

**A CRITIQUE OF MAXIMUM LIKELIHOOD ORDINATION  
USING PRINCIPAL COMPONENTS ANALYSIS**

**Henry S. Lynn**

New England Research Institute

9 Galen Street

Watertown, MA 02172, U.S.A.

**and**

**Charles E. McCulloch**

Biometrics Unit and Statistics Center

338 Warren Hall

Cornell University, Ithaca, NY 14853, U.S.A.

BU-1289-M

May, 1995

## SUMMARY

Principal components analysis is shown to be equivalent to maximum likelihood estimation for a linear model with independent Gaussian error terms. Theoretical properties are discussed and related to results on a linear functional measurement error model with equicorrelated errors. Conditions for consistency of the estimators are established and their finite sample performances are examined via simulations.

*Keywords:* principal components analysis; maximum likelihood ordination; functional measurement error models; incidental parameters; consistency.

### 1. Introduction

Consider an ecological investigation where a researcher samples  $n$  different sites and for each site records the abundances (e.g. biomass, percentage, frequency) of  $m$  different species of flora/fauna. The resulting samples-by-species data matrix can then be examined via ordination. On the one hand, this involves generating low-dimensional summaries of the data, and principal components analysis (PCA) has a long history as such a method (Goodall 1954, Orlóci 1966, Anderson 1971, Gauch 1982). On the other hand, specific models can be postulated and hypothetical species and sites scores estimated using maximum likelihood (ML) or least squares (Gauch *et al.* 1974, Ihm & van Groenewood 1984). Ter Braak and Prentice (1988) further pointed out that "PCA provides the ML-solution to the multidimensional linear ordination problem if the errors are independent and normally distributed with constant variance across species and sites". However, formal discussions concerning the properties of the estimators obtained using PCA or ML ordination seem scarce in ecological applications. Furthermore, the models that have been proposed typically assume species to act independently of each other, when in fact symbiosis and competition effects are common in ecological studies. This paper formally equates PCA with ML estimation for a linear functional measurement error model. Asymptotic and finite sample properties of the estimators are then examined assuming an equicorrelated covariance structure; i.e. when there

is equal correlation between any two species within a site. This particular correlation structure may not truly reflect how species interact in reality but it serves as an initial approximation, and allows for tractable asymptotic results.

## 2. Model

We assume the response (possibly transformed) for each species to be a linear function of the latent environmental variable with Gaussian additive error. Formally, let  $y_{ij}$  be the abundance of the  $j$ th species at site  $i$ ,  $x_i$  the latent environmental value at site  $i$ ,  $a_j$  and  $b_j$  the intercept and slope for species  $j$  respectively, and  $c$  the baseline species slope, such that

$$y_{ij} = a_j + cx_i b_j + e_{ij}, \quad (i = 1, \dots, n; j = 1, \dots, m), \quad (1)$$

where  $e_i = (e_{i1}, \dots, e_{im})' \sim \text{iid } N(0, \mathbf{V})$ ,  $0$  is a  $m \times 1$  vector of zeros,  $\mathbf{V} = \sigma_e^2(1-\rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J}$ ,  $\mathbf{I}$  is a  $m \times m$  identity matrix,  $\mathbf{J}$  is a  $m \times m$  matrix of ones,  $\sigma_e^2$  is the error variance, and  $\rho$  is the correlation between species. The  $x_i$ 's,  $b_j$ 's, and  $c$  are unknown parameters, whereas the  $a_j$ 's and  $\mathbf{V}$  are assumed given. The introduction of the baseline slope  $c$  may seem artificial and letting  $a_j$  and  $\mathbf{V}$  be known may be unduly restrictive, but this is to illustrate how PCA is related to ML estimation in model (1). The objective is to recover the underlying linear model by estimating the  $b_j$ 's. However, the parameters  $c$ ,  $x_i$  and  $b_j$  are only identified as a product, and thus identifiability constraints must be imposed to obtain unique estimates; e.g.

$$\sum_j b_j^2 = 1 \text{ and } \sum_i x_i^2 = 1.$$

With such a normalization,  $b_j$  is identified *up to a change in sign*.

Model (1) belongs to the class of linear functional measurement error models since it can be rewritten as

$$y_{i1} = a_1 + \beta(\mathbf{v}_i - \mathbf{a}^{(1)}) + e_{i1},$$

$$\mathbf{y}_i^{(1)} = \mathbf{v}_i + \mathbf{e}_i^{(1)},$$

where  $\mathbf{y}_i^{(1)} = (y_{i2}, \dots, y_{im})'$ ,  $\mathbf{a}^{(1)} = (a_2, \dots, a_m)'$ ,  $\mathbf{e}_i^{(1)} = (e_{i2}, \dots, e_{im})'$ ,  $\beta = \frac{b_1}{m-1} \left( \frac{1}{b_2}, \dots, \frac{1}{b_m} \right)$ , and

$\mathbf{v}_i = (a_2 + cx_i b_2, \dots, a_m + cx_i b_m)'$  is the latent variable which is only observed as  $\mathbf{y}_i^{(1)}$  because of

measurement error. In another context, model (1) is the factor analysis model where the  $cb_j$ 's are factor loadings and  $\mathbf{x} = (x_1, \dots, x_n)'$  is an unobservable factor but with fixed instead of random components. Both the functional measurement error model and the factor analysis model have been studied extensively especially in the econometrics and psychometrics literature (Gleser 1981; Anderson 1984; Anderson & Ameniya 1988). However most of these results have not been made readily accessible to ecologists partly because of the unfamiliar notation, and independent errors are often assumed. In the following sections, we summarize some of the results that relate specifically to PCA and ML ordination.

