Latent Class Regression & Growth Models

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

www.jeroenvermunt.nl

Introduction

- In the previous videos, I discussed LC models for multiple categorical and/or continuous indicators.
- This video deals with LC models for a *single dependent variable*, which is *observed multiple times* and for which a *regression model* is specified.
- The aim is to identify latent classes for which the intercept and the predictor effects in the regression model differ.
- This model is referred to as LC or mixture regression model.
- When applied to longitudinal data using time variables as predictors, the model is called LC growth, mixture growth, latent trajectory, or group-based trajectory model.

Example: Conjoint.sav

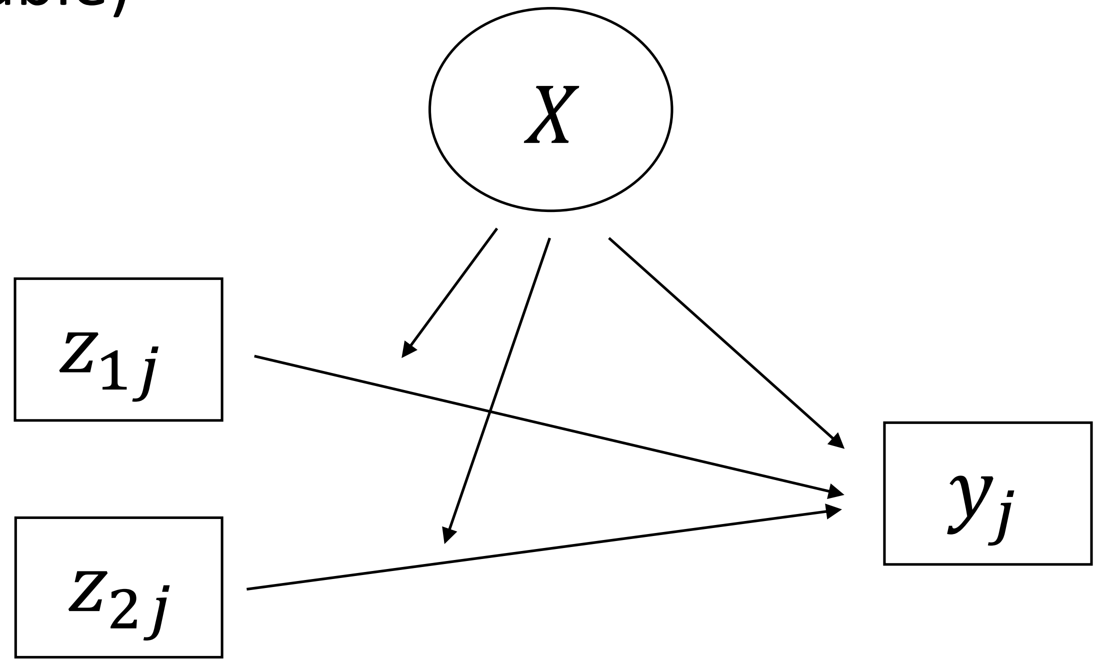
- Fictitious repeated measures experiment
- Rating of 8 pairs of shoes (how likely is it that you will buy this pair measured using a 5 point scale)
- Shoes differ in *fashion*, *quality*, and *price*, which are three within-subject experimental factors
- I will treat the dependent variable as continuous and thus use a linear regression model (though an ordinal logit model would be better)
- Age and gender can be used as covariates (e.g. in a step-3 analysis)

Example: Conjoint.sav (Long format data file)

	id	sex	age	fashion	quality	price	rating
1	1	1	2	1	1	2	1
2	1	1	2	1	1	1	3
3	1	1	2	1	2	2	3
4	1	1	2	1	2	1	5
5	1	1	2	2	1	2	2
6	1	1	2	2	1	1	2
7	1	1	2	2	2	2	5
8	1	1	2	2	2	1	5
9	2	2	1	1	1	2	2
10	2	2	1	1	1	1	3
11	2	2	1	1	2	2	5
12	2	2	1	1	2	1	5
13	2	2	1	2	1	2	1
14	2	2	1	2	1	1	5
15	2	2	1	2	2	2	4
16	2	2	1	2	2	1	5
17	3	1	3	1	1	2	2
18	3	1	3	1	1	1	3

