5

AN EXTENDED STUDY INTO THE RELATIONSHIP BETWEEN CORRESPONDENCE ANALYSIS AND LATENT CLASS ANALYSIS

*Peter G.M. van der Heijden**
Zvi Gilula†
L. Andries van der Ark‡

Researchers dealing with frequency data today can choose from a vast range of methods, descriptive and inferential. Two such well-known and useful methods are correspondence analysis and latent class analysis. Although these two methods were initially used for different research objectives, they are mathematically related to each other. Relations between these methods, however, have only been reported in the literature regarding the bivariate case. In this paper, we extend the study of such relations to the multivariate case. In particular the multivariate X latent class model is shown to imply the (relatively new) joint multivariate correspondence model with X − 1 positive eigenvalues. Such relations allow the underlying methods to be treated as variants of the same conceptual idea, providing some new meaningful aspects, which may help researchers better interpret the findings of their investigation.

1. INTRODUCTION

Researchers in many disciplines, particularly in the behavioral, marketing, and medical sciences, frequently encounter the need to analyze multivar-

*Utrecht University
†Hebrew University of Jerusalem
‡Tilburg University

iate relationships among variables that are nonquantitative (categorical). The abundance of categorical data analytic methods available today to such researchers is impressive. In the last 15 years, efforts have been made to study the possible relationships between such methods. The aim of such efforts is basically a desire to derive a common conceptual frame within which many of these methods can be unified. The advantage of having a unified frame is to provide additional insight into the interpretation of empirical findings. In this paper we show that two well-known methods for analyzing multivariate categorical variables, *correspondence analysis* (CA) and *latent class analysis* (LCA) share a common methodological frame, which is detailed later.

A major problem concerning categorical variables is the lack of a natural scale. Scaling categorical variables has been a scientific challenge for many years, as is evident from the wealth of literature on this subject. The common frame for most scaling methods is known today as *optimal scaling*. Nishisato (1980) and Gifi (1990) provide an introduction to these methods, and a recent state-of-the-art review of the unified frame for scaling methods is found in Michailidis and de Leeuw (1998). CA is a very familiar representative of the optimal scaling family. Officially introduced as a descriptive method by the French school of data analysis founded by Benzécri (Benzécri et al. 1973), CA has become one of the leading data-descriptive methods. Many widely used computer routines and software packages contain such analysis. Most of the marketing research courses in business schools contain CA in their curriculum. A good account of CA is given by Greenacre (1984). Goodman (1985) and Gilula and Haberman (1986, 1988) report the derivation of the relevant theory making CA a family of estimable and statistically testable models.

LCA was officially started by Lazarsfeld (1950) as a method of indirectly measuring unobservable variables, called *latent* variables. Starting from the conjecture that latent variables are highly correlated with certain measurable variables (called *manifest* variables), the latent class theory provides tools of measuring the latent variables through certain patterns of association between the manifest variables. A comprehensive introduction to this analysis is given by Lazarsfeld and Henry (1968), and inferential aspects of it are developed by Goodman (1974a,b), Gilula (1979) and Clogg (1981) among others.

In the bivariate case, both CA and LCA utilize the matrix form of the joint distribution of the manifest variables, and both methods use optimization criteria based on reduced rank joint distributions. It is therefore natural that relations between these two methods were subject to wide investigation (e.g., Gilula 1983, 1984; Goodman, 1985, 1987; de Leeuw and van der Heijden, 1991).

Multivariate versions of both CA and LCA have been developed. In particular *multiple* correspondence analysis (MCA) has become quite popular. Greenacre (1988) has developed a finer multivariate version of CA called *joint* correspondence analysis (JCA), and showed some advantages of JCA over MCA.

The association between multivariate LCA and JCA has not been investigated yet. This is the concern of this paper. In particular we show that latent class models having X latent classes imply joint correspondence models with $X - 1$ positive eigenvalues. We discuss ways of rescaling parameters of LCA models, graphical representations that are implied from such rescaling, and compare these graphical displays with those of JCA. The findings of this study together with their interpretation provide the already popular CA with added interpretative benefit, and applicability. Although this is *not* another review paper comparing the two underlying methods, we first start with the well-known bivariate case for expository purposes. We then report the main result, and exemplify its importance by analyzing two empirical data sets.

2. THE BIVARIATE CASE

In this section we will first define simple CA and LCA for two-way contingency tables. Then we will describe the relation, reviewing the results summarized by de Leeuw and van der Heijden (1991).

2.1. *Simple Correspondence Analysis*

We start with simple CA of an $I \times J$ probability matrix $\mathbf{\Pi}$, with a row variable A with I categories indexed by i and a column variable B with and J categories indexed by j. The elements are π_{ij} $(i = 1, \ldots, I; j = 1, \ldots, J)$, where $\pi_{ij} \geq 0$ and $\sum_i \sum_j \pi_{ij} = 1$. We denote the margins by $\sum_j \pi_{ij} = \pi_{i+}$ and $\sum_i \pi_{ij} = \pi_{+j}$. The decomposition by CA has the following form:

$$\pi_{ij} = \pi_{i+} \pi_{+j} \left(1 + \sum_{m=1}^{M-1} \lambda_m \rho_{im} \gamma_{jm} \right) , \tag{1}$$

with $M \leq \min(I,J)$. Model (1) is restricted by

$$\lambda_m > 0 \ ,$$

$$\sum_i \pi_{i+} \rho_{im} = \sum_j \pi_{+j} \gamma_{jm} = 0 \ ,$$

$$\sum_i \pi_{i+} \rho_{im} \rho_{im'} = \sum_j \pi_{+j} \gamma_{jm} \gamma_{jm'} = \delta^{mm'} \ , \qquad (2)$$

where $\delta^{mm'}$ is Kronecker delta. Greenacre (1988) discusses three interpretations of CA. We discuss here briefly the interpretation of CA as a model providing a reduced rank decomposition of a probability matrix. The other two interpretations focus on CA as a tool that provides graphic representations of a probability matrix (see Benzécri et al. 1973; Gifi 1990; Goodman 1991; Gower and Hand 1996), and on the interpretation on CA as a tool for canonical correlation analysis of categorical data (see Kendall and Stuart 1979). For more details and other interpretations, we refer to overviews such as those by Nishisato (1980), Greenacre (1984), and Gifi (1990).

The interpretation that focuses on the reduced rank emphasizes that (1) gives a reduced rank decomposition of $\mathbf{\Pi}$, in the sense that model (1) defines a matrix of rank M. When $M = \min(I,J)$, then $\mathbf{\Pi}$ has full rank, and the model is saturated. When $M = 1$, (1) is the independence model. Assume that $\mathbf{\Pi}$ has rank M, where $M \leq \min(I,J)$. Then $\mathbf{\Pi}$ can always be decomposed by CA using $M - 1$ sets of parameters indexed by m, $m = 1, \ldots, M - 1$. (For a proof, see de Leeuw and van der Heijden 1991.)

This reduced rank decomposition is closely related to the well-known Pearson chi-square statistic X^2 for testing independence in a two-way contingency table. The relation between (1) and X^2 is through the parameters λ_m in the following way:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (\pi_{ij} - \pi_{i+} \pi_{+j})^2 \Big/ \pi_{i+} \pi_{+j} = \sum_{m=1}^{M-1} \lambda_m^2 \ . \qquad (3)$$

Now consider a two-way contingency table $\mathbf{N}$ with elements n_{ij}, where $n = \sum_i \sum_j n_{ij}$. We derive relative frequencies p_{ij} by $p_{ij} = n_{ij}/n$. The marginal relative frequencies are denoted as p_{i+} and p_{+j}, respectively. The relative frequencies form a probability matrix that can be decomposed with (1). Notice that estimates of expected probabilities under independence are equal to $p_{i+} p_{+j}$, and therefore the left side of equation (3) is then equal to the coefficient of contingency, which is equal to X^2/n. The

quantity $\sum_m^{M-1} \lambda_m^2$ is called the total inertia of a matrix. Equation (3) shows that (1) provides a decomposition of the total inertia into $M - 1$ dimensions.

Model (1) is usually estimated by weighted least squares using a generalized singular value decomposition (for example, see Greenacre 1984), where the parameters λ_m are the singular values. Computer programs are widely available, for example, in software packages such as SPSS, SAS, and BMDP. More recently maximum-likelihood estimation procedures have become available (see Goodman 1985; Gilula and Haberman 1986).

2.2. *The Bivariate Latent Class Model*

The latent class model (LCM) assumes the existence of a categorical latent variable, Z, having categories indexed by x, $x = 1, \ldots, X$. This latent variable explains the relation between manifest categorical variables in the sense that, given the category of the latent variable, the manifest variables are independent. For the bivariate case, the model for the two-way probability matrix $\mathbf{\Pi}$ is

$$\pi_{ij} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} \qquad (4)$$

with all parameters nonnegative and

$$\sum_{x=1}^{X} \pi_x^Z = 1, \ \sum_{i=1}^{I} \pi_{ix}^{\bar{A}Z} = 1, \ \sum_{j=1}^{J} \pi_{jx}^{\bar{B}Z} = 1 \ .$$

Here π_x^Z is the class size—i.e., the probability that an observation falls in latent class x, and $\pi_{ix}^{\bar{A}Z}$ (or $\pi_{jx}^{\bar{B}Z}$) are conditional probabilities indicating the probability of falling in category i (or j) given that the observation falls into latent class x.

Just like CA, model (4) can also be interpreted as a reduced rank decomposition of a probability matrix. Model (4) defines a matrix of rank X. When $X = \min(I,J)$, the matrix $\mathbf{\Pi}$ is of full rank, and the model is saturated. When $X = 1$, (4) reduces to the independence model. When $1 < X < \min(I,J)$, (4) defines a matrix having a reduced rank.

In the bivariate case the LCM can be rewritten as

$$\pi_{ij}/\pi_{i+} = \sum_{x=1}^{X} \pi_{ix}^{A\bar{Z}} \pi_{jx}^{\bar{B}Z} \qquad (5)$$

where

$$\pi_{ix}^{A\bar{Z}} = \frac{\pi_x^Z \pi_{ix}^{\bar{A}Z}}{\sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z}} = \frac{\pi_x^Z \pi_{ix}^{\bar{A}Z}}{\pi_{i+}} \tag{6}$$

(Goodman 1974a), which is also known in the social sciences as the latent budget model (for an overview, see van der Heijden, Mooijaart and de Leeuw 1992; van der Ark 1999) and in geology as the end member model (Renner 1993).

The LCM is usually estimated by maximum likelihood (cf. Goodman 1974a).

2.3. *Relations*

We will first show how models (1) and (4) are related (cf. Gilula 1979, 1983; de Leeuw and van der Heijden 1991; van der Ark, van der Heijden and Sikkel 1999). Then we will discuss the implications for data analysis.

First we note that both models provide rank decompositions of the matrix $\mathbf{\Pi}$. We also note that the rank decomposition is provided in the LCM by nonnegative parameters, whereas the parameters in CA are both positive and negative.

Let the rank of a matrix $\mathbf{\Pi}$ be denoted by R. We already noted that a matrix of rank R can always be decomposed by CA with $M = R$. In terms of the rank of a matrix $\mathbf{\Pi}$ we have the following four situations.

