Multivariate Behavioral Research, 45:910–934, 2010

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273171.2010.532361

Psychology Press
Taylor & Francis Group

Selection Between Linear Factor Models and Latent Profile Models Using Conditional Covariances

Peter F. Halpin and Michael D. Maraun
Simon Fraser University

A method for selecting between K-dimensional linear factor models and $(K + 1)$-class latent profile models is proposed. In particular, it is shown that the conditional covariances of observed variables are constant under factor models but nonlinear functions of the conditioning variable under latent profile models. The performance of a convenient inferential method suggested by the main result is examined via data simulation and is shown to have acceptable error rate control when deciding between the 2 types of models. The proposed test is illustrated using examples from vocational assessment and developmental psychology.

It is usual to make a distinction between continuous and categorical variables in psychometric theory. Models involving latent variables of either one type or the other are well known (e.g., Bartholomew & Knott, 1999), and more recent innovations have allowed both continuous and categorical latent quantities to be specified within a single mixture or hybrid model (e.g., De Boeck, Wilson & Acton, 2005; Muthén, 2007). Consequently, there is a great deal of choice about the appropriate level of measurement when representing substantive concepts as latent variables. The general topic of this article is how to make such choices, and in particular the problem of selecting between linear factor models and latent profile models is addressed.

Correspondence concerning this article should be addressed to Peter F. Halpin, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands. E-mail: p.f.halpin@uva.nl

Approaches to selecting between continuous and categorical latent variables can be roughly categorized into three types: (a) selection based on statistics measuring overall goodness of fit, (b) graphical methods, and (c) taxometrics. In this introduction each of these is reviewed in turn before outlining the present approach. The following discussion employs a terminological distinction between *structures* and *models*. The former refers to a family of parametric probability distributions defined by a common set of equations, for example, the set of all normal distributions. The term *model* is reserved for a particular parameterization of a structure, for example $N(0, 1)$.

Goodness of fit indices do not consistently distinguish continuous and categorical latent variables. In particular, it has been shown that various K-dimensional linear factor structures (KLF) and $(K+1)$-class latent profile structures (KLP) are formally equivalent at the level of the second-order moments of the observed variables (Bartholomew & Knott, 1999, chap. 6; McDonald, 1967, § 4.3; see also the following section of this article). This means that, for any given factor model, a latent profile model with an identical model-implied covariance matrix always exists (P. C. M. Molenaar & von Eye, 1994). For this reason, goodness of fit statistics based on unconditional covariances (e.g., Browne, 1984) cannot be used to discriminate the two structures.

More recently Bauer and Curran (2003, 2004) have addressed the conflation of finite mixtures with normal components and nonnormal homogeneous populations. In this more general context Lubke and Neale (2006, 2008) have argued that goodness of fit based on the joint distribution of the manifest variables, rather than only the second-order moments, should provide a more effective means of discriminating the two types of latent spaces. Their simulation studies indicate that goodness of fit indices based on likelihoods can correctly identify the type of data-generating model. The success of this approach appears to be contingent on large class separation, the number of observations per class, and on the particular fit statistic(s) employed. In cases where the class separations are moderate (e.g., $\leq$1 within-class *SD*), the practical difficulties of detecting mixtures via their marginal distributions are well known (Everitt & Hand, 1981; Frühwirth-Schnatter, 2006). It is desirable to pursue methods that address these difficulties.

McDonald (1967) proposed a graphical method for distinguishing latent profiles from either linear or nonlinear factor models. His suggestion was to examine the modality of factor scores, which are unimodal for factor structures but will asymptotically approach a discrete distribution with $(K+1)$ peaks for KLP. However, as indicated by Steinley and McDonald's (2007) simulation study, for finite sample size the usefulness of this approach is also contingent on class separation and the observed mixing proportions. In practice this approach is not unlike other graphical approaches based on marginal distributions (e.g., Everitt & Hand, 1981, chap. 5).

With regard to selection (as opposed to estimation of model parameters), Meehl's taxometric methods such as MAXCOV and MAXEIG are also largely graphical in nature (e.g., Ruscio, Haslam, & Ruscio, 2006; Waller & Meehl, 1998). A major restriction of such methods is that they are only formulated for the case of $K = 1$. Nonetheless, procedures such as MAXCOV (Meehl & Yonce, 1996) provide an important innovation because they do not employ the marginal distribution of the manifest variables per se but the distribution of two manifest variables conditional upon a third. A similar but more general method for the detection of differential item functioning between two groups was described by Holland and Rosenbaum (1986) under the rubric of conditional association. In the current article, this basic insight is applied to the case of $K \geq 1$ for continuously distributed outcome variables.

The basic outcome of the present research is as follows: When two or more observed variables are conditioned on a linear combination of the remaining observed variables, Y, the resultant covariance functions in Y are constant under KLF but nonlinear under KLP. This is a sharp, observable distinction between KLF and KLP that has the following advantages: (a) it requires only mild and plausible restrictions on the parameter arrangements of particular models, (b) it is insensitive to misspecification of K, and (c) it leads to a convenient asymptotic chi-square test under the null hypothesis of KLF.

It is important to emphasize at the outset the logic of the proposed statistical test. In particular, it is based only on the sufficiency and not the necessity of the two structures considered in this article. This means that the test can only be used as evidence against either one or the other type of model, and the valid form of argument here is *modus tollens*. To elaborate this point, consider the following informal null and alternative hypotheses:

H_o: The conditional covariances of the manifest variables are constant.
H_1: The conditional covariances of the manifest variables are not constant.

Upon rejecting (accepting) H_o, *any* model that implies constant conditional covariances may also be rejected (accepted). In this article we show that the linear factor models are one such structure but that KLP is not. Therefore rejection of H_o implies a rejection of KLF (but not KLP) and rejection of H_1 implies rejection of KLP (but not KLF). This is how the test discriminates the two structures, by providing evidence *against* either one or the other.

The test does not allow one to make a decisive conclusion *for* KLF or KLP. This is because other types of models exist that also imply H_o or H_1. Therefore, to argue from H_1 to KLP is to affirm the consequent, as is the argument from H_o to KLF. For example, it is shown here that heteroscedastic factor models (e.g., Hessen & Dolan, 2009) imply that H_1 is true, and so these are also plausible candidates if H_1 is retained. It is also the case that many regression models with Gaussian error terms will imply H_o. As another example, nonlinear factor models (e.g., Bauer, 2005; McDonald, 1967) can imply constant, linear, or nonlinear conditional covariance functions, depending on the moments of the posterior distribution of the latent variables.

To summarize, the interpretation of H_o and H_1 is not restricted to the two structures under consideration in this article because these structures are only sufficient but not necessary for the hypotheses. When researchers are concerned to distinguish these two structures (i.e., to reject one or the other), we show that a test of H_o against H_1 can be used for that purpose and we provide such a test. As with any null hypothesis test, it should be interpreted within an overall research context. The reader may also find it useful to revisit Bollen's (1989) remarks on statistical tests in the structural equation modeling (SEM), which are equally relevant to the present research.

The remainder of this article is organized as follows: The next section specifies the models of interest and describes the selection problem. Then our solution is presented and illustrated by means of two examples. The first example uses vocational aptitude test data from the Armed Services Vocational Aptitude Battery (ASVAB; see Segall, 2004) and illustrates a case in which the test rejects the latent profile structure. The second example is based on the responses of female children to the Water Level Task (WLT; Piaget & Inhelder, 1969), and it shows that the test rejects KLF when the data are known to be KLP. Finally, a simulation study is used to describe the error rates of the proposed inferential procedure. In the discussion section of this article we consider ways of extending the proposed approach to a broader class of structures.

THE PROBLEM OF DISTINGUISHING LINEAR FACTOR MODELS AND LATENT PROFILES MODELS

Definitions of the K-dimensional linear factor structure and $(K+1)$-class latent profile structure are given directly. In all cases let $\mathbf{x}' = [X_1, \ldots, X_J]$ denote a J-vector of continuous manifest variables. Let $\boldsymbol{\theta}$ denote either a K-vector of continuous latent variables (under KLF) or a $(K+1)$-class latent variable (under KLP). This dual use of $\boldsymbol{\theta}$ simplifies the discussion greatly and context settles its interpretation. The notation f_U represents the density function of U when U is continuous and in the case that U is finite-valued, f_U denotes its discrete mass function. It is assumed that $\mathbf{x}$ and $\boldsymbol{\theta}$ are jointly distributed in the population of interest.

