

# Chapter 10

## LATENT CLASS MODELS

JAY MAGIDSON

JEROEN K. VERMUNT

### 10.1. INTRODUCTION

Latent class (LC) modeling was initially introduced by Lazarsfeld and Henry (1968) as a way of formulating latent attitudinal variables from dichotomous survey items. In contrast to factor analysis, which posits continuous latent variables, LC models assume that the latent variable is categorical, and areas of application are more wide-ranging. The methodology was formalized and extended to nominal variables by Goodman (1974a, 1974b), who also developed the maximum likelihood (ML) algorithm that serves as the basis for many of today's LC software programs. In recent years, LC models have been extended to include observable variables of mixed scale type (nominal, ordinal, continuous, and counts) and covariates, as well as deal with sparse data, boundary solutions, and other problem areas.

In this chapter, we describe three important special cases of LC models for applications in cluster, factor, and regression analyses. We begin by introducing the LC cluster model as applied to nominal variables (the traditional LC model), discuss some limitations of this model, and show how recent extensions can be used to overcome them. We then turn to a formal treatment of the LC factor model and an extensive introduction to LC regression models before returning to show how the LC cluster model, as applied to continuous variables, can be used to improve on the K-means approach to

cluster analysis. We use the Latent GOLD computer program (Vermunt & Magidson, 2003) to illustrate the use of these models as applied to several data sets.

### 10.2. TRADITIONAL LATENT CLASS MODELING

Traditional LC analysis (i.e., Goodman, 1974b) assumes that each observation is a member of one and only one of  $T$  latent (unobservable) classes and that *local independence* exists between the manifest variables. That is, conditional on latent class membership, the manifest variables are mutually independent of each other. This model can be expressed using (unconditional) probabilities of belonging to each latent class and conditional response probabilities as parameters. For example, in the case of four nominal manifest variables  $A$ ,  $B$ ,  $C$ , and  $D$ , we have

$$\pi_{ijklt} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}, \quad (1)$$

where  $\pi_t^X$  denotes the probability of being in latent class  $t = 1, 2, \dots, T$  of latent variable  $X$ ;  $\pi_{it}^{A|X}$  denotes the conditional probability of obtaining the  $i$ th response to item  $A$ , from members of class  $t$ ,  $i = 1, 2, \dots, I$ ; and  $\pi_{jt}^{B|X}$ ,  $\pi_{kt}^{C|X}$ ,  $\pi_{lt}^{D|X}$ ,  $j = 1, 2, \dots, J$ ,  $k = 1, 2, \dots, K$ ,  $l = 1, 2, \dots, L$ , denote the corresponding conditional probabilities for items  $B$ ,  $C$ , and  $D$ , respectively.

Model 1 can be described graphically in terms of a path diagram (or a graphical model) in which manifest variables are not connected to each other directly but indirectly through the common source  $X$ . The latent variable is assumed to explain all of the associations among the manifest variables. A goal of traditional LC analysis is to determine the smallest number of latent classes  $T$  that is sufficient to explain away (account for) the associations (relationships) observed among the manifest variables.

The analysis typically begins by fitting the  $T = 1$  class baseline model ( $H_0$ ), which specifies mutual independence among the variables. Model  $H_0$ :

$$\pi_{ijkl} = \pi_i^A \pi_j^B \pi_k^C \pi_l^D.$$

Assuming that this *null* model does not provide an adequate fit to the data, a one-dimensional LC model with  $T = 2$  classes is then fitted to the data. This process continues by fitting successive LC models to the data, each time adding another dimension by incrementing the number of classes by 1, until the simplest model is found that provides an adequate fit.

### 10.2.1. Assessing Model Fit

Several complementary approaches are available for assessing the fit of LC models. The most widely used approach uses the likelihood ratio chi-squared statistic  $L^2$  to assess the extent to which maximum likelihood (ML) estimates for the expected cell frequencies,  $\hat{F}_{ijkl}$  differ from the corresponding observed frequencies,  $f_{ijkl}$ :

$$L^2 = 2 \sum_{ijkl} f_{ijkl} \ln(\hat{F}_{ijkl}/f_{ijkl}).$$

A model fits the data if the value of  $L^2$  is sufficiently low to be attributable to chance (within normal statistical error limits—generally, the .05 level).

The  $\hat{F}_{ijkl}$  are obtained using the following two-step process. First, ML estimates for the model parameters are obtained and substituted into the right side of equation (1) to obtain ML estimates of the probabilities  $\hat{\pi}_{ijkl}$ . These probability estimates are then summed over the latent classes to obtain estimated probabilities for each cell in the observed table and multiplied by the sample size  $N$  to obtain the ML estimates for the expected frequencies:

$$\hat{F}_{ijkl} = N \sum_{t=1}^T \hat{\pi}_{ijklt}.$$

In the case that  $\hat{F}_{ijkl} = f_{ijkl}$  for each cell  $(i, j, k, l)$ , the model fit will be perfect and  $L^2$  equals zero. To the

extent that the value for  $L^2$  exceeds 0, the  $L^2$  measures lack of model fit, quantifying the amount of association (nonindependence) that remains unexplained by that model. When  $N$  is sufficiently large,  $L^2$  follows a chi-square distribution, and as a general rule,<sup>1</sup> the number of degrees of freedom ( $df$ ) equals the number of cells in the full multiway table minus the number of distinct parameters  $M$  minus 1. For example, in the case of four categorical variables, the number of cells equals  $IJKL$ , and the number of parameters is the following:

$$M = T - 1 + T[(I - 1) + (J - 1) + (K - 1) + (L - 1)].$$

$M$  is obtained by counting the  $T - 1$  distinct LC probabilities and, for each latent class, the  $I - 1$  distinct conditional probabilities associated with the categories of variable  $A$ , the  $J - 1$  distinct conditional probabilities associated with  $B$ , and so on. Because probabilities sum to 1, the probability associated with one category of each variable is redundant (and hence not counted as a *distinct* parameter): It can be obtained as 1 minus the sum of the others.

In situations involving sparse data, the chi-squared distribution should not be used to compute the  $p$ -value because  $L^2$  would not be well approximated. Instead, the bootstrap approach can be used to estimate  $p$  (Langeheine, Pannekoek, & Van de Pol, 1996). Sparse data often occur when the number of observed variables or the number of categories of these variables is large. In such cases, the total number of cells in the resulting multiway frequency table will be large relative to the sample size, resulting in many empty cells. This situation is illustrated below with a data example. Sparse data also result when LC models are extended to include continuous variables, which is illustrated in the last section.

An alternative approach to assessing model fit in the case of sparse data uses an information criterion weighting both model fit and parsimony. Such measures, such as Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), are especially useful in comparing models. The most widely used in LC analysis is the BIC statistic, which can be

<sup>1</sup>According to the general rule, if it turns out that  $df < 0$ , the model is not identifiable, which means that unique estimates are not available for all parameters. For example, for  $I = J = K = L = 2$ ,  $df = -4$  for  $T = 4$ , which means that the four-class model is not identifiable. In some cases, however, this general counting rule may yield  $df > 0$ , yet the model may still not be identifiable. For example, Goodman (1974b) shows that in this situation of four dichotomous variables, the three-class model is also unidentifiable despite the fact that the counting rule yields  $df = 1$ . See also Note 3.

defined as follows:  $BIC_{L^2} = L^2 - \ln(N)df$  (Raftery, 1986). A model with a lower BIC value is preferred over a model with a higher BIC value. A more general definition of BIC is based on the log-likelihood (LL) and the number of parameters ( $M$ ) instead of  $L^2$  and  $df$ ; that is,

$$BIC_{LL} = -2LL + \ln(N)M.$$

Again, a model with a lower BIC value is preferred over a model with a higher BIC value.<sup>2</sup>

If the baseline model ( $H_0$ ) provides an adequate fit to the data, no LC analysis is needed because there is no association among the variables to be explained. In most cases, however,  $H_0$  will not fit the data, in which case  $L^2(H_0)$  can serve as a baseline measure of the total amount of association in the data. This suggests a third approach for assessing the fit of LC models by comparing the  $L^2$  associated with LC models, for which  $T > 1$ , with the baseline value  $L^2(H_0)$  to determine the percent reduction in  $L^2$ . Because the total association in the data may be quantified by  $L^2(H_0)$ , the percent reduction measure represents the total association explained by the model. This less formal approach can complement the more statistically precise  $L^2$  and BIC approaches.

As an example of how these measures are used, suppose that the  $L^2$  suggests that a three-class model falls short of providing an adequate fit to some data (say,  $p = .04$ ) but explains 90% of the total association. Moreover, suppose a four-class model is the simplest model that fits according to the  $L^2$  statistic, but this model only explains 91% of the association. In this case, it may be that, on practical grounds, the three-class model is preferable because it explains almost as much of the total association.

#### 10.2.1.1. Example: Survey Respondent Types

We will now consider a first example that illustrates how these tools are used in practice. It is based on the analysis of four variables from the 1982 General Social Survey given by McCutcheon (1987) to illustrate how traditional LC modeling can be used to study the different types of survey respondents. Two of the variables ascertain the respondent's opinion regarding (A) the purpose of surveys and (B) how accurate they are, and the others are evaluations made by the interviewer of (C) the respondent's levels of understanding of the survey questions and (D) cooperation shown in

answering the questions. McCutcheon initially assumed the existence of two latent classes corresponding to "ideal" and "less than ideal" types.

The study included separate samples of White and Black respondents. Beginning with an analysis of the White sample, McCutcheon (1987) later included data from the Black sample to illustrate a two-group LC analysis. We will use these data to introduce the basics of traditional LC modeling and to illustrate several recent developments that have been made over the past decade. These include allowing for specific local dependencies (Section 10.3.1), the usage of LC factor models (Section 10.3.2), and the inclusion of covariates as well as the methodology for making multigroup comparisons (Sections 10.3.3 and 10.3.4).

Traditional exploratory LC analysis begins by fitting the null model  $H_0$  to the sample of White respondents. Because  $L^2(H_0) = 257.3$  with  $df = 29$  (see Table 10.1), the amount of association (nonindependence) that exists in these data is too large to be explained by chance, so the null model must be rejected ( $p < .001$ ) in favor of  $T > 1$  classes.

Next, we consider McCutcheon's (1987) two-class model ( $H_1$ ). For this model, the  $L^2$  is reduced to 79.5,<sup>3</sup> a 69.1% reduction from the baseline model, but still much too large to be acceptable with  $df = 22$ . Thus, we increment  $T$  by 1 and estimate model  $H_{2C}$ , the three-class model. This model provides a further substantial reduction in  $L^2$  to 22.1 (a 91.5% reduction over the baseline) and also provides an adequate overall fit ( $p > .05$ ). Table 10.1 shows that the four-class LC model provides some further improvement. However, the BIC statistic, which takes parsimony into account, suggests that the three-class model is preferred over the four-class model (see Table 10.1).

The parameter estimates obtained from the three-class model are given in the left-most portion of Table 10.2. The classes are ordered from largest to smallest. Overall, 62% are estimated to be in Class 1, 20% in Class 2, and the remaining 18% in Class 3. Analogous to factor analysis, in which names are assigned to the factors based on an examination of the "factor loadings," names may be assigned to the latent classes based on the estimated conditional probabilities. Like factor loadings, the conditional probabilities provide the measurement *structure* that defines the latent classes.

<sup>2</sup>The two formulations of BIC differ only with respect to a constant. More precisely,  $BIC_{L^2}$  equals  $BIC_{LL}$  minus the  $BIC_{LL}$  corresponding to the saturated model.

<sup>3</sup>This value differs slightly from the value 79.3 reported in McCutcheon (1987) because our models include a Bayes constant set equal to 1 to prevent boundary solutions (estimated model probabilities equal to zero). For further information on Bayes constants, see the technical appendix of the Latent GOLD 3.0 manual (Vermunt & Magidson, 2003, or [www.latentclass.com](http://www.latentclass.com)).

