Title

A Bayesian confirmatory factor model for multivariate observations in the form of two-way tables of data

Journal

Statistics in Medicine, 37(10)

ISSN

0277-6715

Authors

Chen, Qiaolin
Sugar, Catherine A
Weiss, Robert E

Publication Date

2018-05-10

DOI

10.1002/sim.7612

Peer reviewed

HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2019 May 10.

Published in final edited form as:

Stat Med. 2018 May 10; 37(10): 1696–1710. doi:10.1002/sim.7612.

A Bayesian Confirmatory Factor Model for Multivariate Observations in the Form of Two-Way Tables of Data

Qiaolin Chen[a,*], **Catherine A. Sugar**[a], and **Robert E. Weiss**[a]

[a]Department of Biostatistics, UCLA School of Public Health, Los Angeles, California 90095-1772, USA

Abstract

Researchers collected multiple measurements on schizophrenia (SZ) patients and their relatives, as well as control subjects and their relatives, to study vulnerability factors for schizophrenics and their near relatives. Observations across individuals from the same family are correlated, and also the multiple outcome measures on the same individuals are correlated. Traditional data analyses model outcomes separately and thus do not provide information about the interrelationships among outcomes. We propose a novel Bayesian Family Factor Model (BFFM), which extends the classical confirmatory factor analysis (CFA) model to explain the correlations among observed variables using a combination of family-member factors and outcome factors. Traditional methods for fitting CFA models, such as full information maximum likelihood (FIML) estimation using quasi-Newton optimization (QNO) can have convergence problems and Heywood cases (lack-of-convergence) caused by empirical under-identification. In contrast, modern Bayesian Markov chain Monte Carlo handles these inference problems easily. Simulations compare the BFFM to FIML-QNO in settings where the true covariance matrix is identified, close to not identified and not identified. For these settings, FIML-QNO fails to fit the data in 13%, 57% and 85% of the cases, respectively, while MCMC provides stable estimates. When both methods successfully fit the data, estimates from the BFFM have smaller variances and comparable mean squared errors. We illustrate the BFFM by analyzing data on data from schizophrenics and their family members.

Keywords

Confirmatory factor analysis; Multivariate observations; Full information maximum likelihood; Multitrait-multimethod; Schizophrenia; Structural Equation Modeling

1. Introduction

The UCLA NeuroCognitive Study (NCS) [1–3] is a cross-sectional case-controlled study. Multiple cognitive measures were collected on schizophrenia patients and also on their first-degree relatives, as well as on healthy controls and their relatives. Thus the data is essentially a two-way table of correlated observations collected on schizophrenic families, or their matched controls. This two-way table forms the basic data unit that we wish to analyze, with

*Correspondence to: Qiaolin Chen, Department of Biostatistics, UCLA School of Public Health, Los Angeles, California 90095-1772, U.S.A.. qlchen@ucla.edu.

correlations across rows induced by different measures assessed on an individual, and correlations also down columns induced by the same measure assessed on different individuals. The direct product or product-normal covariance model is often used to model two-way tables of correlated data. However, in our experience, it does not accurately model the covariances in two-way tables of data. Thus our goal is to develop an alternative covariance model for two-way table data that better models our data.

The NCS study aimed to identify the outcomes which distinguish schizophrenia patients and their relatives from healthy controls and their relatives, and by examining how these outcomes are differentially expressed among the relatives. Our goal here is to develop a novel covariance model which can help us to compare the degree of abnormality between schizophrenia families and control families, model correlations among multiple measurements from relatives and handle covariates and missing data.

Modeling data such as that from the NCS requires a complex covariance structure. A proband, a schizophrenic or matched control in this case, is an individual who triggers study of other members of the family. Suppose there are K outcome measures for each of J members in a total of N families, so that the observed data for each family, y_i, is a JK vector. In a typical family, there might be $J = 4$ family members: proband, father, mother and sibling. Families with more or fewer than J members or with no siblings can also be handled as described later in the text. It is easier to discuss concepts as if all families have the same family structure however but the algorithms and models can handle families with different structures.

Both the J family member types and the K outcome types contribute to the variation in y_i, which is summarized by a $JK \times JK$ covariance matrix. We assume that the covariances are explained by K unobserved family member factors and J unobserved outcome factors, which induce correlations on the observed measures, both across-family-member within-measure and across-measure within-family-member. As the measurements on individuals from the same family are related due to both genetic effects and unobserved environmental effects, the J family member factors are allowed to be correlated with an unstructured variance covariance matrix. Similarly, the K outcome factors are also assumed correlated because outcome measurements within subjects are correlated.

The relationships among the JK observed variables and $J + K$ factors can be described using a path diagram [4, 5]. Figure 1 shows an example of a path diagram for our two-way table of data with $J = 4$ family members and $K = 5$ outcomes drawn using AMOS [6], an add-on to IBM SPSS for structural equation modeling.

In the figure, the observed variables Y_{ijk} labeled Yjk omitting the i (rectangles in the middle), are assumed to be caused by two sets of factors, the correlated family member factors (Schizophrenic, Sibling, Father and Mother) and correlated outcome factors (Outcome1, …, Outcome5), along with residuals identified as err_jk, which are unique to each observed variable on each family member, for $j = 1, \ldots, 4$ and $k = 1, \ldots, 5$, controlled by variances unique to each observed variable, psi1, …, psi5.

Standard analyses of these sorts of studies usually model outcomes separately, which is potentially less efficient and does not provide information about the relationships among outcomes [2,7]. Approaches for joint analysis of multivariate data include linear mixed models [8, 9], structural equation modeling (SEM) [4, 10, 11] and factor analysis [12, 13, 23]. Classical analysis techniques for multiple outcomes are not designed to take into account associations among family members, which is equivalent to omitting the family member factors in Figure 1 and only considering the right half of the diagram.

Direct product models [14–18] assume $\text{var}(\boldsymbol{y}_i) = \sum_{\text{member}} \otimes \sum_{\text{outcome}}$, where Σ_{member} and Σ_{outcome} are $J \times J$ and $K \times K$ covariance matrices for the two groups of effects, respectively. However, these models are too rigid for most two-way table data as they assume all outcomes have identical correlation matrices across family members and similarly family members have identical correlation matrices across outcomes. Additionally, variance ratios of one family member to the next are the same for all outcomes and this assumption is usually not met by typical data.

Factor analysis (FA) models correlate observed variables using a smaller number of unobservable variables, called latent factors [19]. If some factors are assumed to be independent, the corresponding factor covariances are fixed to zero [13]. Confirmatory factor analysis (CFA) is used to test hypothesized relationships between observed variables and factors [20]. Researchers specify the number of factors beforehand and make *a priori* assumptions about which observed variables are related to which factors based on past evidence and theory. The factor loadings specify the pattern of relationships between the observed variables and the factors. Only loadings corresponding to hypothesized relationships between specific observed variables and factors are allowed to be nonzero. All the others, called cross-loadings, are fixed to zero. The scale of the factors can be defined by fixing factor variances to 1, or by setting the scale of a factor to be the same as one of the observed variables to which it contributes. For standard CFA, parameters are typically estimated using maximum likelihood, EM maximum likelihood or the method of moments [21–23].

The structure of data on a set of relatives with multiple outcomes is similar to that of the multitrait-multimethod (MTMM) data used for studying construct validity: the ability of psychological tests to actually measure the concept being studied [24–26]. For MTMM analysis, a certain number of traits (J) are each assessed by several methods (K) for each of N subjects, resulting in a $JK \times JK$ correlation matrix. The path diagram for an MTMM model is similar to Figure 1, replacing family members with methods and outcomes with traits. Despite the similarity in data structure, the focus of MTMM analyses is quite different from analyses of two-way multivariate data from related family members. MTMM analyses only model the correlation matrix not the mean structure, and mainly focus on estimation and tests of parameters with specific meanings for construct validity. In contrast, in two-way multivariate family data analysis, mean structures may depend on covariates and hypotheses about regression coefficients are of interest. Incomplete data is a significant issue in data collected on family members, as a family may not have all J member types and individual measures can be missing for a particular subject.

The most popular technique for fitting an MTMM model is confirmatory factor analysis (CFA) using the correlated-trait correlated-method (CTCM) structure, which assumes the inter-related trait factors are independent of the inter-related method factors [27, 28]. This model requires a total of at least $J + K \geq 6$ trait and method factors with at least $J \geq 2$ method factors and $K \geq 2$ trait factors to be identified, and it is not empirically identified when the loading matrix has deficient column rank [29], or when all trait or method factor correlations are equal [31]. Wothke [30], Brannick, et. al. [31] and Lance et. al. [32] analyzed 21, 14 and 19 published MTMM matrices, and reported that in 100%, 94% and 100% of the cases, respectively, the algorithm for CFA model failed to converge or gave invalid solutions, such as negative variances or non-positive definite covariance matrices, which are called Heywood cases [29]. The algorithm for fitting CFA models to two-way multivariate family data can have the same identification problems, resulting in non-convergence, fits with invalid solutions, improper estimates such as negative loadings, or unstable estimates with extreme standard errors.

Bayesian factor analysis (BFA) [33–37] can help to mitigate the identification problem by incorporating available knowledge about parameters in the form of prior distributions based on either expert opinions or previous experiments not to mention that typical constraints on variances mean that zero estimates do not occur. Markov chain Monte Carlo (MCMC) methodology has been applied previously in BFA to sample from posterior distributions [38–42]. Bayesian inference using inverse-gamma priors for unique error variances and inverse-Wishart priors for the covariance matrices avoid the problem of Heywood cases (negative variances and non-positive definite covariance matrices) that occur with maximum likelihood approaches. Bayesian methods have not been previously applied to CFA for analyzing familial data with multiple outcomes or for fitting the MTMM models. Bayesian techniques make it possible to solve most problems with standard CFA.

