

Bayesian Path Analysis

Peter Lenk

The University of Michigan Business School

701 Tappan St.

Ann Arbor, MI 48109-1234

734-936-2619; email: plenk@umich.edu

March 2000

Researchers in the social sciences often formulate theoretical models based on constructs that are not directly measured. For instance, marketing researchers hypothesize a psychological process by which customers develop satisfaction for products and services. These theoretical constructs are not directly observed and must be inferred through indirect measurements. Path analysis or LISREL refers to a class of latent variable models that consist of two parts: measurement and structural models. The measurement model relates the observed or manifest variables to their unobserved latent variables. The structural model specifies the relation among the latent variables. The two parts imply a parametric model for the covariance of the manifest variables. Traditional inference estimates the parameters so that the fitted and observed covariances are close, either via maximum likelihood or least squares. This paper proposes Bayesian inference for path analysis and extends the traditional model by including covariates and interactions among the latent variables in the structural model. This extension results in nonlinear structural equations. Bayesian inference also accommodates ordinal manifest variables and imputes the values of the latent variables.

KEY WORDS: Latent variables; LISREL; Ordinal data; Structural equations.

1. Introduction

Social scientist widely utilize path analysis or LISREL, the mnemonic for “linear structural relationships” and Jöreskog and Sörbom’s software. A search of the Institute for Scientific Information’s indices from 1970 to 2000 returned 2463 citations. Latent variable models have a long history, starting with Spearman’s (1904) pioneering work with factor analysis and Wright’s (1918) introduction of path analysis. Jöreskog (1973), Keesing (1972), and Wiley (1973) initiated the modern era of LISREL by synthesizing latent factor models with structural equations. Bollen (1989) and Everitt (1984) provide extensive reviews of latent variable modeling.

These models enable researchers to estimate relations among theoretical constructs even though these constructs can only be indirectly observed. For instance, a questionnaire may require subjects to respond to items on a 7-point Likert scale. Multiple items are designed to probe different aspects of a theoretical construct, and blocks of items are associated with different constructs. The data often consist of a large number of moderately correlated manifest variables, which tend to have higher correlations within blocks than between blocks.

This paper develops Bayesian inference for path analysis. Bartholomew (1994) proposed utilizing Bayesian inference, and Ansari, Jedidi, and Jagpal (1999) implemented a hierarchical Bayes procedure for the case where subjects respond to the same survey instrument at multiple time points. The main difference in Ansari, Jedidi, and Jagpal’s and this paper’s model is in the identifying constraints: they identify the model by setting some of the factor loadings in the measurement model to one, while this paper uses moment constraints on the latent variables. The moment conditions imply constraints on the parameters of the structural model.

Bayesian path analysis provides several extensions to LISREL. The traditional structural model is linear in the latent variables so that the unconditional covariance matrix of

the manifest variables is a closed-form function of the model parameters. The structural model of this paper includes covariates and interactions among both the latent variables and covariates. This modification implies nonlinear structural models, and the unconditional covariance matrix does not have a closed-form. Additionally, the manifest variables frequently are measured on ordinal scales, while maximum likelihood LISREL assumes multivariate normal distributions. Bayesian analysis accommodates ordinal data with cutpoint models (Bradlow and Zaslavsky 1999, Gelfand, Smith and Lee 1992, and Johnson 1996). The Bayesian analysis also imputes the values of the latent variables as part of estimation procedure, thus explicitly accounting for parameter uncertainty. In contrast, LISREL requires a two stage method that ignores the uncertainty in the model parameters when imputing the latent variables.

The path model consists of two components. Suppose that n subjects respond to m manifest variables, which are measured on ratio scales. The modification for ordinal variables will be presented after the model specification. The manifest variables are grouped into J blocks with block j consisting of m_j manifest variables. The variables in block j are jointly determined by a latent variable according to the measurement model or outer relation:

$$U_j = \mathbf{1}_n \mu'_j + Y_j \alpha'_j + \Delta_j \text{ for } j = 1, \dots, J \quad (1)$$

where U_j is a $n \times m_j$ matrix of observations; $\mathbf{1}_n$ is a n vector of ones; μ_j is a m_j vector of intercepts; Y_j is a n vector of latent variables; α_j is a m_j vector of loadings, and Δ_j is a $n \times m_j$ matrix of error terms. The rows of the error term are mutually independent and normally distributed.

The measurement model for all of the variables can be written by concatenating the J blocks of manifest variables. Define $U = [U_1 \dots U_J]$; $\mu' = (\mu'_1, \dots, \mu'_J)$; $Y = [Y_1 \dots Y_J]$; and $\Delta = [\Delta_1 \dots \Delta_J]$. Then the outer model becomes $U = \mathbf{1}_n \mu' + Y A + \Delta$ where A is a $J \times m$ matrix. Row j of A is $(0 \dots 0 \alpha'_j 0 \dots 0)$. The error variance, $\text{var}[\text{vec}(\Delta')]$, is

$\mathbf{I}_n \otimes \Sigma$ where “vec” forms a vector by stacking the columns of a matrix; \mathbf{I}_n is a $n \times n$ identity matrix; “ \otimes ” is the Kronecker product, and Σ is a $m \times m$ positive definite matrix. I will assume that after a permutation of the columns of \mathbf{U} , the permuted columns of Σ will have a block-diagonal structure with K blocks. Let \mathbf{D} be the $m \times m$ permutation matrix such that $\mathbf{D}'\Sigma\mathbf{D}$ is block diagonal with $\{\Sigma_k\}$ along the blocks.