**Table 10.1** Results From Various Latent Class Models Fit to the General Social Survey 1982 Data

Model		$BIC_{LL}$	$L^2$	$df$	$p$ -Value	% Reduction in $L^2(H_0)$
<b>Sample of White respondents</b>						
<i>Traditional</i>						
$H_0$	One-class	5787.0	257.3	29	$2.0 \times 10^{-38}$	0.0
$H_{1C}$	Two-class	5658.9	79.5	22	$2.0 \times 10^{-8}$	69.1
$H_{2C}$	Three-class	5651.1	22.1	15	.11	91.4
$H_{3C}$	Four-class	5685.3	6.6	8	.58	97.4
<i>Nontraditional</i>						
$H_{1C+}$	Two-class + { $CD$ } direct effect	5606.1	12.6	20	.89	95.1
$H_{2F}$	Basic two-factor	5640.1	11.1	15	.75	95.7
<b>Sample of Black respondents</b>						
<i>Traditional</i>						
$H'_0$	One-class	2402.1	112.1	29	$1.0 \times 10^{-11}$	0.0
$H'_1$	Two-class	2389.6	56.9	22	.00006	49.2
$H'_{2C}$	Three-class	2393.8	18.3	15	.25	83.7
$H'_{3C}$	Four-class	2427.6	9.4	8	.31	91.6
<i>Nontraditional</i>						
$H'_{1C+}$	Two-class + { $CD$ } direct effect	2360.2	15.2	20	.77	86.4
$H'_{2F}$	Basic two-factor	2387.0	11.5	15	.72	89.7
<b>Full sample (multiple-group analysis)</b>						
<i>Traditional</i>						
$M_0$	One-class	8185.1	400.0	64	$4.3 \times 10^{-50}$	0
$M_1$	Two-class	8013.8	169.5	56	$2.4 \times 10^{-13}$	57.6
$M_{2C}$	Three-class unrestricted (complete heterogeneity)	8077.4	40.4	30	.10	89.9
$M_{2CR}$	Three-class restricted (partial homogeneity)	7953.0	49.4	48	.42	87.7
$M_{2CRR}$	Three-class restricted (complete homogeneity)	7962.1	73.3	50	.02	81.7
$M_{3CR}$	Four-class restricted (partial homogeneity)	7989.8	27.0	40	.94	93.3
<i>Nontraditional</i>						
$M_{2F}$	Basic two-factor unrestricted	8059.6	22.6	30	.83	94.4
$M_{2FR}$	Basic two-factor restricted	7934.9	31.3	48	.97	92.2

NOTE: BIC = Bayesian information criterion.

McCutcheon (1987) assigned the name “ideal” to latent Class 1, reasoning as follows:

The first class corresponds most closely to our anticipated ideal respondents.

Nearly 9 of 10 in this class believed that surveys “usually serve a good purpose”; 3 of 5 expressed a belief that surveys are either “almost always right” or “right most of the time”; 19 of 20 were evaluated by the interviewer as “friendly and interested” during the interview; and nearly all were evaluated by the interviewer as having a good understanding of the survey questions. (p. 34)

He named the other classes “believers” and “skeptics” based on the interpretations of the

corresponding conditional probabilities for those classes.

### 10.2.2. Testing the Significance of Effects

The next step in a traditional LC analysis is to delete from the model any variable that does not exhibit a significant difference between the classes. For example, to test whether to delete variable  $A$  from a  $T$ -class model, one would test the null hypothesis that the distribution over the  $I$  categories of  $A$  is identical within each class  $t$ :

$$\pi_{i1}^{A|X} = \pi_{i2}^{A|X} = \dots = \pi_{iT}^{A|X} \text{ for } i = 1, 2, \dots, I.$$

**Table 10.2** Parameter Estimates for the Three-Class Latent Class (LC) Model by Sample

	White Sample			Black Sample		
	Class 1 Ideal	Class 2 Believers	Class 3 Skeptics	Class 1 Ideal	Class 2 Believers	Class 3 Skeptics
LC probabilities	0.62	0.20	0.18	0.49	0.33	0.18
Conditional probabilities						
(A) PURPOSE						
Good	0.89	0.92	0.16	0.87	0.91	0.19
Depends	0.05	0.07	0.22	0.08	0.04	0.17
Waste	0.06	0.01	0.62	0.05	0.05	0.65
(B) ACCURACY						
Mostly true	0.61	0.65	0.04	0.54	0.65	0.01
Not true	0.39	0.35	0.96	0.46	0.35	0.99
(C) UNDERSTANDING						
Good	1.00	0.32	0.75	0.95	0.37	0.68
Fair, poor	0.00	0.68	0.25	0.05	0.63	0.32
(D) COOPERATION						
Interested	0.95	0.69	0.64	0.98	0.56	0.64
Cooperative	0.05	0.26	0.26	0.01	0.37	0.25
Impatient/hostile	0.00	0.05	0.10	0.00	0.07	0.11

To implement this test, we make use of the relationship between the conditional response probabilities and the log-linear parameters (see, e.g., Formann, 1992; Haberman, 1979; Heinen, 1996):

$$\pi_{ii}^{AX} = \frac{\exp(\lambda_i^A + \lambda_i^{AX})}{\sum_{i'=1}^I \exp(\lambda_{i'}^A + \lambda_{i'}^{AX})}$$

Standard log-linear modeling techniques can then be used to test the null hypothesis, reexpressed in terms of the log-linear parameters associated with the  $AX$  relationship:

$$\lambda_{i1}^{AX} = \lambda_{i2}^{AX} = \dots = \lambda_{iT}^{AX} = 0 \text{ for } i = 1, 2, \dots, I.$$

One way to test for significance of the four indicators in our three-class model is by means of a  $L^2$  difference test, where  $\Delta L^2$  is computed as the difference between the  $L^2$  statistics obtained under the *restricted* and *unrestricted* three-class models, respectively. The  $\Delta L^2$  values obtained by setting the association parameters corresponding to one of the indicators to zero were 145.3, 125.4, 61.3, and 101.1, for  $A$ ,  $B$ ,  $C$ , and  $D$ , respectively. These numbers are higher than of the corresponding Wald statistics, which took on the values 29.6, 8.4, 7.4, and 19.0. This is because the latter test is uniformly less powerful than the  $\Delta L^2$  statistic. Under the assumption that the unrestricted model is true, both statistics are distributed asymptotically as chi-square with  $df = (I - 1) \cdot (T - 1)$ , where  $I$  denotes the number of categories in the nominal variable. The encountered values show that each of the four indicators

included in the model is significantly related to class membership.

### 10.2.3. Classification

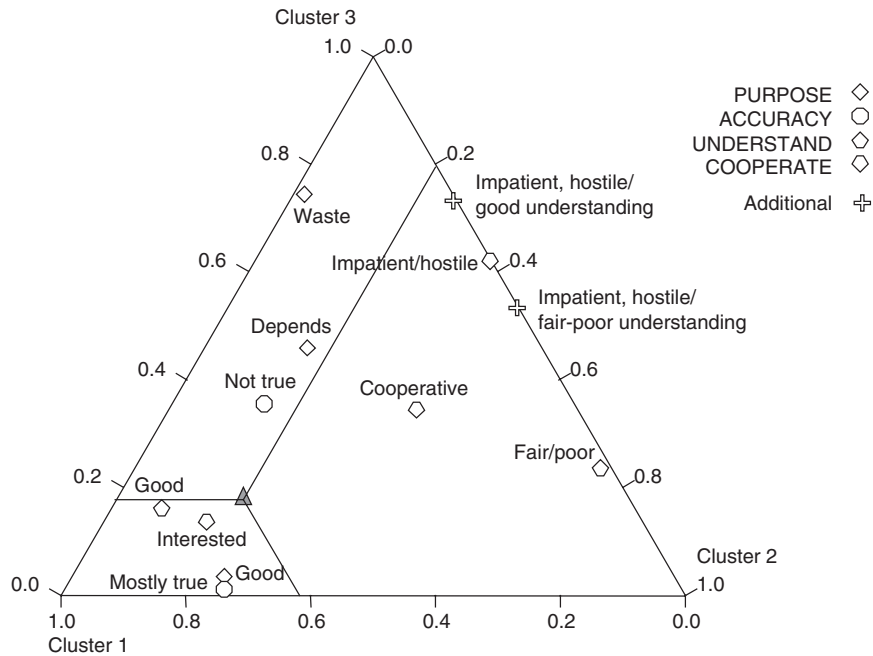
The final step in a traditional LC analysis is to use the results of the model to classify cases into the appropriate latent classes. For any given response pattern  $(i, j, k, 1)$ , estimates for the posterior membership probabilities can be obtained using Bayes theorem as follows:

$$\hat{\pi}_{ijkl}^{X|ABCD} = \frac{\hat{\pi}_{ijkl}^{ABCDX}}{\sum_{t=1}^T \hat{\pi}_{ijkl}^{tABCDX}}, \quad t = 1, 2, \dots, T \quad (2)$$

where the numerator and denominator in equation (2) are obtained by substituting the model parameter estimates in place of the corresponding parameters in equation (1).

Magidson and Vermunt (2001) and Vermunt and Magidson (2002) refer to this kind of model as a LC *cluster* model because the goal of classification into  $T$  homogeneous groups is identical to that of cluster analysis. In contrast to an ad hoc measure of distance used in cluster analysis to define homogeneity, LC analysis defines homogeneity in terms of probabilities. As indicated by equation (1), cases in the same latent class are similar to each other because their responses are generated by the same probability distribution.

Cases are then assigned to the class for which the posterior probability is highest (i.e., the modal

**Figure 10.1** Barycentric Coordinate Display for Three-Class Model

class). For example, according to the three-class LC model, someone with response pattern  $A = 1$  (PURPOSE = “good”),  $B = 1$  (ACCURACY = “mostly true”),  $C = 1$  (UNDERSTANDING = “good”), and  $D = 1$  (COOPERATION = “interested”) has posterior membership probabilities equal to 0.92, 0.08, and 0.00. This means that such a person is assigned to the first class.

#### 10.2.4. Graphical Displays

Because for any given response pattern  $(i, j, k, l)$ , the  $T$ -class membership probabilities sum to 1, only  $T - 1$  such probabilities are required as the probability of belonging to the remaining class can be obtained from the others. Hence, the class membership probabilities  $\hat{\pi}_{ijkl}^{X|ABCD}$  can be used to position each response pattern in  $T - 1$  dimensional space, and for  $T = 3$ , various two-dimensional barycentric coordinate displays can be produced.

Rather than plotting every one of the many response patterns, instructive plots of the kind used in correspondence analysis can be produced, where points are plotted for each category of each variable as well as other meaningful aggregations of these probabilities (Magidson & Vermunt, 2001).

Figure 10.1 depicts the corresponding barycentric coordinate display under the three-class LC model. Points are plotted for each category of each of the four variables in our example. Because these points contain information equivalent to the LC parameter estimates (Van der Heijden, Gilula, & Van der Ark, 1999), this type of plot provides a graphical alternative to the traditional tabular display of parameter estimates and can yield new insights into data. Also displayed in Figure 10.1 are two additional aggregations associated with the response categories UNDERSTANDING = “good” and “fair, poor” ( $k = 1, 2$ ) among those for whom COOPERATION = “hostile/impatient” ( $l = 3$ ).

The horizontal dimension of the plot corresponds to differences between McCutcheon’s (1987) “ideal” and “believer” types (latent Classes 1 and 2). We see that the categories of the variable  $C$  tend to spread out along this dimension. Respondents showing “good” understanding are most likely to belong to the ideal class (the corresponding symbol is plotted closest to the lower left vertex that represents Class 1), whereas those showing only “fair or poor” understanding are plotted closest to the lower right vertex that represents Class 2.

