Online Appendix to Accompany "*Bayesian Analysis of Hierarchical Effects*"

This appendix is divided into two main sections. In the first section we provide the joint distribution of the data and parameters and the estimation procedure of our proposed model. The second section provides an extensive simulation study that mimics our data and provides validation for our modeling approach.

1. Estimation Algorithm

The empirical application discussed in this article is a cross-sectional dataset with i = 1, …, 6177 consumers. There are R = 1, ..., 6 intended consumer action variables (x), M = 1, ..., 7 brand belief variables (b) and P = 1, …, 13 media exposure variables (z). Let K denote the number of latent segments present in the data. Let $\bullet_k$ denote the weights associated with each of the K latent segments and s_i denote the underlying latent indicator variable. The algorithm described below is for the model: [y|X][X|B][B|Z] and can be used for any of the models described in the paper by replacing one set of variables with another. The following steps describe the Gibbs sampler used to obtain the estimates of the model parameters, for a more detailed discussion of the different components refer to, Richardson and Green (1997), George and McCulloch (1993) and Brown et al. (1998).

The joint density for the data and all the parameters in the model is given by:

$$\left\{\prod_{i=1}^{n}\left[y_i \mid x_i,\beta_k,\sigma_k^2,\phi_k\right]\left\{\prod_{k=1}^{K}\left[x_{ik} \mid b_{ik},\Gamma_k,\Sigma_k\right]\left[b_{ik} \mid z_{ik},\Delta_k,\Omega_k\right]\left[z_{ik} \mid \mu_k,\Psi_k\right]\right\}\right\}$$

$$\left\{\prod_{k=1}^{K}\left[\beta_k \mid \theta,\tau,\sigma_k^2,s_i=k\right]\left[\sigma_k^2 \mid \nu,\psi_\beta,s_i=k\right]\left[\Gamma_k \mid \Sigma_k,\gamma_\Gamma\right]\left[\Delta_k \mid \Omega_k,\gamma_\Delta\right]\left[\Sigma_k \mid \lambda_\Sigma,\Psi_\Sigma\right]\right\}$$

$$\left\{\prod_{k=1}^{K}\left[\Omega_k \mid \lambda_\Omega,\Psi_\Omega\right]\left[\phi_k \mid s_i=k\right]\left[s_i=k \mid n_k,\rho_k\right]\right\}\left\{[\theta]\,[\tau]\,[\upsilon]\,[\Psi_\beta]\,[\gamma_\Gamma][\gamma_\Delta]\,[\lambda_\Sigma][\Psi_\Sigma][\lambda_\Omega][\Psi_\Omega][\rho]\right\}$$

1) Generate •

The segment weights • have a Dirichlet distribution with prior parameters ρ_1, …, ρ_K :

$$\left[\phi_k \mid s_i,k\right] \sim D\left(\rho_1+n_1,\rho_2+n_2,...,\rho_K+n_K\right)$$

where, $n_k = \#\{i : s_i = k\}$ is the number of subjects assigned to group k. In our analysis we set $\rho_1 = \rho_1 = \ldots = \rho_K = 2$, given the large sample size for the data used, this prior specification results in a very mild influence on the posterior.

2) Generate β_k , σ_k^2

The segment specific parameters for each of the k = 1, …, K segments are estimated using standard regression, Rossi et al. 2005 (chapter 2).

The full conditionals for β_k , and σ_k^2 for each of the K segments, are given below.

i) $$\left[\beta_k \mid \sigma_k^2, y, X\right] \sim Normal\left(\frac{\sigma_k^{-2}\left[X_k' X_k\right]^{-1} X_k' y_k + \theta\tau}{\sigma_k^{-2} n_k + \theta}, \left[\sigma_k^{-2} X_k' X_k + \theta\right]^{-1}\right)$$

ii) $\left[\sigma_k^2 \mid \beta_k, y, X\right] \sim Inverted\ \ Gamma\left(\frac{\upsilon + n_k}{2}, \frac{|y - X'\beta_k|^2 + \upsilon\psi}{2}\right)$

3) Generate Γ_k

The segment specific parameters Γ_k for each of the k = 1, …, K segments are estimated using Bayesian multivariate variable selection procedure suggested by Brown et al. (1998), which is an extension of the univariate regression procedure of George and McCulloch (1993, 1997). With n_k observations x_i (R × 1) conditional on b_i (M × 1) we have a generalized multivariate regression model. The parameters to be estimated are Γ_k which is of dimensionality (R × M) and the variance-covariance matrix Σ_k. The prior distribution for Γ_k and Σ_k can be written as $[\Gamma_k \mid \Sigma_k][\Sigma_k]$.

Multivariate variable selection is accomplished using a latent binary R-vector λ. The j^{th} element of λ, λ_j is either a 0 or a 1. A 0 implies that the covariance matrix of the corresponding row of Γ_k is very small and 1 implies that it is large, pertaining to the exclusion or inclusion of predictor j respectively.

