Quant Market Econ
DOI 10.1007/s11129-005-9006-5

Bayesian estimation of multivariate-normal models when dimensions are absent

Robert Zeithammer · Peter Lenk

Received: 3 March 2005 / Accepted: 16 December 2005

Abstract Multivariate economic and business data frequently suffer from a missing data phenomenon that has not been sufficiently explored in the literature: both the independent and dependent variables for one or more dimensions are absent for some of the observational units. For example, in choice based conjoint studies, not all brands are available for consideration on every choice task. In this case, the analyst lacks information on both the response and predictor variables because the underlying stimuli, the excluded brands, are absent. This situation differs from the usual missing data problem where some of the independent variables or dependent variables are missing at random or by a known mechanism, and the "holes" in the data-set can be imputed from the joint distribution of the data. When dimensions are absent, data imputation may not be a well-poised question, especially in designed experiments. One consequence of absent dimensions is that the standard Bayesian analysis of the multi-dimensional covariances structure becomes difficult because of the absent dimensions. This paper proposes a simple error augmentation scheme that simplifies the analysis and facilitates the estimation of the full covariance structure. An application to a choice-based conjoint experiment illustrates the methodology and demonstrates that naive approaches to circumvent absent dimensions lead to substantially distorted and misleading inferences.

Keywords Missing data · Data augmentation · Bayesian inference · Covariance estimation · Multinomial probit model · Choice-based conjoint analysis

R. Zeithammer (✉)
University of Chicago, Graduate School of Business,
5807 South Woodlawn Ave, Chicago, IL 60637
e-mail: rzeitham@chicagoGSB.edu

P. Lenk
University of Michigan, Ross Business School,
Ann Arbor, Michigan
e-mail: plenk@bus.umich.edu

JEL classifications C11 · C25 · D12 · M3

1. Introduction

A Bayesian analyst faces a practical difficulty with multivariate economic and business data when both the independent and the dependent variables for some measured dimensions are absent from some observations. The motivating example of this situation occurs in choice-based conjoint (CBC) experiments, where subjects evaluate multiple choice-sets consisting of product alternatives represented by profiles (see Haaijer and Wedel (2002) for a review of the CBC literature). Frequently, the choice-sets comprise subsets from a list of brands under study, and not all brands are included in every choice-set. The usual goal of the experiment is to learn about the inter-related structure of demand for all the brands, so demand for each brand is one dimension of a multivariate dataset, and the brands excluded from a particular choice-task are "absent dimensions" of that choice-task. Absent dimensions also occur in several other marketing contexts: In supermarket scanner-panel data, a category can consist of hundreds of SKUs, but not all are available to consumers in all stores during all weeks. In internet shopbot data, not all major retailers offer every book (Smith and Brynjolfsson, 2001). At the aggregate level, retail sales are usually reported at the product level, and different stores in a given market typically sell different assortments of products (Chintagunta, 2002). Finally, in fractional-factorial surveys, data usually contains individual responses to several related questions, and different versions of the survey may ask different subsets of the questions (Labaw, 1980).

To model the above situations, the multivariate normal distribution is widely used—if not for the observed responses, then for latent constructs—to allow for correlations among the dimensions. The varying absent dimensions of the observations make Bayesian inference about the parameters of the distribution difficult because the Inverted-Wishart prior for the covariance matrix is no longer conveniently conjugate to the likelihood of the data. Conjugacy breaks down because each observation in the data generically consists of a different subvector of the dimensions, and thus only involves a submatrix of the full covariance matrix, whereas the Inverse-Wishart distribution has a density involving the full covariance matrix. This paper proposes a simple data-augmentation method (Tanner and Wong, 1987) that facilitates Bayesian inference about the full covariance matrix under the condition of varying absent dimensions. By augmenting the absent residuals of the multivariate normal model, the procedure restores the conjugacy of the Inverted-Wishart prior and simplifies Bayesian inference. The residual augmentation avoids numerically costly Metropolis steps in Markov chain Monte Carlo (MCMC) and can easily be added to standard procedures.

We develop the residual augmentation procedure in the simplest situation of a multivariate regression model, which illustrates the essential features of the problem. After establishing the feasibility of the approach in that simple case, we extend it to hierarchical Bayes (HB) multivariate regression, which is the basis of our focal model, the HB multinomial-probit (HB-MNP) model (Albert and Chib, 1992; McCulloch and Rossi, 1994), commonly used for marketing choice-data. We then apply the HB-MNP model to a choice-based conjoint dataset where the choice tasks consist of varying subsets of brands. Not only do we illustrate that the procedure is effective but also show that several marketing conclusions are substantially impacted by allowing for error correlations.

Absent dimensions are qualitatively different from well-known missing data models where the process that generates the missing data is ignorable or can be modelled (see Little and Rubin, 2002). Common examples of missing data are nonresponse to items on a survey, or data-collection and processing errors. In missing data models, one assumes that the stimuli are present but the subjects' responses are missing, so the Bayesian procedure is to impute the missing observations from the joint distribution given the observed data. When not just the data but also the stimuli are absent, then imputing the entire dimension for the missing stimuli is not a well-poised problem. For example, in a CBC experiment, it is not meaningful to ask: "Would subject i have selected brand b if brand b was, in fact, part of the choice task?" We show that imputing only the absent residuals is sufficient for a Bayesian analysis of a multivariate-Normal model to proceed. The absent-dimensions situation is also different from Bradlow et al., (2004) who impute subjects' perceptions about absent attribute levels in partial-profile studies. In their model, all brands are included in each choice task, but some of the product-attributes are missing.

The paper is organized as follows: The following Section 2 reviews the existing approaches to Bayesian analysis of marketing data with missing dimensions, and highlights our contribution to the literature. Section 3 then develops the proposed residual augmentation procedure for a multivariate regression model, and extends it to estimate the multinomial probit model. Section 4 then applies the method to choice-based-conjoint data, and Section 5 concludes.

2. Review of the marketing applications of the multinomial probit model

Existing Bayesian MNP-estimation methods circumvent the difficulty posed by absent dimensions, either by throwing away data or by assuming independence across dimensions. The simplest and most widely used approach is to delete the observations with absent dimensions, with a resultant loss of information. This fix is not viable if most or all of the observations have missing dimensions, as in some CBC studies. A second approach is to restrict the analysis to a common subset of dimensions that are present for all observations. For example, in scanner-panel studies Allenby and Lenk (1994, 1995) and McCulloch and Rossi (1994) considered a few major brands of ketchup, detergents, and margarine that were available in all locations throughout the observational period, and eliminated brands with episodic presence. Because the focus of these papers was to develop hierarchical Bayes models and methods, and not to analyze completely the markets concerned, this restriction was not fatal to the papers' theses. However, focusing on a common core of brands negates the purpose of niche marketing and biases the choice structure in favor of brands with large market share. For example, the cross-price elasticities in MNP depend on all brands included in the model (Rossi et al., 1996), so restricting the study to large share brands can lead to biased inference about competition.

Another popular way to circumvent the absent-dimensions problem is to assume that the dimension-specific scalar errors are uncorrelated, and hence avoid the difficult estimation of covariances in the first place. This assumption distorts the resulting inference to a greater or lesser degree that depends on the model and degree of actual correlation. In the multivariate regression, the MLE point-estimate of the parameters is unaffected, but the error correlation manifests itself in probability statements, such as hypothesis testing and prediction. In the multinomial probit model, Hausman and Wise (1978) and Chintagunta (1992) demonstrate that incorrectly assuming independent errors results in inconsistent estimates.

Besides distorting inference, assuming away correlations of dimensions in the MNP eliminates its main advantage over the simpler multinomial logit model (MNL) as a model of utility-maximizing choice: By allowing the random-utilities of choice-alternatives to be mutually correlated, the MNP model allows more realistic substitution patterns than the MNL, which has the restrictive independence-of-irrelevant-alternatives (IIA) property.[1] Nevertheless, many marketing applications of the MNP constrain the random-utility correlations to zero (Haaijer et al., 1998, 2000; Rossi et al., 1996 and several others). All of these papers assume population heterogeneity in response-parameters, thereby escaping the IIA property at least somewhat in the aggregate behavior of the population shares. However, as Chintagunta (1992) points out, the IIA still holds within an individual respondent.

Interestingly, most applications of the MNP to conjoint analysis use the zero-correlation restriction (Haaijer et al., 1998, 2000). One reason for this may be the fact that the existing Bayesian-estimation methods allow only one particular correlation structure which is not theoretically compelling in the CBC setting, namely to associate random utility-shocks with the order of stimuli in the choice-task. This specification of the correlation structure is not compelling because, for example, it is not clear why someone who is more likely to choose the second profile on the list for an unobserved reason should also be systematically more or less likely to choose the fourth profile for an unobserved reason, with the actual attributes of the second and fourth profiles completely changing from task to task.