1. When $R = 1$, the matrix $\mathbf{\Pi}$ can be decomposed by CA with $M = 1$ and by LCA with $X = 1$, since then both CA as well as LCA are equivalent to the independence model.
2. When $R = \min(I,J)$, the matrix $\mathbf{\Pi}$ can always be decomposed both by CA with $M = R$ as well as by LCA with $X = R$, since both models are equivalent to the saturated model.
3. When $1 < R < \min(I,J)$, the matrix $\mathbf{\Pi}$ can always be decomposed by CA with $M = R$ but not always by LCA with $X = R$, since the decomposition provided by LCA is less general than that of CA because of the nonnegativity restrictions on the parameters. It follows that, if LCA of rank $X = R$ is true (i.e., a matrix $\mathbf{\Pi}$ can be decomposed by LCA with $1 < X < \min(I,J)$ latent classes), then CA of rank $M = R$ is true (i.e., a matrix $\mathbf{\Pi}$ can also be decomposed by CA with $M = R$), but the reverse does not hold. There are, therefore, matrices of reduced rank R that can be decomposed by CA with $M = R$ but not by LCA with $X = R$. It follows that the models are not equivalent.
4. When $R = 2$, CA implies the LCM (see de Leeuw and van der Heijden 1991, for a proof). Therefore, for rank 2 the models are equivalent.

For data analytic situations, this implies that if both models are estimated using the same criterion—for example, maximum likelihood—then the fitted values (estimates of expected probabilities) of CA are equal to those of LCA if $M = X = 2$. If $2 < X < \min(I,J)$ it often turns out that fitted values of both models are the same, but this is not necessarily the case. (For more details see van der Ark et al. 1999.)

Assume for the moment a matrix for which the fitted values of CA are equal to those of LCA. Then the CA-parameters are related to the LCA-parameters. This relation is most clearly seen from the LCA parameters rescaled as in equation (6). First a set of rescaled parameters $\pi_{ix}^{A\bar{Z}}$ is obtained from the original parameters $\pi_{ix}^{\bar{A}Z}$, and in a comparable way a set of rescaled parameters $\pi_{jx}^{B\bar{Z}}$ is obtained from the original parameters $\pi_{jx}^{\bar{B}Z}$. Using $\pi_{ix}^{\bar{A}Z} = \pi_{i+}\pi_{ix}^{A\bar{Z}}(\pi_x^Z)^{-1}$ from equation (6), we can now rewrite the LCA model in (4) as

$$\pi_{ij} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} = \pi_{i+}\pi_{+j} \sum_{x=1}^{X} \pi_{ix}^{A\bar{Z}} \pi_{jx}^{B\bar{Z}} (\pi_x^Z)^{-1} \tag{7}$$

Comparing (6) with (1) makes the relation between the rescaled LCA parameters and the CA parameters evident:

$$\left(1 + \sum_{m=1}^{R-1} \lambda_m \rho_{im} \gamma_{jm}\right) = \sum_{x=1}^{R} \pi_{ix}^{A\bar{Z}} \pi_{jx}^{B\bar{Z}} (\pi_x^Z)^{-1} . \tag{8}$$

This shows that, if we collect the CA parameters ρ_{im} in a $I \times R$ matrix $\mathbf{P}$ where the first column is a unit vector, and γ_{jm} in a $J \times R$ matrix $\mathbf{\Gamma}$ where the first column is a unit vector (the unit vector corresponds to the '1' in [1]), and if we collect the rescaled LCA parameters $\pi_{ix}^{A\bar{Z}}$ in a $I \times R$ matrix $\mathbf{\Pi}_i$ and the rescaled LCA parameters $\pi_{jx}^{B\bar{Z}}$ in a $J \times R$ matrix $\mathbf{\Pi}_j$, then

$$\mathbf{P} = \mathbf{\Pi}_i \mathbf{S} \tag{9}$$

and

$$\mathbf{\Gamma} = \mathbf{\Pi}_j \mathbf{T}, \tag{10}$$

where **S** and **T** are $R \times R$ transformation matrices. This shows that linear combinations of the rescaled LCA parameters lead to the CA parameters and the other way around. But we repeat that equation (8) is correct only if for some matrix the fitted values of CA are equal to the fitted values of LCA.

Of course, this result becomes more interesting if the rescaled parameters $\pi_{ix}^{A\bar{Z}}$ and $\pi_{jx}^{B\bar{Z}}$ give us any insight into the latent class model (4). Fortunately they do, as we will explain now. The presentation of the latent class model in (4) is only one way to represent the latent class model, and there exist alternatives. A well-known alternative representation of the latent class model is as a loglinear model for the manifest variables A and B and the latent variable Z. In this presentation the latent variable Z is related to variable A and to variable B, and A and B are not directly related: conditional on X, the variables A and B are independent. Let us define the probabilities in this joint table of variables A, B and X as π_{ijx}. Collapsibility rules from loglinear analysis (see Fienberg 1980) show that, under conditional independence, the dependence between X and A in the three-variable probabilities π_{ijx} is equal to the dependence between X and A in the marginal probabilities π_{i+x}. Dependence can be studied by making a comparison with independence. Probability theory tells us that there are three ways to define independence for elements such as π_{i+x} (cf. Mood Graybill and Boes 1974:40). The first is $\pi_{i++}\pi_{++x}$, and these can be compared with dependence in π_{i+x}. The second way is $\pi_{i+x}/\pi_{++x} = \pi_{i++}$, and dependence can be easily studied by comparing conditional probabilities π_{i+x}/π_{++x} with the marginal probabilities π_{i++}. This is the usual way to study the latent class model, since $\pi_{i+x}/\pi_{++x} = \pi_{ix}^{\bar{A}Z}$. The third way is $\pi_{i+x}/\pi_{i++} = \pi_{++x}$, and now dependence can be easily studied by comparing conditional probabilities π_{i+x}/π_{i++} with the marginal probabilities π_{++x}. This is the alternative way to study the latent class model proposed here, because $\pi_{i+x}/\pi_{i++} = \pi_{ix}^{A\bar{Z}}$. Basically, insight in the LCA model is obtained since parameter $\pi_{ix}^{A\bar{Z}}$ can be interpreted as the mass of category i falling in class x.

We emphasize here that the strength of the relationship between the latent variable and the manifest variable is equally strong, whether we use the original or the rescaled parameters. In other words, if we consider rows i and i', and x and x', then the odds ratio based on four original parameters is equal to the odds ratio based on four rescaled parameters: $\pi_{ix}^{\bar{A}Z}\pi_{i'x'}^{\bar{A}Z}/\pi_{ix'}^{\bar{A}Z}\pi_{x'i}^{\bar{A}Z} = \pi_{ix}^{A\bar{Z}}\pi_{i'x'}^{A\bar{Z}}/\pi_{ix'}^{A\bar{Z}}\pi_{i'x}^{A\bar{Z}}$. (This can be easily checked by rewriting these parameters into elements such as π_{i+x}/π_{i++} and π_{i+x}/π_{++x}.)

Now that we have shown that it can also be useful to interpret LCA by means of the rescaled parameters $\pi_{ix}^{A\bar{Z}}$ and $\pi_{jx}^{B\bar{Z}}$, we will now discuss graphic representations that can be made from them (for details, see van der Ark and van der Heijden 1998). If $X = 2$, then the I categories can be displayed on a line of a two-dimensional space by using the new parameters $\pi_{i1}^{A\bar{Z}}$ (or, equivalently, $\pi_{i2}^{A\bar{Z}}$); if $X = 3$, they can be displayed in two-dimensional subspace of a three-dimensional space, in a so-called ternary diagram, and so on. (A similar interpretation holds for the parameters $\pi_{jx}^{B\bar{Z}}$.) In Sections 2.4 and 3.5 we will give examples of such graphs.

From the relation of the rescaled parameters $\pi_{ix}^{A\bar{Z}}$ and $\pi_{jx}^{B\bar{Z}}$ with the CA parameters, it will be clear that these graphic LCA-representations are very similar to the graphic CA-representation, since the parameters can be obtained from each other by linear transformations. In conclusion, if the fitted values of CA and LCA are equal, the graphical representation of CA for the row parameters can be perfectly matched with the graphical representation of LCA for the row parameters, and similarly for the column. We will illustrate this in the next section.

2.4. *Example*

As an example, we reanalyze the data collected by Srole et al. (1962) (see Table 1). The data are a cross-classification of 1660 adults in Manhattan, obtained from a sample of midtown residents aged 20–59, according to their parental socioeconomic status (SES), with categories 1 = high, . . . ,

TABLE 1
Midtown Manhattan Data

	1 = A	2 = B	3 = C	4 = D	5 = E	6 = F	Total
1 Well	64	57	57	72	36	21	307
2 Mild	94	94	105	141	97	71	602
3 Moderate	58	54	65	77	54	54	362
4 Impaired	46	40	60	94	78	71	389
Total	262	245	287	384	265	217	1660

6 = low, and mental health status with categories "well," "mild" (mild symptom formation), "moderate" (moderate symptom formation), "impaired." They have been studied previously with CA and LCA by Goodman (1987), among others, and many details can be found there. We have the following aims. First, we will illustrate that the fit of simple CA and LCA is identical. Second, we show that the interpretation of rescaled LCA parameters $\pi_{ix}^{A\bar{Z}}$ and $\pi_{jx}^{B\bar{Z}}$ gives a very similar insight into the latent class model as the interpretation of the original LCA parameters $\pi_{ix}^{\bar{A}Z}$ and $\pi_{jx}^{\bar{B}Z}$. Third, we will illustrate the rescaled LCA parameter estimates by a graphic representation. And fourth, we will show the similarity of such a graph with a CA graph.

A first remark is that the results in Section 2.3 hold for probability matrices of an exact rank R. Here we have an observed data matrix, and we can apply the results of Section 2.3 by fitting models to the observed data matrix, so that the fitted values of these models form matrices that have an exact rank R. However, if we want to compare the results from analyses of simple CA and LCA, both models should be fit with the same fitting criterion in order to obtain identical fitted values. Therefore, we fit both models with maximum likelihood. When we fit simple CA with a single dimension (i.e., model [1] with rank $R = M = 2$) the likelihood ratio chi-square equals $G^2 = 2.73$, with 8 degrees of freedom. For rank $R = X = 2$, LCA is equivalent to CA, and LCA with two latent classes has an identical fit. The results in Section 2.3 show that, for rank $R = 2$, this is always the case.

In Table 2 we give the parameter estimates of simple CA and LCA. In column 1 we find the parameter estimates for simple CA. In columns 2 and 3 we find the parameter estimates for LCA. In column 4 and 5 the rescaled parameter estimates are given. Using the interpretation of the simple CA parameters from the perspective of canonical correlation analysis of categorical variables, the CA parameter estimates show optimal category quantifications that lead to a maximized correlation of .163. The quantifications show that this positive correlation is attained by giving positive quantifications for a better mental health and a better parental socioeconomic status, and negative quantifications for a worse mental health and a worse parental socioeconomic status.

Another useful way to interpret the parameter estimates becomes evident when we rewrite (1) as

$$\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} - 1 = \frac{\pi_{ij} - \pi_{i+}\pi_{+j}}{\pi_{i+}\pi_{+j}} = \left(\sum_{m=1}^{M-1} \lambda_m \rho_{im} \gamma_{jm} \right) . \tag{11}$$

TABLE 2
Maximum-Likelihood Parameter Estimates for Srole Data. CA with One Dimension (Column 1), for LCA with 2 Latent Classes (Columns 2 and 3), and Rescaled Parameter Estimates for LCA (Columns 4 and 5)

	CA	LCA		LCA	
	$\hat{\rho}_{i1}$	$\hat{\pi}_{i1}^{\bar{A}Z}$	$\hat{\pi}_{i2}^{\bar{A}Z}$	$\hat{\pi}_{i1}^{A\bar{Z}}$	$\hat{\pi}_{i2}^{A\bar{Z}}$
Well	−1.60	.39	.00	1.00	.00
Mild	−.19	.41	.32	.54	.46
Moderate	.09	.21	.23	.45	.55
Impaired	1.48	.00	.45	.00	1.00
	$\hat{\gamma}_{j1}$	$\hat{\pi}_{j1}^{\bar{B}Z}$	$\hat{\pi}_{j2}^{\bar{B}Z}$	$\hat{\pi}_{j1}^{B\bar{Z}}$	$\hat{\pi}_{j2}^{B\bar{Z}}$
A	−1.09	.20	.12	.62	.38
B	−1.17	.19	.11	.63	.37
C	−.37	.19	.16	.53	.47
D	.05	.23	.23	.47	.53
E	1.01	.12	.20	.35	.65
F	1.80	.07	.19	.25	.75
	$\hat{\lambda}_1$	$\hat{\pi}_1^Z$	$\hat{\pi}_2^Z$	$\hat{\pi}_1^Z$	$\hat{\pi}_2^Z$
	.163	.48	.52	.48	.52

For this example, where $M = 2$, (11) shows that the product of the parameters $\lambda_1 \rho_{i1} \gamma_{j1}$ decomposes the departure from independence (i.e., $\pi_{ij} - \pi_{i+}\pi_{+j}$ scaled by the expected probability under independence). From Table 2 we can then see from the signs of the parameter estimates that rows $i = 1,2$ are positively related to columns $j = 1,2,3$, and rows $i = 3,4$ are positively related to columns $j = 4,5,6$ (i.e., the departure from independence is positive), and for the other cells there are negative relations. For cell $(i, j) = (1,1)$, the scaled departure from independence is $1.60 \times 1.09 \times .163 = .28$.