Using this notation, KLF is defined in terms of the normal factor model of Bartholomew and Knott (1999):

$$\boldsymbol{\theta} \sim N_K(\mathbf{0}, \mathrm{I}) \tag{1}$$

and

$$\mathbf{x}|\boldsymbol{\theta} \sim N_J(\Lambda\boldsymbol{\theta}, C(\mathbf{x}|\boldsymbol{\theta})), \tag{2}$$

where Λ is a $J \times K$ matrix ($J > K$) of rank K and $C(\mathbf{x}|\boldsymbol{\theta})$ is diagonal. For the latent profile structure (e.g., Lazarsfeld & Henry, 1968; Vermunt & Magidson, 2002), two analogous statements are employed:

$$f_{\boldsymbol{\theta}} = \begin{cases} \pi_k, & \boldsymbol{\theta} = k \\ 0, & \boldsymbol{\theta} \neq k \end{cases} \qquad k = 0, \ldots, K \tag{3}$$

with $\pi_k > 0$ and $\sum \pi_k = 1$, and

$$\mathbf{x}|\boldsymbol{\theta} = k \sim N_J(E(\mathbf{x}|\boldsymbol{\theta} = k), C(\mathbf{x}|\boldsymbol{\theta} = k)), \tag{4}$$

with the $E(\mathbf{x}|\boldsymbol{\theta} = k)$ being linearly independent and $C(\mathbf{x}|\boldsymbol{\theta} = k)$ diagonal for $k = 0, \ldots, K$. Linear independence of the conditional expectations implies the more familiar requirement that, for all $k \neq l$,

$$\boldsymbol{\delta}_{kl} \equiv E(\mathbf{x}|\boldsymbol{\theta} = k) - E(\mathbf{x}|\boldsymbol{\theta} = l) \neq \mathbf{0}, \tag{5}$$

which is used later. As described in the following section, a main source of continued interest in *K*LP is that it is indistinguishable from *K*LF when using goodness of fit indices based on the covariance matrix of **x**.

Statement of the Model Selection Problem

The covariance decomposition

$$\Sigma \equiv C(\mathbf{x}) = E_{\boldsymbol{\theta}}[C(\mathbf{x}|\boldsymbol{\theta})] + C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})] \; = \Psi + \Lambda\Lambda', \tag{6}$$

with Ψ diagonal and rank $(\Lambda\Lambda') = K$, is the usual premise for estimation of factor models (Joreskog, 1967). It has also been discussed how this same decomposition is implied by *K*LP (e.g., Bartholomew & Knott, 1999, chap. 6; McDonald, 1967 chap. 4; P. C. M. Molenaar & von Eye, 1994; Steinley & McDonald, 2007). In applied contexts, this means that *K*LF and *K*LP have identical fit at the level of second-order moments, and so covariance-based tests of goodness of fit (e.g., those used in SEM) cannot distinguish the models. We briefly provide a more thorough reiteration of this selection problem.

As with *K*LF, the matrix Ψ is diagonal under *K*LP by local independence. The factorization and rank of $C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})]$ under *K*LP can be established as follows: Define M to be the $J \times (K+1)$ matrix whose rth column is equal to $E(\mathbf{x}|\boldsymbol{\theta} = r - 1)$ and let $\mathrm{D} \equiv \mathrm{diag}\{\pi_k\}$. Then rank(M) = rank(D) = $K + 1$. Also let $\mathbf{1}$ denote a conformable vector whose elements are equal to 1.

Using this notation $E(\mathbf{x}) = \mathrm{MD}\mathbf{1}$ and

$$C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})] = [\mathrm{M} - \mathrm{MD}\mathbf{11}']\mathrm{D}[\mathrm{M} - \mathrm{MD}\mathbf{11}']' \\ = \mathrm{M}[\mathrm{I} - \mathrm{D}\mathbf{11}']\mathrm{D}[\mathrm{I} - \mathrm{D}\mathbf{11}']'\mathrm{M}'. \tag{7}$$

Because the order of $[\mathrm{I} - \mathrm{D}\mathbf{11}']$ is $K + 1$, the rank of the product in Equation (7) is equal to that of $[\mathrm{I} - \mathrm{D}\mathbf{11}']$ (Harville, 1997, §17.5). It is also readily verified that $[\mathrm{I} - \mathrm{D}\mathbf{11}']$ is idempotent. Hence

$$\mathrm{rank}(C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})]) = \mathrm{tr}(\mathrm{I} - \mathrm{D}\mathbf{11}') = (K + 1) - \mathrm{tr}(\mathrm{D}\mathbf{11}') = K. \tag{8}$$

Equation (8) shows that $C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})]$ has the same rank under *K*LP as under *K*LF. Furthermore, because $C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})]$ is Grammian by construction and its rank is K, there exists a $J \times K$ matrix U such that $C_{\boldsymbol{\theta}}[E(\mathbf{x}|\boldsymbol{\theta})] = \mathrm{UU}'$ (Harville, 1997, theorem 14.3.7). Thus *K*LP implies the covariance decomposition given in Equation (6).

This problem was initially posed by McDonald (1967, § 4.3), although in that case the situation was further complicated by considering nonlinear factor models as a third alternative. Bartholomew (1987, § 2.4) provided an independent statement of the problem and interpreted it as demonstrating that latent distributions are poorly determined by observed second-order moments. At this point a satisfactory solution has yet to be proposed (but see Lubke & Neale, 2006; Steinley & McDonald, 2007), and this is the main motivation of this article.

EMPIRICAL DISCRIMINATION OF *K*LF AND *K*LP

In this section the covariance matrix of a partition of manifest variables conditional on a function of the remaining manifest variables is derived under *K*LF and *K*LP. Formulation of conditional covariances of this type has precedent in item response theory (e.g., Holland & Rosenbaum, 1986) and Meehl's taxometric procedures (e.g., Waller & Meehl, 1998). The graphical interpretation of these results is briefly considered before moving on to consider an inferential procedure.

Conditional Covariances Under *K*LF and *K*LP

Begin by partitioning $\mathbf{x}$ as $[\mathbf{x}_p, \mathbf{x}_{J-p}]'$ where $\mathbf{x}_p$ is a selection of $2 < p < J$ manifest variables to be referred to as covariance generating or *output* indicators (Meehl & Yonce, 1996). The remaining $(J - p)$ variables can be used to define the scalar-valued function $Y \equiv f(\mathbf{x}_{J-p})$, which is referred to as an *input* or conditioning indicator.

The following remarks about input and output indicators are relevant to further developments. First, in the sequel it is required that $Y|\boldsymbol{\theta}$ is normally distributed under both structures. As the structures are specified earlier, this is assured when Y is a linear combination of one or more manifest variables. In practice, the use of linear combinations with a large number of indicators is advisable when the normality of $Y|\boldsymbol{\theta}$ is in question. It is also required that, under KLP, at least one class separation is nonnull on both Y and $\mathbf{x}_p$ (see Equation (14) later). This is the only restriction we place on the parameters of either structure. Intuitively, it means that Y and $\mathbf{x}_p$ are also indicators of KLP. For KLF, this requirement follows directly from its specification. Finally, note that $[\mathbf{x}_p, Y]'$ satisfies the requirement of local independence for both KLF and KLP, which follows from the local independence of $\mathbf{x}$ (Loeve, 1963, p. 224).

Because the following analysis treats the manifest variables mainly in terms of $\mathbf{x}_p$ and Y, the subscript p will be dropped for notational simplicity where this is unambiguous (e.g., when conditioning on Y). The consequence is merely to consider a p-vector rather than a J-vector of manifest variables under the notation of $\mathbf{x}$.