Compare the conjoint situation with the existing applications of the MNP to scanner data, where alternatives, and hence the random components of utility, are usually associated with brands (McCulloch and Rossi, 1994; Elrod and Keane, 1995) or brand-attribute combinations (Chintagunta (1992) considers brand-size combinations of ketchup, and Rossi et al. (1996), consider brand-liquid combinations of canned tuna). These correlations have the natural and compelling interpretation of unobserved pairwise similarities between products (SKUs) at the brand or brand-attribute level. These papers report significant non-zero correlations, so it seems that restricting the correlations is not only theoretically but also empirically problematic. Moreover, McCulloch and Rossi (1994) and Allenby and Rossi (1999) find positive correlations even in the presence of population heterogeneity in response-parameters, so merely allowing for population heterogeneity is not enough to capture real-world substitution patterns.

[1] The IIA property occurs when the preference comparison of two alternatives does not depend on the other alternatives that are available. One implication of IIA is that the introduction of a new alternative reduces the choice probabilities of existing alternatives on a proportional basis, which is particularly unrealistic when subsets of alternatives are close substitutes. Taken to its natural conclusion if IIA were true, a company could drive competitors out of business merely by offering superficial variations of its products.

In the standard Bayesian estimation methodology (McCulloch and Rossi, 1994), the utility-shocks must be associated with choice-alternatives in a one-to-one mapping. The aforementioned interpretability of these shocks in existing scanner-data analyses thus follows from the fortunate fact that the choice-alternatives modeled happen to be brands or brand-attribute combinations. The main contribution of this paper is to allow Bayesian analysts of the MNP to decouple the specification of random utility-shocks from the specification of choice-alternatives. In particular, the proposed method can be used to associate random shocks with any discrete variable that takes on mutually exclusive values in each choice-task. To keep the estimated covariance matrix small, such a variable should not take on too many values across all observations. For example, when each brand appears only once in each choice-set and there is only a small total number of brands, the random utility-components can be associated with brands even though the choice-alternatives are conjoint profiles. Therefore, a Bayesian CBC-analyst can associate random utility-components with brands as is standard in the scanner literature. Alternatively, the position of products on the shelf/webpage may be changing from one choice-occasion to another, with not all positions necessarily occupied at all times. Then, it is also possible to associate the utility-shocks with the position. We explore both possibilities in our conjoint empirical application and conclude in favor of associating utility-shocks with brands on the basis of superior holdout performance. The general empirical goal of characterizing the ideal specification of utility-shocks, both in the conjoint and in the scanner-data domains, should be a fascinating avenue for future research. For example, the scanner-data situation offers a whole range of possibilities: in the limit, random utility-shocks can always be associated with SKUs, but the resulting covariance matrices may be too large, warranting aggregation to, for example, brand-size level.

3. Residual augmentation for absent dimensions

Residual augmentation of absent dimensions is most easily seen in a multivariate regression model. After detailing the results for multivariate regression, we adapt the model to hierarchical Bayes (HB) multivariate regression, which is then used as the basis of the procedure for HB multinomial-probit models.

3.1. Multivariate regression

Consider the multivariate regression model:

$$Y_i = X_i\beta + \epsilon_i \text{ for } i = 1, \ldots, n$$

where Y_i is a m-vector of dependent variables for unit i; X_i is a $m \times p$ matrix of independent variables for unit i, β is a p-vector of regression coefficients; and ϵ_i is the error term. The error terms are mutually independent random variables from a multivariate normal distributionwith mean zero and covariance matrix Σ:

$$[\epsilon_i] = N_m(0, \Sigma).$$

We will use the bracket notation to signify the distribution of the argument. The prior distributions are:

$$[\beta] = N_p(b_0, V_0) \quad \text{and} \quad [\Sigma] = IW_m(f_0, S_0)$$

where $IW_m(f_0, S_0)$ is the m-dimensional, inverted Wishart distribution with f_0 prior degrees of freedom and scale parameters S_0. Under our specification of the Inverse Wishart distribution, the prior mean of Σ is $S_0/(f_0 - m - 1)$. By correctly defining the design matrices X_i, this model includes the standard multivariate regression model that has different regression coefficients for each component of Y_i and the seemingly unrelated regression models of Zellner (1971).

The case that we will study is when the same dimensions of Y_i and X_i are absent. Define $\mathcal{P}(i)$ to be the vector of indices corresponding to the present or absent dimensions for unit i, and $\mathcal{A}(i)$ to be a vector of indices corresponding to the absent dimensions for unit i. $\mathcal{A}(i)$ is the null vector if all of the dimensions are present. Define m_i to be the number of present dimensions for unit i. $Y_{\mathcal{P}(i)}$ and $X_{\mathcal{P}(i)}$ are the observations for unit i. The error covariance matrix is partitioned according to $\mathcal{P}(i)$ and $\mathcal{A}(i)$: $\Sigma_{\mathcal{P}(i),\mathcal{P}(i)}$ is the $\mathcal{P}(i)$ rows and columns of Σ corresponding to the present dimensions; $\Sigma_{\mathcal{A}(i),\mathcal{A}(i)}$ is the $\mathcal{A}(i)$ rows and columns of Σ corresponding to the absent dimensions, and $\Sigma_{\mathcal{A}(i),\mathcal{P}(i)}$ is the $\mathcal{A}(i)$ rows and $\mathcal{P}(i)$ columns of Σ.

The likelihood function of the data with absent dimensions is:

$$L(\beta, \Sigma) = \prod_{i=1}^{n} |\Sigma_{\mathcal{P}(i),\mathcal{P}(i)}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(Y_{\mathcal{P}(i)} - X_{\mathcal{P}(i)}\beta\right)' \Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)} \left(Y_{\mathcal{P}(i)} - X_{\mathcal{P}(i)}\beta\right)\right].$$

For the error covariances to be identified, every pair of dimensions must appear together at least once.

While a frequentist can simply maximize the above likelihood function over the parameter space, the absent dimensions create technical problems for Bayesian MCMC procedures. The full conditional distribution for β is a multivariate normal andnot difficult to generate:

$$[\beta|\text{Data}, \Sigma] = N_p(b_n, V_n)$$

$$V_n = \left(V_0^{-1} + \sum_{i=1}^{n} X'_{\mathcal{P}(i)} \Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)} X_{\mathcal{P}(i)}\right)^{-1}$$

$$b_n = V_n \left(V_0^{-1} b_0 + \sum_{i=1}^{n} X'_{\mathcal{P}(i)} \Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)} Y_{\mathcal{P}(i)}\right)$$

However, the full conditional distribution for Σ is no longer an inverted Wishart distribution due to the varying dimensions of $Y_{\mathcal{P}(i)}$ and $X_{\mathcal{P}(i)}$: If all dimensions are present, then the full conditional distribution for Σ given β depends on the residuals:

$$R_i = Y_i - X_i\beta, \text{ for } i = 1, \ldots, n.$$

Given full-dimensional residuals R_i, the aforementioned conjugacy of the inverted-Wishard to the likelihood in the full-conditional distribution follows because the density function of the inverted-Wishart has the same functional form, up to normalizing constant, as the likelihood for Σ given β:

$$[\Sigma] \propto |\Sigma|^{-(f_0+m+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S_0)\right]$$

$$L(\Sigma|\beta) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}SSE)\right]$$

$$SSE = \sum_{i=1}^{n}(Y_i - X_i\beta)(Y_i - X_i\beta)'$$

so the full conditional of Σ given β and the data is:

$$[\Sigma|\beta, Y] \propto |\Sigma|^{-(f_n+m+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S_n)\right]$$

$$f_n = f_0 + n \text{ and } S_n = S_0 + SSE.$$

This conditional conjugacy breaks down when only sub-vectors of the residuals are available:

$$R_{\mathcal{P}(i)} = Y_{\mathcal{P}(i)} - X_{\mathcal{P}(i)}\beta, \text{ for } i = 1, \ldots, n.$$

The conditional likelihood of Σ given β becomes:

$$L(\Sigma|\beta) \propto \prod_{i=1}^{n} |\Sigma_{\mathcal{P}(i),\mathcal{P}(i)}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)}SSE_i\right)\right]$$

$$SSE_i = (Y_{\mathcal{P}(i)} - X_{\mathcal{P}(i)}\beta)(Y_{\mathcal{P}(i)} - X_{\mathcal{P}(i)}\beta)'$$

where the outer products of the residuals SSE_i have different dimensions. The full conditional of Σ is a proper distribution, but not in the inverted-Wishart family.

