Report Series

The Heterogeneity Model and its Special Cases - an Illustrative Comparison

Regina Tüchler
Sylvia Frühwirth-Schnatter
Thomas Otter

Report No. 73
June 2002

Report Series

June 2002

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series
are preliminary versions of journal articles
and not for quotations.

The Heterogeneity Model and its Special Cases - an Illustrative Comparison

by

Regina Tüchler [1], Sylvia Frühwirth-Schnatter [2],

and Thomas Otter [3]

May 2002

Abstract

In this paper we carry out fully Bayesian analysis of the general heterogeneity model, which is a mixture of random effects model, and its special cases, the random coefficient model and the latent class model. Our application comes from Conjoint analysis and we are especially interested in what is gained by the general heterogeneity model in comparison to the other two when modeling consumers' heterogeneous preferences.

Keywords: Heterogeneity Model; Finite Mixture of Random Effects Model; Gibbs Sampling; MCMC Methods; Conjoint Analysis

1 The Substantive Problem and the Data

Our application comes from Conjoint analysis, a procedure that is focused on obtaining the importance of certain product attributes and their significance in motivating a consumer torward purchase from a holistic appraisal of attribute combinations. Our data come from a brand - price trade off study in the mineral-water category. Each of 213 Austrian consumers evaluated their likelihood of purchasing 15 different product-profiles offering five different brands of mineral-water at different prices on 20 point rating scales. The goal of the modeling exercise is to find a model describing consumers' heterogeneous preferences towards the different brands of mineral water and their brand-price trade offs.

Applying the general heterogeneity model to these data we follow up previous work on the same data using the random coefficient model (Frühwirth-Schnatter *et al.*, 1999) - this work was presented at the 14th IWSM conference - and the latent class model (Otter *et al.*, 2001).

2 Statistical Modeling Tools

2.1 The Heterogeneity Model

The data are described by a *mixture of random effects model*:

$$y_i = X_i^1 \alpha + X_i^2 \beta_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2 I), \tag{1}$$

where y_i is a vector of T_i observations for subject $i = 1, \ldots, N$, X_i^1 is the $T_i \times d$ design matrix for the $d \times 1$ vector of the fixed effects α and X_i^2 is the design matrix of dimension $T_i \times r$ for the $r \times 1$ random effects vector β_i. I is the identity matrix. Due to unobserved heterogeneity the random effects β_i are different for each subject i. The unknown distribution $\pi(\beta_i)$ of heterogeneity is approximated by a mixture distribution $\beta_i \sim \sum_{k=1}^{K} \eta_k N(\beta_k^G, Q_k^G)$ with the unknown group means $\beta_1^G, \ldots, \beta_K^G$, the unknown group covariance matrices $Q_1^G, \ldots, Q_K^G$ and the unknown group probabilities $\eta = (\eta_1, \ldots, \eta_K)$.

This model includes as special case the *aggregate model*, for $K = 1$, $Q^G \equiv 0$, the *latent class model* (LCM), for $K > 1$, $Q_1^G \equiv \ldots \equiv Q_K^G \equiv 0$ and the *random coefficient model* (RCM), for $K = 1$, $Q^G \neq 0$.

[1]Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: `regina.tuechler@wu-wien.ac.at`

[2]Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: `sfruehwi@isis.wu-wien.ac.at`

[3]Department of Advertising and Marketing Research, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: `thomas.otter@wu-wien.ac.at`

Verbeke and Lesaffre (1996) study this model with the groups covariances being the same for all groups. Bayesian estimation of heterogeneity models is discussed in Allenby *et al.* (1998) for models without fixed effects and in Lenk and DeSarbo (2000) for observations from distributions of general exponential families.

