

How collinearity affects mixture regression results

Jan-Michael Becker · Christian M. Ringle ·
Marko Sarstedt · Franziska Völckner

Published online: 7 May 2014
© Springer Science+Business Media New York 2014

Abstract Mixture regression models are an important method for uncovering unobserved heterogeneity. A fundamental challenge in their application relates to the identification of the appropriate number of segments to retain from the data. Prior research has provided several simulation studies that compare the performance of different segment retention criteria. Although collinearity between the predictor variables is a common phenomenon in regression models, its effect on the performance of these criteria has not been analyzed thus far. We address this gap in research by examining the performance of segment retention criteria in mixture regression models characterized by systematically increased collinearity levels. The results have fundamental implications and provide guidance for using mixture regression models in empirical (marketing) studies.

Electronic supplementary material The online version of this article (doi:10.1007/s11002-014-9299-9) contains supplementary material, which is available to authorized users.

J.-M. Becker · F. Völckner (✉)
Department of Marketing and Brand Management, University of Cologne, Albertus Magnus Platz 1,
50923 Cologne, Germany
e-mail: voelckner@wiso.uni-koeln.de

J.-M. Becker
e-mail: j.becker@wiso.uni-koeln.de

C. M. Ringle
Institute of Human Resource Management and Organizations (HRMO),
Hamburg University of Technology (TUHH), Schwarzenbergstraße 95, 21073 Hamburg, Germany
e-mail: c.ringle@tuhh.de

M. Sarstedt
Institute of Marketing, Otto-von-Guericke-University Magdeburg,
Universitätsplatz 2, 39106 Magdeburg, Germany
e-mail: Marko.Sarstedt@ovgu.de

C. M. Ringle · M. Sarstedt
School of Business and Law, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

Keywords Market segmentation · Segment retention · Mixture regression · Collinearity

1 Introduction

When researchers use regression models, they typically assume that the regression coefficients are constant and that the same regression equation can satisfactorily describe all the population members. More often than not, this assumption of homogeneous data structures does not hold (Wedel et al. 1999). If not addressed adequately, heterogeneity leads to inaccurate results and inappropriate conclusions (Jedidi et al. 1997). To account for heterogeneity, researchers usually rely on observable characteristics, which allow for partitioning data into subsets, or they introduce interaction terms to consider group-related effects. However, in the light of the manifold limitations of those approaches—especially in the context of model-based clustering (Wedel and Kamakura 2000; Jedidi et al. 1997)—research interest during the last decade has been devoted to uncovering unobserved heterogeneity. Numerous researchers emphasize the importance of accounting for unobserved heterogeneity and call for a routine application of appropriate analysis techniques (e.g., Hutchinson et al. 2000; Grewal et al. 2010).

In this context, mixture regression models have gained increasing prominence. They help researchers uncover and treat unobserved heterogeneity and decide on a suitable number of segments to retain from the data. This method clusters observations of a pre-determined number of segments, while simultaneously estimating the parameters, thus avoiding well-known biases, which occur when segment-specific models are evaluated independently (e.g., DeSarbo and Cron 1988; McLachlan and Peel 2000). Not surprisingly, studies using mixture regression models are abundant in marketing research and other business disciplines (Andrews et al. 2007; Grewal et al. 2008; Mantrala et al. 2007). The increasing dissemination of finite mixture conjoint models (Andrews et al. 2002a, b; Jagpal et al. 2007) and the finite mixture partial least squares approach (Hahn et al. 2002; Sarstedt and Ringle 2010), both of which are based on the mixture regression concept, has further bolstered their popularity. With recent research supporting mixture regression models' ability to reveal segmentation structures in markets and the nature of regression relationships in segments (Andrews et al. 2010), their importance for empirical research is expected to further increase.

In the application of mixture regression models, a fundamental challenge is the selection of the number of segments to retain from the data. A priori, this number is unknown in most empirical applications, but has a substantial effect on results' interpretation. A misspecified number of segments results in under- or oversegmentation, which easily leads to inaccurate management decisions regarding, for example, customer targeting, product positioning, or determining the optimal marketing mix (Andrews and Currim 2003a). Specifically, if managers underestimate the number of customer segments in a market, they fail to identify distinct segments. If certain segments, such as lucrative niches, are ignored, companies may miss the opportunity to gain revenues from these customers due to their inability to address them separately and to satisfy their varying needs more precisely (Boone and Roehm 2002). In contrast, oversegmentation is likely to cause specific marketing activities that target non-relevant segments, resulting in a misallocation of resources. Given that companies spend considerable

sums developing and targeting market segments, the costs of overestimating the appropriate number of segments can be substantial.

In order to determine the number of segments to retain from the data, researchers can use a broad range of segment retention criteria (i.e., information and classification criteria) to compare different segmentation solutions in terms of their model fit (based on likelihood or entropy measures; Claeskens and Hart 2009; Hawkins et al. 2001). Previous research has provided simulation studies that compare the efficacy of different segment retention criteria to identify a pre-specified number of segments in different situations that could potentially affect the criteria's performance (e.g., Hawkins et al. 2001). In this line of research, Andrews and Currim (2003b) provide the most comprehensive study by investigating the performance of seven segment retention criteria. Their results reveal that researchers should revert to AIC_3 (Akaike's Information Criterion with a penalty factor of 3; Bozdogan 1994), as it yields the highest success rate and only a minor underfitting across a wide variety of data constellations.

A key aspect that neither Andrews and Currim (2003b) nor other studies in this research stream (e.g., Hawkins et al. 2001; Sarstedt 2008) have considered is how collinearity affects the retention criteria's performance. Researchers working with regression-based marketing models usually face situations with collinearity between two or more predictor variables. Collinearity leads to analytical problems such as unstable estimates of the regression coefficients and inflated standard errors (Mason and Perreault 1991; Ofir and Khuri 1986). Mixture regression models intensify these collinearity effects, since estimating the regression model in each segment relies on fewer observations than estimating it at the aggregate data level (DeSarbo et al. 2004). However, collinearity not only affects the mixture regression coefficients and the estimates' standard errors—a characteristic that is problematic in studies that aim at identifying segment structures—but also the likelihood of the model and the observations' probabilities of segment membership, because collinearity between the predictor variables masks some of the population heterogeneity (DeSarbo et al. 2004). Since information criteria primarily build on the model's likelihood values, increased levels of collinearity negatively affect their ability to determine the underlying number of segments. The same holds true for classification criteria that primarily rely on the entropy, which is a function of the segment membership probabilities (McLachlan and Peel 2000). Consequently, in the presence of collinearity, segment retention criteria are likely to misspecify the number of segments, fostering inaccurate management implications.