Next to these two interpretations for CA parameters many others exist, but it is beyond the scope of this paper to discuss these here. Instead we refer to Greenacre (1984), Goodman (1987, 1991), Gifi (1990), and Gower and Hand (1996), and we go on with the interpretation of the LCA solution.

For LCA of a two-way matrix, the parameters are unidentified, and in column 2 and 3 of Table 2 we have chosen solution H5 from Table 6 in Goodman (1987). For more details on this identification problem, we refer to Gilula (1979, 1983, 1984) and Goodman (1987); van der Ark, van der Heijden and Sikkel (1999) discuss the identification for the equivalent LBA model (5), and van der Ark (1999) discusses the relation between both identification problems, and the solutions proposed by Gilula, Goodman and van der Ark, van der Heijden and Sikkel. Column 2 of Table 2 shows that, given that someone is in latent class 1, his probabilities are relatively high to have a better mental health (the probability to fall in levels 1 and 2 add up to .80) and his probabilities are relatively high to have a better parental socioeconomic status (the probability to fall in levels 1, 2, and 3 add up to .58). In latent class 2 the probabilities are relatively high to have a worse mental health (levels 3 and 4 add up to .68), and his probabilities are relatively high to have a worse parental socioeconomic status (the probability to fall in levels 4, 5, and 6 add up to .62). This led Goodman (1987) to interpret the first latent class as the class for those individuals who are favorably endowed (.48 of the individuals are estimated to fall in this class), and the second latent class as the class for those individuals who are unfavorably endowed (.52 of the individuals are estimated to fall in this class).

We now interpret the rescaled LCA parameter estimates in columns 4 and 5 of Table 2. This is most simple by comparing the estimates of the parameters $\pi_{ix}^{A\bar{Z}}$ and $\pi_{jx}^{B\bar{Z}}$ with the estimates of the latent class sizes π_x^Z (see Section 2.3 for a theoretical justification). The estimate of latent class size 1 is .48 and that of latent class 2 is .52. Columns 4 and 5 show that, given that one belongs to the "well" mental health status, the probability to fall in class 1 is 1.00 and the probability to fall in class 2 is .00. This shows that being in the category "well" is closely related to the first latent class because 1.00 is much larger than .48, so knowledge about this particular row category increases the probability to fall in class 1 enormously. This relation is reversed for the "impaired" mental health status, showing that having impaired mental health is closely related to falling in the second latent class. There is a weaker relation for the intermediate categories showing that individuals having mental health status 2 fall in class 1 a bit more than average (the average being the estimate for class size, .48), and individuals having mental health status 3 fall in class 2 a bit more than average (the average being the estimate .52). For parental socioeconomic status we find a rather strong relation between levels 1 and 2 and the first latent class (estimates .61 and .61 being much higher than the average .48) and a strong relation of levels 5 and 6 with latent class 2 (estimates .64 and .75 compared with .52). This shows, first, that an interpretation of the rescaled LCA parameter estimates also leads to an interpretation of the latent classes in terms of (un)favorable endowment, and second, that the rescaled parameter estimates in columns 4 and 5 are related to the latent classes in the same way as the original parameter estimates in columns 2 and 3.

What remains is to show the similarity of a graph of the CA parameter estimates with a graph of the rescaled LCA parameter estimates in columns 4 and 5. We can compare the graph of the CA row parameters with the graph of the LCA row parameters, and we can compare the graph of the CA column parameters with the graph of the LCA column parameters. For simple CA, there are various ways to make these graphs (for example, see Goodman 1987, 1991), and we choose to make a graph using $\hat{\rho}_{i1}$ as coordinates for the rows and of $\hat{\gamma}_{j1}$ as coordinates for the columns. We will use the parameter estimates in column 5 as coordinates for the LCA graph (see Figure 1)—i.e., we make a graph for the second latent class. It will be clear from the graphs that the relative distances between points on a CA line are equal to the relative distances between points on an LCA line. This is due to (9) and (10)—i.e., if we compare, for example, the distance between rows 1 and 2 with the distance between rows 3 and 4, both for CA as well as for LCA, the ratio of their differences is $(1.60 - .19)/(-.09 - -1.48) = (1.00 - .54)/(.46 - .00)$. In both types of graphs there is a point of reference, which is 0.0 in the CA graphs and $\hat{\pi}_2^X = .52$ in the LCA graphs. In CA this point represents the point for the estimates of marginal probabilities—for example, in the CA row graph it represents the vector of the estimated

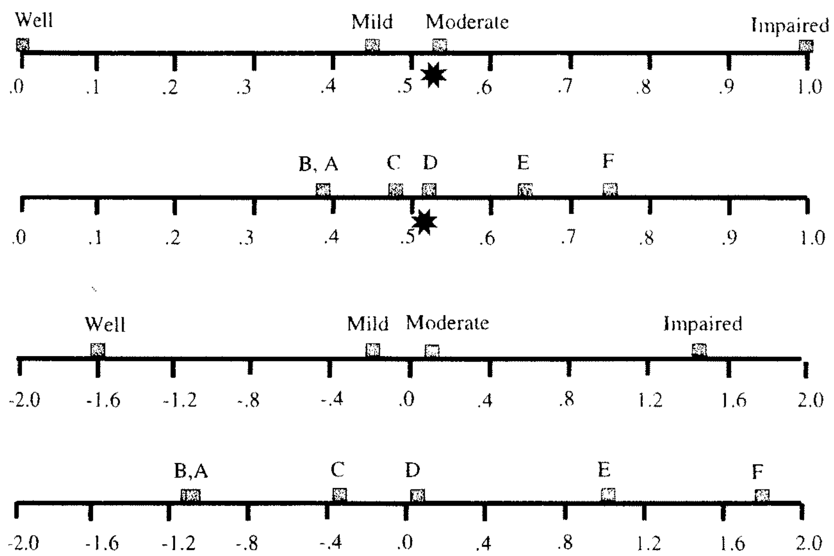

FIGURE 1. Graphic display for Srole data (1) the rescaled latent class parameter estimates with two latent classes (column 5 of Table 2) in line 1 and 2, and (2) the CA estimates (column 1 of Table 2) in line 3 and 4.

column probabilities $\hat{\pi}_{+j}$ (this is the weighted mean of all row vectors with fitted values). By comparing the CA row graph with the CA column graph, this point of reference is helpful because it shows that, for example, mental health status 4 goes together more often than average (i.e., the marginal probabilities) with levels 4, 5, and 6 of parental socioeconomic status. Similarly, in the LCA graphs the point with coordinate .52 is helpful because it shows that mental health status 4 is positively related with latent class 2, and similarly for parental socioeconomic status 4, 5, and 6. A difference between the CA and the LCA graphs is that the LCA graphs have interpretable endpoints—namely, 0.0 and 1.0. These are absent in CA. But on the whole the similarities between the graphs of CA and LCA are large.

Concluding, we have seen an illustration that simple CA and LCA are closely related. This does not mean that it does not matter which model is chosen when analyzing a contingency table. A proper choice depends on the question asked about the data. If the question is formulated in terms of a categorical latent variable explaining the interaction of the two manifest variables, LCA is the model of choice. For CA, there is more than one interpretation, but if, for example, the question is framed in terms of a maximized correlation between the categorical row and column variable, then simple CA should be the model of choice. However, it is clear that both models focus on the same aspects in the data (namely, a factorization of the relation between the variables) and therefore that the interpretations of both models are far from unrelated.

3. THE MULTIVARIATE CASE

In Section 2 we discussed the relation between CA and LCA for two-way tables. We will now discuss this relation when the number of variables is greater than two. Usually CA is then generalized into either multiple correspondence analysis (MCA) or joint correspondence analysis (JCA), and LCA is extended by including one extra set of parameters for each extra variable.

To keep the exposition simple, we will discuss the relation between the models for the situation of three variables. The relation for more than three variables is identical to the relation for three variables. We switch from probabilities π_{ij} to π_{ijk}, where k $(k = 1, \ldots, K)$ indexes the level of the third variable, C.

3.1. *Multiple Correspondence Analysis*

MCA can be introduced in many ways (for example, see Tenenhaus and Young 1985; Gifi 1990). For the purpose of this paper, where we want to compare MCA and JCA with LCA, it is easiest if we introduce MCA using the so-called Burt matrix.

The Burt matrix is a matrix of order $(I + J + K) \times (I + J + K)$ and is a concatenation of matrices consisting of the bivariate margins π_{ij+}, π_{i+k} and π_{+jk}, and diagonal matrices consisting of the univariate margins of π_{ijk} (see Figure 2). The generalization of the Burt matrix to more than three variables, and its decomposition, is straightforward: for each additional variable, the relevant bivariate and diagonal univariate matrices are concatenated to the Burt matrix in Figure 2.