The latent variables follow an inner relation or structural model that is specified through a series of regression models:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j \text{ for } j = 1, \dots, J \quad (2)$$

where \mathbf{X}_j is a $n \times p_j$ design matrix; $\boldsymbol{\beta}_j$ is a p_j vector of regression coefficients; and $\boldsymbol{\epsilon}_j$ is a n vector of error terms. The error terms are a random sample from a normal distribution with zero mean and standard deviation τ_j . The design matrix \mathbf{X}_j can consist of other latent variables \mathbf{Y}_k for $k \neq j$, covariates, and interactions among the latent variables or between the latent variables and the covariates. Equation (2) specifies the distribution of \mathbf{Y}_j given $\mathbf{Y}_{(j)}$, which is \mathbf{Y} without column j . If \mathbf{Y}_j does not have a regression model, then \mathbf{Y}_j had a normal distribution with zero mean and variance matrix \mathbf{I}_n .

The model in (1) and (2) is not identified, as can be seen by multiplying $\boldsymbol{\alpha}_j$ and dividing \mathbf{Y}_j by a constant. One method of identifying the model is to assume that a component of each $\boldsymbol{\alpha}_j$ is one, which implies that the corresponding manifest variable is \mathbf{Y}_j observed with error. In so far as this manifest variable matches the latent construct that the researcher is attempting to measure, this identifying assumption is very sensible and greatly simplifies the analysis because the full conditionals are standard distributions (Ansari, Jedidi, and Jagpal 1999). One limitation is that the imputed latent variables depend on the loading that is set to one.

This paper identifies the model by constraining the first two moments: $n^{-1}E(\mathbf{1}'_n \mathbf{Y}_j | \mathbf{Y}_{(j)}, \boldsymbol{\beta}_j, \tau_j^2) = 0$ and $n^{-1}E(\mathbf{Y}'_j \mathbf{Y}_j | \mathbf{Y}_{(j)}, \boldsymbol{\beta}_j, \tau_j^2) = 1$. These moment constraints result in imputed latent

variables that have the same scale. The constraint on the first moment can be easily realized by mean-centering the design matrix \mathbf{X}_j and setting the intercept to zero. The constraint on the second moment implies: $n^{-1}\boldsymbol{\beta}'_j\mathbf{X}'_j\mathbf{X}_j\boldsymbol{\beta} + \tau_j^2 = 1$. This constraint is built into the joint distribution of $\{\mathbf{Y}_j, \boldsymbol{\beta}_j, \tau_j^2\}$, which is implicitly specified through J conditional distributions (Gelfand and Smith 1990):

$$[\mathbf{Y}_j|\mathbf{Y}_{(j)}, \boldsymbol{\beta}_j, \tau_j^2][\boldsymbol{\beta}_j|\mathbf{Y}_{(j)}, \tau_j^2][\tau_j^2] \propto N_n(\mathbf{Y}_j|\mathbf{X}_j\boldsymbol{\beta}_j, \tau_j^2 I) \\ N_{p_j}(\boldsymbol{\beta}_j|\mathbf{b}_{j,0}, \tau_j^2 B_{j,0})\chi\left(\frac{1}{n}\boldsymbol{\beta}'_j\mathbf{X}'_j\mathbf{X}_j\boldsymbol{\beta}_j = 1 - \tau_j^2\right) IG(\tau_j^2|r_{j,0}, s_{j,0})\chi(\tau_j^2 \leq 1).$$

$N_q(\cdot|\boldsymbol{\xi}, \Lambda)$ is the q -variate normal density with mean $\boldsymbol{\xi}$ and variance Λ ; $\chi(\cdot)$ is the indicator function, and $IG(\tau_j^2|r_{j,0}, s_{j,0}) \propto \tau_j^{-r_{j,0}-2} \exp(-0.5s_{j,0}/\tau_j^2)$ is the inverted gamma density.

An additional restriction is to assume that one or more components of each loading vector $\boldsymbol{\alpha}_j$ are positive. If some of the manifest variables in \mathbf{U}_j are scaled so that their correlations with \mathbf{Y}_j have the same sign, then the corresponding loadings can be forced to be positive without overly restricting the model. In fact, doing so is a sensible use of prior information.

The parameters of the outer model have the following prior distributions. The outer model's means $\boldsymbol{\mu}$ are normally distributed: $N_{m_j}(\boldsymbol{\mu}|\mathbf{v}_0, V_0)$. The loadings $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_J)'$ have truncated normal distributions: $N_m(\boldsymbol{\alpha}|\mathbf{v}_0, \mathbf{A}_0)\chi(\boldsymbol{\alpha}^+ > 0)$ where $\boldsymbol{\alpha}^+$ is the positive components of $\boldsymbol{\alpha}$. The prior for each $d_k \times d_k$, positive definite $\boldsymbol{\Sigma}_k$ is the inverted Wishart distribution: $IW_{d_k}(\boldsymbol{\Sigma}_k|r_{U,k,0}, \mathbf{S}_{U,k,0}) \propto |\boldsymbol{\Sigma}_k|^{-(d_k+r_{U,k,0}+1)/2} \exp[-0.5\text{tr}(\boldsymbol{\Sigma}_k^{-1}\mathbf{S}_{U,k,0})]$ for $r_{U,k,0} > d_k \geq 1$ and positive definite matrix $\mathbf{S}_{U,k,0}$.

A cutpoint model is used with ordinal variables: subject i selects category t if a latent variable falls between two cutpoints. Suppose that sets of ordinal variables share the same cutpoints. Define \mathbf{W}^k to be the $n \times j_k$ matrix of manifest variables that use the same cutpoints for a T_k -point scale. If ordinal manifest variable \mathbf{W}_κ belongs to \mathbf{W}^k , then subject i responds $w_{i,\kappa}$ if $c_k(w_{i,\kappa} - 1) < u_{i,\kappa} \leq c_k(w_{i,\kappa})$ for the cutpoints: $-\infty = c_k(0) < \dots < c_k(T_k)$

$= \infty$. U_κ is a latent variable that follows the outer model, and W_κ is the ordinal manifest variable.