The rest of the paper proceeds as follows: Section 2 describes the proposed Bayesian Family Factor Model (BFFM), including the basic model structure, prior specification, a Gibbs algorithm to sample from posterior distributions, and methods for missing data imputation and hypothesis testing. Section 3 discusses simulation studies comparing the BFFM with the full information likelihood estimation of CFA model using quasi-Newton optimization by the lavaan [43] package in R. In Section 4 we fit the BFFM to the motivating UCLA Neurocognitive Family Study data. Implications and possible extensions are discussed in Section 5.

2. A Bayesian Family Factor Model

To model correlations among these multivariate observations in the form of two-way tables of data, classical CFA model using a combination of family-member factors and outcome factors can be used, but may have convergence problem and Heywood cases (lack-of-convergence) caused by empirical under-identification. To handle these inference problems, we propose a novel Bayesian Family Factor Model (BFFM) using informative priors, which has the advantages of being able to handle missing data, incorporating mean structure and test hypotheses easily. The basic structure for a BFFM is described below.

Suppose K normally distributed outcomes are collected on each of J members in N families. Let i, j and k index family, member type and outcome, respectively, with $i = 1, \ldots, N$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Then y_{ijk} is the k^{th} outcome for the j^{th} member in the i^{th} family, and $\boldsymbol{y}_i = (y_{i11}, \ldots, y_{i1K}, \ldots, y_{iJ1}, \ldots, y_{iJK})^T$ is the $JK \times 1$ vector of observations for all J members in the i^{th} family. The relationships among the JK observed variables are characterized by a factor analysis model

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}_A\boldsymbol{f}_{Ai} + \boldsymbol{\Lambda}_B\boldsymbol{f}_{Bi} + \varepsilon_i, \quad (1)$$

where $\boldsymbol{X}_{i(JK \times P)}$ is a matrix of known covariates for family i; $\boldsymbol{\beta}_{P \times 1} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_P^T)^T$ is a vector of regression coefficients; $\boldsymbol{f}_{Ai(J \times 1)} = (f_{Ai1}, \ldots, f_{AiJ})^T$ and $\boldsymbol{f}_{Bi(K \times 1)} = (f_{Bi1}, \ldots, f_{BiK})^T$ are a priori independent vectors of family member factors and outcome factors, respectively, with corresponding unstructured variance-covariance matrices $\boldsymbol{\Phi}_{A(J \times J)}$ and $\boldsymbol{\Phi}_{B(K \times K)}$

$$\boldsymbol{f}_{Ai} \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Phi}_A),$$

$$\boldsymbol{f}_{Bi} \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Phi}_B);$$

$\boldsymbol{\Lambda}_{A(JK \times J)} = [\boldsymbol{A}_{jk}]$ is a family member factor loading matrix with $A_{jj} = \boldsymbol{a}_j = (1, a_{j2}, \ldots, a_{jK})^T$ and $A_{jk} = \boldsymbol{0}_{(K \times 1)}$ when $j \neq k$; and $\boldsymbol{\Lambda}_{B(JK \times K)} = [\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_K]^T$ is an outcome factor loading matrix, where $\boldsymbol{B}_1 = \boldsymbol{I}_K$, $\boldsymbol{B}_j = \text{diag}(b_{j1}, \ldots, b_{jK})$, for $j = 2, \ldots, J$. Here $\boldsymbol{a}_j$ is a vector of non-zero family factor loadings for the j^{th} family member specific effects and $\boldsymbol{\Lambda}_{Bj}$ is a diagonal matrix of outcome factor loadings for the j^{th} family member. Furthermore, $\boldsymbol{\varepsilon}_i$ is a $JK \times 1$ vector of unique errors independent of $\boldsymbol{f}_{Ai}$ and $\boldsymbol{f}_{Bi}$

$$\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{\Psi}),$$

with diagonal error variance matrix $\boldsymbol{\Psi}_{(JK \times JK)} = \text{diag}(\psi_1, \ldots, \psi_K, \ldots, \psi_1, \ldots, \psi_K)$, as we assume that unique variances for the same outcome are equal across family members.

The variance-covariance matrix of the outcomes, y_i, unconditional on the factors is

$$\sum = \text{var}(\boldsymbol{y}_i|\boldsymbol{\beta}) = \boldsymbol{\Lambda}_A\boldsymbol{\Phi}_A\boldsymbol{\Lambda}_A^T + \boldsymbol{\Lambda}_B\boldsymbol{\Phi}_B\boldsymbol{\Lambda}_B^T + \boldsymbol{\Psi},$$

with diagonal element

$$\text{var}(y_{ijk}|\boldsymbol{\beta}) = a_{jk}^2\phi_{Ajj} + b_{jk}^2\phi_{Bkk} + \psi_k, \quad (2)$$

where ϕ_{Ajj} is the j^{th} diagonal element of $\mathbf{\Phi}_A$ and ϕ_{Bkk} is the k^{th} diagonal element of $\mathbf{\Phi}_B$, a_{jk} is the k^{th} element of α_j and b_{jk} is the k^{th} diagonal element of B_j.

For the 1^{st} outcome, as $b_{j1} = 1$, the overall variance var($y_{ij1}|\beta$) is

$$\text{var}(y_{ij1}|\beta) = 1\phi_{Ajj} + b_{j1}^2\phi_{Bkk} + \psi_{k,} \quad (3)$$

which indicates that family factor variances, ϕ_{Ajj}, must be smaller than the overall variances of the first outcome, var(y_{ij1}). Therefore, it is better to scale observed variables to make var(y_{ij1}) similar to other overall variances, so that ϕ_{Ajj} will not be forced to be small, which can cause precision problems in computing when very small values are rounded to 0. Eqn 3 will help in setting priors for factor loadings and for the factor covariance matrix.

Next, we set size or scale for factors and factor loadings. The scales of all family member factors, f_{Aij}, for $j = 1, \ldots, J$, are set to be the same as the first outcome, y_{ij1}, by fixing the first nonzero loading in each column of $\mathbf{\Lambda}_A$ to 1, $a_{j1} = 1$. The unit of a family member factor loading, a_{jk}, is the ratio of the units of the k^{th} outcome to that of the first outcome, for $k = 2, \ldots, K$. Factor loading a_{jk} is the amount of change in y_{ijk} associated with a 1 unit increase in f_{Aij} all else held constant. In addition, because $a_{jk}/a_{j1} = a_{jk}/1 = a_{jk}$, loading a_{jk} is also the ratio of the effect of f_{Aij} on y_{ijk} to the effect of f_{Aij} on y_{ij1}.

Similarly, the scale for an outcome factor, f_{Bik}, is specified to be the same as that of outcome y_{i1k}, the first family member (proband), by fixing the first nonzero loading in each column of $\mathbf{\Lambda}_B$ to 1, $b_{1k} = 1$. Therefore, the scale of the k^{th} outcome is passed on to the k^{th} outcome factor, f_{Bik}. Similarly, outcome loading b_{jk} is the amount of change in y_{ijk} associated with a 1 unit increase in f_{Bik} and as $b_{jk}/b_{1k} = b_{jk}$ for $j = 2, \ldots, J$, b_{jk} is also the ratio of the effect of f_{Bik} on y_{ijk} to that on y_{i1k}.

The total number of free hyper-parameters in the model is ($2JK + J^2/2 + K^2/2 - J/2 + K/2 + P$), as there are P regression coefficients, $J(K - 1)$ family member factor loadings, $(J - 1)K$ outcome factor loadings, $J(J + 1)/2$ unique parameters in the family factor variance matrix, $K(K + 1)/2$ unique parameters in the outcome factor variance matrix, and K unique error variance parameters. Similar to the CFA model for MTMM data, this model requires at least a total $J + K \geq 6$ family member and outcome factors with at least $J \geq 2$ family member and $K \geq 2$ outcome factors to be identified. It is not empirically identified when the loading matrix has deficient column rank [29], or when all family member or outcome factor correlations are equal [31].

There is a one-to-one correspondence between model parameters and lines on the path diagram in Figure 1. Unstructured factor variances matrices, $\mathbf{\Phi}_A$ and $\mathbf{\Phi}_B$, correspond to bidirectional arrows among the $J = 4$ family member factors on the left and among the $K = 5$ outcome factors on the right, respectively. The non-zero elements of $\mathbf{\Lambda}_A$, namely $\alpha_1, \ldots, \alpha_J$, correspond to unidirectional arrows from family member factors on the left to the JK observed variables, y_i. The non-zero elements of $\mathbf{\Lambda}_B$, namely diagonal elements of B_j, correspond to unidirectional arrows from family member factors on the right to y_i.