### 3. Existence of MLE

It is well known that the existence of maximum likelihood estimators (MLEs) for functional measurement error models hinges on the structure of the covariance matrix. If  $\mathbf{V}$  were unknown and totally arbitrary, the likelihood would be unbounded and no maximum is attainable (Anderson & Rubin 1956; Kendall & Stuart 1979; Gleser 1981). The log-likelihood for model (1) is

$$\log L = -\frac{mn}{2} \log(2\pi) - \frac{n}{2} \log|\mathbf{V}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{z}_i - c\mathbf{x}_i \mathbf{b})' \mathbf{V}^{-1} (\mathbf{z}_i - c\mathbf{x}_i \mathbf{b}), \quad (2)$$

where  $\mathbf{b} = (b_1, \dots, b_m)'$ ,  $\mathbf{z}_i = (z_{i1}, \dots, z_{im})'$ , and we write  $z_{ij} = y_{ij} - a_j$  for notation convenience.

Now,

$$|\mathbf{V}| = \sigma_e^{2m} |(1-\rho)\mathbf{I} + \rho\mathbf{J}| = \sigma_e^{2m} (1-\rho)^{m-1} (1 + (m-1)\rho),$$

and

$$\mathbf{V}^{-1} = \frac{1}{\sigma_e^2(1-\rho)} \left( \mathbf{I} - \frac{\rho}{1-\rho+m\rho} \mathbf{J} \right).$$

Therefore  $\log L$  is proportional to

$$-\frac{n}{2} (m \log \sigma_e^2 + (m-1) \log(1-\rho) + \log(1 + (m-1)\rho)) - \frac{1}{2\sigma_e^2(1-\rho)} \sum_{i=1}^n \text{tr} \left( \left( \mathbf{I} - \frac{\rho}{1-\rho+m\rho} \mathbf{J} \right) (\mathbf{z}_i - c\mathbf{x}_i \mathbf{b})(\mathbf{z}_i - c\mathbf{x}_i \mathbf{b})' \right).$$

By inspection, we see that  $c$ ,  $x_i$  and  $b_j$  cannot be chosen such that  $(\mathbf{z}_i - cx_i \mathbf{b})(\mathbf{z}_i - cx_i \mathbf{b})'$  is identically zero because of identifiability constraints on the  $x_i$ 's and  $b_j$ 's. This implies that the log-likelihood is bounded away from positive infinity since the terms involving  $\rho$  and  $\sigma_e^2$  that can drive the likelihood towards infinity compensate each other. Hence, for an equicorrelated  $\mathbf{V}$  the existence of MLEs in model (1) is guaranteed, although in general this must be checked for each individual covariance structure. In ML ordination, interest is traditionally on estimating the mean structure and so we assume  $\mathbf{V}$  to be given. This avoids the issue of nonexistence of MLEs and illustrates the connection between ML estimation and PCA.

#### 4. MLE as an Eigenvector Solution

With  $\mathbf{V}$  known, the likelihood equations for the parameters  $c$ ,  $\mathbf{x}$ , and  $\mathbf{b}$  are respectively

$$c(\mathbf{b}'\mathbf{V}^{-1}\mathbf{b})(\mathbf{x}'\mathbf{x}) = \mathbf{x}'\mathbf{Z}\mathbf{V}^{-1}\mathbf{b}, \quad (3a)$$

$$c\mathbf{x}(\mathbf{b}'\mathbf{V}^{-1}\mathbf{b}) = \mathbf{Z}\mathbf{V}^{-1}\mathbf{b}, \text{ and} \quad (3b)$$

$$c\mathbf{b}(\mathbf{x}'\mathbf{x}) = \mathbf{Z}'\mathbf{x}, \quad (3c)$$

where  $\mathbf{Z} = \{z_{ij}\}$  is the  $n \times m$  matrix of species abundances. Equation (3a) can be derived from either (3b) or (3c) and hence is redundant. According to (3b) and (3c), the MLE of  $x_i$  is obtained by a weighted regression of the rows of  $\mathbf{Z}$  on  $\mathbf{b}$ , whereas the MLE of  $b_j$  is obtained by regressing the columns of  $\mathbf{Z}$  on  $\mathbf{x}$ . This suggests that instead of jointly maximizing the likelihood (2), one can find the ML estimates by iterating between the individual likelihood equations (3b) and (3c) until convergence. This algorithm is known as the Gauss-Seidel-Newton (GSN) method (Ortega & Rheinboldt 1970; Thisted 1988). Equations (3b) and (3c) can also be combined to give

$$\mathbf{Z}'\mathbf{Z}\mathbf{V}^{-1}\mathbf{b} = c^2\mathbf{b} \text{ and} \quad (4)$$

$$\mathbf{Z}\mathbf{V}^{-1}\mathbf{Z}'\mathbf{x} = c^2\mathbf{x}, \quad (5)$$

using the normalizations  $\mathbf{b}'\mathbf{V}^{-1}\mathbf{b} = 1$  and  $\mathbf{x}'\mathbf{x} = 1$  to ensure identifiability of the parameters. This shows that  $c^2$  is an eigenvalue, and  $\mathbf{x}$  and  $\mathbf{b}$  are eigenvectors. In fact the following proposition holds.