We use prior distributions as proposed in Brown et al. (1998):

$[\Sigma_k] \sim Inverted\ Wishart(\Omega_0, V)$

$[\Gamma_k \mid \Sigma_k] \sim N(D_\lambda C_\lambda D_\lambda, \Sigma_k)$

Here D_λ is diagonal matrix and C_λ is a correlation matrix, the j^{th} element of D_λ^2 is either m_{0j}, a small constant, when $\lambda_j = 0$ or m_{1j}, a large constant, when $\lambda_j = 1$. George and McCulloch (1997) provide further discussion on the different choices of m_{0j}, m_{1j} and C_λ. We set $m_{0j} = 0.001$ and $m_{1j} = 5$ for our analysis and C_λ as the identity matrix.

An efficient algorithm for sampling from a multivariate regression can be carried out using the procedure suggested in Rossi et al. 2005 (chapter 2).

$$[\Gamma_k \mid \lambda_j, \Sigma_k, X, B] \sim Normal(\tilde{\gamma}_k, \Sigma_k \otimes (B'B + \Lambda_k)^{-1})$$

i) $\tilde{\gamma}_k = vec(\tilde{\Gamma}_k)$, $\tilde{\Gamma}_k = (B'B + \Lambda_k)^{-1}(B'B\hat{\Gamma}_k + \Lambda_k \bar{\Gamma}_k)$,

$\hat{\Gamma}_k = (B'B)^{-1}B'X$, $\Lambda_k = D_\lambda C_\lambda D_\lambda$

ii) $[\Sigma^k \mid \Gamma_k, X, B, S] \sim Inverted\ Wishart(\Omega_0 + n_k, V_0 + S)$

$S = (X - B\ \tilde{\Gamma}_k)'(X - B\ \tilde{\Gamma}_k) + (\tilde{\Gamma}_k - \bar{\Gamma}_k)'\Lambda_k(\tilde{\Gamma}_k - \bar{\Gamma}_k)$

iii) $[\lambda_{kj} \mid \Gamma_k, \lambda_{-kj}]$, the prior on λ_k is multivariate Bernoulli with probability similar to that given in the univariate case and involves evaluation of multivariate normal densities.

4) Generate Δ_k

The segment specific parameters Δ_k for each of the k = 1, …, K segments are estimated using Bayesian multivariate variable selection procedure as described in step 3. With n_k observations, b_i (M × 1) conditional on z_i (P × 1) we have a generalized multivariate regression model. The parameters to be estimated are Δ_k which is of dimensionality (M × P) and the variance-covariance matrix Ω_k. The prior distribution for Δ_k and Ω_k can be written as $[\Delta_k \mid \Omega_k][\Omega_k]$. The parameters are estimated as shown in the previous step.

5) Generate s_i

The multinomial probabilities for the latent variable s_i, where $s_i = k$ if subject i belongs to segment k (for k = 1, …, K) is given by:

$$p(s_i = k \mid \phi_k, \sigma_k, \beta_k, y_i, x_i) \propto \frac{\phi_k}{\sigma_k} \exp\left(-\frac{(y_i - x_i \beta_k)^2}{\sigma_k^2} \right)$$

2. Simulation Study

The following simulation study is used to demonstrate the ability of our model to recover parameters and identify the segments correctly using cross-sectional data. In this study, we demonstrate the ability of our approach to identify the right model and deal with high dimensionality by incorporating Bayesian variable selection. We also find that a standard aggregate model that assumes one response segment and no hierarchy does not recover the parameters in the model. Similarly, model with just the mixture specification and no hierarchical component also fails to recover the true parameters. Thus, we demonstrate that models that do not incorporate the option to recognize both the segments and the hierarchy lead to incorrect conclusions. We also find that log marginal likelihood can aid in identifying the correct model.

Our simulation strategy involves the following steps. 1) Generate data based on a specific hierarchical model. 2) Identify the best fitting model (based on Chib 1995, marginal likelihood estimation) from all possible combinations of the conditional relationships used to identify response segments. 3) Pick the best fitting model and consider sub-models for the best fitting model and pick the best sub-model. 4) Compare the posterior estimates of the best fitting sub-model with the data generated in step 1. 5) Demonstrate that estimates from standard aggregate models can be misleading.

Step 1: A sample size of 5000 observations were generated according to the conditional relationship [y|X][X|B][B|Z][Z] with three latent segments (K=3). For each of the latent segments, values of Z were generated from a multivariate normal and used in equation (3) to generate values for B. The values for X were then generated from the values of B generated from the previous step using equation (2). Finally, equation (1) is used to generate the values for y.

Steps 2 and 3: Table 1 provides log marginal likelihood computed based on Chib 1995 (See Appendix C for further details). We find that model 1 fits the data the best. We then consider five sub-models for the best fitting model. A comparison of the model fit statistics across five sub-models is presented in table 2. We find that our simulation correctly identifies the best fitting model to be the one where the factorization is given by [y|X][X|B][B|Z][Z]. This demonstrates that the log marginal likelihood for model comparison can identify the correct model in our study.