2.1. Prior Distributions for the Bayesian Family Factor Model

In the absence of strong theoretical or empirical beliefs to the contrary, we specify conditionally conjugate priors for all parameters. The prior distributions for the regression coefficients, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_P)^T$, and free elements a_{jk} and b_{jk} in the factor loading matrices, $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$, are independent normal

$$\beta_p \overset{\text{iid}}{\sim} \mathcal{N}(\beta_{0p}, \sigma^2_{\beta 0p}), \quad \text{for } p = 1, \ldots, P,$$

$$a_{jk} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{ajk}, \sigma^2_{ajk}), \quad \text{for } j = 1, \ldots, J, k = 2, \ldots, K,$$

$$b_{jk} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_{bjk}, \sigma^2_{bjk}), \quad \text{for } j = 2, \ldots, J, k = 1, \ldots, K$$

The factor variance matrices, $\boldsymbol{\Phi}_A$ and $\boldsymbol{\Phi}_B$, follow independent inverse Wishart distributions

$$\boldsymbol{\Phi}_A \sim \mathcal{IW}(\boldsymbol{W}_A, \nu_A),$$

$$\boldsymbol{\Phi}_B \sim \mathcal{IW}(\boldsymbol{W}_B, \nu_B),$$

where ν_A and ν_B are the degrees of freedom parameters, $\boldsymbol{W}_{A(J\times J)} = (\nu_A - J - 1)\boldsymbol{D}_A\boldsymbol{C}_A\boldsymbol{D}_A$ and $\boldsymbol{W}_{B(K\times K)} = (\nu_B - K - 1)\boldsymbol{D}_B\boldsymbol{C}_B\boldsymbol{D}_B$ are location parameters, $\boldsymbol{C}_{A(J\times J)}$ and $\boldsymbol{C}_{B(K\times K)}$ are prior factor correlation matrices, and $\boldsymbol{D}_{A(J\times J)} = \text{diag}(d_{A1}, \ldots, d_{AJ})$ and $\boldsymbol{D}_{B(K\times K)} = \text{diag}(d_{B1}, \ldots, d_{BK})$ are matrices with factor standard deviations d_{Aj} and d_{Bk} as diagonal elements to be specified shortly. Independent inverse-gamma priors are specified for the K distinct diagonal elements of $\boldsymbol{\Psi}$

$$\psi_k \overset{\text{ind}}{\sim} \mathcal{IG}\left(\frac{\vartheta_{\psi k}}{2}, \frac{\varsigma_{\psi k}}{2}\right),$$

for $k = 1, \ldots, K$.

2.2. Specification of Prior Hyper-parameters

We specify prior hyper-parameters based on model interpretation and subject matter knowledge. The basic assumptions are that the variances of the K outcomes are distinct due to scale differences and that the variances across family members of the k^{th} outcome are similar. The first step is to obtain estimated values for the overall variances of the K outcomes, $\widehat{\text{var}}(y..1), \ldots, \widehat{\text{var}}(y..k)$, either from the literature, from previous studies or from expert opinion. When no other information is available, 1⁄4 of the range of the k^{th} outcome variable in the data set under study is a plausible value of $\widehat{\text{var}}(y..k)^{1/2}$.

To specify priors for factor variance matrices, we use Equation (3), which implies the minimum of $\widehat{\text{var}}(y..1), \ldots, \widehat{\text{var}}(y..k)$ can be used as an upper bound for the prior mean d^2_{A1} of ϕ_{A11}, and we set $d^2_{A1} = p_{\phi a1} \min \widehat{\text{var}}(y..1), \ldots, \widehat{\text{var}}(y..k))$ for $0 < p_{\phi a1} \leq 1$, a scaling constant to be specified later. Furthermore,

$$\text{var}(y_{i11}) = \phi_{A11} + \phi_{B11} + \psi_1 \quad (4)$$

implies $\widehat{\text{var}}(y..1)$ can be used as an upper bound for the prior mean of ϕ_{B11}, $d^2_{B1} = p_{\phi b1}\widehat{\text{var}}(y..1)$, again for a scaling constant $0 < p_{\phi b1} \leq 1$ to be specified later. As the scale of the k^{th} outcome is passed on to the k^{th} outcome factor, f_{Bik}, the prior means of outcome factor variances, ϕ_{Bkk}, are set to be proportional to the estimated overall variances,

$$\frac{d^2_{B1}}{\widehat{\text{var}}(y..1)} = \cdots \frac{d^2_{BK}}{\widehat{\text{var}}(y..K)} = p_{\phi b1}$$

As the scale of the first outcome is passed on to all family member factors, f_{Aij}, the prior means of the factor variances are set to be equal, $d^2_{A1} = \ldots = d^2_{AJ}$.

Information on correlations within outcome across family members and among outcomes within subjects can help to specific C_A and C_B, the prior factor correlation matrices. Some information about theoretical associations among family members are available. For example, the genetic correlations between father and mother, between parent and child and between siblings are 0, 0.5 and 0.5, respectively. In addition, some outcome measures are known to be more closely related than others. For example, correlations among sub-scales from the same test will be similar and higher than correlations coming from sub-scales of different tests, which can be reflected in the prior factor correlation matrix C_B.

Prior means of factor loadings are elicited as follows. For a particular outcome, effects of the different family member factors on the observed variables are likely to be similar, so we assume that the prior means of loadings for the same outcome are equal across members,

$$\mu_{a1k} = \mu_{a2k} = \cdots \mu_{aJk},$$

for $k = 2, \ldots, K$. As the scale of a_{jk} is the ratio of the scale of the k^{th} outcome to the scale of the first outcome. we set prior means of the loadings proportional to the square root of the estimated overall variances

$$\frac{1}{\widehat{\text{var}}(y..1)^{1/2}} = \frac{\mu_{aj2}}{\widehat{\text{var}}(y..2)^{1/2}} = \ldots = \frac{\mu_{ajk}}{\widehat{\text{var}}(y..K)^{1/2}},$$

for $j = 1, \ldots, J$. For outcome factor loadings, because effects of the same outcome factor on observed variables are likely to be similar across family members, and $b_{11} = \ldots = b_{1K} = 1$, we set prior means of all outcome factor loadings to 1

$$\mu_{bjk} \equiv 1,$$

for $j = 2, \ldots, J$ and $k = 1, \ldots, K$. To specify prior means for the unique error variance, ψ_k, for $k = 1, \ldots, K$, $\widehat{\text{var}}(y..k)$ can also be used as an upper bound, $E(\psi_k) = \vartheta_{\psi k}/(\mathcal{X}_{\psi k} - 2) = p_\psi \widehat{\text{var}}(y..k)$, for some scaling constant p_ψ. From Equation (4), we can decompose the estimate of $\text{var}(y_{i11})$ into 3 parts by setting $p_{\phi a1} + p_{\phi b1} + p_\psi = 1$. When data from a previous study are available, these scaling constants can be further estimated using the proportions of variances explained by unique variances, outcome factors and/or family factors in standard CFA models.

To specify the priors for regression coefficients, it is necessary to identify plausible values for the covariate effects on each outcome from previous studies or expert opinion. For the special case where covariates are indicators of diagnostic or treatment groups, the estimated means of outcomes in the general population or in patients from earlier studies are useful guides for choosing prior means.

2.3. Gibbs Sampling from the Posterior Distribution

We use conditionally conjugate priors so that simulation of the posterior distribution proceeds via a Gibbs sampling algorithm [44, 45]. To reduce autocorrelation and improve efficiency, we use a blocked Gibbs sampler to sample the regression coefficients, $\boldsymbol{\beta}$ and the factor scores, $\boldsymbol{f}_i$ from their joint conditional distributions, respectively.

Missing y_{ijk} are imputed at each iteration of the MCMC algorithm with a data augmentation (DA) algorithm [46]. This approach has the advantage of using BFFM for both imputation and data analysis. Because the missing and observed data are jointly normal given parameters, the conditional distribution of the missing data given the observed data and the unknown parameters is also normal. Thus missing outcome data or families without certain members are not a problem for our method. We implement [47]'s sweep operator algorithm for imputation of multivariate normal data. Full details of the Gibbs sampler are given in Web Appendix A.

2.4. Hypothesis Testing using Bayes Factors (BF)

Besides estimating model parameters, it is of interest to test various hypotheses about these parameters. In Bayesian inference, testing a null hypothesis against an alternative can be regarded as comparing two corresponding models, $\mathcal{M}_0$ and $\mathcal{M}_1$. A Bayes factor [48–50] is a summary of evidence provided by the data in favor of $\mathcal{M}_0$ as opposed to $\mathcal{M}_1$

$$B_{01} = \frac{p(\boldsymbol{Y}|\mathcal{M}_0)}{p(\boldsymbol{Y}|\mathcal{M}_1)} \quad (5)$$

where $p(\boldsymbol{Y}|\mathscr{M}_\ell) = \int p(\boldsymbol{y}|\boldsymbol{\Theta}_\ell)p(\boldsymbol{\Theta}_\ell)d\boldsymbol{\Theta}_\ell$ for $\ell = 0, 1$ is the marginal likelihood of the data $\boldsymbol{Y}$ given model $\mathscr{M}_\ell$.