LC regression model: graphical representation

- Multiple measurements of the same dependent variable; predictors represent, for example, *within-subject* experimental conditions
- Predictor effects differ across latent classes (which serve as a moderator variable)



LC regression model: formulae

- A model with two predictors z_{1ij} and z_{2ij}

$$P(y_{i1},\ldots,y_{iJ_i}\mid \mathbf{z}_{i1},\mathbf{z}_{i2})=\sum_{c=1}^{C}P(X=c)\prod_{j=1}^{J_i}P(y_{ij}\mid X=c,z_{ij1},z_{ij2})$$

with, for example, a binary logistic regression model for y_{ij}

$$P(y_{ij}=1\mid X=c,z_{ij2},z_{ij2})=\frac{\exp(\alpha_c+\beta_{1c}z_{ij1}+\beta_{2c}z_{ij2})}{1+\exp(\alpha_c+\beta_{1c}z_{ij1}+\beta_{2c}z_{ij2})}$$

or a linear regression model

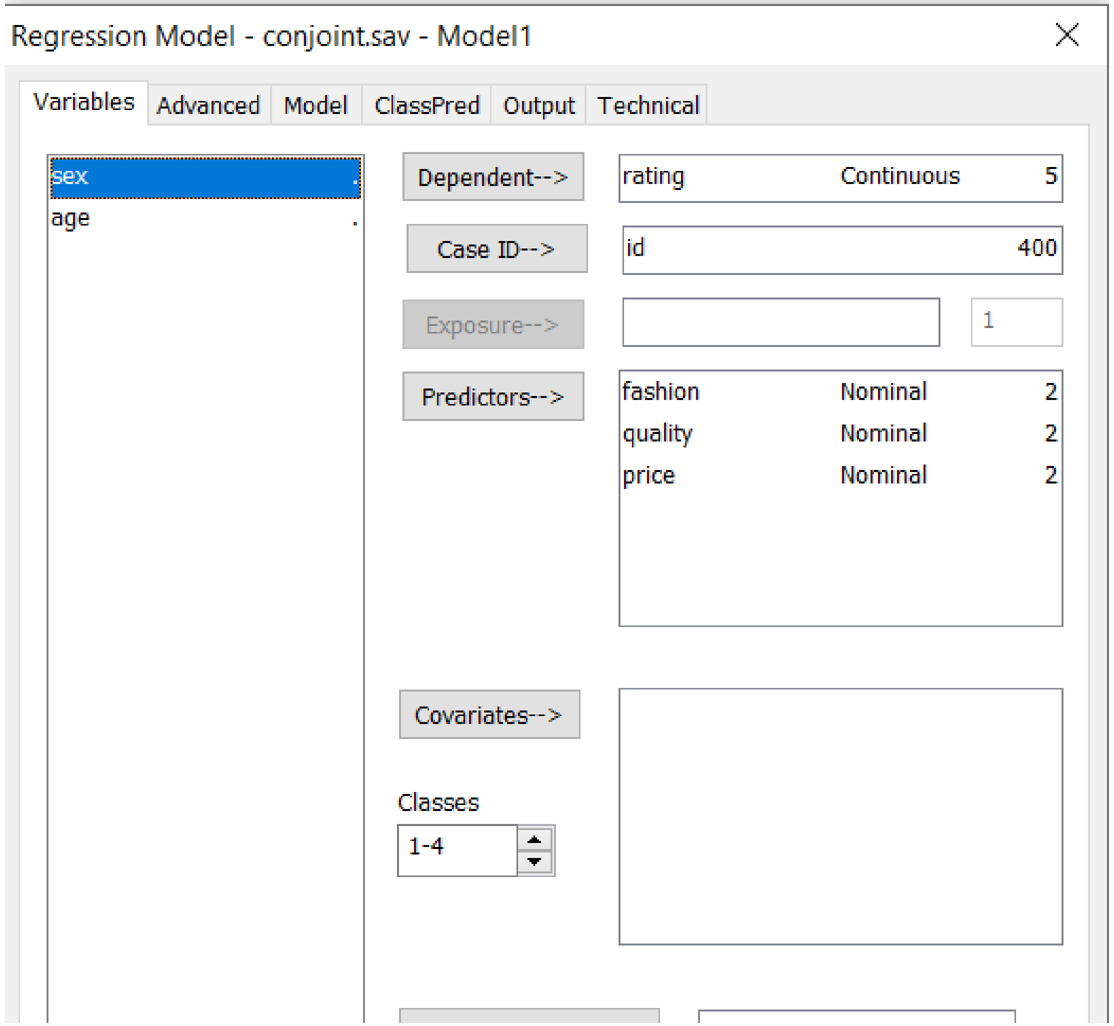
$$E(y_{ij}\mid X=c,z_{ij1},z_{ij2})=\alpha_c+\beta_{1c}z_{ij2}+\beta_{2c}z_{ij2}$$