Denote the covariance matrix of a set of p output indicators conditional on an input indicator by $\Sigma_{\mathbf{x}|Y} \equiv C(\mathbf{x}|Y)$. Letting the subscript $\boldsymbol{\theta}|Y$ denote expectation over the posterior density of $\boldsymbol{\theta}$, the usual method of covariance decomposition yields

$$\begin{aligned}\Sigma_{\mathbf{x}|Y} &= E_{\mathbf{x}|Y}[C(\mathbf{x}|Y,\boldsymbol{\theta})] + C_{\boldsymbol{\theta}|Y}[E(\mathbf{x}|Y,\boldsymbol{\theta})] \\ &= E_{\boldsymbol{\theta}|Y}[C(\mathbf{x}|\boldsymbol{\theta})] + C_{\boldsymbol{\theta}|Y}[E(\mathbf{x}|\boldsymbol{\theta})]\end{aligned} \tag{9}$$

with the second equality in Equation (9) following from local independence of $[\mathbf{x}_p, Y]'$. Note that both latent structures imply that the matrix $\Psi_{\mathbf{x}|Y} \equiv E_{\boldsymbol{\theta}|Y}[C(\mathbf{x}|\boldsymbol{\theta})]$ is diagonal and hence that the off-diagonal elements of $\Sigma_{\mathbf{x}|Y}$ may be written using only $C_{\boldsymbol{\theta}|Y}[E(\mathbf{x}|\boldsymbol{\theta})]$.

Under KLF it is readily shown that

$$C_{\boldsymbol{\theta}|Y}[E(\mathbf{x}|\boldsymbol{\theta})] = \Lambda_p[C(\boldsymbol{\theta}|Y)]\Lambda'_p, \tag{10}$$

where Λ_p is the $p \times K$ matrix of loadings corresponding to $\mathbf{x}_p$. Letting $\pi_k(Y) \equiv p(\boldsymbol{\theta} = k|Y)$, it is tedious although again straightforward to demonstrate that, under KLP,

$$C_{\boldsymbol{\theta}|Y}[E(\mathbf{x}|\boldsymbol{\theta})] = \sum_{k<1}\sum \pi_k(Y)\pi_l(Y)\boldsymbol{\delta}_{kl}\boldsymbol{\delta}'_{kl}. \tag{11}$$

Equations (10) and (11) show that the distribution of $\boldsymbol{\theta}|Y$ is required in order to obtain an explicit expression for the elements of $\Sigma_{\mathbf{x}|Y}$ under either structure.

Considering the *K*LF case first, we write $Y|\boldsymbol{\theta} \sim N(\Lambda_Y \boldsymbol{\theta}, \sigma^2_{Y|\theta})$, where Λ_Y is the $1 \times K$ vector of loadings on Y. Then it is a routine application of multinormal theory to show that

$$\boldsymbol{\theta}|Y \sim N_K\left(\frac{Y}{\Lambda_Y \Lambda_Y' + \sigma^2_{Y|\theta}} \cdot \Lambda_Y', \sigma^2_{Y|\theta}[\Lambda_Y' \Lambda_Y + \mathrm{I}]^{-1}\right). \tag{12}$$

Substitution from Equation (12) into Equation (10) yields

$$\Sigma_{\mathbf{x}|Y} = \Psi_{\mathbf{x}|Y} + \sigma^2_{Y|\theta}[\Lambda_p[\Lambda_Y' \Lambda_Y + \mathrm{I}]^{-1} \Lambda_p']. \tag{13}$$

Inspection of Equation (13) shows that, under *K*LF, the elements of $\Sigma_{\mathbf{x}|Y}$ are constant over Y. If the factors are not orthogonal the identity matrix in Equation (13) is replaced by a nondiagonal matrix, and this matrix is also constant over Y. However, if $\sigma^2_{Y|\theta}$ varies with $\boldsymbol{\theta}$ (e.g., Hessen & Dolan, 2009), so will the conditional covariances. This point is revisited later.

Next consider the case of *K*LP. From Equations (9) and (11) it follows that

$$\Sigma_{\mathbf{x}|Y} = \Psi_{\mathbf{x}|Y} + \sum_{k<1}\sum\left(\frac{\pi_k f_{Y|\boldsymbol{\theta}=k} \cdot \pi_l f_{Y|\boldsymbol{\theta}=l}}{\left(\sum_{m=0}^{K} \pi_m f_{Y|\boldsymbol{\theta}=m}\right)^2}\right) \boldsymbol{\delta}_{kl} \boldsymbol{\delta}_{kl}' \tag{14}$$

with the explicit function obtained from requirement that the $Y|\boldsymbol{\theta} = k$ are normal. Here the $\pi_k(Y)$ are written using Bayes's rule and the $\boldsymbol{\delta}_{kl}$ are in $\mathbf{x}_p$ rather than $\mathbf{x}$. In Equation (14) is seen that off-diagonal elements $\sigma_{X_i X_j|Y}$, $i \neq j$, are weighted sums over the pairwise products of the posterior distributions of the latent classes. These are nonlinear functions of Y given that at least one class separation is not null on both Y and $\mathbf{x}_p$.

Equation (14) was analyzed by Maraun and Slaney (2005) for the case of $K = 1$. They showed that (a) the conditional covariances are single-peaked functions of Y when the within-class variances are equal, and (b) the conditional covariances can be either one- or two-peaked when the within-class variances are unequal. Their analysis was facilitated by the fact that in the two-class case, only one of the $\pi_k(Y)$ is independent. When $K > 1$ the first and second derivates (with respect to Y) of Equation 14 cannot be analytically solved for zero, so its behavior cannot be described by conventional methods. In the next section we consider some graphical examples to show that off-diagonal elements of $\Sigma_{\mathbf{x}|Y}$ are not, in general, well approximated by a constant.

In summary, this section has show that $\Sigma_{\mathbf{x}|Y}$ is constant under Gaussian factor models but nonlinear under latent profile models with Gaussian components. Note that we have not shown that these latent structures are the *only* ones with these implications for $\Sigma_{\mathbf{x}|Y}$, and indeed it has been discussed how heteroscedastic factor models can also imply nonlinear conditional covariances. We now turn to consider the usefulness of these results for selecting between KLF and KLP.

Graphical Methods

Covariance-based methods for distinguishing unidimensional linear factor models from the two-class latent profiles have been proposed by Meehl (e.g., Meehl & Yonce, 1996, Waller & Meehl, 1998). The results given in the previous section allow these methods to be extended to $K > 1$. In particular, each of the $\sigma_{X_i X_j|Y}$, $i \neq j$, can be plotted as a function of Y to visually inspect the form of the curve. Examples of the 2LP and 3LP structures were generated using Equation (14) and are shown in Figure 1. For examples of the 1PL case see Maraun and Slaney (2005). Analytic examples of factor models are not shown because their (constant) behavior is obvious from inspection of Equation (13).

Each curve in Figure 1 represents a single conditional covariance function from a different latent profile model. Functions with matching types of lines in either panel have similar weighting schemes given by the δ_{kl}. The details of the curves are summarized in the caption. It is important to note that there are

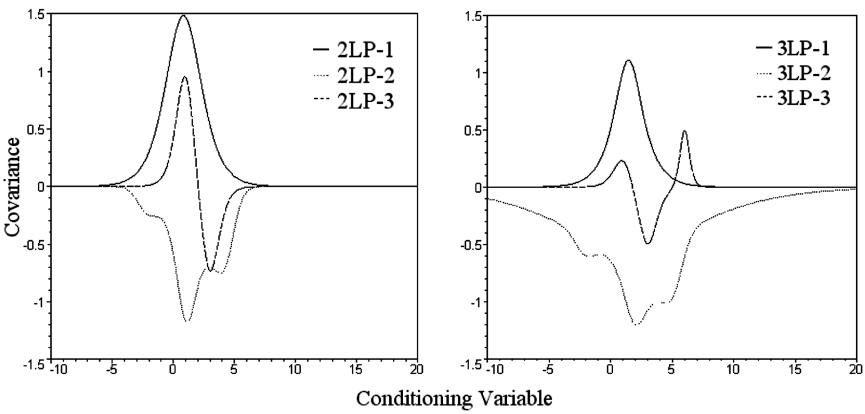

FIGURE 1 Analytic conditional covariance functions of some latent profile models. For each curve the mixture proportions are equal. On the left panel, the conditioning variables' within-class means range from 0 to 4 and on the right panel these range from 0 to 8. Within-class variances of the conditioning variables are between 1 and 2.

14 parameters to vary for the 2LP covariances, and 19 under 3LP, so the figure should not be taken to illustrate the full range of possible curves.

