

DECOMPOSING INDIVIDUAL AND GROUP
DIFFERENCES OF CATEGORICAL
VARIABLES WITH GENETIC FACTOR
MODEL

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts

By
SEUNG BIN CHO

Dr. Phillip K. Wood, Thesis Supervisor

AUGUST 2007

The undersigned, appointed by the dean of the Graduate School, have examined the
thesis entitled

DECOMPOSING INDIVIDUAL AND GROUP DIFFERENCES OF
CATEGORICAL VARIABLES WITH GENETIC FACTOR MODEL

presented by Seung Bin Cho,

a candidate for the degree of master of arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Phillip Wood

Professor Denis McCarthy

Professor Joan Hermsen

ACKNOWLEDGEMENTS

First of all, this thesis would not have been completed without enlightening inspiration, support, and patience of Dr. Phillip K. Wood, my thesis advisor. Also, I would like to thank my committee members, Dr. Denis McArthy and Dr. Joan Hermsen for their helpful comments and suggestions and sharing their precious time spent reviewing my thesis.

Finally, thanks to my wife for her love and endurance and, most of all, I would like to thank my parents for their endless support and love that I can always rely on.

This research was supported by grants to KJS (R37 AA07231) and ACH (P50 AA11998) from the National Institute on Alcohol Abuse and Alcoholism.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
ABSTRACT.....	vi
Chapter	
1. INTRODUCTION.....	1
2. METHOD.....	5
Description of Basic Genetic Factor Model	
Application of Latent Response Variable Formulation	
Decomposing the Group Differences of Ordered Categorical Variables	
3. ILLUSTRATIVE EXAMPLE.....	23
Data from Australian Twin Registry	
The Model Description	
Results	
4. DISCUSSIONS.....	35
5. REFERENCES.....	40
APPENDIX	
1. APPENDIX A.....	46
2. APPENDIX B.....	51
3. TABLES.....	53
4. FIGURES.....	63

LIST OF TABLES

Table	Page
1. Items Included in Each Dimension	53
2. Estimated means and 95% confidence intervals	54
3. Result Summary for <i>Racial Prejudice</i> dimension.....	55
4. Result Summary for <i>Religious Fundamentalism</i> dimension.....	56
5. Result Summary for <i>Right-wing Orientation</i> dimension	57
6. Result Summary for <i>Social Permissiveness</i> dimension	58
7. Estimated means of latent continuous variables for males.....	59
B1. The excerpt of Mplus program.....	60

LIST OF FIGURES

Figure	Page
1. Univariate genetic factor model	63
2. Multivariate genetic factor model	64
3. Relationship between y_i^* and y_i	65
4. Different thresholds are estimated for different response proportions based on same distribution.....	66
5. Different mean and variances are estimated based on same thresholds.....	67
6. Plot of composite score and age	68
7. The path-diagram of the model used in example	69
8. Plot of the mean and its 95% confidence interval of common factor for males	70

Abstract

The method of decomposing individual and group differences on ordered categorical variables into genetic and environmental factors is introduced. Although the genetic factor model is used to identify the relative contributions of genetic and environmental factors on individual differences of phenotypic behavior, Dolan (1989) and Dolan, Molenaar, and Boomsma (1992) developed the method to decompose phenotypic means of variables with the genetic and environmental factors. Often, psychological constructs are measured by ordered categorical variables which can restrict the application of genetic factor model. Assuming an ordered categorical variable as a discretization of underlying latent continuous variable, the latent continuous variable can be modeled by applying Latent Response Variable (LRV) formulation. The present paper proposes a genetic factor decomposition similar to that proposed by Dolan and colleagues but for ordered polytomous variables in which mean structures of phenotype indicators are included. An example of the approach is described using data from Australian Twin Registry on conservatism scale. Mean differences between men and women are modeled with genetic and environmental factors. Although the patterns of factor loadings and factor means are different for each sub-dimension of conservatism, the model with means fits well to each sub-dimension which indicates that the source of individual and group differences of conservatism measure are similar.

Decomposing Individual and Group Differences of Categorical Variables with Genetic Factor Model

The genetic factor model, also known as ACE model, is used in behavior genetics to determine the relative contributions of genetic and environmental components on phenotypic behavior (Martin & Eaves, 1977; Neale & Cardon, 1992). In its simplest form, the genetic factor model, as a structural equation model of variance components, is a variance decomposition of the variance/covariance matrix from observed genetically informative data. Although the genetic factor model is often used to explain a single phenotypic behavior of interest, the genetic, shared environmental and unique environmental effects can be estimated using analogous latent variables when multiple indicators of the construct of interest are available (Heath, Eaves, & Martin, 1989; Heath, Neale, Hewitt, Eaves, & et al., 1989; Neale & Cardon, 1992). Elaborations of the genetic factor models are used to further refine estimated variance components to adjust for a variety of other factors influencing the analysis such as assortative mating and extended families (Eaves et al., 1999; Truett et al., 1994) or gene-environment interactions (Dick, Rose, Viken, Kaprio, & Koskenvuo, 2001; Neale & Cardon, 1992; Purcell, 2002). Although useful in determining the relative proportion of variance due to genetic and environmental effects on individual differences of phenotype behaviors, these genetic variance component models are not without their limitations. Specifically, without including the mean structure into the model genetic factor model misses the possibility of modeling the between group variations of the levels of phenotype behavior or constructs with genetic and environmental factors.

In order to model group differences of the levels of phenotype behavior or constructs, the information about the elevation or drop of the levels of phenotype variables should present in the form of the differences of factor means in the data. Although this may be accomplished in some models by including estimates of elevation at the individual level, this is more often accomplished within structural models by basing the analysis on the adjoined sum of squares and cross-products (SSCP) matrix instead of the variance/covariance matrix. For example, Dolan (1989) proposed an identified model for the genetic factor analysis model based on the SSCP in which the intercepts of phenotypic indicators are set to zeroes in a single group analysis and factor means are freely estimated. The model in Dolan (1989) is on the assumptions that phenotype indicators are in commensurate scales, which is not necessarily correct (Dolan, Molenaar, & Boomsma, 1992). Fixing this problem intercepts of phenotype indicators and factor means should be estimated but both intercepts and factor means cannot be identified simultaneously. Dolan, Molenaar, & Boomsma (1992) solved this problem by estimating the factor means as contrast to the reference group in which factor means are set to zeroes in multiple group setting, as suggested in Sörbom (1974). This model has the particular advantage of allowing the researcher to simultaneously estimate both the means of latent variable level as well as intercepts of the individual phenotype indicators. Such models permit the researcher to investigate the direction and extent of change in latent construct due to genetic and environmental factors without having to assume that the commensurate scales of individual phenotypic indicators, which is not always correct to assume in many cases.

As mentioned above, one of the analytic limitations of the models based on variance/covariance matrix considered so far is the assumption that the manifest indicators under consideration are continuous, and are multivariate normally distributed. Many measurements used in behavior genetics, however, involve simple categorizations (e.g., diagnosis status) or represent, at best, ordered categorical scales. The data measured by ordered categorical data have been analyzed using polychoric correlation matrix as input data for structural equation model estimation. Polychoric correlation matrix is the correlation matrix among the latent continuous variables that are assumed to underlie observed categorical variables. Even though this approach has been used in previous studies with evidently satisfactory results (Eaves et al., 1999; Loehlin, 1993; Truette et al. 1992), there are some limitations of using polychoric correlation matrix as input data. It has been known that estimated polychoric correlation matrices are not necessarily positive semi definite (Wothke, 1993) and no constraints can be applied on the model parameters, such as thresholds parameters. Also, no information regarding the group differences of the levels of latent continuous variables because, using polychoric correlation matrix, the distributions of latent continuous variables are assumed to be standard normal. Increasingly in structural equation models, *Latent Response Variable* (LRV) formulations, also known as threshold model or latent trait analysis, are used to specify structural equation models for variables which are measured as ordered polytomies. More detailed description of LRV formulation can be found in Muthén & Asparouhov (2002) and Skrondal & Rabe-Hesketh (2004). The LRV formulation assumes the latent continuous variables that underlie the observed ordered categorical variables of interest, and enables the application of structural equation modeling on

ordered polytomous variables by relating the latent continuous variables with observed polytomous variables with thresholds and distributional assumption for latent continuous variables. Also, explicit constraints on thresholds and other parameters that are assumed for genetic factor model can be applied using LRV formulation. Constraints on the threshold parameters are especially important in current study because thresholds should be constrained to be the same across twins in the same pair (Prescott, 2004) and threshold constraints across groups enables the estimation of level differences of latent continuous variables among different groups. The distribution of latent continuous response variables can take any form of distribution, but, usually, the distribution of latent continuous variable is assumed to be normal and the probit function, the inverse of cumulative normal distribution function, is employed as a link function. Alternatively, the logit function, can be used and probit function and logit function give almost identical result with appropriate scaling factor (Embretson & Reise, 2000).

In this paper, we propose a genetic factor model of ordered categorical data that can be used to decompose the level differences between groups of the construct measured by ordered categorical variables. In contrast to the case of continuous variables, mean differences of underlying latent continuous variables should be estimated based on the differences of proportions of responses on categorical variables. Unlike the polychoric correlation approach, latent response variable formulation enables estimation of different means of latent continuous variables across groups with equality constraints of thresholds. After introducing models which include the mean for categorical data, these models are applied to data from Australian Twin Registry on conservatism scale (Wilson & Patterson, 1968) which were originally proposed based on analysis of a polychoric correlation

matrix (Eaves et al., 1999; Martin, Eaves, Heath, Jardine, & Feingold, 1986). The responses from men and women are compared and the difference is decomposed with the same genetic and environmental factors used for decomposing individual differences within groups.

Method

Description of basic genetic factor model

The genetic factor model, shown in Figure 1, is an application of structural equation model used in the analysis of data gathered on twin pairs designed to assess the relative contributions of genetic and environmental constructs on phenotypic behavior (Neale & Cardon, 1992). This model is often called as ACE model because three constructs are supposed to explain the individual differences in phenotypic behaviors; additive genetic, common environmental, and individual-specific environmental factors, denoted by A, C, and E, respectively.

In this model, each twin pair's phenotype behavior is determined by additive genetic, common environmental, and unique environmental factors according to the following measurement model;

$$Y_1 = \lambda_A A_1 + \lambda_C C_1 + \lambda_E E_1$$

$$Y_2 = \lambda_A A_2 + \lambda_C C_2 + \lambda_E E_2$$

in which Y_1 and Y_2 are the observed phenotype behavior for each twin, and A , C , and E represent the additive genetic factor, common environmental factor, and individual-specific factors, respectively. Correlations among latent factors within each individual are assumed to be zero and unit variances of the latent variables are assumed to secure an

identified solution. λ_A , λ_C , λ_E are factor loadings from additive genetic, common environmental, and individual specific environmental factors, respectively. Factor loadings for each factor are constrained to be the same across twins. The genetic factor model may be written in matrix form as follows:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \lambda_A & \lambda_C & \lambda_E & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_A & \lambda_C & \lambda_E \end{pmatrix} \begin{pmatrix} A_1 \\ C_1 \\ E_1 \\ A_2 \\ C_2 \\ E_2 \end{pmatrix} \quad (1)$$

Because the factors within a same twin are assumed to be uncorrelated each other the variance of variable y_j , σ_j^2 , is determined by sum of square of factor loadings from each factor:

$$\sigma_j^2 = \lambda_A^2 + \lambda_C^2 + \lambda_E^2 \quad (2)$$

In the genetic factor model, the correlation between shared environmental factors is assumed to be one and unique environmental factors are assumed to be uncorrelated between twin pairs. Because monozygotic (MZ, identical) twins share the same genotype and dizygotic (DZ, fraternal) twins have half of their genotype in common (Neale & Cardon, 1992) correlations between twins due to additive genetic factors are assumed to be 1 for MZ twins and .5 for DZ twins. Thus the correlation structure among genetic and environmental factors is:

$$\begin{array}{c}
A_1 \ C_1 \ E_1 \ A_2 \ C_2 \ E_2 \\
A_1 \left(\begin{array}{cccccc}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
r & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{array} \right)
\end{array} \tag{3}$$

If these assumptions are substituted into the genetic factor model, it can be shown that the covariances between different types of twin pairs are;

$$\sigma_{MZ} = \lambda_A^2 + \lambda_C^2, \text{ for MZ twin pairs and}$$

$$\sigma_{DZ} = 0.5\lambda_A^2 + \lambda_C^2, \text{ for DZ twin pairs.}$$

This model forms a modeled covariance matrix between twins as:

$$\left(\begin{array}{cc} \lambda_A^2 + \lambda_C^2 + \lambda_E^2 & \lambda_A^2 + \lambda_C^2 \\ \lambda_A^2 + \lambda_C^2 & \lambda_A^2 + \lambda_C^2 + \lambda_E^2 \end{array} \right) \text{ for MZ twin pairs, and}$$

$$\left(\begin{array}{cc} \lambda_A^2 + \lambda_C^2 + \lambda_E^2 & .5\lambda_A^2 + \lambda_C^2 \\ .5\lambda_A^2 + \lambda_C^2 & \lambda_A^2 + \lambda_C^2 + \lambda_E^2 \end{array} \right) \text{ for DZ twin pairs.}$$

In this model, error variance cannot be separated from unique environmental factor because, mathematically, unique environmental factor is equivalent to error variance.