Against this background, we provide the first study on the effect of collinearity in mixture regression models. Specifically, we analyze several segment retention criteria and the impact that increased collinearity levels between the predictor variables have on their performance. The results show that collinearity has a substantial impact on the segment retention criteria's performance and identify four particularly well-performing criteria, as well as a critical level of collinearity for their effective application. Our study's findings contribute to the knowledge of mixture regression models and their application in the presence of collinearity levels that marketing researchers and practitioners are likely to encounter in empirical applications.

2 Study and simulation design

Reverting to prior studies (e.g., Claeskens and Hart 2009; Hawkins et al. 2001) allows us to identify and select nine information criteria (i.e., AIC, AIC_3 , AIC_4 , BIC, CAIC,

HQ, ICOMP, MDL₂, and MDL₅) and eight classification criteria (i.e., AWE, CLC, EN, ICL-BIC, PC, PE, NFI, and NEC) to decide on the number of segments in mixture regression models.¹ To evaluate their performance, we systematically manipulate seven data characteristics (i.e., factors) on different levels. The selection of factors and factor levels draws primarily on Andrews and Currim's (2003b) seminal study on the performance of segment retention criteria. Specifically, we consider the following factors and levels: *number of segments* [2; 3; 4], *sample size* [5×100; 10×100; 5×300; 10×300],² *explained variance* R^2 [40 %; 60 %; 80 %], *mean separation between (standardized) segment-specific coefficients* [0.2; 0.3; 0.4],³ and *relative segment size* [balanced; unbalanced; very unbalanced].⁴ In line with Mason and Perreault's (1991) seminal study on the effects of collinearity in regular regression models, we consider four continuous predictor variables and one dependent variable.

Most importantly, our research considers the effects of collinearity between the predictor variables on the segmentation outcome. We generate data with different collinearity levels, following Mason and Perreault's (1991) study on regular regression models and Grewal et al.'s (2004) study on structural equation models. To summarize these authors' approaches, data is generated on the basis of a pre-specified correlation matrix of the four independent variables in the regression model. The higher the correlation between the independent variables, the more severe the presence of collinearity (Kim et al. 2013). In line with Mason and Perreault (1991), we use the variance inflation factor (VIF) to characterize the *level of collinearity*. Our study considers the situation without collinearity and eight levels of collinearity, representing levels that marketing researchers and practitioners regularly encounter in (mixture) regression studies. In particular, the pre-specified correlation matrices of these nine situations, which we use for data generation, result in VIF values of [1.00; 1.35; 1.80; 2.40; 3.00; 3.80; 5.50; 7.30; 10.70].⁵

In addition, we consider whether the segments exhibit the same correlation matrix in terms of the independent variables or not. Usually, heterogeneity between segments is represented in terms of the regression coefficients' heterogeneity and, therefore, in terms of different correlations between the independent and the dependent variables. However, heterogeneity can also occur in the correlations between the independent variables, thus resulting in different correlation matrices across segments in terms of the independent variables. This situation is not unlikely, because the entire covariance matrix between all the variables can be segment-specific and not only the covariance

¹ For the mathematical specification of the criteria, see Table A1 in the Online Supplement of this paper.

² These factor levels combine Andrews and Currim's (2003b) two factors "number of individuals" (100 or 300) and "number of observations per individual" (5 or 10).

³ Note that prior studies used unstandardized mean separations with a random distribution of coefficients, which makes a full replication impossible as detailed information on the specified variances is missing.

⁴ The balanced factor level involves equally sized segments, while the unbalanced factor levels characterize the existence of one segment that is considerably larger than the other segments. Specifically, the unbalanced segments exhibit the following relative sizes: 65 %/35 % (unbalanced) and 80 %/20 % (very unbalanced) in a situation with two segments, 50 %/25 %/25 % (unbalanced) and 66.66 %/16.66 %/16.66 % (very unbalanced) in the case of three segments, and 40 %/20 %/20 %/20 % (unbalanced) and 55 %/15 %/15 %/15 % (very unbalanced) in the case of four segments.

⁵ For the correlation matrices, see Table A2 in the Online Supplement of this paper.

between the dependent and the independent variables (Marcoulides et al. 2012). To account for this phenomenon, we distinguish between two situations: consistent versus inconsistent *between-segment correlation matrices* [consistent; inconsistent]. In the first condition, the correlation matrix is the same across all segments, independently of the true segment differences. Thus, the independent variables (e.g., X_1 and X_2) always exhibit the same correlation in all the segments that depend on the collinearity level. In contrast, in the second condition, the correlation matrix differs across segments, because it is aligned with the pre-specified, segment-specific differences; consequently, high correlations occur between those variables that separate the segments more strongly. Hence, two specific independent variables (e.g., X_1 and X_2) could have a high correlation in one segment and a low correlation in another. Even though we use the same level of collinearity (i.e., the same correlation pattern, but assigned to different independent variables) in both situations (i.e., consistent and inconsistent), we expect that it is more difficult to identify the segments in the second condition, leading to a decline in performance of the segment retention criteria.⁶

The full factorial design of the study results in $2 \times 3^4 \times 4 \times 9 = 5,832$ different combinations of the design factors. We created a program with the statistical software R (R Core Team 2014) to generate sets of data for the pre-specified factor levels. To analyze the simulated data, we use the R package FlexMix (Grün and Leisch 2008), which uses an expectation-maximization (EM) algorithm to solve the maximum likelihood estimation. To ensure the robustness of the results, 30 datasets per factor combination are generated and analyzed. Moreover, since the EM algorithm does not always converge to the global optimum, we select the best log-likelihood result of ten FlexMix runs with different starting values per dataset (Wedel and Kamakura 2000).