The two plots in Figure 1 were constructed to demonstrate that these conditional covariance functions are not well suited to distinguish the number of latent classes. In particular, this has implications for applications of MAXCOV. A single-peaked conditional covariance curve is not specific to two-class latent profile models but is also compatible with three- and four-class mixtures (see the solid lines in Figure 1). Our work with these covariance functions supports the conjecture that when the within-class variances are equal, the number of extrema is between 1 and $(K-1)$ inclusive; when the within-class variances are unequal, the number of extrema is between 1 and $2(K-1)$ inclusive. If this conjecture is correct, a single-peaked curve is possible for any value of K. Of course, a similar consideration holds under KLF: the conditional covariances are constant for all values of K. The implication is that, although these functions can be used to distinguish KLF and KLP, they cannot be used to determine the number of factors in a factor model or the number of classes in a latent profile model.

Figure 1 provides some insight at the "population level," but sample-based procedures must be employed in practice. In Figure 2, the 2LP models from Figure 1 were used for data generation and the covariance curves were plotted. The data generation protocol is described later in our simulation study, and Ruscio et al. (2006) discuss how to obtain sliding window plots of conditional

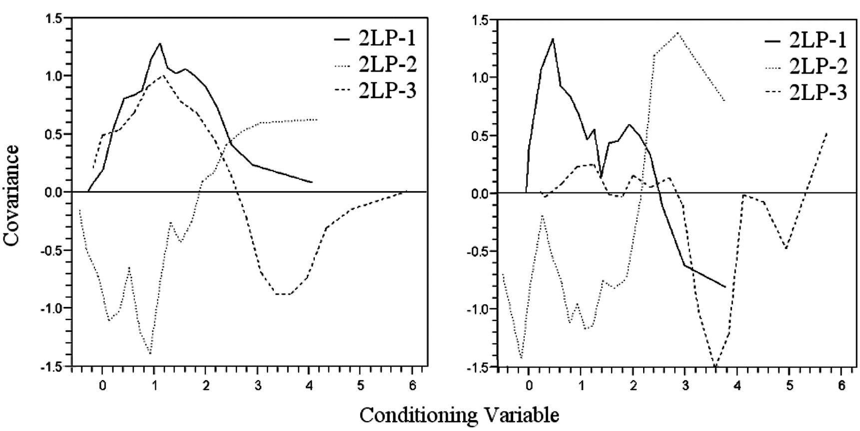

FIGURE 2 Empirical conditional covariance functions of some latent profile models. The same parameters from the 2LP curves in Figure 1 were used for data simulation. In the left panel each covariance function is obtained from a total sample size of $N = 2000$ and $i = 1, \ldots, 18$ sliding windows with a per window sample size of $n_i = 200$ and an overlap between windows of $m = 100$ observations. In the right panel $N = 300$, $n_i = 30$, and $m = 15$.

covariances. In the left panel of Figure 2 a total of 2,000 observations were used, and in the right panel 300 observations. The plots to the left reproduce the relevant properties of the analytic curves from the right panel of Figure 1 quite well (i.e., the number and location of extrema). When the sample size is decreased to 300, the shape of the plots becomes more irregular and more difficult to interpret. A sample size of 300 is the minimum sample size for MAXCOV advised by Meehl (Meehl & Yonce, 1996).

Sample-based graphical procedures also depend on their performance under factor models. Some examples for 2LF are given in Figure 3. The interpretation of Figure 3 is analogous to that of Figure 2, and the parameter arrangements are described in the caption. In both panels of Figure 3, the overall fluctuation of the covariances is notably less than that found in Figure 2. However, flatness is only a marked property of the 2LF-1 and 2LF-2 curves in the left panel. It is difficult to judge the apparent fluctuations without knowledge of the sampling variance of the covariances, and, as with all graphs, the resolution used on the axes affects the apparent shape of the curves.

Many other numerical examples are available in the taxometric literature and these have made use of more sophisticated graphical methods (see Ruscio et al., 2006). The examples given here serve to show that covariance plots can be useful for distinguishing *K*LF and *K*LP at larger sample sizes. Because the graphical approach is more difficult to interpret at moderate sample sizes, other means for selecting between *K*LF and *K*LP are desirable.

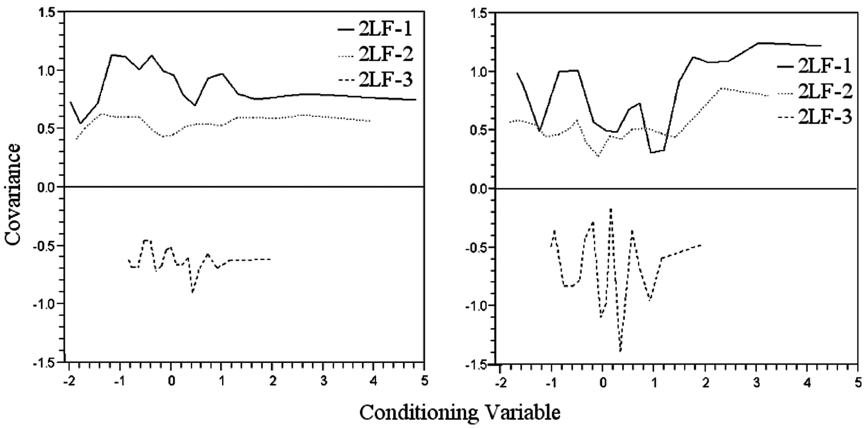

FIGURE 3 Empirical conditional covariance functions of some linear factor models. The parameter arrangements for three linear factor models were randomly selected and used for simulation. Covariance functions in the left and right panels were obtained using the sample size specifications described in Figure 2.

A CHI-SQUARE TEST FOR MODEL SELECTION

This section outlines the derivation of an inferential procedure suggested by the foregoing results. Begin by recalling from Equation (13) that, under KLF, the $\sigma_{X_i X_j|Y}$, $i \neq j$, are constant over Y. Next consider a partition of a closed interval $[a, b]$ of Y

$$a = y_o < y_1 \cdots < y_R = b. \qquad (15)$$

Let F_Y denote the cumulative distribution function of Y and write $z_r = F_Y(y_r)$ for $r = 1, \ldots, R$. Then requiring that

$$\max\{z_r - z_{r-1} | r = 1, \ldots, R\} = \min\{z_r - z_{r-1} | r = 1, \ldots, R\} \qquad (16)$$

ensures that the intervals $Y_r = [y_{r-1}, y_r]$ all have equal probability under F_Y. KLF then implies

$$E_{Y \in Y_r}[\sigma_{X_i X_j|Y}] = E_{Y \in Y_s}[\sigma_{X_i X_j|Y}] \qquad (17)$$

for any interval $[a, b]$ of Y, any choice of R, and all $r, s = 1, \ldots, R$. The main significance of Equation (17) is as follows: If two or more groups of output variables are formed by selecting on equiprobable subsets of Y, the population covariances within each group are always equal under KLF.

We now explain why Equation (17) is false under KLP, the basic idea being to let R become large for an arbitrary choice of $[a, b]$. In order for Equation (17) to be true, it must be the case that the integrand of $E_{Y \in Y_r}[\sigma_{X_i X_j|Y}]$, namely, $f_Y . \sigma_{X_i X_j|Y}$, is constant over all choices of $[a, b]$ and R. From Equation (14) we have

$$f_Y . \sigma_{X_i X_j|Y} = \sum_{k<1}\sum \left(\frac{\pi_k f_{Y|\boldsymbol{\theta}=k} \cdot \pi_l f_{Y|\boldsymbol{\theta}=l}}{\sum_{m=0}^{K} \pi_m f_{Y|\boldsymbol{\theta}=m}} \right) \delta_{ikl} \delta_{jkl}, \qquad (18)$$

where δ_{ikl} is the univariate analogue of $\boldsymbol{\delta}_{kl}$. Equation (18) is not constant over Y, which can be confirmed by taking its derivative in Y. Therefore Equation (17) is false under KLP. Roughly, this means that if we form a sufficient number of equiprobable groups by selecting on Y, we are sure to obtain at least two groups whose population covariances differ when KLP is true.

The foregoing considerations show that Equation (17) provides a means of discriminating the two structures—it is true under KLF and false under KLP. Note again that only sufficiency, not necessity, of the structures has been considered. In the next section we outline the derivation of a statistical test of Equation (17) when KLF is true and for any integer $R \in [2, \infty)$.

