4.

References

Abul Naga R. and G. Antille (1990): "Stability of Robust and Non-Robust Principal Component Analysis", *Computational Statistics and Data Analysis* 10, 169-174.

Abul Naga R. (1994), "Identifying the Poor: a Multiple Indicator Approach," Distributional Analysis Research Programme discussion paper no. 9, London School of Economics.

Amemiya T. (1985): *Advanced Econometrics*, New York, Blackwell.

Anderson T. W. and H. Rubin (1956): "Statistical Inference in Factor Analysis", *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability*, 111-150.

Bartholomew D. (1981): "Posterior Analysis of the Factor Model", *British Journal of Mathematical and Statistical Psychology* 34, 93-99.

Bartholomew D.(1984): "The Foundations of Factor Analysis", *Biometrika* 71, 221-32.

Bartlett M. S. (1937): "The Statistical Conception of Mental Factors", *British Journal of Psychology* 28, 97-104.

Basilevsky A. (1994): *Statistical Factor Analysis and Related Methods,* New York, John Wiley.

Burgess R. and M. Murthi (1996): "Calorie-Expenditure Linkages in Rural China", Development Economics Programme discussion paper, Sticerd, London School of Economics.

Chamberlain G. and Z. Griliches (1975): "Unobservables with a Variance Components Structure: Ability, Schooling, and the Economic Success of Brothers", *International Economic Review* 16, 422-50.

Everitt B. (1984): *An Introduction to Latent Variable Models,* London, Chapman and Hall.

Garratt A. and S. Hall (1996): "Measuring Underlying Economic Activity", *Journal of Applied Econometrics* 11, 135-151.

Geweke J. (1977): "The Dynamic Factor Analysis of Economic Time-Series Models" in D. Aigner and A. Goldberger (eds.) *Latent Variables in Socio-Economic Models*, New York, North Holland.

Geweke J. and K. Singleton (1981): "Maximum Likelihood 'Confirmatory' Factor Analysis of Economic Time-Series", *International Economic Review* 22, 37-54

Goldberger A. S. (1972): "Structural Equations Methods in the Social Sciences", *Econometrica* 40, 979-1001.

Gourieroux C. and A. Monfort (1995): *Statistics and Econometric Models,* Cambridge, Cambridge University Press.

Lawley D. N. and A. E. Maxwell (1971): *Factor Analysis as a Statistical Method,* London, Butterworths.

Mercader M. (1995): Identifying Low Standards of Living: Evidence from Spain", mimeo, Universitat Pompeu Fabra, Barcelona.

Mills T. and N. Crafts (1996): "Modelling Trends in Economic History", *The Statistician* 45, 153-59.

Mood A., F. Graybill and D. Boes (1974): *Introduction to the Theory of Statistics*, Auckland. McGraw-Hill.

Van Praag B., Hagenaars A., and W. Van Eck (1983): "The Influence of Classification and Observation Errors on the Measurement of Income Inequality," *Econometrica* 51, 1093-1108.

Wegge L. (1996): "Local Identifiability of the Factor Analysis and Measurement Error Model Parameter", *Journal of Econometrics* 70, 351-382.

Zimmerman D. (1992): "Regression Toward Mediocrity in Economic Stature", *American Economic Review* 82, 409-429.