A simple way to work around the difficulty, while preserving the usefulness of the Inverted Wishart prior, is to augment (Tanner and Wong, 1987) the residuals for the absent dimensions. Since a subset of multivariate Normal variables is multivariate Normal, residual augmentation is theoretically straightforward, and it is achieved by drawing the absent residuals $R_{\mathcal{A}(i)}$ for every observation i:

$$\left[R_{\mathcal{A}(i)}|R_{\mathcal{P}(i)}, \Sigma, \beta\right] = N_{m-m_i}\left(\mu_{\mathcal{A}(i)|\mathcal{P}(i)}, \Sigma_{\mathcal{A}(i)|\mathcal{P}(i)}\right)$$

$$\mu_{\mathcal{A}(i)|\mathcal{P}(i)} = \Sigma_{\mathcal{A}(i),\mathcal{P}(i)}\Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)}R_{\mathcal{P}(i)}$$

$$\Sigma_{\mathcal{A}(i)|\mathcal{P}(i)} = \Sigma_{\mathcal{A}(i),\mathcal{A}(i)} - \Sigma_{\mathcal{A}(i),\mathcal{P}(i)}\Sigma^{-1}_{\mathcal{P}(i),\mathcal{P}(i)}\Sigma_{\mathcal{P}(i),\mathcal{A}(i)}.$$

Together with the present residuals $R_{\mathcal{P}(i)}$, the augmentation produces full-dimensional residual vectors that can be used as pseudo-observations in the conditional posterior draw of Σ as if there were no absent dimensions:

$$[\Sigma|\text{Rest}] = IW_m(f_n, S_n)$$
$$f_n = f_0 + n$$
$$S_n = S_0 + \sum_{i=1}^{n} R_i R_i'$$

where R_i is a concatenation of $R_{\mathcal{P}(i)}$ and $R_{\mathcal{A}(i)}$.

The conditional posterior draw of β must now account for the augmented residuals $R_{\mathcal{A}(i)}$ because $R_{\mathcal{P}(i)}$ and $R_{\mathcal{A}(i)}$ are correlated. $R_{\mathcal{P}(i)}$ given $R_{\mathcal{A}(i)}$, the error of observation i, follows a Normal distribution because the distribution of a subset of multivariate Normal variables conditional on it complement is again Normal. The conditional distribution is thus still Normal as in the standard "full-dimensional" case, but the computation of the conditional mean and variance of β may at first seem cumbersome because each observation involves errors with generically different means and variances depending on the pattern of absence in that observation. We utilize a convenient implementation of the conditional posterior draw of β that uses existing "full-dimensional" code, and does not require the computation of conditional distributions of $R_{\mathcal{P}(i)}$ given $R_{\mathcal{A}(i)}$ observation by observation. In the standard code for the "full-dimensional" draw, set $X_{\mathcal{A}(i)} \leftarrow 0$, and $Y_{\mathcal{A}(i)} \leftarrow R_{\mathcal{A}(i)}$. The left arrow is used to indicate an assignment of a value to a variable to distinguish it from a mathematical identity. Then, the mean and variance of the conditional distribution of can be computed exactly as if there were no absent dimensions, i.e. using existing "full-dimensional" code:

$$[\beta|\text{Rest}] = N_p(b_n, V_n)$$
$$V_n = \left(V_0^{-1} + \sum_{i=1}^{n} X_i' \Sigma^{-1} X_i\right)^{-1}$$
$$b_n = V_n \left(V_0^{-1} b_0 + \sum_{i=1}^{n} X_i \Sigma^{-1} Y_i\right)$$

The implementation works because of the properties of inverses of partitioned matrices. This concludes the description of the modified Gibbs sampler.

We demonstrate this procedure with four simulations. The design of the first simulation is not very challenging to the algorithm: there are three dimensions and pairs of dimensions are randomly present. This simulation confirms that the algorithm works as the above computations indicate. The second and third simulations are designed to stress the algorithm. In the second simulation, one of the covariances is not identified because the two dimensions for that covariance are never jointly present: we will term this as "data nonidentification" because the nonidentification is due to the data structure. This simulation indicates that the lack of identification for one covariance

Table 1 Posterior parameters for multivariate regression simulation one and two

		Simulation one all pairs present		Simulation two pairs {1, 2} & {2, 3} present	
Regression coefficient	True	Posterior Mean	STD	Posterior Mean	STD
β_1	1.0	1.057	0.036	1.062	0.042
β_2	−1.0	−0.958	0.033	− 0.953	0.040
Error Covariance	True	Posterior Mean	STD	Posterior Mean	STD
$\sigma_{1,1}$	1.0	0.990	0.074	0.900	0.082
$\sigma_{1,2}$	0.6	0.622	0.078	0.586	0.076
$\sigma_{1,3}$	−0.5	−0.445	0.059	**0.072**	**0.451**
$\sigma_{2,2}$	1.4	1.358	0.105	1.517	0.096
$\sigma_{2,3}$	0.0	0.132	0.080	0.100	0.064
$\sigma_{3,3}$	0.8	0.809	0.062	0.724	0.065
Error Correlation	True	Posterior Mean	STD	Posterior Mean	STD
$\rho_{1,2}$	0.507	0.536	0.050	0.501	0.047
$\rho_{1,3}$	−0.559	−0.497	0.052	**0.088**	**0.554**
$\rho_{2,3}$	0.000	0.125	0.075	0.094	0.059

does not adversely affect estimation of the rest of the model parameters. The third simulation shows that the algorithm performs well even with a high degree of absent dimensions, and the fourth simulation illustrates the mixing of the chain.

In the first and second simulation there are $m = 3$ dimensions and $n = 500$ observations. There are two independent variables: X_1 was randomly sampled from a uniform distribution between plus and minus one, and X_2 was a dummy variable for the first dimension. That is, the intercepts for dimensions two and three were zero. Simulated data were generated from the model:

$$Y_{i,1} = X_{i,1}\beta_1 + \beta_2 + \epsilon_{i,1}, Y_{i,2} = X_{i,1}\beta_1 + \epsilon_{i,2}, \quad \text{and} \quad Y_{i,3} = X_{i,1}\beta_1 + \epsilon_{i,3}.$$

The true parameters were $\beta_1 = 1$ and $\beta_2 = -1$, and the covariance of $\{\epsilon_i\}$ was:

$$\text{cov}(\epsilon_i) = \begin{bmatrix} 1.0 & 0.6 & -0.5 \\ 0.6 & 1.4 & 0.0 \\ -0.5 & 0.0 & 0.8 \end{bmatrix}.$$

We purposely used a simple regression equation to focus attention on estimating the error covariance matrix. The priors were $[\beta] = N_2(0, 100I)$ and $[\Sigma] = IW_2(5, S_0)$ with S_0 set to center the prior on the identity matrix.

In the first simulation, all pairs of dimensions {1, 2}, {1, 3}, and {2, 3} were present with equal probabilities. In the second simulation, half of the dimensions were {1, 2} and half were {2, 3}, so that $\sigma_{1,3}$ is not identified with this data. The MCMC used a total of 100,000 iterations, discarding the first 50,000. Table 1 presents the pos-

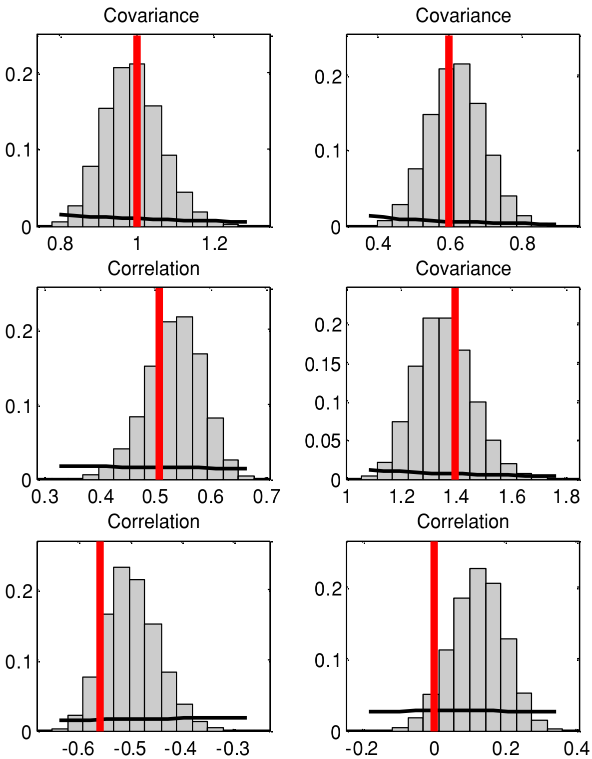

Fig. 1 Posterior distributions for the covariance and correlation for the first simulation of the multivariate regression model. Correlations are below the diagonal. Vertical solid lines are the true parameter values; the solid curves are prior densities, and histograms are posterior distributions

terior parameters for both cases. In simulation one, the residual imputation for the absent dimension results in accurate estimates for all of the parameters. The same is true for simulation two with the exception of the data unidentified covariance $\sigma_{1,3}$. Because the $\{1, 3\}$ pair was never present, the posterior mean is inaccurate and the posterior standard deviation is large. This result is not surprising. A more interesting result is that the other parameters are accurately estimated so that the data unidentifiability for $\sigma_{1,3}$ does not adversely effect the other estimators. Figures 1 and 2 display the prior and posterior distribution for the error covariances and correlations for the two simulations. These figures confirm the results from Table 1. The "U" shape for the posterior distribution of $\sigma_{1,3}$ in Fig. 2 was unexpected and remains unexplained.