It is a well-known result from the CA literature that MCA is equivalent to simple CA applied to the Burt matrix. Simple CA of the Burt matrix leads to the simultaneous decomposition of each of the sub matrices of the Burt matrix. Thus the three bivariate margins in the Burt matrix are decomposed as follows:

$$\pi_{ij+} = \pi_{i++}\pi_{+j+}\left(1 + \sum_{m=1}^{M-1} \phi_m \eta_{im} \upsilon_{jm}\right), \tag{12}$$

$$\pi_{i+k} = \pi_{i++}\pi_{++k}\left(1 + \sum_{m=1}^{M-1} \phi_m \eta_{im} \omega_{km}\right), \tag{13}$$

$$\left[\begin{array}{ccccc|ccccc|ccccc}
\pi_{1++} & & 0 & \cdots & 0 & \pi_{11+} & \cdots & \pi_{1j+} & \cdots & \pi_{1J+} & \pi_{1+1} & \cdots & \pi_{1+k} & \cdots & \pi_{1+K} \\
 & \ddots & & \ddots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
0 & & \pi_{i++} & & 0 & \pi_{i1+} & \cdots & \pi_{ij+} & \cdots & \pi_{iJ+} & \pi_{i+1} & \cdots & \pi_{i+k} & \cdots & \pi_{i+K} \\
\vdots & \ddots & & \ddots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
0 & \cdots & 0 & & \pi_{I++} & \pi_{I1+} & \cdots & \pi_{Ij+} & \cdots & \pi_{IJ+} & \pi_{I+1} & \cdots & \pi_{I+k} & \cdots & \pi_{I+K} \\
\hline
\pi_{11+} & \cdots & \pi_{i1+} & \cdots & \pi_{I1+} & \pi_{+1+} & & 0 & \cdots & 0 & \pi_{+11} & \cdots & \pi_{+1k} & \cdots & \pi_{+1K} \\
\vdots & & \vdots & & \vdots & & \ddots & & \ddots & \vdots & \vdots & & \vdots & & \vdots \\
\pi_{1j+} & \cdots & \pi_{ij+} & \cdots & \pi_{Ij+} & 0 & & \pi_{+j+} & & 0 & \pi_{+j1} & \cdots & \pi_{+jk} & \cdots & \pi_{+jK} \\
\vdots & & \vdots & & \vdots & \vdots & \ddots & & \ddots & & \vdots & & \vdots & & \vdots \\
\pi_{1J+} & \cdots & \pi_{iJ+} & \cdots & \pi_{IJ+} & 0 & \cdots & 0 & & \pi_{+J+} & \pi_{+J1} & \cdots & \pi_{+Jk} & \cdots & \pi_{+JK} \\
\hline
\pi_{1+1} & \cdots & \pi_{i+1} & \cdots & \pi_{I+1} & \pi_{+11} & \cdots & \pi_{+j1} & \cdots & \pi_{+J1} & \pi_{++1} & & 0 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \ddots & & \ddots & \vdots \\
\pi_{1+k} & \cdots & \pi_{i+k} & \cdots & \pi_{I+k} & \pi_{+1k} & \cdots & \pi_{+jk} & \cdots & \pi_{+Jk} & 0 & & \pi_{++k} & & 0 \\
\vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots & \ddots & & \ddots & \\
\pi_{1+K} & \cdots & \pi_{i+K} & \cdots & \pi_{I+K} & \pi_{+1K} & \cdots & \pi_{+jK} & \cdots & \pi_{+JK} & 0 & \cdots & 0 & & \pi_{++K}
\end{array}\right]$$

FIGURE 2. Burt matrix.

and

$$\pi_{+jk} = \pi_{+j+}\pi_{++k}\left(1 + \sum_{m=1}^{M-1} \phi_m \upsilon_{jm}\omega_{km}\right) \quad (14)$$

where

$$\lambda_m > 0 \ ,$$

$$\sum_i \pi_{++i}\eta_{im} = \sum_j \pi_{+j+}\upsilon_{jm} = \sum_k \pi_{++k}\omega_{km} = 0 \ ,$$

$$\sum_i \pi_{++i}\eta_{im}\eta_{im'} + \sum_j \pi_{+j+}\upsilon_{jm}\upsilon_{jm'} + \sum_k \pi_{++k}\omega_{km}\omega_{km'} = \delta^{mm'} \ . \quad (15)$$

The simultaneity of the three decompositions reveals itself by the fact that the parameters ϕ_m are found in both (12) and (13) as well as in (14), that the parameters η_{im} in (12) are also found in (13), that the parameters υ_{jm} in (12) are also found in (14), and, lastly, that the parameters ω_{km} in (13) are also found in (14). Because of the way the Burt matrix is built up, its maximal rank is $1 + (I - 1) + (J - 1) + (K - 1)$. The above three decompositions show that, when M is chosen such that attention is restricted to the first few sets (indexed by m) of parameters of the solution, the three observed bivariate margins are simultaneously approximated by matrices of rank M. MCA is computed using a singular value decomposition, and due to the properties of the singular value decomposition, the rank M approximation of the complete Burt matrix is optimal in a least squares sense.

One way to motivate MCA is as a generalization of principal component analysis (PCA) to nominal variables. This is similar to the interpretation of simple CA as a generalization of correlation analysis to nominal variables: There the parameters ρ_{i1} and γ_{j1} could be interpreted as quantifications of the categories that yield a maximal correlation between the quantified row and column variable. In MCA the generalization of PCA is as follows: if the parameters η_{i1}, υ_{j1}, and ω_{k1} are used as quantifications of the categories, then a 3×3 correlation matrix can be derived that has the property that the first eigenvalue is maximized (this property of MCA also holds for larger numbers of variables). We refer to van de Geer (1993) for a discussion of the second and higher dimensional interpretation of MCA seen from the perspective of PCA of nominal variables.

The final observation to be made about this perspective is that it explains why interest goes out to bivariate margins, which may come as a surprise to those researchers who are less familiar with MCA but more familiar with modeling approaches that model the multivariate frequencies. A justification for modeling bivariate margins is that, if correlations are an appropriate way to model the interaction between variables, only the second order moments in the data are needed. Using this interpretation of MCA the model is often used for the analysis of a large number of variables, often having a large number of categories. Sparseness of data then plays a relatively minor role irrespective of the number of variables, because the cells of the bivariate margins are usually reasonably filled.

Multidimensional graphic representations play an important role in the interpretation of MCA. The categories of the distinct variables can be represented in a space of dimension $M - 1$ with coordinates $(\eta_{im}\phi_m^{1/2})$ for the categories of variable A, $(\upsilon_{jm}\phi_m^{1/2})$ for the categories of variable B, and $(\omega_{k1}\phi_m^{1/2})$ for the categories of variable C. Thus, if we consider category i of variable A and category j of variable B, for example, their inner product in $M - 1$-dimensional space is equal to $(\pi_{ij+} - \pi_{i++}\pi_{+j+})/\pi_{i++}\pi_{+j+}$ (compare [12]). If $(\pi_{ij+} - \pi_{i++}\pi_{+j+})/\pi_{i++}\pi_{+j+} > 0$, then these two categories will have an angle with the origin less than 90 degrees. Thus we can see from the multidimensional representation that π_{ij+} is greater than $\pi_{i++}\pi_{+j+}$—in other words, that i and j are positively related; given a certain angle, the further i and j are away from the origin, the greater this departure will be. If the angle is greater than 90 degrees, $(\pi_{ij+} - \pi_{i++}\pi_{+j+})/\pi_{i++}\pi_{+j+} < 0$, so i and j are negatively related. And if they form a right angle, $\pi_{ij+} = \pi_{i++}\pi_{+j+}$, so i and j are unrelated. So inner products between categories of different variables give a clear and interpretable relation between the graph and the data. Now consider two categories of the same variable, say category i and i' of variable A. In the Burt matrix the element where i and i' cross equals zero. The result is that, in full-dimensional space, the inner product should equal $(0 - \pi_{i++}\pi_{i'++})/\pi_{i++}\pi_{i'++} = -1$, so category i and i' have an angle that is greater than 90 degrees. Here the inner products between the categories of the same variable give a clear relation between the graph and the data, but it does not show an aspect of the data we are interested in.

We conclude that, in going from simple CA to MCA, distinct interpretations of simple CA can be generalized to the MCA setting. We have seen that this is successful for the principal component analysis interpretation, and there are more interpretations that can be applied successfully that we have not discussed here (for example, see Tenenhaus and Young [1985] for an interpretation of MCA as generalized canonical correlation analysis, and Gifi [1990] for an interpretation of MCA as homogeneity analysis). The inner product interpretation is less useful: This last interpretation showed that the inertias in each of the subtables of the Burt matrix are decomposed (compare Section 2.1). For the subtables with bivariate margins, this is a useful property of MCA, since it shows how the bivariate margins depart from marginal independence. For the diagonal matrices with univariate margins, this interpretation is not very insightful, because we are not interested in these particular inner products, and yet they can have a large impact on the full-dimensional solution. This does not imply that this effect will always be clearly visible on the first few dimensions, to which attention is usually restricted in MCA, but it is clear that the inner products for the univariate margins play a role, and for this purpose an alternative is proposed for MCA, which is called JCA. This alternative will be discussed in Section 3.2.

Before we discuss JCA, we pay some attention to individual response patterns. An individual response pattern consists of one category of every variable, and for each dimension m each response pattern is quantified separately by averaging the quantifications of the categories it has. So, if a particular response pattern consists of categories i of variable A, j of variable B, and k of variable C, the quantification for dimension m is

$$x_{(ijk)m} = (\eta_{im} + \upsilon_{jm} + \omega_{km})/3\sigma_m \ , \tag{16}$$

where σ_m is chosen such that the weighted variance of $\sum_{m=1} \pi_{ijk} x^2_{(ijk)m} = 1$. Due to restrictions (15) $\sum_{m=1} \pi_{ijk} x_{(ijk)m} = 0$.

3.2. *Joint Correspondence Analysis*

We have just seen that the interpretation of MCA from the perspective of decomposition of inertia is only partly successful. This is used as a starting point for the proposal of a different generalization of simple CA to the multivariate case. This generalization is called joint correspondence analysis (JCA; Greenacre 1988).

In JCA the rank of the approximation has to be fixed at a prespecified $m^* = M$. Then a rank-m^* approximation is obtained that is optimal in a generalized least squares sense for the bivariate margins in the Burt matrix only. So, just as in MCA this results in the simultaneous decompositions (12), (13), and (14). However, where in MCA the rank-m^* approximation is also optimal for the diagonal matrices with univariate margins, in JCA the rank-m^* approximation is optimal only for the bivariate margins. In other words, where MCA decomposes the total inertia of the complete Burt matrix, JCA concentrates on the total inertia of each of the off-diagonal submatrices only.

Greenacre (1988) and Boik (1996) point out and discuss an analogy between MCA and principal component analysis of a correlation matrix, where both the diagonal and the off-diagonal elements are approximated by a matrix of lower rank, and an analogy between JCA and factor analysis, where the approximation holds only for the off-diagonal elements of the correlation matrix. More details and examples can be found in a series of papers by Greenacre—for example, Greenacre (1988, 1990, 1991, 1994)—in Gower and Hand (1996), and in Boik (1996).

These papers also explain how JCA can be estimated by performing MCA iteratively by updating the diagonal submatrices after each iteration. Let the prefixed rank be m^*. Let $d^{(q)}_{ii'}$ be the updated element using the parameter estimates found in iteration $q-1$ for categories i and i'. In the first step, when $q = 1$, $d^{(1)}_{ii'} = 0$. After the first step, when MCA is performed for the first time, $d^{(2)}_{ii'}$ is obtained as $d^{(2)}_{ii'} = \pi_{i++}\pi_{i'++}(1 + \sum_{m=1}^{m^*-1} \phi^{(1)}_m \eta^{(1)}_{im} \eta^{(1)}_{i'm})$, where the superscripts (1) indicate that these estimates are found by doing a generalized singular value decomposition on the Burt matrix with modified marginal elements in step 1. In general,

$$d^{(q+1)}_{ii'} = \pi_{i++}\pi_{i'++}\left(1 + \sum_{m=1}^{m^*-1} \phi^{(q)}_m \eta^{(q)}_{im} \eta^{(q)}_{i'm}\right) , \tag{17}$$

and this procedure is iterated until convergence. Boik (1996) has recently proposed a more efficient algorithm for JCA.

In JCA graphic displays are made in the same way as in MCA. So in a rank-m^* JCA, a $m^* - 1$-dimensional graphic display can be made using coordinates $(\eta_{im}\phi_m^{1/2})$ for the categories of variable A, $(\upsilon_{jm}\phi_m^{1/2})$ for the categories of variable B, and $(\omega_{k1}\phi_m^{1/2})$ for the categories of variable C. This $m^* - 1$-dimensional display can be interpreted using inner products for the bivariate margins of the Burt matrix, where it should be noted that the inner products are not equal to the bivariate marginal probabilities, but approximate them, unless $m^* = (I-1) + (J-1) + (K-1)$. The quality of the approximation can be expressed in terms of percentage of inertia displayed by the $m^* - 1$-dimensional solution. In general, compared to MCA of the Burt matrix, the percentage of inertia displayed is dramatically higher in JCA because the diagonal submatrices of the Burt matrix are not approximated. For more details, we refer to Gower and Hand (1996: chs. 4 and 10).

We conclude by mentioning that in JCA quantifications of individual response patterns are obtained in the same way as in MCA.