There is a trade-off between the mean and variance of U_κ and the location and spread of the cutpoints. This paper identifies the model by setting the mean μ_κ for U_κ in the outer model to zero, and by using a two-point prior for the last cutpoint: $P[c_k(T_k - 1) = a_k] = 1 - \pi$ and $P[c_k(T_k - 1) = b_k] = \pi$ where $a_k < 0 \leq b_k$ and $0 \leq \pi \leq 1$ are known constants. The two-point prior allows the probability of the last category to be greater than or less than 0.5. The prior distribution for the remainder of cutpoints given the last one is proportional to $\exp[\zeta_k c_k(1)] \chi[c_k(1) < \dots < c_k(T_k - 1)]$ for $\zeta_k > 0$, which is the product of truncated exponential and uniform distributions.

The remainder of the paper is organized as follows. Section 2 briefly presents the Markov chain Monte Carlo (MCMC) algorithm (Gelfand and Smith 1990). Most of the full conditional distributions are standard computations, except for the identifying constraints. Section 3 analyzes synthetic data and demonstrates that the procedure accurately recovers the parameters and latent variables, and Section 4 analyzes a customer satisfaction survey.

2. MCMC

U has a matrix normal distribution $N_{n \times m}(U | \mathbf{1}_n \mu' + Y A, I_n, \Sigma)$:

$$\begin{aligned} &\propto |\Sigma|^{-n/2} \exp \left\{ \frac{1}{2} \text{tr} \left[\Sigma^{-1} (U - \mathbf{1}_n \mu' - Y A)' I_n (U - \mathbf{1}_n \mu' - Y A) \right] \right\} \\ &\propto |\Sigma|^{-n/2} \exp \left\{ \frac{1}{2} [\text{vec}(U') - \mathbf{1}_n \otimes \mu - \text{vec}(A' Y')]' \right. \\ &\quad \left. (I_n \otimes \Sigma^{-1}) [\text{vec}(U') - \mathbf{1}_n \otimes \mu - \text{vec}(A' Y')] \right\}. \end{aligned}$$

W^k signifies the $n \times j_k$ matrix of ordinal manifest variables that are bonded to a common vector of cutpoints c_k with T_k categories. The last cutpoint takes one of two values $a_k < 0 \leq b_k$. Initialize the Markov chain by setting the last cutpoint to b_k if the proportion

of observations in the last category is less than 0.5, else to a_k . After assigning values to the remaining cutpoints, initialize U^k to the intervals between cutpoints according to their ordinal responses W^k . This initialization improves the performance of the algorithm by starting in a region such that the model and data are consistent.

If the ordinal manifest variable W_κ is contained in the matrix W^k , then generate U_κ conditional on $U_{(\kappa)}$, the $n \times m - 1$ matrix with column κ removed. Set $E(U) = \Theta = \mathbf{1}_n \mu' + Y A$. Partition Θ and Σ according to U_κ and $U_{(\kappa)}$: $E(U_\kappa) = \Theta_\kappa$, $E[U_{(\kappa)}] = \Theta_{(\kappa)}$, $\text{var}(U_\kappa) = \sigma_{\kappa,\kappa} I_n$, $\text{var}[U_{(\kappa)}] = I_n \otimes \Sigma_{(\kappa,\kappa)}$, and $\text{cov}[U_{(\kappa)}, U_\kappa] = I_n \otimes \Sigma_{(\kappa),\kappa}$. The conditional mean and variance of U_κ given $U_{(\kappa)}$ are $\Theta_{\kappa|(\kappa)} = \Theta_\kappa + [U_{(\kappa)} - \Theta_{(\kappa)}][I_n \otimes (\Sigma_{(\kappa,\kappa)}^{-1} \Sigma_{(\kappa),\kappa})]$ and $\sigma_{\kappa|(\kappa)} I_n = [\sigma_{\kappa,\kappa} - \Sigma_{\kappa,(\kappa)} \Sigma_{(\kappa,\kappa)}^{-1} \Sigma_{(\kappa),\kappa}] I_n$. Generate U_κ from $N_n(U_\kappa | \Theta_{\kappa|(\kappa)}, \sigma_{\kappa|(\kappa)} I_n)$ subject to $\prod_{i=1}^n \chi[c_k(w_{i,\kappa} - 1) < u_{i,\kappa} \leq c_k(w_{i,\kappa})]$. One method of generating truncated normals is with the inverse cdf transform (Devroye 1986 and Gelfand, Smith, and Lee 1992)

Define $v_t = \min\{U^k : W^k = t\}$ and $\Upsilon_t = \max\{U^k : W^k = t\}$. The last cutpoint $c_k(T_k - 1)$ is a_k with probability proportional to $\chi(\Upsilon_{T_k-1} < a_k \leq v_{T_k}) (1 - \pi)$ and b_k with probability proportional to $\chi(\Upsilon_{T_k-1} < b_k \leq v_{T_k}) \pi$. The full conditional distribution of the remaining cutpoints are proportional to $\exp[\zeta_k c_k(1)] \prod_{t=1}^{T_k-2} \chi[\Upsilon_t < c_k(t) \leq v_{t+1}]$.

Generate μ from $N_m(\mu | v_n, V_n)$ with $V_n = (V_0^{-1} + n \Sigma^{-1})^{-1}$ and $v_n = V_n [V_0^{-1} v_0 + \Sigma^{-1}(U' - A'Y') \mathbf{1}_n]$. If the j th manifest variable is ordinal, set the j th component of μ to zero.