3 Results

3.1 Performance of segment retention criteria

Table 1 shows the average success (S), the underestimation (U), and overestimation (O) rates of the information and classification criteria. According to these results, AIC_4 , for example, identifies the correct number of segments in 82 % of all simulation runs where the number of segments is held constant at $S=2$ (Factor 1) and all the other factors are varied according to the factor levels described above. Similarly, at this factor level, AIC_4 underestimates (overestimates) the correct number of segments in 11 % (6 %) ⁷ of all simulation runs.

Overall, our results are similar to those of Andrews and Currim (2003b) when *no collinearity* is present (i.e., $VIF=1.0$). For instance, Andrews and Currim (2003b) find that AIC identifies the pre-specified number of segments in 59 % of the cases, while it underestimates (overestimates) this number in 10 % (30 %) of the cases. These results compare well with ours, where AIC's success rates, as well as its under- and

⁶ For an illustration of the difference between consistent and inconsistent correlation matrices between segments, see Table A3 in the Online Supplement of this paper.

⁷ Note that the numbers do not always add to 100 % because of rounding inaccuracies. The more precise numbers of 82.38, 11.21, and 6.41 % add to 100 %.

Table 1 Performance results

	AIC (%)			AIC ₃ (%)			AIC ₄ (%)			BIC (%)			CAIC (%)			ICOMP (%)			HQ (%)			MDL ₂ (%)			MDL ₅ (%)				
	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S
Groups	3	51	45	7	78	15	11	82	6	20	80	1	22	77	1	9	74	17	11	83	7	34	66	0	59	41	0		
	19	44	37	36	57	7	46	53	1	60	40	0	63	37	0	30	56	15	45	54	1	75	25	0	90	10	0		
	32	33	34	55	39	6	65	34	1	77	23	0	80	20	0	43	41	16	64	34	1	88	12	0	96	4	0		
Sample Size	27	31	43	50	41	9	60	38	2	70	29	0	73	27	0	35	41	24	57	39	3	83	17	0	95	5	0		
	19	51	30	32	63	5	39	60	1	50	50	0	53	47	0	27	62	12	38	61	1	65	35	0	83	17	0		
	18	38	44	33	56	11	41	55	4	54	45	0	57	43	0	29	54	18	41	55	4	68	32	0	84	16	0		
	9	52	39	17	72	11	22	73	5	34	66	1	36	64	0	18	72	11	23	73	4	46	54	0	65	35	0		
Relative Segment Size	16	40	43	30	59	11	38	59	4	49	50	1	52	48	0	24	58	17	37	59	4	63	37	0	79	21	0		
	21	46	33	37	57	7	45	54	2	56	44	0	59	41	0	31	55	14	44	54	2	70	30	0	86	14	0		
	18	42	40	32	58	10	40	57	3	51	49	0	54	46	0	26	57	17	39	58	3	65	35	0	81	19	0		
Mean Separation	33	39	28	52	42	5	61	37	1	73	27	0	76	24	0	48	41	11	61	38	1	85	15	0	96	4	0		
	15	44	41	29	61	10	37	60	3	49	51	0	52	48	0	22	60	18	36	60	3	64	36	0	82	18	0		
	7	46	48	17	70	13	24	72	4	34	65	1	37	63	0	11	70	20	23	73	5	48	52	0	68	32	0		
R ²	28	41	31	46	47	7	56	43	2	69	31	0	71	29	0	3	44	19	55	43	2	81	19	0	93	7	0		
	19	43	38	34	57	9	42	55	3	54	45	0	57	43	0	28	54	17	42	55	3	69	31	0	85	15	0		
	8	44	48	18	70	12	24	72	4	34	66	0	36	63	0	16	72	12	23	73	4	48	52	0	68	32	0		
VIF	3	52	44	10	81	9	15	83	2	25	75	0	28	72	0	6	82	12	14	84	2	42	58	0	68	32	0		
	7	50	43	16	75	9	23	76	2	34	66	0	37	62	0	12	74	14	22	76	2	51	49	0	74	26	0		
	11	47	42	23	69	9	30	69	2	42	58	0	45	55	0	17	68	15	29	69	2	59	41	0	79	21	0		
	14	44	42	28	62	10	36	61	2	49	51	0	52	48	0	22	61	17	35	62	3	64	36	0	81	19	0		
	17	42	41	33	57	10	41	55	3	54	45	0	57	43	0	26	55	18	41	56	3	68	32	0	83	17	0		
	22	40	39	38	52	10	47	50	4	59	41	0	62	38	0	31	50	18	46	50	4	72	28	0	86	14	0		

Table 1 (continued)

	AIC (%)		AIC ₃ (%)		AIC ₄ (%)		BIC (%)		CAIC (%)		ICOMP (%)		HQ (%)		MDL ₂ (%)		MDL ₅ (%)										
	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S									
5.50	26	38	36	44	46	10	53	43	4	65	35	1	67	33	0	38	45	18	52	43	4	76	24	0	88	12	0
7.30	30	37	33	49	42	9	58	39	4	69	31	1	71	29	0	42	41	17	57	39	4	79	21	0	89	11	0
10.70	35	36	30	55	37	8	63	33	3	73	26	1	75	25	1	49	37	14	63	34	3	82	18	0	90	10	0
Consistent	13	49	38	26	67	7	34	65	1	45	55	0	47	53	0	20	66	14	33	66	2	59	41	0	77	23	0
Inconsistent	23	37	39	39	49	12	48	48	4	60	40	1	62	37	0	34	48	18	47	48	5	73	27	0	87	13	0
Overall	18	43	39	33	58	9	41	57	3	52	47	0	55	45	0	27	57	16	40	57	3	66	34	0	82	18	0