Differences along the vertical dimension of the plot are best shown by the categories of  $A$  and  $B$ . For

**Table 10.3** Descriptive Information and Parameter Estimates From Three-Class and Two-Factor Latent Class (LC) Models Obtained With the Landis and Koch (1977) Data

	Descriptive Information			Two Factors (Joint Probabilities)						
				Three Classes			Factor 1=1 (True Negative)		Factor 1 = 2 (True Positive)	
	% of Slides Rated Positive	% of Ratings That Agree With		Class 1	Class 2	Class 3	Factor 2 = 1 (Negative Bias)	Factor 2 = 2 (Positive Bias)	Factor 2 = 1 (Negative Bias)	Factor 2 = 1 (Positive Bias)
Class size		5+	6+	0.44	0.37	0.18	0.36	0.19	0.30	0.16
F	21	64	58	<b>0.47</b>	0.00	0.00	0.00	0.01	<b>0.23</b>	0.86
D	27	70	62	<b>0.59</b>	0.00	0.06	0.00	0.05	<b>0.37</b>	0.92
C	38	80	64	0.85	0.00	0.01	0.00	0.00	0.83	0.83
A	56	82	64	1.00	0.06	0.51	0.06	<b>0.47</b>	0.99	1.00
G	56	85	66	1.00	0.00	0.63	0.01	<b>0.58</b>	0.99	1.00
E	60	80	64	1.00	0.06	0.76	0.06	<b>0.72</b>	0.99	1.00
B	67	75	61	0.98	0.15	0.99	0.13	<b>0.99</b>	0.97	1.00

NOTE: False-negative and false-positive rates are highlighted in bold.

example, respondents agreeing that the purpose of surveys is “good” are plotted close to the lower left (Class 1) vertex. Those who say “it depends” are plotted somewhat midway between the Class 1 and Class 3 (top) vertex. Those who say “it’s a waste of time and money” are most likely to be in Class 3 and are positioned near the top vertex. The fact that the positioning of categories for both *A* and *B* spreads out over the vertical dimension suggests a high degree of association between these variables. In contrast, the categories of *C* are spread over the horizontal dimension, suggesting that the association between *C* and the two variables *A* and *B* is close to nil.

The categories of the variable *D* form an interesting diagonal pattern. Respondents showing they are “interested” in the questions are most likely to be in Class 1 (“ideal”), whereas those who are only “co-operative” or exhibit “impatience/hostility” are plotted closer to Classes 2 and 3. This suggests the hypothesis that impatience and hostility may arise for either of two different reasons: (a) disagreement that surveys are accurate and serve a good purpose (indicated by the vertical dimension of the plot) and/or (b) lack of understanding (indicated by the horizontal dimension).

The additional points plotted deal with the relationship between variables *C* and *D*. The positioning of these points suggest that among impatient/hostile respondents, those who show good understanding of the questions tend to be more in Class 3, whereas those whose understanding is fair/poor tend to be about equally likely to be in Class 2 or 3.

We will revisit these data and obtain further insights later when we examine an alternative *nontraditional* two-dimensional LC model, the two-factor LC model.

#### 10.2.4.1. Example: Sparse Multirater Agreement Data

We next consider an example with sparse data in which seven pathologists each classified 118 slides as to the presence or absence of carcinoma in the uterine cervix (Landis & Koch, 1977) that was also analyzed by Agresti (2002). LC modeling will be used here to estimate the false-positive and false-negative rates for each pathologist and to use multiple ratings to distinguish between slides that indicate carcinoma and those that do not (for similar medical applications, see Rindskopf & Rindskopf, 1986; Uebersax & Grove, 1990). The second column of Table 10.3 shows that the raters vary from classifying only about 1 of every 5 slides as positive (Rater D) to classifying more than 2 of every 3 as positive (Rater B). The next two columns indicate for which percentage of slides the ratings agree among five or more and six or more raters. This information shows that agreement is highest among Raters C, A, G, and E.

As a starting point, Agresti (2002) formulated a model containing two latent classes, in an attempt to confirm the hypothesis that slides are either “true positive” or “true negative.” The assumption of local independence in the two-class model means that rater agreement is caused solely by the differing

**Table 10.4** Results From Various Latent Class (LC) Models Fit to Landis and Koch (1977) Data

Model		$BIC_{LL}$	$L^2$	Bootstrap $p$ -Value	% Reduction in $L^2(H_0)$
<i>Traditional</i>					
$H_0$	One-class	1082.3	476.8	.00	0.0
$H_1$	Two-class	707.9	64.2	.00	86.5
$H_{2C}$	Three-class	699.6	17.7	.49	96.3
$H_{3C}$	Four-class	729.4	9.3	.79	98.0
<i>Nontraditional</i>					
$H_{2C+}$	Three-class + $\{DF\}$ direct effect	698.0	11.3	.83	97.6
$H_{2FR}$	Restricted basic two-factor	688.4	11.3	.90	97.6

NOTE: BIC = Bayesian information criterion.

characteristics between these two types of slides. That is, given that a slide is in the class of “true positive” (“true negative”), any similarities and differences between raters represent pure error. However, in his analysis of these data, he found that three classes were necessary to obtain an acceptable fit.

Although there are  $2^7 = 128$  possible response patterns, because of the large amount of inter-rater agreement, 108 of these patterns were not observed at all. As mentioned above, sparse data such as these cause a problem in testing model fit because the  $L^2$  statistic does not follow a chi-square distribution. For this reason, Agresti (2002) simply alluded to the obvious discrepancy between the expected frequencies estimated under the two-class model and the observed frequencies and speculated that this model does not provide an adequate fit to these data. He then compared estimates obtained from the three-class model and suggested that the fit of this model was adequate.

We report the bootstrap  $p$ -value in Table 10.4, which confirms Agresti’s (2002) speculation that the fit of the two-class model is poor and that of the three-class model is adequate. It also shows that the three-class model is preferred over the four-class model according to the BIC criteria.

The parameter estimates obtained with the three-class model are given in the middle portion of Table 10.3. The largest class (44%) refers to slides that all pathologists (except for D and F) almost always agree show carcinoma (“true positive”). Class 2 (37%) refers to slides that all pathologists (except occasionally B) agree show no carcinoma. The remaining class of slides (18%) shows considerable disagreement between pathologists—B, E, and G usually diagnose carcinoma, whereas C, D, and F rarely do, and A diagnoses carcinoma half the time.

If we assume that Class 1 represents cases of true carcinoma, the results reported in Table 10.3 show that those pathologists who rated the *fewest* slides as positive (D and F) have the highest false-negative rates (42% and 53%, respectively, highlighted in bold). Similarly, under the assumption that Class 2 represents cases free from carcinoma, the results show that the pathologist who rated the *most* slides as positive, Rater B, shows a false-positive rate (15%) that is substantially larger than the other pathologists.

The traditional model-fitting strategy requires us to reject our two-class hypothesis in favor of a three-class alternative in which the third latent class consists of slides that cannot be classified as either “true positive” or “true negative” for cancer. Next we consider some nontraditional LC models that provide classification of each slide according to its likelihood of carcinoma. In particular, we will show that a two-factor LC model provides an attractive alternative whereby Factor 1 classifies all slides as either “true positive” or “true negative,” and Factor 2 classifies slides according to a tendency for ratings to be biased toward false-positive or false-negative error.

### 10.3. NONTRADITIONAL LATENT CLASS MODELING

Rejection of a traditional  $T$ -class LC model for lack of fit means that the local independence assumption does not hold with  $T$  classes. In such cases, the traditional LC model-fitting strategy is to fit a  $T + 1$  class model to the data. In both of our examples, theory supported a two-class model but because this model failed to provide an adequate fit, we formulated a three-class model. In this section, we consider some alternative



strategies for modifying a model. In both cases, we will see the nontraditional alternatives lead to models that are more parsimonious than traditional models, as well as models that are more congruent with our initial hypotheses. The alternatives considered are as follows:

1. adding one or more direct effects,
2. deleting one or more items,
3. increasing the number of latent variables.

Alternative 1 is to include “direct-effect” parameters in the model (Hagenaars, 1988) that account for the residual association between the observed variables that is responsible for the local dependence. This approach is particularly useful when some external factor, unrelated to the latent variable, creates an irrelevant association between two variables. Examples of such external factors include similar question wording used in two survey items, as well as two raters using the same incorrect criterion in evaluating slides.

Alternative 2 also deals with the situation in which two variables are responsible for some local dependency. In such cases, rather than add a direct effect between two variables, it may make more sense to eliminate the dependency by simply deleting one of the two items. This variable reduction strategy is especially useful in situations in which there are *many* redundant variables.

Alternative 3 is especially useful when a group of several variables accounts for a local dependency. Magidson and Vermunt (2001) show that by increasing the dimensionality through the addition of latent variables rather than latent classes, the resulting LC *factor* model often fits data substantially better than the traditional LC cluster models having the same number of parameters. In addition, LC factor models are identified in some situations when the traditional LC model is not.<sup>4</sup>

In the next section, we introduce a diagnostic statistic called the *bivariate residual* (BVR) and illustrate its use to develop some nontraditional alternative models for our two data examples. The BVR helps pinpoint those *bivariate* relationships<sup>5</sup> that fail to be adequately

**Table 10.5** Values for Bivariate Residuals Obtained Under Various Models for the Sample of White Respondents

Two-Way Table	Model					
	$H_0$	$H_1$	$H_{2C}$	$H_{3C}$	$H_{2C+}$	$H_{2F}$
{AB}	61.6	0.1	0.1	0.0	0.0	0.0
{AC}	0.5	0.7	0.1	0.0	0.2	0.0
{AD}	10.6	0.0	0.1	0.0	0.2	0.1
{BC}	0.3	1.1	0.0	0.0	0.0	0.0
{BD}	8.6	0.4	0.3	0.2	0.2	0.4
{CD}	43.4	32.3	2.4	0.0	0.0	0.2

explained by the LC model and can help determine which of the three alternative strategies to employ. We will see that even in situations when the  $L^2$  statistic reports that the model provides an adequate *overall* fit, the fit in one or more two-way tables may not be adequate and may indicate a flaw or weakness in the model.

### 10.3.1. Bivariate Residuals and Direct Effects

A formal measure of the extent to which the observed association between two variables is reproduced by a model is given by the BVR statistic (Vermunt & Magidson, 2003). Each BVR corresponds to a Pearson chi-square statistic (divided by the degrees of freedom) in which the observed frequencies in a two-way cross-tabulation of the variables are contrasted with those expected counts estimated under the corresponding LC model.<sup>6</sup> A BVR value substantially larger than 1 suggests that the model falls somewhat short of explaining the association in the corresponding two-way table.

#### 10.3.1.1. Example: Survey Respondent Types (Continued)

Table 10.5 reports BVRs for each variable pair under each of several models estimated in our first example. Because model  $H_0$  corresponds to the model of mutual independence, each BVR for this model provides a measure of the overall association in the corresponding observed two-way table; that is, each BVR equals the usual Pearson chi-square statistic used to test for independence in the corresponding two-way table divided by the degrees of freedom. The results show that except for the nonsignificant relationships in

<sup>4</sup>For example, with four dichotomous variables, an LC two-factor model (composed of four latent classes) is identified, whereas a traditional three-class model is not (Goodman, 1974b).

<sup>5</sup>Traditional factor analysis, through the assumption of multivariate normality, limits its focus to bivariate relationships (i.e., the correlations) because higher order relationships are assumed not to exist. In contrast, LC models do not make strict distributional assumptions and hence attempt to explain higher order associations as well. Nevertheless, the two-way (bivariate) associations are generally the most prominent, and the ability to pinpoint specific two-way tables in which lack of fit may be concentrated can be useful in suggesting alternative models.

<sup>6</sup>These residuals are similar to Lagrangian statistics. A difference is that they are limited-information fit measures: Dependencies with parameters corresponding to other items are not taken into account.

the {AC} and {BC} tables, all of the remaining BVRs are quite large, attesting to several significant associations (local dependencies) that exist among these variables. The BVR is especially large for {AB} and for {CD}. For example, in Table {CD}, a Pearson chi-square test confirms that the observed relationship is highly significant ( $\chi^2 = 86.8$ ,  $df = 2$ ,  $p < .001$ ;  $BVR = 86.8/2 = 43.4$ ).