Let $\mathscr{M}_1$ denote a general model indexed by $\boldsymbol{\Theta} = (\boldsymbol{\omega}^T, \boldsymbol{\Upsilon}^T)^T$, where $\boldsymbol{\omega}$ denotes the vector of parameters of interest, $\boldsymbol{\Upsilon}$ denotes the vector of all the remaining "nuisance parameters", $p(\boldsymbol{\Theta}|\mathscr{M}_1)$ denotes the prior density under $\mathscr{M}_1$ and $p(\boldsymbol{Y}|\boldsymbol{\Theta}, \mathscr{M}_1)$ denote the sampling density under $\mathscr{M}_1$. A nested model, denoted $\mathscr{M}_0$, is constructed by setting $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, while leaving $\boldsymbol{\Upsilon}$ unconstrained. The marginal likelihood of $\boldsymbol{Y}$ under $\mathscr{M}_0$ is $p(\boldsymbol{Y}|\mathscr{M}_0) = p(\boldsymbol{Y}|\boldsymbol{\omega} = \boldsymbol{\omega}_0, \mathscr{M}_1)$. The prior density under $\mathscr{M}_0$ satisfies $p(\boldsymbol{\Upsilon}|\mathscr{M}_0) = p(\boldsymbol{\Upsilon}|\boldsymbol{\omega} = \boldsymbol{\omega}_0, \mathscr{M}_1)$ and the sampling density under $\mathscr{M}_0$ is $p(\boldsymbol{Y}|\boldsymbol{\Upsilon}, \mathscr{M}_0)$. From Bayes Theorem,

$$p(\boldsymbol{Y}|\boldsymbol{\omega} = \boldsymbol{\omega}_0, \mathscr{M}_1) = \frac{p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\boldsymbol{Y}, \mathscr{M}_1)p(\boldsymbol{Y}|\mathscr{M}_1)}{p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\mathscr{M}_1)}, \quad (6)$$

so the Bayes factor can be expressed as the Savage-Dickey density ratio [51–53]

$$B_{01} = \frac{p(\boldsymbol{Y}|\mathscr{M}_0)}{p(\boldsymbol{Y}|\mathscr{M}_1)} = \frac{p(\boldsymbol{Y}|\boldsymbol{\omega} = \boldsymbol{\omega}_0, \mathscr{M}_1)}{p(\boldsymbol{Y}|\mathscr{M}_1)} = \frac{p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\boldsymbol{Y}, \mathscr{M}_1)}{p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\mathscr{M}_1)}. \quad (7)$$

The marginal prior density $p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\mathscr{M}_1)$ can be easily calculated from the prior. The marginal posterior, $p(\boldsymbol{\omega} = \boldsymbol{\omega}_0|\boldsymbol{Y}, \mathscr{M}_1)$, can be estimated using MCMC outputs from the unrestricted model, which provide draws from the marginal posterior $p(\boldsymbol{\omega}|\boldsymbol{Y}, \mathscr{M}_1)$. Different methods to calculate the marginal posteriors include the usual normal approximation, and conditional marginal density estimation (CMDE) [53].

In the UCLA Family Study, hypotheses of interest include whether group means are equal (1) between schizophrenia and control families and (2) across family member types within schizophrenia families. Both scenarios are equivalent to testing whether particular linear combinations of the regression coefficients are simultaneously equal to zero

$$\boldsymbol{\omega}_{(l\times 1)} = \boldsymbol{L}_{(l\times P)}\boldsymbol{\beta}_{(P\times 1)} = \boldsymbol{0}_{(l\times 1)},$$

where $\boldsymbol{L}$ is a full rank matrix. Let $\boldsymbol{M}$ denote any $(P - l) \times P$ full rank matrix so that rank($[\boldsymbol{L}^T, \boldsymbol{M}^T]$) = P and let $\boldsymbol{\omega}^\perp = \boldsymbol{M}\boldsymbol{\beta}$. Let $\mathscr{M}1$ denote a general model where $\boldsymbol{\beta}$ is freely estimated, which has parameters $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\Lambda}_A, \boldsymbol{\Lambda}_B, \boldsymbol{\Phi}_A, \boldsymbol{\Phi}_B, \boldsymbol{f}_{A1}, \ldots, \boldsymbol{f}_{AN}, \boldsymbol{f}_{B1}, \ldots, \boldsymbol{f}_{BN}, \boldsymbol{\Psi})$ and let $\boldsymbol{\Upsilon} = (\boldsymbol{\omega}^\perp, \boldsymbol{\Lambda}_A, \boldsymbol{\Lambda}_B, \boldsymbol{\Phi}_A, \boldsymbol{\Phi}_B, \boldsymbol{f}_{A1}, \ldots, \boldsymbol{f}_{AN}, \boldsymbol{f}_{B1}, \ldots, \boldsymbol{f}_{BN}, \boldsymbol{\Psi})$. Then $\boldsymbol{\Theta}$ can be reparametrized as $\boldsymbol{\Theta}^* = (\boldsymbol{\omega}, \boldsymbol{\Upsilon})$. The nested null model, $\mathscr{M}_0$, is constructed by setting $\boldsymbol{\omega} = \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$.

As a priori $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\beta 0}, \boldsymbol{\Sigma}_{\beta 0})$, the prior distribution of $\boldsymbol{\omega} = \mathbf{L}\boldsymbol{\beta}$ is

$$\boldsymbol{\omega} | \mathcal{M}_1 \sim \mathcal{N}(\boldsymbol{L}\boldsymbol{\mu}_{\beta 0}, \boldsymbol{L}\boldsymbol{\Sigma}_{\beta 0}\boldsymbol{L}^T),$$

so we can obtain the denominator of the Savage-Dickey density ratio, $p(\boldsymbol{\omega} | \mathcal{M}_1)|_{\boldsymbol{\omega}} = \mathbf{0}$.

Let $\hat{\boldsymbol{\mu}}_{\beta}$ and $\hat{\boldsymbol{\Sigma}}_{\beta}$ denote the posterior mean and variance of $\boldsymbol{\beta}$ estimated from the MCMC output. The marginal posterior distribution can be approximated as

$$\boldsymbol{\omega} | \boldsymbol{Y}, \mathcal{M}_1 \overset{\text{approx.}}{\sim} \mathcal{N}(\boldsymbol{L}\hat{\boldsymbol{\mu}}_{\beta}, \boldsymbol{L}\hat{\boldsymbol{\Sigma}}_{\beta}\boldsymbol{L}^T),$$

which gives an approximation to $p(\boldsymbol{\omega} | \mathcal{M}_1)|_{\boldsymbol{\omega} = \mathbf{0}}$.

The conditional marginal density estimator (CMDE) approximates $p(\boldsymbol{\omega} | \boldsymbol{Y}, \mathcal{M}_1)|_{\boldsymbol{\omega} = \mathbf{0}}$ using an average of the full conditional posterior density of $\boldsymbol{\omega}$ evaluated at $\boldsymbol{\omega} = 0$ over all T MCMC iterations

$$p(\boldsymbol{\omega} = \mathbf{0} | \boldsymbol{Y}, \mathcal{M}_1) \approx \frac{1}{T} \sum_{t=1}^{T} p(\boldsymbol{\omega} | \boldsymbol{\Upsilon}, \boldsymbol{Y})|_{\boldsymbol{\omega} = \mathbf{0}, \boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}^{(t)}}, \quad (8)$$

where $\boldsymbol{\Upsilon}^{(t)}$ is value of $\boldsymbol{\Upsilon}$ from the t^{th} MCMC sample. The full conditional posterior density of $\boldsymbol{\omega}$ is

$$p(\boldsymbol{\omega} | \boldsymbol{\Upsilon}, \boldsymbol{Y}) = p(\boldsymbol{\omega} | \boldsymbol{M}\boldsymbol{\beta}, \boldsymbol{\sum}, \boldsymbol{Y}),$$

where $\boldsymbol{\sum} = \text{var}(\boldsymbol{y}_i | \boldsymbol{\beta}) = \boldsymbol{\Lambda}_A \boldsymbol{\Phi}_A \boldsymbol{\Lambda}_A^T + \boldsymbol{\Lambda}_B \boldsymbol{\Phi}_B \boldsymbol{\Lambda}_B^T + \boldsymbol{\Psi}$ is the variance of $\boldsymbol{y}_i$ given $\boldsymbol{\beta}$.

3. Analysis of Simulated Data

To assess the performance of the Bayesian Family Factor Model (BFFM) in different scenarios, simulation studies are used to compare BFFM with CFA estimated by full information maximum likelihood (FIML) using quasi-Newton optimization (QNO), on the basis of ability to fit the data, as well as examining mean squared errors (MSE), squared biases and variances of parameters estimated by the two methods.

Grayson *et. al.* [29] proved that a CFA model is not identified when the true factor loading matrix, $\boldsymbol{\Lambda}$, is not full rank. One sufficient condition for deficient column rank is $\boldsymbol{\Lambda} = [\boldsymbol{C} \otimes \boldsymbol{a}_0 | \boldsymbol{d} \otimes \boldsymbol{B}_0]$, where $\boldsymbol{C}_{(J \times J)}$ and $\boldsymbol{B}_{0(K \times K)}$ are diagonal full rank matrices, and a_0 and d are $K \times 1$ and $J \times 1$ vectors, respectively [29]. We generate data under three scenarios where the true covariance matrix is identified, close to not identified and not identified, by

specifying different true factor loading matrices, $[\Lambda_A/\Lambda_B]$, which were far from equal to, almost equal to, and equal to $[\boldsymbol{C} \otimes \boldsymbol{a}_0 | \boldsymbol{d} \otimes \boldsymbol{B}_0]$.

Two hundred data sets were simulated for each scenario. Each data set has $N = 200$ families, $K = 5$ outcomes and $J = 4$ members: proband, sibling, father and mother. True regression coefficients $\boldsymbol{\beta}$ and true unique error variances $\boldsymbol{\Psi}$ are the same for the three scenarios,

$$\boldsymbol{X}_{i(JK \times 2JK)} = \left[d_i \boldsymbol{I}_{JK} \;\; (1 - d_i)\boldsymbol{I}_{JK}\right],$$

where $d_i = 0, 1$ for control and SZ families respectively, and $\boldsymbol{I}_{JK}$ is a $JK \times JK$ identity matrix. The corresponding $\boldsymbol{\beta}_{(2JK \times 1)}$ is $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $JK \times 1$ vectors of means of all K outcomes on the J family members in the control and SZ families. For $J = 4$ and $K = 5$, the total number of parameters is 101. The observations are set to be missing completely at random with probability $p = 0.15$ and the missingness pattern is the same across all 200 data sets for each scenario. True values for all parameters are given in Web Appendix B.