Therefore random error variance cannot be separated from systematic variance due to

unique environmental factor. Error variance can be separated from unique environmental factor in multivariate genetic factor model which is described below. Genetic factor model described above is based on several assumptions (Neale & Cardon, 1992). First, no correlations are assumed among additive genetic, common environmental, and unique environmental factors within a twin. Second, it is assumed that there is no genotype \times environment interaction. Finally, random mating of parents is assumed.

More often, a psychological construct is measured by several indicators than by a single variable and relative contribution of genetic and environmental components can be estimated by multivariate extension of univariate genetic factor model. Suppose a set of phenotype indicators that measure one phenotype behavior. Each observed variable can be represented as a function of genetic and environmental factors. For example, supposing four phenotype indicator variables, this model can be represented in matrix form as:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{pmatrix} = \begin{pmatrix} \lambda_{A1} & \lambda_{C1} & \lambda_{E1} & 0 & 0 & 0 \\ \lambda_{A2} & \lambda_{C2} & \lambda_{E2} & 0 & 0 & 0 \\ \lambda_{A3} & \lambda_{C3} & \lambda_{E3} & 0 & 0 & 0 \\ \lambda_{A4} & \lambda_{C4} & \lambda_{E4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{A1} & \lambda_{C1} & \lambda_{E1} \\ 0 & 0 & 0 & \lambda_{A2} & \lambda_{C2} & \lambda_{E2} \\ 0 & 0 & 0 & \lambda_{A3} & \lambda_{C3} & \lambda_{E3} \\ 0 & 0 & 0 & \lambda_{A4} & \lambda_{C4} & \lambda_{E4} \end{pmatrix} \begin{pmatrix} A_1 \\ C_1 \\ E_1 \\ A_2 \\ C_2 \\ E_2 \end{pmatrix} + \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} \quad (4)$$

where y_{11}, \dots, y_{15} and y_{21}, \dots, y_{24} represent the observed variable from twin 1 and twin 2, respectively. Unlike the univariate model, multivariate genetic factor model has

additional error term, θ_j , which is specific for each variable and assumed to be equal across each twin pair. This error term is distinguished from unique environmental factor which is supposed to affect the construct measured by phenotype indicators specific for each twin. The genetic and environmental common factors have the same covariance structure as in univariate genetic factor model, as in Equation 3, and it makes the covariance structure between y_{j1} and y_{j2} for monozygotic twins as:

$$\begin{pmatrix} \lambda_{A_j}^2 + \lambda_{C_j}^2 + \lambda_{E_j}^2 + \theta_j & \lambda_A^2 + \lambda_C^2 \\ \lambda_A^2 + \lambda_C^2 & \lambda_{A_j}^2 + \lambda_{C_j}^2 + \lambda_{E_j}^2 + \theta_j \end{pmatrix} \quad (5)$$

and for dizygotic twins as:

$$\begin{pmatrix} \lambda_{A_j}^2 + \lambda_{C_j}^2 + \lambda_{E_j}^2 + \theta_j & .5\lambda_A^2 + \lambda_C^2 \\ .5\lambda_A^2 + \lambda_C^2 & \lambda_{A_j}^2 + \lambda_{C_j}^2 + \lambda_{E_j}^2 + \theta_j \end{pmatrix} \quad (6)$$

Figure 2 illustrates this model in path notation form. The assumptions applied in univariate model also apply to the multivariate genetic factor model.

The model described above is mainly focused on the decomposition of individual differences, in terms of variance/covariance structure, into genetic and environmental factors. However mean structure of phenotype indicators can be incorporated in genetic factor model. Dolan et al. (1992) included mean structure of phenotype indicators by adding intercepts of variables and estimating the factor means as contrasts from the reference group. By adding intercepts and group indicator (k) Equation 4 becomes as:

$$\begin{pmatrix} y_{11}^{(k)} \\ y_{12}^{(k)} \\ y_{13}^{(k)} \\ y_{14}^{(k)} \\ y_{21}^{(k)} \\ y_{22}^{(k)} \\ y_{23}^{(k)} \\ y_{24}^{(k)} \end{pmatrix} = \begin{pmatrix} \alpha_1^{(k)} \\ \alpha_2^{(k)} \\ \alpha_3^{(k)} \\ \alpha_4^{(k)} \end{pmatrix} + \begin{pmatrix} \lambda_{A1} & \lambda_{C1} & \lambda_{E1} & 0 & 0 & 0 \\ \lambda_{A2} & \lambda_{C2} & \lambda_{E2} & 0 & 0 & 0 \\ \lambda_{A3} & \lambda_{C3} & \lambda_{E3} & 0 & 0 & 0 \\ \lambda_{A4} & \lambda_{C4} & \lambda_{E4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{A1} & \lambda_{C1} & \lambda_{E1} \\ 0 & 0 & 0 & \lambda_{A2} & \lambda_{C2} & \lambda_{E2} \\ 0 & 0 & 0 & \lambda_{A3} & \lambda_{C3} & \lambda_{E3} \\ 0 & 0 & 0 & \lambda_{A4} & \lambda_{C4} & \lambda_{E4} \end{pmatrix} \begin{pmatrix} A_1^{(k)} \\ C_1^{(k)} \\ E_1^{(k)} \\ A_2^{(k)} \\ C_2^{(k)} \\ E_2^{(k)} \end{pmatrix} + \begin{pmatrix} \theta_1^{(k)} \\ \theta_2^{(k)} \\ \theta_3^{(k)} \\ \theta_4^{(k)} \\ \theta_1^{(k)} \\ \theta_2^{(k)} \\ \theta_3^{(k)} \\ \theta_4^{(k)} \end{pmatrix} \quad (7)$$

where α_j is the intercept of variable y_j . Added to the variance/covariance structure, this model has the mean structure as:

$$\mu_j = \alpha_j + \lambda_{A_j} \kappa_A + \lambda_{C_j} \kappa_C + \lambda_{E_j} \kappa_E$$

where κ 's are the factor means. The variance/covariance structures for twin pairs are the same as in the model without mean structure, Equation 5 through 6, because the mean structure does not affect variance/covariance structure. However this model cannot be identified because there are more parameters to estimate than the number of the components in variance/covariance matrix. Dolan et al. identified this model by restricting the factor means as contrasts from a reference group in multiple group setting, which is suggested by Sörbom (1974). The factor means are set to be zeroes in reference group, and are estimated in non-reference group. Thus this model has mean structure in addition to the variance/covariance structure as follows.

$$\mu_j^{(r)} = \alpha_j^{(r)}, \text{ for reference group, and}$$

$$\mu_j^{(k)} = \alpha_j^{(k)} + \lambda_{A_j} \kappa_A + \lambda_{C_j} \kappa_C + \lambda_{E_j} \kappa_E, \text{ for non-reference group.}$$

Therefore the factor means estimated represent the differences of phenotype variables from a reference group due to genetic and environmental factors.

Application of Latent Response Variable Formulation

As mentioned above, the models discussed so far assume continuous observed variables drawn from a multivariate normal distribution. In many cases the psychological constructs of interest are measured by dichotomous or polytomous variables. Categorical variables violate the assumptions of continuity and multivariate normality that are needed in structural equation modeling and the direct application of structural equation modeling to categorical variables is usually misleading (Babakus, 1987; Johnon & Creech, 1983; Lubke & Muthén, 2004). When the categorical variable is ordinal it can be thought of as the discretized version of unobserved continuous response and Latent Response Variable (LRV) formulation provides the method that enables the modeling of the latent continuous response measured by ordered categorical variable within structural equation modeling (Christofferson, 1977; Muthén, 1984; Muthén & Asparouhov, 2002). In LRV formulation latent continuous variable that underlies each of the observed ordered polytomous variables is assumed and latent continuous variable and observed categorical variable are related using distributional assumption of latent continuous variable and thresholds. Denoting y_j^* as latent continuous variable that underlies corresponding observed ordered categorical variable y_j with C_j response categories, with single latent factor assumed for simplicity of illustration the measurement model for y_j^* is as follows:

$$y_j^* = \nu_j + \lambda_j \eta + \theta_j \quad (8)$$

From Equation 8 the mean and variance of y_j^* are given as:

$$\mu_j^* = \nu_j + \lambda_j \kappa \quad (9)$$

$$\sigma_j^{*2} = \lambda_j^2 \varphi + \theta_j \quad (10)$$

y_j^* is mapped onto y_j using threshold parameter τ 's.

$$y_j = \begin{cases} 1, & \text{if } y_j^* \leq \tau_{j,1} \\ 2, & \text{if } \tau_{j,1} < y_j^* \leq \tau_{j,2} \\ \vdots & \\ C_j - 1, & \text{if } \tau_{j,C_j-2} < y_j^* \leq \tau_{j,C_j-1} \\ C_j, & \text{if } \tau_{j,C_j-1} < y_j^* \end{cases} \quad (11)$$

where $\tau_{j,1}, \tau_{j,2}, \dots, \tau_{j,C_j-1}$ are thresholds which are cut-offs that discretize the continuous latent variable y_j^* into y_j . Figure 3 shows the mapping of y_j^* onto y_j with four ordered response categories. Combining Equation 8 through 11 the probability of the response of y_j is less or equal to category c can be expressed as:

$$P(y_j \leq c | \eta) = \Phi \left[\frac{\tau_c - (v_\phi + \lambda_j \eta)}{\sqrt{\theta_j}} \right] \quad (12)$$

and the probability of the response of y_j is c is that:

$$P(y_j = c | \eta) = \Phi \left[\frac{\tau_c - (v_\phi + \lambda_j \eta)}{\sqrt{\theta_j}} \right] - \Phi \left[\frac{\tau_{c-1} - (v_\phi + \lambda_j \eta)}{\sqrt{\theta_j}} \right] \quad (13)$$

where Φ is cumulative distribution function of standard normal distribution. This model introduces additional $\sum_{j=1}^p (C_j - 1)$ parameters of thresholds to be estimated. Thresholds can be estimated from the marginal distribution of y_j^* . Given that the distribution of y_j^* is normal with mean of μ_j^* and variance of σ_j^{*2} , thresholds can be obtained from,

$$P(y_j \leq c) = \int_{-\infty}^{(\sigma_j^*)^{-1}(\tau_c - \mu_j^*)} \phi_1(y_j^*) dy_j^* \quad (14)$$

where $\phi_1(\cdot)$ is standard univariate normal density function. Sample estimate of $P(y_j \leq c)$ is estimated from sample proportion of response on y_j is less or equal to c . Thus Equation 14 means that c -th threshold is z-score corresponding to the observed proportion of less or equal to c -th response category. However, from this formulation, thresholds cannot be identified unless the mean and variance of the distribution of y_j^* are not specified. The distribution of y_j^* is usually set to standard normal distribution to identify threshold in a

single group analysis. By setting the distribution of y_j^* as standard normal thresholds are estimated as z-score that corresponds to the observed cumulative response proportions:

$$\tau_c = \Phi^{-1}[p(y_j \leq c)] \quad (15)$$

where $\Phi^{-1}(\cdot)$ is the inverse of cumulative distribution function of standard normal distribution. By setting the σ_j^{*2} to one, unlike in the case of continuous variable, the residual variance of y_j^* , θ_j , is not a free parameter to be estimated because θ_j is obtained as residua from Equation 10:

$$\theta_j = 1 - \lambda_j^2 \varphi \quad (16)$$

where φ is the variance of factor η . Muthén and Asparaouhov (2002) parameterized the variance of y_j^* , in Mplus software, with the scaling factor Δ which is the inverse of σ_j^{*2} . Scaling factor is especially useful in multiple-group setting because it can be freed across groups to capture the difference of residual variance, factor loadings, and factor variances. Using the scaling factor instead of setting σ_j^{*2} to 1 residual variance is determined as:

$$\theta_j = \Delta_j^{-2} - \lambda_j^2 \varphi \quad (17)$$

The relationship between y^* variables is represented on polychoric correlation matrix.