The third simulation illustrates a high degree of absence. The number of dimensions is $m = 8$, and $n = 2000$ random pairs were generated using the same mean structure as in simulations one and two. The error variances were set as $\sigma_{k,k} = k$, and the error correlations were randomly drawn from a uniform distribution on $[-0.5, 0.5]$. The resulting covariance matrix was not well conditioned with the smallest eigenvalue of 0.0122. The sample used the same number of iterations as the first two simulations. The β parameters were estimated quite well at 1.022 (STD 0.030) and -0.972 (STD

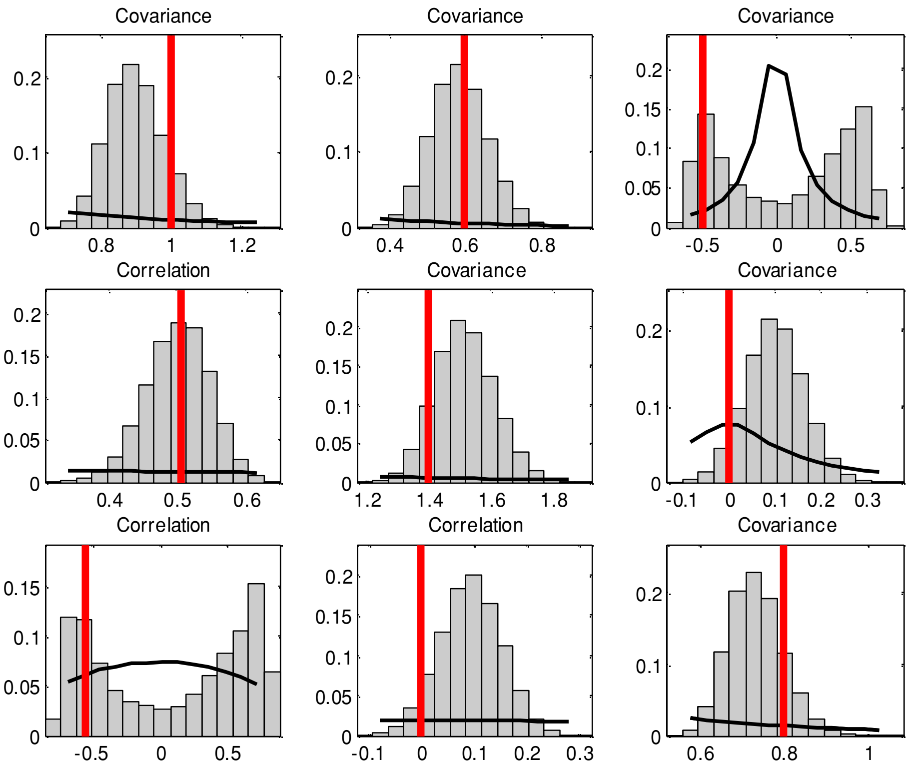

Fig. 2 Posterior distributions for the covariance and correlation for the second simulation of the multivariate regression model. Correlations are below the diagonal. Vertical solid lines are the true parameter values; the solid curves are prior densities, and histograms are posterior distributions

0.024). The error variances were also well recovered, which was expected given the large sample size. Figure 3 presents the posterior means and ± 2 posterior standard deviation error bars for the 28 covariances and correlations plotted against their true values. The 45^0 line provides a reference for the true parameters and their posterior means.

These simulations illustrate that the imputation of the absent dimensions facilitates the estimation of the covariance matrix. However, there is a small price to pay in terms of the mixing of the MCMC algorithm, which is demonstrated with the following simulation. Once again, $n = 500$ observations were generated from a multivariate regression model with $m = 3$ dependent measures and one independent variable. The independent variable was generated from a standard normal distribution with different values for each dependent measures. Table 2 displays the true parameters and their Bayes estimators for a full dataset and one with absent dimensions. One third of the dimensions were randomly deleted in the incomplete dataset. Once again, the posterior means for the full and incomplete data are close to their true values. The posterior standard deviations tend to be smaller for the full dataset, which has a total of 1500 dependent observations, than the incomplete dataset, which only has 1000 observations. Figure 4 graphs the autocorrelation function for the MCMC

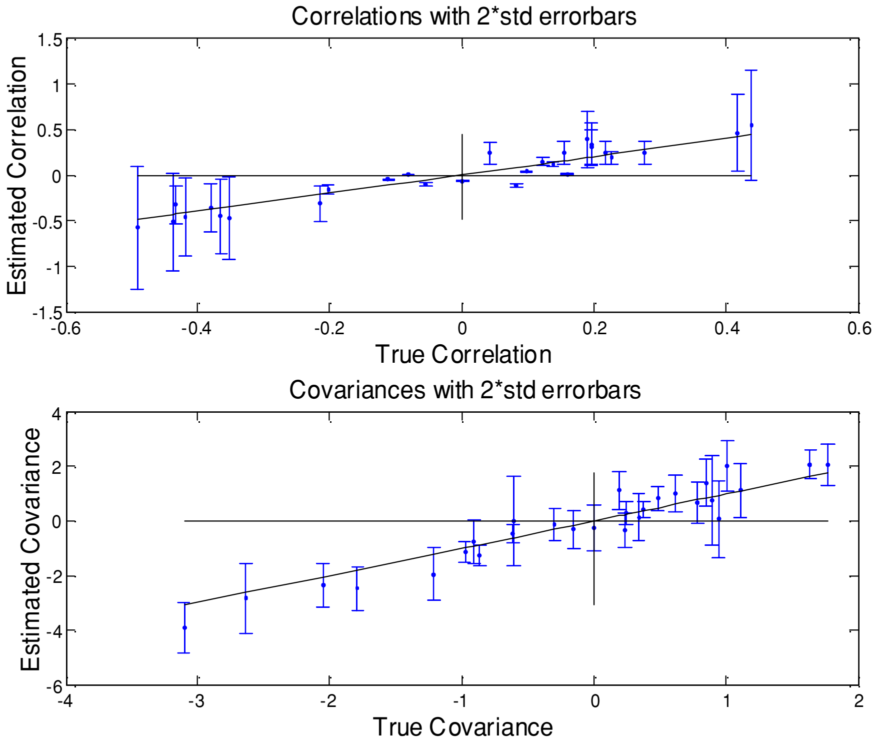

Fig. 3 Posterior distributions of the off-diagonal covariances and correlations in simulation three of the multivariate regression model

iterations. In each panel, the maximum, minimum, and median autocorrelation of six parameters (six regression coefficients and six covariances) are plotted against their lags. For the absent dimensions, the first few autocorrelations, especially for the covariance matrix, are substantial, especially compared to the full dataset. However, the comparison is not "fair" in that the analyst may not have the luxury of full-dimensional data, as pointed out in the Introduction. In any case, all autocorrelations rapidly decline with lag number, making inference feasible even with absent dimensions.

3.2. HB multivariate regression

The extension of the data augmentation method to hierarchical Bayes, multivariate regression is straight forward. There are repeated observations on each subject or sampling unit. The "within" units model is:

$$Y_{i,j} = X_{i,j}\beta_i + \epsilon_{i,j} \text{ for } i = 1, \ldots N; \text{ and } j = 1, \ldots, n_i$$

where $Y_{i,j}$ is a m-vector of dependent variables for observation j of unit i. The error terms $\{\epsilon_{i,j}\}$ are mutually independent from $N_m(0, \Sigma)$. Conceptually, the heterogeneity

Table 2 Multivariate regression: Parameter estimates for simulation 4

	True	Posterior mean		Posterior STD DEV	
		Full	Absent	Full	Absent
Constant 1	2.0	1.951	1.988	0.062	0.079
Constant 2	−1.0	−1.169	−1.300	0.092	0.112
Constant 3	3.0	2.867	2.863	0.122	0.144
X1	1.0	0.952	0.971	0.059	0.073
X2	0.0	−0.167	−0.202	0.095	0.118
X3	−2.0	−2.026	−1.888	0.119	0.144
$\sigma_{1,1}$	2.0	1.948	2.059	0.122	0.159
$\sigma_{2,1}$	0.5	0.509	0.396	0.131	0.238
$\sigma_{2,2}$	4.0	4.339	4.186	0.275	0.331
$\sigma_{3,1}$	−1.0	−1.175	−1.325	0.178	0.276
$\sigma_{3,2}$	0.0	−0.426	−0.760	0.259	0.436
$\sigma_{3,3}$	7.0	7.494	7.334	0.480	0.560

Simulation generated 500 observations for the full dataset. One third of the dimensions were randomly deleted. The full and incomplete datasets have a total of 1500 and 1000 dependent observations, respectively.