LC regression model: notes

- Note the use of the indices *i* and *j*, referring to individual and response/measurement, respectively
- Note the use of the index *c* referring to a latent class: as you can see regression parameters are allowed to vary across latent classes

- Issues like model selection, classification, relating classes to external variables (say using a step-3 analysis) remain the same as in other LC models

Example: Conjoint.sav (Regression in LG)



Example: Conjoint.sav (Parameters Output)

Model for Dependent									
	Class1	**Class2**	**Class3**	**Overall**					
R^2	0.5528	0.4851	0.6551	0.6032					
rating	**Class1**	**Class2**	**Class3**	**Wald**	**p-value**	**Wald(=)**	**p-value**	**Mean**	**Std.Dev.**
Intercept									
	2.6886	3.2675	3.2665	22375.4390	2.1e-4857	250.1108	4.9e-55	2.9191	0.2832
Predictors	**Class1**	**Class2**	**Class3**	**Wald**	**p-value**	**Wald(=)**	**p-value**	**Mean**	**Std.Dev.**
fashion									
Traditional	-0.9641	-0.3951	0.0633	1467.0511	8.3e-318	395.0477	1.6e-86	-0.6538	0.4075
Modern	0.9641	0.3951	-0.0633					0.6538	0.4075
quality									
Low	-0.0671	-0.6349	-1.1648	510.3738	2.7e-110	396.4347	8.2e-87	-0.3900	0.4302
High	0.0671	0.6349	1.1648					0.3900	0.4302
price									
Lower	0.4668	0.6092	0.3047	680.7587	3.1e-147	21.5233	2.1e-5	0.4681	0.0958
Higher	-0.4668	-0.6092	-0.3047					-0.4681	0.0958
Error Variances	**Class1**	**Class2**	**Class3**					**Mean**	**Std.Dev.**
rating	0.9322	0.9879	0.7662					0.9140	0.0732
Model for Classes									
Intercept	**Class1**	**Class2**	**Class3**	**Wald**	**p-value**				
	0.7388	-0.2833	-0.4555	60.8273	6.2e-14				