2.2 Bayesian Estimation of the Heterogeneity Model

MCMC Sampling Steps. Estimation of the heterogeneity model follows the ideas described in (Frühwirth-Schnatter *et al.*, 2002). It is carried out for a fixed number K of groups using Markov Chain Monte Carlo methods. Let $y^N = (y_1, \ldots, y_N)$ denote all observations. We introduce discrete latent group indicators $S^N = (S_1, \ldots, S_N)$, with S_i taking values in $\{1, \ldots, K\}$ and thereby indicating which group subject i belongs to, with the unknown probability distribution $Pr(S_i = k) = \eta_k$. Following the principle of data augmentation we augment the parameter vector of the unknown model parameters $\phi = (\alpha, \beta_1^G, \ldots, \beta_K^G, \eta, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2)$ by the individual parameters $\beta^N = (\beta_1, \ldots, \beta_N)$ and the group indicators S^N. We apply standard Gibbs sampling to sample η from $\pi(\eta|S^N)$ and $Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2$ from $\pi(Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2|\beta^N, S^N, \alpha, \beta_1^G, \ldots, \beta_K^G, y^N)$. Deviating from standard full conditional Gibbs sampling the marginal heteroscedastic model, where the individual parameters β^N are integrated out, serves to obtain S^N from $\pi(S^N|\phi, y^N)$. Finally we use the model's representation as a switching random effect model and apply a blocked Gibbs sampler to derive $\alpha, \beta_1^G, \ldots, \beta_K^G$ and β^N from $\pi(\beta_1^G, \ldots, \beta_K^G, \alpha, \beta^N|S^N, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2, y^N)$. Therefore we derive $\alpha, \beta_1^G, \ldots, \beta_K^G$ from $\pi(\alpha, \beta_1^G, \ldots, \beta_K^G|S^N, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2, y^N)$ in one step and β^N from $\pi(\beta^N|S^N, \beta_1^G, \ldots, \beta_K^G, \alpha, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2, y^N)$ in another step.

Label Switching. As our model includes a discrete latent structure, we have to identify a unique labeling subspace to avoid biased estimates of the group specific parameters $\beta_1^G, .., \beta_K^G, Q_1^G, .., Q_K^G, \eta_1, .., \eta_K$ and S^N. To achieve a unique labeling we apply the method of *Permutation sampling* described in (Frühwirth-Schnatter, 2001a). The sampler is restricted to a unique labeling subspace by introducing a constraint $R_g : g(\beta_1^G, Q_1^G, \eta_1) < \ldots < g(\beta_K^G, Q_K^G, \eta_K)$, where g is an appropriate function of the group specifics. For a lot of estimation problems arising in the empirical analysis of the heterogeneity models it is not necessary to identify a unique labeling. Such problems are for example the estimation of the individual parameters $\beta_i \sim \sum_{k=1}^K \eta_k N(\beta_k^G, Q_k^G) = \sum_{k=1}^K \eta_{\rho(k)} N(\beta_{\rho(k)}^G, Q_{\rho(k)}^G)$, (with some permutation ρ of the labels $1, \ldots, K$), and of the moments of the distribution of heterogeneity. Finally, it is possible to predict the behaviour of each subject under designs $X_i^{1,\star}, X_i^{2,\star}$ different from the ones used for estimation.

2.3 Model Comparison

We compare our models by their model likelihoods, which are computed from the MCMC outputs. For computing the model likelihoods we apply the method of bridge sampling, which has proved to be robust against label switching and more efficient than other methods (Frühwirth-Schnatter, 2001b).

3 Main Results of our Application from Conjoint analysis

Our fully parameterized design matrix consists of 15 columns corresponding to the constant, four brand contrasts (of the brands Römerquelle - *RQ*, Vöslauer - *VOE*, Juvina - *JU*, Waldquelle - *WA*), a linear and a quadratic price effect, four brand by linear price and four brand by quadratic price interaction effects, respectively. We used dummy-coding for the brands. The fifth brand Kronsteiner (*KR*) was chosen as the baseline. We subtracted the smallest price from the linear price column, and computed the quadratic price contrast from the centered linear contrast. Therefore, the constant corresponds to the purchase likelihood of Kronsteiner at the lowest price level, if quadratic price effects are not present. Earlier investigations of these data indicated that a specification with fixed brand by quadratic price interactions is preferable (Otter *et al.*, 2001) and is therefore chosen in this paper.