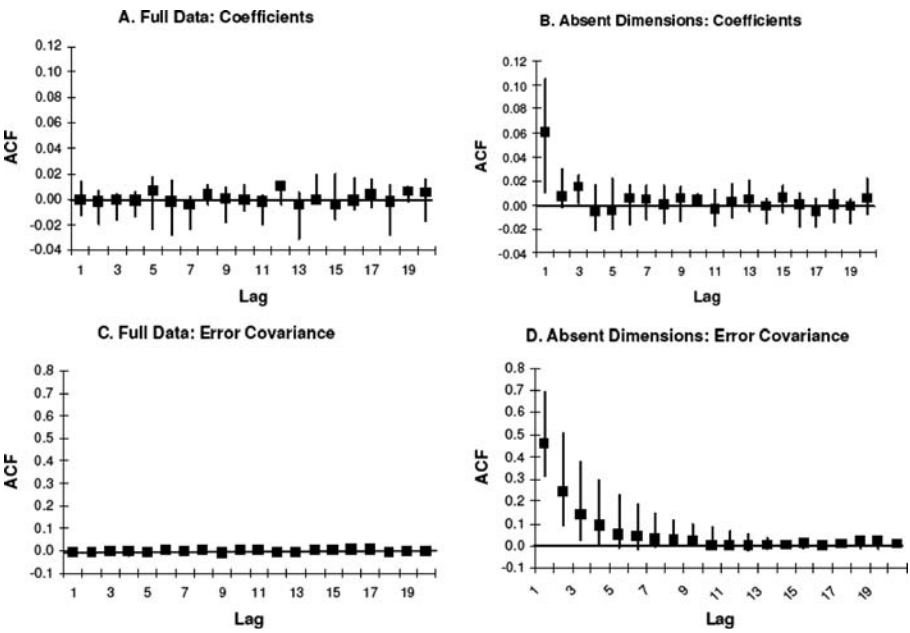

Fig. 4 Autocorrelation function from multivariate regression simulation 4. Minimum, median, and maximum ACF at each lag for the regression coefficients and covariances with full and absent dimensions

for the β_i can be any model; the exact form is not pertinent for our arguments. Once the present dimensions are augmented with residuals for the absent dimensions as in the pervious sub-section, the generation of Σ and β_i proceeds as standardly performed with complete data.

Define $\mathcal{A}(i, j)$ to be the absent dimensions of $Y_{i,j}$ and $X_{i,j}$, and $\mathcal{P}(i, j)$ to be the present dimensions. Let $m_{i,j}$ be the number of present dimensions and $m - m_{i,j}$ be the number of absent dimensions. Once again, the MCMC augments the present data with residuals from the absent dimensions. As before, assign zeros to absent independent variables $X_{\mathcal{A}(i,j)}$, and the augmented residuals to the absent dimensions of $Y_{i,j}$. Given β_i and Σ, generate the residuals for the absent dimensions:

$$\left[R_{\mathcal{A}(i,j)}|\Sigma, \beta_i\right] = N_{m-m_{i,j}}\left(\mu_{\mathcal{A}(i,j)|\mathcal{P}(i,j)}, \Sigma_{\mathcal{A}(i,j)|\mathcal{P}(i,j)}\right) \tag{1}$$

$$\mu_{\mathcal{A}(i,j)|\mathcal{P}(i,j)} = \Sigma_{\mathcal{A}(i,j),\mathcal{P}(i,j)}\Sigma^{-1}_{\mathcal{P}(i,j),\mathcal{P}(i,j)}R_{\mathcal{P}(i,j)} \tag{2}$$

$$R_{\mathcal{P}(i,j)} = Y_{\mathcal{P}(i,j)} - X_{\mathcal{P}(i,j)}\beta_i \tag{3}$$

$$\Sigma_{\mathcal{A}(i,j)|\mathcal{P}(i,j)} = \Sigma_{\mathcal{A}(i,j),\mathcal{A}(i,j)} - \Sigma_{\mathcal{A}(i,j),\mathcal{P}(i,j)}\Sigma^{-1}_{\mathcal{P}(i,j),\mathcal{P}(i,j)}\Sigma_{\mathcal{P}(i,j),\mathcal{A}(i,j)}. \tag{4}$$

Augment the present dimensions $Y_{\mathcal{P}(i,j)}$ with $R_{\mathcal{A}(i,j)}$ to obtain a complete-m vector $Y_{i,j}$.

After imputing the absent residuals, the rest of the MCMC follows from standard arguments, regardless of the rest of the model specification. In the empirical example, we will use a multivariate regression model (Lenk et al., 1996) for the "between" units model of heterogeneity of $\{\beta_i\}$:

$$\beta_i = \Theta' z_i + \delta_i \text{ for } i = 1, \ldots, n \tag{5}$$

where Θ is a $q \times p$ matrix of regression coefficients; z_i is a q-vector of covariates for unit i, and δ_i are mutually independent error terms from $N_p(0, \Lambda)$. Suppose, moreover, that the prior distributions are:

$$[\Sigma] = IW_m\left(f_{0,\Sigma}, S_{0,\Sigma}\right); [\Lambda] = IW_p\left(f_{0,\Lambda}, S_{0,\Lambda}\right);$$
$$\text{and } [\text{vec}(\Theta')] = N_{pq}\left(u_0, V_0\right) \tag{6}$$

where vec(Θ') stacks the rows of Θ. Appendix A details the full conditional distributions for the MCMC algorithm. One could just as easily use different forms of heterogeneity, such as mixture models (Lenk and DeSarbo, 2000). The main point is the residual augmentation is conditionally independent of the heterogeneity specifications. We highlight the multivariate regression specification because we will use it in the application.

We illustrate the algorithm with a simulated data set where the between-units model is in Eq. (5). In the simulation, the number of dimensions is $m = 4$; $N = 500$; and n_i varied from 11 to 20 with an average of 15.6 and standard deviation of 2.9. Each $Y_{i,j}$ and $X_{i,j}$ had one or two absent dimensions. In addition to the four intercepts, there were two additional predictor variables. The covariate z_i consisted of an intercept and two other variables. Table 3 reports fit statistics between the true

Table 3 HB multivariate regression: Fit statistic for individual-level coefficients

	Correlation	RMSE
Intercept 1	0.972	1.824
Intercept 2	0.732	1.970
Intercept 3	0.692	2.140
Intercept 4	0.864	2.319
X1	0.998	0.364
X2	0.969	0.662

Correlation and RMSE are the correlation and root mean squared error between the true regression coefficient and the posterior mean.

β_i from the simulation and its posterior mean, and Table 4 displays the true model parameters and their posterior means and standard deviations. Both tables indicates that the procedure does recover the model structure even when nearly half of the dimensions where absent.

3.3. HB multinomial-probit

The data imputation method for the multivariate models in Sections 3.1 and 3.2 extends to multinomial-probit (MNP) models, which assume that the observed, nominal responses arise form latent or unobserved, multivariate normal random variables. Because our application uses consumer choice data from a choice-based conjoint analysis (CBC), we will develop the MNP with that application in mind. In choice experiments like CBC, subjects are presented with sets of alternatives also called "profiles," each profile describing a product or service using brands and other attributes. From each set of profiles, a subject is asked to pick the one he or she most prefers.

The multinomial probit model captures the choice-behavior by assuming the each subject maximizes latent utility over the set of profiles. The latent utility of each profile is given by a linear combination of the attributes $X\beta$ plus a "random component" term ε that captures an idiosyncratic unobserved random shock. Suppose that the experiment design involves m brands, and each brand appears at most once in each choice-set of profiles. A theoretically meaningful way to specify the covariance between any pair of ε is to treat brands as dimensions of the underlying multivariate regression of latent utilities on attributes. Since not all brands are present in all choice-tasks,the underlying regression has absent dimensions, and the augmentation methodology discussed in Section 3.1 can be used in the MNP context almost immediately.