Step 4: We investigate the parameter estimates for the best fitting model. Parameter estimates with their corresponding true values are presented in tables 3 and 4. Results include estimates of the posterior mean, posterior standard deviation (in parenthesis), and a 95% credible interval (in brackets). A closer investigation at table 6 indicates that parameters with effect sizes smaller than 0.05 were shut down to 0 due to our proposed variable selection procedure. We only incorporate variable selection at the second and third level of the hierarchy because the number of parameters (in the multivariate regressions) tends to explode with an increase in the number of segments. Thus, we find that our procedure correctly shuts down parameters with effect sizes less than 0.05 while leaving the other parameters unaffected. Since this simulation pertains to small dimension of data use of multivariate variable selection is to primarily demonstrate that our procedure works. In the case of the real automobile data used in the study the dimensionality of the parameters is large. If we investigate the impact of B on X the number of parameters is equal to 6 X 8 = 48 for each segment i.e. 192 (48 X 4) and Z on B the number of parameters is equal to 14 X 7 = 98 for each segment i.e. 392 parameters (98 X 4). Thus, the total number of parameters to be estimated equals 584 (192 + 392), which provides justification for the use of multivariate variable selection.

Estimates based on the hierarchical model correctly identify the existence of three segments, and successfully recovers the true parameters. Figure 1 illustrates the trace plot and ability of the Bayesian estimation procedure to successfully identify the number of latent mixing distributions. Plotted are the last 200 draws of the Markov chain. Label-switching was handled by estimating the parameters without any restrictions and then post-processing the draws from the output[1]. We used several different mixture weight combinations to generate our data. The results presented are for segment weights of 0.33, 0.33 and 0.34. We also employ an approach where we over-specify the number of latent segments in the data, and let the estimation routine shut down redundant segments by estimating small segment weights (i.e., $\bullet_k = 0$) where appropriate. These plots can be obtained from the authors and have been omitted to avoid clutter.

Step 5: We compare our modeling approach against an: i) Aggregate regression model that does not incorporate the mixture or the hierarchical specification. ii) Aggregate mixture model that does not incorporate the hierarchy in the data but does include the mixture specification.

Table 5 provides the parameter estimates for the aggregate regression model ([y|X,B,Z]). This model does not incorporate the mixture or the hierarchical specification used in the data generating process. Thus, we find that an aggregate regression model does not recover the parameters. Table 6 provides the parameter estimates for an aggregate mixture model without the hierarchical specification and is given by:

$$y_i \mid x_i, b_i, z_i, \{\phi_k, \beta_k, \sigma_k^2\} \sim \sum_{k=1}^{K} \phi_k N\left(w_i' \beta_k, \sigma_k^2\right) \; where \; w_i = \{x_i, b_i, z_i\}$$

Again, we find that the parameters used to generate the data are not recovered.

Table 1
(Model Fit for all different models considered)
(Simulated Data)

Hierarchical Models		Log Marginal Likelihood
1	[y|X][X][B][Z]	-20064
2	[y|B][X][B][Z]	-21002.8
3	[y|Z][X][B][Z]	-23215.9
4	[y|X,B][X][B][Z]	-21098
5	[y|X,Z][X][B][Z]	-21697
6	[y|B,Z][X][B][Z]	-21495.1
7	[y|X,B,Z][X][B][Z]	-22852.6

[1] We thank an anonymous reviewer for this suggestion.

Table 2
(Model Fit for Sub-Models)
(Simulated Data)

Sub Models		Log Marginal Likelihood
1	[y|X][X|B][B][Z]	-19607.4
2	[y|X][X|B,Z][B][Z]	-19444.2
3	[y|X][X|B][B|Z][Z]	**-19359.6**
4	[y|X,B][X|B][B|Z][Z]	-21355.2
5	[y|X,B,Z][X|B][B|Z][Z]	-19423.1

Table 3
Coefficient Estimates for Hierarchical Mixture Model
(Simulated Data)

Variable	Hierarchical Mixture Model (K = 3) $y_i \mid x_i, \left\{ \phi_k, \beta_k, \sigma_k^2 \right\} \sim \sum_{k=1}^{K} \phi_k N\left(x_i' \beta_k, \sigma_k^2 \right)$		
	Segment 1	Segment 2	Segment 3
X0	-1.01 (0.03) [-1.06, -0.95] True Val: **-1.0**	-1.99 (0.21) [-2.36, -1.70] True Val: **-2.0**	-3.15 (0.48) [-3.90, -2.79] True Val: **-3.0**
X1	0.91 (0.08) [0.79, 1.03] True Val: **1.0**	2.01 (0.40) [1.23, 2.49] True Val: **2.0**	2.87 (0.61) [2.49, 3.34] True Val: **3.0**
X2	0.57 (0.05) [0.48, 0.64] True Val: **0.5**	1.01 (0.20) [0.63, 1.41] True Val: **1.0**	1.62 (0.47) [1.02, 2.13] True Val: **1.5**
Sigma	0.99 (0.07) [0.91, 1.12] True Val: **1.0**	1.04 (0.10) [0.97, 1.13] True Val: **1.0**	1.95 (0.11) [1.81, 2.03] True Val: **2.0**
Weights	0.33 (0.01) [0.31, 0.34] True Val: **0.33**	0.34 (0.02) [0.33, 0.35] True Val: **0.34**	0.33 (0.02) [0.32, 0.35] True Val: **0.33**

Parameter estimates are the posterior means, posterior standard deviation (in parenthesis) and 95% credible interval (in brackets).