Given the thresholds identified from Equation 14 or 15 and bivariate normality

assumption on y_i^* and y_j^* , the polychoric correlation coefficient between y_i^* and y_j^* can be estimated from:

$$P(y_j \leq c_j, y_i \leq c_i) = \int_{-\infty}^{(\sigma_j^*)^{-1}(\tau_{c_j} - \mu_j^*)} \int_{-\infty}^{(\sigma_i^*)^{-1}(\tau_{c_i} - \mu_i^*)} \phi_2(y_j, y_i) dy_j^* dy_i^* \quad (18)$$

where $\phi_2(\cdot)$ is bivariate standard normal density with zero mean vector and covariance matrix,

$$\begin{pmatrix} 1 & (\sigma_i^*)^{-1} \sigma_{ij}^* (\sigma_j^*)^{-1} \\ (\sigma_j^*)^{-1} \sigma_{ij}^* (\sigma_i^*)^{-1} & 1 \end{pmatrix} \quad (19)$$

where σ_{ij}^* is covariance between y_i^* and y_j^* , which is $\lambda_i \phi \lambda_j$. In terms of scaling factor Equation 19 can be written as:

$$\begin{pmatrix} 1 & \Delta_i \sigma_{ij}^* \Delta_j \\ \Delta_i \sigma_{ij}^* \Delta_j & 1 \end{pmatrix}, \quad (20)$$

Using LRV formulation genetic factor model can be applied to the ordered categorical variables (Prescott, 2004; Prescott & Kendler, 1999). Ordered categorical variables can be used in genetic factor model by applying the model in Equation 4 through 6 to the latent continuous variables:

$$\begin{pmatrix} y_{11}^* \\ y_{12}^* \\ y_{13}^* \\ y_{14}^* \\ y_{21}^* \\ y_{22}^* \\ y_{23}^* \\ y_{24}^* \end{pmatrix} = \begin{pmatrix} \lambda_{A1} & \lambda_{C1} & \lambda_{E1} & 0 & 0 & 0 \\ \lambda_{A2} & \lambda_{C2} & \lambda_{E2} & 0 & 0 & 0 \\ \lambda_{A3} & \lambda_{C3} & \lambda_{E3} & 0 & 0 & 0 \\ \lambda_{A4} & \lambda_{C4} & \lambda_{E4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{A1} & \lambda_{C1} & \lambda_{E1} \\ 0 & 0 & 0 & \lambda_{A2} & \lambda_{C2} & \lambda_{E2} \\ 0 & 0 & 0 & \lambda_{A3} & \lambda_{C3} & \lambda_{E3} \\ 0 & 0 & 0 & \lambda_{A4} & \lambda_{C4} & \lambda_{E4} \end{pmatrix} \begin{pmatrix} A_1 \\ C_1 \\ E_1 \\ A_2 \\ C_2 \\ E_2 \end{pmatrix} + \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} \quad (21)$$

Latent continuous variable y_j^* s' are determined by genetic and environmental factors and, from the Equation 12 and 13, the probability that the response on the variable y_j is less than c -th category given genetic and environmental factor is as follows:

$$P(y_j \leq c | A, C, E) = \Phi \left[\frac{\tau_{j,c} - (\lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E)}{\sqrt{\theta_j}} \right] \quad (22)$$

Thus the probability of the response on the variable y_j is c is:

$$\begin{aligned}
P(y_j = c | A, C, E) &= P(y_j \leq c | A, C, E) - P(y_j \leq c-1 | A, C, E) \\
&= \Phi \left[\frac{\tau_{j,c} - (\lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E)}{\sqrt{\theta_j}} \right] - \Phi \left[\frac{\tau_{j,c-1} - (\lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E)}{\sqrt{\theta_j}} \right] \quad (23)
\end{aligned}$$

The variance/covariance matrix between twin pair is the same as in the Equation 5 and 6, but the elements in variance/covariance matrix are in terms of latent continuous variables.

Estimating the model is usually based on the assumption of standardized latent

continuous variables. Different threshold for each standardized latent continuous variable and polychoric correlation matrix is estimated, and this information is used as input data of structural equation modeling. Resulting parameter estimates can be interpreted with the same fashion as in the model with continuous variables: factor loadings from genetic and environmental factors represent the effect of genetic and environmental factors on individual differences of phenotype behavior.

Decomposing the Group Differences of Ordered Categorical Variables

Although the application of behavior genetic model on ordered categorical variable as described above have been used to identify the relative contribution of genetic and environmental factors on individual differences of phenotypic behavior (Eaves et al., 1999; Heath, Eaves, & Martin, 1989; Prescott & Kendler, 1999; Truett, Eaves, Meyer, & Heath, 1992), use of polychoric correlation matrix as input data restricts the modeling of between group differences which needs mean structure in the model. Because analyzing the polychoric correlation matrix is equivalent to analyzing the standardized version of latent continuous variables there is no information regarding the elevation or drop of the levels of latent continuous variables across different groups. However the assumption of standardized variables can be relaxed with explicit constraints on thresholds, and the model described above can be modified to incorporate the mean structure. In this section incorporating mean structure with genetic factor model using ordered categorical variables is described. The levels of latent continuous variables can be modeled with the constraints on threshold parameters and it needs further consideration of model identification. The conditions of identification are also described.

In the case of continuous variables the change of item means can be decomposed with genetic and environmental factors by setting factor means in a reference group to be zero and estimating the differences of factor means from the reference group (Dolan, Molenaar, & Boomsma, 1992; Sörbom, 1974). However this cannot be directly applied to the case of ordered categorical variables without further consideration of additional parameters, thresholds and distributional parameters of latent continuous variables, in the analysis of ordered categorical variables. The changes in categorical variables are observed by different proportions of response categories. For example, in an item from the conservatism scales of Wilson and Patterson (1968) to be discussed in our example below, if women are more supportive of the item “bible truth” than men, the proportion answering “Yes” in women must be higher in women than in men. Under LRV formulation the proportional differences can be reflected either in threshold differences or parameters of latent continuous variables. With the same distributions for latent continuous variables the proportional difference between group would be reflected in different thresholds. Figure 4 illustrates different thresholds estimated for different proportion of responses. The upper panel of Figure 4 shows the thresholds based on standard normal distribution for the items with 3 ordered response categories with the cumulative proportions of 40% and 90%. The estimated thresholds are estimated as corresponding z-scores to the cumulative proportions as in Equation 13 and are -0.2533 and 1.2816. The lower panel shows the thresholds for the cumulative proportion of 20% and 80%. On the same distributional assumption, standard normal, for the latent continuous variable the thresholds are changed to -0.8416 and 0.8416. To model the proportional difference with genetic factor model, proportional differences should be

reflected in parameters of the distribution of latent continuous variables. Instead of fixing the distributions of latent continuous variables as standard normal, by constraining the thresholds across groups the differences of the distributional parameters of latent continuous variables can be estimated. Figure 5 illustrates distributional difference between the variables using the same cumulative proportions as used in Figure 5 (40% and 90% for upper panel, and 20% and 80% for lower panel) while thresholds are constrained to be the same. The thresholds in upper and lower panel are constrained to the values computed from standard normal distribution. The upper panel is standard normal distribution so that the mean is 0 and the variance is 1. The mean and variances in lower panel are computed based on constrained thresholds. With the thresholds constrained to be the same, the distribution of lower panel is moved to the right with mean of 0.5141 and the variance of 0.9119. This change in latent continuous variable can be modeled with the genetic factor model. Replacing the phenotype variable y' in Equation 7 with latent continuous variable y^* ,

$$\begin{pmatrix} y_{11}^{*(k)} \\ y_{12}^{*(k)} \\ y_{13}^{*(k)} \\ y_{14}^{*(k)} \\ y_{21}^{*(k)} \\ y_{22}^{*(k)} \\ y_{23}^{*(k)} \\ y_{24}^{*(k)} \end{pmatrix} = \begin{pmatrix} \alpha_1^{(k)} \\ \alpha_2^{(k)} \\ \alpha_3^{(k)} \\ \alpha_4^{(k)} \end{pmatrix} + \begin{pmatrix} \lambda_{A1} & \lambda_{C1} & \lambda_{E1} & 0 & 0 & 0 \\ \lambda_{A2} & \lambda_{C2} & \lambda_{E2} & 0 & 0 & 0 \\ \lambda_{A3} & \lambda_{C3} & \lambda_{E3} & 0 & 0 & 0 \\ \lambda_{A4} & \lambda_{C4} & \lambda_{E4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{A1} & \lambda_{C1} & \lambda_{E1} \\ 0 & 0 & 0 & \lambda_{A2} & \lambda_{C2} & \lambda_{E2} \\ 0 & 0 & 0 & \lambda_{A3} & \lambda_{C3} & \lambda_{E3} \\ 0 & 0 & 0 & \lambda_{A4} & \lambda_{C4} & \lambda_{E4} \end{pmatrix} \begin{pmatrix} A_1^{(k)} \\ C_1^{(k)} \\ E_1^{(k)} \\ A_2^{(k)} \\ C_2^{(k)} \\ E_2^{(k)} \end{pmatrix} + \begin{pmatrix} \theta_1^{(k)} \\ \theta_2^{(k)} \\ \theta_3^{(k)} \\ \theta_4^{(k)} \\ \theta_1^{(k)} \\ \theta_2^{(k)} \\ \theta_3^{(k)} \\ \theta_4^{(k)} \end{pmatrix}. \quad (24)$$

Reference group and non-reference group have same covariance structures as in Equation 5 and 6.

Unlike in the continuous variable case in Equation 7 the intercepts of each variable and factor means cannot be simultaneously identified even if the factor mean is restricted as the contrasts from the reference group because the phenotype indicators in this case are also latent variables, so the intercepts of each indicator are set to zeroes as in Dolan (1989). Setting the distributions of variables in reference group as standard normal it has added mean structure of latent continuous variables as,

$$\begin{aligned}\mu_j^{(r)*} &= 0, \text{ for reference group, and} \\ \mu_j^{(k)*} &= \lambda_{Aj}\kappa_A + \lambda_{Cj}\kappa_C + \lambda_{Ej}\kappa_E, \text{ for non-reference group}\end{aligned}\tag{25}$$

where, μ_j^* is the mean of latent continuous variable for observed j -th variable, κ are factor means, and λ are factor loadings that are the same across groups. The subscript for twins is omitted because twins in the same family would have same mean structure.

As pointed out by Dolan et al. (1992) setting intercepts of indicators as zeroes is equivalent to assuming that all the indicators are in commensurate scales, and it is not necessarily true in many cases. However, using latent response variable formulation, scale difference, if there is any, is absorbed into different thresholds estimates among variables. If the intercepts of variables are not set to be zeroes the measurement model for latent continuous variable y_j^* becomes as follow.

$$y_j^* = \alpha_j + \beta_j x_{age} + \lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E + \theta_j$$

Intercepts and thresholds cannot be identified simultaneously, so, with same reasoning for constraining the thresholds to be the same across the groups, thresholds should be constrained to be the same across the variables to identify intercepts of variables. Therefore the probability of the response is less or equal to the category c is determined as,

$$P(y_j \leq c | x_{age}, A, C, E) = \Phi \left[\frac{\tau_c - \alpha_j - (\beta_j x_{age} + \lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E + \theta_j)}{\sqrt{\theta_j}} \right] \quad (26)$$

The subscript j is omitted from τ , because τ 's have been set to be the same across indicator variables. From Equation 30 the threshold for c -th category of variable j is adjusted from τ_c by α_j , which is $\tau_c - \alpha_j$. Without the intercept and freely estimated thresholds the Equation 30 can be written as:

$$P(y_j \leq c | x_{age}, A, C, E) = \Phi \left[\frac{\tau_{cj} - (\beta_j x_{age} + \lambda_{Aj}A + \lambda_{Cj}C + \lambda_{Ej}E + \theta_j)}{\sqrt{\theta_j}} \right] \quad (27)$$

Subscript j is not omitted from τ_c because different thresholds are estimated across variables. Therefore the adjusted threshold from Equation 26, $\tau_c - \alpha_j$, and freely estimated threshold from Equation 27, τ_{cj} should be equivalent and the differences of intercepts across variables are absorbed to different estimates of thresholds.

Implementing this model the identification of the model parameters are more complicated than in continuous variable case due to the additional parameters of thresholds. The conditions of identifications should be considered based on the specific structure of model (Millsap, 2001; Millsap & Yun-Tein, 2004). The general identification conditions from Millsap and Yun-Tein (2004) can be applied with some modifications according the specific structure of genetic factor model. In genetic factor model factor variances are constrained to be 1 in all groups because correlation constraints are imposed for common factors. Other constraints are; a) the correlations between common factors within each twin are zero, b) the correlations of additive genetic and common environmental factor between twins in the same family are constrained to be 1 and 0.5, respectively, c) the correlation of unique environmental factor between twins from the same family is constrained to be zero, and d) factor loadings are invariant across twins and different relationships (monozygotic and dizygotic) of twins. With this conditions factor means and mean and variance of latent continuous variables can be compared across groups with following conditions. A) The distributions of latent continuous variables, y^* 's, underlying observed variables, y 's, are set to be standard normal distributions, i.e., the mean and variances, μ^* and σ^{2*} are set to be zero and one, respectively, in one reference group. B) One of the thresholds of all indicator variables is constrained to be invariant across group. C) For three variables one more threshold other than the thresholds constrained in condition B) is constrained to be invariant across groups. Condition C) implicates that at least three indicators are needed. D) Factor means, κ 's are set to be zeroes in a reference group. E) The factor loadings are constrained to be invariant across groups. Factor loadings set to be invariant across group makes the

variables measures the same common factors across groups. Conditions A), B), and C) are modified from the identification conditions from Millsap and Yun-Tein by which identify the distributions of latent continuous variables, and conditions D) and E) identify the distribution of latent factors as in Dolan et al. (1992). The detail of identification of the model with these conditions is described in Appendix A.