**Proposition 1.** Let model (1) hold. The MLEs of  $\mathbf{x}$  and  $\mathbf{b}$  are then the dominant eigenvectors (i.e. those associated with the largest eigenvalue) of (4) and (5).

**Proof:** See Appendix.

## 5. Principal Components Analysis

Introduced to ecologists by Goodall (1954), PCA provides an orthogonal least squares approximation to the data via a singular value decomposition (Greenacre 1984). Let  $\mathbf{W} = \{y_{ij} - \bar{y}_{.j}\}$  be the  $n \times m$  matrix of species abundances with the mean of each species subtracted from its columns. If  $\mathbf{W}$  has rank  $r$ , the singular value decomposition of  $\mathbf{W}$  is

$$\mathbf{W} = \sum_{k=1}^r \lambda_k \mathbf{p}_k \mathbf{q}'_k, \text{ such that} \quad (6)$$

$$\lambda_k \mathbf{q}_k = \mathbf{W}' \mathbf{p}_k \text{ and } \lambda_k \mathbf{p}_k = \mathbf{W} \mathbf{q}_k, \quad (7)$$

$$\mathbf{W} \mathbf{W}' \mathbf{p}_k = \lambda_k^2 \mathbf{p}_k, \quad (8)$$

$$\mathbf{W}' \mathbf{W} \mathbf{q}_k = \lambda_k^2 \mathbf{q}_k, \quad (9)$$

$$\mathbf{p}'_k \mathbf{p}_l = \mathbf{q}'_k \mathbf{q}_l = \delta_{kl}, \text{ where } \delta_{kl} \text{ is Kronecker's delta.}$$

PCA is usually used by ecologists as a data reduction technique, with the output summarized in a 2-dimensional plot. The first two terms in (6) are selected and the coordinates  $(\lambda_1 p_{1i}, \lambda_2 p_{2i})$  ( $i = 1, \dots, n$ ) and  $(q_{1j}, q_{2j})$  ( $j = 1, \dots, m$ ) or its other rescaled versions are then plotted together for a visual appraisal of the relationship between species and sites (Gabriel 1971). This application is mainly descriptive and does not explicitly require a model for the data. However, PCA can also be viewed as multivariate linear regression with latent predictors. Choose a rank one approximation to  $\mathbf{W}$  by selecting only the first term in (6), then

$$\lambda \mathbf{q} = \mathbf{W}' \mathbf{p} \text{ and } \lambda \mathbf{p} = \mathbf{W} \mathbf{q}, \text{ yielding} \quad (10)$$

$$\mathbf{W}' \mathbf{W} \mathbf{q} = \lambda^2 \mathbf{q}, \text{ and} \quad (11)$$

$$\mathbf{W} \mathbf{W}' \mathbf{p} = \lambda^2 \mathbf{p}. \quad (12)$$

Observe that (11) and (12) have the same form as the likelihood equations (4) and (5) but with  $\mathbf{V} = \mathbf{I}$ . Now since

$$w_{ij} \equiv y_{ij} - \bar{y}_{.j} = c(x_i - \bar{x})b_j + (e_{ij} - \bar{e}_{.j}), \quad (13)$$

and  $(e_{i1} - \bar{e}_{.1}, \dots, e_{im} - \bar{e}_{.m})' \sim \text{iid } N(\mathbf{0}, (1 - \pi^{-1})\mathbf{V})$ , the likelihood for (13) has the same form as that for model (1). Therefore, solving for  $\lambda$ ,  $p_i$ , and  $q_j$  in (10) is equivalent to finding the MLEs of  $c$ ,  $x_i - \bar{x}$ , and  $b_j$  respectively in (1) but assuming  $\rho = 0$ .

## 6. Consistency

The previous section showed that PCA is equivalent to ML estimation of the species and sites parameters in model (1), but the number of site parameters,  $x_i$ , increases with the sample size  $n$ . This leads to the infinite incidental parameters problem (Neyman & Scott 1948), and causes the ML approach to give inconsistent estimators unless some 'population' assumption is imposed on the  $x_i$ 's. In order to establish consistency we make use of the following lemma.

**Lemma 1** (Ortega 1972). An eigenvalue is a continuous function of the matrix elements. If it is also simple (i.e. has multiplicity one) then its associated eigenvector is a continuous function of the matrix elements.

**Proposition 2.** Let model (1) hold. If  $\lim_{n \rightarrow \infty} \frac{\sum x_i^2}{n} = h$ , where  $h \in (0, \infty)$ , and  $\lambda_{\max}$ , the maximum eigenvalue of  $\mathbf{Z}'\mathbf{Z}\mathbf{V}^{-1}$  is simple, then the MLE of  $\mathbf{b}$  consistently estimates  $s\mathbf{b}$  for some nonzero constant  $s$ .

**Proof:** See Appendix.

The condition that  $\lim_{n \rightarrow \infty} \frac{\sum x_i^2}{n}$  be positive and finite is in fact the univariate version of Gleser's (1981) assumption C for establishing consistency for estimators in errors in variables regression models. Anderson & Taylor (1975) also required a similar condition for proving strong consistency of the least squares estimator when the number of independent variables increases with sample size. Intuitively, this suggests that we regard  $x_i$  as coming from some population with finite variance, which is reminiscent of Kiefer & Wolfowitz's (1956) approach. For the PCA estimator; i.e. the eigenvector of  $\mathbf{W}'\mathbf{W}$ , additional conditions on the eigenvalues are required for consistency.