Let $y_{i,j}$ be the value for latent variable j and subject i . Define

$$B_i = \begin{bmatrix} y_{i,1} I_{m_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & y_{i,J} I_{m_J} \end{bmatrix} \text{ and } B = \begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix}.$$

Then $\text{vec}(U') = \mathbf{1}_n \otimes \mu + B \alpha + \text{vec}(\Delta)$. Generate α from $N(\alpha | v_n, A_n)$ constrained to $\chi(\alpha^+ > 0)$ where $A_n = [A_0^{-1} + B' (I_n \otimes \Sigma^{-1}) B]^{-1}$, and $v_n = A_n \{A_0^{-1} v_0 + B' (I_n \otimes \Sigma^{-1})\}$

$[\text{vec}(U') - \mathbf{1}_n \otimes \boldsymbol{\mu}]$. Sequentially generate each component of $\boldsymbol{\alpha}$ from a univariate normal distribution or truncated normal distribution by conditioning on the current values of the other components.

Define $SSE = D'(U - \mathbf{1}_n \boldsymbol{\mu}' - Y A)'(U - \mathbf{1}_n \boldsymbol{\mu}' - Y A)D$ where D is the $m \times m$ permutation matrix that defines the blocks of Σ . Let SSE_k be the rows and columns of SSE that corresponds to Σ_k . Generate Σ_k from $IW_{d_k}(\Sigma_k | r_{U,k,n}, S_{U,k,n})$ where $r_{U,k,n} = r_{U,k,0} + n$, and $S_{U,k,n} = S_{U,k,0} + SSE_k$.

The conditional distribution of U given Y can be expressed as $[U_j | U_{(j)}, Y] [U_{(j)} | Y_{(j)}]$. Only the conditional distribution of U_j given $U_{(j)}$ depends on Y_j . Partition the error covariances matrix Σ according to the block structure of U : $\text{var}(U_j) = I_n \otimes \Sigma_{j,j}$; $\text{cov}(U_j, U_k) = I_n \otimes \Sigma_{j,k}$; $\text{var}[U_{(j)}] = I_n \otimes \Sigma_{(j,j)}$; $\text{cov}[U_j, U_{(j)}] = I_n \otimes \Sigma_{j,(j)}$. Define $\boldsymbol{\mu}_{(j)}$ to be the vector of means $\boldsymbol{\mu}$ with $\boldsymbol{\mu}_j$ removed; and $A_{(j)}$ to be the matrix of loadings A with the row j deleted. The conditional distribution of U_j given $U_{(j)}$ and Y is $N_{n \times m_j}[U_j | Y_j \boldsymbol{\alpha}'_j + \boldsymbol{\xi}_{j|(j)}, I_n, \Sigma_{j|(j)}]$ where $\boldsymbol{\xi}_{j|(j)} = \mathbf{1}_n \boldsymbol{\mu}'_j + (U_{(j)} - \mathbf{1}_n \boldsymbol{\mu}'_{(j)} - Y_{(j)} A_{(j)}) \Sigma_{(j,j)}^{-1} \Sigma_{(j),j}$ and $\Sigma_{j|(j)} = \Sigma_{j,j} - \Sigma_{j,(j)} \Sigma_{(j,j)}^{-1} \Sigma_{(j),j}$. The full conditional of Y_j is $N_n(Y_j | \mathbf{f}_{j,n}, \mathbf{G}_{j,n})$. If X_j is not null in Equation (2), then $\mathbf{G}_{j,n} = (\boldsymbol{\alpha}'_j \Sigma_{j|(j)}^{-1} \boldsymbol{\alpha}_j + \tau_j^{-2})^{-1} I_n$ and $\mathbf{f}_{j,n} = \mathbf{G}_{j,n} [(U_j - \boldsymbol{\xi}_{j|(j)}) \Sigma_{j|(j)}^{-1} \boldsymbol{\alpha}_j + \tau_j^{-2} X_j \boldsymbol{\beta}_j]$; else $\mathbf{G}_{j,n} = (\boldsymbol{\alpha}'_j \Sigma_{j|(j)}^{-1} \boldsymbol{\alpha}_j + 1)^{-1} I_n$ and $\mathbf{f}_{j,n} = \mathbf{G}_{j,n} [(U_j - \boldsymbol{\xi}_{j|(j)}) \Sigma_{j|(j)}^{-1} \boldsymbol{\alpha}_j]$.

Given the latent variables, the prior distributions for τ_j^2 and $\boldsymbol{\beta}_j$ are the natural conjugate priors for multiple regression with the constraints that $\tau_j^2 < 1$ and $n^{-1} \boldsymbol{\beta}'_j X'_j X_j \boldsymbol{\beta}_j = 1 - \tau_j^2$. Define $\mathbf{B}_{j,n} = [\mathbf{B}_{j,0}^{-1} + X'_j X_j]^{-1}$; $\mathbf{b}_{j,n} = \mathbf{B}_{j,n} [\mathbf{B}_{j,0}^{-1} \mathbf{b}_{j,0} + X'_j Y_j]$; $r_{j,n} = r_{j,0} + n$; and $s_{j,n} = s_{j,0} + Y'_j Y_j + \mathbf{b}'_{j,0} \mathbf{B}_{j,0}^{-1} \mathbf{b}_{j,0} - \mathbf{b}'_{j,n} \mathbf{B}_{j,n}^{-1} \mathbf{b}_{j,n}$. Generate τ_j^2 from $IG(\tau_j^2 | r_{j,n}, s_{j,n}) \chi(\tau_j^2 \leq 1)$, a truncated inverted gamma distribution, and then generate $\boldsymbol{\beta}_j$ from $N_{p_j}(\boldsymbol{\beta}_j | \mathbf{b}_{n,j}, \tau_j^2 \mathbf{B}_{n,j}) \chi(\boldsymbol{\beta}'_j \mathbf{Q} \boldsymbol{\beta}_j = 1 - \tau^2)$ where $\mathbf{Q} = n^{-1} X'_j X_j$. Define $\mathbf{Z} = \mathbf{Q}^{\frac{1}{2}} \boldsymbol{\beta}_j$ where $\mathbf{Q}^{\frac{1}{2}}$ is the Cholesky decomposition of \mathbf{Q} . The constraint becomes $\mathbf{Z}' \mathbf{Z} = 1 - \tau^2$. \mathbf{Z} is generated one component at a time by conditioning on the previous draws. Then each component, except the last, is a