	NEC (%)		PE (%)		EN (%)		NFI (%)		PC (%)		AWE (%)		CLC (%)		ICL-BIC (%)									
	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S								
Groups	*	75	25	*	92	8	*	56	44	*	67	33	*	93	7	51	49	0	19	41	40	34	66	0
	53	24	32	86	11	3	36	23	41	54	24	22	86	12	3	85	15	0	37	20	43	70	30	0
	61	15	24	94	4	1	53	14	33	73	13	14	94	4	1	93	7	0	45	11	44	83	17	0
Sample Size	26	27	47	59	33	8	24	24	52	35	28	37	59	34	8	91	9	0	26	10	65	77	23	0
	40	40	20	59	38	4	32	33	35	44	37	19	59	38	3	74	26	0	16	17	67	56	44	0
	40	38	22	63	34	3	27	27	46	43	32	25	63	35	3	80	20	0	59	26	15	70	30	0
	46	47	7	60	39	1	36	39	25	48	41	11	60	39	1	59	41	0	35	44	22	46	54	0
Relative Segment Size	38	37	25	62	34	4	31	27	42	47	30	23	62	34	4	74	26	0	33	21	46	60	40	0
	38	40	22	58	38	4	29	36	35	38	40	22	57	39	4	79	21	0	36	28	36	65	35	0
	38	37	25	60	36	4	29	30	41	42	34	24	60	36	4	75	25	0	33	23	44	62	38	0
Mean Separation	32	31	38	61	33	6	25	22	53	38	29	33	61	33	5	93	7	0	56	16	28	85	15	0

Table 1 (continued)

	NEC (%)			PE (%)			EN (%)			NFI (%)			PC (%)			AWE (%)			CLC (%)			ICL-BIC (%)		
	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O	U	S	O
0.30	38	40	22	61	36	3	30	31	39	43	36	21	61	36	3	75	25	0	30	27	44	60	40	0
0.40	44	44	12	59	39	3	34	39	27	46	40	15	58	39	3	59	41	0	16	29	55	43	57	0
0.40	37	31	32	62	33	5	28	24	48	42	29	28	62	33	5	90	10	0	48	20	32	80	20	0
0.60	37	38	25	61	35	4	29	30	41	42	34	24	61	36	3	79	21	0	36	24	40	65	35	0
0.80	39	45	15	57	39	3	32	39	29	42	41	17	57	40	3	59	41	0	17	28	55	42	58	0
1.00	8	63	29	48	47	5	6	45	49	18	51	31	47	48	5	56	44	0	11	30	59	35	65	0
1.35	19	53	28	54	42	4	14	40	46	29	44	27	53	42	4	65	35	0	18	30	52	44	56	0
1.80	32	43	25	59	37	4	24	34	41	39	38	23	59	37	3	70	30	0	23	29	48	53	47	0
2.40	40	37	23	62	35	3	30	31	38	45	34	21	62	35	3	76	24	0	29	26	44	60	40	0
3.00	44	33	23	63	33	4	34	28	37	48	31	21	63	33	3	79	21	0	35	24	41	65	35	0
3.80	48	31	22	63	33	4	37	27	37	50	30	21	64	33	3	81	19	0	39	23	38	70	30	0
5.50	50	29	22	63	33	4	39	25	35	51	29	21	64	33	4	84	16	0	45	20	35	75	25	0
7.50	50	28	22	63	32	4	40	24	36	51	28	21	64	33	4	86	14	0	50	18	32	78	22	0
10.70	51	27	22	64	32	4	41	24	35	51	27	22	64	32	4	87	13	0	55	16	29	81	19	0
Consistent	45	37	18	61	35	3	40	28	33	49	31	19	61	36	3	70	30	0	26	25	49	55	45	0
Inconsistent	31	39	30	59	37	5	19	34	46	35	38	27	59	37	4	82	18	0	42	23	35	70	30	0
Overall	38	38	24	60	36	4	30	31	40	42	35	23	60	36	4	76	24	0	34	24	42	62	38	0

Bold is the column which represents the percentage of successful segment retention

U underestimation, *S* success, *O* overestimation

*Underestimation not possible because criterion does not provide a solution for one segment

overestimation rates, are 52, 3, and 44 % when no collinearity is present. More pronounced differences can be observed regarding Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC).

Taking the full range of design factors into account, the results show that AIC_3 , AIC_4 , HQ, and ICOMP are the best-performing criteria with overall success rates of 57 to 58 %. With regard to deviations from the pre-specified number of segments, information complexity criterion (ICOMP) exhibits both over- and underestimation tendencies, whereas AIC_3 , AIC_4 , and HQ (Hannan-Quinn criterion) show clear underestimation tendencies. AIC, BIC, and CAIC achieve success rates of between 43 and 47 %, whereas all the other criteria have overall success rates of only 38 % and lower (Table 1).

The classification criteria that do not allow for computing a one-segment solution (i.e., NEC, PE, EN, NFI, and PC) also achieve relatively high overall success rates regarding two groups, which is not surprising, as no underestimation is possible in this situation. However, the success rates of these criteria regarding three and four groups decline strongly, regardless of the other data characteristics. Hence, it is usually not advisable to apply these criteria in mixture regression model analyses to determine the number of segments.

In the following, we will focus our analysis of the collinearity effects on the best-performing information and classification criteria in this study; that is, AIC_3 , AIC_4 , ICOMP, and HQ. Our study reveals two major results. First, the segment retention criteria’s overall performance declines significantly at higher levels of collinearity.