Table 4
Coefficient Estimates for Hierarchical Mixture Model
(Simulated Data)

$$x_i \,|\, b_i, \Gamma_k, \Sigma_k, (s_i = k) \sim N(\Gamma_k b_i, \Sigma_k)$$

	Segment 1			Segment 2		
	B_0	B_1	B_2	B_0	B_1	B_2
X_1	0.49 (0.01) [0.47, 0.51] TrueVal:**0.5**	0.49 (0.01) [0.47,0.50] True Val: **0.5**	1.49 (0.01) [1.47, 1.51] True Val: **1.5**	0.99 (0.02) [0.96, 1.01] TrueVal:**1.0**	0.98 (0.02) [0.94, 1.02] True Val: **1.0**	2.98 (0.04) [2.89, 3.01] True Val: **3.0**
X_2	1.49 (0.01) [1.47, 1.51] True Val: **1.5**	-1.01 (0.01) [-1.03, -0.99] True Val: **-1.0**	-0.51 (0.01) [-0.53, -0.49] True Val: -**0.5**	2.99 (0.02) [2.96, 3.01] True Val: **3.0**	-2.02 (0.02) [-2.06, -1.99] True Val: **-2.0**	-1.03 (0.04) [-1.11, -0.99] True Val: **-1.0**

	Segment 3		
	B_0	B_1	B_2
X_1	1.49 (0.07) [1.35, 1.58] True Val: **1.5**	1.48 (0.10) [1.29, 1.62] True Val: **1.5**	4.43 (0.15) [4.12, 4.61] True Val: **4.5**
X_2	4.49 (0.07) [4.35, 4.58] True Val: **4.5**	-3.02 (0.10) [-3.20, -2.88] TrueVal: **-3.0**	-1.57 (0.15) [-1.88, -1.39] True Val: -**1.5**

$$b_i \,|\, z_i, \Delta_k, \Omega_k, (s_i = k) \sim N(\Delta_k z_i, \Omega_k)$$

	Segment 1			Segment 2		
	Z_0	Z_1	Z_2	Z_0	Z_1	Z_2
B_1	0.00 (0.02) [-0.01, 0.01] True Val: **0.02**	0.99 (0.02) [0.78, 0.89] True Val: **1.0**	1.50 (0.02) [0.67, 0.89] True Val: **1.5**	-0.01 (0.02) [-0.01, 0.01] True Val: **0.03**	2.01 (0.01) [1.99, 2.03] True Val: **2.0**	3.00 (0.02) [2.99, 3.02] True Val: **3.0**
B_2	0.00 (0.01) [-0.01, 0.01] True Val: **0.04**	-1.01 (0.05) [-1.12, -0.93] True Val: **-1.0**	-1.50 (0.02) [-1.57, -1.42] True Val: **-1.5**	0.00 (0.01) [0.00, 0.01] True Val: **0.05**	-1.99 (0.01) [-2.01, -1.98] True Val: **-2.0**	-3.01 (0.01) [-3.02, -3.0] True Val: **-3.0**

	Segment 3		
	Z_0	Z_1	Z_2
B_1	0.00 (0.01) [-0.01, 0.01] True Val: **0.02**	2.84 (0.12) [2.84, 3.1] True Val: **3.0**	4.43 (0.06) [4.37, 4.61] True Val: **4.5**
B_2	0.00 (0.01) [0.00, 0.00] True Val: **0.04**	-2.81 (0.20) [-3.31, -2.79] True Val: **-3.0**	-4.46 (0.08) [-4.62, -4.34] True Val: -**4.5**

Parameter estimates are the posterior means, posterior standard deviation (in parenthesis) and 95% credible interval (in brackets).

Figure 1
Trace Plot of Segment Weights for a Three-Component Mixture Model with K = 3
(Simulated Data)