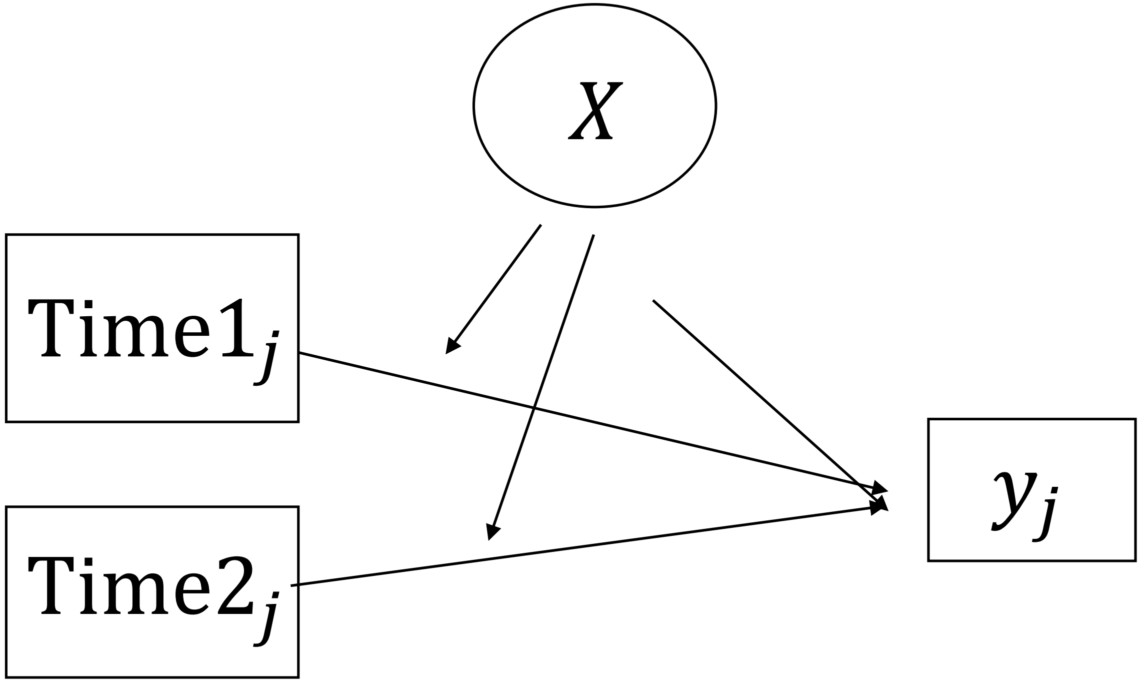
Other applications of the LC regression model

- Two-level data sets:
  - Individuals nested within groups
  - Intercept and effects of individual-level predictors may differ across latent classes of groups
  - For example, pupils nested within schools; latent classes of schools differ in average performance and in the effects of pupil-level predictors on performance

- Longitudinal data sets:
  - Repeated measurements nested within individuals
  - Time variables are used as predictors
  - Latent classes differ in the intercept and the time effects; thus in their initial/starting position and in the way they change over time
  - We call this a LC growth model

LC growth model

- LC regression model for repeated measures over time and predictors which are time variables



Example: dfg.sav

- Data from an experience sampling study (164 students, 7 days per student, 8 measurements per day)
- I will use “content” as the (binary) dependent variable (I recode category 2 within LatentGOLD to 1)
- I will use DayID as the id variable, so the aim is to identify latent classes of days (note that I ignore we have multiple days per person)
- I will use time (in minutes) as predictor, with the B-spline(3,2) option (thus a rather flexible time function)