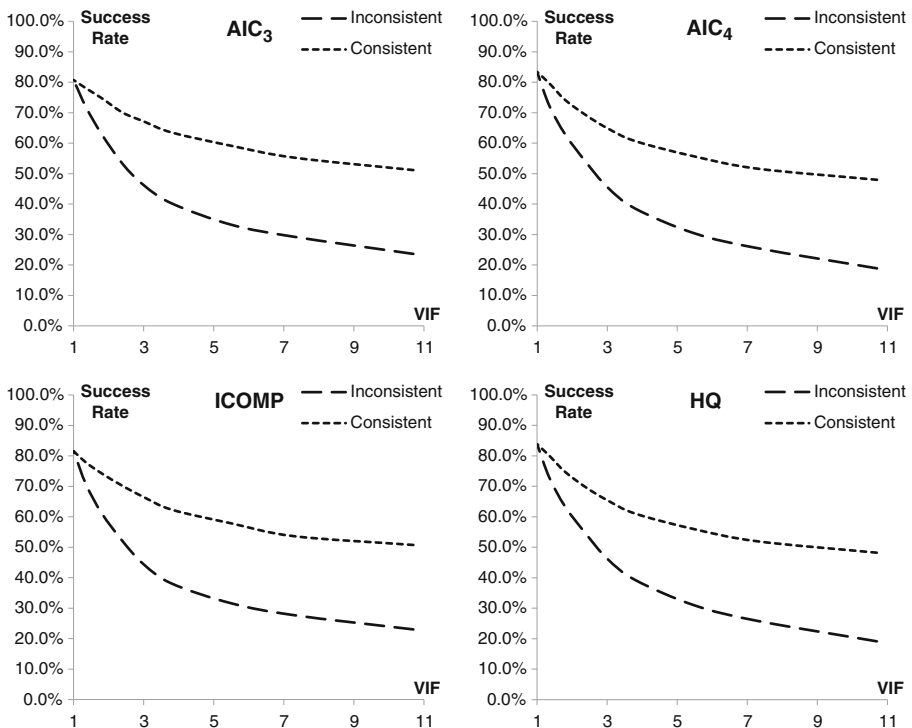


Fig. 1 Success rates for different levels of collinearity and consistent vs. inconsistent correlation matrix (four best-performing Criteria)

Second, higher levels of collinearity substantially increase the criteria's underestimation rates, while their overestimation rates generally remain unaffected.

On average, the success rates of the best-performing criteria (i.e., AIC₃, AIC₄, ICOMP, and HQ) decrease by 21, 32, and 47 percentage points when the VIF increases from 1.0 to low (VIF=2.4), intermediate (VIF=3.8), and large (VIF=10.7) values. While the success rates of AIC₃ and ICOMP decrease by 44 and 45 percentage points, those of AIC₄ and HQ even decline by 50 percentage points. These declines in the success rates go hand in hand with pronounced increases in the criteria's underestimation tendencies. Thus, when collinearity is present, the criteria offer only limited guidance in terms of clearly pinpointing the number of segments to extract; instead, their underestimation tendency should be taken as evidence of a higher number of segments than actually indicated.

In addition, we also analyze the effect of inconsistent versus consistent correlation matrices across segments. When correlation matrices are inconsistent across segments, the segment retention criteria's ability to correctly detect the pre-specified number of segments is considerably reduced. On average, the success rate of the best-performing criteria (i.e., AIC₃, AIC₄, HQ, and ICOMP) declines by 18 percentage points when the correlation matrix is inconsistent instead of consistent across segments (Table 1).

Moreover, we find a strong interaction effect between collinearity and the consistency of the correlation matrix across segments (Fig. 1). Correlation matrices that are inconsistent across segments reinforce collinearity's negative effect on the success rates, while correlation matrices that are consistent across segments only moderately affect the success rates. A possible explanation for collinearity's strong effect in situations with inconsistent correlation matrices across segments might be that these models have not been properly identified (Hennig 2000). One of the identification conditions for mixture regression models requires the independent variable matrix in each segment to not be singular (DeSarbo et al. 2007). Moreover, the predictor variables should not be collinear but should also be prevented from being too close to collinearity to derive consistent estimates (Hennig 2000).

To shed further light on the simulation design factors' effects on the criteria's performance, we meta-analyzed the results of the four best-performing criteria (AIC₃, AIC₄, ICOMP, and HQ) by fitting multinomial logistic regression models to the success data. We coded the dependent variable as -1 if the criterion underestimated the correct number of segments, 0 if the criterion identified the correct number of segments (reference category), and 1 if the criterion overestimated the correct number of segments. We confirmed the above results as increasing the number of segments, or the level of collinearity, significantly increases the log-odds of underestimating the correct number of segments, while increasing the sample size, the explained variance, or the mean separation between segment-specific coefficients, significantly reduces the log-odds of underestimating the correct number of segments. Moreover, an inconsistent correlation matrix across segments significantly increases the log-odds of underestimating the correct number of segments.⁸

3.2 Quality of correctly extracted segments

The next step of the analysis involves assessing the clustering accuracy of the identified segments.⁹ A well-known criterion for this purpose is the adjusted Rand index (ARI;

⁸ For detailed results, see Table A4 in the Online Supplement of this paper.

⁹ We thank an anonymous reviewer for this suggestion.

Table 2 Adjusted rand index for the four best-performing criteria

		Pre-specified	AIC ₃	AIC ₄	ICOMP	HQ
Groups	2	<i>0.631</i>	0.602	0.608	0.596	0.609
	3	<i>0.460</i>	0.453	0.448	0.446	0.450
	4	<i>0.366</i>	0.359	0.351	0.354	0.353
Sample size	5×100	<i>0.360</i>	0.339	0.330	0.331	0.334
	10×100	<i>0.583</i>	0.572	0.567	0.564	0.569
	5×300	<i>0.392</i>	0.379	0.380	0.373	0.380
	10×300	<i>0.609</i>	0.595	0.598	0.594	0.598
Relative segment size	Balanced	<i>0.476</i>	0.456	0.455	0.452	0.456
	Unbalanced	<i>0.482</i>	0.466	0.464	0.461	0.466
	Very unbalanced	<i>0.500</i>	0.491	0.487	0.483	0.489
Mean separation	0.20	<i>0.288</i>	0.270	0.259	0.264	0.262
	0.30	<i>0.510</i>	0.496	0.494	0.491	0.496
	0.40	<i>0.660</i>	0.647	0.653	0.641	0.653
R ²	0.40	<i>0.346</i>	0.330	0.323	0.321	0.325
	0.60	<i>0.461</i>	0.447	0.443	0.439	0.445
	0.80	<i>0.650</i>	0.636	0.640	0.636	0.641
VIF	1.00	<i>0.669</i>	0.657	0.655	0.654	0.656
	1.35	<i>0.608</i>	0.594	0.590	0.590	0.592
	1.80	<i>0.558</i>	0.543	0.540	0.537	0.542
	2.40	<i>0.513</i>	0.497	0.495	0.491	0.497
	3.00	<i>0.478</i>	0.460	0.458	0.453	0.460
	3.80	<i>0.443</i>	0.425	0.423	0.418	0.424
	5.50	<i>0.402</i>	0.386	0.385	0.379	0.386
	7.30	<i>0.369</i>	0.355	0.353	0.348	0.354
	10.70	<i>0.333</i>	0.322	0.320	0.317	0.321
	Correlation matrix	Consistent	<i>0.568</i>	0.559	0.555	0.554
Inconsistent		<i>0.403</i>	0.383	0.383	0.376	0.385
Overall		<i>0.486</i>	0.471	0.469	0.465	0.470