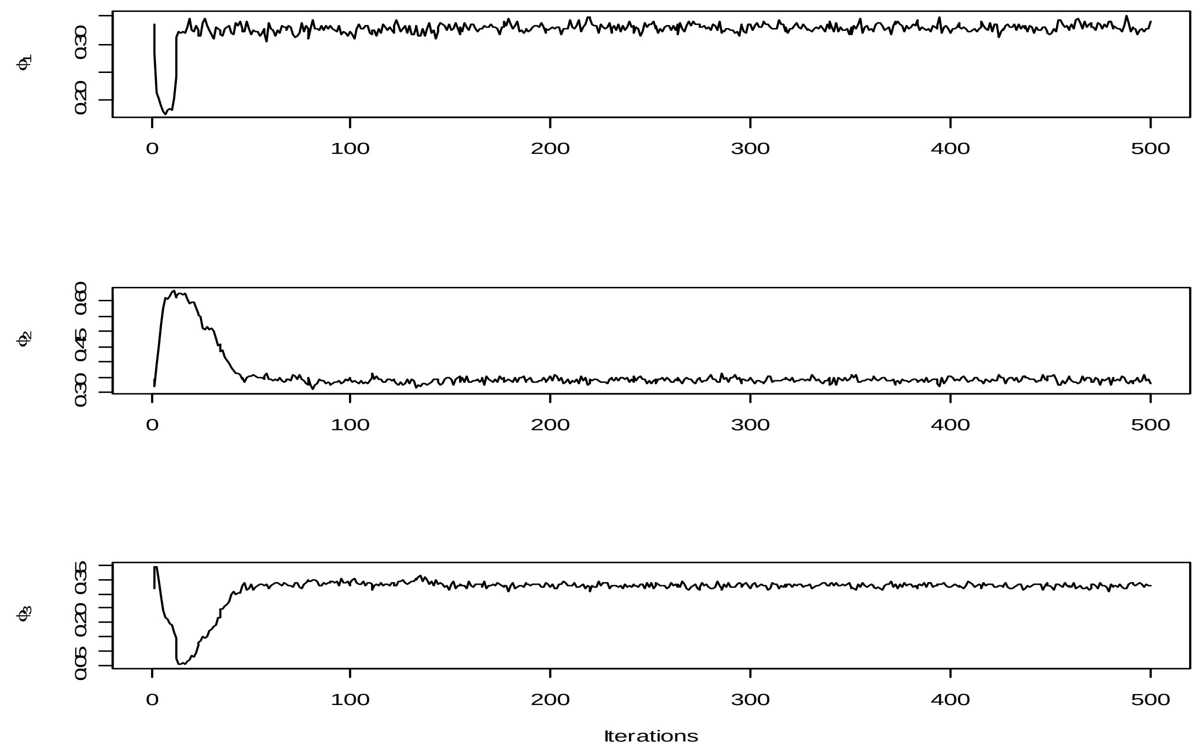


Table 5
Coefficient Estimates for Homogenous Regression Model (No Mixture and No Hierarchy)

Variable	Aggregate Regression Model (y | X, B, Z)
X0	0.06 (0.71) [0.01, 1.04]
X1	0.89 (0.21) [0.53, 1.1]
X2	-7.66 (0.30) [-8.27, -7.49]
B1	-12.01 (0.71) [-12.85, -11.15]
B2	14.53 (0.27) [14.03, 14.97]
Z1	8.35 (0.65) [7.85, 8.87]
Z2	4.26 (0.41) [4.01, 4.92]

Parameter estimates are the posterior means, posterior standard deviation (in parenthesis) and 95% credible interval (in brackets).

Table 6
Coefficient Estimates for Pooled Mixture Model (No Hierarchy)

Variable	Mixture Model (K = 3) $y_i \mid x_i, b_i, z_i, \{\phi_k, \beta_k, \sigma_k^2\} \sim \sum_{k=1}^{K} \phi_k N(w_i^{'}\beta_k, \sigma_k^2)$ *where* $w_i = \{x_i, b_i, z_i\}$		
	Segment 1	Segment 2	Segment 3
X0	0.49 (0.61) [-0.6, 0.55] True Val: **-1.0**	0.98 (0.51) [0.21, 1.42] True Val: **-2.0**	2.5 (1.06) [1.03, 3.84] True Val: **-3.0**
X1	2.84 (0.07) [2.62, 2.93] True Val: **1.0**	-1.89 (0.34) [-2.21, -1.35] True Val: **2.0**	0.71 (0.73) [0.11, 1.65] True Val: **3.0**
X2	3.19 (0.17) [2.81, 3.53] True Val: **0.5**	2.35 (0.03) [2.31, 2.47] True Val: **1.0**	3.98 (0.31) [3.44, 4.38] True Val: **1.5**
B1	1.83 (0.16) [1.37, 2.08]	3.71 (0.13) [3.24, 3.98]	-3.03 (0.21) [-3.22, -2.79]
B2	0.04 (0.01) [-0.02, 0.08]	-1.26 (0.25) [-1.96, -1.06]	-1.81 (0.23) [-2.26, -1.36]
Z1	-6.12 (0.78) [-7.02, -5.92]	6.06 (0.84) [-6.85, -5.98]	-3.76 (0.29) [-4.04, -3.32]
Z2	3.21 (0.42) [2.81, 3.96]	2.85 (0.09) [2.77, 2.99]	0.62 (0.12) [0.41, 0.73]

Parameter estimates are the posterior means, posterior standard deviation (in parenthesis) and 95% credible interval (in brackets).

Online Appendix B

In this part of the appendix we analyze our data using a pooled or homogenous model that assumes one response segment ($K = 1$) and then compare estimates to a model with multiple segments ($K > 1$). Table 1 reports coefficient estimates for the one segment model in which intended actions (X) and media exposure (Z) are directly related to likelihood of purchase (y). The left side of table 1 corresponds to the model of purchase intention regressed on intended actions alone (i.e., $y \mid X$), and the right side of the table corresponds to the model that includes the media variables (i.e., $y \mid X, Z$). Reported are means and standard deviations of the posterior distribution. Entries in bold correspond to variables with at least 95% of their posterior mass centered away from zero. The results indicate a fairly strong relationship between intended actions and purchase likelihood, but that the media variables add little to improve model fit. The media coefficients are estimated to be small in magnitude, with all estimates being near zero.

[Table 1]

We continue the naïve analysis of our data in two parts. First, we attempt to identify response segments using the marginal distribution of purchase intentions (y). Table 2 provides coefficients estimates of this model where response segments are formed by partitioning the dependent variable (segment 1: $y = 0 – 2$; segment 2: $y = 3 – 5$; segment 3: $y = 6 – 7$; segment 4: $y = 8 – 10$). We recognize that partitioning based on the dependent variable is not the right approach, but is done to demonstrate its shortcomings and contrast it with what we believe is the right modeling approach. The coefficients displayed in table 2 are estimated to be small in magnitude, with almost all estimates being near zero.