Under the two-class model ( $H_1$ ), note that the BVRs are all near or less than 1, except for one very large value of 32.3 for {CD}. This suggests that the overall lack of fit for this model can be traced to this single large BVR. The traditional way to account for the lack of fit is by adding another latent class. However, Table 10.5 shows that even after the addition of a third class, the BVR for {CD} under the three-class model  $H_{2C}$  remains unacceptably high ( $BVR = 2.4$ ). Although the inclusion of the third class does add a second dimension that causes the overall fit to be adequate, it is not until we add a fourth class (model  $H_{3C}$ ) that all BVRs are at acceptable levels.

Below, we consider the alternative approach of adding a “direct effect” to the model to account for the residual correlation. In addition, we consider use of the two-factor LC model and further explore the differences between the three- and four-class models.

10.3.1.2. Example: Sparse Multirater Agreement Data (Continued)

Turning now to our second example, Table 10.6 shows that all of the BVRs under the one-class model of mutual independence (model  $H_0$ ) are very large,<sup>7</sup> indicating that the amount of agreement between each pair of raters is highly significant. Under the two-class model, many BVRs remain large. Although the three-class model provides an acceptable overall fit to these data, again we see that there is a single BVR that remains unacceptably large— $BVR = 4.5$  for Raters D and F, the two pathologists who rated the fewest slides positive (recall Table 10.3). This large BVR suggests that Raters D and F may be using some rating criterion not shared by the other raters.

To account for this large residual association, we will use nontraditional Alternative 1 and modify the three-class model by adding the D through F direct-effect parameter  $\lambda^{DF}$  into the model (Hagenaars, 1988; for a slightly different formulation, see Uebersax, 1999). Formally, this new model,  $H_{2C+}$  is expressed as

$$\pi_{ijklmpt} = \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{mt}^{E|X} \pi_{lpt}^{DF|X},$$

<sup>7</sup>The smallest BVR under model  $H_0$  is 20.8, which occurs in table {EF}.

**Table 10.6** Bivariate Residuals Obtained Under Various Models for Landis and Koch (1977) Data

Two-Way Table <sup>a</sup>	Model					
	Traditional			Nontraditional		
	$H_0$	$H_1$	$H_{2C}$	$H_{2C+}$	$H_{2FR}$	$H_{2FRC}$
{BE}	66.4	8.4	0.0	0.0	0.0	0.1
{DF}	38.0	7.2	4.5	0.0	0.0	0.0
{BG}	66.7	5.2	0.0	0.0	0.1	0.1
{EG}	77.2	3.3	0.1	0.1	0.2	0.2
{AB}	54.5	1.7	0.1	0.0	0.1	0.1
{CF}	28.0	1.3	0.0	0.0	0.0	0.0
{CE}	47.7	1.1	0.1	0.1	0.2	0.1
{DE}	24.5	0.0	0.7	0.6	0.6	1.2

a. These are the two-way tables for which the bivariate residuals were larger than 1 under any of the reported models (other than  $H_0$ ).

where the probabilities  $\pi_{lpt}^{DF|X}$  are constrained as follows:

$$\pi_{lpt}^{DF|X} = \frac{\exp(\lambda_l^D + \lambda_p^F + \lambda_{lp}^{DF} + \lambda_{lt}^{DX} + \lambda_{pt}^{FX})}{\sum_{l=1}^L \sum_{p=1}^P \exp(\lambda_l^D + \lambda_p^F + \lambda_{lp}^{DF} + \lambda_{lt}^{DX} + \lambda_{pt}^{FX})}$$

By relaxing the local independence assumption between Raters D and F, model  $H_{2C+}$  is able to account for the excessive association between D and F that is not explainable by the latent classes. The  $\Delta L^2$  test shows that inclusion of the direct-effect parameter provides a significant improvement over the traditional model  $H_{2C}$  ( $\Delta L^2 = 17.7 - 11.3 = 6.4$ ;  $p = .01$ ).

From a practical perspective, models  $H_{2C}$  and  $H_{2C+}$  do not differ much as both models assign the 118 slides to the same classes under the modal assignment rule. This occurs despite the fact that model  $H_{2C+}$  gives D and F less weight than model  $H_{2C}$  during the computation of the posterior probabilities. The primary benefit of model  $H_{2C+}$  is to suggest the possibility that Raters D and F share a bias when evaluating Class 1 slides, those slides that D and F often rate negative but that the other pathologists almost always rate positive (recall Table 10.3). The implication of including the direct effect is that model  $H_{2C+}$  provides higher predictions of agreement between Raters D and F than model  $H_{2C}$  on Class 1 slides.<sup>8</sup>

<sup>8</sup>Because model  $H_{2C}$  assumes local independence, the expected probability of both raters agreeing that a given Class 1 slide is free from cancer can be computed by multiplying the corresponding conditional probabilities. Using the estimates from Table 10.3, the probability of both agreeing that a Class 1 slide is negative is  $.42 \times .53 = .22$ , and similarly, the probability of both agreeing that it was positive is  $.59 \times .47 = .28$ . In contrast, model  $H_{2C+}$  predicts higher probabilities (.31 and .35, respectively) for Raters D and F agreeing in both cases. Under the assumption that Class 1 slides are “true positive,” the results from model  $H_{2C+}$  mean that Raters D and F both tend to share a bias toward committing a false-negative error.

Returning to the first data example for a moment, we might now expect to find similar insights by the inclusion of the direct-effect parameters,  $\lambda_{kl}^{CD}$ , in the two-class model. Table 10.1 shows that this model ( $H_{1C+}$ ) provides a good fit to the data. However, under this model, the parameter measuring the contribution of  $C$  to the latent classes is no longer significant, and therefore  $C$  can be deleted from the LC model completely. As this amounts to deleting an association simply because it could not be explained by a model with two latent classes, Alternative 1 does not provide a desirable solution here.

### 10.3.2. LC Factor Models

Next we consider Alternative 3, in which we use LC factor models to include more than one latent variable in the model. LC factor models were proposed as a general alternative to the traditional exploratory LC modeling by Magidson and Vermunt (2001). For both examples, the results (given in Table 10.1 and Table 10.4) show that a two-factor model is preferable to the other models. We shall see that the two-factor model is actually a restricted four-class model. In both cases, the fit is almost as good as the (unrestricted) four-class solution but is more parsimonious and parameterized in a manner that allows easier interpretation of the results.

LC factor models were initially proposed by Goodman (1974a) in the context of confirmatory latent class analysis. Certain traditional LC models containing four or more classes can be interpreted in terms of two or more component latent variables by treating those components as a joint variable (see, e.g., Hagenars, 1990; McCutcheon, 1987). For example, a latent variable  $X$  consisting of  $T = 4$  classes can be reexpressed in terms of two dichotomous latent variables  $V = \{1, 2\}$  and  $W = \{1, 2\}$  using the following correspondence:

	$W = 1$	$W = 2$
$V = 1$	$X = 1$	$X = 2$
$V = 2$	$X = 3$	$X = 4$

Thus,  $X = 1$  corresponds with  $V = 1$  and  $W = 1$ ,  $X = 2$  with  $V = 1$  and  $W = 2$ ,  $X = 3$  with  $V = 2$  and  $W = 1$ , and  $X = 4$  with  $V = 2$  and  $W = 2$ .

Formally, for four nominal variables, the four-class LC model can be reparameterized as a LC

factor model with two dichotomous latent variables as follows:

$$\begin{aligned}\pi_{ijklrs} &= \pi_{rs}^{VW} \pi_{ijklrs}^{ABCD|VW} \\ &= \pi_{rs}^{VW} \pi_{irs}^{A|VW} \pi_{jrs}^{B|VW} \pi_{krs}^{C|VW} \pi_{lrs}^{D|VW}.\end{aligned}$$

Magidson and Vermunt (2001) consider various restricted factor models. They use the term *basic* LC factor models to refer to certain LC models that contain two or more dichotomous latent variables that are mutually independent of each other and that exclude higher order interactions from the conditional response probabilities. Such a model is analogous to the approach of traditional factor analysis in which multiple latent variables are used to model multidimensional relationships among manifest variables.

It turns out that by formulating the model in terms of  $R$  mutually independent, dichotomous latent factors, the basic LC factor model has the same number of distinct parameters as a traditional LC model with  $R+1$  classes. That is, the LC factor parameterization allows specification of a  $2^R$ -class model with the same number of parameters as a traditional LC model with only  $R+1$  classes! This offers a great advantage in parsimony over the traditional  $T$ -class model as the number of parameters is greatly reduced by natural restrictions.

As mentioned previously, the basic two-factor model provides an excellent fit to both of our example data sets. For the first example, Table 10.1 shows that this model (model  $H_{2F}$ ) is preferred over any of the LC cluster models according to the BIC. In addition, this model explains all bivariate relationships in the data (see Table 10.5). We will interpret the results from this model in the next section in conjunction with a more extensive analysis, including both the White and Black sample.

#### 10.3.2.1. Example: Sparse Multirater Agreement Data (Continued)

Regarding our second example, Table 10.4 shows that the basic two-factor model is preferred over all the other models according to the BIC criteria. The right-most portion of Table 10.3 provides the parameter estimates<sup>9</sup> that we used to name the factors. These are joint latent class and conditional response probabilities for combinations of factor levels. We assigned the names “true negative” and “true positive” to Levels 1 and 2 of Factor 1, respectively. Each of these levels is split again into two levels by Factor 2, which we

<sup>9</sup>The two-factor model in Table 3 was further restricted by setting the effect of indicator  $C$  on Factor 2 to zero because this effect was not significant.

named “tendency toward ratings bias.” We named the two levels of Factor 2 “tend to negative bias” and “tend to positive bias,” respectively.

Comparing the four factor cells (right-most portion of Table 10.3) to the classes in the three-class model (middle portion of Table 10.3), we see the following similarities. First, note that Class 1 of the three-class solution (representing 44% of the slides mostly rated positive) corresponds primarily to Factor 1, Level 2 slides (those named “true positive”), which account for 46% of all slides. These “true-positive” slides are divided according to Factor 2 into cell (2, 1), accounting for 30% of all slides, and cell (2, 2), accounting for 16% of the slides. Note that the former slides show a clear tendency toward a false-negative error, especially among Raters D and F.

Next, notice the similarity between Class 2 of the three-class solution, representing 37% of the slides rated mostly negative, and factor cell (1, 1), accounting for 36% of the slides rated mostly negative. In addition, from Table 10.3, we can also see the strong similarity between Class 3 of the three-class solution and factor cell (1, 2), identified in the table as “true-negative” slides that are prone to false-positive error, especially by Raters A, B, E, and G.

In conclusion, we have shown that the two-factor LC model fits better than the traditional three-class model and offers two substantive advantages. First, it provides a clear way to classify slides as “true positive” or “true negative.” Second, it provides a further grouping of slides that may be useful in pinpointing the reasons for rater disagreement. Of course, whether Factor 1 *actually* distinguishes between “true negative” and “true positive” and whether the error characterization given by Factor 2 is accurate are important questions that could be addressed in future research.

### 10.3.3. Multigroup Models

Multigroup LC models can be used to compare models across groups. A completely unrestricted multigroup LC model, referred to by Clogg and Goodman (1984) as the model of complete heterogeneity, is equivalent to the estimation of a separate  $T$ -class LC model for each group. The fit of such a model can be obtained by simply summing the  $L^2$  values (and corresponding degrees of freedom) for the corresponding models in each group.

Let  $G$  denote a categorical variable representing membership in group  $g$ . The model of *complete heterogeneity* is expressed as (model  $M_{2C}$ )

$$\pi_{ijkl|g}^{ABCDX|G} = \pi_{i|g}^{X|G} \pi_{j|g}^{A|X,G} \pi_{k|g}^{B|X,G} \pi_{l|g}^{C|X,G} \pi_{l|g}^{D|X,G}.$$

#### 10.3.3.1. Example: Survey Respondent Types (Continued)

The second part of Table 10.1 provides the results of repeating our Example 1 analyses for the sample of *Black* respondents. These results turn out to be very similar to those obtained for the White respondents (see first part Table 10.1). As in our analysis for the White sample, we again reject the one- and two-class models in favor of three classes to obtain a model that provides an overall fit to the data that is adequate. The right-most portion of Table 10.2 presents the parameter estimates obtained from the three-class model (model  $H'_{2C}$ ) as applied to the sample of Blacks. As in our earlier analysis, the classes are ordered from largest to smallest.