**Proposition 3.** Let model (1) hold. If  $\lim_{n \rightarrow \infty} \frac{\sum x_i^2}{n}$  is positive and finite,  $\mathbf{b}'\mathbf{1} = 0$ ,  $\lambda_{\max}$  of  $\mathbf{W}'\mathbf{W}$  is simple, and  $\rho\sigma_e^2(\mathbf{1}'\mathbf{1}) < c^2\sigma_x^2(\mathbf{b}'\mathbf{b})$ , where  $\sigma_x^2 \equiv \frac{\sum(x_i - \bar{x})^2}{n}$ . Then the PCA estimator consistently estimates  $s\mathbf{b}$  for some nonzero constant  $s$ .

**Proof:** See Appendix.

In ordination, sometimes only the ordering of the species is important and thus it suffices to estimate the parameters up to a location and scale change. In such cases, the PCA estimator is consistent up to a location and scale change given only the 'population' assumption on the  $x_i$ 's.

**Corollary.** Let model (1) hold. If  $\lim_{n \rightarrow \infty} \frac{\sum x_i^2}{n}$  is positive and finite, and  $\lambda_{\max}$  of  $\mathbf{W}'\mathbf{W}$  is simple, then the PCA estimator consistently estimates  $(s_1\mathbf{1} + s_2\mathbf{b})$  for nonzero constants  $\{s_1, s_2\}$ .

The previous results on consistency are derived assuming that the number of species  $m$  remains fixed but the number of sites  $n$  increases to infinity. This scenario seems appropriate in most ecological studies, but in other applications where it is reasonable for both  $m$  and  $n$  to increase to infinity, consistency can be attained without requiring any condition on the  $x_i$ 's (Haberman 1977; Portnoy 1988). The main condition for consistency is that the ratio of the parameter size over the samples size tends to zero. This implies that consistent estimates of both  $\mathbf{x}$  and  $\mathbf{b}$  in model (1) can be obtained when  $mn^{-1} \rightarrow 0$ . Nevertheless, the question still remains as to how large  $n$  needs to be relative to  $m$  before accurate estimates can be attained.

## 7. Simulations

We compared the finite sample performance of the PCA estimator of  $\mathbf{b}$  (i.e. the MLE assuming independence) with the MLE assuming equicorrelated species. The latter represents the ideal (albeit unrealistic) case when we know the species covariance structure. Data were generated according to model (1), with  $x_i$  simulated from a random uniform (0,1) generator. (Without loss of generality,  $\mathbf{x}$  was normalized to have zero mean.) The number of species,  $m$ , was fixed at 5, and 4 levels of  $\rho$  (.8, .2, -.1, -.2) and 2 levels of  $n$  (30,200) were chosen. Each



simulation used 1000 replicates. Tables 1a and 1b contrast how the behavior of the estimates of  $\mathbf{b}$  depend on the value of  $\sum x_i^2$ .

In Table 1a with  $\sum x_i^2$  set equal to  $n$ , the PCA estimates behaved according to Proposition 3. When  $m\rho\sigma_e^2$  was larger the estimates converged to  $\mathbf{1}$ . When  $c^2\sigma_x^2 \sum b_j^2$  was larger they converged to  $\mathbf{b}$ , and this will be the case whenever  $\rho < 0$ . (Note that absolute values of the estimates were taken since estimates are only determined up to a sign-change.) The mean square errors for the ML estimates were smaller than those for the PCA estimates when  $\rho > 0$ , but were larger when  $\rho < 0$ . This suggests that naively assuming independence may have its advantages under special circumstances.

In Table 1b,  $\sum x_i^2$  was fixed at 4 and  $\sum b_j^2$  was increased such that the variability in the  $y$ 's was comparable to that in Table 1a. As expected, the mean square errors of both the PCA and the ML estimates no longer diminish with sample size, indicating the inconsistency of the estimators. However, the behavior of the PCA estimates can still be predicted by the magnitudes of  $m\rho\sigma_e^2$  and  $c^2\sigma_x^2 \sum b_j^2$ . When  $\rho > 0$ ,  $c^2\sigma_x^2 \sum b_j^2$  can be less than  $m\rho\sigma_e^2$  for large sample sizes, thereby causing the PCA estimates to converge to  $\mathbf{1}$ . As in Table 1a, the mean square errors of the PCA estimates were larger if  $\rho > 0$  but were smaller if  $\rho < 0$ .

Table 2 compares PCA and ML in terms of a summary statistic which is location and scale invariant, and thus without having to specify the identifiability constraints used in model (1). The particular statistic chosen was

$$\frac{1}{(\hat{b}_{[1]} - \hat{b}_{[m]})^2} \sum_{j=1}^m (\hat{b}_j - \hat{b}_m)^2,$$

where  $\hat{b}_{[1]}$  and  $\hat{b}_{[m]}$  are respectively the minimum and maximum elements of the estimate of  $\mathbf{b}$ . As predicted by theory, the biases of the PCA estimates approached zero with increasing sample size regardless of the correlation. The standard deviations of the ML estimates were smaller, although the differences were minimal except for large correlations.