We carried out 30000 MCMC iterations and based our inference on the last 6000. The group specific means β_k^G and the fixed effects α are a priori normally distributed with $N(b_0, B_0)$ and $N(a_0, A_0)$, respectively. The prior means b_0 and a_0 are equal to the population mean of the RCM model reported in Frühwirth-Schnatter *et al.*, (1999) and for the information matrices we choose $A_0^{-1} = B_0^{-1} = 0.04 \cdot I$. The prior distribution of the groups covariances is an inverted Wishart distribution $IW(\nu_0^Q, S_0^Q)$. We choose $\nu_0^Q = 10$ and then derive S_0^Q from $E(Q_k^G) = (\nu_0^Q - (d+1)/2)^{-1} S_0^Q$, where $E(Q_k^G)$ was computed by

individual OLS estimation and d is the dimensionality of Q_k^G. The prior on η is the commonly used Dirichlet distribution $D(1,\dots,1)$. We stay noninformative about the error variances σ_ε^2 and choose the inverted Gamma distribution $IG(0,0)$.

3.1 Model Selection

We estimated various models for our data, the general heterogeneity model, varying the number of groups K, the special case of the LCM, also varying the number of groups K and the special case of the RCM.

Table 1 shows estimates of the logarithm of model likelihoods for all these models. We see that the RCM (column $Q \neq 0$, line $K = 1$) is clearly prefered to all LCMs (column $Q = 0$). Under the assumption of fixed brand by quadratic price interactions the optimal *latent* class model has seventeen classes. The optimal model out of *all* models under consideration is a general heterogeneity model with $K = 3$.

	$\log L(y^N|\text{Model})$			MSE	
K	$Q_k^G \neq 0$	$Q_k^G = 0$		$Q_k^G \neq 0$	$Q_k^G = 0$
1	-9222.36 (0.05)	-10077.31 (0.00)		19.69	38.98
2	-9165.66 (0.06)	-9881.49 (0.01)		19.89	34.55
3	-9161.27 (0.06)	-9733.98 (0.02)		19.80	31.11
4	-9165.73 (0.08)	-9669.98 (0.05)		19.88	29.32
5	-	-9596.61 (0.04)		-	27.75
$\vdots$		$\vdots$			$\vdots$
17	-	-9460.61 (1.19)		-	23.78
18	-	-9465.79 (1.33)		-	23.21

Table 1: Left-hand side: estimates of the $log\,L(y^N|Model)$, (rel. std. errors in parenthesis); right-hand side: mean squared errors for the holdout data
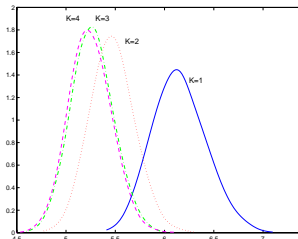

Figure 1: Marginal densities of σ_ε^2 for $K = 1,\dots,4$ classes, general model

Within the general heterogeneity specification this choice is also supported by the exploratory inspection of the model error variances σ_ε^2 plotted in Figure 1. We see that the model error variance decreases only up to $K = 3$.