Standard non-Bayesian CFA models are fit to the simulated data using the lavaan package in R [43], which uses full information maximum likelihood (FIML) estimation to handle missing data and uses a quasi-Newton optimization algorithm to estimate parameters. Full information maximum likelihood estimation with quasi-Newton optimization (FIML-QNO) is defined as successful in fitting the data if the algorithm converges and provides a valid solution, e.g. having positive-definite covariance matrices and positive variances. In many cases, FIML-QNO fails to find a fit to the data due to empirical under-identification. The percentages of data sets for which FIML-QNO was successful in fitting in the 3 scenarios are 87%, 43% and 15%, respectively. When CFA model using FIML-QNO was fit to the same simulated data but with no missing observation, the percentages increase slightly to 92.5%, 50.5% and 21%, respectively, suggesting that the missing data was not the major cause of the failure of FIML-QNO.

Next, a BFFM is fit to the same 200 data sets in each scenario, with 10, 000 iterations after an initial burn-in of 1000 iterations. Priors are chosen to be partially informative and centered at true values with large dispersions. Trace plots, density plots and autocorrelation plots show no obvious evidence of bad mixing, non-convergence or high autocorrelations. BFFM successfully fit all 600 data sets and the resulting posterior means were always valid solutions (i.e. positive variances and positive definite covariance matrices).

We compare the performance characteristics of BFFM and FIML-QNO, when FIML-QNO was successful in fitting the data sets. The mean squared error (MSE) of an estimator $\hat{\theta}$ for a parameter θ, $\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$, measures the average squared distance between the estimator $\hat{\theta}$ and the true parameter value θ. The $\text{MSE}(\hat{\theta})$ is the sum of the variance of the estimator, $\text{Var}(\hat{\theta})$ and the squared bias, $[\boldsymbol{E}(\hat{\theta}) - \theta]^2$. Denote $\hat{\theta}_l$ as the posterior mean of θ from the MCMC outputs of the l^{th} data set, for $l = 1, \ldots, 200$, then the variance of $\hat{\theta}$ is estimated by

$$\mathrm{var}(\hat{\theta}) = \frac{1}{200}\sum_{l=1}^{200}\left(\hat{\theta}_l - 200^{-1}\sum_{l=1}^{200}\hat{\theta}_l\right)^2.$$

We compare methods on the relative MSE, relative variance and relative squared bias defined as $\mathrm{MSE}(\hat{\theta})/\theta^2$, $\mathrm{var}(\hat{\theta})/\theta^2$ and $[E(\hat{\theta}) - \theta]^2/\theta^2$, respectively.

In the scenario where the true covariance matrix is close to not identified, we compare relative mean squared errors (RMSE), relative variances and relative squared biases of all parameters estimated by fitting BFFM and FIML-QNO to the 43% of the data sets which FIML-QNO was successful in fitting (Web Appendix C). Overall, parameter estimates from BFFM and FIML-QNO are similar and are close to the true values. Figure 2(a) plots on a log-log scale the relative MSEs of parameters estimated by BFFM against relative MSEs of parameters estimated by FIML-QNO. There are 101 dots representing all parameters. Different symbols represent different groups of parameters (factor loadings, factor variance-covariance parameters, regression coefficients and unique error variances). For a given parameter, if RMSEs estimated by two models are the same, the dot will lie on a diagonal line with slope 1, which is drawn in every plot; when the RMSE estimated by BFFM is smaller, the dot will lie above the diagonal line; and when the RMSE estimated by FIML-QNO is smaller, the dot will lie below the diagonal line. For more than 60% of the parameters, the RMSEs estimated by BFFM are smaller. For most parameters, the RMSEs estimated by both methods are small (RMSE < 0.1, dots in the lower left corner). However, for some factor loadings and factor variance-covariances, the RMSEs estimated by FIML-QNO are much larger than those estimated by BFFM (dots in the upper half).

Figure 2(b) plots relative variances of parameters estimated by BFFM against those estimated by FIML-QNO. Almost all dots lie above the diagonal line, where BFFM has smaller relative variances for almost all parameters. Similarly, Figure 2(c) plots the relative squared biases ($[E(\hat{\theta}) - \theta]^2/\theta^2$) of parameters estimate by BFFM and FIML-QNO. For most parameters, relative squared biases are small (< 0.1) using both methods. For about 40% of the parameters, the relative squared biases estimated by BFFM are smaller, but the FIML-QNO has smaller relative squared biases when both methods perform well (relative squared biases < 0.1). However, as with the relative MSEs, for some factor variance-covariances and factor loadings, the squared biases estimated by FIML-QNO are larger than those estimated by BFFM.

It is important to check whether BFFM performs worse when FIML-QNO failed as compared to when FIML-QNO succeeded.

The plot of the relative MSEs in Figure 2(d) indicates that BFFM works equally well for both kinds of data sets where FIML-QNO did or did not converge, because almost all dots are close to the diagonal line with slope 1. The plots of relative squared biases and relative variances are similar. Plots comparing the BFFM and FIML-QNO in the scenarios where the true covariance matrix is identified or not identified are presented in Web Appendix C, and show similar patterns.

4. Application to the UCLA Family Study Data

We fit the BFFM to the UCLA NFS data. The $J = 4$ family member types, proband, sibling, father and mother are indexed by $j = 1, \ldots, 4$, respectively. The $K = 5$ cognitive outcome measures, MANIPA, CPTDSD, CPT37D, SPAN10 and logTRLB, corresponding to $k = 1, \ldots, 5$, are scaled and transformed as described in Table 1 to make the scales similar and all correlations positive. The sample consists of 625 subjects and $N = 206$ families, about half of which are schizophrenia families. The covariate is an indicator of being in a schizophrenia or control family, as in the simulation study. For families with two or more siblings, data on one sibling are randomly chosen to use in this analysis.

To fit BFFM, partially informative priors are specified using the approach described in Section 2.1. In particular, estimates of overall means and variances for the five outcome measures are obtained from Phase 1 data of the UCLA NFS [2, 3], which collected four of the five outcomes in the current study, and from some previous literature reporting these outcome measures [54–56]. The estimates of overall means and standard deviations from all those various sources are summarized in Table 1. Hyper-parameters for variances or dispersion are set to produce large variances so that the prior is uninformative. As the raw correlations of four outcomes from the phase 1 data ignoring the family structure were between 0.3 and 0.5, the prior correlation means among all 5 outcomes are set to be 0.3. Based on the phase 1 data correlations and the theoretical correlations among the family members, the prior means of correlations among family members are set to be 0.15 between father and mother and 0.2 otherwise.

When the classic CFA model using full information likelihood estimation and the full information maximum likelihood estimation using quasi-Newton optimization (FIML-QNO) is fit to the UCLA NFS data using the lavaan package in R, the algorithm fails to converge. The AMOS add-on in SPSS failed to fit the data because it can not handle missing data, while the CALIS procedure in SAS failed to converged.

The BFFM estimation procedure using a Gibbs sampling algorithm is implemented in R, with a total of 100, 000 iterations after excluding 10, 000 initial burn-in iterations. Trace plots, density plots and autocorrelation plots show no obvious evidence of bad mixing, non-convergence or high autocorrelation.

Tables 1, 2 and 3 in web Appendix A4 present summaries of the posterior distributions for all 101 parameter estimates, including means, standard deviations (SD), and posterior probabilities $p(\theta < 0/\mathbf{Y})$. Table 2 presents posterior means of factor variances, factor correlations and factor loadings for family member and outcome factors. The posterior means of all family member factor correlations are positive and vary from a low of 0.034 between mother and sibling to a high of 0.390 between proband and sibling. Similarly, the posterior means of all outcome factor correlations are positive and range from 0.29 to 0.61. The posterior means of all factor loadings are also all positive. Thus the observed variables are positively associated with the factors they load on.

Table 3 lists the $2JK = 40$ posterior means of regression coefficients, β_{pjk} (top) and posterior means of the differences in group means, control minus SZ, $\beta_{1jk} - \beta_{2jk}$ (bottom). SZ

probands performed worse than the control probands for all five outcomes, while the sign of the differences in mean outcomes between siblings of the two groups are not well determined by looking at the posterior probabilities $p(\theta < 0/\boldsymbol{Y})$. Parents of schizophrenia probands did worse in span of apprehension and trails B than control parents.

Figure 3 plots the posterior distribution of group means of CPT37D for probands, siblings, fathers and mothers in the control and SZ families (left) and the differences between two groups. These plots show that the means of CPT37D for SZ probands are much smaller than those for control probands, while there are no obvious differences in means between the two groups for fathers and siblings. Mothers in the control families have larger means CPT37D than all others, including the SZ mothers. Additional plots of posterior distributions for group means and factor loadings are provided in Tables 11, 12 and 13 and Figures 6, 7 and 8 in Web Appendix A4. By comparing posterior means of five outcomes plotted in Appendix A4 Figure 6, psychologists can identify CPT37D as the cognitive deficit which differs most between relatives of schizophrenics and relatives of control probands. Furthermore, for some cognitive impairments such as SPAN10 and logTRLBA, parents of schizophrenics seem to perform worse than parents of control probands, while differences between control and SZ siblings are not obvious. These results suggest different patterns of neurocognitive deficits aggregation in relatives of schizophrenics and motivate further research to investigate potential heritable vulnerability factors for schizophrenia.