In choice task j, subject i evaluates the brands given by $\mathcal{P}(i, j)$, which is a subset of the m brands. The absent brands are indexed by $\mathcal{A}(i, j)$, which is the complement of $\mathcal{P}(i, j)$. $Y_{\mathcal{P}(i,j)}$ a $m_{i,j}$-vector of unobserved random utilities for the brands in the choice set $\mathcal{P}(i, j)$. Subject i selects that brand that maximizes his or her utility for the brands. If the subject picks brand $c_{i,j}$:

$$Y_{i,j,c_{i,j}} = \max_{u\in\mathcal{P}(i,j)} Y_{i,j,u} \text{ for } i = 1, \ldots N, \quad j = 1, \ldots, n_i, \text{ and } c_{i,j} \in \mathcal{P}(i, j). \tag{7}$$

The observed data are the indices of the preferred alternatives $\{c_{i,j}\}$ given the choice sets $\{\mathcal{P}(i, j)\}$. The analysis of the HB-MNP model follows that of the HB multivariate regression model, with the additional generation of the utilities for the brands in the choice task. As in HB multivariate regression, we will assume that the utility of the

Table 4 HB multivariate regression: Parameter estimates

	Y1	Y2	Y3	Y4		
True Σ						
Y1	1.0	0.1	0.0	1.0		
Y2	0.1	4.0	0.0	4.1		
Y3	0.0	0.0	9.0	0.0		
Y4	1.0	4.1	0.0	21.0		
Posterior mean Σ						
Y1	1.004	0.068	0.154	0.935		
Y2	0.068	4.052	0.180	4.111		
Y3	0.154	0.180	9.131	0.166		
Y4	0.935	4.111	0.166	21.529		
Posterior STD Σ						
Y1	0.022	0.048	0.070	0.115		
Y2	0.048	0.092	0.153	0.198		
Y3	0.070	0.153	0.203	0.335		
Y4	0.115	0.198	0.335	0.469		
	Intercept 1	Intercept 2	Intercept 3	Intercept 4	X1	X2
True Θ						
Constant	−15.0	−5.0	5.0	20.0	−5.0	3.0
Z1	2.0	1.0	0.0	−2.0	1.0	−0.2
Z2	−1.0	−0.5	0.0	1.0	−0.2	0.5
Posterior mean Θ						
Constant	−14.778	−6.497	5.521	18.754	−4.168	−2.199
Z1	1.745	0.920	−0.203	−2.148	0.951	0.282
Z2	−0.798	−0.295	0.070	1.333	−0.186	0.530
Posterior STD Θ						
Constant	1.650	1.750	1.838	1.920	1.545	1.625
Z1	0.498	0.529	0.557	0.578	0.472	0.498
Z2	0.267	0.277	0.295	0.294	0.243	0.256
True Λ						
Intercept 1	0.250	−0.500	0.750	0.000	0.125	−0.150
Intercept 2	−0.500	2.000	−1.000	0.000	−0.750	−0.700
Intercept 3	0.750	−1.000	4.750	0.000	1.625	−0.875
Intercept 4	0.000	0.000	0.000	4.000	0.000	0.000
X1	0.125	−0.750	1.625	0.000	7.563	2.975
X2	−0.150	−0.700	−0.875	0.000	2.975	11.093
Posterior mean Λ						
Intercept 1	0.277	−0.002	−0.107	0.251	0.432	0.586
Intercept 2	−0.002	2.160	−1.571	−0.421	0.034	0.252
Intercept 3	−0.107	−1.571	3.363	−1.207	2.255	−0.377
Intercept 4	0.251	−0.421	−1.207	3.951	0.726	0.678
X1	0.432	0.034	2.255	0.726	8.586	3.281
X2	0.586	0.252	−0.377	0.678	3.281	10.414
Posterior STD Λ						
Intercept 1	0.136	0.189	0.161	0.208	0.275	0.490
Intercept 2	0.189	0.376	0.284	0.347	0.395	0.570
Intercept 3	0.161	0.284	0.477	0.484	0.455	0.549
Intercept 4	0.208	0.347	0.484	0.642	0.542	0.695
X1	0.275	0.395	0.455	0.542	0.562	0.478
X2	0.490	0.570	0.549	0.695	0.478	0.721

product description or profile is linear in the attributes:

$$Y_{\mathcal{P}(i,j)} = X_{\mathcal{P}(i,j)}\beta_i + \epsilon_{i,j}. \quad (8)$$

Nonlinear specifications are possible with the addition of Metropolis steps. We identify the model by setting $Y_{i,j,m}$ to zero, and the first scalar variance to one. We use McCulloch and Rossi's, (1994) method of postprocessing the draws to enforce the restriction that one error variance is one.

The Gibbs sampler for MNP differs from the multivariate regression model in that the latent utilities $Y_{\mathcal{P}(i,j)}$ of the present dimensions are also augmented. The conditional draws of $R_{\mathcal{A}(i,j)}$, β, and Σ remain exactly as in Section 3.1. In particular, existing code can again be used to draw β using the implementation trick of setting $(X_{\mathcal{A}(i,j)} \leftarrow 0, Y_{\mathcal{A}(i,j)} \leftarrow R_{\mathcal{A}(i,j)}.)$ The new draw of $Y_{\mathcal{P}(i,j)}$ follows the same"Gibbs-through" procedure as in McCulloch and Rossi (1994), but the conditional mean and variance need to be adjusted for the augmented absent residuals $R_{\mathcal{A}(i,j)}$ observation by observation:

$$\left[Y_{\mathcal{P}(i,j)}|\text{Rest}\right] \propto N_{m_i}\left(\mu_{\mathcal{P}(i,j)|\mathcal{A}(i,j)}, \Sigma_{\mathcal{P}(i,j)|\mathcal{A}(i,j)}\right)\chi\left(Y_{i,j,c_{i,j}} = \max_{u\in\mathcal{P}(i,j)} Y_{i,j,u}\right)$$

$$\mu_{\mathcal{P}(i,j)|\mathcal{A}(i,j)} = X_{\mathcal{P}(i,j)}\beta_i + \Sigma_{\mathcal{P}(i,j),\mathcal{A}(i,j)}\Sigma^{-1}_{\mathcal{A}(i,j),\mathcal{A}(i,j)}R_{\mathcal{A}(i,j)}$$

$$\Sigma_{\mathcal{P}(i,j)|\mathcal{A}(i,j)} = \Sigma_{\mathcal{P}(i,j),\mathcal{P}(i,j)} - \Sigma_{\mathcal{P}(i,j),\mathcal{A}(i,j)}\Sigma^{-1}_{\mathcal{A}(i,j),\mathcal{A}(i,j)}\Sigma_{\mathcal{A}(i,j),\mathcal{P}(i,j)}$$

$$\chi(A) = \text{indicator function for the set } A.$$

A hierarchical specification can be implemented as in the case of the multivariate regression. For example, if part-worth heterogeneity is described by Eq. (5) and if the prior distributions are specified by Eq. (6), then the updating follows that of Section 3.2.

McCulloch et al. (2000) and Imai and van Dyk (2005) provide alternative models and algorithms for handling the error variance constraints in MNP models. As is well known (Imai and van Dyk, 2005; McCulloch et al., 2000), the ACFs in the full-dimensional probit samplers are much greater than in the linear regression, often dissipating only after 100 lags or so. We do not find a significant detrimental effect of absent dimensions beyond this. Our algorithm for absent dimensions does not depend on the method of enforcing the variance constraints, and it can be used with any of these alternatives.

4. Stated preference experiment

We illustrate the absent dimension problem with a choice-based conjoint (CBC) dataset provided by Bryan Orme of Sawtooth Software. We will use the HB-MNP model in Section 3.3. In this example, the brands included in the choice sets vary across choice tasks. The absent dimensions correspond to the brands that are excluded from each choice task. One method of modelling this data is to assume independence of the random utilities. We will show that this assumption distorts the posterior inference because the specification with full error covariance detects substantial error correlations in the random utilities. We also consider a specification that associates utility-shocks with order of presentation, and we find that the model with unrestricted Σ associated with brands outperforms both alternative specifications in terms of out-of-sample predictive ability.

The 326 subjects in the experiment were purchasing managers of information technology hardware and software. The experiment consisted of five brands of personal computers. Each subject was presented with eight choice tasks. Each choice tasks consisted of three descriptions of personal computers, and each description included one of the five brands. Subjects could also choose "None of the Above," which we used as the base choice by setting its utility to zero. If "None of the Above" was selected, we assumed that the utilities for the PC profiles in the choice task were less than zero. If a PC was selected, the utility of its profile was greater than zero. In addition to brand, there were four product attributes: Performance, Channel, Warrantee, and Service, each with three levels, and there were four price levels.

Table 5 describes the attributes and their dummy variable coding. The attribute levels were balanced across profiles. Each choice task was randomly constructed from the brands and attributes with some utility balancing. That is, none of the profiles in a choice task obviously dominated the others, such as a high performance computer with a low price. Each brand appeared in approximately 60% of the choice tasks. Brands A and B are generally considered to be premium brands, while Brands D and E are value brands. The aggregate choice shares were 24.5% for Brand A, 19.1% for Brand B, 17.4% for Brand C, 14.1% for Brand D, 12.0% for Brand E, and 12.8% for "None of the Above." The brands were also randomized across presentation order in a choice task. The choice shares by presentation order were approximately equal with 2% spread between the first and third choices. The survey also collected demographic information and information about the purchasing manager's firm. The subset of these covariates that we use in this example are also presented in Table 5.