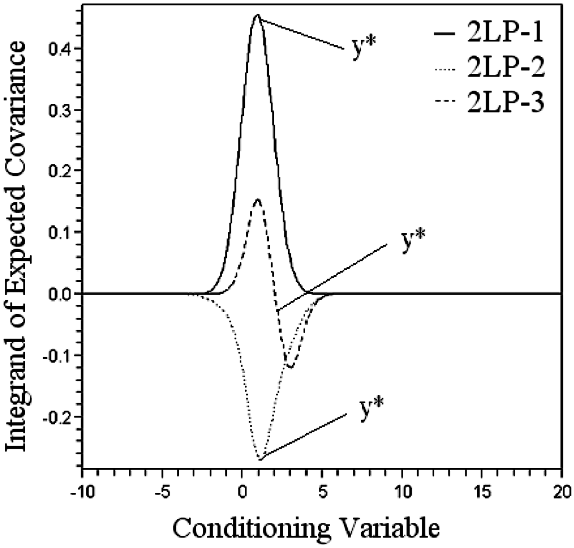

FIGURE 4 Integrand of expected conditional covariances. The same parameter arrangements from the 2LP curves in Figure 1 were used to obtain integrands of the expected conditional covariance functions. The median value of the conditioning variable is denoted on each function by y^*. Subtracting the integrals to the left of y^* from those to their right yields the following values: 2LP-1: 0.0724; 2LP-2: 0.1381; 2LP-3: 0.4271.

Before presenting the details of the proposed test, it is worthwhile to note that R does not have to be large in order for Equation (17) to be false under KLP. In particular, Figure 4 serves to illustrate that $R = 2$ is sufficient in some cases. The figure shows Equation (18) for the 2LP curves considered in Figure 1 together with the median values of Y. As described in the caption, for all three cases the median split on Y is sufficient for Equation (17) to be false. However, it is not clear whether such differences are large enough to be detected inferentially. We address this question in more detail in the examples and simulation study that follow.

Outline of the Derivation of a Test Statistic

Begin by defining the row vector

$$\mathbf{q}_{ij} = [E_{Y \in Y_1}[\boldsymbol{\sigma}_{X_i X_j|Y}] \cdots E_{Y \in Y_R}[\boldsymbol{\sigma}_{X_i X_j|Y}]] \tag{19}$$

for each pair $i \neq j$, and $i, j = 1, \ldots, p$. Then let Q denote the $(p(p-1)/2) \times R$ matrix with rows $\mathbf{q}_{ij}$ and let $\mathbf{w}$ denote an R-dimensional column vector of constants w_r such that $\mathbf{w}'\mathbf{1} = 0$. We propose to test the following null hypotheses under KLF:

$$H_o : \mathrm{Q}\mathbf{w} = \mathbf{0}. \tag{20}$$

Equation (20) is the formal statement of the null hypothesis considered in the introduction to this article.

To this end we consider the distribution of $\mathbf{x}|Y$ under KLF, which is

$$\mathbf{x}|Y \sim N_P(Y \cdot \mathbf{b}, \Sigma_{\mathbf{x}|Y}). \tag{21}$$

Here $\mathbf{b}$ is the p-vector whose ith element is given by $\sigma_{X_i|Y}/\sigma_Y^2$. Because $E(\mathbf{x}|Y)$ varies with Y, it is useful to formulate the test statistic in terms of the vector of residuals $\mathbf{e} \equiv \mathbf{x}_p - Y \cdot \mathbf{b}$. Then $\mathbf{e}|Y \sim N_P(\mathbf{0}, \Sigma_{\mathbf{x}|Y})$ and hence the $\mathbf{e}|y \in Y_r$, $r = 1, \ldots, R$, are i.i.d. multinormal.

From here a test of the hypothesis given in Equation (20) is readily obtained. The test is adapted from Seber (1984, chap. 3) and Steiger (1980). It is a quadratic form statistic based on Fisher-transformed correlations computed on random samples from $\mathbf{e}|Y \in Y_r$, $r = 1, \ldots, R$, and has an asymptotic chi-square distribution.

Let P denote the $(p(p-1)/2)$ by R matrix of Fisher-transformed correlations corresponding to Q. For $r = 1, \ldots, R$ let $\boldsymbol{\rho}_r$ denote the rth column of P, let $\hat{\boldsymbol{\rho}}_r$ denote the estimated Fisher-transformed correlations computed on n_r independent observations from $\mathbf{e}|Y \in Y_r$, and let Γ_r denote the covariance matrix of $(n_r - 3)^{1/2}\hat{\boldsymbol{\rho}}_r$. Because the $\mathbf{e}|Y \in Y_r$ are i.i.d. multivariate normal under KLF, it follows that $\Gamma_r = \Gamma$ and as the $n_r \to \infty$ (see, e.g., Steiger, 1980),

$$\hat{\mathrm{P}}\mathbf{w} \sim N_{p(p-1)/2}\left(\mathbf{0}, \sum_{r=1}^{R} \frac{w_r^2}{(n_r - 3)}\Gamma\right). \tag{22}$$

Next let A be any consistent estimator of Γ. Then, $p\ \lim_{n_r \to \infty} \mathrm{A}^{-1}\Gamma = \mathrm{I}$, and the asymptotic distributional result can be obtained using theorem 2.5.2 of Searle (1971):

$$\left(\sum_{r=1}^{R} \frac{w_r^2}{(n_r - 3)}\right)^{-1} \mathbf{w}'\hat{\mathrm{P}}'\mathrm{A}^{-1}\hat{\mathrm{P}}\mathbf{w} \sim \chi^2(p(p-1)/2). \tag{23}$$

The distributional result in Equation (23) can be used to test the null hypothesis in Equation (20). We now provide two examples of its application.

TABLE 1
Summary of Five ASVAB Subtests for $N = 500$ Respondents

	AR	*PC*	*AI*	*MK*	*AO*
M	−0.42	−0.35	−1.74	−0.11	−0.37
SD	1.05	1.01	0.93	1.03	1.03
Skew ($SE = 0.11$)	−0.35	−0.157	0.592	−0.170	0.079
Kurtosis ($SE = 0.22$)	−0.29	−0.807	0.790	−0.590	−1.016

Note. AI = Auto Information; AO = Assembling Objects; AR = Arithmetic Reasoning; MK = Mathematics Knowledge; PC = Paragraph Comprehension.

Numerical Example of the Proposed Test: Rejection of *K*LP

In this section an example from aptitude testing is used to illustrate a scenario in which *K*LP is rejected. The purpose of this example is twofold. First, we describe the steps used to compute the test statistic in Equation (23). Each step can be easily performed with available software, and a simple Fortran 95 algorithm that combines all the necessary steps is also available from the authors. Second, multiple choices of the input variable Y are considered in conjunction with multiple values of R, which serves to illustrate the consistency of the test across these parameters.

We analyzed a random sample of $N = 500$ complete responses to five subtests of the ASVAB (see Segall, 2004). The battery was administered to a total of 7,127 youths between ages 12 and 18 as part of the National Longitudinal Survey of Youth 1997.[1] The subtests employed were Arithmetic Reasoning (AR), Paragraph Comprehension (PC), Auto Information (AI), Mathematics Knowledge (MK), and Assembling Objects (AO). The data fit a unidimensional factor model reasonably well based on second-order moments (root mean square error of approximation = 0.07; 95% confidence interval = [0.04, 0.10]). Based on marginal maximium likelihood, a separate sample from the same five subtests of the ASVAB of 1997 was found to fit a factor model with heteroscedastic errors on AR and AO (D. Molenaar, Dolan, & Verhelst, 2010). Also, as reported in Table 1, the present data show mild to moderate univariate skew and kurtosis. In light of the observed violations of the assumptions of *K*LF and the relatively large sample size, the present example can be taken to illustrate the robustness of the proposed procedure.

The analysis proceeded by selecting each of the $i = 1, \ldots, 5$ subtests as an input variable Y. For each choice of Y, the following steps were followed: First the four output variates were regressed on Y in order to obtain the residuals

[1]Retrieved March 30, 2008, from http://www.nlsinfo.org/web-investigator/.

described below Equation (21); this required four univariate linear regressions for each value of i. The input variate was then ordered and the quantile of the nth ranked observation on Y was estimated as $n/N + 1$. Subsequently a series of $R = 2, \ldots, 5$ groups of residuals was formed, the maximum value of R based on sample size considerations. For each value of R, the groups were formed by selecting on adjacent, disjoint intervals of Y. The requirement of equiprobability of each group under F_Y was obtained via equality of group sample sizes, although the Rth group was required to have a smaller sample size when N/R was not an integer. The Fisher-transformed correlations were then estimated in the usual manner for each group of residuals.