## 8. Summary

ML ordination seeks to uncover the underlying relationship between species and environmental data by alternately estimating the species and site parameters for a specific model. We addressed several important aspects of this technique. First, the model was extended by allowing for interaction between species. Second, the issue between identifiability constraints and uniquely defined estimates was delineated. Third, conditions for consistent estimators of the species parameters were established when the number of latent site parameters increased with sample size. We demonstrated how model (1) is equivalent to a functional linear measurement error model. For this model, PCA is the same as ML estimation of the  $x_i$ 's and  $b_j$ 's for independent errors. It should be noted however that ML allows the estimation of certain error structures but PCA assumes the covariance structure to be known. The MLE of  $\mathbf{b}$  is inconsistent unless some 'population' assumption is imposed on the site parameters. The PCA estimator requires in addition that  $\sum b_j = 0$  and  $c^2 \sigma_x^2 \sum b_j^2 > m \rho \sigma_\epsilon^2$ , although in general it is consistent up to a location and scale change. For finite samples, simulations indicated that the mean square errors of the PCA estimates were larger than those of the ML estimates for positive correlations but can be smaller for negative correlations. Therefore, using PCA may be advantageous when there is reason to believe that species have a negative equicorrelated covariance structure. Whether this would hold true for other covariance structures remains to be verified separately. When the results were summarized in terms of a location and scale invariant statistic, the performances of PCA and ML were generally similar, although the ML estimates tended to have slightly smaller mean square errors.

**Table 1a. Absolute values and mean square errors of ML and PCA estimates.**

$\sum x_i^2 = n$ ,  $\sum b_j^2 = 4$ ,  $\sum b_j = 0$ ,  $\sigma_e^2 = 2$ ,  $c = 1$ , and  $\mathbf{b}' = (-0.853 \ -0.689 \ -0.361 \ 0.295 \ 1.607)$

$\rho$	$n$	ML		PCA		$c^2 \sigma_x^2 \sum b_j^2$	$m\rho\sigma_e^2$
		$ \hat{b}_j $	$\text{mse}(\hat{b}_j)$	$ \hat{b}_j $	$\text{mse}(\hat{b}_j)$		
0.8	30	0.817	0.065	0.854	0.065	4	8
		0.660	0.060	0.850	0.072		
		0.368	0.049	0.850	0.261		
		0.321	0.042	0.847	0.323		
		1.545	0.061	0.884	0.698		
	200	0.851	0.010	0.891	0.010		
		0.689	0.010	0.892	0.048		
		0.362	0.010	0.892	0.284		
		0.291	0.011	0.888	0.353		
		1.592	0.009	0.883	0.552		
0.2	30	0.822	0.073	0.796	0.123	4	2
		0.646	0.078	0.648	0.114		
		0.386	0.062	0.435	0.093		
		0.318	0.048	0.388	0.084		
		1.534	0.051	1.440	0.164		
	200	0.850	0.012	0.844	0.028		
		0.690	0.013	0.686	0.027		
		0.361	0.013	0.363	0.026		
		0.299	0.013	0.299	0.025		
		1.589	0.007	1.570	0.024		
-0.1	30	0.816	0.080	0.818	0.076	4	-1
		0.654	0.087	0.654	0.083		
		0.392	0.065	0.389	0.062		
		0.349	0.062	0.345	0.059		
		1.520	0.043	1.526	0.037		
	200	0.844	0.012	0.845	0.011		
		0.679	0.013	0.680	0.012		
		0.361	0.013	0.362	0.013		
		0.296	0.014	0.296	0.014		
		1.597	0.005	1.598	0.004		
-0.2	30	0.793	0.085	0.794	0.081	4	-2
		0.662	0.089	0.667	0.086		
		0.384	0.065	0.382	0.064		
		0.352	0.060	0.346	0.056		
		1.529	0.038	1.533	0.034		
	200	0.845	0.013	0.845	0.013		
		0.689	0.016	0.689	0.015		
		0.353	0.015	0.353	0.015		
		0.292	0.015	0.292	0.014		
		1.593	0.004	1.594	0.004		

**Table 1b. Absolute values and mean square errors of ML and PCA estimates.**

$\sum x_i^2 = 4$ ,  $\sum b_j^2 = 100$ ,  $\sum b_j = 0$ ,  $\sigma_e^2 = 2$ ,  $c = 1$ , and  $\mathbf{b}' = (-4.263 \ -3.443 \ -1.804 \ 1.476 \ 8.034)$

$\rho$	$n$	ML		PCA		$c^2 \sigma_x^2 \sum b_j^2$	$m\rho\sigma_e^2$
		$ \hat{b}_j $	$\text{mse}(\hat{b}_j)$	$ \hat{b}_j $	$\text{mse}(\hat{b}_j)$		
0.8	30	4.208	0.453	4.064	2.033	13	8
		3.407	0.493	3.314	1.967		
		0.771	0.482	1.928	1.431		
		1.466	0.448	1.683	1.232		
		7.944	0.427	7.544	2.729		
	200	4.220	0.547	4.469	0.106	2	8
		3.414	0.528	4.468	1.098		
		1.808	0.525	4.466	7.118		
		1.469	0.536	4.469	8.989		
		7.905	0.512	4.450	13.016		
0.2	30	4.232	0.468	4.227	0.546	13	2
		3.425	0.525	3.421	0.604		
		1.798	0.509	1.801	0.571		
		1.432	0.487	1.434	0.538		
		7.925	0.290	7.907	0.374		
	200	4.149	0.786	3.849	4.160	2	2
		3.361	0.810	3.546	3.459		
		1.832	0.772	3.034	3.993		
		1.492	0.727	2.983	4.609		
		7.898	0.471	5.828	13.162		
-0.1	30	4.221	0.467	4.222	0.457	13	-1
		3.446	0.503	3.447	0.493		
		1.779	0.533	1.778	0.526		
		1.479	0.534	1.479	0.527		
		7.920	0.193	7.923	0.183		
	200	4.207	0.824	4.213	0.748	2	-1
		3.357	0.922	3.362	0.840		
		1.784	0.856	1.780	0.818		
		1.512	0.736	1.503	0.704		
		7.868	0.358	7.885	0.272		
-0.2	30	4.232	0.463	4.233	0.458	13	-2
		3.401	0.498	3.402	0.490		
		1.784	0.559	1.785	0.552		
		1.489	0.522	1.489	0.513		
		7.934	0.147	7.935	0.140		
	200	4.157	0.819	4.163	0.776	2	-2
		3.373	0.956	3.372	0.902		
		1.823	0.922	1.822	0.900		
		1.566	0.852	1.556	0.847		
		7.858	0.325	7.869	0.273		