truncated normal distribution. The last component Z_{p_j} can have one of two values: $\pm d$ for $d = \left(1 - \tau_j^2 - z_1^2 - \dots - z_{p_j-1}^2\right)^{\frac{1}{2}}$ with probabilities proportional to $\exp[-(\pm d - \psi_{p_j|p_j-1})^2 / (2\omega_{p_j|p_j-1}^2)]$ where $\psi_{p_j|p_j-1}$ and $\omega_{p_j|p_j-1}^2$ are the conditional mean and variance of the unconstrained normal Z_{p_j} given the previous components. Set $\beta_j = Q^{-\frac{1}{2}}Z$.

3. Simulated Example

The simulation consists of 500 “subjects,” 12 manifest variables, four latent variables, and two covariates C1 and C2. Table 1 gives the outer model and the manifest variable’s scales. All variables that are measured on the 3–point scale share the same set of cutpoints, as do all variables with the 5–point scale. W01 and W02 are the manifest variables for Y1; W03, U04, and W05 for Y2; W06, W07, U08, and U09 for Y3; and W10, W11, and U12 for Y4. All of the error covariances in the outer model are zero except for $\text{cov}(W03, W10)$, $\text{cov}(U04, W11)$, and $\text{cov}(W05, U12)$. This covariance structure occurs in applications when a panel of subjects respond at two points in time (Sörbom 1979).

The MCMC algorithm had 100,000 iterations: 50,000 for the initial transition period, and every tenth of the last 50,000 for estimation. Plots of the random draws versus iteration indicated that the chain converged to its stationary distribution. Independent chains with shorter runs had nearly the same results. Table 1 reports the true parameters and posterior means and standard deviations for the outer model; Table 2 reports the cutpoints, and Table 3 reports the inner model. These tables indicate that the Bayesian analysis accurately recovers the true parameter values: the posterior standard deviations are small, and the posterior means are within one or two posterior standard deviations of the true parameters.

A standard measure of fit in LISREL is the root mean squared error between the observed and fitted correlations of the manifest variables. On each iteration of the MCMC, predictive values for the manifest variables were generated given the current parameter values. The

correlation matrix was then computed for the these predicted values and averaged over MCMC iterations. The observed and estimated correlations are within one or two posterior standard deviations, and the RMSE between the observed and estimated correlations is 0.038.

The imputed latent variables were fairly accurate. The correlation and RMSE between the actual and imputed latent variables are: 0.97 and 0.23 for Y1, 0.97 and 0.25 for Y2, 0.97 and 0.25 for Y3, and 0.95 and 0.33 for Y4. Table 4 demonstrates that the correlations among the imputed latent variables accurately estimate the correlations among the known, but unobserved, variables from the simulation.

4. ACSI Survey

The American Customer Satisfaction Index (ACSI) survey is conducted by the National Quality Research Center, The University of Michigan Business School. Fornell *et al* (1996) describes the survey methodology and presents the theoretical basis for the model, which is shown in Figure 1. Customers' Expectations, perceived Quality, and perceived Value are antecedents to customers' Satisfaction, and consequences of Satisfaction are Complaints and Loyalty.

One hundred and seventy five customers rated their experience with a local telephone carrier by answering 15 items on a questionnaire. Respondents were asked what their expectations were prior to purchase, i.e. their *ex ante* expectations. Three expectation measures are overall expectation, the expectation of the provider's ability to customize the service to meet the customer's personal requirements, and expectation of service reliability. Three quality measures are post-purchase evaluations of perceived quality: overall quality, ability to meet personal requirements, and reliability. Two value measures rated the quality of the service relative to its price and the price relative to quality. Three satisfaction measures

are overall satisfaction, the difference between service performance and expectations, and the difference between actual and ideal service. Two complaint measures are the number of formal and informal complaints to the service provider. The first loyalty measure is the likelihood of repurchasing the service, and the second measures the customer's price tolerance. If a customer discontinued the service, price tolerance is the percentage that the price would have to drop to induce repurchase. If a customer continues the service, it is the percentage that the price would have to increase before cancelling the service. This item could be viewed as a respondent's reported, not actual, reservation price, the highest price that he or she is willing to pay for the service.

The variables used in the analysis have been transformed. All of the variables except the two complaint measures and price tolerance were measured on a 10-point ordinal scale. Fornell *et al* (1996) justify the 10-point scale "to reduce the statistical problems of extreme skewness" with partial least squares and "to allow customers to make better discriminations." PLS is designed for ratio data, and the 10-point scale attempts to satisfy this assumption. The data exhibited considerable skewness. Among the 13 ordinal variables, the modal response was 10 with 37% of all responses, 11% of the responses were nine, 16% were eight, and 9% were seven. Less than 15% of the responses were less than five. Skewed ordinal data does not present a problem when using the cutpoint procedure. However, with so few observations in the first four categories, the data are not very informative about the lower cutpoints. I reduced the 10-point scale to six points by grouping categories one to four and combining five and six.