4.1. Hypothesis Testing

For the UCLA NFS, researchers are interested in differences in cognitive measurements both between the control and SZ families and among different members in the SZ family. First we test whether the means of all outcomes for SZ and control probands are equal (the number of contraints, NC = $K = 5$). The linear combination of interest is

$$\boldsymbol{L}_{1(K\times 2JK)} = \left[\boldsymbol{I}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ -\boldsymbol{I}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\right],$$

where $\boldsymbol{0}_{K\times K}$ is a $K \times K$ matrix of 0's. Next, we test whether the means of SZ probands are equal to the means of the average of siblings, fathers and mothers of SZ probands (NC = K = 5). The linear combination of interest is

$$\boldsymbol{L}_{2(K\times 2JK)} = \left[\boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{0}_{K\times K}\ \boldsymbol{I}_K\ -\tfrac{1}{3}\boldsymbol{I}_K\ -\tfrac{1}{3}\boldsymbol{I}_K\ -\tfrac{1}{3}\boldsymbol{I}_K\right],$$

where $\boldsymbol{0}_{K\times K}$ is a $K \times K$ matrix of 0's. Third, we test whether for a particular variable, say CPT37D ($k = 3$), the means of SZ and control family members are equal (NC = $J = 4$). Let d = (0, 0, 1, 0, 0) be a 1 × 5 row vector, then the linear combination of interest is

$$\boldsymbol{L}_3 = \left[\boldsymbol{I}_J \otimes \boldsymbol{d}\ \ -\boldsymbol{I}_J \otimes \boldsymbol{d}\right]$$

Similar Bayes factors (BF) are obtained using the three methods, (the normal approximation, KDE and CMDE). The results using the normal approximation from multiple MCMC runs

are summarized here and in Figure 9 of Appendix A4 using boxplots. The Bayes factor estimated for $\boldsymbol{L}_1$ (mean = 6.1, s.d.= 1.4) suggests the 5 × 1 vectors of means of the probands between the two groups are not different, though the support in the data for this hypothesis is "barely worth a mention" [50]. For $\boldsymbol{L}_2$ (NC= 5), the mean Bayes factor is 0.00009 (s.d. = 0.00004), suggesting that the means of SZ probands are quite different from the average of their relatives. As for $\boldsymbol{L}_3$ (NC= 4), the mean of Bayes factor is 0.0014 (s.d. = 0.0008), suggesting strong evidence against equality in means of CPT37D between SZ and control families, which is consistent with the 1-sided posterior probability, $p(\beta_{113} - \beta_{213} > 0/\boldsymbol{Y})$ for testing the difference in means between the two groups.

5. Discussion

BFFM is a novel covariance model using $2JK + (J^2 + K^2 - J + K)/2$ parameters to model a $JK \times JK$ covariance matrix. Factor analysis is not the primary goal of this example, but rather used as a convenient way to model dependencies. We take a Baysian approach to facilitate this modeling through the use of informative priors and consistent ability to fit the model to any data set. An interesting path for future research is to expand the analysis to include more than one factor per outcome and/or more than one factor per family member.

This model is not limited to small nuclear family (J=4) or balanced family structure with exactly J members. It can be extended for families of different sizes wit h different types of members. (1) If one or more members do not exist in a family, we treat the data as missing and impute them as described in Section 2.3. (2) If there is more than one sibling, we can include more than one family member factors for different siblings and specify the family member factor variance matrix $\boldsymbol{\Phi}_A$ as: (i) Factor variances are equal across siblings; (ii) Factor covariances between siblings are all equal and (iii) Factor covariances between a sibling and another family members are equal across siblings. Instead of Gibbs sampling from a Wishart distribution for $\boldsymbol{\Phi}_A$, other method will be used to sample $\boldsymbol{\Phi}_A$.

This model does not have scalability problem and can be extend to model data with larger family sizes and larger numbers of outcomes. Due to the flexibility of allowing one factor per family member and one factor per outcome, this is a general model not limited to familial data and can be easily applied to any two-way table data with similar correlation structure, such as MTMM data to study construct validity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Keith H. Nuechterlein and Dr. Robert F. Asarnow for access to the data and for helpful discussions. This work was partially funded by a Dissertation Year Fellowship from the Graduate Division at UCLA and by NIMH Grants MH041953, MH049716, MH045112, MH037705 and MH066286.

References

1. Asarnow RF, Nuechterlein KH, Fogelson D, Subotnik KL, Payne DA, Russell AT, Asamen J, Kuppinger H, Kendler KS. Schizophrenia and schizophrenia-spectrum personality disorders in the first-degree relatives of children with schizophrenia: the UCLA family study. Archives of General Psychiatry. 2001; 58(6):581–8. [PubMed: 11386988]
2. Asarnow RF, Nuechterlein KH, Subotnik KL, Fogelson DL, Torquato RD, Payne DL, Asamen J, Mintz J, Guthrie D. Neurocognitive impairments in nonpsychotic parents of children with schizophrenia and attention-deficit/hyperactivity disorder: The University of California, Los Angeles Family Study. Arch Gen Psychiatry. 2002; 59(11):1053–60. [PubMed: 12418939]
3. Nuechterlein KH, Asarnow RF, Subotnik KL, Fogelson DL, Payne DL, Kendler KS, Neale MC, Jacobson KC, Mintz J. The structure of schizotypy: Relationships between neurocognitive and personality disorder features in relatives of schizophrenic patients in the UCLA Family Study. Schizophrenia Research. 2002; 54(1):121–130. [PubMed: 11853986]
4. Bollen, KA. Structural Equation Models. Wiley; New York: 1998.
5. Loehlin, JC. Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. Psychology Press; 2004.
6. Arbuckle, JL. IBM®SPSS® AMOS 20 Users Guide. IBM Corporation; 2011.
7. Hamsten A, de Faire U. Risk factors for coronary artery disease in families of young men with myocardial infarction. The American Journal of Cardiology. 1987; 59(1):14–19. [PubMed: 3812225]
8. Sammel M, Lin X, Ryan L. Multivariate linear mixed models for multiple outcomes. Statistics in Medicine. 1999; 18(17–18):2479–2492. [PubMed: 10474154]
9. McCulloch, CE. Generalized Linear Mixed Models. Wiley; New York: 2006.
10. Byrne, BM. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming. CRC Press; 2009.
11. Kline, RB. Principles and Practice of Structural Equation Modeling. Guilford Press; 2011.
12. Thompson, B. Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. American Psychological Association; 2004.
13. Brown, TA. Confirmatory Factor Analysis for Applied Research. The Guilford Press; 2006.
14. Browne MW. The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology. 1984; 37(1):1–21. [PubMed: 6733052]
15. Cudeck R. Analysis of correlation matrices using covariance structure models. Psychological Bulletin. 1989; 105(2):317–327.
16. Wothke W, Browne MW. The direct product model for the MTMM matrix parameterized as a second order factor analysis model. Psychometrika. 1990; 55(2):255–262.
17. Naik DN, Rao SS. Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. Journal of Applied Statistics. 2001; 28(1):91–105.
18. Srivastava MS, von Rosen T, von Rosen D. Models with a Kronecker product covariance structure: Estimation and testing. Mathematical Methods of Statistics. 2008; 17(4):357–370.
19. Harman, HH. Modern Factor Analysis. University of Chicago Press; 1960.
20. Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika. 1969; 34(2):183–202.
21. Rubin DB, Thayer DT. EM algorithms for ML factor analysis. Psychometrika. 1982; 47(1):69–76.
22. Basilevsky, AT. Statistical Factor Analysis and Related Methods: Theory and Applications. Wiley; New York: 2009.
23. Bartholomew, DJ., Knott, M., Moustaki, I. Latent Variable Models and Factor Analysis: A Unified Approach. Wiley; New York: 2011.
24. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. 1959; 56(2):81–105. [PubMed: 13634291]
25. Marsh HW. Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. Applied Psychological Measurement. 1989; 13(4):335–361.

26. Eid M, Lischetzke T, Nussbeck FW. Structural Equation Models for Multitrait-multimethod Data. Handbook of Multimethod Measurement in Psychology. 2006; (xiv):283–299.
27. Marsh HW, Hocevar D. A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. Journal of Applied Psychology. 1988; 73(1):107–117.
28. Kenny DA, Kashy DA. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. Psychological Bulletin. 1992; 112(1):165–172.
29. Grayson D, Marsh HW. Identification with deficient rank loading matrices in confirmatory factor analysis: Multitrait-multimethod models. Psychometrika. 1994; 59(1):121–134.
30. Wothke, W. PhD Dissertation. University of Chicago; 1984. The Estimation of Trait and Method Components in Multitrait-Multimethod Measurement.
31. Brannick MT, Spector PE. Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. Applied Psychological Measurement. 1990; 14(4):325–339.
32. Lance CE, Noble CL, Scullen SE. A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. Psychological Methods. 2002; 7(2): 228–244. [PubMed: 12090412]
33. Press, SJ., Shigemasu, K. Contributions to Probability and Statistics. Vol. 15. Springer; 1989. Bayesian Inference in Factor Analysis; p. 271-287.
34. West M. Bayesian factor regression models in the “large p, small n” paradigm. Bayesian Statistics. 2003; 7(2003):723–732.
35. Lopes HF, West M. Bayesian model assessment in factor analysis. Statistica Sinica. 2004; 14(1): 41–68.
36. Ghosh J, Dunson DB. Default prior distributions and efficient posterior computation in Bayesian factor analysis. Journal of Computational and Graphical Statistics. 2009; 18(2):306–320. [PubMed: 23997568]
37. Press, SJ. Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference. Courier Dover Publications; 2012.
38. Geweke J, Zhou G. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies. 1996; 9(2):557–587.
39. Press, SJ., Shigemasu, K. Technical report 243. University of California; Riverside: 1997. Bayesian Inference in Factor Analysis (Revised).
40. Rowe, D. PhD dissertation. University of California; Riverside: 1998. Correlated Bayesian Factor Analysis.
41. Aguilar O, West M. Bayesian dynamic factor models and portfolio allocation. Journal of Business & Economic Statistics. 2000; 18(3):338–357.
42. Rowe, DB. Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing. CRC Press; 2003.
43. Rosseel Y. Iavaan: An R package for structural equation modeling. Journal of Statistical Software. 2012; 48(2):1–36.
44. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association. 1990; 85(410):398–409.
45. Robert, CP., Casella, G. Monte Carlo Statistical Methods. Springer; 2004.
46. Little, RJA., Rubin, DB. Statistical Analysis with Missing Data. Wiley; New York: 2002.
47. Schafer, JL. Analysis of Incomplete Multivariate Data. CRC press; 1997.
48. Berger, JO. Statistical Decision Theory and Bayesian Analysis. Springer; 1985.
49. Chib S. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association. 1995; 90(432):1313–1321.
50. Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association. 1995; 90(430):773–795.
51. Dickey JM, Lientz BP. The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. The Annals of Mathematical Statistics. 1970; 41(1):214–226.
52. Verdinelli I, Wasserman L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. Journal of the American Statistical Association. 1995; 90(430):614–618.