**Table 2. Biases and standard deviations of ML and PCA estimates of a location and scale invariant statistic.  $\sum x_i^2 = n$ ,  $\sigma_e^2 = 2$ ,  $c = 1$ , and  $\mathbf{b}' = (0.054 \ 0.108 \ 0.217 \ 0.433 \ 0.866)$ .**

$\rho$	$n$	ML		PCA	
		Bias	Std	Bias	Std
0.8	10	.086	.266	.154	.356
	20	.035	.213	.074	.257
	40	.012	.160	.031	.197
	80	.001	.128	.009	.153
	160	.000	.090	.002	.113
0.2	10	.310	.438	.317	.442
	20	.199	.357	.202	.356
	40	.078	.272	.082	.272
	80	.044	.217	.044	.222
	160	.019	.173	.021	.174
-0.1	10	.391	.475	.396	.485
	20	.220	.403	.218	.402
	40	.118	.317	.119	.317
	80	.064	.249	.065	.252
	160	.020	.186	.021	.187
-0.2	10	.354	.456	.370	.487
	20	.212	.373	.231	.380
	40	.109	.307	.123	.309
	80	.062	.251	.070	.256
	160	.029	.190	.034	.192

## REFERENCES

- Anderson, A. J. B. (1971). Ordination methods in ecology. *Journal of Ecology* **59**:713-726.
- Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of statistics* **5**, 1-45.
- Anderson, T. W. & Ameniya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *Annals of statistics* **16**, 759-771.
- Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics & Probability* **5**, 111-150.
- Anderson, T. W. & Taylor, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Annals of statistics* **4**, 788-790.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal components analysis. *Biometrika* **58**, 453-467.
- Gauch, H. G., Chase, G. B. & Whittaker, R. H. (1974). Ordinations of vegetation samples by Gaussian species distributions. *Ecology* **55**, 1382-1390.
- Gauch, H. G. (1982). *Multivariate Analysis in Community Ecology*. Cambridge, England: Cambridge University Press.
- Gleser, L. J. (1981). Estimation in a multivariate 'errors-in-variables' regression model, Large sample results. *Annals of Statistics* **9**, 24-44.
- Goodall, D. W. (1954). Objective methods for the classification of vegetation. III. A essay in the use of factor analysis. *Australian Journal of Botany* **1**, 39-63.
- Haberman, S. J. (1977). Maximum likelihood estimation in exponential response models. *Annals of Statistics* **5**, 815-841.
- Ihm, P. & van Groenewoud, H. (1984). Correspondence analysis and Gaussian ordination. *COMPSTAT lectures* **3**, 5-60.
- Kendall, M. G. & Stuart, A. (1979). *The Advanced Theory of Statistics*, vol. 2, 4th edition. New York: Macmillan.
- Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals of Mathematical Statistics* **27**, 887-906.

- Neyman, J. & Scott, E. L. (1948). Consistent estimators based on partially consistent observations. *Econometrica* **16**, 1-32.
- Orlóci, L. (1966). Geometric models in ecology. I. The theory and application of some ordination methods. *Journal of Ecology* **54**, 193-215.
- Ortega, J. & Rheinboldt, W. (1970). *Iterative solution of nonlinear equations in several variables*. New York: Academic Press.
- Ortega, J. M. (1972). *Numerical Analysis. A Second Course*. New York: Academic Press.
- Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics* **16**, 356-366.
- Ter Braak, C. J. F. & Prentice, I. C. (1988). A theory of gradient analysis. *Advances of Ecological Research* **18**, 271-317.
- Thisted, R. A. (1988). *Elements of Statistical Computing*. London: Chapman & Hall.

## APPENDIX

*Proof of Proposition 1.* From (2) the negative log-likelihood is proportional to

$$\begin{aligned} & \sum_{i=1}^n \text{tr}((\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})'\mathbf{V}^{-1}(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})) \\ &= \sum_{i=1}^n \text{tr}(\mathbf{V}^{-1}(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})(\mathbf{z}_i - c\mathbf{x}_i\mathbf{b})') = \text{tr}(\mathbf{V}^{-1}(\mathbf{Z} - c\mathbf{x}\mathbf{b}')'(\mathbf{Z} - c\mathbf{x}\mathbf{b}')) \\ &= \text{tr}(\mathbf{V}^{-1}\mathbf{Z}'\mathbf{Z}) + \text{tr}(c^2\mathbf{b}'\mathbf{V}^{-1}\mathbf{b}\mathbf{x}'\mathbf{x}) - 2\text{tr}(c\mathbf{V}^{-1}\mathbf{b}\mathbf{x}'\mathbf{Z}). \end{aligned}$$