The number of formal and informal complaints are highly skewed with means of 1.0 and 1.2, standard deviations of 2.5 and 4.7, and maxima of 20 and 52. Seventy two percent and 75% of the respondents reported zero complaints, and 17% and 15% reported one or two complaints. I recoded the complaints to a three points scale: one for no complaints, two

for one or two complaints, and three for three or more complaints. Price tolerance ranged from -50% to 90% with a mean of 19% and a standard deviation of 25% . Five percent of the respondents gave negative numbers, and no respondents reported 0% . Fourteen percent indicated that a 1% price increase would be sufficient for them to cancel their phone service. All of the responses after 10% were stated in increments of 5% . The lumpiness of the responses implies that the scale is not truly ratio. I converted it to a five point scale by coding negative responses as one; 1% to 10% as two; 15% and 20% as three; 25% , 30% , 25% , and 40% as four, and 45% or more as five.

In addition to the transformed data, I analyzed the original data by assuming that all of the variables are on ratio scales and by using the original 10-point ordinal scale and ratio scales for complaints and price tolerance. The outer model depends on the scaling of the variables, but the estimated inner relations were remarkably robust with most of the estimated parameters for different analyses within $\pm 5\%$ of each other. Using the 10-point scale seemed to increase the amount of autocorrelation within the Markov chain.

Bryant and Cha (1996) found evidence of an impact of demographics on ACSI. They used a two-stage method: first, they computed individual ACSI scores, which is the PLS imputed satisfaction, and then averaged these scores by industry and demographic segment. They found that gender, age, socioeconomic status, and urban/rural effects across 40 industries.

The following analysis uses five covariates. Age is the respondents age in years, which ranged from 18 to 83 years with a mean of 41 and standard deviation of 15. College is a 0/1 variable that indicates whether the respondent attended college, with or without receiving a degree: 64% of the sample reported some college. Rich is a 0/1 variable that indicates whether the respondent's income was above $\$50,000$ per year; 25% of the sample reported incomes above $\$50,000$. Male and Minority are 0/1 variables that indicates whether the respondent reported as being male or as being African-American or Hispanic. Thirty four

percent of the sample are male, and 20% are minority.

I augmented the basic ACSI model in Figure 1 with demographic variables and interactions. There is evidence that not only does Expectation influence Satisfaction, but also that Satisfaction strongly effects Expectation. One reason may be that all of the items on the survey were jointly determine. Subjects were asked to report their *ex ante* expectations at the same time that they reported their *post hoc* perceptions of quality, value, and satisfaction. The reported expectations may not be purely *ex ante*.

The MCMC algorithm ran for 100,000 iterations. The first 50,000 iterations were discarded, and every tenth iteration from the remaining 50,000 were used for estimation. Independent runs using shorter chains resulted in estimates similar to those reported in the following tables. Table 5 reports the outer model for the augmented ACSI model. For manifest variables that use the same cutpoints, the loadings and error standard deviations can be compared to each other. The predictive correlation squared is the correlation squared between the observed and predicted manifest variables. The algorithm generated predicted values of U on each MCMC iteration according to the outer model and used the cutpoints to create ordinal variables. Correlations between these predicted and the manifest variables were squared and averaged over iterations. The predicted correlation squared is one indicator of model fit. Loyalty 2, which is price tolerance, has the smallest measure of fit; and Loyalty 1, which is likelihood of repurchase, has the largest.

Table 6 gives the inner model for the both the basic and augmented ACSI model. Because the error standard deviations are forced to be less than or equal to one, their magnitude provides some indication of model fit. A pseudo R-squared is one minus the error variance. For both models Satisfaction has the largest R-squared of 77% and 84%, and Complaints has the smallest of 27% and 32%.

The estimated basic ACSI broadly supports with the hypothesized model, except that the

effects of Expectation on Satisfaction and Complaints on Loyalty are negligible. In the basic ACSI model, Expectation does not have a regression relation, so its variance is set to one. The effects of Expectation on Quality, Value, and Satisfaction are larger in the augmented ACSI than the basic ACSI. In the augmented ACSI model, the coefficient for Satisfaction on Expectation is large, both absolutely and also relative to its posterior standard deviation. There is moderate evidence that older customers have higher expectations and that college education reduces perceived quality. Expectation and Quality strongly effect Value, with the more affluent having higher perceived value. This result may be due to the wealthy either being less price sensitive or using a greater degree of customized services. Expectation, Quality and Value are strong determinants of Satisfaction with some evidence of a negative interaction between Quality and Value. As predicted by the ACSI model, satisfied customers tend to complain less frequently. Perhaps more surprising, at least to men if not women, is that for a given level of satisfaction, men tend to complain more frequently than women. Loyalty is positively related to Satisfaction with minorities being less Loyal for a given level of Satisfaction. The RMSE of the residual correlation matrix is 0.056 for the augmented ACSI, which is 27% less than the RMSE (0.077) for the basic ACSI.

5. Discussion

An alternative estimation method for latent variable models is partial least squares (Frank and Friedman 1993 and Wold 1966 and 1989). PLS uses projective geometry and fixed-point methods to iteratively estimate linear functions of the manifest variables that have maximal predictive correlation. Fornell and Bookstein (1982) and Fornell and Cha (1994) contrast LISREL and PLS. Goutis (1996) shows that PLS estimators shrink ordinary least squares estimates. The Bayesian imputation of the latent variables use a similar idea: the means of the full conditionals of the latent variables are linear functions of the manifest variables.

Decomposing complex, joint distributions by conditioning on unknown quantities or “latent variables” has a long history in Bayesian inference.