3.3. *The Multivariate Latent Class Model*

Moving from the bivariate case to the multivariate case, the latent class model is extended in a straightforward way by including an extra set of parameters for each additional variable. For the trivariate case, LCA is written as

$$\pi_{ijk} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} \pi_{kx}^{\bar{C}Z} \tag{18}$$

with restrictions

$$\sum_{x=1}^{X} \pi_x^Z = 1, \sum_{i=1}^{I} \pi_{ix}^{\bar{A}Z} = 1, \sum_{j=1}^{J} \pi_{jx}^{\bar{B}Z} = 1, \sum_{k=1}^{K} \pi_{kx}^{\bar{C}Z} = 1 \ .$$

Just as we did for the two-variable case discussed in Sections 2.3 and 2.4, we propose to rescale the parameters $\pi_{ix}^{\bar{A}Z}$, $\pi_{jx}^{\bar{B}Z}$, and $\pi_{kx}^{\bar{C}Z}$ into parameters $\pi_{ix}^{A\bar{Z}}$, $\pi_{jx}^{B\bar{Z}}$, and $\pi_{kx}^{C\bar{Z}}$. For the parameter $\pi_{ix}^{\bar{A}Z}$, the way to obtain such a rescaled parameter $\pi_{ix}^{A\bar{Z}}$ is shown in (6). We showed there that the rescaled parameters give us an alternative insight into the latent class model, and we showed that these rescaled parameters could be used to make graphic representations. We refer to Section 2.3 for a detailed discussion of these issues. In Section 2.4 we illustrated that these rescaled parameters led to basically the same interpretation of the latent class model as the original parameters.

In multivariate LCA it is possible to obtain probabilities that a particular individual response pattern (i,j,k) is falling in latent class x by

$$\pi_{ijkx}^{ABC\bar{Z}} = \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} \pi_{kx}^{\bar{C}Z} / \pi_{ijk} \tag{19}$$

using (18). These probabilities can be used to classify individual response patterns as belonging to latent classes. They can also be used to make a graphic display, following the principles discussed in Section 2.3.

3.4. *Relations*

The relation between LCA and MCA and JCA becomes simple when we add up π_{ijk} in (18) over k, j, and i respectively. We then get the three equations

$$\pi_{ij+} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} \ , \tag{20}$$

$$\pi_{i+k} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{kx}^{\bar{C}Z} \ , \tag{21}$$

and

$$\pi_{+jk} = \sum_{x=1}^{X} \pi_x^Z \pi_{jx}^{\bar{B}Z} \pi_{kx}^{\bar{C}Z} \ . \tag{22}$$

What this shows is that LCA implies that the three two-way matrices with bivariate margins have reduced rank X, and that equations (20) to (22) have parameters in common, showing that they should be considered simultaneously.

This is the key to our comparison of MCA and JCA with LCA. If we compare equation (12) with (20), (13) with (21), and (14) with (22), we can interpret the relation in a way similar to what we did for simple CA and LCA of two-way matrices in Section 2. For example, if we compare equation (12) with (20), we can conclude from the results of Section 2 that if the decomposition (20) is true (the bivariate margin has rank X), then the decomposition (12) with $X = M$ is true. It follows that if LCA with X latent classes is true, then the decompositions of the margins defined in (20), (21), and (22) with X latent classes are simultaneously true, and then the decompositions (12), (13), and (14) with $M = X$ are also simultaneously true.

In Section 3.1 we saw that an important distinction between MCA and JCA was that MCA is defined in terms of both the decompositions of the bivariate margins as defined in (12), (13), and (14), as well as the decomposition of the diagonal matrices with univariate margins. On the other hand, JCA is defined solely in terms of the decomposition of the bivariate margins (12), (13), and (14). An important result of this paper now follows directly: LCA with X latent classes implies JCA with $m^* = X$.

For two-way matrices, we saw in Section 2 that there was also a reverse relation for rank 2. However, it is evident that the reverse relation does not hold, because a set of lower rank two-way margins (JCA) does not have clear implications for a higher-way table (LCA).

The precise relation between LCA and JCA does not hold for LCA and MCA due to the fact that the decomposition of MCA is also defined in terms of the univariate margins.

For the bivariate case we saw in Section 2.3 that, if LCA is true, the LCA parameters can be transformed into CA parameters and the other way around. This result also holds for the multivariate case, as we will see here. To show this, consider equations (20), (21), and (22). In Section 2.3 we have shown that the original parameters $\pi_{ix}^{\bar{A}Z}$, $\pi_{jx}^{\bar{B}Z}$, and $\pi_{kx}^{\bar{C}Z}$ can be rescaled into parameters $\pi_{ix}^{A\bar{Z}}$, $\pi_{jx}^{B\bar{Z}}$, and $\pi_{kx}^{C\bar{Z}}$ using equation (6) for variable A and similar equations for variable B and C. These rescaled parameters can be used to rewrite equations (20), (21), and (22) into

$$\pi_{ij+} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{jx}^{\bar{B}Z} = \pi_{i++}\pi_{+j+} \sum_{x=1}^{X} \pi_{ix}^{A\bar{Z}} \pi_{jx}^{B\bar{Z}} (\pi_x^Z)^{-1} , \quad (23)$$

$$\pi_{i+k} = \sum_{x=1}^{X} \pi_x^Z \pi_{ix}^{\bar{A}Z} \pi_{kx}^{\bar{C}Z} = \pi_{i++}\pi_{++k} \sum_{x=1}^{X} \pi_{ix}^{A\bar{Z}} \pi_{kx}^{C\bar{Z}} (\pi_x^Z)^{-1} , \quad (24)$$

and

$$\pi_{+jk} = \sum_{x=1}^{X} \pi_x^Z \pi_{jx}^{\bar{B}Z} \pi_{kx}^{\bar{C}Z} = \pi_{+j+}\pi_{++k} \sum_{x=1}^{X} \pi_{jx}^{B\bar{Z}} \pi_{kx}^{C\bar{Z}} (\pi_x^Z)^{-1} , \quad (25)$$

(compare [7]). Assume now that the LCA model with X latent classes is true. Then, if we compare Equation (23) to (12), (24) to (13), and (25) to (14), we find the following relation between the rescaled LCA parameters and the CA parameters:

$$\left(1 + \sum_{m=1}^{R-1} \phi_m \eta_{im} v_{jm}\right) = \sum_{x=1}^{R} \pi_{ix}^{A\bar{Z}} \pi_{jx}^{B\bar{Z}} (\pi_x^Z)^{-1} , \quad (26)$$

$$\left(1 + \sum_{m=1}^{R-1} \phi_m \eta_{im} \omega_{km}\right) = \sum_{x=1}^{R} \pi_{ix}^{A\bar{Z}} \pi_{kx}^{C\bar{Z}} (\pi_x^Z)^{-1} , \quad (27)$$

and

$$\left(1 + \sum_{m=1}^{R-1} \phi_m v_{jm} \omega_{km}\right) = \sum_{x=1}^{R} \pi_{jx}^{B\bar{Z}} \pi_{kx}^{C\bar{Z}} (\pi_x^Z)^{-1} . \quad (28)$$

(compare [8]). As in Section 2.3, collect the LCA-parameters $\pi_{ix}^{A\bar{Z}}$ in a $I \times R$ matrix $\mathbf{\Pi}_i$, $\pi_{jx}^{B\bar{Z}}$ in a $J \times R$ matrix $\mathbf{\Pi}_j$, and $\pi_{kx}^{C\bar{Z}}$ in a $K \times R$ matrix $\mathbf{\Pi}_k$; and collect the CA-parameters η_{im} in a $I \times R$ matrix $\mathbf{H}$, v_{jm} in a $J \times R$ matrix $\mathbf{Y}$, and ω_{km} in a $K \times R$ matrix $\mathbf{\Omega}$, where the first columns of $\mathbf{H}$, $\mathbf{Y}$, and $\mathbf{\Omega}$ are

unit vectors. Then equation (26) shows that $\mathbf{H} = \mathbf{\Pi}_i\mathbf{U}$, (27) shows that $\mathbf{Y} = \mathbf{\Pi}_j\mathbf{U}$, and (28) shows that $\mathbf{\Omega} = \mathbf{\Pi}_k\mathbf{U}$, where $\mathbf{U}$ is a $R \times R$ transformation matrix. This shows that, if the LCA model is true, linear combinations of the rescaled LCA parameters lead to JCA parameters.

The relations just shown are interesting, because they give more insight into JCA for those who are accustomed to LCA, and vice versa. This is particularly interesting because the two tools for data analysis stem from distinct schools: LCA stems from the more traditional school of statistical modeling, where concepts like maximum likelihood and model fit are important; JCA stems from a school of exploratory data analysis, where models are fitted by least squares and interpreting graphical displays are central. This is of theoretical interest. This close relation does not lead to a preference of one model over the other, as we already argued at the end of Section 2.4. Both LCA as well as JCA give answers to different questions, and depending on the question that is asked about the data, a researcher could choose one of those models. The interesting point is that JCA focuses on the bivariate margins, and that the decompositions used in JCA are closely related to bivariate decompositions that are fitted as byproducts of the fitting of multivariate frequencies.

The relations also lead to the following practical implication, that will be illustrated in the next section. First, in the analysis of a set of data it is possible to do an LCA with R latent classes, and then this analysis can be illustrated by a JCA-graph of dimensionality $R - 1$. This JCA graph should then be obtained from a JCA on the fitted values found in LCA. The JCA-graph can then be supplemented with points for the R latent classes as well as with the point for the class sizes. The interpretation of the JCA graph can be both in terms of inner products, which is the interpretation discussed in Section 3.2, as well as in terms of the rescaled LCA-parameters. An example will be shown in the next section.

A last point to be discussed is how the representations of the response patterns are related. These representations are not as closely related as the parameters of LCA and JCA. The reason can be seen by comparing (16) with (19). This shows that in JCA the response patterns are quantified by a linear operation on the category parameters, whereas in LCA the response patterns are quantified by a multiplicative operation on the category parameters (rescaling the category parameters does not change this). Interestingly, a linear operation on the rescaled LCA-parameters would make them very similar to the quantification found for JCA in (16). For latent class m and response pattern (i, j, k) this linear operation would be

$(\pi_{ix}^{A\bar{Z}} + \pi_{jx}^{B\bar{Z}} + \pi_{kx}^{C\bar{Z}})/3$, which could be interpreted as an average conditional probability. It is not clear to us yet how useful this measure is over the usual measure (19) and we leave this as a topic for further study.

3.5. *Examples*

We will now illustrate our results by two examples. The data in the first example, shown in Table 3, were analyzed earlier with LCA by McCutcheon (1987).

There are four items on political campaign participation from the 1980 National Election Study: whether a respondent (1) voted in the election, (2) tried to influence people, (3) attended any political meetings, and (4) worked for one of the parties or candidates. For more details, see McCutcheon (1987). McCutcheon tried out a series of models for scale construction. Here we perform only an ordinary LCA. We will fit a latent class model with $X = 2$ latent classes, and first interpret the original LCA parameters. Then we will show that the rescaled LCA parameters lead to the same interpretation of the latent class model. Subsequently the rescaled LCA parameters are displayed in a one-dimensional graph. A JCA is performed on the fitted values found under LCA, and we show that this leads to a JCA graph that is very similar to the LCA graph. This JCA of fitted values is compared with a JCA and an MCA of the data.

The latent class model does not fit adequately because $G^2 = 41.6$ (df is 7). We have intentionally chosen this nonfitting example because it makes the comparison between the JCA performed on the fitted values

TABLE 3
McCutcheon Data, Four Dichotomous Items

			Vote	
Work	Attend	Influence	Yes	No
Yes	Yes	Yes	27	0
		No	2	0
	No	Yes	16	0
		No	4	1
No	Yes	Yes	40	3
		No	32	2
	No	Yes	339	83
		No	543	310

under LCA with the JCA on the data interesting: if the model would have fit the data, then the fitted values would be approximately equal to the data, and those two JCAs would necessarily lead to approximately the same solutions. Now we can see to what extend these JCAs differ.