Italic is the column with the pre-specified number of groups

Hubert and Arabi 1985), which depicts the similarity between the clustering of the identified segments and the pre-specified groups. An ARI value of 0 indicates the expected value of random (uniform) clustering solutions, whereas an ARI value close to its maximum value of 1 indicates that the segmentation solution substantially matches the expected data grouping, and thus has a high clustering accuracy (Hubert and Arabi 1985).

Table 2 shows the average ARI values of the four best-performing segment retention criteria and all the factor levels analyzed in this study.¹⁰ We computed ARI values for (1) the number of segments that each segment retention criterion indicated and (2) for

¹⁰ For the complete table with all criteria's results, see Table A5 in the Online Supplement of this paper.

the pre-specified number of segments. The ARI values of the pre-specified number of segments serve as benchmarks for the criteria's ARI values, as this value would result if the correct number of segments were chosen. Analogous to the success rates, AIC₃, AIC₄, ICOMP, and HQ show the best overall performance with ARI values between 0.465 and 0.471, which are close to the benchmark value of 0.486.

Our analyses show that those factors that negatively affect the success rates of the segment retention criteria also negatively affect their ARI values. Most notably, we find that at higher levels of collinearity, the ARI declines considerably. For example, the ARI value of AIC₃ declines from 0.657 in situations without collinearity (VIF=1.0) to 0.322 in situations with very high collinearity (VIF=10.7). On average, the ARI values of the best-performing criteria (i.e., AIC₃, AIC₄, ICOMP, and HQ) show an overall decline (i.e., the difference between the situation without collinearity and the highest

Table 3 Parameter recovery accuracy

		ME	MAE	RMSE
Groups	2	-0.0027	0.051	0.065
	3	-0.0019	0.080	0.107
	4	-0.0016	0.105	0.145
Sample Size	5×100	-0.0025	0.122	0.166
	10×100	-0.0013	0.065	0.088
	5×300	-0.0028	0.083	0.111
	10×300	-0.0016	0.044	0.059
Relative Segment Size	Balanced	-0.0020	0.073	0.097
	Unbalanced	-0.0019	0.076	0.103
	Very unbalanced	-0.0023	0.086	0.118
Mean Separation	0.20	-0.0020	0.097	0.130
	0.30	-0.0020	0.077	0.104
	0.40	-0.0023	0.060	0.083
R ²	0.40	-0.0029	0.113	0.151
	0.60	-0.0021	0.081	0.109
	0.80	-0.0012	0.042	0.058
VIF	1.00	-0.0001	0.027	0.035
	1.35	-0.0022	0.039	0.051
	1.80	-0.0025	0.051	0.067
	2.40	-0.0023	0.064	0.084
	3.00	-0.0022	0.075	0.099
	3.80	-0.0023	0.087	0.117
	5.50	-0.0024	0.104	0.141
	7.30	-0.0025	0.119	0.163
10.70	-0.0022	0.140	0.194	
Correlation Matrix	Consistent	0.0000	0.059	0.083
	Inconsistent	-0.0042	0.098	0.129
Overall		-0.0020	0.078	0.106

collinearity level) of 0.336. Furthermore, settings with inconsistent correlation matrices entail considerably lower ARI values.

However, collinearity also affects the benchmark ARI values that one would obtain if the correct pre-specified number of segments were chosen. This result implies that even if one identified the correct number of segments, the collinearity between the predictors would negatively influence the mixture regression results. Hence, collinearity not only influences the segment retention negatively, but also the quality of the mixture regression results in general.

We also ran separate ANCOVAs for each of the four best-performing criteria, using the ARI values as dependent variables. The analysis results are highly consistent regarding AIC_3 , AIC_4 , $ICOMP$, and HQ , showing that all factors have a significant impact on the ARI.¹¹ Moreover, the vast majority of the two-way interaction effects between the factors are also significant ($p \leq 0.01$), but of minor relevance (based on their partial η^2 , which is generally below 0.05), with the exception of the interaction between the number of segments and the consistent vs. inconsistent correlation matrix, which has a partial η^2 of around 0.065 as well as VIF and the consistent vs. inconsistent correlation matrix, which has a partial η^2 of around 0.17. This result further illustrates the importance of the interaction between the level of collinearity and the consistency of the correlation matrix between segments, which we also show in Fig. 1.

3.3 Parameter recovery

To further investigate the effect of collinearity on the mixture regression results, we examine the parameter recovery errors that collinearity induces by comparing the estimated mixture regression coefficients with their pre-specified values. Table 3 shows the mean error (ME), mean absolute error (MAE), and root mean squared error (RMSE) of the parameter estimates.

We find a slightly negative ME (systematic negative bias) across all factor levels, indicating an underestimation of the mixture regression coefficients. In terms of collinearity effects, the ME remains relatively stable across all levels of collinearity but increases considerably in size when the correlation matrices are inconsistent rather than consistent. Furthermore, we find that the MAE and RMSE, which indicate the magnitude of the estimation error, increase considerably with stronger levels of collinearity. Specifically, the MAE (RMSE) increases from 0.027 (0.035) in situations without collinearity to 0.140 (0.190) in situations with very high collinearity (VIF=10.7). This result implies that the variance of the estimated regression coefficients strongly increases with increased levels of collinearity. This result is one we would also expect from a normal regression case (Mason and Perreault 1991). While the difference between the MAE and RMSE is relatively small in situations of low levels of collinearity, it increases with higher levels of collinearity. Hence, at higher levels of collinearity, the likelihood of obtaining segmentation results that differ significantly from the true population results increases.