Illustrative Example

Data from Australian Twin Registry

The data collected from Australian Twin Registry on conservatism scale (Wilson & Patterson, 1968) is used for application of the model described above. 3808 pairs of twins are divided into 5 groups by their zygosity and gender. There are 1202 pairs of monozygotic female twins, 567 pairs of monozygotic male twins, 747 pairs of dizygotic female twins, 350 pairs of dizygotic male twins, and 912 pairs of opposite sex twin pairs. Conservatism scale developed by Wilson and Patterson consists of 50 items with three possible response categories each. Respondents are instructed to indicate whether they agree or not about the controversial topics by checking one of three response categories of “Yes”, “?”, and “No”. Three response categories are assumed to be ordered because the response categories can reflect the degree of a respondent’s supportive attitude on the item. Odd-numbered items are designed to have positive relationship with conservative attitude (e.g. White Superiority, Church Authority, etc.) and even-numbered of items are designed to have negative relationship with conservative attitude (e.g. White Superiority, Evolution Theory, etc.). Rather than analyzing all of fifty items in one model, items are factor analyzed beforehand to find sub-dimensions of one’s conservative attitude and each subset of items is analyzed separately. This approach has two main advantages. First,

it can allow different attitude and genetic structure for each sub-dimension of conservative attitude. Originally developed by Wilson and Patterson the conservatism scale used in this study based on the items collected from several sub-dimensions of conservative attitude and those sub-dimensions are consistently found in subsequent researches using factor analysis (Eaves et al., 1999; Truett, Eaves, Meyer, & Heath, 1992). One can have different attitudes on different sub-dimensions. For example one may have relatively liberal attitude on patriotic claims while he/she has relatively conservative attitude on religious claims. Moreover each sub-dimension may have different genetic structure. It is possible that shared environmental factor is more important in determining the conservative attitude on religious claims than in determining the conservative attitude on racial prejudice. Thus applying genetic factor model treating all the items as indicators of single conservatism attitude can be misleading because it cannot incorporate respondents' different attitude on different sub-dimensions and differential genetic structure for each dimension. Second, it can save computing time of running the model on structural equation software. Since estimating model parameters using ordered categorical variables is computing-intensive, increasing numbers of variable in analysis escalates computing time exponentially. Dividing items into several subsets and running separate analyses for each subset can save time and computing resource.

Exploratory factor analysis is performed on polychoric correlation matrix. Factor loadings are rotated to find meaningful pattern and four oblique sub-dimensions are found. Each sub-dimension is named considering the items included in each sub-dimension: a) *Racial Prejudice*, b) *Religious Fundamentalism*, c) *Right-wing Orientation*,

and d) *Social Permissiveness*. For each dimension items with absolute value of loadings over 0.4 are retained. Items included in each dimension is in Table 1 The patterns of items included in each identified dimensions are consistent with previous researches (Eaves et al., 1999; Truett, Eaves, Meyer, & Heath, 1992), and the characteristics on which Wilson and Patterson (1968) based when they developed the scale.

One of the traditional beliefs about the conservatism is that it changes over time: people become more conservative as they become older. Though Norval (1974) reviewed the empirical data of age change in conservatism, and pointed out that the age change of conservatism is not evident, there are evidence of age-cohort effect on conservatism (Eaves et al., 1999; Truett, 1993). While controlling possible covariates, Truett (1993) found strong cohort effect of age on Wilson and Patterson conservatism score that conservatism scores are larger in older respondents than in younger respondents and this escalate is more rapid after fifth decade of life. This may indicate that linear and quadratic effect of age should be used as covariate for conservatism measures (Eaves et al., 1999). In Figure 6, The sum of the scores of conservatism measures are plotted with age for each of four dimensions. Figure 6 shows that conservatism scores are larger in older respondents than in younger respondents, but it does not show dramatic change at the age between 40 and 50 as found by Truett. It may be due to the fact that the samples used in Truett and in this study are from different populations: Truett used data from the United States and data used in this study is from Australia. Truett pointed out that remarkable change at fifth decade of life can be due to political changes associated World War II, but Australian sample might not experience same change. Therefore, in this study, the linear effect of age is used as covariate for conservatism measures.

The Model Description

As described in *Latent Response Variable* section, a variable that measures of one's phenotype behavior is assumed to be discretization of underlying latent continuous variable. Since j -th observed variable y_j have 3 ordered categories and they are coded as 0, 1, and 2, j -th latent continuous variable $y_j^{*(g)}$ in group g is discretized by 2 thresholds $\tau_{j,1}$ and $\tau_{j,2}$.

$$y_j^{(g)} = \begin{cases} 0, & \text{if } y_j^{*(g)} \leq \tau_{j,1} \\ 1, & \text{if } \tau_{j,1} < y_j^{*(g)} \leq \tau_{j,2} \\ 2, & \text{if } \tau_{j,2} < y_j^{*(g)} \end{cases} \quad (28)$$

Superscript (g) represents group identity and, in this case, there two groups are compared: men and women. Group identities are omitted from $\tau_{j,1}$ and $\tau_{j,2}$ because thresholds are constrained to be the same across groups. Each of $y_j^{*(g)}$ is determined by its covariate and genetic and environmental factors.

$$y_j^{*(g)} = \alpha_j^{(g)} + \beta_j x_{age} + \lambda_{Aj} A^{(g)} + \lambda_{Cj} C^{(g)} + \lambda_{Ej} E^{(g)} + \theta_j^{(g)}. \quad (29)$$

A, C, and E are additive genetic factor, common environmental factor, and unique environmental factor, respectively. $\theta_j^{(g)}$ is residual variance of $y_j^{*(g)}$. Since the intercept of y_j^* is assumed to be zero, as discussed above, Equation 29 above becomes as,

$$y_j^{*(g)} = \beta_j x_{age} + \lambda_{Aj} A^{(g)} + \lambda_{Cj} C^{(g)} + \lambda_{Ej} E^{(g)} + \theta_j^{(g)}. \quad (30)$$

After the effect of age is covariates out, the mean of y_j^* is determined by genetic and environmental factors. Means of genetic and environmental factors are set to be zeroes for reference group, and are estimated in non-reference group. Because female groups are set to be reference groups the mean of y_j^* for female groups and male groups are determined as below.

$$E(y_j^{*(f)}) = 0 \text{ for female groups, and}$$

$$E(y_j^{*(m)}) = \lambda_{Aj} \kappa_A^{(m)} + \lambda_{Cj} \kappa_C^{(m)} + \lambda_{Ej} \kappa_E^{(m)} \text{ for male groups,}$$

where $\kappa_A^{(g)}$, $\kappa_C^{(g)}$, and $\kappa_E^{(g)}$ are the means of additive genetic factor, common environmental factor, and unique environmental factor, respectively. Because the variance of genetic and environmental factors are set to be zero, the variance of y_j^* is determined as,

$$\text{var}(y_j^{*(g)}) = \lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j^{(g)}$$

From Equation 28 and 30, with the normality assumption on y_j^* , the probability of the response to y_j falls into category c given respondents' age and genetic and environmental factors can be written as,

$$\begin{aligned}
P(y_j = c | x_{age}, A, C, E) &= P(y_j \leq c | x_{age}, A, C, E) - P(y_j \leq c-1 | x_{age}, A, C, E) \\
&= \Phi \left[\frac{\tau_{j,c} - (\beta_j x_{age} + \lambda_{Aj} A + \lambda_{Cj} C + \lambda_{Ej} E)}{\sqrt{\theta_j}} \right] \\
&\quad - \Phi \left[\frac{\tau_{j,c-1} - (\beta_j x_{age} + \lambda_{Aj} A + \lambda_{Cj} C + \lambda_{Ej} E)}{\sqrt{\theta_j}} \right]
\end{aligned} \tag{31}$$

where Φ is the cumulative distribution function of standard normal distribution and σ_j is the standard deviation of y_j^* . Superscript (g) is omitted from Equation 31 to avoid unnecessary complication of the equation. Given age the c -th threshold is adjusted as $\tau_{j,c} - \beta_j x_{age}$. After covariating out the effect of age y_j^* 's are modeled with genetic factor model. Let y_{j1}^* and y_{j2}^* be the latent continuous variables from the first and second twin in the same family.

$$\begin{pmatrix} y_{j1}^* \\ y_{j2}^* \end{pmatrix} = \begin{pmatrix} \beta \\ \beta \end{pmatrix} age + \begin{pmatrix} \lambda_{Aj} & \lambda_{Cj} & \lambda_{Ej} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{Aj} & \lambda_{Cj} & \lambda_{Ej} \end{pmatrix} \begin{pmatrix} A_1 \\ C_1 \\ E_1 \\ A_2 \\ C_2 \\ E_2 \end{pmatrix} + \begin{pmatrix} \theta_{j1} \\ \theta_{j2} \end{pmatrix}$$

Genetic and environmental factors have covariance structure as in Equation 3:

$$\begin{array}{c}
A_1 \ C_1 \ E_1 \ A_2 \ C_2 \ E_2 \\
\begin{array}{l}
A_1 \\
C_1 \\
E_1 \\
A_2 \\
C_2 \\
E_2
\end{array}
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
r & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}$$

where r is the correlation of additive genetic factors between twin pairs, which is 1 for monozygotic twin pairs and 0.5 for dizygotic twin pairs. Residual variance θ_{j1} and θ_{j2} are independent. Therefore the correlation matrix between y_{j1}^* and y_{j2}^* for monozygotic twin pair is

$$\begin{pmatrix}
\lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j & \lambda_{Aj}^2 + \lambda_{Cj}^2 \\
\lambda_{Aj}^2 + \lambda_{Cj}^2 & \lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j
\end{pmatrix},$$

and

$$\begin{pmatrix}
\lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j & 0.5\lambda_{Aj}^2 + \lambda_{Cj}^2 \\
0.5\lambda_{Aj}^2 + \lambda_{Cj}^2 & \lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j
\end{pmatrix}$$

for dizygotic twin pairs. The square of factor loading from each factor represents the proportion of variance of each variable explained by each factor. Path diagram of Figure 7 illustrates this model with four indicator variables. For illustrative purpose the left side of diagram is the model for female groups and right side of the diagram is the model for

male groups. The triangle is the imaginary variable whose mean and variance are set to one and zero to represent mean factor. Factor loadings, denoted by arrows, from the triangle to each factor represent the factor mean. Note that the factor means for female part are set to be zeroes, and are freed for male part. Latent continuous variables are connected with their observed categorical variables with filled circles to represent the probit transformation.

Mplus version 4 (Muthén & Muthén, 2006) is used to estimate model parameters. Estimation method is weighted least square using a diagonal weight matrix with standard errors and mean- and variance- adjusted chi-square test statistic that use a full weight matrix (WLSMV) which is the default estimation method in Mplus for a model with categorical variables and mean structure (Muthén & Muthén, 2006). Parameterization method is Delta parameterization which allows the estimation of scale factors while residual variance is determined based on the estimated scale factor (Muthén & Asparouhov, 2002). An example of Mplus program and detailed description is in Appendix B.

Results

Before applying the model the overall differences between men and women are estimated using multi-group confirmatory factor analysis. Table 2 and Figure 8 show the estimated mean and its 95% confidence interval of males on a single common factor extracted from the items included in each dimension. Because factor means for female are set to zeroes the means for males represent the differences of factor mean from females. It is shown that men have more conservative attitude on *Racial Prejudice* dimension, and have less conservative attitude on *Religious Fundamentalism, Right-wing*

Orientation, and *Social Permissiveness* dimensions. It turned out that the difference on Religious Fundamentalism dimension is not significant between men and women.

Table 3 through 6 summarizes the result of application of the model described above for each sub-dimension of conservatism. Even though chi-square test for model for each sub-dimension is significant due to the large sample size, other fit indices, TLI, CFI, and RMSEA, indicate that the model fits well to the data, which means that the individual differences and group differences can be explained by the same genetic and environmental factors. Relatively poorer fit within *Social Permissiveness* dimension (CFI = 0.868, TLI = 0.840, and RMSEA = 0.044) may imply that different factors operate on the individual and group differences. Poorer fit can be due to other source of model, such as factorial invariance, and examining a series of nested model would be helpful to find the source of misfit.

For each dimension, factor loadings indicate the contribution of each latent factor on the individual differences of responses on items because the variance of a latent continuous variable is determined by factor loadings and residual variance.

$$\sigma_j^2 = \lambda_{Aj}^2 + \lambda_{Cj}^2 + \lambda_{Ej}^2 + \theta_j$$

Therefore environmental or genetic factors which have a bigger factor loading than other factors on each item are supposed to be more important to determine the individual difference of the item. The signs of factor loadings inform the direction of how the factors work on conservative attitude. As mentioned in *Data Description* section, odd-numbered items are positively worded and even-numbered of items are negatively

worded for conservative attitude. Thus factors that are positively loaded on positively worded items and negatively loaded on negatively worded items are supposed to contribute more conservative attitude. In the same way, factors that are negatively loaded on positively worded items and positively loaded on negatively worded items are supposed to less conservative attitude.

Factor mean in each of Table 3 through Table 6 are the estimated factor means for males and they represent the mean difference on genetic and environmental factors between males and females because means of factors in female groups are set to be zeroes. Factor means inform the magnitude and the direction of the differences on factors between groups. The absolute values of factor means represent the magnitude of difference of genetic and environmental factors between men and women. The signs of factor means indicate the direction of the differences on factor means for male groups relative to female groups. A positive factor mean indicates that the level of the factor for male groups is above the level of the same factor in female groups and vice versa.