We will use the HB-MNP model from Section 3.3, which is derived from the random utility model in Eqs. (7) and (8). We used the first seven choice tasks for calibration and the last choice task, which varies across subjects, for hold-out predictions. The model for subject-level heterogeneity is given by Eq. (5), and the prior distributions are given by Eq. (6). We fitted three models to the data. Model 1 used a full error covariance matrix ($\mathrm{var}(\epsilon_{i,j}) = \Sigma$) for all five brands concepts. This implies that a subject's brand preference consists of fixed (partworths for brand) and random (unobserved utility-shocks) components. A subject's brand partworths are assumed to be fixed through out the conjointstudy, while the utility-shock changes with choice task. One possible reason for associating the utility-shocks with brands is to related the CBC model-specification to standard scanner-data specifications reviewed in Section 2. The goal underlying such aspecification is to model unobserved similarities between brands while allowing additional unobserved variables to perturb each subject's utility task by task. Such random-seeming perturbations can arise from pure white-noise error in judgement, or from the possibility that attribute levels modify the subject's perception for the brand. For example, if a premium brand lacks premium features or has a low price in a particular profile, subjects may view these profiles in a much different light

Table 5 CBC experiment: Attributes and covariates

Conjoint design		Covariates	
Dummy variable	Attribute level	Dummy variable	Attribute level
	Brand		*Expect to pay*
Brand A	A	Pay Low	Low
Brand B	B		Medium – base
Brand C	C	Pay High	High
Brand D	D		
			Expertise for PC purchases
Brand E	E		Novice – base
	None – base	Expert	Expert
	Performance		
			Respondent gender
Low Prf	Below average		Man – base
	Average – base	Female	Woman
High Prf	Above average		
			Company Size
	Sales Channel	Small firm	Small
Tele buy	Telephone		Medium – base
	Retail store – base	Large firm	Large
Site Buy	Sale on site		
	Warrantee		
Short Wrt	90 day		
	1 year – base		
Long Wrt	5 year		
	Service and Repair		
MFG Fix	Ship back to manufacturer		
	Local dealer – base level		
Site Fix	On-site service		
	Price		
	Low – base		
Price 2	Medium-low		
Price 3	Medium-high		
Price 4	High		

than if the premium brand has premium features and high price. That is, the overall evaluations are different than what would would expect if the fixed brand preference are added to the attribute partworths. These differences then become part of the error term, which varies among all profiles.

Model 2 assumed that the covariances between utility-shocks in Model 1 were zero, and Model 3 associated the utility-shocks with the presentation-order of the profiles instead, using a full unrestricted 3×3 Σ. Except for the specification the error variance, the three models were identical. In Model 2, inverted gamma priors were used for the error variances, and their parameters were set to induce roughly the same prior mean and variance as the marginal priors from Model 1 and 3, so that potential differences in the estimation results can not be attributed to differences in prior specifications.

Table 6 CBC probit: Posterior mean of rror variances and covariances

	Brand A	Brand B	Brand C	Brand D	Brand E
Model 1					
Brand A	0.889	0.174	– 0.156	– 0.716	0.040
Brand B	0.174	0.860	0.055	0.037	– 0.564
Brand C	– 0.156	0.055	0.961	– 0.247	– 0.754
Brand D	– 0.716	0.037	– 0.247	0.875	0.135
Brand E	0.040	– 0.564	– 0.754	0.135	1.000
Model 2					
Brand A	1.042	0.000	0.000	0.000	0.000
Brand B	0.000	1.041	0.000	0.000	0.000
Brand C	0.000	0.000	1.053	0.000	0.000
Brand D	0.000	0.000	0.000	1.036	0.000
Brand E	0.000	0.000	0.000	0.000	1.000
Model 3	Order 1	Order 2	Order 3		
Order 1	1.386	– 0.569	– 0.617		
Order 2	– 0.569	1.107	– 0.535		
Order 3	– 0.617	– 0.535	1.000		

Model 1 uses a full, error covariance matrix for brand concepts; Model 2 assumes that error are independent; and Model 3 uses a full, error covariance for presentation order.

The MCMC used 50,000 iterations for burn-in, and 50,000 iterations for estimation. After the burn-in every fifth iteration was saved for estimation purposes for a total of 10,000 iterations. The point of comparing the three models is to demonstrate that the assumption about the random utilities' error covariance matrix does matter and has practical implications.

Table 6 reports the estimated variances and covariances for the three models. The models were identified by setting the error variance for Brand E to one. In Model 1, the error variances are close to one. Brand A and B have a slightly positive correlation. Brand E has sizeable, negative covariances with Brands B and C, and Brands D and A also have negative covariances. These covariances are important in interpreting the model and can greatly impact market share. For example, if a purchasing manager likes Brand A more than what the model predicts from his or her part-worths, then this manager would tend to dislike Brand D more than predicted from his or her part-worths. If Brand E was introduced into a market that only has Brands A, B, and C, then Brand E would tend to draw more heavily from customers whose utilities for Brands B or C are below their expectations. In contrast, Model 2, with uncorrelated utilities, has the IIA property, and Brand E would draw from Brands A, B, and C proportional to their market shares. In Model 3, the correlations of the error terms, which are associated with presentation order, are negative. This implies that if a subject prefers the first profile more than attributes alone would predict, then he or she tends to prefers the second and third profiles less than the attributes would predict. This result holds true for all of the profiles.

Table 7 displays the estimated regression coefficients, Θ in Eq. (5), that relate the individual-level part-worths to the subject or firm level covariates. To simplify the comparison between the three models, only posterior means that where greater than

Table 7 CBC Probit: Explained heterogeneity in selected subject-level part-worths

	Constant	Low pay	High pay	Expert	Female	Small firm	Large firm
Model 1							
Brand A	0.790						
Brand B	0.890					−0.394	
Brand C	0.518				0.365	−0.350	
Brand D		0.298	−0.391				
Brand E			−0.421				
Price 2							− 0.236
Price 3	−0.864		0.329				
Price 4	−1.118	−0.431	0.253			0.240	
Model 2							
Brand A	1.361						
Brand B	1.421		−0.701				
Brand C	1.052		−0.709			−0.417	
Brand D	1.021		−0.940			−0.548	
Brand E	0.690		−0.965				
Price 2				−0.309			
Price 3	−1.031		0.421				
Price 4	−1.456	−0.578	0.475				
Model 3							
Brand A	0.644						
Brand B	0.511						
Brand C		0.476		0.383	0.410		
Brand D		0.351					
Brand E			−0.494				
Price 2					−0.396		
Price 3	−0.705						
Price 4	−1.120	−0.374					

Model 1 uses a full, error covariance matrix for brand concepts; Model 2 assumes that errors for brand concepts are independent; and Model 3 uses a full, error covariance matrix for presentation order. Posterior means that are less than two posterior standard deviations in absolute value are suppressed.

two posterior standard deviations in absolute value are displayed, and to keep the table the a manageable size, only the coefficients for Brand and Price are shown. (Coefficients for other partworths were significant, but their interpretation is not particularly relevant to this paper's theme.) The direction of the results is similar for the two models. The coefficients for Model 2 tends to be somewhat larger than from Models 1 and 3. However, it is difficult to discern generalizable results about the impact of the error structure on the explained heterogeneity.

Table 8 presents the estimated population variances, the diagonal elements of Λ, of the error terms in Eq. (5). The covariances for Models 1 and 2 are significantly different from zero, but not very large. A distinguishing feature of the models is that the variances for Model 2 tend to be larger than for Models 1 and 3. In fact, the natural logarithms of the determinants of the full covariance matrix Λ are –64.85 and –65.10 for Models 1 and 2 and –39.32 for Model 2. The ln-determinant is a measure of the error variation with larger numbers representing greater variation in unexplained heterogeneity. It appears as though incorrectly assuming that the latent utilities are independent forces over estimation of unexplained heterogeneity in the between-subjects model to compensate for the lack of fit in the within-subjects model.

Table 8 CBC probit: Unexplained heterogeneity in subject-level part-worths

	Model 1	Model 2	Model 3
Brand A	0.307	0.687	0.407
Brand B	0.117	0.290	0.081
Brand C	0.030	0.187	0.052
Brand D	0.173	0.507	0.214
Brand E	0.254	0.617	0.340
Low Perfomance	0.421	0.833	0.072
High Performance	0.144	0.317	0.077
Telephone Channel	0.015	0.048	0.019
Site Channel	0.022	0.089	0.026
Short Warranty	0.122	0.255	0.076
Long Warranty	0.129	0.181	0.109
MFG Fix	0.165	0.452	0.100
Site Fix	0.020	0.139	0.021
Price 2	0.022	0.066	0.042
Price 3	0.033	0.105	0.042
Price 4	0.094	0.200	0.065
Ln Determinant	−64.854	−39.322	−65.099

Diagonal elements (variances) of the posterior means of Λ, the error covariance matrix for the partworths. Model 1 uses a full, error covariance matrix associated with brand concepts; Model 2 assumes that brand-concept errors are independent; and Model 3 uses a full, error covariance matrix associated with presentation order.