The original parameter estimates for a solution with $X = 2$ latent classes are in columns 1 and 2 of Table 4. In the first latent class ($x = 1$), the probabilities to work, attend, influence, and vote are much lower than in latent class 2 (.002, .022, .297, and .675 respectively, as compared with .274, .452, .822, and 1.000), so that we interpret the first latent class as the class of nonactive people and the second class as the class of active people. These conditional probabilities also show that, in both classes, the probabilities to work, attend, influence, and vote are increasing in the same order.

We will now show that the rescaled parameter estimates will lead to the same interpretation of the latent classes. In column 3 (and 4) we find the estimated probabilities to fall in latent class 1 (and 2), given the response to a specific category. Given that someone works for a candidate, the estimated probability to fall in latent class 1 is .039 (and for latent class 2 it is .961), and given that one does not work for a candidate, the probability to fall in latent class 1 is .906 (and for latent class 2 it is .094). In Section 2.3 it has been shown that these estimates are to be compared with the class size estimates of .875 and .125. This shows that the parameter estimate for "yes" on working is very high indeed in class 2 (.961 respondents who have answered "yes" compared with .125 overall). Comparable interpretations for the parameter estimates of the other variables lead to the

TABLE 4
Parameter Estimates for Latent Class Analysis with Two Latent Classes

		Original Parameters		Rescaled Parameters	
		$x = 1$	$x = 2$	$x = 1$	$x = 2$
Work	Yes	.002	.274	.039	.961
	No	.998	.726	.906	.094
Attend	Yes	.022	.452	.253	.747
	No	.978	.547	.926	.074
Influence	Yes	.297	.822	.717	.283
	No	.703	.178	.965	.035
Vote	Yes	.675	1.000	.825	.175
	No	.325	.000	1.000	.000
Class size		.875	.125	.875	.125

interpretation of the first latent class as the class for nonactive people and the second class as the class of active people. Notice that the voting "yes" category is not very indicative of being in the active class, because the estimate of .175 is not much higher than the overall class size of .125. We conclude that the interpretation of the original LCA parameters leads to the same conclusion as the rescaled parameters. Notice that the rescaled parameter estimates are also ordered from work to vote, for the "yes" categories (in class 2 .961, .747, .283, and .175) as well as for the "no" categories (.094, .074, .035, and .000).

In the upper line of Figure 3, we find a graph of the rescaled LCA parameter estimates for class 2. On the left we find the "no" categories, having low probabilities, and more to the right we find the "yes" categories, having higher probabilities. The graph should be interpreted in terms of the coordinates of the points, which are the rescaled parameter estimates that we have just interpreted without having a graph. The usefulness of this graph should be that it helps in finding the same interpretation more quickly, or that it illustrates the interpretation found. It is possible to relate the parameter estimates in this graph to the bivariate margins of the fitted values, using equations like (23), but this is not easy because of the number

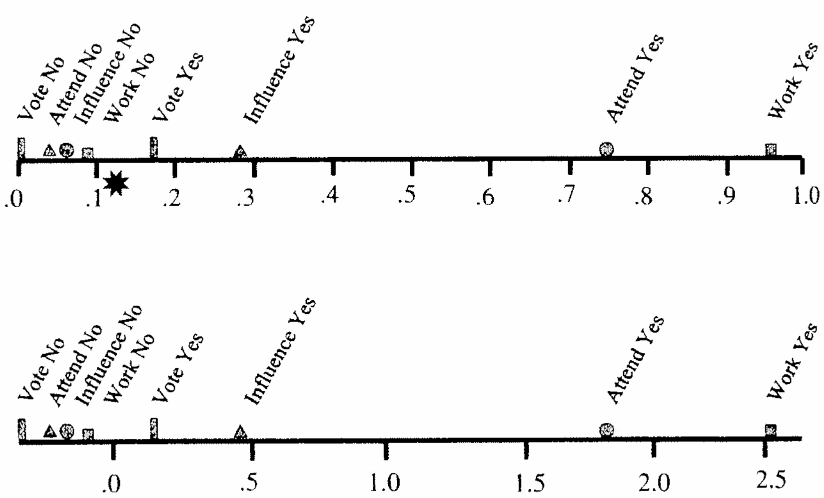

FIGURE 3. Graphic display for McCutcheon data in Table 3 (1) rescaled latent class parameter estimates with two latent classes (column 4 of Table 4) in line 1, and (2) the JCA estimates (column 1 of Table 6) in line 2.

TABLE 5
Burt Matrix Derived from Parameter Estimates of Latent Class Analysis with Two Latent Classes

.03566	.00000	.01553	.02013	.02857	.00709	.03521	.00045
.00000	.96434	.06008	.90426	.33377	.63057	.68019	.28414
.01553	.06008	.07561	.00000	.05208	.02352	.06939	.00622
.02013	.90426	.00000	.92439	.31026	.61414	.64602	.27838
.02857	.33377	.05208	.31026	.36234	.00000	.27788	.08446
.00709	.63057	.02352	.61414	.00000	.63766	.43753	.20013
.03521	.68019	.06939	.64602	.27788	.43753	.71541	.00000
.00045	.28414	.00622	.27838	.08446	.20013	.00000	.28459

of operations involved. We come back to a further interpretation of this graph after we have discussed the JCA graph of the fitted values of LCA.

In Table 5 the Burt matrix derived from the fitted values under LCA is given. In Table 6, first column, we provide the estimates obtained by JCA of this table. Since the Burt matrix in Table 5 has rank 2, JCA with one dimension provides a perfect fit of the solution. The estimates are displayed in a graph shown in the lower line of Figure 3. We have multiplied the estimates in the first column with $\phi_1^{1/2}$ so that we can relate the inner products of the coordinates of the points using equa-

TABLE 6
Parameter Estimates for Joint CA of LCA Estimates (Column 1), for JCA of Observed Data (Column 2), and MCA of Observed Data (Column 3)

		JCA of LCA	JCA of Data	MCA of Data
Work	Yes	5.989	6.381	5.824
	No	−.222	−.236	−.215
Attend	Yes	4.456	4.372	3.948
	No	−.365	−.358	−.323
Influence	Yes	1.134	1.017	1.237
	No	−.644	−.578	−.703
Vote	Yes	.356	.370	.489
	No	−.895	−.929	−1.229
	$\hat{\phi}$	.1783	.2213	.4034
	prop.	1.0000	.8796	.4034

tions like (12) (see Section 3.1 for details). For example, for this four-variable example the bivariate margin of the first two variables is π_{ij++}, and equations such as (12), (13), and (14) show that the coordinates of the points for the first two variables are related to the data (here: the fitted values of LCA) by $(\pi_{ij++} - \pi_{i+++}\pi_{+j++})/\pi_{i+++}\pi_{+j++} = (\eta_{i1}\phi_1^{1/2})(\upsilon_{j1}\phi_1^{1/2})$. For example, for the categories work-yes, attend-yes $(\eta_{i1}\phi_1^{1/2})(\upsilon_{j1}\phi_1^{1/2}) = .1783 \times (5.989 \times 4.456) = 4.752$, and this is equal to $(\pi_{ij++} - \pi_{i+++}\pi_{+j++})/\pi_{i+++}\pi_{+j++} = (.01553 - .03566 \times .07561)/.03566 \times .07561 = 4.752$. As discussed in Section 3.1, when the inner products are positive, then $\pi_{ij++} > \pi_{i+++}\pi_{+j++}$; when they are negative, then $\pi_{ij++} < \pi_{i+++}\pi_{+j++}$. It follows that points are positively related in the bivariate margins to points on the same side of the origin, and points are negatively related to points on the opposite side of the origin. This shows that in the bivariate margins, all the yes-categories are positively related, and all the no-categories are positively related, and every yes-category is negatively related to every no-category.

Due to the fact that the Burt matrix analyzed by JCA consisted of the fitted values from JCA, the relative distances between the categories in the JCA graph are identical to the relative distances between the categories in the LCA graph. But this also means that some properties derived from the inner product interpretation of the JCA graph also hold for the LCA graph. In particular, where we concluded for JCA that points on one side of the origin are positively related in the bivariate margins and points on opposite sides of the origin are negatively related in the bivariate margins, we can now conclude for the LCA graph that points on one side of the class size point (indicated with a star in Figure 3) are positively related in the bivariate margins and points on opposite sides of the class size points are negatively related. Thus the straightforward aspects of the interpretation of both graphs can supplement each other. The coordinates of the JCA plot are more easily related to the bivariate margins using an inner product interpretation; yet their precise values are more difficult to interpret. For the LCA graph, it is the other way around: here the precise values are easily interpreted, since they are conditional probabilities; yet these conditional probabilities are more difficult to relate to the bivariate margins of the fitted values.

We now discuss the JCA of the observed Burt matrix and compare it with the JCA of the Burt matrix with fitted LCA values. In the observed Burt matrix most relations between the yes-yes and the no-no combinations are a bit stronger than in the Burt matrix with fitted LCA values, in particular between the variables Influence and Vote, where the departure is

.023. The total inertia for the observed Burt matrix is higher (.2515, whereas it is .1783 for the Burt matrix with fitted values), of which .2213/.2515 = 87.96 percent displayed. The parameter estimates are given in column 2 of Table 6. Apart from the larger first eigenvalue (.2213 compared with .1783), that reflects the stronger association between the yes-yes and the no-no combinations, no interpretable differences can be found between the JCA of the fitted LCA values and the JCA of the data.

In column 3 of Table 6 we find the scores obtained with MCA of the Burt matrix from the observed data (eigenvalues are .4034, .2391, .2004, and .1570). The total inertia is much higher, due to the fact that the inertia of the diagonal submatrices is now also decomposed. Yet the estimates for the categories in the MCA solution are rather similar to the estimates of the two JCA solutions.

We conclude that the estimates of the three solutions are similar, and this is remarkable, because the fit of LCA was not very good; yet the maximum likelihood approximation provided by LCA to the bivariate margins is very similar to the least-squares approximation to the bivariate margins provided by JCA and MCA.

This similarity is not found in the second example, also taken from McCutcheon (1987). Table 7 shows the observed frequencies for four categorical variables from the 1982 General Social Survey. Two are white

TABLE 7
Cross-tabulation of Observed Variables for White Respondents of 1982 General Social Survey

Purpose	Accuracy	Understanding	Interested	Cooperative	Impatient/ Hostile
Good	Mostly true	Good	419	35	2
		Fair, poor	71	25	5
	Not true	Good	270	25	4
		Fair, poor	42	16	5
Depends	Mostly true	Good	23	4	1
		Fair, poor	6	2	0
	Not true	Good	43	9	2
		Fair, poor	9	3	2
Waste	Mostly true	Good	26	3	0
		Fair, poor	1	2	0
	Not true	Good	85	23	6
		Fair, poor	13	12	8

respondents' evaluations of surveys and two are interviewers' evaluations of these respondents.

The first variable asks about the perceived purpose of surveys, the second about the accuracy of survey results, the third about the perceived cooperation of the respondent in the survey and the fourth about the perceived understanding (for more details, see McCutcheon 1987). LCA with $X = 3$ latent classes has an acceptable fit ($G^2 = 21.89$, df is 16). The parameter estimates are presented in the first three columns of Table 8.