Because the factor loadings are constrained to be the same between men and women the estimated means of latent factors represent the elevation or drop of the levels of conservative attitude due to genetic and environmental factors. Estimated mean for men for each item is in Table 6. Mean of item j is computed as a linear combination of factor means:

$$\mu_j^* = \lambda_{Aj} \kappa_A^{(m)} + \lambda_{Cj} \kappa_C^{(m)} + \lambda_{Ej} \kappa_E^{(m)}$$

The scores of items in which the responses are negative direction are reversed in Table 6. Thus, in Table 6, higher scores represent more conservative attitude. The average of items included in each dimension on the last row show the consistent pattern of direction as the factor means from confirmatory factor analysis even though the magnitudes of differences are different.

For the *Racial Prejudice* dimension, in Table 3, factor loadings indicate a substantially larger effect of additive genetic factor than of shared or unique environmental factors on the variance of each item. Because all of three factors are positively loaded on positively worded items and negatively loaded on negatively loaded item, all of the factors work positively to conservative attitude. Factor mean of unique environmental factor is positive and the largest and the levels of factor means are negative in genetic and shared environmental factors. Thus men are characterized as less genetic and shared environmental factors which contribute to the higher conservative attitude and more unique environmental factor which contributes to higher conservative attitude. Even though the factor means of additive genetic factor and shared environmental factors are negative in men, the estimated means of most of the items are positive, except “Death Penalty” and “Apartheid”, and the average of the means of latent continuous variables is positive which indicate that men are more conservative on racial prejudice dimension.

For *Religious Fundamentalism* dimension, in Table 4, the items related with marriage and reproduction are affected relatively more by genetic factor and items related with religious claims are largely determined by environmental factors. The signs of factor loadings indicate that genetic factor and shared environmental factor are negative and

unique environmental factor is positive for conservative attitude. The mean of genetic factor is negative and the mean shared environmental factor is positive. The mean of unique environmental factor is negligible. Even though it turned out that the difference of common factor between men and women on this dimension from confirmatory factor analysis is not significant (Table 2), it is possible to find the significant difference of genetic and environmental factors because overall difference is decomposed into genetic and environmental factors. Because both genetic factor and shared environmental factor are negative for conservative attitude men are characterized as less genetic factor and more shared environmental factor which contribute to less conservative attitude. This pattern of factor means and factor loadings produce the means of latent continuous variables in Table 6. It shows that men are more conservative on the items related with reproductive rights and women are more conservative on religious claims. The average of means is slightly higher in men.

For *Right-wing Orientation* dimension, the variances of most items included are mainly determined by genetic factor and shared environmental factors except “Licensing Law” and “Censorship” which is mainly determined by environmental factors. Items included in this dimension are all positively worded and all three factors are positively loaded on each item, so three factors are positively functioning on conservative attitude. Factor means are negative and significantly different from zero for genetic and unique environmental factors and negligible for shared environmental factor. Therefore the difference between men and women is mainly determined by genetic factor and unique environmental factor and men are characterized as less genetic and unique environmental

factors that contribute to higher conservative attitude. Estimated means of all items are positive in men which indicate that men are less conservative on all items.

Factor loadings for items included in *Social Permissiveness* dimension do not show noticeable pattern of dominant factor(s). It may indicate that Social Permissiveness dimension can further be divided with its sub dimensions. Individual differences of responses of items are determined by three factors. Factor means are negative for genetic factor and unique environmental factor and the mean of shared environmental factor is negligible which means that the difference between men and women are mainly determined by genetic factor and unique environmental factor. Men can be characterized as less genetic factor that contribute to liberal attitude. Genetic and shared environmental factors are positively functioning for conservative attitude, but the interpretation of unique environmental factor is not as clear as for genetic factor because unique environmental factor has significant negative factor loadings on some positively worded items. This may indicate that unique environmental factor is a combination of several factors that have different effects on each item. Resulting means of items indicate that men are more conservative on most items.

Discussions

By integrating the model for incorporating the mean structure with genetic factor model, suggested by Dolan et al. (1992), and the Latent Response Variable formulation, the model described in this study enables the estimation of individual and group differences with genetic and environmental factors. The model from Dolan et al., which incorporates the method for estimating factor means suggested by Sörbom (1974), provides the method of estimating means of genetic and environmental factors as

contrasts to a specified reference group. LRV formulation is modified in this context to incorporate the mean structure of latent continuous variables. Differences of proportions on response categories between groups are converted into the differences of means of latent continuous variables and these mean differences of latent continuous variables are modeled by the genetic and environmental factors which are used to decompose individual differences within groups.

The main advantage of the model suggested in this study is the ability of testing the hypothesis regarding the origins of within and between group variations on phenotypic behavior that is measured by ordered categorical variables. As utilized and suggested by subsequent studies (Cleveland, Wiebe, Oord, & Rowe, 2000; Dolan & Molenaar, 1994; Heiman, Stallings, Hofer, & Hewitt, 2003; Rowe & Cleveland, 1996; Rowe & Rodgers, 1997), the model suggested by Dolan et al. (1992), on which current model is based, can be used to test hypothesis regarding the origins of within and between group variations measured by continuous variables. Conventional method of analyzing the ordered categorical variables with genetic factor model is not able to incorporate this model because the group differences cannot be estimated. Proposed model extends the utility of the model suggested by Dolan et al. (1992) to the constructs measured by ordered categorical variables by providing the method to estimate and model the differences across groups in terms of latent continuous variables underlying observed categorical variables. Because the same genetic and environmental factors are used to decompose within and between group variations, when the model with group differences fits the data, it can be interpreted that within and between group variations have common causal determinants (Cleveland, Wiebe, Oord, & Rowe, 2000; Rowe &

Cleveland, 1996; Rowe & Rodgers, 1997). In the example illustrated in this study, fit indices indicate that current model fits well to the data from with Australian Twin Registry on conservatism scale, which means that the same genetic and environmental factors operate on within and between group variations.

The framework of current model can also be applied to categorical longitudinal data. In the analyses of longitudinal data, such as the models suggested by McArdle and Hamagami (2003), McArdle (1986), and Dolan, Molenaar, and Boomsma (1991), it is required that the information about the differences among repeated measurements should present in the data. With ordered categorical variables these differences are represented in the differences of proportions of response categories between repeated measurements and they can be modeled in terms of the means of latent continuous variables with additional set of constraints on thresholds. Thus, to model the change over time, one of the repeated measurements would be set as a reference point and the difference from the reference point of the means of latent continuous variables can be estimated by applying the set of constraints of parameters, as applied across groups in current model, but across repeated measurements. Applying this framework on the longitudinal data from the same source of data as in the example of this study, conservatism measures from Australian Twin Registry, it would be possible to model the differences of latent continuous variables by setting the first measurement as a reference point in addition to the current setting of the model.

Although the magnitude and the direction of the difference of the levels of genetic and environmental factors between the groups that are being compared, they do not inform what constitute the estimated differences of each factor. For example, as

suggested from Truett et al. (1992), church attendance can be an important part of shared environmental factor for religious dimension and less mean of shared environmental factor in men may indicate that women tend to go to church more than men do. However it should be noted that factor means only provide the information about the comparative levels of factors across different groups and the substantive interpretation of factor means would require further information.

Some technical features of current model may restrict its use from wider range applications. First, as described in previous sections and Appendix A, at least three indicators for a phenotype construct are needed to identify the model. As in other types of factor analysis model, the number of variables needed is determined by the number of factors and the structure of the model. Considering the structure of current model, the same number of equations as the number of factors are needed to identify factor means. Because there are three factors, additive genetic, shared environmental, and unique environmental factors, at least three indicator variables are needed. Second, the identification conditions suggested that current model cannot be applied when the indicator variables are dichotomous. At least two thresholds should be identified to identify both mean and variance of latent continuous variables, but a dichotomous variable has only a single threshold. The model with dichotomous variable can be identified by setting the standard deviation of latent continuous variable to a certain value, usually one, as suggested in Millsap and Yun-Tein (2004). Finally, current model is based on the assumption of factorial invariance between groups to be compared which means that both men and women have same structures, factor loadings, same thresholds, and same factor variances. What vary across male and female groups are the factor means

and the residual variances. The interpretation of differences of factor means as the causal determinants of between group differences is meaningful provided the factorial invariance across the groups of interest. Factorial invariance can be tested with series of nested models. If the groups are not known in advance, latent class mixture approach can be combined with genetic factor model to find potential sub population by allowing the model parameters to vary across latent classes (Muthén & Asparouhov, 2006).

Another important consideration would include the selection of the link function for ordered categorical variables. Specifically, other types of distribution for latent continuous variables than normal distribution can be more appropriate for some variables. Using probit function as a link function current model assumes the distribution of latent continuous variables to be normal distributions, but, in some case, it may not be the case that the latent continuous variable is symmetric, let alone normally distributed. Future direction of current model would involve the assessment of different types of link function, such as the logit function, for different types of measurement variables. Logit function is readily available in Mplus and it may be more appropriate in research contexts such as a substance use and clinical diagnoses.

While some limitations apply the model proposed in this study is an advance over the conventional method within behavior genetics of using polychoric correlation matrix to analyze ordered categorical variables. By extending a well-known method of analyzing mean structure developed by Dolan et al. (1992) to be applied to the psychological constructs measured by categorical indicators, the model proposed in this study can provide methodological benefit for theory construction within behavior genetic context.

References

- Babakus, E. (1987). The Sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing research*, 24, 222-228.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Christoffersson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika* 42, 433-438.
- Cleveland, H., Wiebe, R. P., van den Oord, E. J., & Rowe, D. C. (2000). Behavior problems among children from different family structures: The influence of genetic self-selection. *Child Development*, 71, 733-751.
- Dick, D. M., Rose, R. J., Viken, R. J., Kaprio, J., & Koskenvuo, M. (2001). Exploring gene-environment interactions: Socioregional moderation of alcohol use. *Journal of Abnormal Psychology*, 110, 625-632.
- Dolan, C. V. (1989). LISREL analysis of twin data with structured means. *Behavior Genetics*, 19, 51-62.
- Dolan, C. V., Molenaar, P., & Boomsma, D. (1994). Simultaneous genetic analysis of means and covariance structure: Pearson-Lawley selection rule. *Behavior Genetics*, 24, 17-24.
- Dolan, C. V., & Molenaar, P. C. (1994). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research*, 29, 203-222.

- Dolan, C. V., Molenaar, P. C., & Boomsma, D. I. (1991). Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data. *Behavior Genetics*, 21, 49-65.
- Dolan, C. V., Molenaar, P. C. M., & Boomsma, D. I. (1992). Decomposition of multivariate phenotypic means in multigroup genetic covariance structure analysis. *Behavior Genetics*, 22, 319-335.
- Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K., et al. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30 000 study of twins and their relatives. *Twin Research*, 2, 62-80.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Erlbaum.
- Heath, A., Eaves, L., & Martin, N. (1989). The genetic structure of personality III. Multivariate genetic item analysis of the EPQ scales. *Personality and Individual Differences*, 10, 877-888.
- Heath, A., Neale, M., Hewitt, J., Eaves, L., & et al. (1989). Testing structural equation models for twin data using LISREL. *Behavior Genetics*, 19, 9-35.
- Heiman, N., Stallings, M. C., Hofer, S. M., & Hewitt, J. K. (2003). Investigating age differences in the genetic and environmental structure of the tridimensional personality questionnaire in later adulthood. *Behavior Genetics*, 33, 171-180.
- Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48, 398-407.

- Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling, 12*, 1-27.
- Loehlin, J. C. (1993). Nature, nurture, and conservatism in the Australian twin study. *Behavior Genetics, 23*, 287-290.
- Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11*, 514-534.
- Martin, N. G., & Eaves, L. J. (1977). The genetic analysis of covariance structure. *Heredity, 38*, 79-95.
- Martin, N. G., Eaves, L. J., Heath, A. C., Jardine, R., & Feingold, L. M. (1986). Transmission of social attitudes. *Proceedings of the National Academy of Sciences, 83*, 4364-4368.
- McArdle, J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*, 163-200.
- McArdle, J. J., & Hamagami, F. (2003). Structural equation models for evaluating dynamic concepts within longitudinal twin analyses. *Behavior Genetics, 33*, 137-159.
- Millsap, R. E. (2001). When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling, 8*, 1-17.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research, 39*, 479-515.
- Muthén, B. (1979). A Structural probit model with latent variables. *Journal of the American Statistical Association, 74*, 807-811.

- Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 43-65.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics*, 9, 313-324.
- Muthén, B., & Asparouhov. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes: No. 4, Version 5*
- Muthén, L., & Muthén, B. (2006). *Mplus User's Guide* (4th ed.). Los Angeles: Muthén & Muthén.
- Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*, Dordrecht, Netherland: Kluwer.
- Olsson, U. (1979a). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, 14, 485-500.
- Olsson, U. (1979b). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.
- Prescott, C. A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. *Behavior Genetics*, 34, 17-40.