De Finetti’s representation (De Finetti 1930 cf. Bernardo and Smith 1994) for infinitely exchangeable sequences can be viewed as a latent variable model. Tanner and Wong (1987) used data augmentation to simplify complex posterior distributions; and Damien, Wakefield, and Walker (1999) push the concept to its natural extreme. In this sense, latent variables are not new to Bayesian inference. This paper continues this tradition to analyze covariance matrices that have complex parameterizations.

An open question concerns the specification of the structural model. Traditionally, it has been based on theory, while including covariates and interactions may be driven by empirical analysis. What is the effect of adding covariates or interactions to different parts of the inner model? In the ACSI survey, is it better to introduce Age in the equation for Expectation so that Expectation is a mediating variable for the effect of Age on Quality and Value or in the equations for Quality and Value so that Expectation is a moderating variable? Bayes factors can be used to provide a statistical answer, but what are the substantive implications?

Acknowledgement: I wish to thank Gene Anderson and the National Quality Research Center for providing the ACSI survey data.

References

- Ansari, Asim, Kamel Jedidi, and Sharan Jagpal (1999), "A General Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models," working paper, Columbia University.
- Bernardo, José, and Adrian F. M. Smith (1994), *Bayesian Theory*, John Wiley & Sons, New York, 172–181.
- Bradlow, Eric T. and Alan M. Zaslavsky (1999) "A Hierarchical Latent Variable Model for Ordinal Data from a Customer Satisfaction Survey with 'No Answer' Responses," *J. Amer. Stat. Assoc.*, **94**, 43–52.
- Bartholomew, D. J. (1994), "Bayes' Theorem in Latent Variable Modelling," in *Aspects of Uncertainty*, editors P. R. Freeman and A. F. M. Smith, John Wiley & Sons, New York, 41–50.
- Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, John Wiley & Sons, New York.
- Damien, P., Wakefield, J. C., and Walker, S. (1999). Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables. *J. Roy. Stat. Soc. Ser. B*, **61**, 331–344.
- De Finetti, B. (1930), "Funzione caratteristica di un fenomeno aleatorio," *Mem. Acad. Naz. Lincei*, **4**, 86–133.
- Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- Bryant, Barbara Everitt and Jaesung Cha (1996), "Crossing the Threshold," *Marketing Research*, Winter, 21–28.

- Everitt, B. S. (1984), *An Introduction to Latent Variable Models*, Chapman and Hall, London.
- Frank, Ildiko E. and Jerome H. Friedman (1993), "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, **35**, 109–135.
- Fornell, Claes and Fred L. Bookstein (1982), "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research*, Vol. XIX (November), 440-52.
- Fornell, Claes and Jaesung Cha (1994), "Partial Least Squares," in *Advanced Methods in Marketing*, editor Richard Bagozzi, Blackwell Business, Cambridge, 52–78.
- Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant (1996), "The American Customer Satisfaction Index: Nature, Purpose, and Findings," National Quality Research, University of Michigan Business School.
- Gelfand, Alan E. and Adrian F. M. Smith (1990), "Sampling-based approaches to calculating marginal densities," *J. Amer. Stat. Assoc.* **85**, 398–409.
- Gelfand, Alan E., Adrian F. M. Smith, and Tai-Ming Lee (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems using Gibbs Sampling," *J. Amer. Stat. Assoc.*, **87**, 523–532.
- Goutis, Constantinos (1996), "Partial Least Squares Algorithm Yields Shrinkage Estimators," *Ann. Stat.*, **24**, 816–824.
- Johnson, Valen E. (1996), "On Bayesian Analysis of Multivariate Ordinal Data: An Application to Automated Essay Grading," *J. Amer. Stat. Assoc.*, **91**, 42–51.
- Jöreskog, Karl G. (1973), "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, eds. A. S. Goldberg and O. D. Duncan, Sminar Press, New York, 85–112.

- Keesing, W. (1972), *Maximum Likelihood Approaches to Causal Flow Analysis*, Ph.D. Thesis, University of Chicago.
- Sörbom, Dag (1979), "Detection of Correlated Errors in Longitudinal Data," *Advances in Factor Analysis and Structural Equation Models*, eds. K. G. Jöreskog and D. Sörbom, Abt Associates, Cambridge, MA, 171–184.
- Spearman, C. (1904), "'General Intelligence' Objectively Determined and Measured," *American Journal of Psychology*, **15**, 201–293.
- Tanner, M. A., and W. H. Wong (1987), "The Calculation of Posterior Distribution by Data Augmentation," *J. Amer. Stat. Assoc.*, **82**, 542–543.
- Wiley, D. E. (1973), "The Identification Problem for Structural Equation Models with Unmeasured Variables," in *Structural Equation Models in the Social Sciences*, eds. A. S. Goldberg and O. D. Duncan, Sminar Press, New York, 69–83.
- Wold, Herman (1966), "Nonlinear Estimation by Iterative Least Squares Procedures," in *Research Papers in Statistics. Festschrift for Jerzy Neunam*, John Wiley & Sons, New York, 411–444.
- Wold, Herman (1989), *Theoretical Empiricism: A General Rationale for Scientific Model Building*, Paragon House, New York.
- Wright, S. (1918), "On the Nature of Size Effects," *Genetics*, **3**, 367–374.

Table 1. Outer model for the simulated data. The estimates are the posterior means, and the numbers in parentheses are the posterior standard deviations.