3.2 Capturing Heterogeneity

We give bivariate plots of expected individual coefficients for the linear *price* against the *RQ* parameter in Figure 2 to illustrate to which extend the general heterogeneity model with three classes, the RCM and the LCM with seventeen classes capture heterogeneity. For comparison we add estimates for the aggregate model i.e. a model with no heterogeneity over the consumers and for individual OLS estimation, which proceeds with zero degrees of freedom for our data. From the triangular form of the three heterogeneity models' plots we conclude that the more price sensitive consumers tend to behave more homogeneously towards the *RQ* brand. This fact is also captured by the RCM despite its assumption of normally distributed random effects. In contrast to the RCM the general model reflects three clusters - one with high price sensitivity and low preference for *RQ*, one with low price sensitivity and low preference for *RQ* and the last one with low price sensitivity and high preference for *RQ*. In line with expectations and previous research the LCM captures least of the preference variation between consumers.
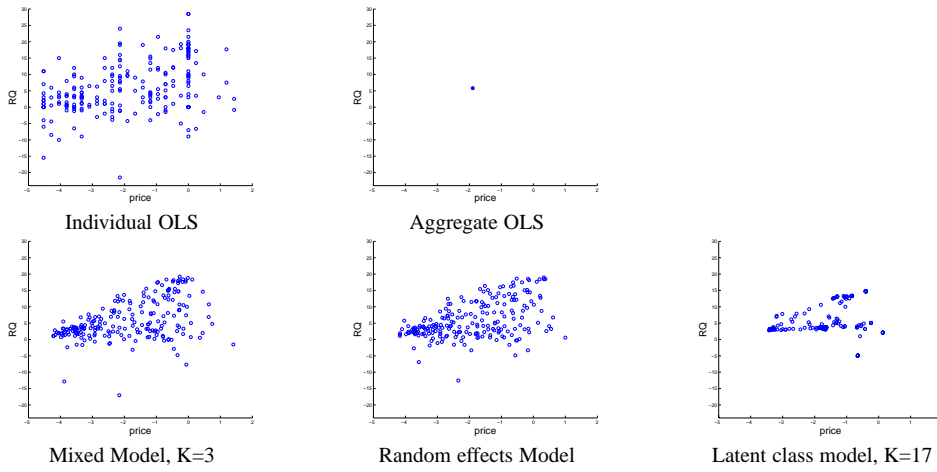
3

Figure 2: Scatterplots of individual parameters β_i, *price* against *RQ*

3.3 Model Identification

We are now going to illustrate how to achieve a unique labeling for the general heterogeneity model with $K = 3$ classes. First we analyze the output of the *Random Permutation Sampler* (Frühwirth-Schnatter, 2001a) graphically. The *Random Permutation Sampler* explores the unconstrained posterior distribution sampling from each labeling subspace with equal probability $1/K$. This can be seen in the left plot of Figure 3, where the group specific mean of the *price* is plotted against the one of *RQ*. Though there is no association between individual MCMC chains and group specific parameters by definition of the *Random Permutation Sampler* - estimates from any chain integrate over between group differences - three clusters and possible constraints to separate these may be found by visual inspection. In the middle of Figure 3 we see the output of the model that has been identified by separating the first group from the remaining two by the constraint $price_1 < price_{2,3}$ and by dividing the second group from the third one by $RQ_2 < RQ_3$. In Table 2 we give resulting estimates for the group specific means and the group weights. We have two big groups of nearly equal size, one collecting very price sensitive consumers whereas the consumers of the other group tend to value the "high-image" brands *RQ* and *VOE*. Moreover, they are less price sensitive. The smallest group consists of consumers, who are neither price sensitive nor brand conscious. The right plot of Figure 3 is a plot for the general model with $K = 4$ that again supports our choice of $K = 3$. We find the same three clusters as before but the data do not support a fourth cluster. The widely spread simulations overlaying the three clusters indicate that parameters are sampled from their prior because a fourth class is empty on many iterations.
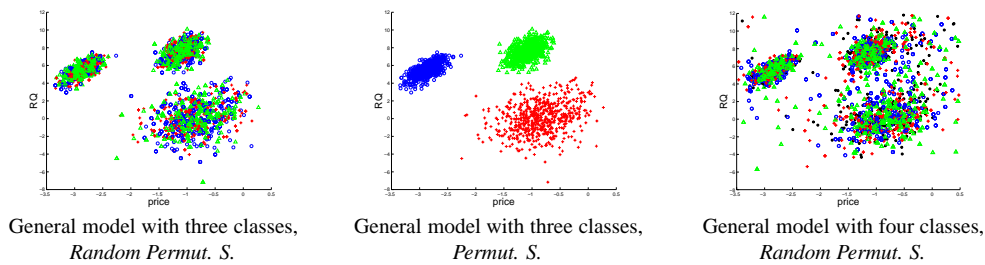

Figure 3: Posterior group specific means for the *price* against *RQ*

3.4 Prediction

In addition to the 15 products our estimation was based on, the 213 consumers evaluated 5 further products. These evaluations serve as holdout data to compare our models through their capability to predict. We give the traditional measure mean squared error (MSE) in Table 1 and illustrate the differences between predictive densities of selected models for one selected consumer in Figure 4. We include the estimates of individual OLS estimation. The MSE as well as the predictive density plots clearly advocate for a model