Each of the foregoing procedures can be readily performed with the usual built-in utilities of most general statistical packages, whereas the following will additionally require some easily programmable matrix operations.

For each value of R, the vech of the Fisher-transformed correlation matrix was computed for each group and stored as a column of the matrix P in Equation (21). Then covariance matrix of the Fisher transforms was computed using the pooled estimates (Seber, 1984, chap. 3) and inverted. Next a normed vector of weights with alternating signs was produced. For even values of R the weights were chosen such that $|w_r| = w$ for all r and so the general form of the weight vector is $\mathbf{w} = [w, -w, w, -w, \ldots]'$. For odd values of R, the negative weights were equal to $(1 - w/w^*)$, where w is the positive weight and w^* is equal to the floor function of $R/2$. Finally, the quadratic form in Equation (21) was computed and its tail probability evaluated under the testing distribution.

The results of the analysis are presented in Table 2. It can be seen that most of the observed tail probabilities under the testing distribution are quite large. Two exceptions occur for the case of $R = 4$, yet these values of the test statistic are also not entirely out of keeping with the null distribution. The overall weight of the evidence suggests that there is no violation of the factor structure. Note that because all 20 tests reported in Table 2 are computed on the same data,

TABLE 2
Chi-Square Values Using Each ASVAB Subtest as Input Variable

	AR	*PC*	*AI*	*MK*	*AO*
$R = 2$	1.29 (.97)	9.09 (.17)	7.86 (.25)	5.79 (.45)	5.39 (.49)
$R = 3$	5.33 (.50)	2.57 (.85)	2.90 (.82)	2.70 (.84)	2.96 (.73)
$R = 4$	12.9 (.05)	10.93 (.09)	4.58 (.60)	2.37 (.88)	2.96 (.81)
$R = 5$	5.15 (.52)	9.04 (.17)	1.74 (.94)	3.11 (.79)	2.88 (.82)

Note. AI = Auto Information; AO = Assembling Objects; AR = Arithmetic Reasoning; MK = Mathematics Knowledge; PC = Paragraph Comprehension. Table entries are the observed chi-square statistic ($df = 6$) followed in parentheses by its tail-probability under Equation (23).

interpretation of error rates is not appropriate in this context, and we defer discussion of error rates until the simulation study of the following section. The present example only serves to illustrate a case where the proposed test leads to acceptance of a factor structure for a variety of combinations of R and Y when the data are known to fit 1LF reasonably well.

Numerical Example of the Proposed Test: Rejection of *K*LF

In this section we discuss data from the WLT (Piaget & Inhelder, 1969). The task has been traditionally used to discriminate among children at different phases of cognitive development. Success on the task requires the respondent to correctly indicate that the surface of still water is always horizontal regardless of the orientation of the container in which the water sits. The task is well known to produce bimodal distributions among children of different Piagetian stages of development (e.g., operational and preoperational) and to also have a strong gender effect. This example therefore serves to illustrate the performance of the test statistic when the factor would be implausible.

Our sample consisted of 285 observations from children age 8 to 12 years. Responses to the WLT were recorded as angles (in degrees) of the water level drawn by the children for bottles whose orientations ranged between 0° and 90° from upright. The present example employs the $N = 160$ complete responses of the female participants to five different bottle orientations. The fit of these data to the unidimensional factor structure was tentative but not decisively unacceptable (RMSEA = 0.08; 95% CI = [0.01, 0.15]). Descriptive statistics are presented in Table 3, and a more thorough discussion of the full data set can be found in Bringman, Buitenweg, van der Kooij, and Sierksma (2006).

The analysis followed the aforementioned procedure. Due to the relatively small sample size, the test was only conducted for $R = 2$ and 3, and the latter should be interpreted with caution. The results of the analysis are presented in Table 4, where it can be seen that the test rejects the null hypothesis in most cases. There are, however, two cases where the observed tail probability is quite

TABLE 3
Summary of Water Level Task for $N = 160$ Female Respondents

	15°	*30°*	*45°*	*60°*	*75°*
M	23.00	29.91	34.91	35.05	36.87
SD	31.80	34.77	34.91	43.09	40.63
Skew (*SE* = 0.19)	0.62	0.28	0.19	0.46	0.28
Kurtosis (*SE* = 0.38)	0.37	−0.55	−0.82	−1.66	−0.75

TABLE 4
Chi-Square Values Using Each Water Level Task Angle as Input Variable

	15°	*30°*	*45°*	*60°*	*75°*
$R = 2$	20.90 (.002)	15.70 (.015)	2.32 (.88)	10.99 (.09)	27.63 (< .001)
$R = 3$	3.21 (.78)	14.74 (.02)	17.58 (.008)	10.08 (.09)	17.63 (.007)

Note. Table entries are the observed chi-square statistic ($df = 6$) followed in parentheses by its tail-probability under Equation (23).

large, which may be attributable to a variety of reasons including the small values of R and the large sampling error due to small within-group sample size. Nonetheless, the weight of the evidence allows for the conclusion that a factor model does not fit the data even though the fit of the marginal covariance structure was not conclusive in this respect. Because a latent profile model is in keeping with theories that posit discrete stages of development, whereas the factor model is not, the conclusion to accept KLP has clear theoretical implications.

This example also illustrates that the test can produce the correct rejection of KLF when using a simple median split on the input variable. The following simulation study sheds more light on the question of error rate control for the case of $R = 2$.

Simulation Study: Error Rates of the Proposed Test Statistic When $R = 2$

As explained earlier, the intended purpose of the proposed test is to make inferences about KLF and KLP when sample size is not adequate to make reliable use of graphical methods. Therefore this section considers the performance of the test for the minimal value of R (i.e., the maximal value of n_r) and for small ($N = 30$) to moderate ($N = 300$) sample sizes. If the power of the test is adequate for the case of a median split then there is little reason to consider larger values of R. The study leads to the favorable conclusion that the test can be reliably applied under these circumstances.

Data were generated and analyzed using Fortran 95. To generate data under the KLF structure, factor loadings and specific variances were assigned pseudorandom values in the ranges [0, 1] and ± 1.5, respectively. In order to generate data under KLP the conditional mean vectors were assigned pseudorandom numbers between ± 2 and the conditional variances were obtained from the diagonal of the inner product of a $J \times J$ matrix of pseudorandom numbers between ± 1. The maximum value of the within-class variances is therefore equal to the number of observed variables, and so it was always possible for class separations to be less than the within-class variances. The π_k were also assigned pseudorandom numbers, and the observed mixture was given by comparing realizations of a pseudorandom $U(0, 1)$ variable with the π_k. The analysis used one randomly selected input indicator to conduct one test for each data set.

The four factors addressed in this simulation study were (a) model type (KLF or KLP), (b) number of latent variate/components ($K = 1, \ldots, 4$), (c) number of covariance generating variables ($p = 2, \ldots, 7$), and (d) sample size ($N = 30, 50, 100, 200, 300$). For each combination of the four factors, 5,000 independent trials were recorded, yielding a total of 1,200,000 data simulations. The proportion of times out of 5,000 that the observed chi-square statistic led to a rejection of the null hypothesis at the .01 and .05 alpha levels was recorded. Under KLF this proportion is the empirical alpha level of the test, and under KLP it is the empirical power of the test. Together these values provide a comprehensive description of the test's error rate control under the two structures. Tables 5 and 6 report the observed proportions for KLF at the .01 and .05 alpha levels, respectively. Tables 7 and 8 present this information for KLP.

Considering Tables 5 and 6 first, the empirical alpha levels correspond quite well to the theoretical probability of Type I error when only a single covariance is considered (i.e., when $p = 2$). This is the case even if $N = 30$. Rather surprisingly, when the number of covariance-generating variables increases, the empirical alpha levels decrease. This is likely attributable to the weighting coefficients $(n_r - 3)^{-1}$, which have been reported to be inaccurate for high dimensional quadratic form statistics based on Fisher-transformed correlations, although improved weighting schemes have not been published (see Fouladi & Steiger, 1999). In the present context the observed departure from the theoretical probabilities implies that the number of Type I errors is less than expected. This may suggest the use of a more liberal alpha level in applications with large values of p. In general, Tables 5 and 6 provide an initial indication that the proposed test performs satisfactorily under the null hypothesis: it does not reject the linear factor model more often than it should.