In comparing results across these two groups, it is important to be able to interpret the three classes obtained from the Black respondents as representing the same latent constructs (“ideal,” “believers,” and “skeptics”) as in our analysis of the White respondents. Otherwise, any between-group comparisons would be like comparing apples with oranges. Although it is tempting to interpret Class 1 for both samples as representing the “ideal” respondents, this is not appropriate without first restricting the measurement portion of the models (the conditional probabilities) to be equal. These restrictions are accomplished using the model of *partial homogeneity* (model  $M_{2CR}$ ):

$$\pi_{ijkl|g}^{ABCDX|G} = \pi_{i|g}^{X|G} \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}. \quad (3)$$

Estimates from this model are given in the left-most portion of Table 10.7. The third part of Table 10.1 compares the fit of the unrestricted model  $M_{2C}$  and restricted model  $M_{2CR}$ . The  $\Delta L^2$  statistic can be used to test the restrictions made under model  $M_{2CR}$ . Because  $\Delta L^2 = 9.0$  with 18  $df$  is *not* significant, we are free to use this restricted model for our group comparisons.

The model of complete homogeneity (model  $M_{2CRR}$ ) imposes the further restriction that the latent class probabilities across the groups are identical:  $\pi_{t|1}^{X|G} = \pi_{t|2}^{X|G}$ , for  $t = 1, 2, 3$ . Because these restrictions yield a significant increase in  $L^2$ , we reject the model of complete homogeneity in favor of the model of partial homogeneity and conclude that there are significant differences in latent class membership between the White and Black samples.

Table 10.1 also includes results obtained from the LC factor model counterparts to the models of complete heterogeneity and partial heterogeneity. Because these models contain two dichotomous and independent factors, they contain the same number of parameters

**Table 10.7** Parameter Estimates for the Three-Class Latent Class (LC) Model of Partial Homogeneity (Model  $M_{2CR}$ ) and the Corresponding LC Two-Factor Model  $M_{2FR}$ 

	<i>Three Classes</i>			<i>Two Factors (Marginal Probabilities)</i>			
	<i>Class 1</i>	<i>Class 2</i>	<i>Class 3</i>	<i>Factor V</i>		<i>Factor W</i>	
	<i>Ideal</i>	<i>Believers</i>	<i>Skeptics</i>	<i>Level 1</i>	<i>Level 2</i>	<i>Level 1</i>	<i>Level 2</i>
LC probabilities							
Whites	0.68	0.15	0.17	0.81	0.19	0.85	0.16
Blacks	0.51	0.30	0.19	0.79	0.21	0.70	0.31
Conditional probabilities							
PURPOSE							
Good	0.89	0.90	0.16	0.90	0.20	0.76	0.78
Depends	0.06	0.06	0.21	0.06	0.21	0.09	0.07
Waste	0.05	0.04	0.63	0.05	0.59	0.15	0.15
ACCURACY							
Mostly true	0.60	0.64	0.01	0.63	0.02	0.50	0.55
Not true	0.40	0.36	0.99	0.37	0.98	0.50	0.45
UNDERSTANDING							
Good	0.94	0.32	0.74	0.79	0.76	0.92	0.26
Fair, poor	0.06	0.68	0.26	0.21	0.24	0.08	0.74
COOPERATION							
Interested	0.95	0.57	0.65	0.86	0.66	0.90	0.50
Cooperative	0.05	0.35	0.25	0.12	0.24	0.09	0.38
Impatient/hostile	0.00	0.08	0.10	0.02	0.10	0.01	0.12

as the three-class models  $M_{2C}$  and  $M_{2CR}$ . The lower part of Table 10.1 shows that these models fit better than the corresponding LC cluster models according to the BIC criteria. Also, the smaller BVRs than the LC cluster counterpart confirm that the LC factor model fits the data better.

The parameter estimates from the two-factor model  $M_{2FR}$  are presented in the right-most portion of Table 10.7. These are marginal latent class and conditional response probabilities for factors  $V$  and  $W$ , which are obtained by summing over the other factor. Note that variable  $D$  is strongly related to both factors  $V$  and  $W$ . That is, respondents at Level 1 of each factor have a higher probability (.90 or .91) of being “interested” than those at Level 2. Variables  $A$  and  $B$  relate only to factor  $V$ , and variable  $C$  relates only to factor  $W$ . That is, for factor  $V$ , those at Level 1 are substantially more likely to agree that surveys serve a good purpose and are more accurate than those at Level 2, but the two levels are about equal in showing a good understanding of the questions. For factor  $W$ , Level 1 shows good understanding, but Level 2 does not.

Moreover, Table 10.7 shows that group differences exist primarily with respect to Factor 2 (observed group differences on factor  $V$  are not significant). Black respondents are twice as likely as Whites to be at Level 2 of Factor 2 (30% vs. 15%). These results allow us to formulate a more rigorous test of our earlier hypothesis

that cooperation may be due to two separate factors—one associated with the belief that surveys serve a good purpose and are accurate (as assessed by LC Factor 1) and the second related to understanding the questions (as assessed by LC Factor 2).

Before concluding this section, we note that thus far, we have treated the trichotomous variables COOPERATE ( $A$ ) and PURPOSE ( $C$ ) as nominal. Alternatively, they can be treated as ordinal, which serves to simplify the model by reducing the number of parameters. The most straightforward approach is to restrict the log-linear parameters by using uniform scores  $v_i^A$  and  $v_k^C$  for the categories of  $A$  and  $C$ , implying the following constraints:  $\lambda_{ir}^{AV} = \lambda_r^{AV} v_i^A$  and  $\lambda_{is}^{AW} = \lambda_s^{AW} v_i^A$  (see, e.g., Formann, 1992; Heinen, 1996).

The use of these restrictions in our example increased the  $L^2$  by very little, indicating that variables  $A$  and  $C$  may in fact be treated as ordinal. In the next section, we present the results of a modified two-factor model in which variables  $A$  and  $C$  are treated as ordinal.

#### 10.3.4. Covariates

The parameters in the traditional LC model consist of unconditional and conditional probabilities. The conditional probabilities comprise the measurement

portion of the model. They characterize the distribution among the observed variables (indicators) conditional on the latent classes. The *unconditional* probabilities describe the distribution of the latent variable(s). To obtain improved description/prediction of the latent variable(s), we use a multinomial logit model to express these probabilities as a function of one or more exogenous variables  $\mathbf{Z}$ , called covariates (Dayton & Macready, 1988).

The multigroup model described in the previous section is an example of the use of a single nominal covariate ( $\mathbf{Z} = G$ ). For example, the term  $\pi_{ig}^{X|G}$  in equation (3) can be expressed as

$$\pi_{ig}^{X|G} = \frac{\exp(\gamma_i^X + \gamma_{ig}^{XG})}{\sum_{t=1}^T \exp(\gamma_t^X + \gamma_{it}^{XG})}$$

Although the latent variable(s) explain all of the associations among the indicators, associations between the covariates are *not* explained by the latent variables. This is what distinguishes the indicators from the covariates.

10.3.4.1. Example: Survey Respondent Types (Continued)

Regarding the interpretation of the three-class solution, McCutcheon (1987) questioned whether some of the difference in latent class membership between Black and White respondents might be explained by education, a question that falls outside the scope of traditional LC modeling. We address this question below by including  $E$ : EDUCATION as a second covariate in the two-factor model— $\mathbf{Z} = (G, E)$ .

The model provides a good fit to the data. The results indicate that the effect of education *does* explain most, but not all, of the group effect on factor  $W$ . The logit parameter estimates are given in Table 10.8, in which nonsignificant estimates were set to zero. The multinomial model used for the covariates was

$$\pi_{rsge}^{VW|GE} = \frac{\exp(\gamma_r^V + \gamma_s^W + \gamma_{gs}^{GW} + \gamma_{es}^{EW})}{\sum_{r=1, s=1}^{R, S} \exp(\gamma_r^V + \gamma_s^W + \gamma_{gs}^{GW} + \gamma_{es}^{EW})}$$

The gamma parameters in Table 10.8 indicate that the higher the educational level, the lower the score on factor  $W$ . The race effect is very weak: Blacks have a slightly higher score on factor  $W$  than Whites.

The results for this two-factor restricted multigroup model are also displayed in the biplot display (Magidson & Vermunt, 2001) given in Figure 10.2. Like the barycentric coordinate display in Figure 10.1,

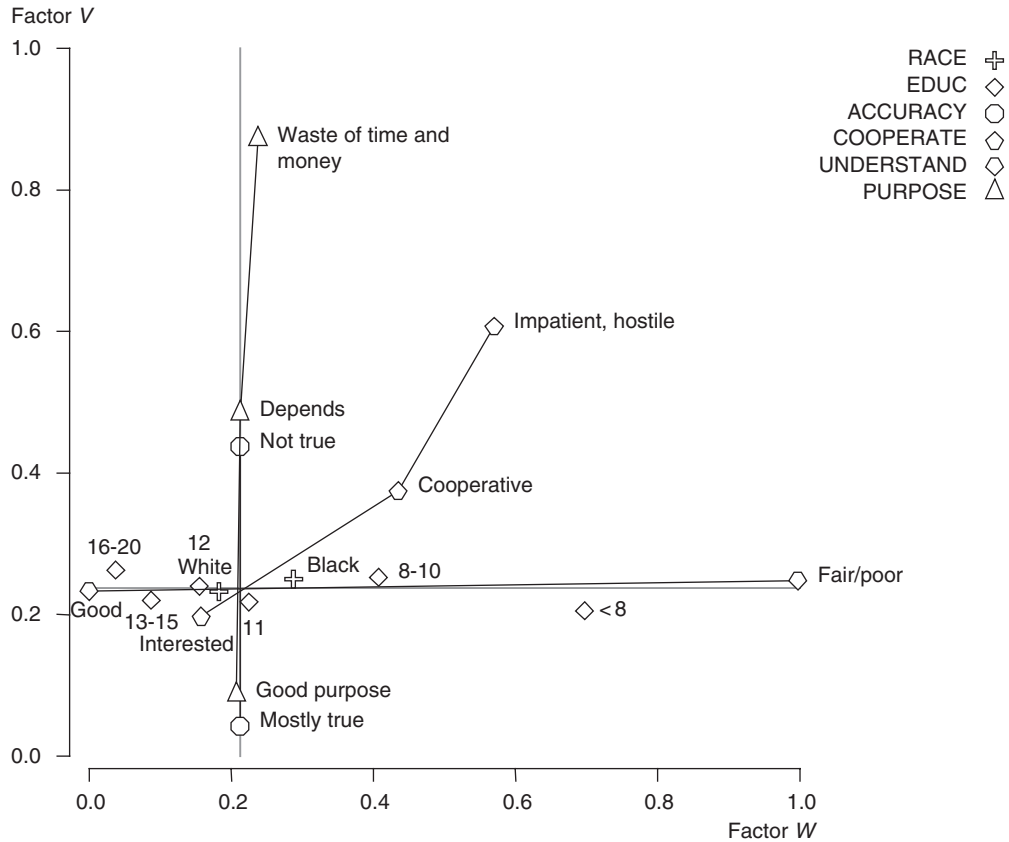
**Table 10.8** Parameter Estimates for the Two-Factor Restricted Multigroup Latent Class (LC) Model With Covariates

	Factor	
	V	W
<b>Covariates (gammas)</b>		
<i>G</i> : Group		
WHITE	0	-0.20
BLACK	0	0.20
<i>E</i> : Years of education		
< 8	0	2.19
8–10	0	0.97
11	0	0.08
12	0	-0.34
13–15	0	-1.01
16–20	0	-1.89
<b>Indicator variables (lambdas)</b>		
A: PURPOSE	2.26	0
<i>B</i> : ACCURACY		
Mostly true	-1.34	0
Not true	1.34	0
<i>C</i> : UNDERSTANDING		
Good	0	-5.14
Fair/poor	0	5.14
<i>D</i> : COOPERATION	0.98	1.26

we see that the horizontal axis, corresponding to factor  $W$ , is associated with UNDERSTANDING. Overall, respondents having a good understanding are highly likely to be at Level 1 of factor  $W$ , whereas those with a fair/poor understanding are highly likely to be at Level 2. The figure makes it clear that education is much more related to this factor than race. The vertical dimension is highly related to PURPOSE. Figure 10.2 shows more clearly than Figure 10.1 that COOPERATION is related to both factors. In particular, those rated as impatient/hostile tend to include two different types of respondents—those whose understanding is fair/poor, as well as those who view the purpose of surveys as a “waste of time and money.”