One way to ascertain the overall effect of these differences in error structure is to examine the predictive performance for the models. The eighth and last choice task was used as a hold-out sample for prediction. Because the choice tasks and profiles were randomly constructed, different subjects had different hold-out choice tasks. We compute two measures of predictive performance: the hit rate and the Brier score. The hit rate is the proportion of time that the maximum of the predicted utilities corresponds to the subjects' selected profile in the hold-out sample. The Brier score (c.f. an Gordon and Lenk, 1991, 1992) is a root mean squared error measure between the profile selections (1 if selected and 0 if not) and predictive probabilities for the profiles. The Brier scores are computed within the MCMC iterations and averaged over iterations, thus reflecting parameter uncertainty. The following table presents these predictive measures for the three models. On both measures, Model 1 predicts the best; Model 2 the second best, and Model 3 is the worst. Model 1 performs substantially better than other two models. The hit rate for Model 1 is 8.4% better than Model 2 and 16.1% better than Model 3. These improvements are striking given that the only change in the models is the structure of the error terms for the random utilities.

Predictive Performance

	Hit rate	Improvement	Brier score	Reduction
Model 1	56.6%		0.377	
Model 2	52.2%	8.4%	0.479	21.4%
Model 3	48.8%	16.1%	0.508	25.8%

5. Conclusion

Absent dimensions occur when both the independent and dependent variables are absent for one or more components in multivariate data. This situation is qualitatively different from standard missing data problems and frequently is not considered to be a problem. However, absent dimensions can negatively effect researcher's modelling choices. Naive methods of circumventing absent dimensions include deleting observations with absent dimensions, restricting the dimensions to a subset of dimensions that is fully present, or assuming independence among the dimensions. All three of these approaches represent compromises: deleting partial observations wastes information; restricting the phenomenon to a common subset of dimensions limits the ability to generalize the inference; and simplifying assumptions, such as independence, distort the inference.

In the models considered in this paper, absent dimensions make it difficult to estimate covariance matrices because the inverted Wishart prior for the covariance matrix is no longer conjugate. We propose a data imputation scheme that facilitates covariance estimation by making the inverted Wishart distribution conjugate for the completed data matrix. The Metropolis algorithm is an alternative to data imputation; however, data imputation is fast and effective and circumvents implementation issues associated with Metropolis algorithms, such as choice of proposal distribution, acceptance rates, and mixing of that omnibus method.

The residual augmentation method outlined in this paper is a minor modification of existing MCMC procedure and only requires an additional vector to denote present and absent dimensions for each observation. The paper presents simulation studies that document the effectiveness of the algorithm and a choice based conjoint study that illustrates how skirting the problem by assuming independent utility-shocks can distort the results.

The main contribution of this paper to Bayesian estimation of MNP choice-models is the new ability of the analyst to decouple utility-shocks from choice-alternatives. While prior Bayesian estimation methods required that utility-shocks be associated with alternatives in a one-to-one mapping, the proposed method allows them to be associated with almost any discrete attribute of the alternatives.

Appendix

A: MCMC algorithm

This appendix gives the remainder of the MCMC algorithm for the HB-Multivariate regression model in Section 3.2. The "within" units model is:

$$Y_{i,j} = X_{i,j}\beta_i + \epsilon_{i,j} \text{ for } i = 1, \ldots N; \text{ and } j = 1, \ldots, n_i,$$

and the "between" units model is:

$$\beta_i = \Theta' z_i + \delta_i \text{ for } i = 1, \ldots, n$$

The prior distributions are:

$$[\Sigma] = IW_m\left(f_{0,\Sigma}, S_{0,\Sigma}\right); [\Lambda] = IW_p\left(f_{0,\Lambda}, S_{0,\Lambda}\right);\ \text{and}\ [\text{vec}(\Theta')] = N_{pq}\left(u_0, V_0\right)$$

where $\text{vec}(\Theta')$ stacks the rows of Θ.

Then the MCMC draws are:

$$\begin{aligned}
[\beta_i|\text{Rest}] &= N_p(b_{N,i}, B_{N,i}) \\
B_{N,i} &= \left(\Lambda^{-1} + \sum_{j=1}^{n_i} X'_{i,j}\Sigma^{-1}X_{i,j}\right)^{-1} \\
b_{N,i} &= B_{N,i}\left(\Lambda^{-1}\Theta' z_i + \sum_{j=1}^{n_i} X'_{i,j}\Sigma^{-1}Y_{i,j}\right)
\end{aligned}$$

$$\begin{aligned}
[\text{vec}(\Theta')|\text{Rest}] &= N_{pq}\left(u_N, V_N\right) \\
V_N &= \left(V_0^{-1} + \left[\sum_{i=1}^{N} z_i z'_i\right] \otimes \Lambda^{-1}\right)^{-1} \\
u_N &= V_N\left(V_0^{-1}u_0 + \sum_{i=1}^{N} z_i \otimes \Lambda^{-1}\beta_i\right)
\end{aligned}$$

$$\begin{aligned}
[\Sigma|\text{Rest}] &= IW_m\left(f_{N,\Sigma}, S_{N,\Sigma}\right) \\
f_{N,\Sigma} &= f_{0,\Sigma} + \sum_{i=1}^{N} n_i \\
S_{N,\Sigma} &= S_{0,\Sigma} + \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[Y_{i,j} - X_{i,j}\beta_i\right]\left[Y_{i,j} - X_{i,j}\beta_i\right]' \\
[\Lambda\mid\text{Rest}] &= IW_m\left(f_{0,\Lambda}, S_{N,\Lambda}\right) \\
f_{N,\Lambda} &= f_{0,\Lambda} + N \\
S_{N,\Lambda} &= S_{0,\Lambda} + \sum_{i=1}^{N}\left[\beta_i - \Theta' z_i\right]\left[\beta_i - \Theta' z_i\right]'
\end{aligned}$$

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669–679.

Allenby, G., & Lenk, P. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association*, *83*(428), 1218–1231.
Allenby, G., & Lenk, P. (1995). Reassessing brand loyalty, price sensitivity, and merchandizing effects on consumer brand choice. *Journal of Business and Economic Statistics*, *13*(3), 281–290.
Allenby, G., & Rossi, P. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, *89*, 57–78.
Bradlow, E. T., Hu, Y., & Ho, T. H. (2004). A learning-based model for imputing missing levels in partial conjoint profiles. *Journal of Marketing Research, November*, (XLI), 369–381.
Chintagunta, P. K. (1992). Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science*, *11*, 386–407.
Chintagunta, P. K. (2002). Investigating category pricing behavior at a retail chain. *Journal of Marketing Research*, *39*(2), 141–154.
Elrod, T., & Keane, M. P. (1995). A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, *32*(1), Feb., 1–16.
Gordon, M., & Lenk, P. (1991). A utility theoretic examination of the probability ranking principle in information retrieval. *Journal of the American Society for Information Science*, *42*, 703–714.
Gordon, M., & Lenk, P. (1992). When is the probability ranking principle sub–Optimal? *Journal of the American Society for Information Science*, *43*, 1–14.
Haaijer, R., Kamakura, W. A., & Wedel, M. (2000). The information content of response latencies in conjoint choice experiments. *Journal of Marketing Research*, *37*(3), 376–382.
Haaijer, R., & Wedel, M. (2002). Conjoint choice experiments: General characteristics and alternative model specifications. In A. Gustafsson, A. Herrmann, and F. Huber (Eds.), *Conjoint measurement* (pp. 317–360). Berlin: Springer.
Haaijer, R., Wedel M., Vriens, M., & Wansbeek, T. (1998). Utility covariances and context effects in conjoint MNP models. *Haaijer, Wedel, Vriens, and Wasbeek, Marketing Science*, *17*(3), 236–252.
Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *46*, 403–426.
Imai, K., & van Dyk, D. A. (2005). A bayesian analysis of the multinomial probit model using marginal data argumentation. *Journal of Econometrics*, *124*, 311–334.
Labaw, P. J. (1980). *Advanced questionnaire design*. Cambridge, MA: Abt Books.
Lenk, P., & DeSarbo, W. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, *65*(1), 93–119.
Lenk, P., DeSarbo, W., Green, P., & Young, M. (1996). Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, *15*(2), 173–191.
Little, R. J. A., & Rubin, D. (2002). *Statistical analysis with missing data*, Second Edition. New Jersey: John Wiley and Sons.
McCulloch, R., & Rossi, P. (1994). An exalt likelihood analysis of the multinomial probit model. *Journal of Econometrics*, *64*, 207–240.
McCulloch, R., Polson, N. G., & Rossi, P. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, *99*, 173–193.
Rossi, P., McCulloch, R., & Allenby, G. (1996). On the value of household purchase information in target marketing. *Marketing Science*, *15*, 321–340.
Smith, M. D., & Brynjolfsson, E. (2001). Consumer decision-making at an internet shopbot: Brand still matters. *Journal of Industrial Economics*, *49*(4), 541–558.
Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *81*, 82–86.
Zellner, A. (1971). *An introduction to bayesian inference in econometrics*. New York: John Wiley & Sons.