53. Morey RD, Rouder JN, Pratte MS, Speckman PL. Using MCMC chain outputs to efficiently estimate Bayes factors. Journal of Mathematical Psychology. 2011; 55(5):368–378.

54. Kim J, Glahn DC, Nuechterlein KH, Cannon TD. Maintenance and manipulation of information in schizophrenia: Further evidence for impairment in the central executive component of working memory. Schizophrenia Research. 2004; 68(2):173–187. [PubMed: 15099601]

55. Nuechterlein KH, Subotnik KL, Green MF, Ventura J, Asarnow RF, Gitlin MJ, Yee CM, Gretchen-Doorly D, Mintz J. Neurocognitive predictors of work outcome in recent-onset schizophrenia. Schizophrenia Bulletin. 2011; 37(suppl 2):S33–S40. [PubMed: 21860045]

56. Koide T, Aleksic B, Kikuchi T, Banno M, Kohmura K, Adachi Y, Kawano N, Iidaka T, Ozaki N. Evaluation of Factors Affecting Continuous Performance Test Identical Pairs Version Score of Schizophrenic Patients in a Japanese Clinical Sample. Schizophrenia Research and Treatment. 2012; 2012:1–5.

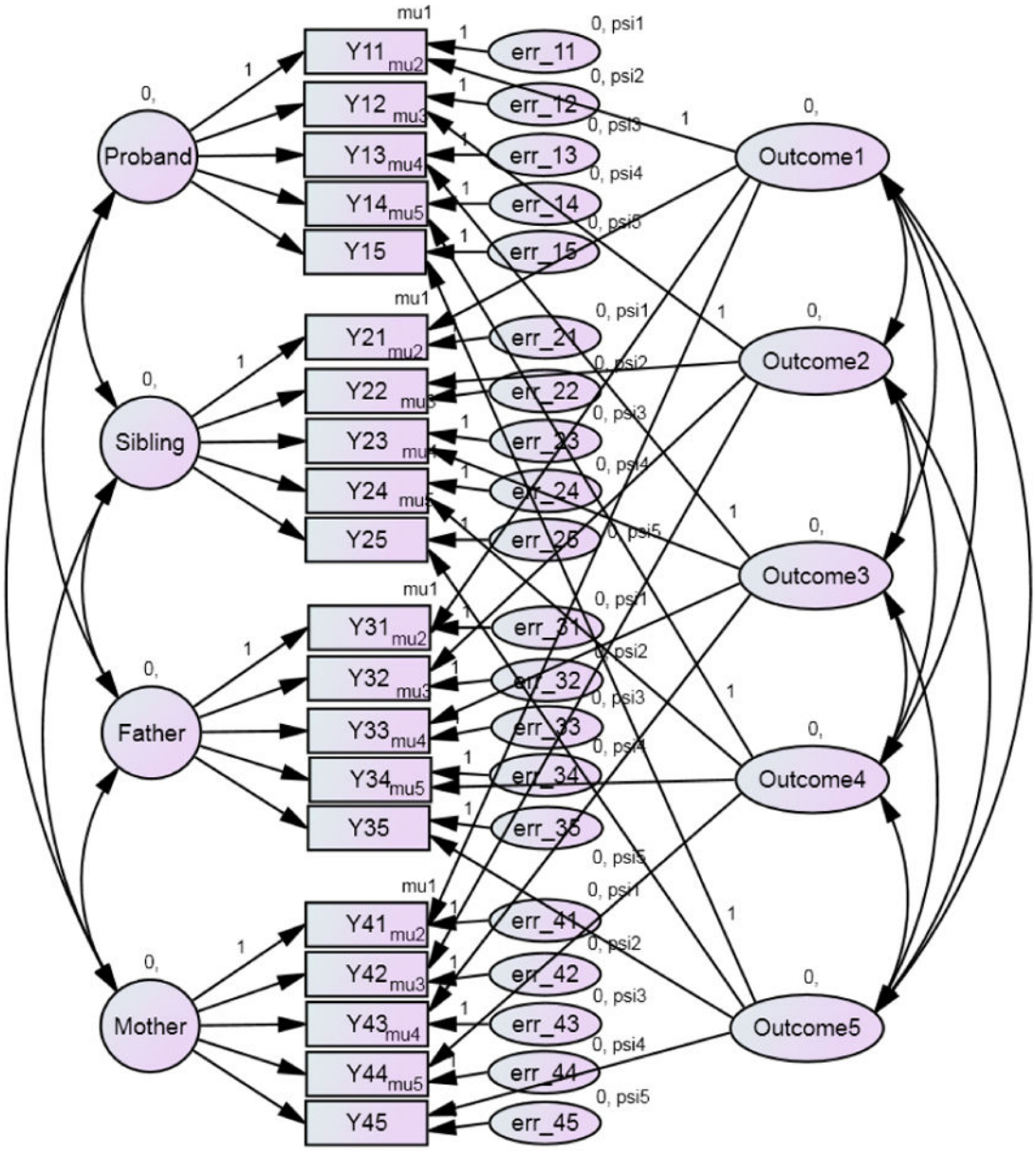

Figure 1.

A path diagram for the Bayesian Family Factor Model (BFFM). Responses variables Y_{ijk} labeled Yjk omitting the *i* (rectangles in the middle), for $j = 1, \ldots, 4$ and $k = 1, \ldots, 5$, are caused by two sets of factors, family member specific factors (circles on the left) and outcome specific factors (ovals on the right), along with a residual error, err_jk, which is unique to each item. Rectangles represent measured variables. Circles and ovals represent latent variables and residual errors. Covariances are represented by bidirectional arrows. Causal effects are represented by single-headed arrows in the path diagram. Means and

variance parameters are labeled on the rectangles and ovals/circles, before and after commas, respectively.

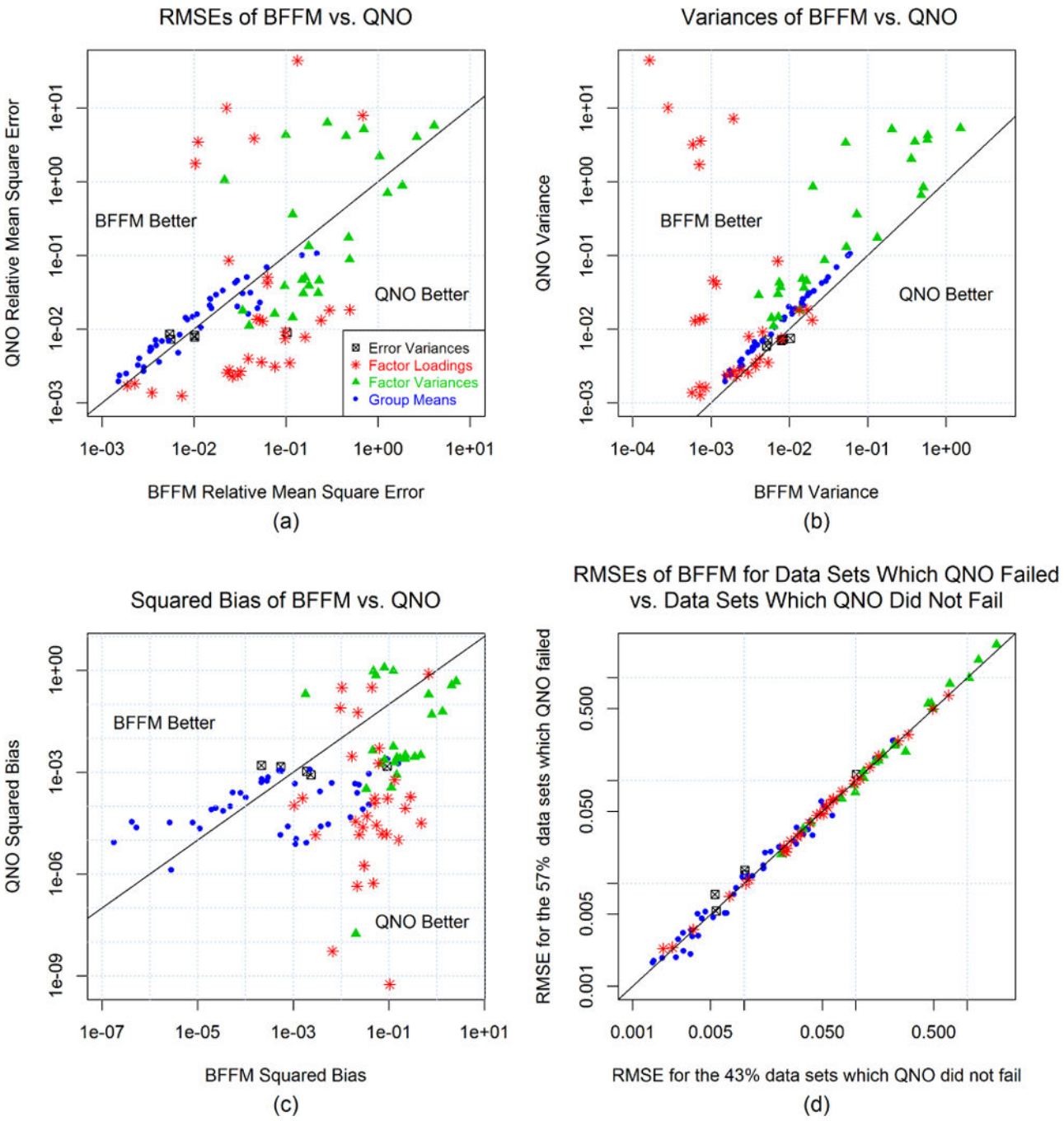

Figure 2.