10.4. OTHER TYPES OF LATENT CLASS MODELS

Thus far, we have focused on the traditional LC modeling approach, including some important extensions such as covariates, several latent variables, and local dependencies. Some common characteristics of these models are that they serve as scaling methods or tools for dealing with measurement error, that indicators are nominal or ordinal, and that local

**Figure 10.2** Biplot for Two-Factor Model With Covariates

independence between indicators is the primary model assumption. In this section, we discuss other types of LC models. They are not used as scaling tools but as clustering methods, tools for dealing with unobserved heterogeneity, density estimation methods, or random-coefficients models (McLachlan & Peel, 2000). Moreover, indicators or dependent variables can be of scale types other than nominal or ordinal, and local independence is no longer the basic model assumption. As we will see, in some cases, there is only one indicator or dependent variable.

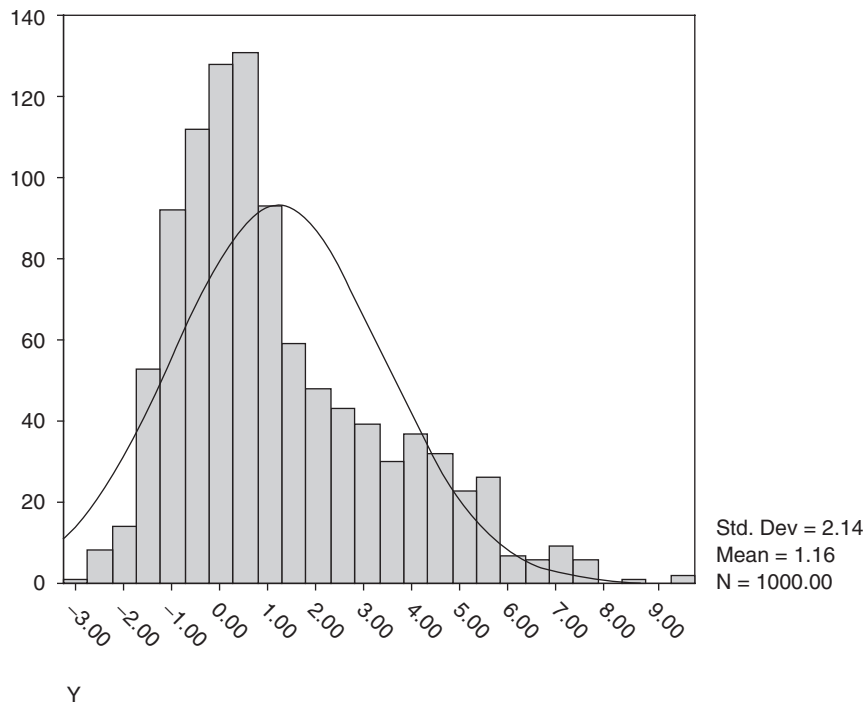
The next section presents simple mixture models for univariate distributions, with examples of mixtures of normals and mixtures of Poisson distributions. Then, we extend this basic model by including predictors, yielding what is called mixture regression or LC regression models. We present an example of a mixed linear regression model and show how the method can deal with various types of repeated measurements. Special attention is given to the relationship with hierarchical or multilevel models. Then, we present

another extension of the simple mixture model, that is, a mixture model for multivariate distributions. As will be shown, the resulting LC model can be seen as a model-based alternative to standard hierarchical clustering methods such as K-means. We end with a short overview of LC methods that were not discussed in detail.

#### 10.4.1. Simple Mixture Models

Consider the histogram depicted in Figure 10.3. This generated data set of 1,000 cases is obtained from a population consisting of a mixture of two normal distributions. For 60% of the population, the variable of interest follows a normal distribution with a mean of 0 and a variance of 1,  $N(0, 1)$ ; for the other 40%, the mean equals 3 and the variance 4,  $N(3, 4)$ . The normal curve that is drawn through the histogram shows that the resulting mixture is clearly not normally distributed.

**Figure 10.3** Simulated Distribution From a Mixture of Two Normals



A model that can be used to describe such a phenomenon is a finite mixture model (Everitt & Hand, 1981; McLachlan & Peel, 2000), which is a particular kind of LC model. The basic formula for a mixture of univariate distributions is

$$f(y|\boldsymbol{\vartheta}) = \sum_{t=1}^T \pi_t^X f(y|\boldsymbol{\varphi}_t). \quad (4)$$

The left-hand side of equation (4) indicates that we are interested in describing the distribution of a random variable  $y$ , which depends on a set of unknown parameters  $\boldsymbol{\vartheta}$ . The right-hand side contains two terms:  $\pi_t^X$  is the probability of belonging to latent class or mixture component  $t$ , and  $f(y|\boldsymbol{\varphi}_t)$  is the distribution of  $y$  within latent class  $t$ , given some unknown parameters  $\boldsymbol{\varphi}_t$ . The class-specific distribution of  $y$  is assumed to belong to a particular parametric family. Depending on the scale type of  $y$ , this can, for instance, be a normal, Poisson, binomial, exponential, or gamma distribution. The summation on the right-hand side indicates that the distribution of  $y$  is a weighted mean of the class-specific distributions, where the latent class proportions serve as weights.

Mixture models such as these have two important types of applications. The first is density estimation: Complicated distributions can be approximated by a

**Table 10.9** Test Results for Generated Mixture of Normals Data

<i>Model</i>	<i>Log-Likelihood</i>	<i>BIC<sub>LL</sub></i>	<i>Number of Parameters</i>
Equal variances			
One-class	-2177.75	4369.31	2
Two-class	-2066.99	4161.61	4
Three-class	-2050.78	4143.00	6
Four-class	-2046.25	4147.75	8
Unequal variances			
Two-class	-2048.14	4130.81	5
Three-class	-2047.78	4150.83	8
Four-class	-2045.41	4166.80	11

NOTE: BIC = Bayesian information criterion.

mixture of simple parametric distributions. Another important application type is clustering, in which case the class-specific parameters are used to define the clusters, and the posterior membership probabilities are used to classify cases into the most appropriate cluster.

Table 10.9 presents test results for various models fitted to the data depicted in Figure 10.3. We estimated one- to four-class mixtures of normal distributions with equal and unequal within-cluster variances. As can be seen, the BIC measure identifies the correct model, the two-cluster model with unequal



**Table 10.10** Observed and Estimated Frequency Distribution of Packs of Hard Candy Purchased During Past 7 Days Under the One-Class and Three-Class Poisson Model

Number of Packages	Frequencies		
	Observed	One-Class Model	Three-Class Model
0	102	8.43	101.67
1	54	33.63	54.63
2	49	67.11	50.03
3	62	89.28	53.89
4	44	89.09	47.25
5	25	71.11	34.14
6	26	47.30	22.00
7	15	26.97	14.37
8	15	13.46	11.02
9	10	5.97	10.18
10	10	2.38	10.17
11	10	0.86	9.97
12	10	0.29	9.20
13	3	0.09	7.90
14	3	0.03	6.32
15	5	0.01	4.72
16	5	0.00	3.30
17	4	0.00	2.18
18	1	0.00	1.36
19	2	0.00	0.80
20	1	0.00	0.45

within-cluster variances, as best. The three-class model with equal within-cluster variances fits almost as well, showing that a simpler parametric form can sometimes be compensated by a larger number of mixture components.

In the two-class model with unequal variances, the estimated probability of belonging to Class 1 is .64. This class has an estimated mean of  $-0.03$  and a variance of 1.01. The mean and variance of the other class equal 3.24 and 3.95. Note that these estimates are close to the population values we used to generate this data set.

Table 10.10 provides a data set taken from Dillon and Kumar (1994) that we will use as a second example. It gives the observed frequency distribution of the number of packs of hard candy consumed by 456 respondents during the 7 days prior to the survey. Because the outcome variable is a count without a fixed maximum, it is most natural to assume that it follows a Poisson distribution. The table also reports the estimated frequency distribution obtained with a standard, or one-class, Poisson model, as well as with a three-class mixture Poisson model. As can be seen, the standard Poisson model does not fit the

empirical distribution at all, whereas the three-class Poisson describes the data almost perfectly. This shows that a mixture of simple parametric distributions can be used to describe a quite complicated empirical distribution.

Test results obtained when applying mixture Poisson models to the hard candy data set show that models with two and three mixture components perform much better than the standard Poisson model. As is typical, there is a saturation point at which increasing the number of classes no longer increases the log-likelihood function: In this case, it occurs at four classes. The three-cluster solution is the one that is preferred according to the BIC criterion.

The estimated latent class proportions in the three-class model are 0.54, 0.28, and 0.18, and the Poisson rates are 3.48, 0.29, and 11.21. This means that we identified a small cluster of heavy users (more than 11 packs in 7 days), a cluster containing slightly more than a quarter of the respondents with almost no usage, and a large group of moderate users.

#### 10.4.2. LC Regression Models

In the simple mixture models discussed above, it was assumed that the mean of the chosen parametric distribution differs across latent classes. This can also be expressed by specifying a linear regression model for the mean of the distribution of interest,  $\mu_t$ , after applying some transformation or link function  $g(\cdot)$  that depends on the scale type of the  $y$  variable. For the mean of a binomial or multinomial distribution, we use a logit transformation; for a Poisson mean, a log transformation; and for a normal mean, no transformation or an identity link. The regression model has the form

$$g(\mu_t) = \beta_{0t}.$$

As can be seen, this regression model contains only an intercept, and this intercept is class specific.

Let  $w$  denote a set of predictors or explanatory variables. Suppose we are no longer interested in the unconditional distribution of  $y$  but in the conditional distribution of  $y$  given  $w$ ,  $f(y|w, \phi_t)$ . A natural way to express the dependency of  $y$  on  $w$  is by the inclusion of the set of predictors  $w$  on the right-hand side of the regression equation. In the case of a single predictor  $w$ , the resulting LC regression model (Wedel & DeSarbo, 1994) has the form

$$g(\mu_t) = \beta_{0t} + \beta_{1t}w,$$

where  $\beta_{0t}$  and  $\beta_{1t}$  are the class-specific regression coefficients.

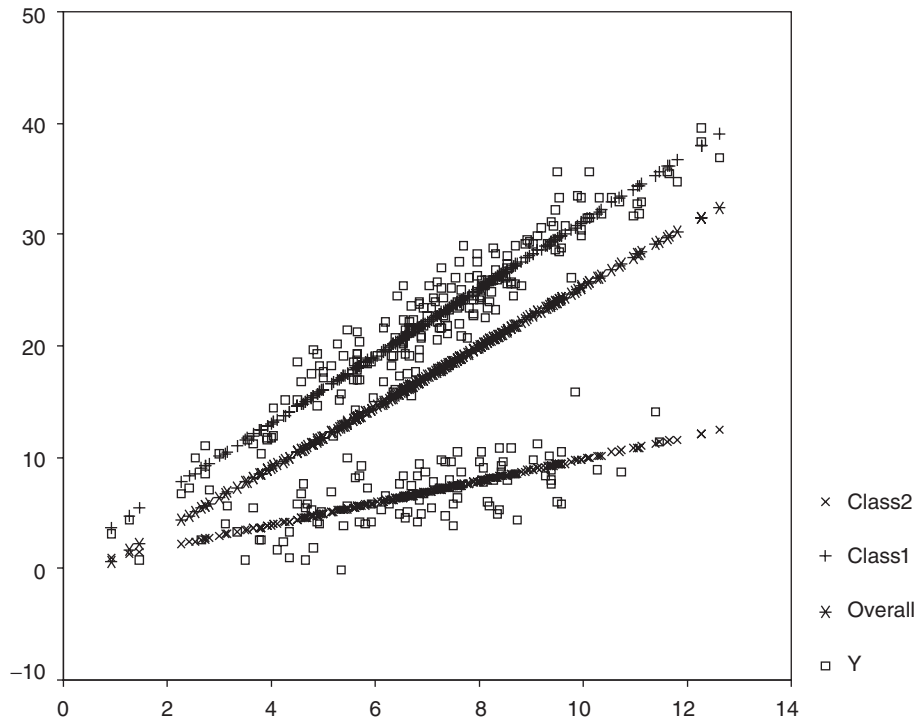
**Figure 10.4** Simulated Two-Class Latent Class (LC) Regression Model

Figure 10.4 depicts a data set generated from a population consisting of two latent classes, with class-specific regression models equal to  $\mu_1 = 1 + 3w$  and  $\mu_2 = 0 + 1w$ . It also compares the estimated  $y$  values for the two-class model (YLC2) with the standard one-class regression model (YLC1). As can be seen, the description given by the standard regression model is very poor compared to the two-class model. The LC regression modeling procedure has no problem identifying the two regression lines without preknowledge of class membership.