¹¹ For the ANCOVA results, see Table A6 in the Online Supplement of this paper.

4 Conclusion and recommendations

This study extends the research on mixture regression models, most notably the study by Andrews and Currim (2003b), by examining the effect of collinearity on the performance of a broad range of segment retention criteria. When using mixture regression models, researchers should not apply (popular) segment retention criteria blindly—even at collinearity levels much smaller than those that popular textbooks on research methods judge as critical (e.g., a VIF value of 10; Hair et al. 2010). In particular, our analysis reveals that VIF levels as low as 2.4 affect the segment retention criteria's performance by dramatically increasing their underestimation tendencies. Our results also show that collinearity effects differ clearly across the criteria. Collinearity affects some criteria, such as AIC₃, AIC₄, BIC, CAIC, ICOMP, and HQ, more strongly than it does others (e.g., AIC, MDL₅, and CLC).

The segment retention criteria's underfitting tendencies in the presence of high collinearity are particularly problematic when the segment retention criteria indicate a one-segment solution, although the group-specific regression coefficients differ significantly across the segments. In these cases, ignoring heterogeneity and analyzing the data on the aggregate level can have adverse consequences for any conclusions drawn from such marketing models.

The negative effect of collinearity is even more pronounced if the correlation matrices are inconsistent across segments (i.e., when high correlations occur between those variables that separate the segments more strongly). This kind of situation is not unlikely, as the entire covariance matrix between all the variables can be segment-specific—not only the covariance between the dependent and the independent variables (Marcoulides et al. 2012). The overall sample covariance (correlation) matrix might not reveal the strong underlying group-specific collinearity because the correlations are equaled out when the two groups are combined. Researchers should therefore not only analyze the collinearity on the aggregate-data level when conducting a mixture regression analysis, but also analyze the group-specific correlations to assess whether collinearity threatens the segment retention. Only if the collinearity diagnostics suggest that multicollinearity is not an issue for the aggregate data level or the segment-specific outcomes, should researchers further interpret the mixture regression results and draw conclusions from them (Grewal et al. 2013).

It seems common practice for authors to justify their segment decision solely on the basis of segment retention criteria (e.g., Cortiñas et al. 2010; Dubois et al. 2005). Our results show that such an approach is not without problems and suggest three main recommendations for working with mixture regression models. *First*, given the criteria's increased underestimation tendencies even in the presence of low collinearity levels, researchers are well-advised to consider higher segment solutions than the criteria actually indicate, even if all criteria point at the same number of segments. This especially holds true when criteria point at a one-segment solution, thus suggesting that heterogeneity is not a problem. In such a situation, researchers should carefully evaluate whether using a segment-specific model provides plausible insights and an increased predictive ability (Andrews et al. 2007). *Second*, the risk of collinearity increases especially for survey-based research studies that do not use an orthogonal (experimental) design (e.g., cross-sectional studies with complex models that feature three- or four-way interactions), as higher-order interaction terms correlate increasingly with lower-order interaction terms and the main effects (Grewal et al. 2013). Hence, in mixture regression models with interaction terms (e.g., in moderated regression analyses), researchers should be extra cautious about increasing levels of collinearity. *Third*, the criteria should be perceived as providing a reasonable range of segment solutions that researchers

can evaluate by following guidelines such as those that Kotler and Keller (2012) provide (i.e., the segments should be actionable, differentiable, and substantial). In doing so, researchers need to separately evaluate the collinearity between the predictors in each segment when comparing different segment solutions. The corresponding results need to be routinely provided (e.g., the VIF values of each segment regarding different segment solutions), along with the segment retention criteria values, to increase confidence in the findings.

This research is subject to limitations, which future studies should address. *First*, researchers should develop segment retention criteria that perform well under varying levels of those design factors that are relevant for mixture regression models, including collinearity. *Second*, our study only provides evidence of collinearity effects on the performance of segment retention criteria regarding the multivariate normal mixture case. Future research could examine how collinearity affects their performance in other types of exponential family mixture models. *Third*, examining the effects of collinearity not only on parameter recovery but also on statistical power would be a fruitful research endeavor. Replicating Mason and Perreault's (1991) work in a *mixture* regression context might be a promising first step toward calibrating the conditions under which collinearity affects the interpretation of mixture regression results. *Fourth*, recent publications have presented alternative methods for applying mixture regression models that are more robust against irregularities in the data, such as strong collinearity between the predictor variables (Kim et al. 2012, 2013).¹² However, future research should assess the effect of collinearity on the segment retention capabilities of such novel approaches in greater detail.¹³ For example, as extant research on these approaches solely relies on the log marginal likelihood (LML) for model selection. Future studies should more broadly examine the efficacy of the LML under different model conditions (including collinearity), and, if necessary, develop alternative model selection heuristics.

Acknowledgments The authors would like to thank Jörg Henseler (Radboud University Nijmegen) and Edward E. Rigdon (Georgia State University) for their comments on earlier versions of the paper.

References

- Andrews, R. L., & Currim, I. S. (2003a). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, 40(20), 235–243.
- Andrews, R. L., & Currim, I. S. (2003b). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, 20(4), 315–321.

¹² For example, Kim et al. (2013) extend the new Bayesian latent structure regression model by Kim et al. (2012) by implementing model constraints and illustrating these in comparative analyses that contrast the performance of the proposed methodology with standard latent class finite mixture regression, as well as with traditional Bayesian finite mixture regression. The authors show that the new Bayesian regression model is more robust against collinearity problems than both the finite mixture regression models and traditional Bayesian finite mixture models in terms of the RMSE and ARI. In addition, the new Bayesian regression model can also be used to simultaneously select the number of segments and select the variables to retain per segment.