Plots of relative mean squared errors (RMSE, a), relative variances (b) and relative squared biases (c) for parameters estimated by BFFM against those estimated by FIML-QNO, and plot of the relative mean squared errors by BFFM for the 43% of the data sets which FIML-QN failed vs for the 57% of the data sets which FIML-QN was successful (d), in the scenario where the true covariance matrix is close to not identified, on a log-log scale.

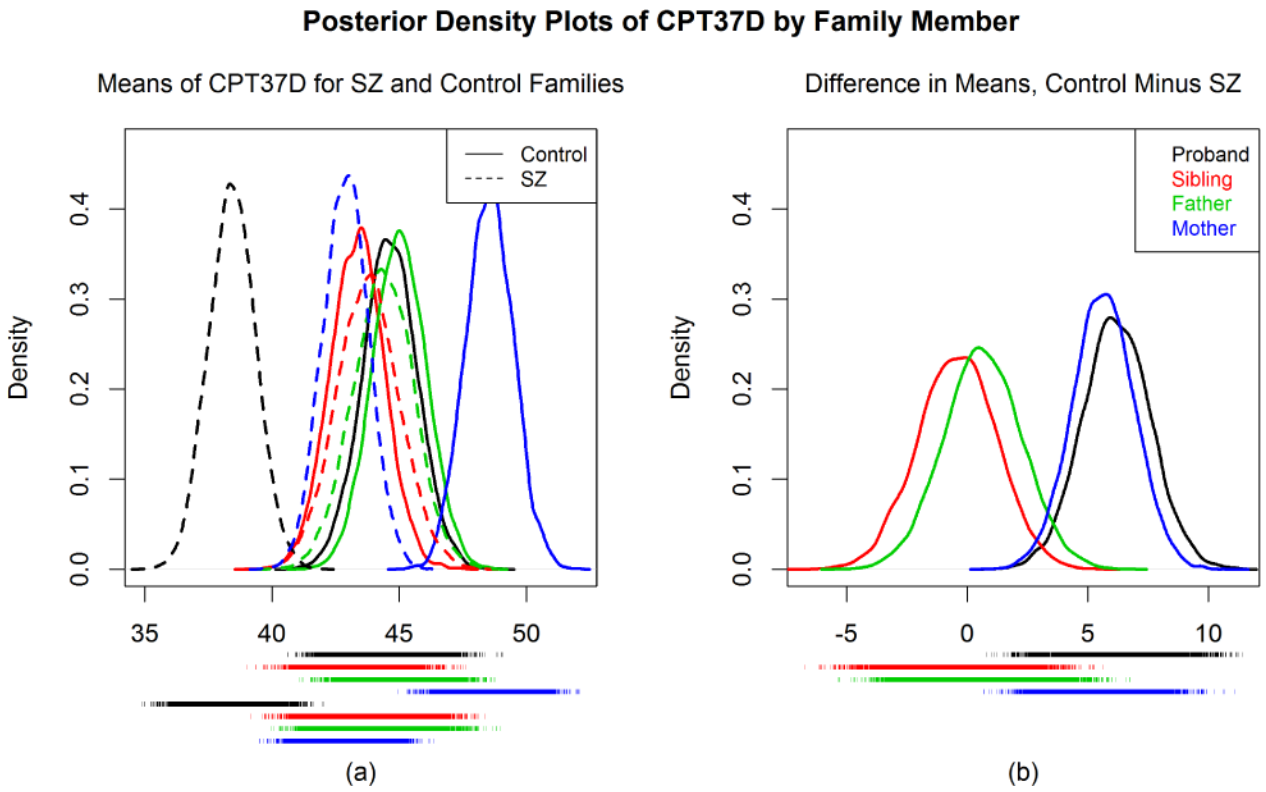

Figure 3.

(a) Posterior density of means of CPT37D for probands, siblings, fathers and mothers in the control (black) and SZ (grey) families. The 8 1-dimensional density plots at the bottom represent locations of posterior samples for probands, siblings, fathers and mothers in control and SZ families, from top to bottom. (b) Posterior densities for differences in means of CPT37D between two groups, control minus SZ. The 4 1-dimensional density plots at the bottom represent locations of posterior samples for probands, siblings, fathers and mothers, from top to bottom.

Table 1

Descriptions, transformations and summary of prior information for the five outcome

The first four outcomes are scaled while the logTRLBA with a skewed distribution and negatively correlated with other outcomes are log-transformed and has its sign reversed.

k	Variable (Description)	Transformation	Mean	SD	SD Ratio
1	MANIPA (Maintenance and Manipulation test accuracy during manipulation)	$100 * y$	70	13	1
2	CPTDSD (Degraded Stimulus Continuous Performance Test block sum d prime)	$10 * y$	28	11	0.85
3	CPT37D (Memory-load Continuous Performance Test block sum d prime)	$10 * y$	41	9	0.69
4	SPAN10 (Forced-choice Span of Apprehension test 10-letter accuracy)	$100 * y$	50	5	0.38
5	logTRLBA (Trail making test b time in sec)	$-100 * \log_{10}(y)$	−140	20	1.54

Estimates of overall means and standard deviation (SD) are obtained from Phase 1 data of the UCLA Family Study and from previous literature [54–56]. SD ratio is the ratio of the SD of an outcome to the SD of the first outcome.

Table 2

Posterior means of factor variances, factor correlations, factor loadings and unique error variances estimated by BFFM.

Family Member		**Proband**	**Sibling**	**Father**	**Mother**	
Factor Variance		**50.57**	**50.81**	**33.31**	**37.14**	
Factor Correlations	Proband	1.000				
	Sibling	0.390 *	1.000			
	Father	0.173	0.070	1.000		
	Mother	0.081	0.034	0.104	1.000	
Factor Loadings	MANIPA	1	1	1	1	
	CPTDSD	0.88 *	0.61 *	0.92 *	0.77 *	
	CPT37D	1.01 *	0.99 *	0.73 *	0.90 *	
	SPAN10	0.61 *	0.46 *	0.47 *	0.52 *	
	logTRLBA	2.13 *	2.05 *	1.77 *	2.36 *	
Outcome		MANIPA	CPTDSD	CPT37D	SPAN10	logTRLBA
Factor Variance		50.20	24.07	16.70	4.67	97.69
Factor Correlations	MANIPA	1.000				
	CPTDSD	0.594 *	1.000			
	CPT37D	0.493 *	0.609 *	1.000		
	SPAN10	0.349 *	0.354 *	0.291	1.000	
	logTRLBA	0.590	0.526	0.529	0.470	1.000
Factor Loadings	Proband	1	1	1	1	1
	Sibling	1.18 *	1.23 *	0.84 *	1.52 *	0.81 *
	Father	0.18	0.66 *	0.60 *	1.08 *	0.84 *
	Mother	0.54 *	0.65 *	0.65 *	0.98 *	0.76 *
Unique Error		MANIPA	CPTDSD	CPT37D	SPAN10	logTRLBA
Variances		150.17	65.41	28.57	15.77	229.25

For a parameter with *, the posterior probability of being negative, is less than 0.05, $p(\theta < 0/\mathbf{Y}) < 0.05$.

Table 3

Posterior means of regression coefficients for $J = 4$ family members and $K = 5$ outcomes per family in the control and schizophrenia (SZ) families (top), and the differences between groups (bottom).

Group	**Control**				**SZ**			
Outcome	**Proband**	**Sibling**	**Father**	**Mother**	**Proband**	**Sibling**	**Father**	**Mother**
MANIPA	76.01	70.25	77.00	74.39	66.56	72.37	75.59	71.84
CPTDSD	25.27	22.85	20.83	26.07	21.38	23.44	24.02	23.25
CPT37D	44.63	43.36	45.01	48.55	38.44	43.77	44.36	42.90
SPAN10	56.05	54.79	53.88	53.75	53.32	55.33	51.13	51.68
logTRLBA	−139.4	−143.0	−139.5	−140.8	−151.1	−145.1	−150.0	−151.8

	Difference (Control – SZ)			
Outcome	Proband	Sibling	Father	Mother
MANIPA	9.46 *	−2.12	1.41	2.55
CPTDSD	3.89 *	−0.59	−3.19	2.82 *
CPT37D	6.19 *	−0.42	0.65	5.65 *
SPAN10	2.73 *	−0.54	2.75 *	2.06 *
logTRLBA	11.8 *	2.1	10.5 *	11.0 *

For a parameter with *, the posterior probability being smaller than zero, is smaller than 0.05, $p(\theta < 0 | \mathbf{Y}) < 0.05$.