Manifest Variable	Scale	Means		Loadings		Error STD	
		True	Estimate	True	Estimate	True	Estimate
W01	3-point			0.546	0.558 (0.034)	0.447	0.414 (0.027)
W02	3-point			0.890	0.913 (0.063)	0.707	0.669 (0.048)
W03	5-point			0.981	0.946 (0.065)	0.548	0.539 (0.035)
U04	ratio	-0.500	-0.555 (0.044)	0.862	0.762 (0.107)	1.000	0.946 (0.037)
W05	5-point			0.797	0.741 (0.057)	0.707	0.662 (0.031)
W06	3-point			0.807	0.845 (0.109)	0.894	0.950 (0.065)
W07	5-point			0.649	0.527 (0.085)	1.225	1.267 (0.065)
U08	ratio	4.476	4.461 (0.018)	0.612	0.613 (0.026)	0.316	0.341 (0.022)
U09	ratio	3.120	3.118 (0.065)	0.611	0.545 (0.069)	1.414	1.412 (0.045)
W10	5-point			0.640	0.514 (0.102)	0.548	0.570 (0.052)
W11	5-point			0.929	0.752 (0.136)	1.000	0.989 (0.072)
U12	ratio	6.237	6.292 (0.036)	0.637	0.624 (0.178)	0.707	0.789 (0.094)
Error Covariances							
		W03 & W10		U04 & W11		W05 & U12	
		True	Estimate	True	Estimate	True	Estimate
		0.200	0.202 (0.035)	0.800	0.726 (0.078)	-0.200	-0.203 (0.038)

Table 2. Cutpoints for the ordinal scales in the simulated data.

	Five Point Scale			
True	-1.000	-0.250	0.250	1
Posterior Mean	-1.014	-0.252	0.261	1
Posterior STD	0.036	0.025	0.021	0
	Three Point Scale			
True	0.000	1		
Posterior Mean	0.022	1		
Posterior STD	0.026	0		

Table 3. Inner model for the simulated data. The estimates are the posterior means, and the numbers in parentheses are the corresponding posterior standard deviations.

	Y1		Y2		Y3		Y4	
	True	Estimate	True	Estimate	True	Estimate	True	Estimate
Y1			0.200	0.211 (0.020)	-0.280	-0.236 (0.038)		
Y2					0.150	0.118 (0.035)		
Y3							0.050	0.009 (0.052)
Y1×Y2					-0.446	-0.482 (0.025)		
C1	0.201	0.197 (0.003)						
C2							0.128	0.141 (0.024)
Y1×C1			0.108	0.104 (0.007)				
Y1×C2			0.137	0.126 (0.008)				
Y3×C1							-0.078	-0.055 (0.038)
Error STD	0.239	0.299 (0.040)	0.138	0.373 (0.034)	0.187	0.345 (0.041)	0.176	0.286 (0.034)

Table 4. Correlation matrix for the latent variables in the simulated data.

		Y1	Y2	Y3
Y2	True	0.233		
	Posterior Mean	0.253		
	Posterior STD	0.028		
Y3	True	-0.688	-0.150	
	Posterior Mean	-0.644	-0.187	
	Posterior STD	0.022	0.044	
Y4	True	0.076	0.661	-0.276
	Posterior Mean	0.058	0.436	-0.301
	Posterior STD	0.061	0.277	0.086

Table 5. Outer model for the augmented ACSI model.

Manifest Variable	Loadings Posterior		Error STD Posterior		Predictive Correlation Squared
	Mean	STD	Mean	STD	
Expectation 1	3.600	0.417	2.467	0.298	0.369
Expectation 2	2.758	0.305	1.953	0.262	0.374
Expectation 3	2.984	0.482	3.928	0.403	0.100
Quality 1	3.132	0.308	1.282	0.207	0.628
Quality 2	3.151	0.315	1.342	0.210	0.607
Quality 3	3.454	0.400	2.338	0.260	0.365
Value 1	4.953	0.470	1.802	0.452	0.658
Value 2	4.150	0.379	1.266	0.385	0.728
Satisfaction 1	2.847	0.211	1.029	0.167	0.711
Satisfaction 2	2.846	0.313	2.616	0.224	0.267
Satisfaction 3	3.408	0.316	2.124	0.213	0.444
Complaints 1	0.913	0.117	0.326	0.053	0.592
Complaints 2	0.805	0.100	0.320	0.049	0.546
Loyalty 1	4.188	0.467	1.283	0.670	0.745
Loyalty 2	0.644	0.107	0.878	0.104	0.057

Table 6. Inner model for the basic and augmented ACSI models.

Basic ACSI	Expectation	Quality	Value	Satisfaction	Complaints	Loyalty
Expectation		0.130 (0.063)	0.048 (0.042)	0.004 (0.027)		
Quality			0.443 (0.096)	0.415 (0.079)		
Value				0.549 (0.082)		
Satisfaction					-0.513 (0.143)	0.627 (0.088)
Complaints						-0.023 (0.120)
Error STD	1.0 (0.0)	0.803 (0.075)	0.771 (0.066)	0.483 (0.050)	0.856 (0.079)	0.758 (0.071)
Augmented ACSI	Expectation	Quality	Value	Satisfaction	Complaints	Loyalty
Expectation		0.690 (0.084)	0.429 (0.112)	0.171 (0.087)		
Quality			0.354 (0.113)	0.354 (0.083)		
Value				0.515 (0.082)		
Satisfaction	0.715 (0.068)				-0.480 (0.153)	0.614 (0.091)
Complaints						-0.025 (0.123)
Value× Quality				-0.072 (0.055)		
Age	0.010 (0.010)					
College		-0.206 (0.138)				
Rich			0.196 (0.144)			
Male× Satisfaction					0.425 (0.180)	
Minority× Satisfaction						-0.542 (0.217)
Error STD	0.625 (0.073)	0.700 (0.065)	0.688 (0.062)	0.399 (0.045)	0.825 (0.080)	0.709 (0.084)

Figure 1. The American Satisfaction Index Model