McCutcheon (1987) interprets the first latent class as the class of "ideal respondents," because they often find the purpose of surveys "good" (estimated probability is .888), the results are "mostly true" (.613), they always have a "good" understanding (1.000), and their cooperation is almost always "interested" (.943); the third class has "skeptics," they often answer that surveys are a "waste" of time (.633), the results are often found "not true" (.969), but they are often found "interested" (.641) and have often a "good" understanding (.753). The second class is interpreted as the class for "believers": they often answer that the purpose is "good" (.910), and that the results are "mostly true" (.648), their cooperation is often "interested" (.688) but their understanding is "fair/poor" (.686).

In columns 4, 5, and 6 of Table 8, we find the rescaled parameter estimates. These lead to the same interpretation of the latent classes. To

TABLE 8

Parameter Estimates for Latent Class Analysis with Three Latent Classes. First Three Columns Give Original Parameter Estimates, and Columns 4 to 6 Give Rescaled Parameter Estimates Discussed in Section 2.2

		$x = 1$	$x = 2$	$x = 3$	$x = 1$	$x = 2$	$x = 3$
Purpose	Good	.888	.912	.143	.721	.247	.032
	Depends	.053	.072	.225	.382	.171	.447
	Waste	.059	.017	.633	.245	.023	.732
Accuracy	Mostly true	.613	.648	.031	.732	.258	.010
	Not true	.387	.352	.969	.500	.152	.348
Understanding	Good	1.000	.313	.753	.761	.079	.159
	Fair, poor	.000	.687	.247	.000	.770	.230
Cooperation	Interested	.943	.690	.641	.698	.170	.132
	Cooperative	.057	.255	.256	.267	.400	.333
	Impatient/ hostile	.000	.055	.103	.000	.391	.609
Class Size		.621	.207	.172	.621	.207	.172

find the correct interpretation, we have to compare these conditional probabilities with the class sizes (see Section 2.3). This shows that given that the purpose is judged as "good," respondents fall more than average in latent class 1 (estimated probability .721 compared with the average—i.e. class size estimate, .621), similarly for accuracy "mostly true" (.732), "good" understanding (.761), and "interested" cooperation (.698). This leads again to the labeling of the first latent class as "ideal respondents." For the second latent class, the following categories have higher estimated probabilities than the average—i.e., the class size estimate, .207: "good" purpose (.247), accuracy "mostly true" (.258), "fair, poor" understanding (.770), cooperation "cooperative" (.400), and "impatient/hostile" (.391). This leads again to the labeling of the second latent class as "believers." For the third latent class, the categories "depends" and "waste" of Purpose, "not true" of Accuracy, "fair, poor" of Understanding and "cooperative" and "impatient/hostile" are larger than the class size estimate; this leads again to the interpretation of class 3 as a class of "skeptics."

Figure 4 gives a graphic display of the rescaled parameter estimates in columns 4, 5, and 6 of Table 8. We could graph them as points in a three-dimensional space, but because these rescaled estimates add up to 1 for each category, they are all lying in a two-dimensional subspace of this three-dimensional space; and because the estimates are all nonnegative, they fall in a triangle. Such a triangle is called a ternary diagram. The top of the triangle is the point with coordinates (1,0,0), and this shows that it is the point for latent class 1 (i.e., there the mass is falling completely in class 1); the bottom-right point has coordinates (0,1,0) and it is the point for latent class 2; the bottom-left point has coordinates (0,0,1) and it is the point for latent class 3. As an example of a point in the triangle, we will discuss the coordinates of the point for the class sizes, which has coordinates (.621, .207 and .172) (see Table 8). We use the edge of the triangle going from bottom left (first coordinate is zero) to the top (first coordinate is one) as the coordinate axis for the first latent class. From point .621, a dotted line is found going from the first coordinate axis to the right. We use the edge going from the top to bottom-right as the coordinate axis for the second latent class. From point .207, a dotted line is going parallel to coordinate axis 1, and the edge going from the bottom-right to the bottom-left is the coordinate axis for the third latent class. At point .172, a dotted line is going parallel to the second coordinate axis. The three dotted lines meet in the point for class size.

We have already interpreted the rescaled parameter estimates, and Figure 4 is simply a graph of these estimates. We will now show how the

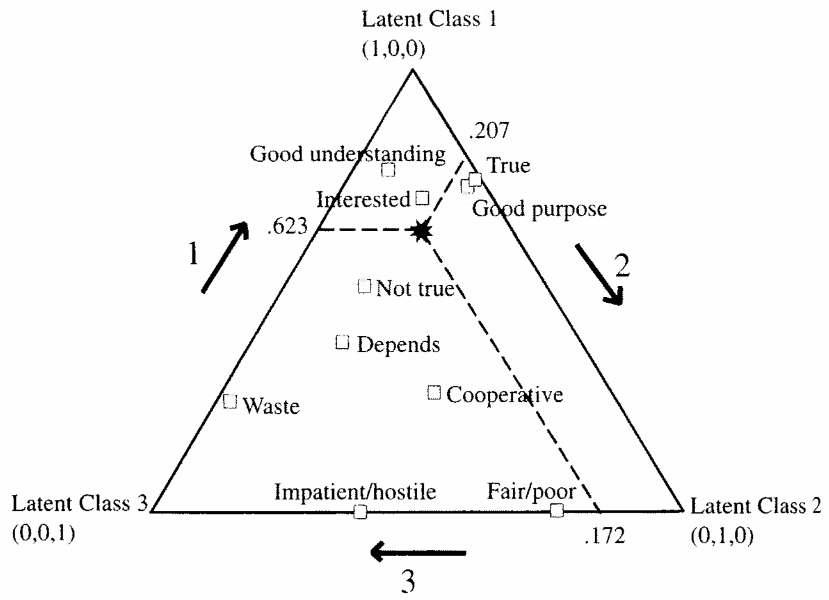

FIGURE 4. Graphic display of the rescaled estimates of the latent class model with three latent classes (columns 4, 5, and 6 of Table 8).

interpretation made of the rescaled parameter estimates can also be obtained from the graph. In our interpretation of the rescaled parameter estimates, the first latent class was characterized by relatively higher estimates for "good" purpose, "good" understanding, mostly "true" accuracy, and "interested" cooperation, and points for these categories are all in-between the point for the class sizes and the point for the first latent class (the top corner of the triangle). The second latent class was characterized by "good" purpose and "mostly true" (but their estimates of .247 and .258 are only marginally higher than the average .207), "fair, poor" understanding and "cooperative" cooperation; this is also revealed by Figure 4. The third latent class was characterized by "waste" and "depends," by purpose "not true," marginally by "fair/poor" and "cooperative," but particularly by "impatient/hostile." These characterizations are all rather clear from Figure 4, and this figure can give support in the interpretation and presentation of the results of the LCA.

In Table 9 we give the Burt matrix derived from the fitted values of LCA. Table 10 gives in the first two columns the estimates for JCA of

TABLE 9
Burt Matrix Derived from Parameter Estimates of Latent Class Analysis with Three Latent Classes

.765	.000	.000	.461	.304	.629	.136	.666	.086	.013
.000	.087	.000	.031	.055	.067	.020	.066	.016	.005
.000	.000	.149	.028	.121	.120	.029	.107	.031	.011
.461	.031	.028	.520	.000	.427	.093	.455	.057	.008
.304	.055	.121	.000	.480	.389	.091	.384	.075	.017
.629	.067	.120	.427	.389	.815	.000	.713	.085	.017
.136	.020	.029	.093	.091	.000	.185	.125	.047	.012
.666	.066	.107	.455	.384	.713	.125	.839	.000	.000
.086	.016	.031	.057	.075	.085	.047	.000	.132	.000
.013	.005	.011	.008	.021	.017	.012	.000	.000	.029

Table 9 (eigenvalues are .1970 and .1329, and the other eigenvalues are equal to zero). Notice that 100 percent of the inertia is displayed in two dimensions, and this follows directly from the result that LCA of three latent classes implies JCA with two dimensions.

In Figure 5 we find the JCA graph of the fitted LCA values. As in the first example in this section, we have scaled the parameter estimates in such a way that we can interpret inner products between the categories. For example, the category "waste" has coordinates (1.346, −.631) and the category "impatient/hostile" has coordinates (1.346, .284) (these coordinates are found by multiplying parameter estimates for "waste" and "impatient/hostile" in the first two columns of Table 10 with the estimates $\hat{\phi}_m^{1/2}$). The inner product between these points is (1.346 × 1.346 + (−.631) × .284) = 1.633, and this is equal to [.011 − (.149 × .029)]/(.149 × .029) (see Table 9). This means that when we consider the "waste" and "impatient/hostile" in the bivariate margins for the variables "purpose" and "cooperation," the difference between the LCA-estimates and marginal independence (i.e., [.011 − (.149 × .029)]) is 1.633 times as large as the estimate under marginal independence (.149 × .029), so there is a strong positive dependence between "waste" and "impatient/hostile."

By a linear transformation, the first and second dimension of the configuration of points in Figure 5 one could obtain precisely the configuration of points in the ternary diagram in Figure 4. Graphically, this linear transformation is equivalent to rotating and stretching the three axes of the triangle in Figure 4.

TABLE 10
Parameter Estimates for JCA of LCA Estimates (Columns 1 and 2), for JCA of Observed Data (Columns 3 and 4), and MCA of Observed Data (Columns 5 and 6)

		LCA		JCA		MCA	
		$m = 1$	$m = 2$	$m = 1$	$m = 2$	$m = 1$	$m = 2$
Purpose	Good	−.772	.393	−.066	−.863	−.613	.385
	Depends	1.600	−.495	.167	1.658	1.222	−.577
	Waste	3.032	−1.730	.240	3.468	2.437	−1.644
Accuracy	Mostly true	−.882	.486	−.069	−.983	−.996	.878
	Not true	.955	−.526	.075	1.065	1.079	−.951
Understanding	Good	−.339	−.817	−.935	.110	−.327	−.642
	Fair, poor	1.497	3.608	4.126	−.488	1.444	2.834
Cooperation	Interested	−.322	−.200	−.147	−.216	−.434	−.392
	Cooperative	1.372	1.099	.643	.843	1.769	2.103
	Impatient/hostile	3.033	.778	1.328	2.396	4.463	1.727
	$\hat{\phi}_m$	.1970	.1329	.4231	.1920	.3709	.2858
	Proportion	.5972	.4028	.6787	.3080	.2473	.1906

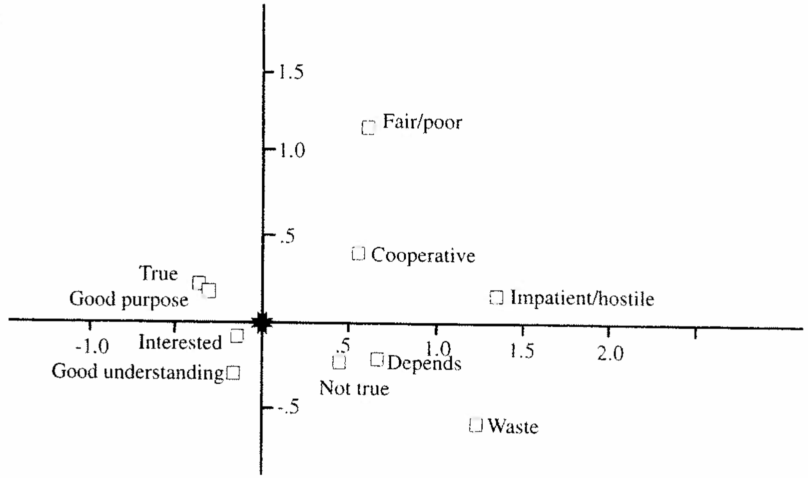

FIGURE 5. Graphic display of two principal axes of JCA (columns 1 and 2 of Table 10).