[Table 2]

The second part of our naïve analysis attempts to investigate hierarchical effects in an aggregate analysis of the X, B and Z variables. Table 3 displays coefficient estimates for a Bayesian multivariate regression of X given B and B given Z without accounting for heterogeneous segments. The results indicate that brand beliefs have a significant impact on consumer actions. Media coefficients, however, are estimated to be small in magnitude, with almost all estimates being near zero. Thus, the impact of media on brand beliefs is estimated to be insignificant. The results of our naïve analysis of the data is that media effects are estimated to be insignificant in any kind of aggregate analysis, and that brand beliefs and consumer actions do not have large effects on purchase intent. Moreover, the naïve analysis does not provide measures of model fit needed to assess alternative hypothesized model structures.

[Table 3]

Table 1
Homogenous Regression Coefficient Estimates
Automobile Data

Variables (Dependent variable y: Likelihood of purchase)	Estimates (y | X)	Estimates (y | X, Z)
Intermediate Variables (x)	Posterior Mean (std. dev.)	Posterior Mean (std. dev.)
Go to a retailer	**0.29** (0.02)	**0.27** (0.02)
Seek info directly	**0.16** (0.02)	**0.16** (0.02)
Seek info from objective Source	0.03 (0.02)	0.02 (0.02)
Recommend to a Friend	**0.20** (0.01)	**0.19** (0.01)
Read mail	**0.14** (0.01)	**0.14** (0.01)
Take a test Drive	0.01 (0.02)	0.02 (0.02)
Media Exposure Variables (z)		
Magazine Advertisement		0.0 (0.01)
News Advertisement		0.01(0.00)
Radio Advertisement		0.0 (0.00)
Television Advertisement		0.01 (0.00)
Sponsor Advertisement		-0.01 (0.01)
Internet Advertisement		0.00 (0.01)
Direct Mailing		0.04 (0.03)
Brochure		-0.03 (0.02)
Focal Brand's Website		0.02 (0.02)
Dealership Website		-0.04 (0.03)
Independently published article		0.02 (0.02)
Independent Website		0.01 (0.02)
Display in Public Place		0.00 (0.01)
Intercept	0.63 (0.05)	0.57 (0.05)
Sigma	5.76 (0.11)	5.4 (0.10)
Model Fit (R^2)	0.54	0.54

Table 2
Naïve Segment Regression Estimates

	Partition y*	Segment 1	Segment 2	Segment 3	Segment 4
X	Intercept	0.07	**3.82**	**5.96**	**6.39**
	Go to a dealer	**0.05**	0.02	0.02	0.04
	Seek Info directly	0.02	0.02	0	0.03
	Seek Info from Obj. Source	0.02	0.02	-0.03	-0.03
	Recommend to a Friend	-0.01	0.01	0	0.04
	Read mail	0.03	-0.03	0.01	-0.02
	Take a test drive	0.02	0	0.01	0.01
B	Overall Impression of each make	0	0	0	0
	Durability	0	0	0	0
	Security	0	0	0	0
	Excitement	0	0	0	0
	Design	0	0	-0.02	0
	Innovation	0	0	-0.01	0
	Manufacturing Quality	0.02	-0.02	-0.02	0.04
Z	Magazine Ad	0	0.02	0.03	0
	Newspaper Ad	-0.03	0.12	0.04	0
	Radio Ad	0.01	0	-0.01	0.01
	TV Ad	0.01	0.01	0.01	0
	Sponsorship Event	0.01	0	0	0
	Internet Ad	0	0	0.01	0
	Direct Mail	0.02	0.02	**0.07**	**0.23**
	Brochure	0.02	-0.01	0	-0.07
	Company Website	0	0.01	0	0
	Dealer Website	-0.02	0.03	-0.01	0.03
	Independent Article	0	0	-0.02	0.01
	Independent Website	0	0	0.01	0.02
	Public Display	-0.01	0	-0.01	0.02

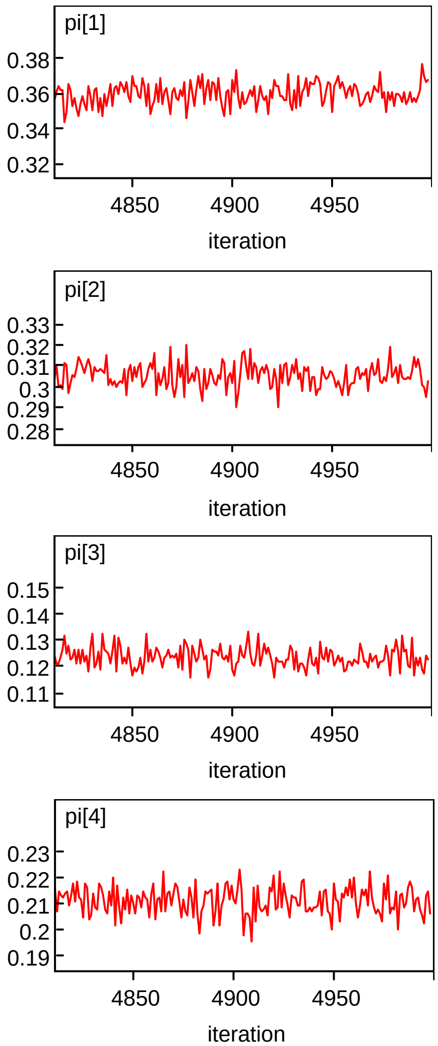
Note *: segment 1: y = 0 – 2; segment 2: y = 3 – 5; segment 3: y = 6 – 7; segment 4: y = 8 – 10