- Prescott, C. A., & Kendler, K. S. (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *American Journal of Psychiatry*, *156*, 34-40.
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, *5*(6), 554-571.
- Rose, R. J. (1995). Genes and human behavior. *Annual Review of Psychology*, *46*, 625-654.
- Rowe, D. C., & Cleveland, H. H. (1996). Academic achievement in Blacks and Whites: Are the developmental processes similar? *Intelligence*, *23*, 205-228.
- Rowe, D. C., & Rodgers, J. L. (1997). Poverty and behavior: Are environmental measures nature and nurture? *Developmental Review*, *17*, 358-375.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling : multilevel, longitudinal, and structural equation models*. London: Chapman & Hall.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.
- Stallings, M. C., Hewitt, J. K., Cloninger, C., Heath, A. C., & Eaves, L. J. (1996). Genetic and environmental structure of the Tridimensional Personality Questionnaire: Three or four temperament dimensions? *Journal of Personality and Social Psychology*, *70*, 127-140.
- Truett, K., Eaves, L., Meyer, J., & Heath, A. C. (1992). Religion and education as mediators of attitudes: A multivariate analysis. *Behavior Genetics*, *22*, 43-62.

- Truett, K. R. (1993). Age differences in conservatism. *Personality and Individual Differences, 14*, 405-411.
- Truett, K., Eaves, L., Walters, E., Heath, A., Hewitt, J., Meyer, J., et al. (1994). A model system for analysis of family resemblance in extended kinships of twins. *Behavior Genetics, 24*, 35-49.
- Wilson, G. D., & Patterson, J. R. (1968). A new measure of conservatism. *British Journal of Social & Clinical Psychology, 7*, 264-269.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 257-293). Newbury Park, CA: Sage.

Appendix A: Mathematical Details of Identification Conditions

The minimal conditions that can identify current model are:

- A) The distributions of latent continuous variables, y^* , underlying observed variables, y 's, are set to be standard normal distributions, i.e., the mean and variances, μ^* and σ^{2*} are set to be zero and one, respectively, in one reference group.
- B) One of the thresholds of all indicator variables is constrained to be invariant across group.
- C) For three variables one more threshold other than the threshold constrained in condition B) is constrained to be invariant across groups. Condition C) implies that at least three indicators are needed.
- D) Factor means, κ 's are set to be zero in reference group.
- E) The factor loadings are constrained to be invariant across groups. Factor loadings set to be invariant across group makes the variables measures the same common factors across groups.

To show the identification conditions denote j -th observed variable each twin as $y_{j1}^{(k)}$ and $y_{j2}^{(k)}$, and their underlying continuous variables as $y_{j1}^{(k)*}$ and $y_{j2}^{(k)*}$, respectively. The variances of common factors are set to 1 in all groups and the distributions of latent continuous variables, $y_j^{(k)*}$, are set to standard normal in reference group by condition A). Therefore by setting the factor means, κ 's, as 0 by condition B) factor loadings can be identified in reference group without any restriction on the factor loading matrix as discussed in Millsap and Yun-Tein (2004) and Millsap (2001). Minimally, for non-

reference group h the means and variances of $y_j^{(h)*}$ can be identified as follow by setting two thresholds of three variables to be invariant by condition B) and C). Denote those three variables in group h as $y_1^{(h)}$, $y_2^{(h)}$, and $y_3^{(h)}$. The subscript that identify twin 1 or twin 2 in the same pair is omitted in $y_j^{(h)*}$ because the order of twin is assumed to be randomly selected and the means and variances from each twin in the same pair are assumed to be the same across twins (Neale & Cardon, 1992). For the first observed variable $y_1^{(h)}$ denote two thresholds constrained to be invariant as $\tau_{11}^{(h)}$ and $\tau_{12}^{(h)}$, and the mean and variance of $y_1^{*(h)}$, $\mu_1^{*(h)}$ and $\sigma_1^{*(h)}$, respectively, can be identified.

$$\frac{\tau_{11} - \mu_1^{*(k)}}{\sigma_1^{*(k)}} = Z_{11}^{(k)} \quad (\text{A1})$$

where $Z_{11}^{(h)}$ is $\Phi^{-1}(P(y_1^{(h)} = C_1))$, i.e. the corresponding Z score of the probability of $y_j^{(h)}$ is in the first category C_1 . For the second threshold,

$$\frac{\tau_{12} - \mu_1^{*(h)}}{\sigma_1^{*(h)}} = Z_{12}^{(h)} \quad (\text{A2})$$

where $Z_{12}^{(h)}$ is $\Phi^{-1}(P(y_1^{(h)} \leq C_2))$. The superscript on τ_{11} and τ_{12} are omitted because they are set to be invariant across groups. Because τ_{11} and τ_{12} are given from invariance condition, and $Z_{11}^{(h)}$ and $Z_{12}^{(h)}$ are given from data, the Equation A1 and A2 contain two unknowns $\mu_1^{(h)*}$ and $\sigma_1^{(h)*}$ with two equations. Therefore $\mu_1^{(h)*}$ and $\sigma_1^{(h)*}$ can be identified as follow.

$$\mu_1^{*(k)} = \frac{\tau_{11}Z_{12}^{(h)} - \tau_{12}Z_{11}^{(h)}}{Z_{12}^{(h)} - Z_{11}^{(h)}}, \text{ and} \quad (\text{A3})$$

$$\sigma_1^{*(h)} = \frac{\tau_{12} - \tau_{11}}{Z_{12}^{(h)} - Z_{11}^{(h)}}. \quad (\text{A4})$$

Means and variances of $y_2^{(h)}$ and $y_3^{(h)}$ can be identified with the same way. Equation A3 and A4 can be written in terms of scaling factor $\Delta_j^{(h)}$, which is the inverse of $\sigma_j^{*(h)}$ (Muthén & Asparouhov, 2002).

$$\Delta_j^{(h)}(\tau_{j1}^{(h)} - \mu_j^{*(h)}) = Z_{j1}^{(h)} \quad (\text{A5})$$

$$\Delta_j^{(h)}(\tau_{j2}^{(h)} - \mu_j^{*(h)}) = Z_{j2}^{(h)} \quad (\text{A6})$$

With $\mu_1^{*(k)}$, $\mu_2^{*(k)}$, and $\mu_3^{*(k)}$ identified above, factor means can be identified with the means of three variables. Denoting means of three underlying continuous variables as $\mu_1^{*(h)}$, $\mu_2^{*(h)}$, and $\mu_3^{*(h)}$ identified above, the means of three common factors can be expressed in terms of these three means.

$$\begin{aligned} \mu_1^{(h)*} &= \lambda_{A1}K_A^{(h)} + \lambda_{C1}K_C^{(h)} + \lambda_{E1}K_E^{(h)} \\ \mu_2^{(h)*} &= \lambda_{A2}K_A^{(h)} + \lambda_{C2}K_C^{(h)} + \lambda_{E2}K_E^{(h)} \\ \mu_3^{(h)*} &= \lambda_{A3}K_A^{(h)} + \lambda_{C3}K_C^{(h)} + \lambda_{E3}K_E^{(h)} \end{aligned} \quad (\text{A7})$$

There are three equations with three unknowns because factor loadings are constrained to be the same by condition E), so $\kappa^{(h)}$'s can be identified. For the variables in which only one threshold is invariant the mean of those variables can be identified with given factor means and factor loadings. For the fourth variable $y_4^{(h)}$, the mean is,

$$\mu_4^{(h)*} = \lambda_{A4}\kappa_A + \lambda_{C4}\kappa_C + \lambda_{E4}\kappa_E \quad (\text{A8})$$

and standard deviation $\sigma_4^{*(h)}$ can be identified from

$$\sigma_4^{*(h)} = \frac{\tau_{41} - \mu_4^{*(h)}}{Z_{41}^{(h)}} \quad (\text{A9})$$

τ_{41} is identified from invariant condition. Remaining parameters are the thresholds which are not constrained to be invariant across group. Denoting the third threshold of variable $y_1^{(h)}$ as $\tau_{13}^{(h)}$, it can be identified using the mean and variance identified from A8 and A9:

$$\begin{aligned} \frac{\tau_{13}^{(h)} - \mu_1^{*(h)}}{\sigma_1^{*(h)}} &= Z_{13}^{(h)} \\ \tau_{13}^{(h)} &= Z_{13}^{(h)} \sigma_1^{*(h)} + \mu_1^{*(h)} \end{aligned} \quad (\text{A10})$$

Denoting the second threshold of $y_4^{(h)}$ as $\tau_{42}^{(h)}$, it can also be identified from the mean and variance identified from A8 and A9.

$$\begin{aligned}\frac{\tau_{42}^{(h)} - \mu_4^{*(h)}}{\sigma_4^{*(h)}} &= Z_{42}^{(h)} \\ \tau_{42}^{(h)} &= Z_{42}^{(h)} \sigma_4^{*(h)} + \mu_4^{*(h)}\end{aligned}\tag{A11}$$

Thus all parameters are identified.

Appendix B: Description of Mplus Program

The excerpt of Mplus program for Racial Prejudice dimension is shown as a sample program. The details of Mplus language is described in Mplus User's Guide (Muthén & Muthén, 2006). Rather than going over general Mplus programming the application of current model is described. Mplus program starts with **Data** statement which specifies the location of the data file. Before **Data** statement users can add title of the program. Variable statement specifies the variables to be used and missing values and grouping variables. Five groups are specified in terms of the zygosity and gender: mzf for monozygotic females, monozygotic males, dizygotic females, dizygotic males, and opposite sex twins. Model statement specifies the overall model and model specific for each group can be made after defining overall model. Latent factors are defined by **By** statement followed by indicator variables. The latent variable F1 through F3 are the additive genetic factor, shared environmental factor, and unique environmental factor, respectively, for the first twin in the same family. Factors F4 through F6 are the same factors for the second twin in the same family. The numbers in parenthesis represent the constraints. The parameters with the same number in parenthesis are constrained to be the same. Because the factor loadings for the first twin and the second twin are assumed to be the same, the same numbers are used for the first twin and the second twin. Asterisks after the variable name force the Mplus to estimate the parameters. Asterisks are needed for the factor loadings because Mplus constrains the factor loading for the first indicator as one by default which is not desirable for genetic factor model. The statement of F1-F6@1 constraints the variances of factors as 1. Symbol @ fixes the parameter to the following value. **With** statement specifies the correlation between variables. The

correlation between F2 and F5 is constrained to be one because the correlation between shared environments between twins in the same family is assumed to be the same. Regression on covariate variable is specified by On statement. Each of phenotype indicator variables is regressed on age at the time of measurement (variable aget1). Regression coefficients for the same variable for the first and second twin are constrained to be the same by the number in the parenthesis. Next several lines are zero correlation constraints between factors. Because, unless being specified, Mplus tries to estimate the correlation between exogeneous variables, zero correlations should be specified explicitly. Thresholds are specified by the variable name and \$ sign in the bracket []. The numbers following the \$ sign indicate the order of threshold. There are two thresholds (\$1 and \$2) because each item has three response categories. From the numbers in parentheses, thresholds for the same variables are constrained to be the same across twins in the same family, group dzf, correlation between additive genetic factors is set to be 0.5. Because factor means are estimated for male groups factor means are freed to be estimated, but are constrained to be the same for twins from same family. From the minimal condition of identification one threshold for three variables are freely estimated for male groups. Scale factors for each variable for male groups are freely estimated and constrained to be the same across twins from same family. The constraints of factor means, thresholds, and scale factors are specified with the same manner for the rest of the groups.

Table 1. *Items Included in Each Dimension*

Racial Prejudice	Death Penalty, White Superiority, Apartheid, Women Judges, Caning, Colored immigration,
Religious Fundamentalism	Evolution, Sabbath Observance, Birth Control, Divine Law, Legalized Abortion, Church Authority, Divorce
Right-Wing Orientation	Patriotism, Licensing Laws, Royalty, Censorship, Strict Rules
Social Permissiveness	Modern Art, Working Mothers, Suicide, Fluoridation, Disarmament, White Lies

Table 2. *Estimated means and 95% confidence intervals*

	Racial Prejudice	Religious Fundamentalism	Right-wing Orientation	Social Permissiveness
upper 95%	0.8942	0.20264	-0.28524	-0.11464
Factor Mean	0.659	-0.011	-0.542	-0.391
lower 95%	0.4238	-0.22464	-0.79876	-0.66736

Table 3. *Result Summary for Racial Prejudice dimension*

** factor mean is the difference from female group in which factor means are set to be zeroes*

Item	Additive Genetic	Common Environment	Unique Environment
	Loading (SE)	Loading (SE)	Loading (SE)
Death Penalty	0.533 (0.041)	-0.343 (0.052)	0.077 (0.036)
White Superiority	0.611 (0.031)	0.207 (0.057)	0.364 (0.030)
Women Judges	-0.172 (0.037)	-0.089 (0.034)	-0.341 (0.047)
Apartheid	0.499 (0.035)	0.278 (0.048)	0.203 (0.034)
Caning	0.339 (0.043)	-0.367 (0.038)	0.056 (0.032)
Colored Immigration	-0.511 (0.022)	-0.103 (0.048)	-0.304 (0.025)
Factor Mean*	-1.800 (0.625)	-1.401 (0.673)	5.655 (1.288)
Fit Statistics			
χ^2 (df)	445.208 (210)		
CFI	0.952		
TLI	0.946		
RMSEA	0.038		