For the first example in this section, we concluded that the graphic LCA in Figure 4 and the graphic JCA representation in Figure 5 each have their strong points: the JCA graph is rather easily interpreted in terms of inner products, and the LCA graph is interpreted rather easily by its coordinates that are conditional probabilities. This conclusion also holds for the second example. In Figure 6 we combine the two advantages by adding to the JCA graph the three points for the latent classes. It should be noticed that through this joint representation the origin in the JCA solution coincides with the point for latent class size, so this point now has a double interpretation.

Examples such as Figure 6 can very well supplement an LCA. In Figure 6 the rescaled LCA parameter estimates are displayed graphically, and this can be helpful in the interpretation of the latent classes. The JCA interpretation relates inner products between the coordinates to bivariate margins of the fitted values. Thus JCA shows how LCA explains the dependence in the bivariate margins by assuming a latent variable.

We now compare the estimates for the JCA of the fitted values with the estimates of JCA and MCA of the observed data. (Eigenvalues are .4231, .1920, .0056, and .0026 for JCA, and the first two dimensions display 98.48 percent of the inertia; eigenvalues are .3709, .2858, .2505, .2486,

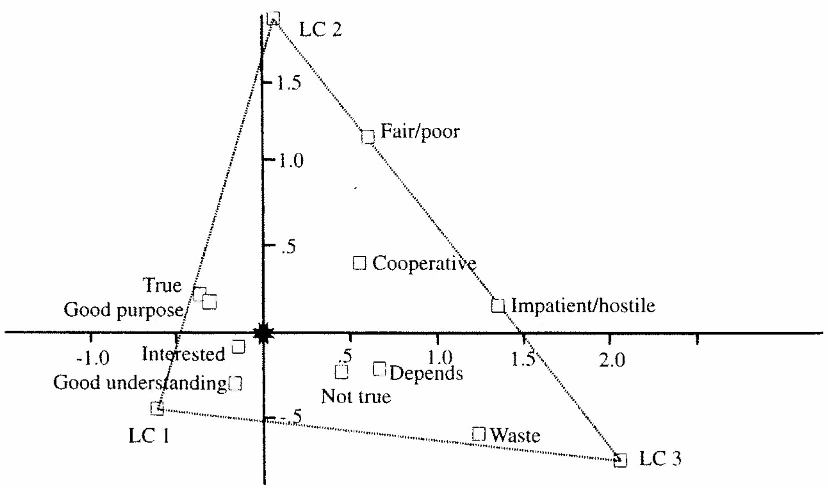

FIGURE 6. Graphic display combining aspects of Figure 4 (triangle) and Figure 5.

.1806, and .1636 for MCA, and the first two dimensions display 43.97 percent of the total inertia.) We can thus conclude that we have here an example that the adequate maximum likelihood approximation provided by LCA to the bivariate margins is similar to the least-squares approximation to the bivariate margins provided by MCA but rather different from the least-squares approximation provided by JCA. The reason is that the approximation provided for the diagonal matrix of the third variable has negative elements (cf. Boik 1996, who uses the parallel between JCA and factor analysis by calling this a Heywood case). This makes it possible for the JCA of the data to find much higher inertias than in the JCA of the fitted LCA values by using rather different parameter estimates.

4. CONCLUSIONS AND DISCUSSION

We have seen that LCA and JCA are closely related models, and their similarity makes it possible to use the same type of graphical representations of the parameters in both models. The relation just shown between LCA and JCA is remarkable, because both models have very different interpretations. First, LCA uses a latent variable, whereas JCA does not. Second, the latent variable in LCA is categorical and finds X classes of objects, whereas JCA finds $M = X - 1$ quantifications of each of the objects; in other words, in LCA the objects differ in a discrete way, whereas in JCA the objects differ in a quantitative way.

There are a few points that deserve further study, and we will discuss them shortly.

A first point deals with representations of individual response patterns. At the end of Section 3.4, we concluded that the representations used for LCA and JCA are not closely related, but we also proposed a different representation for LCA that was closely related to the representation of JCA. Further study should lead to an answer to the question of what the properties are of this different representation, and in what circumstances it is useful.

A second point for further study has to do with the comparison between the JCA of fitted values from LCA, the JCA of the observed data, and the MCA of the observed data. A central question to be asked is under what circumstances these three analyses are similar or dissimilar. For the first example, we found a remarkable resemblance between the estimates for the categories for all three analyses, although the fitted values from JCA differed significantly from the observed data. In the second example, the LCA model fitted well, and therefore we expected resemblance of the JCA of fitted values from LCA and the JCA of the observed data. However, due to negative elements in the JCA of the observed data the solutions differed considerably. Similarly, when can we expect JCA of the observed data and MCA of the observed data to be similar or dissimilar? This seems to be an open question.

And last, of natural interest is to extend JCA to a full multivariate model, so that it starts off from multivariate probabilities instead of from the bivariate marginal probabilities. This could possibly lead to an even greater similarity between JCA and LCA than we have found thus far. In particular a referee suggested the following trivariate JCA model, which indeed has a great appeal:

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}\left(1 + \sum_{m=1} \phi_m \eta_{im} \upsilon_{jm} \eta_{km}\right) . \qquad (29)$$

Such models however go beyond the scope of the current paper, which compares models for distributions that have canonical forms. As is well known, joint distributions of three or more variables do not have a canonical representation but bivariate distributions do. A canonical form is a saturated model for a joint distribution of a given dimension, and in our study represents a hierarchy of models starting from full independence (for the lowest possible dimension) on the one end, up to the saturated model for that given dimension on the other end. One very important aspect of a canonical form is that it decomposes the association into orthogonal components, and each component accounts for part of the association. In that respect each component must include an interpretable measure of association between the underlying variables. In bivariate forms we have ϕ_m as a correlation measure between the two variables. All CA models contain correlation measures. In model (29) the parameter does not have such an interpretation. Gilula (1986) deals with some of these aspects, where a variety of trivariate models based on two-dimensional association is proposed. Gilula and Haberman (1988) also use bivariate CA representations for multivariate prediction problems. Also it was illustrated by Gilula (1984) that comparing latent class models with CA models in terms of reparameterization is far more complex in the multivariate case than in the bivariate case. For all these reasons, we do not consider model (29) in this paper, but we intend to investigate in the near future models such as model (29) which could add more insight into the analysis of multivariate categorical variables.

REFERENCES

Benzécri, J.-P. et al. 1973. *L'Analyse des Données. L'Analyse de Correspondence.* Paris: Dunod.

Boik, R. J. 1996. "An Efficient Algorithm for Joint Correspondence Analysis." *Psychometrika* 61:255–69.

Clogg, C. C. 1981. "Latent Structure Models of Mobility." *American Journal of Sociology* 86:836–68.

de Leeuw, J., and P. G. M. van der Heijden. 1991. "Reduced Rank Models for Contingency Tables." *Biometrika* 78:229–32.

Fienberg, S. E. 1990. *The Analysis of Cross-classified Categorical Data.* Cambridge, MA: MIT Press.

Gifi, A. 1990. *Non-linear Multivariate Analysis.* New York: Academic Press.

Gilula, Z. 1979. "Singular Value Decomposition of Probability Matrices: Probabilistic Aspects of Latent Dichotomous Variables." *Biometrika* 66:339–44.

———. 1983. "Latent Conditional Independence in Two-way Contingency Tables: A Diagnostic Approach." *British Journal of Mathematical and Statistical Psychology* 36:114–22.

———. 1984. "On Some Similarities Between Canonical Correlation Models and Latent Class Models for Two-way Contingency Tables." *Biometrika* 71:523–29.

———. 1986. "Grouping and Association in Contingency Tables: An Exploratory Canonical Correlation Approach." *Journal of the American Statistical Association* 81:773–79.

Gilula, Z., and S. J. Haberman. 1986. "Canonical Analysis of Contingency Tables by Maximum Likelihood." *Journal of the American Statistical Association* 81:780–88.

———. 1988. "The Analysis of Multivariate Contingency Tables by Restricted Canonical and Restricted Association Models." *Journal of the American Statistical Association* 83:760–71.

Goodman, L. A. 1974a. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.

———. 1974b. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. *I.* A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–259.

———. 1985. "The Analysis of Cross-classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables with or Without Missing Entries." *The Annals of Statistics* 13:10–69.

———. 1987. "New Methods for Analyzing the Intrinsic Character of Qualitative Variables Using Cross-classified Data." *American Journal of Sociology* 93:529–83.

———. 1991. "Measures, Models, and Graphical Displays in the Analysis of Cross-classified Data." *Journal of the American Statistical Association* 86:1085–138.

Gower, J. C., and D. J. Hand. 1996. *Biplots.* London: Chapman and Hall.

Greenacre, M. J. 1984. *Theory and Applications of Correspondence Analysis.* London: Academic Press.

———. 1988. "Correspondence Analysis of Multivariate Categorical Data by Weighted Least-squares." *Biometrika* 75:457–67.

———. 1990. "Some Limitations of Multiple Correspondence Analysis." *Computational Statistics Quarterly* 3:249–56.

———. 1991. "Interpreting Multiple Correspondence Analysis." *Applied Stochastic Models and Data Analysis* 7:195–210.

———. 1994. "Multiple and Joint Correspondence Analysis." Pp. 141–61 in *Correspondence Analysis in the Social Sciences: Recent Developments and Applications,* edited by M. Greenacre and J. Blasius. London: Academic Press.

Kendall, M. G., and A. Stuart. 1979. *The Advanced Theory of Statistics,* Vol. 2, 4th ed. New York: Hafner.

Lazarsfeld, P. F. 1950. "The Logical and Mathematical Foundations of Latent Structure Analysis." Pp. 362–412 in *Measurement and Prediction,* edited by S. A. Stouffer et al. Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis.* New York: Houghton Mifflin.

McCutcheon, A. L. 1987. *Latent Class Analysis. Series: Quantitative Applications in the Social Sciences, 64.* London: Sage.

Michailides, G., and J. de Leeuw. 1998. "The Gifi Sytem for Nonlinear Multivariate Analysis." *Statistical Science* 13(4):307–36.

Mood, A. M., R. A. Graybill, and D. C. Boes. 1974. *Introduction to the Theory of Statistics,* 3rd ed. London: MacGraw-Hill.

Nishisato, S. 1980. *Analysis of Categorical Data: Dual Scaling and Its Applications.* Toronto: University of Toronto Press.

Renner, R. M. 1993. "The Resolution of a Compositional Data Set into Mixtures of

Fixed Source Compositions." *Applied Statistics, Journal of the Royal Statistical Society Series C*, 42, no. 4, 615–31.

Srole, L., T. S. Langner, S. T. Michael, M. K. Opler, and T. A. C. Rennie. 1962. *The Midtown Manhattan Study*. New York: McGraw-Hill.

Tenenhaus, M., and F. W. Young. 1985. "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis, and Other Methods for Quantifying Categorical Multivariate Data." *Psychometrika* 50:91–119.

van de Geer, J. P. 1993. *Multivariate Analysis of Categorical Data.* (2 vols.) Sage: Newbury Park.

van der Ark, L. A. 1999. "Contributions to Latent Budget Analysis; A Tool for the Analysis of Compositional Data." Leiden, Netherlands: DSWO Press.

van der Ark, L. A., and P. G. M. van der Heijden. 1998. "Graphical Display of Latent Budget Analysis and Latent Class Analysis, with Special Reference to Correspondence Analysis." Pp 489–509 in *Visualization of Categorical Data*, edited by J. Blasius, and M. Greenacre. San Diego: Academic Press.

van der Ark, L. A., P. G. M. van der Heijden, and D. Sikkel. (1999). "On the Geometry and Identifiability in the Latent Budget Model." *Journal of Classification* 16:117–37.

van der Heijden, P. G. M., A. Mooijaart, and J. de Leeuw. 1992. "Constrained Latent Budget Analysis." Pp. 279–320 in *Sociological Methodology 1992*, Cambridge, MA: Blackwell Publishers.