Turning to Tables 7 and 8, the most notable trend is due to sample size. For $N = 30$, the empirical power is inadequate. However, by the time the total sample size has increased to 300, the average power has reached more acceptable levels (approximately 0.83 for $\alpha = .05$ and 0.75 for $\alpha = .01$). It is also the case that the average power to detect a 1LP structure at a given sample size is generally less than that when $K > 1$.

An interaction between sample size and the number of covariance-generating variables is also apparent. As N increases, the value of p for which power is a maximum also increases. For example, at $N = 100$, the average empirical power increases from $p = 2$ to $p = 4$ and then decreases again for $p > 4$. At lower values of N the number of output variables that maximizes power is less than four. Based on Tables 7 and 8, the best testing situation for correctly

TABLE 5
Empirical Probabilities for Linear Factor Structure at .01 Alpha Level

Sample Size	*Factors*	p = *2*	p = *3*	*Output* p = *4*	*Variables* p = *5*	p = *6*	p = *7*	*Average*
$N = 30$	$K = 1$	.0142	.0096	.0054	.0038	.0018	.0012	.0060
	$K = 2$	.0104	.0072	.0040	.0014	.0004	.0004	.0040
	$K = 3$	.0120	.0108	.0046	.0024	.0006	.0002	.0051
	$K = 4$	.0152	.0070	.0036	.0012	.0004	.0000	.0046
	Average	.0130	.0087	.0044	.0022	.0008	.0005	.0049
$N = 50$	$K = 1$	.0118	.0076	.0036	.0028	.0014	.0008	.0047
	$K = 2$	.0112	.0086	.0042	.0024	.0006	.0002	.0045
	$K = 3$	.0112	.0078	.0036	.0012	.0000	.0004	.0040
	$K = 4$	.0094	.0076	.0052	.0016	.0002	.0002	.0040
	Average	.0109	.0079	.0041	.0020	.0005	.0004	.0043
$N = 100$	$K = 1$	.0096	.0108	.0042	.0022	.0018	.0008	.0049
	$K = 2$	.0112	.0064	.0028	.0016	.0008	.0004	.0039
	$K = 3$	.0094	.0090	.0036	.0020	.0008	.0000	.0041
	$K = 4$	.0090	.0056	.0040	.0008	.0006	.0002	.0034
	Average	.0098	.0080	.0036	.0017	.0010	.0003	.0041
$N = 200$	$K = 1$	.0102	.0080	.0056	.0030	.0024	.0012	.0051
	$K = 2$	.0104	.0066	.0040	.0018	.0000	.0002	.0038
	$K = 3$	.0120	.0048	.0024	.0002	.0000	.0000	.0032
	$K = 4$	.0110	.0054	.0030	.0010	.0002	.0000	.0034
	Average	.0109	.0062	.0038	.0015	.0007	.0004	.0039
$N = 300$	$K = 1$	.0112	.0080	.0032	.0034	.0024	.0024	.0051
	$K = 2$	.0108	.0040	.0030	.0006	.0002	.0000	.0031
	$K = 3$	.0090	.0078	.0022	.0006	.0006	.0000	.0034
	$K = 4$	.0102	.0072	.0024	.0014	.0004	.0000	.0036
	Average	.0103	.0067	.0027	.0015	.0009	.0006	.0038

Note. Each table entry is the proportion out of 5,000 that a data simulation led to rejection of the null hypothesis in Equation (20) at the specified alpha level.

retaining a latent profile model would be one where $N = 300$ and the number of output variables used is between four and six. From inspection of Tables 5 and 6, this situation is also favorable for Type I error control.

In summary, conclusions based on Tables 5 through 8 include the following: The minimal sample size for which a power of 0.80 may be obtained when using an alpha level of .05 is $N = 200$, or 100 observations per group. At this sample size, Type I and II error rates are well controlled when using between four and six output variables. Therefore the test is most useful with at least 200 observations on five or more indicators. Further discussion is provided in the next section.

TABLE 6
Empirical Probabilities for Linear Factor Structure at .05 Alpha Level

				Output	*Variables*			
Sample Size	*Factors*	p = *2*	p = *3*	p = *4*	p = *5*	p = *6*	p = *7*	*Average*
$N = 30$	$K = 1$	.0526	.0368	.0264	.0152	.0114	.0084	.0251
	$K = 2$	.0468	.0348	.0186	.0076	.0044	.0010	.0189
	$K = 3$	.0530	.0358	.0186	.0082	.0040	.0014	.0202
	$K = 4$	.0542	.0346	.0186	.0088	.0030	.0006	.0200
	Average	.0517	.0355	.0206	.0099	.0057	.0029	.0210
$N = 50$	$K = 1$	.0476	.0408	.0222	.0174	.0108	.0060	.0241
	$K = 2$	.0544	.0382	.0214	.0102	.0030	.0012	.0214
	$K = 3$	.0498	.0356	.0174	.0092	.0036	.0018	.0196
	$K = 4$	.0484	.0342	.0180	.0078	.0038	.0006	.0188
	Average	.0501	.0372	.0197	.0112	.0053	.0024	.0210
$N = 100$	$K = 1$	.0504	.0424	.0294	.0164	.0138	.0070	.0266
	$K = 2$	.0562	.0306	.0182	.0090	.0052	.0022	.0202
	$K = 3$	.0486	.0346	.0178	.0092	.0026	.0004	.0189
	$K = 4$	.0474	.0328	.0204	.0060	.0026	.0004	.0183
	Average	.0507	.0351	.0215	.0102	.0060	.0025	.0210
$N = 200$	$K = 1$	.0558	.0434	.0298	.0164	.0130	.0096	.0280
	$K = 2$	.0520	.0348	.0206	.0126	.0026	.0012	.0206
	$K = 3$	.0536	.0342	.0150	.0060	.0030	.0006	.0187
	$K = 4$	.0526	.0320	.0164	.0066	.0018	.0002	.0183
	Average	.0535	.0361	.0204	.0104	.0051	.0029	.0214
$N = 300$	$K = 1$	.0472	.0372	.0226	.0212	.0118	.0102	.0250
	$K = 2$	.0470	.0288	.0178	.0094	.0058	.0018	.0184
	$K = 3$	.0492	.0338	.0158	.0070	.0032	.0006	.0183
	$K = 4$	.0496	.0320	.0156	.0062	.0034	.0006	.0179
	Average	.0483	.0329	.0179	.0110	.0060	.0033	.0199

Note. Each table entry is the proportion out of 5,000 that a data simulation led to rejection of the null hypothesis in Equation (20) at the specified alpha level.

DISCUSSION

This article has provided a method for the empirical discrimination of Gaussian linear factor models and latent profile models with Gaussian components. Graphical approaches were shown to be useful with large sample sizes, but the interpretation of such graphs in taxometric applications should be modified. In particular, a single-peaked conditional covariance function is not necessarily indicative of two classes. In place of graphical methods, we also developed an asymptotic chi-square test based on Fisher-transformed correlations. This test

TABLE 7
Empirical Probabilities for Latent Profile Structure at .01 Alpha Level

Sample Size	*Latent Classes*	p = *2*	p = *3*	*Output* p = *4*	*Variables* p = *5*	p = *6*	p = *7*	*Average*
$N = 30$	$K = 1$	.1432	.1354	.1006	.0666	.0450	.0300	.0868
	$K = 2$	.1504	.1442	.1148	.0836	.0486	.0334	.0958
	$K = 3$	.1422	.1294	.1060	.0806	.0506	.0336	.0904
	$K = 4$	.1346	.1324	.0970	.0688	.0566	.0316	.0868
	Average	.1426	.1353	.1046	.0749	.0502	.0322	.0900
$N = 50$	$K = 1$	.2226	.2416	.2044	.1686	.1258	.0874	.1751
	$K = 2$	.2140	.2504	.2264	.1838	.1406	.1040	.1865
	$K = 3$	.2108	.2436	.2200	.1874	.1464	.0970	.1842
	$K = 4$	.2040	.2224	.2198	.1814	.1444	.1014	.1789
	Average	.2129	.2395	.2176	.1803	.1393	.0975	.1812
$N = 100$	$K = 1$	.3392	.4254	.4420	.4044	.3530	.2998	.3773
	$K = 2$	.3368	.4326	.4624	.4470	.4034	.3488	.4052
	$K = 3$	.3212	.4358	.4636	.4392	.4010	.3576	.4031
	$K = 4$	.3118	.4168	.4564	.4316	.3854	.3572	.3932
	Average	.3273	.4277	.4561	.4306	.3857	.3408	.3947
$N = 200$	$K = 1$	.4800	.6310	.6630	.6592	.6368	.5902	.6100
	$K = 2$	.4724	.6436	.7096	.7188	.7166	.6962	.6595
	$K = 3$	.4490	.6318	.6988	.7264	.7240	.7024	.6554
	$K = 4$	.4478	.6150	.7098	.7284	.7114	.6902	.6504
	Average	.4623	.6303	.6953	.7082	.6972	.6697	.6439
$N = 300$	$K = 1$	.5456	.7228	.7600	.7806	.7624	.7414	.7188
	$K = 2$	.5460	.7572	.8220	.8388	.8350	.8170	.7693
	$K = 3$	.5212	.7440	.8238	.8432	.8430	.8360	.7685
	$K = 4$	.5196	.7276	.8142	.8416	.8458	.8276	.7627
	Average	.5331	.7379	.8050	.8260	.8215	.8055	.7548