With  $\mathbf{b}'\mathbf{V}^{-1}\mathbf{b} = 1$  and  $\mathbf{x}'\mathbf{x} = 1$ ,  $\text{tr}(c^2\mathbf{b}'\mathbf{V}^{-1}\mathbf{b}\mathbf{x}'\mathbf{x}) = c^2$  and from (3b)

$$2\text{tr}(c\mathbf{V}^{-1}\mathbf{b}\mathbf{x}'\mathbf{Z}) = 2\text{tr}(c^2\mathbf{x}\mathbf{x}') = 2c^2.$$

Therefore, the log-likelihood is proportional to  $c^2$  and attains its maximum for the largest  $c^2$ , which implies that the dominant eigenvectors of (4) and (5) are the MLEs.

*Proof of Proposition 2.* Recall that the MLE is the dominant eigenvector of (4). Under model (1), the  $(j,k)$ th element  $(j,k = 1, \dots, m)$  of  $\mathbf{Z}'\mathbf{Z}$  is

$$\begin{aligned} \sum_i z_{ij}z_{ik} &= \sum_i (cb_jx_i + e_{ij})(cb_kx_i + e_{ik}) \\ &= c^2b_jb_k\sum x_i^2 + cb_j\sum x_ie_{ij} + cb_k\sum x_ie_{ik} + \sum e_{ij}e_{ik}. \end{aligned}$$

According to the iid Weak Law of Large Numbers,

$$\frac{1}{n} \sum_i e_{ij}e_{ik} \rightarrow \rho\sigma_e^2 \text{ for } j \neq k, \text{ and } \frac{1}{n} \sum_i e_{ik}^2 \rightarrow \sigma_e^2.$$

Furthermore, when  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_i x_i^2 = h$ , for  $h \in (0, \infty)$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \text{Var}(\sum_i x_ie_{ik}) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_i x_i^2 \sigma_e^2 = \lim_{n \rightarrow \infty} \frac{h\sigma_e^2}{n} = 0,$$

and thus  $\lim_{n \rightarrow \infty} \frac{\sum_i x_ie_{ik}}{n} \rightarrow 0$ . Define  $\mu_{2x} \equiv \frac{\sum x_i^2}{n}$ . It follows that

$$n^{-1}\mathbf{Z}'\mathbf{Z}\mathbf{V}^{-1} \rightarrow \Sigma, \text{ where } \Sigma = (\sigma_e^2(1 - \rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J} + c^2\mu_{2x}\mathbf{b}\mathbf{b}')\mathbf{V}^{-1}.$$

Observe that

$$\begin{aligned} \Sigma &= (\sigma_e^2(1 - \rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J} + c^2\mu_{2x}\mathbf{b}\mathbf{b}') \frac{1}{\sigma_e^2(1 - \rho)} \left( \mathbf{I} + \frac{\rho}{1 - \rho + m\rho} \mathbf{J} \right) \\ &= (\mathbf{I} + k\mathbf{b}\mathbf{b}' - \frac{k(\mathbf{b}'\mathbf{1})\rho}{(1 - \rho + m\rho)} \mathbf{b}\mathbf{1}'), \text{ where } k = \frac{c^2\mu_{2x}}{\sigma_e^2(1 - \rho)}. \end{aligned}$$



The eigenvectors of  $\Sigma$  are equivalently those defined by the equation

$$(k\mathbf{b}\mathbf{b}' - \frac{k(\mathbf{b}'\mathbf{1})\rho}{(1-\rho+m\rho)}\mathbf{b}\mathbf{1}')\mathbf{u} = (\lambda - 1)\mathbf{u},$$

where  $\lambda$  is an eigenvalue of  $\Sigma$  and  $\mathbf{u}$  some vector. Now  $k\mathbf{b}\mathbf{b}' - \frac{k(\mathbf{b}'\mathbf{1})\rho}{(1-\rho+m\rho)}\mathbf{b}\mathbf{1}'$  has rank 1, and it is easily verified that  $\mathbf{b}$  is an eigenvector with eigenvalue  $k(\mathbf{b}'\mathbf{b} - \frac{\rho(\mathbf{b}'\mathbf{1})^2}{1-\rho+m\rho})$ . The

other  $m - 1$  eigenvalues all equal zero. This implies that  $\Sigma$  has eigenvalues

$$1 + k(\mathbf{b}'\mathbf{b} - \frac{\rho(\mathbf{b}'\mathbf{1})^2}{1-\rho+m\rho}), \text{ and } 1 \text{ with multiplicity } m - 1.$$

Furthermore, since

$$k(\mathbf{b}'\mathbf{b} - \frac{\rho(\mathbf{b}'\mathbf{1})^2}{1-(m-1)\rho}) > 0 \text{ within the range of } \rho,$$

$\mathbf{b}$  is the dominant eigenvector. To complete the proof, let  $\hat{\mathbf{b}}$  be the dominant eigenvector of  $\mathbf{Z}'\mathbf{Z}\mathbf{V}^{-1}$  with eigenvalue  $\lambda_{\max}$ . It follows by Lemma 1 that since  $n^{-1}\mathbf{Z}'\mathbf{Z}\mathbf{V}^{-1} \rightarrow \Sigma$ ,  $n^{-1}\lambda_{\max}$  converges to the largest eigenvalue of  $\Sigma$ , and  $n^{-1}\hat{\mathbf{b}}$  converges to its dominant eigenvector; i.e.  $s\mathbf{b}$  for some nonzero constant  $s$ .