¹³ We thank an anonymous reviewer for these comments.

- Andrews, R. L., Ainsle, A., & Currim, I. S. (2002a). An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *Journal of Marketing Research*, 39(4), 479–487.
- Andrews, R. L., Ansari, A., & Currim, I. S. (2002b). Hierarchical Bayes versus finite mixture conjoint analysis models: a comparison of fit, prediction and partworth recovery. *Journal of Marketing Research*, 39(1), 87.
- Andrews, R. L., Currim, I. S., Leeflang, P., & Lim, J. (2007). Estimating the SCAN*PRO model of store sales: HB, FM or just OLS? *International Journal of Research in Marketing*, 25(1), 22–33.
- Andrews, R. L., Brusco, M. J., Currim, I. S., & Davis, B. (2010). An empirical comparison of methods for clustering problems: are there benefits from having a statistical model? *Review of Marketing Science*, 8(1), 1–32.
- Boone, D. S., & Roehm, M. (2002). Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion. *Marketing Letters*, 13(4), 317–333.
- Bozdogan, H. (1994). Mixture-model cluster analysis using model selection criteria in a new information measure of complexity. *Paper presented at the Proceedings of the First US/Japan Conference on Frontiers of Statistical Modelling: An Information Approach*.
- Claeskens, G., & Hart, J. D. (2009). Goodness-of-fit tests in mixed models. *Test*, 18(2), 213–239.
- Core Team, R. (2014). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Cortiñas, M., Chocarro, R., & Villanueva, M. L. (2010). Understanding multi-channel banking customers. *Journal of Business Research*, 63(11), 1215–1221.
- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(2), 249–282.
- DeSarbo, W. S., Kamakura, W., & Wedel, M. (2004). Applications of multivariate latent variable models in marketing. In Y. Wind & P. E. Green (Eds.), *Market Research and Modeling: Progress and Prospects. A Tribute to Paul E. Green* (pp. 43–68). Boston: Kluwer Academic Publishers. et al.
- DeSarbo, W. S., Benedetto, C. A., & Song, M. (2007). A heterogeneous resource based view for exploring relationships between firm performance and capabilities. *Journal of Modelling in Management*, 2(2), 103–130.
- Dubois, B., Czellar, S., & Laurent, G. (2005). Consumer segmentation based on attitudes toward luxury: empirical evidence from twenty countries. *Marketing Letters*, 16(2), 115–128.
- Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: implications for theory testing. *Marketing Science*, 23(4), 519–529.
- Grewal, R., Chakravarty, A., Ding, M., & Liechty, J. (2008). Counting chickens before the eggs hatch: associating new product development portfolios with shareholder expectations in the pharmaceutical sector. *International Journal of Research in Marketing*, 25(3), 261–272.
- Grewal, R., Chandrashekar, M., & Citrin, A. V. (2010). Customer satisfaction heterogeneity and shareholder value. *Journal of Marketing Research*, 47(4), 612–626.
- Grewal, R., Chandrashekar, M., Johnson, J. L., & Mallapragada, G. (2013). Environments, unobserved heterogeneity, and the effect of market orientation on outcomes for high-tech firms. *Journal of the Academy of Marketing Science*, 41(2), 206–233.
- Grün, B., & Leisch, F. (2008). Flexmix 2: finite mixtures with concomitant variables and varying constant parameters. *Journal of Statistical Software*, 28(4), 1–35.
- Hahn, C., Johnson, M. D., Herrmann, A., & Huber, F. (2002). Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review*, 54(3), 243–269.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Englewood Cliffs: Prentice Hall.
- Hawkins, D. S., Allen, D. M., & Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, 38(1), 15–48.
- Hennig, C. (2000). Identifiability of models for clusterwise linear regression. *Journal of Classification*, 17(2), 273–296.
- Hubert, L., & Arabi, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218.
- Hutchinson, J. W., Kamakura, W. A., & Lynch, J. G. (2000). Unobserved heterogeneity as an alternative explanation for “reversal” effects in behavioral research. *Journal of Consumer Research*, 27(3), 324–344.
- Jagpal, S., Jedidi, K., & Jamil, M. (2007). A multibrand concept-testing methodology for new product strategy. *Journal of Product Innovation Management*, 24(1), 34–51.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16(1), 39–59.
- Kim, B.-D., Fong, D. K. H., & DeSarbo, W. S. (2012). Model-based segmentation featuring simultaneous segment-level variable selection. *Journal of Marketing Research*, 49(5), 725–736.

- Kim, S., Blanchard, S. J., Desarbo, W. S., & Fong, D. K. H. (2013). Implementing managerial constraints in model-based segmentation: extensions of Kim, Fong, and DeSarbo (2012) with an application to heterogeneous perceptions of service quality. *Journal of Marketing Research*, 50(5), 664–673.
- Kotler, P., & Keller, K. L. (2012). *Marketing management* (14th ed.). Pearson: Prentice-Hall.
- Mantrala, M. K., Naik, P. A., Sridhar, S., & Thorson, E. (2007). Uphill or downhill? Locating the firm on a profit function. *Journal of Marketing*, 71(2), 26–44.
- Marcoulides, G. A., Chin, W. W., & Saunders, C. (2012). When imprecise statistical statements become problematic: a response to Goodhue, Lewis, and Thompson. *MIS Quarterly*, 36(3), 717–728.
- Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, 28(3), 268–280.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
- Ofir, C., & Khuri, A. (1986). Multicollinearity in marketing models: diagnostics and remedial measures. *International Journal of Research in Marketing*, 3(3), 181–205.
- Sarstedt, M. (2008). Market segmentation with mixture regression models: understanding measures that guide model selection. *Journal of Targeting, Measurement and Analysis for Marketing*, 16(3), 228–246.
- Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modelling: a comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics*, 37(8), 1299–1318.
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: conceptual and methodological foundations* (2nd ed.). Boston: Kluwer.
- Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., et al. (1999). Discrete and continuous representations of unobserved heterogeneity in choice modeling. *Marketing Letters*, 10(3), 219–232.