Note. Each table entry is the proportion out of 5,000 that a data simulation led to rejection of the null hypothesis in Equation (20) at the specified alpha level.

requires both a moderate sample size ($N \geq 200$) and a moderate number of covariance-generating variables ($4 \leq p \leq 6$) for good error rate control. A rather surprising finding was that the observed Type I error rate was less than expected under the theoretical null distribution when $p > 2$. Research along the lines of Fouladi and Steiger (1999) would be beneficial to shed light on this phenomenon, but in the present application it is rather unproblematic. More pressingly, it would be useful to devise an optimal strategy for selecting input and output indicators and for interpreting discrepant results when multiple tests are conducted on a single data set.

TABLE 8
Empirical Probabilities for Latent Profile Structure at .05 Alpha Level

Sample Size	*Latent Classes*	p = *2*	p = *3*	*Output* p = *4*	*Variables* p = *5*	p = *6*	p = *7*	*Average*
$N = 30$	$K = 1$	.2502	.2554	.2012	.1544	.1074	.0714	.1733
	$K = 2$	.2590	.2714	.2290	.1800	.1196	.0898	.1915
	$K = 3$	.2442	.2546	.2184	.1790	.1294	.0890	.1858
	$K = 4$	.2414	.2552	.2114	.1744	.1326	.0870	.1837
	Average	.2487	.2591	.2150	.1720	.1223	.0843	.1836
$N = 50$	$K = 1$	.3404	.3788	.3418	.2874	.2294	.1720	.2916
	$K = 2$	.3294	.3952	.3668	.3160	.2632	.2032	.3123
	$K = 3$	.3326	.3860	.3650	.3304	.2694	.2150	.3164
	$K = 4$	.3216	.3678	.3674	.3266	.2640	.2128	.3100
	Average	.3310	.3820	.3602	.3151	.2565	.2007	.3076
$N = 100$	$K = 1$	.4598	.5596	.5768	.5360	.4868	.4280	.5078
	$K = 2$	.4602	.5770	.6138	.5942	.5538	.4930	.5487
	$K = 3$	.4428	.5722	.6098	.5924	.5526	.5106	.5467
	$K = 4$	.4358	.5620	.5972	.5872	.5466	.5216	.5417
	Average	.4496	.5677	.5994	.5774	.5350	.4883	.5362
$N = 200$	$K = 1$	.5826	.7326	.7616	.7622	.7308	.6860	.7093
	$K = 2$	.5874	.7498	.8018	.8186	.8126	.7880	.7597
	$K = 3$	.5562	.7402	.8060	.8230	.8096	.7934	.7547
	$K = 4$	.5670	.7342	.8122	.8256	.8120	.7906	.7569
	Average	.5733	.7392	.7954	.8074	.7913	.7645	.7452
$N = 300$	$K = 1$	.6440	.8036	.8332	.8462	.8232	.8116	.7936
	$K = 2$	.6462	.8324	.8866	.8962	.8882	.8730	.8371
	$K = 3$	.6228	.8254	.8866	.8998	.8954	.8892	.8365
	$K = 4$	.6194	.8188	.8816	.9008	.9008	.8816	.8338
	Average	.6331	.8201	.8720	.8857	.8769	.8638	.8253

Note. Each table entry is the proportion out of 5,000 that a data simulation led to rejection of the null hypothesis in Equation (20) at the specified alpha level.

It is important to recall that the hypotheses on which the test is premised are only implied by the sufficiency, not the necessity, of the two latent structures considered in this article. This means that, although the test can be used to reject either a factor structure or a latent profile structure, it can not lead to a decisive conclusion in favor of either. Thus it would be useful to conduct further research on the functional form of conditional covariances under other alternative models of interest and devise means of making inferences about these functions. In particular, it would be useful to consider the conditional covariances of factor mixture models.

In summary, this article has demonstrated that the choice between continuous and categorical latent variables is not merely a matter of theoretical preference but also has empirical, testable implications. By deriving such implications, latent variable methodology can be brought to bear on problems of theory formation in domains where the distinction between categories and continua is of importance.

ACKNOWLEDGMENT

This research was supported by the doctoral fellowship program of the Social Science and Humanities Research Council of Canada.

REFERENCES

Bartholomew, D. J. (1987). *Latent variable models and factor analysis* (1st ed.). London: Arnold.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London: Arnold.

Bauer, D. J. (2005). The role of nonlinear factor-to-indicator relationships in tests of measurement equivalence. *Psychological Methods, 10,* 305–316.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promise opportunities. *Psychological Methods, 9,* 3–29.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bringman, L., Buitenweg, J., van der Kooij, C., & Sierksma, J. (2006). *Detecting bimodality using a multivariate mixture model: Switching and hysteresis on the horizontality task.* Unpublished manuscript, University of Amsterdam, Amsterdam, The Netherlands.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review, 112,* 129–158.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman-Hall.

Fouladi, R. T., & Steiger, J. H. (1999). Tests of an identity correlation structure. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 168–196). Thousand Oaks, CA: Sage.

Frühwirth-Schnatter, M. (2006). *Finite mixture and Markov switching models.* New York: Springer.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.

Hessen, D. J., & Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology, 62,* 57–77.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14,* 1523–1543.

Joreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443–482.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* New York: Houghton Mifflin.

Loeve, M. (1963). *Probability theory* (3rd ed.). Princeton, NJ: Van Nostrand.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41,* 499–532.

Lubke, G., & Neale, M. C. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research, 43,* 592–620.

Maraun, M. D., & Slaney, K. (2005). An analysis of Meehl's MAXCOV-HITMAX procedure for the case of continuous indicators. *Multivariate Behavioral Research, 40,* 489–518.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monographs, 15.*

Meehl, P. E., & Yonce, L. (1996). Taxometric analysis II: Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV PROCEDURE) [Monograph supplement]. *Psychological Reports,* 1091–1227.

Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modeling non-normality with the one-factor model. *British Journal of Mathematical and Statistical Psychology, 63,* 293–317.

Molenaar, P. C. M., & von Eye, A. (1994). On the arbitrary nature of latent variables. In W. van der Linden & R. K. Hambleton (Eds.), *Latent variable analysis* (pp. 226–242). New York: Springer-Verlag.

Muthén, B. (2007). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child.* New York: Basic Books.

Ruscio, J., Haslam, N., & Ruscio, A. Y. (2006). *Introduction to the taxometric method: A practical guide.* Mahwah, NJ: Erlbaum.

Searle, S. R. (1971). *Linear models*. New York: Wiley.

Seber, G. A. F. (1984). *Multivariable observations.* New York: Wiley.

Segall, D. O. (2004). *Development and evaluation of the 1997 ASVAB score scale* (Tech. Rep. No. 2004-002). Seaside, CA: Defense Manpower Data Center.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research, 15,* 335–352.

Steinley, D., & McDonald, R. P. (2007). Examining factor score distributions to determine the nature of latent spaces. *Multivariate Behavioral Research, 42,* 133–156.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class models* (pp. 89–106). Cambridge, UK: Cambridge University Press.

Waller, N. G., & Meehl, P. E. (1998). *Multivariable taxometric procedures: Distinguishing types from continua.* Thousand Oaks, CA: Sage.