Table 4. *Result summary for Religious Fundamentalism dimension*

** factor mean is the difference from female group in which factor means are set to be zeroes*

Item	Additive Genetic	Common Environment	Unique Environment
	Loading (SE)	Loading (SE)	Loading (SE)
Evolution Theory	0.280(0.042)	0.485(0.026)	-0.033(0.026)
Sabbath Observation	-0.206(0.052)	-0.504(0.023)	0.448(0.023)
Birth Control	0.755(0.028)	0.182(0.062)	-0.097(0.045)
Divine Law	-0.165(0.053)	-0.537(0.020)	0.433(0.022)
Legalized Abortion	0.735(0.033)	0.354(0.055)	-0.168(0.033)
Church Authority	-0.256(0.053)	-0.530(0.025)	0.449(0.023)
Divorce	0.618(0.030)	0.291(0.048)	-0.151(0.031)
Factor Mean*	-0.568(0.146)	0.789(0.234)	0.136(0.260)
Fit Statistics			
χ^2 (df)	380.22(192)		
CFI	0.984		
TLI	0.984		
RMSEA	0.036		

Table 5. Result summary for Right-wing Orientation dimension

* factor mean is the difference from female group in which factor means are set to be zeroes

Item	Additive Genetic	Common Environment	Unique Environment
	Loading (SE)	Loading (SE)	Loading (SE)
Patriotism	0.305 (0.059)	0.415 (0.043)	0.162 (0.037)
Licensing Law	0.190 (0.032)	0.058 (0.035)	0.575 (0.095)
Royalty	0.511 (0.057)	0.412 (0.061)	0.060 (0.035)
Censorship	0.574 (0.024)	-0.028 (0.068)	0.334 (0.054)
Strict Rules	0.524 (0.026)	-0.103 (0.063)	0.094 (0.031)
Factor Mean*	-0.527 (0.144)	0.250 (0.185)	-0.722 (0.211)
Fit Statistics			
χ^2 (df)	245.705 (157)		
CFI	0.973		
TLI	0.967		
RMSEA	0.027		

Table 6. *Result summary for Social Permissiveness dimension*

* *Factor mean is the difference from female group in which factor means are set to be zeroes*

Item	Additive Genetic	Common Environment	Unique Environment
	Loading (SE)	Loading (SE)	Loading (SE)
Modern Art	0.346(0.035)	0.253(0.038)	0.275(0.050)
Working Mothers	0.361(0.039)	0.326(0.038)	0.085(0.035)
Suicide	0.031(0.058)	0.530(0.024)	-0.056(0.042)
Fluoridation	0.598(0.032)	-0.202(0.056)	-0.311(0.066)
Disarmament	0.106(0.030)	0.203(0.023)	0.134(0.033)
White Lies	0.117(0.047)	0.374(0.027)	-0.246(0.047)
Factor Mean*	-0.531(0.144)	0.074(0.178)	-0.442(0.209)
Fit Statistics			
χ^2 (df)	613.678(248)		
CFI	0.868		
TLI	0.840		
RMSEA	0.044		

Table 7. *Estimated means of latent continuous variables for males*

Racial		Religious		Political		Social	
Death Penalty	-0.043	Evolution Theory	-0.219	Patriotism	-0.174	Modern Art	-0.287
White Superiority	0.669	Sabbath Observation	-0.220	Licensing Law	-0.501	Working Mother	-0.205
Women Judges	1.494	Birth Control	0.298	Royalty	-0.210	Suicide	0.048
Apartheid	-0.140	Divine Law	-0.271	Censorship	-0.551	Fluoridation	-0.195
Caning	0.221	Legalized Abortion	0.161	Strict Rules	-0.370	Disarmament	-0.100
Colored Immigration	0.655	Church Attendance	0.212			White Lies	0.074
		Divorce	-0.142				
Average	0.476		-0.026		-0.361		0.111

Table B1. The excerpt of Mplus program

Title: Genetic Factor Model with the Means of Genetic and Environmental Factors
Data:

```
File is 'u:\data\ozpairw.txt';  
Variable:  
Names are xfam xid sexwlt1 zygwlt1 wlt1a1 wlt1a2...;  
Usevariables are wlt1a1 ... wlt2A48 aget1;      ! list of variables to be used;  
Categorical are wlt1a1 ... wlt2A48;           ! list of categorical variables;  
Missing are .;  
Grouping is  
zygwlt1(1=mzf, 2=mzm, 3=dzf, 4=dzm, 5=dzo); ! specifying grouping variable;
```

```
Analysis:  
Type=general meanstructure H1;                ! specifying analysis options;  
Param=delta;
```

```
output:  
standardized samp res;                        ! Requesting the outputs to display;
```

```
Model:                                         ! Model specification is started;  
F1 by                                         ! F1 is additive genetic factor;  
wlt1a1* (1)                                  ! for the first twin;  
wlt1A17*(2)  
wlt1A30*(3)  
wlt1A33*(4)  
wlt1A39*(5)  
wlt1A48*(6);  
  
F2 by                                         ! F2 is shared environmental factor;  
wlt1a1* (11)                                 ! for the first twin;  
wlt1A17*(12)  
wlt1A30*(13)  
wlt1A33*(14)  
wlt1A39*(15)  
wlt1A48*(16);  
  
F3 by                                         ! F3 is the unique environmental factor;  
wlt1a1* (21)                                 ! for the first twin;  
wlt1A17*(22)  
wlt1A30*(23)  
wlt1A33*(24)  
wlt1A39*(25)  
wlt1A48*(26);  
  
F4 by                                         ! F4 is the additive genetic factor;  
wlt2A1* (1)                                  ! for the second twin;  
wlt2A17*(2)  
wlt2A30*(3)  
wlt2A33*(4)  
wlt2A39*(5)  
wlt2A48*(6);  
  
F5 by                                         ! F5 is the shared environmental factor;  
wlt2A1* (11)                                 ! for the second twin;  
wlt2A17*(12)  
wlt2A30*(13)  
wlt2A33*(14)  
wlt2A39*(15)  
wlt2A48*(16);  
  
F6 by                                         ! F6 is the unique shared environmental;
```

```

wlt2A1* (21)          ! factor for the second twin;
wlt2A17*(22)
wlt2A30*(23)
wlt2A33*(24)
wlt2A39*(25)
wlt2A48*(26);

F1-F6@1;             ! Constraining the factor variance to be one;

F2 with F5@1;       ! Correlation between shared environment
                    ! between shared environment of the ;
                    ! first and second twin is one;
wlt1a1 wlt2a1 on aget1 (col); ! Covariating out the age variable;
wlt1A17 wlt2A17 on aget1 (col7);
.
.

F1 with F2@0 F3@0 F5@0 F6@0; ! Constraints of correlations which are;
F2 with F3@0 F4@0 F6@0;      ! supposed to be zeroes;
F3 with F4@0 F5@0 F6@0;
F4 with F5@0 F6@0;
F5 with F6@0;

[wlt1a1$1](t11);          ! Constraints of the first thresholds;
[wlt1A17$1](t171);
[wlt1A30$1](t301);
[wlt1A33$1](t331);
[wlt1A39$1](t391);
[wlt1A48$1](t481);
[wlt2a1$1](t11);
[wlt2A17$1](t171);
[wlt2A30$1](t301);
[wlt2A33$1](t331);
[wlt2A39$1](t391);
[wlt2A48$1](t481);

[wlt1a1$2](t12);          ! Constraints of the second thresholds;
[wlt1A17$2](t172);
[wlt1A30$2](t302);
[wlt1A33$2](t332);
[wlt1A39$2](t392);
[wlt1A48$2](t482);
[wlt2a1$2](t12);
[wlt2A17$2](t172);
[wlt2A30$2](t302);
[wlt2A33$2](t332);
[wlt2A39$2](t392);
[wlt2A48$2](t482);

{wlt1a1@1};             ! Constraints of scale factors;
{wlt1A17@1};
{wlt1A30@1};
{wlt1A33@1};
{wlt1A39@1};
{wlt1A48@1};
{wlt2a1@1};
{wlt2A17@1};
{wlt2A30@1};
{wlt2A33@1};
{wlt2A39@1};
{wlt2A48@1};

Model dzm;             ! Specifying the model for;

```

```

F1 with F4@.5;           ! monozygotic female twins;
[F1 F4] (fm1);
[F2 F5] (fm2);
[F3 F6] (fm3);

[wlt1A33$2*] (tm332);   ! For minimal constraints one of thresholds;
[wlt1A39$2*] (tm392);   ! for three variables are freely estimated;
[wlt1A48$2*] (tm482);
[wlt2A33$2*] (tm332);
[wlt2A39$2*] (tm392);
[wlt2A48$2*] (tm482);

{wlt1a1*} (sm1);        ! Scale factors are freed for males;
{wlt1A17*} (sm17);
{wlt1A30*} (sm30);
{wlt1A33*} (sm33);
{wlt1A39*} (sm39);
{wlt1A48*} (sm48);
{wlt2a1*} (sm1);
{wlt2A17*} (sm17);
{wlt2A30*} (sm30);
{wlt2A33*} (sm33);
{wlt2A39*} (sm39);
{wlt2A48*} (sm48);

```

Figures

Figure 1. Univariate genetic factor model

Note: Circles represent latent variables and rectangles represent observed variables.

Arrows from circles to rectangles represent factor loadings, and slings connect circles

represent the correlations. Note that correlation between additive genetic factors is 1 for monozygotic twin pairs and .5 for dizygotic twin pairs. The correlation between common