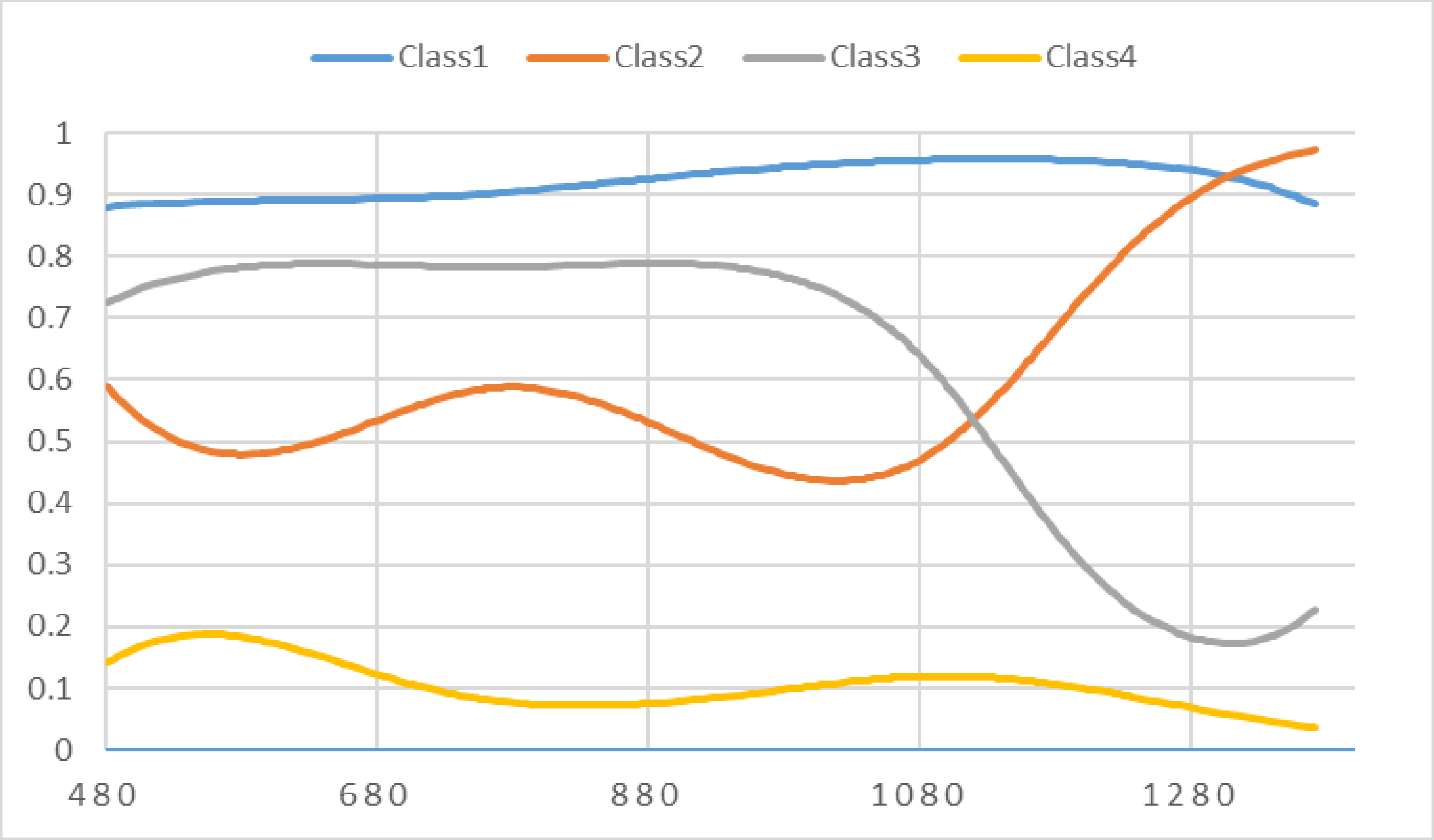
Example: dfg.sav (fit measures)

- Content 0 versus 1/2; B-spline(3,2) for time variable

	LL	BIC(LL)	AIC(LL)	AIC3(LL)	Npar
1-Class Regression	-4138.0780	8318.3939	8288.1559	8294.1559	6
2-Class Regression	-3699.9696	7491.4547	7425.9391	7438.9391	13
3-Class Regression	-3653.4492	7447.6916	7346.8984	7366.8984	20
4-Class Regression	-3619.6106	7429.2919	7293.2211	7320.2211	27
5-Class Regression	-3599.5734	7438.4953	7267.1468	7301.1468	34
6-Class Regression	-3587.1041	7462.8343	7256.2083	7297.2083	41

Example: dfg.sav (EstimatedValues)

- 4 content trajectories/classes during the day:



Example: dfg.sav (extensions)

- Covariates gender, age, and day of the week
- Using multiple dependent variables
- Taking into account measurement error in the dependent variable(s)
- Taking into account that days are nested within persons, for example, with a multilevel version of the LC growth model
- Taking into account that persons tend to stay in the same state, for example, using a (latent) Markov structure
- Crayen et al. (2012) does all of these