In an LC regression model, the latent variable is a predictor that interacts with the observed predictors, which means that it serves as a moderator variable. Compared to a standard regression model in which all predictors are observed, this basic LC regression model provides several useful functions. First, it can be used to weaken standard regression assumptions about the nature of the effects (linear, no interactions) and the error term (independent of predictors, particular distribution, homoskedastic). Second, it makes it possible to identify and correct for sources of unobserved heterogeneity. As explained below, this is especially useful in situations when there are repeated measurements

or other types of dependent observations. Longitudinal data applications are sometimes referred to as LC or mixture growth models (each latent class has its own growth curve). Third, it can be used to detect outliers because these are cases for which the primary regression model does not hold.

An important application area for LC regression modeling is clustering or segmentation (Wedel & Kamakura, 1998). In particular, ratings- and choice-based conjoint studies are designed to identify subgroups (segments) that react differently to product characteristics, which is the same as saying that these groups have different regression coefficients. This type of application is illustrated in more detail below with an empirical example.

#### 10.4.2.1. Example: Repeated Measurements or Clustered Observations

As explained below, the LC regression model can be viewed as a random-coefficients model that, similar to multilevel or hierarchical models, can take dependencies between observations into account. This extends the application of LC regression models to

**Table 10.11** Parameter Estimates for the Abortion Example

<i>Parameter</i>	<i>Class 1</i>	<i>Class 2</i>	<i>Class 3</i>	<i>Class 4</i>	<i>Mean</i>	<i>Standard Deviation</i>
Class size	0.30	0.28	0.24	0.19		
Intercept	-0.34	0.60	3.33	1.59	1.16	1.38
Year						
1983	0.14	0.26	0.47	-0.58	0.12	0.35
1984	-0.12	-0.46	-0.35	-1.11	-0.45	0.34
1985	0.04	-0.44	-0.26	1.43	0.10	0.66
1986	-0.06	0.64	0.14	0.26	0.24	0.27
Religion						
Roman Catholic	-0.53	-0.53	-0.53	-0.53	-0.53	0.00
Protestant	0.20	0.20	0.20	0.20	0.20	0.00
Other	-0.10	-0.10	-0.10	-0.10	-0.10	0.00
No religion	0.42	0.42	0.42	0.42	0.42	0.00

situations with repeated measurements or other types of dependent observations.

We will illustrate LC regression with repeated measurements using an application to longitudinal survey data. This is, therefore, an example of an LC growth model. The data set consists of 264 participants in the 1983 to 1986 yearly waves of the British Social Attitudes Survey (McGrath & Waterton, 1986). The dependent variable is the number of yes responses on seven yes/no questions as to whether it is a woman's right to have an abortion under specific circumstances. Because this is a count variable with a fixed total, it is most natural to work with a logit link and binomial error function. The predictors that we used are the year of measurement (1 = 1983, 2 = 1984, 3 = 1985, 4 = 1986) and religion (1 = Roman Catholic, 2 = Protestant, 3 = other, 4 = no religion). The effect of year of measurement is assumed to be class dependent, and the effect of religion is assumed to be the same for all classes.

We estimated models with one to five classes, and the four-class model turned out to perform best in terms of the BIC criterion. We also estimated more restricted models in which the time effect is assumed to be linear and/or the time effect is assumed to be class independent. These models did not describe the data as well as our four-class model, which indicates that the time trend is nonlinear and heterogeneous.

The parameters obtained with the four-class model appear in Table 10.11. The parameter means across classes indicate that the attitudes are most positive at the last time point and most negative at the second time point. Furthermore, the effects of religion show that people without religion are most in favor and Roman Catholics and others are most against abortion.

Protestants have a position that is close to the no-religion group.

The class-specific parameters indicate that the four latent classes have very different intercepts and time patterns. The largest Class 1 is most against abortion, and Class 3 is most in favor of abortion. Both latent classes are very stable over time. The overall level of latent Class 2 is somewhat higher than of Class 1, and it shows somewhat more change of the attitude over time. People belonging to latent Class 4 are very instable: At the first two time points they are similar to Class 2, at the third time point to Class 4, and at the last time point again to Class 2 (this can be seen by combining the intercepts with the time effects). Class 4 could therefore be labeled as random responders. It is interesting to note that in a three-class solution, the random-responder class and Class 2 are combined. Thus, by going from a three- to a four-class solution, one identifies the interesting group with less stable attitudes.

Vermunt and Van Dijk (2001) used the same empirical example to illustrate the similarity between LC regression models and random-coefficients, multilevel, or hierarchical models. Using terminology from multilevel modeling, the time variable is a Level 1 predictor and religion a Level 2 predictor. The effect of the Level 1 predictor time is allowed to vary across Level 2 units—in this case, individuals. The LC regression output can be transformed into the usual output produced by a standard multilevel or hierarchical model—means, variances, covariances of the intercept, and the three time effects—by elementary statistical operations. The most important part of this multilevel output is what appears in the last two columns of Table 10.11.

A difference between LC regression analysis and standard hierarchical models is that the former does not make strong assumptions about the distribution of the random coefficients. LC regression models can, therefore, be seen as nonparametric hierarchical models in which the distribution of the random coefficients is approximated by a limited number of mass points (= latent classes). As shown by Vermunt and Van Dijk (2001), the LC approach has the practical advantage of being much less computationally intensive than parametric models, and substantively easier to interpret results are often obtained.

#### 10.4.2.2. Example: Application to Choice-Based Conjoint Studies

The LC regression model is a popular tool for the analysis of data from conjoint experiments in which individuals rate or choose between sets of products having different attributes (Wedel & Kamakura, 1998). The objective is to determine the effect of product characteristics on the rating or the choice probabilities. LC analysis is used to identify subgroups or market segments for which these effects differ.

For illustration of LC analysis of data obtained from choice-based conjoint experiments, we use a generated data set. The products are 10 pairs of shoes that differ on three attributes: fashion (0 = traditional, 1 = modern), quality (0 = low, 1 = high), and price (ranging from 1 to 5). Eight choice sets offer 3 of the 10 possible alternative products to 400 individuals. Each choice task consists of indicating which of the 3 alternatives they would purchase, with the response “none of the above” allowed as a fourth choice option.

The model that is used is a multinomial logit model with choice-specific predictors, also referred to as the conditional logit model. Let  $M$  be the number of choice sets,  $K$  the number of choices per set, and  $J$  the number of predictors. A particular set, choice, and predictor are denoted by  $m$ ,  $k$ , and  $j$ , respectively. The regression model of interest is

$$\pi_{mkt} = \frac{\exp(\sum_{j=1}^J \beta_{jt} w_{mjk})}{\sum_{k=1}^K \exp(\sum_{j=1}^J \beta_{jt} w_{mjk})}$$

Here,  $\pi_{mkt}$  denotes the probability that someone belonging to class  $t$  selects choice alternative  $k$  in choice set  $m$ . The predictors we use are the three product attributes (fashion, quality, and price), as well as a dummy variable for the “none” category.

The BIC values indicated that the three-class model is the model that should be preferred. The parameter estimates obtained with the three-class model

**Table 10.12** Parameter Estimates for Conditional Logit Model in Conjoint Study Example

	Class 1	Class 2	Class 3	Wald for No Effect	Wald for Equal Effects
Fashion	3.03	-0.17	1.20	494.74	216.37
Quality	-0.09	2.72	1.12	277.96	171.16
Price	-0.39	-0.36	-0.56	144.48	3.58
None	1.29	0.19	-0.43	82.39	59.26

**Table 10.13** Parameter Estimates for the Latent Variable Regression for Conjoint Study Example

	Class 1	Class 2	Class 3	Wald
Intercept	0.37	0.00	-0.37	8.22
SEX				
Male	-0.66	-0.34	1.01	24.15
Female	0.66	0.34	-1.01	
AGE				
16-24	1.02	-0.15	-0.87	62.76
25-39	-0.59	-0.37	0.96	
40+	-0.43	0.52	-0.09	

are reported in Table 10.12. As can be seen, fashion has a major influence on choice for Class 1, quality for Class 2, and both fashion and quality for Class 3. The price effect is similar for all three classes. The Wald test for the equality of effects between classes indicates that the difference in price effects across classes is not significant. The price effects could, therefore, be assumed to be class independent.

In addition to the conditional logit model, which shows how the predictors affect the likelihood of choosing one alternative over another, differentially for each class, we specified a second logit model to describe the latent class variable as a function of the covariates sex and age. Table 10.13 shows that females turn out to belong more often to Class 1 and males to Class 3. Younger persons have a higher probability of belonging to Class 1 (emphasize fashion in choices), and older persons are most likely to belong to Class 2 (emphasize quality in choices).

In conclusion, the LC regression model offers computational and interpretive advantages over the more traditional hierarchical modeling approach that tends to overfit data (Andrews, Ansari, & Currim, 2002). In our example, we used the BIC criteria to select a parsimonious number of classes. However, researchers who prefer the results to show *higher* levels of individual variation in regression coefficients can obtain such

with LC regression models by simply increasing the number of latent classes to produce the desired amount of variation.

#### 10.4.3. LC Analysis as an Alternative to K-Means Clustering

An important application of LC analysis is clustering (Banfield & Raftery, 1993; McLachlan & Peel, 2000; Vermunt & Magidson, 2002). Actually, we already saw several cluster-like applications. The traditional LC model was used to construct a typology of survey respondents using a set of categorical indicators. We also showed that simple mixture models such as mixtures of normals or mixtures of Poisson distributions could be used for clustering purposes.