Table 3
Coefficient Estimates for Naïve Aggregate Model using Multivariate Regression

(X|B) - Aggregated	Go to a dealer	Seek Info directly	Seek Info from Obj. Source	Recommend to a Friend	Read mail	Take a test drive
Intercept	**-1.21**	**-1.22**	**-1.18**	**-2.71**	**-1.77**	**-1.11**
Overall Impression of each make	**0.25**	**0.29**	**0.28**	**0.44**	**0.45**	**0.23**
Durability	0.06	0.05	0.04	0.1	0.12	0.07
Security	**0.14**	**0.17**	**0.17**	**0.15**	**0.2**	0.09
Excitement	**0.16**	**0.13**	**0.11**	**0.12**	**0.1**	**0.15**
Design	**0.11**	**0.1**	0.08	**0.13**	0.06	**0.11**
Innovation	0.06	**0.11**	**0.14**	**0.14**	**0.12**	0.07
Manufacturing Quality	**-0.28**	**-0.3**	**-0.27**	**-0.22**	**-0.21**	**-0.23**

(B|Z) - Aggregated	Overall Impression of each make	Durability	Security	Excitement	Design	Innovation	Manufacturing Quality
Intercept	**0.76**	**0.75**	**0.73**	**0.62**	**0.66**	**0.67**	**0.76**
Magazine Ad	0.02	**0.03**	0.04	0.01	0.01	0.03	0.03
Newspaper Ad	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Radio Ad	0	0	0	0	0	0	0
TV Ad	0.01	0.01	0.01	0	0.01	0.01	0.01
Sponsorship Event	0	0.02	0.01	-0.01	-0.01	0.02	0.03
Internet Ad	0	0	-0.01	0	0	0	-0.01
Direct Mail	-0.01	0	0.02	0.01	0	-0.01	0
Brochure	0.03	0	-0.04	0.03	0.02	0	-0.02
Company Website	0.09	0.05	0.07	0.1	0.09	**0.12**	0
Dealer Website	**0.04**	0.03	0.03	0.04	0.04	**0.04**	0.03
Independent Article	0	0.01	0.04	0.01	0.01	0.03	0.01
Independent Website	0.02	0.01	0.01	0.02	0.01	0	0
Public Display	0.01	0.01	0	0.01	0.01	0.01	0.01

Figure 1
Segment weights for a finite mixture SEM with K = 4
(Automobile Data)



Appendix C
(Computation of Marginal Likelihood)

The following steps describe the computation of marginal likelihood. For a more detailed discussion please refer to the general case of Chib 1995 (section 2.1.3, pg. 1315).
This method requires the computation of log-likelihood, prior and an estimate of the posterior ordinate. Chib (1995) discusses how this computation does not lead to any instability problems that can result from log marginal density calculations using Newton Raftery approximations. This method uses density value that is averaged and no inverse calculations are involved as in log marginal density computations.
The Basic Marginal Likelihood Identity and the proposed estimate of the marginal density is given by (Chib equation 4 and 6):

$$m(y) = \frac{f(y \mid \theta)\pi(\theta)}{\pi(\theta \mid y)}$$

Where, 'y' are the data and '•' are the parameters of the model.

The proposed estimate of the marginal density is given by:

$$\ln m^*(y) = \ln f(y \mid \theta^*) + \ln \pi(\theta^*) - \ln \pi^*(\theta^* \mid y)$$

Thus, the estimation of the marginal likelihood reduces to evaluating the above equation for a given value of '•' (say •*). We choose "•*" to be the posterior mean. The likelihood function and the prior can be evaluated at the posterior mean for our proposed model very easily, since the likelihood is a mixture of normals and the priors are normals, inverted Wishart and Dirichlet distributions. The only laborious part is the computation of the posterior ordinate because we have more than three parameter blocks. Using Chib's method for three or more parameter blocks we can break it up as:

$$\pi(\theta_1^*, \theta_2^*, \theta_3^*, ..., \theta_m^* \mid y) = \pi(\theta_1^* \mid y)\pi(\theta_2^* \mid \theta_1^*, y)\pi(\theta_3^* \mid \theta_1^*, \theta_2^*, y)..\pi(\theta_m^* \mid \theta_1^*, \theta_2^*, ..., \theta_{m-1}^*, y)$$

For our proposed model:

$$\{\theta_1^*, \theta_2^*, \theta_3^*, ..., \theta_m^*\}\ correspond\ to\ \{\beta^*, \sigma^{2*}, \phi^*, \Gamma^*, \Sigma^*, \Delta^*, \Omega^*, \mu^*, \Psi^*, \tau^*\} \tag{1}$$