environmental factor is 1 for both monozygotic and dizygotic twin pairs

1 for MZ twin pairs/.5 for DZ twin pairs

1

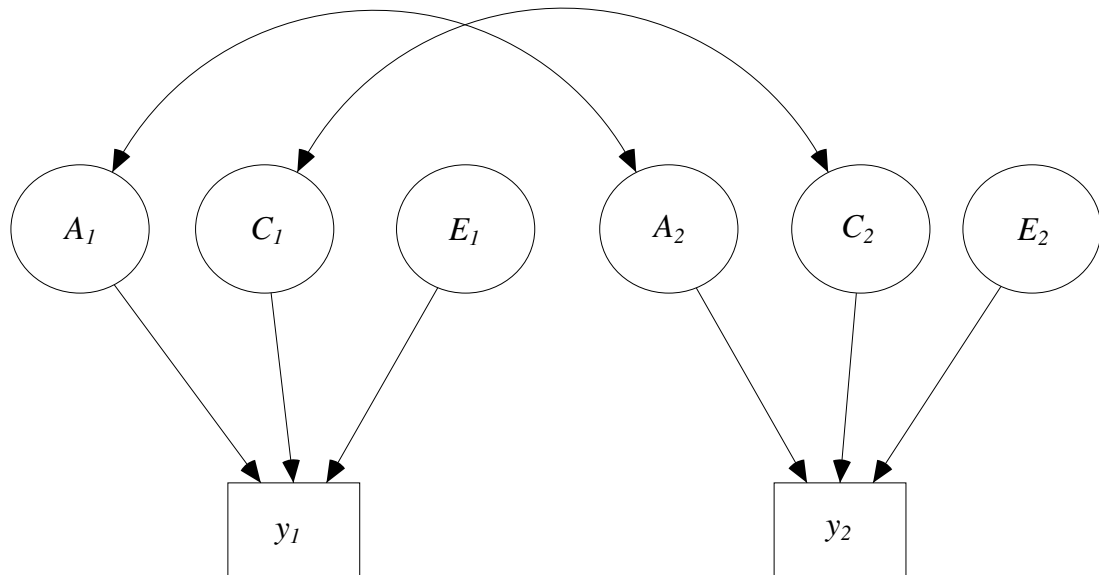


Figure 2. Multivariate genetic factor model

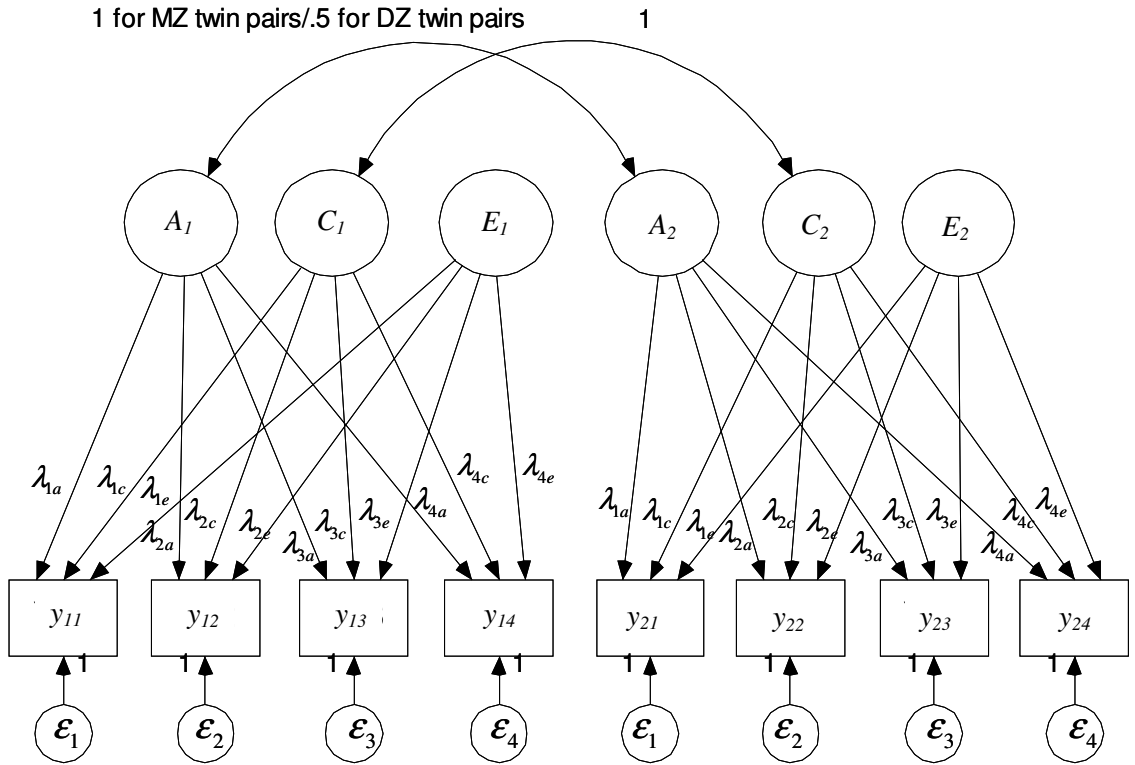


Figure 3. Relationship between y_i^* and y_i

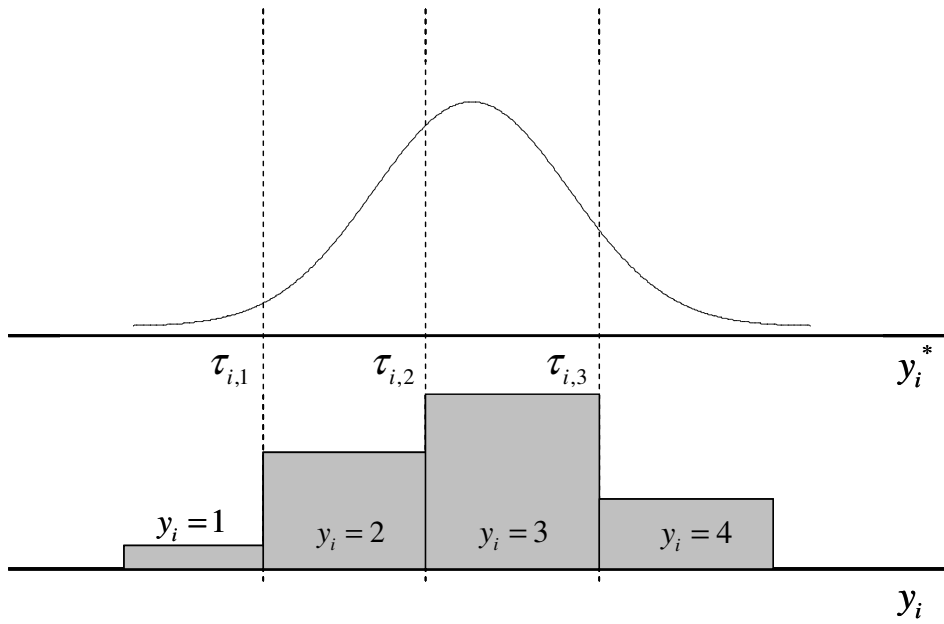


Figure 4 . Different thresholds are estimated for different response proportions based on same distribution

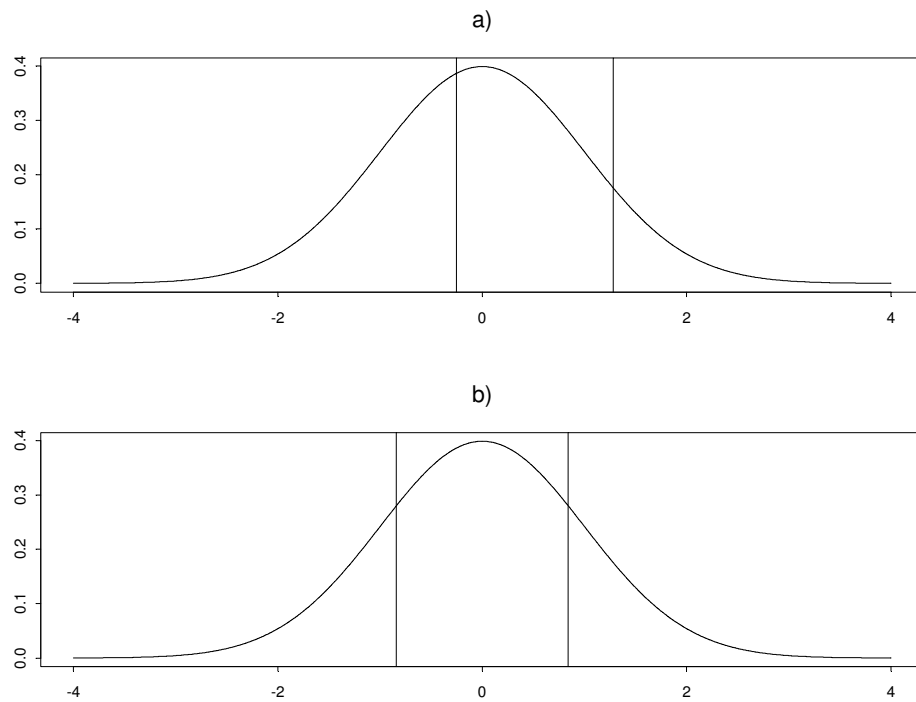


Figure 5 . Different mean and variances are estimated based on same thresholds

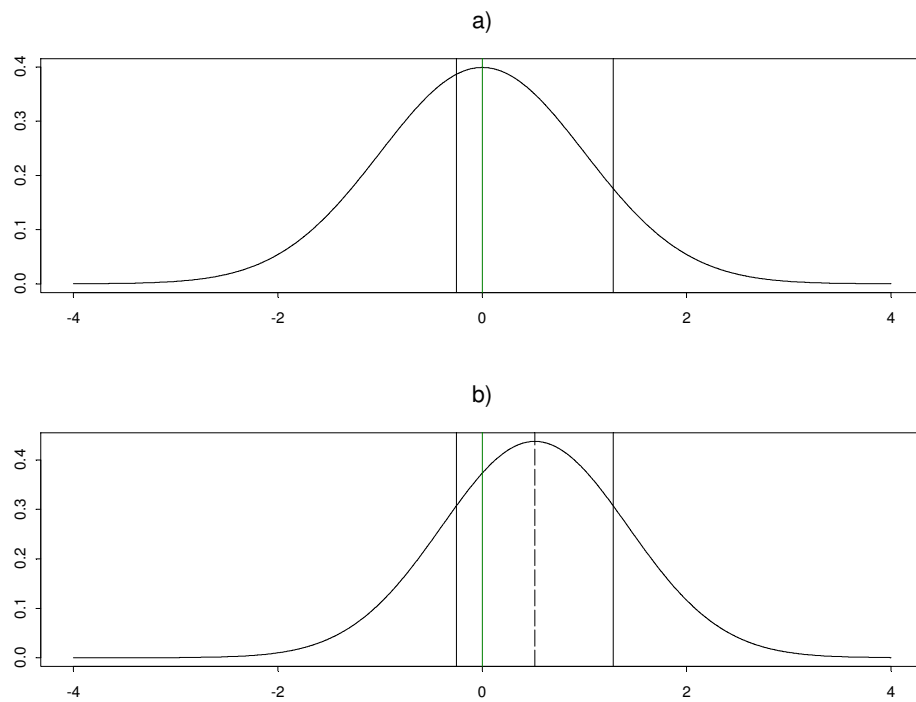


Figure 6. Plot of composite score and age

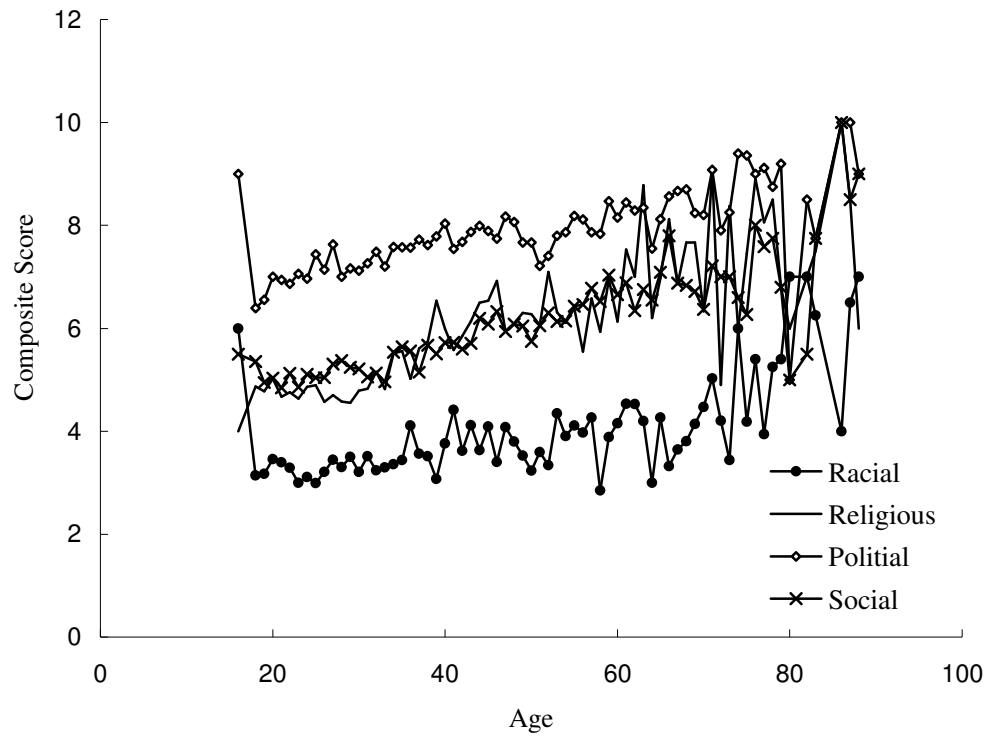


Figure 7. The path-diagram of the model used in example

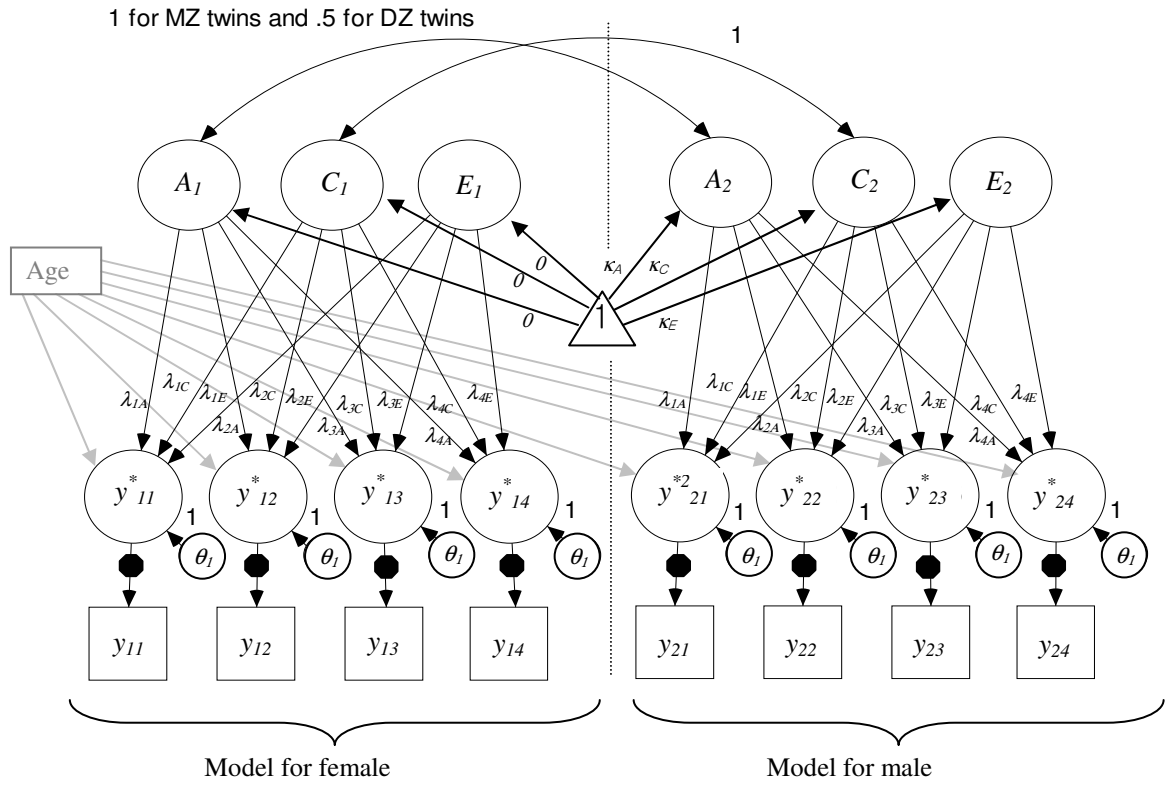


Figure 8. Plot of the mean and its 95% confidence interval of common factor for males

Note: Means for females are set to be zeroes.