In this section, we will concentrate on LC analysis as a tool for cluster analysis with *continuous* indicators. These LC models can be seen as multivariate extensions of the mixtures of univariate normals discussed above. Instead of assuming a univariate normal distribution, we assume multivariate normal distributions within latent classes. The most general form of the mixture model concerned assumes that each latent class has its own set of means, variances, and covariances. More formally,

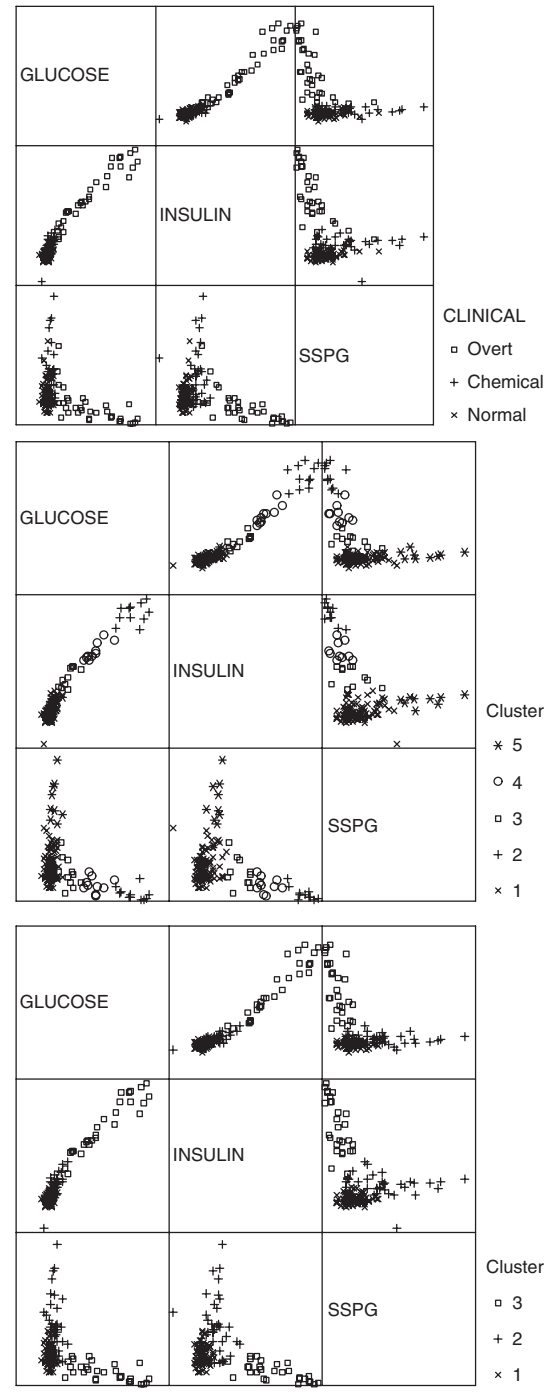
$$f(\mathbf{y}|\boldsymbol{\vartheta}) = \sum_{t=1}^T \pi_t^X f(\mathbf{y}|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t).$$

Here,  $\boldsymbol{\mu}_t$  denotes the vector with class-specific means and  $\boldsymbol{\Sigma}_t$  the class-specific variance-covariance matrix. Note that, contrary to traditional LC modeling, it is not necessary to assume local independence between the indicators.

The above LC cluster model is similar to the model used in discriminant analysis. An important difference is, of course, that in cluster analysis, group membership is unobserved or latent, which is the reason that LC cluster analysis is sometimes referred to as latent discriminant analysis.

The first part of Figure 10.5 depicts a data set that we will use to illustrate the LC cluster model for continuous variables. Three measures are available to diagnose diabetes: glucose, insulin, and steady-state plasma glucose (SSPG) (see Fraley & Raftery, 1998). In addition to these measures, we have information on the clinical diagnosis consisting of the three categories “normal,” “chemical diabetes,” and “overt diabetes.” However, in practice, a gold standard is not available in cluster applications. Our objective here is to construct a mixture model that yields a classification that is close to the clinical diagnosis, without use of the

**Figure 10.5** Matrix Scatter Plot of Diabetes Data Set for the Clinical Classification, the K-Means-Like Five-Cluster Solution, and the Final Three-Cluster Solution



information on the clinical diagnosis. We use this data set to demonstrate the flexibility of LC clustering compared to other clustering methods. The gold

**Table 10.14** Test Results for Diabetes Data

<i>Model</i>	<i>Log-Likelihood</i>	<i>BIC<sub>LL</sub></i>	<i>Number of Parameters</i>
Equal and diagonal			
One-cluster	-2750.13	5530.13	6
Two-cluster	-2559.88	5169.52	10
Three-cluster	-2464.78	4999.24	14
Four-cluster	-2424.46	4938.49	18
Five-cluster	-2392.56	<b>4894.60</b>	22
Unequal and diagonal			
One-cluster	-2750.13	5530.13	6
Two-cluster	-2446.12	4956.94	13
Three-cluster	-2366.92	4833.38	20
Four-cluster	-2335.38	<b>4805.13</b>	27
Five-cluster	-2323.13	4815.47	34
Unequal and full			
One-cluster	-2546.83	5138.46	9
Two-cluster	-2359.12	4812.80	19
Three-cluster	-2308.64	<b>4761.61</b>	29
Four-cluster	-2298.13	4790.34	39
Five-cluster	-2284.97	4813.79	49
Unequal and $y_1 - y_2$ free			
One-cluster	-2560.40	5155.64	7
Two-cluster	-2380.27	4835.19	15
Three-cluster	-2320.57	<b>4755.61</b>	23
Four-cluster	-2303.14	4760.56	31
Five-cluster	-2295.05	4784.19	39

NOTE: Bold numbers (minimum Bayesian information criterion [BIC]) indicate the model that would be selected according to the BIC criterion.

standard makes it possible to judge whether the methods do what we want them to do.

LC cluster analysis is a model-based clustering procedure. As such, it is a probabilistic and more flexible alternative to K-means clustering. K-means clustering performs well under very strict conditions—that is, if indicators are locally independent and if error variances are cluster invariant and equal across indicators ( $\Sigma_t = \sigma^2 \mathbf{I}$ ). These implicit assumptions of K-means imply that in a three-dimensional scatter plot, each cluster has the form of a sphere with the same radius, and in each two-dimensional plot, each cluster will have the form of a sphere with the same radius. The assumption of equal error variances across indicators is the reason that in K-means clustering, it is advised to standardize the variables prior to the analysis. Although standardization often improves the situation, it does not solve the problem because equating the variance in the total sample is not the same as equating the within-group variances (Magidson & Vermunt, 2002).

Having a closer look at Figure 10.5, it can easily be seen that it is impossible to describe the shape of the three diabetes clusters by a K-means model, that is, by three spheres with the same radius. The within-cluster variances are very different across clusters and

across indicators. Moreover, the glucose and insulin indicators are strongly correlated within the group with overt diabetes. Nevertheless, because the clusters are well separated, a reliable cluster method should be able to yield a three-cluster solution that is similar to the clinical classification.

The problems associated with K-means are confirmed by the test results reported in Table 10.14. We estimated one- to five-cluster models, each with four different specifications of the variance-covariance matrix: diagonal (= local independence) and equal across classes, diagonal and unequal, glucose-insulin covariance and unequal, and all covariances and unequal. It can be seen that when the specifications are too restrictive, one needs five and four clusters, respectively. Actually, with the first K-means-like specification, even more than five clusters are needed.

Although the BIC values indicate that the two additional local dependencies ( $y_1 - y_3$  and  $y_2 - y_3$ ) in the full model are not needed (compare the three-cluster solutions for the last two specifications), the fit measures also show that both the model with the fully unrestricted covariance matrix and the model with only the glucose-insulin covariance detect the correct three-cluster solution. This means that working with a

model with insufficient restrictions does not harm in this example, but this is not always the case.

The middle part of Figure 10.5 shows the five nearly spherical clusters identified with the most restricted specification we used. Similar results would have been obtained with K-means. The lower part of Figure 10.5 depicts the three-cluster solution that turned out to be the best according to the BIC criterion. It can be seen that the three clusters identified by this model are very similar to the clinical classification. Our three-cluster solution is smoother in the sense that some of the overlap between the clinical classes disappears, which is, of course, what can be expected from a statistical model. The correspondence between the three-cluster and the clinical classification is 87%, which is only slightly lower than the 93% correct classifications of a quadratic discriminant analysis (in which cluster membership is treated as known).

The LC cluster model can be applied not only with continuous indicators but also with indicators of other scale types and different combinations of scale types. Depending on the scale type, one will specify the most appropriate within-cluster distribution for the indicator concerned. This yields a general cluster model for mixed-model data (Hunt & Jorgensen, 1999; Vermunt & Magidson, 2002). Note that the traditional LC model is the special case in which all indicators are categorical variables.

#### 10.4.4. Other Developments in LC Modeling

In this chapter, we presented what we believe to be the most important types of LC models. We did not discuss LC models for transition, survival, or event history data (Vermunt, 1997). Most of these models are mixture regression models and can, therefore, be handled within the LC regression framework. Another important class of models for transition data are latent or hidden Markov models that can be used to separate true change from measurement error in the outcome variable of interest (see, e.g., Langeheine & Van de Pol, 1994). The structure of latent Markov models is similar to the LC models with several latent variables discussed in the previous section.

In the previous section, we presented LC models that can be used for scaling. There also exist more sophisticated LC scaling models, which can be obtained by imposing certain constraints on the parameters of the traditional LC model. Examples are LC models for probabilistic Guttman scaling, LC models with order constraints, LC Rasch models, LC models for preference data, and LC models for distance data

(see Böckenholt, 2002; Croon, 2002; Dayton, 1998; Heinen, 1996).

Another more advanced type of LC model we would like to mention is the LISREL-type framework for categorical variables developed by Hagenaars (1990) and extended by Vermunt (1997). Any type of LC models with categorical indicators, including LC models for transition data and sophisticated LC scaling models, are special cases of this general model. A limitation of this approach is that it is restricted to categorical indicators.

A final recent development that we would like to mention is the development of more sophisticated restricted mixtures of multivariate normals than that discussed above. LC models have been proposed in which the class-specific covariance matrices are constrained by means of principal component (Fraley & Raftery, 1998) or factor-analytic (Yung, 1997) structures or by structural equation models (Jedidi, Jagpal, & DeSarbo, 1997).

#### REFERENCES

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley.
- Andrews, R. L., Ansari, A., & Currim, I. S. (2002). Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *Journal of Marketing Research*, 39, 87–98.
- Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, 803–821.
- Böckenholt, U. (2002). Comparison and choice: Analyzing discrete preference data by latent class scaling models. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 163–182). Cambridge, UK: Cambridge University Press.
- Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762–771.
- Croon, M. A. (2002). Ordering the classes. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 137–162). Cambridge, UK: Cambridge University Press.
- Dayton, C. M. (1998). *Latent class scaling models*. Thousand Oaks, CA: Sage.
- Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173–178.
- Dillon, W. R., & Kumar, A. (1994). Latent structure and other mixture models in marketing: An integrative survey and overview. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge, UK: Blackwell.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476–486.

- Fraley, C., & Raftery, A. E. (1998). *How many clusters? Which clustering method? Answers via model-based cluster analysis* (Tech. Rep. No. 239). Seattle: Department of Statistics, University of Washington.
- Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I. A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.
- Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–231.
- Haberman, S. J. (1979). *Analysis of qualitative data: Vol. 2. New developments*. New York: Academic Press.
- Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods & Research*, 16, 379–405.
- Hagenaars, J. A. (1990). *Categorical longitudinal data—loglinear analysis of panel, trend and cohort data*. Newbury Park, CA: Sage.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.
- Hunt, L., & Jorgensen, M. (1999). Mixture model clustering using the MULTIMIX program. *Australian and New Zealand Journal of Statistics*, 41, 153–172.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39–59.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Langeheine, R., Pannekoek, J., & Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods & Research*, 24, 492–516.
- Langeheine, R., & Van de Pol, F. (1994). Discrete-time mixed Markov latent class models. In A. Dale & R. B. Davies (Eds.), *Analyzing social and political change: A casebook of methods* (pp. 171–197). London: Sage.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31, 223–264.
- Magidson, J., & Vermunt, J. K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20, 37–44.
- McCutcheon, A. L. (1987). *Latent class analysis*. Newbury Park, CA: Sage.
- McGrath, K., & Waterton, J. (1986). *British social attitudes, 1983–1986 panel survey* (Technical report). London: Social and Community Planning Research.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: John Wiley.
- Raftery, A. E. (1986). Choosing models for cross-classifications. *American Sociological Review*, 51, 145–146.
- Rindskopf, R., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, 5, 21–27.
- Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement*, 23, 283–297.
- Uebersax, J. S., & Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, 9, 559–572.
- Van der Heijden, P. G. M., Gilula, Z., & Van der Ark, L. A. (1999). On a relationship between joint correspondence analysis and latent class analysis. *Sociological Methodology*, 29, 147–186.
- Vermunt, J. K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.
- Vermunt, J. K., & Magidson, J. (2003). *Latent GOLD 3.0 user's guide*. Belmont, MA: Statistical Innovations, Inc.
- Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, 13, 6–13.
- Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge, UK: Blackwell.
- Wedel, M., & Kamakura, W. A. (1998). *Market segmentation: Concepts and methodological foundations*. Boston: Kluwer Academic.
- Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297–330.