4

	$E(\beta_k^G|y^N)$				$E(\beta_k^G|y^N)$		
Effect	k=1	k=2	k=3	Effect	k=1	k=2	k=3
const	14.99	12.16	13.38	RQ·p	–0.78	–0.29	0.11
	(0.79)	(0.79)	(1.49)		(0.20)	(0.22)	(0.46)
RQ	5.45	7.57	0.17	VOE·p	–0.89	–0.29	0.49
	(0.77)	(0.85)	(1.72)		(0.20)	(0.24)	(0.54)
VOE	5.23	6.91	–0.46	JU·p	–0.54	0.16	–0.18
	(0.78)	(0.94)	(2.08)		(0.21)	(0.21)	(0.51)
JU	1.83	0.02	1.76	WA·p	–0.67	–0.08	–0.38
	(0.81)	(0.94)	(1.64)		(0.18)	(0.18)	(0.38)
WA	2.35	1.09	2.66				
	(0.81)	(0.92)	(1.72)				
p	–2.87	–1.09	–0.85		$E(\eta_k|y^N)$		
	(0.18)	(0.18)	(0.42)				
p^2	0.01	–0.08	–0.15		0.46	0.44	0.10
	(0.09)	(0.07)	(0.18)		(0.05)	(0.05)	(0.03)

Table 2: Posterior estimates of the group specific means β_k^G and the group specific weights η_k for the general model with K=3, (std.dev. in paranthesis)
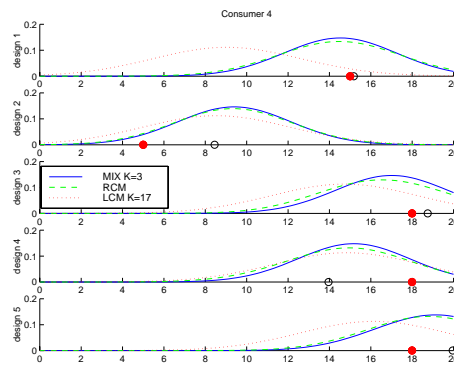

Figure 4: Predicitive densities for various models, (the full point indicates the true value, the circle indicates the OLS estimator)

with random effects being included. As it was to be expected from Section 3.2 the general models with $K > 1$ do not differ much from the RCM with $K = 1$ and the LCMs are clearly outperformed.

References

Allenby, G. M., Arora, N., Ginter, J. L. (1998): On the Heterogeneity of Demand, *Journal of Marketing Research*, **35**, 384-89.

Frühwirth-Schnatter, S. and Otter, Th. (1999): Conjoint-Analysis Using Mixed Effect Models, In: Friedl, H., Berghold, A. and Kauermann, G. (Eds.), *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, Graz 1999, pp.181–191.

Frühwirth-Schnatter, S. (2001a): Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, *Journal of the American Statistical Association,* **96**, 194-209.

Frühwirth-Schnatter, S. (2001b): Model Likelihoods and Bayes Factors for Switching and Mixture Models. Preprint **71**, Department of Statistics, Vienna University of Economics and Business Administration, revised version submitted for publication.

Frühwirth-Schnatter, S., Tüchler, R. and Otter, Th. (2002): Bayesian Analysis of the Heterogeneity

Model. Preprint **82**, Department of Statistics, Vienna University of Economics and Business Administration, revised version submitted for publication.

Lenk, P. J. and DeSarbo, W. S. (2000): Bayesian Inference for Finite Mixture of Generalized Linear Models with Random Effects. *Psychometrika*, **65**, 93-119.

Otter, Th., Tüchler, R. and Frühwirth-Schnatter, S. (2001): Bayesian Latent Class Metric Conjoint Analysis - A Case Study from the Austrian Mineral Water Market. *Proceedings of the 25th annual Conference of the German Classification Society (GfKl)*, Munich, 14-16th March 2001.

Verbeke, G. and Lesaffre, E. (1996): A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association,* **91**, 217-221.