*Proof of Proposition 3.* Under model (13), the  $(j, k)$ th element of  $\mathbf{W}'\mathbf{W}$  is

$$\sum_i w_{ij}w_{ik} = c^2b_jb_k\Sigma(x_i^2 - \bar{x}) + cb_j\Sigma(x_i - \bar{x})\varepsilon_{ij} + cb_k\Sigma(x_i - \bar{x})\varepsilon_{ik} + \Sigma\varepsilon_{ij}\varepsilon_{ik},$$

where  $\varepsilon_{ij} = e_{ij} - \bar{e}_{.j}$ . Now

$$\frac{1}{n} \sum_i \varepsilon_{ij}\varepsilon_{ik} \rightarrow \rho\sigma_e^2 \text{ for } j \neq k, \text{ and } \frac{1}{n} \sum_i \varepsilon_{ik}^2 \rightarrow \sigma_e^2.$$

When  $\lim_{n \rightarrow \infty} \frac{\Sigma x_i^2}{n}$  is positive and finite,  $\bar{x}$  is finite and thus  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_i (x_i - \bar{x})\varepsilon_{ik} \rightarrow 0$ . With

$$\sigma_x^2 \equiv \frac{\Sigma(x_i - \bar{x})^2}{n}, \text{ it follows that}$$

$$n^{-1}\mathbf{W}'\mathbf{W} \rightarrow \Sigma, \text{ where } \Sigma = \sigma_e^2(1-\rho)\mathbf{I} + \rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}'.$$

The eigenvectors of  $\Sigma$  are equivalently those defined by

$$(\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}')\mathbf{u} = (\lambda - \sigma_e^2(1-\rho))\mathbf{u}.$$

Now  $(\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}')\mathbf{b} = \rho\sigma_e^2(\mathbf{1}'\mathbf{b})\mathbf{1} + c^2\sigma_x^2(\mathbf{b}'\mathbf{b})\mathbf{b}$  which cannot be a multiple of  $\mathbf{b}$  unless  $\mathbf{b}'\mathbf{1} = 0$ . The matrix  $\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}'$  has rank 2. When  $\mathbf{b}'\mathbf{1} = 0$ , it has eigenvectors  $\mathbf{1}$  and  $\mathbf{b}$  with eigenvalues  $\rho\sigma_e^2(\mathbf{1}'\mathbf{1})$  and  $c^2\sigma_x^2\mathbf{b}'\mathbf{b}$  respectively. The other  $m - 2$  eigenvalues all equal zero. This implies that  $\Sigma$  has eigenvalues

$$m\rho\sigma_e^2 + (1 - \rho)\sigma_e^2, c^2\sigma_x^2(\mathbf{b}'\mathbf{b}) + (1 - \rho)\sigma_e^2, \text{ and } (1 - \rho)\sigma_e^2 \text{ with multiplicity } m - 2.$$

Let  $\hat{\mathbf{b}}$  be the dominant eigenvector of  $\mathbf{W}'\mathbf{W}$  with eigenvalue  $\lambda_{\max}$ . It follows by Lemma 1 that when  $c^2\sigma_x^2(\mathbf{b}'\mathbf{b}) > m\rho\sigma_e^2$ ,  $n^{-1}\hat{\mathbf{b}}$  converges to  $s\mathbf{b}$  for some nonzero constant  $s$ .

**Proof of Corollary.** According to Proposition 3,  $n^{-1}\hat{\mathbf{b}}$  converges to the dominant eigenvector of  $\Sigma$ , which is also the dominant eigenvector of  $\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}'$ . Now since

$$(\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}')\mathbf{u} = \rho\sigma_e^2(\mathbf{1}'\mathbf{u})\mathbf{1} + c^2\sigma_x^2(\mathbf{b}'\mathbf{u})\mathbf{b},$$

all nontrivial eigenvectors of  $\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}'$  are of the form  $(s_1\mathbf{1} + s_2\mathbf{b})$  for nonzero constants  $s_1$  and  $s_2$ . Observe that

$$(\rho\sigma_e^2\mathbf{J} + c^2\sigma_x^2\mathbf{b}\mathbf{b}')(s_1\mathbf{1} + s_2\mathbf{b}) = \rho\sigma_e^2(ms_1 + s_2(\mathbf{b}'\mathbf{1}))\mathbf{1} + c^2\sigma_x^2(ms_1 + s_2(\mathbf{b}'\mathbf{b}))\mathbf{b}.$$

This implies that  $(s_1\mathbf{1} + s_2\mathbf{b})$  is an eigenvector whenever

$$\rho\sigma_e^2(ms_1 + s_2(\mathbf{b}'\mathbf{1})) = \lambda s_1 \text{ and } c^2\sigma_x^2(ms_1 + s_2(\mathbf{b}'\mathbf{b})) = \lambda s_2,$$

which solving for  $\lambda$  gives

$$mc^2\sigma_x^2s_1^2 + s_1s_2(m\rho\sigma_e^2 - c^2\sigma_x^2(\mathbf{b}'\mathbf{b})) + \rho\sigma_e^2(\mathbf{b}'\mathbf{1})s_2^2 = 0.$$

This conic equation always has a nonzero solution set  $\{s_1, s_2\}$ , and thus the PCA estimator is consistent up to a location and scale change.