Therefore the marginal likelihood is given by:

$$\ln m^*(y) = \ln f(y \mid \theta^*) + \ln \pi(\theta^*) - \sum_{m=1}^{M} \ln \pi^*(\theta_m^* \mid y, \theta_r^*(r < m)) \tag{2}$$

In the above equation we can evaluate the likelihood and also the analytical form of the priors is known, but not the posterior. Chib proposes a simulation method to evaluate the posterior ordinate. To demonstrate the steps involved in the simulation, we choose the first three parameters (β^*, σ^{2*}, φ^*) from (1):

$$\pi\left(\beta^*,\sigma^{2*},\phi^* \mid y\right) = \pi\left(\phi^* \mid \beta^*,\sigma^{2*},y\right)\pi\left(\beta^*,\sigma^{2*} \mid y\right)$$
$$= \pi\left(\beta^* \mid y\right)\pi\left(\sigma^{2*} \mid \beta^*,y\right)\pi\left(\phi^* \mid \beta^*,\sigma^{2*},y\right) \tag{3}$$

There are three components to be evaluated in equation (3), they are:

First component:

$$\pi\left(\beta^* \mid y\right) \approx \frac{1}{R}\sum_{r=1}^{R}\pi\left(\beta^* \mid \sigma_r^2,\phi_r,y\right) \tag{4}$$

can be computed directly using the draws of the original Gibbs sampler for (σ^2_r, φ_r). At each draw of the original Gibbs sampler we compute the conditional posterior $\pi(\beta \mid \sigma^2_r, \varphi_r, y)$ fixing β at β^* and then computing the average over the resulting density values.

Second component:

$$\pi\left(\sigma^{2*} \mid \beta^*,y\right) = \int \pi\left(\sigma^{2*} \mid \beta^*,\phi,y\right)\pi\left(\phi \mid \beta^*,y\right)d\phi \tag{5}$$

Since we do not have the draws from the original Gibbs sampler for the last term in the integral in equation (4), we can approximate it using a separate Gibbs sampler. Equation (5) can be shown to be easily approximated by[1]:

$$\pi\left(\sigma^{2*} \mid \beta^*,y\right) = \frac{1}{R}\sum_{r=1}^{R}\pi\left(\sigma^{2*} \mid \beta^*,\phi_r,y\right)$$

Third component:

$\pi\left(\phi^* \mid \beta^*,\sigma^{2*},y\right)$ can be evaluated using the known conditional posterior $\pi\left(\phi \mid \beta,\sigma^2,y\right)$ at $\phi^*,\beta^*,\sigma^{2*}$.

Therefore, the log marginal likelihood is given by:

$$\log\pi(y) =$$
$$\log\pi\left(y \mid \beta^*,\sigma^{2*},\phi^*\right) + \log\pi\left(\beta^*,\sigma^{2*},\phi^*\right) - \left(\log\pi\left(\beta^* \mid y\right) + \log\pi\left(\sigma^{2*} \mid \beta^*,y\right) + \log\pi\left(\phi^* \mid \beta^*,\sigma^{2*},y\right)\right)$$

As can be seen this procedure is straightforward (albeit a bit laborious) and for each additional parameter block we need to incorporate a separate Gibbs sampler to compute the last expression in equation (2).

---

[1] Details of this approximation can be found at:
http://www.ag.unr.edu/moeltner/RECO777/word/Lecture%20notes%20part%206.doc

Note:
For our regression models we have the posterior made up of factors such that:

$$p\left(\theta^* \mid y\right) = Normal\left(\beta^* \mid \tilde{\beta}, \left(\Sigma \otimes \left(X'X + A\right)^{-1}\right)\right) \times Inverted\,Wishart\left(\Sigma^* \mid \upsilon_0 + n, V_0 + S\right)$$

where exact expressions for the arguments of these distributions are provided on p.34 of the Rossi et al. 2005, equation (2.8.43).
If we take $\left(\beta^*, \Sigma^*\right) = \left(E\left[\beta\right], E\left[\Sigma\right]\right)$ as the reported means of our model, then at each assignment of data to segments we compute the theoretical mean and variance using (2.8.43), and evaluate the densities using the bayesm routine lndMvn(x,mu,rooti):

The (log) Inverted Wishart density (including normalizing constants) can be obtained with the bayesm subroutine: lndIWishart(nu,V,IW), where:

$$nu = \left(\upsilon_0 + n\right)$$

$$V = \left(V_0 + S\right)$$

IW = mean of draws of • = •*

References

Brown, P.J., M. Vannucci and T. Fearn (1998), "Multivariate Bayesian variable selection and prediction," *Journal of the Royal Statistical Society*. Series B, 60(3), 627-41.

Chib, Siddhartha (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*. 432(90), 1313-1321.

George, Edward I. and Robert E. McCulloch (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88 (September), 881-89.

George, Edward I. and Robert E. McCulloch (1997), "Approaches for Bayesian Variable Selection," *Statistical Sin.*, 7, 339-73.

Richardson, S. and P.J. Green (1997), "On Bayesian Analysis of mixtures with an unknown number of components," *Journal of the Royal Statistical Society*. Series B, 59, 731-92.

Rossi, Peter E., Greg M. Allenby and Robert McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, New York.