The Autoregressive Structural Model for Multivariate Longitudinal Data Analysis

A Thesis

Presented in Partial Fulfilment of the Requirements for the

Degree of Master of Science

with a

Major in Statistical Science

in the

College of Graduate Studies

University of Idaho

by

Yazhuo Deng

Major Professor: Audrey Q. Fu, Ph.D.

Committee Members: Timothy R. Johnson, Ph.D.; Stephen S. Lee, Ph.D.

Department Administrator: Christopher J. Williams, Ph.D.

December 2019

Authorization to Submit Thesis

This thesis of Yazhuo Deng, submitted for the degree of Master of Science with a major in Statistical Science and titled "The Autoregressive Structural Model for Multivariate Longitudinal Data Analysis", has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ______________________________ Date______________
Audrey Q. Fu, Ph.D.

Committee Members: ______________________________ Date______________
Timothy R. Johnson, Ph.D.

______________________________ Date______________
Stephen S. Lee, Ph.D.

Department Administrator: ______________________________ Date______________
Christopher J. Williams, Ph.D.

Abstract

Although a variety of statistical methods are available for analyzing longitudinal data, modeling the dynamics within a complex system remains a difficult methodological challenge. In this thesis, we develop an Autoregressive Structural Model (ASM) to examine these factors on depressive symptoms while accounting for temporal dependence. The ASM builds on an autoregressive model for repeated measurements and incorporates a structural equation model that delineates the mechanism among the factors and outcome. We elucidate the impact of social activity, physical activity and functional health status (factors) on depressive symptoms (outcome) in the China Health and Retirement Longitudinal Study (CHARLS), a multi-year study of aging involving 20,000 participants 45 years of age and older. The results from applying the ASM to the CHARLS data indicate that social and physical activity independently and consistently mitigated depressive symptoms over the course of five years, by mediating through functional health status.

Acknowledgements

I would like to thank Dr. Audrey Fu for her support and guidance in serving as my major professor and mentor. I would like to thank my committee members, Dr. Timothy Johnson and Dr. Stephen Lee for being wonderful mentors during my time at the University of Idaho. Thank you for all of your help, advice and feedback! I would also like to thank the Department of Movement Sciences and the Department of Statistical Sciences for supporting me to pursue two degrees at the same time.

Dedication

I dedicate this work to my parents and family for their love and support.

Table of Contents

Authorization to Submit Thesis ..... **ii**

Abstract ..... **iii**

Acknowledgements ..... **iv**

Dedication ..... **v**

Table of Contents ..... **vi**

List of Tables ..... **viii**

List of Figures ..... **ix**

1 Introduction **1**

2 Literature review **4**

2.1 Structural equation modeling 4

2.2 The simplex model 5

2.3 The Cross-lagged Panel Model 6

2.4 The Autoregressive Mediation Model 7

3 Methods **11**

3.1 The longitudinal measurement model 11

3.2 The Autoregressive Structural Model (ASM) 11

3.3 Covariate adjustment 13

3.4 Inference and model fit 13

3.5 Longitudinal measurement invariance 15

4 Application **17**

4.1 The CHARLS study 17

4.2 Results 21

5 Discussion 29

References 31

Appendix A: Mplus code for the Autoregressive Structural Model using the CHARLS data 34

List of Tables

4.1 Goodness-of-fit metrics of the ASM with different levels of invariance. The factor loadings of observed variables of functional health status and depressive symptoms were constrained to be invariant in the weak invariance model. The intercepts of observed variables of depressive symptoms were constrained to be invariant in the strong invariance model. df stands for degree of freedom, CFI comparative fit index, RMSEA root mean square error of approximation, and SRMR standardized root mean square residual. 21

4.2 Observed means ($\bar{y}$), estimated intercepts (μ) and factor loadings (λ) in the measurement model of the strong invariance ASM. 25

4.3 Coefficient estimates (βs and πs) in the structural model of our strong invariance ASM. 99% bootstrap confidence intervals are obtained for standardized estimates. SA stands for social activity, PA physical activity, FHS functional health status, and DS depressive symptoms. 26

4.4 Standardized coefficient estimates for the covariates in our strong invariance ASM (denoted by c in Equation 4.2). Italic coefficients are significant. URB stands for rural/urban, MAR marital status, and EDU educational attainment. 27

List of Figures

2.1 The simplex model ($t = 1, 2, 3 \ldots, T$). 6

2.2 The cross-lagged panel model ($t = 1, 2, 3 \ldots, T$). 8

2.3 The autoregressive mediation model ($t = 1, 2, 3 \ldots, T$). 9

4.1 The graph representation the order-2 Autoregressive Structural Model for three time points. SA = social activity; PA = physical activity; FHS = functional health status; DS = depressive symptoms. 20

4.2 Coefficient estimates in the structural model of our order-2 strong invariance ASM over three time points. Black estimates are standardized β coefficients, which are the effects among factors, and gray estimates are standardized π coefficients, which are the autoregressive effects (see Equation 4.2). Solid lines indicate statistically significant coefficient estimates, whereas dashed lines indicate insignificant ones. SA stands for social activity, PA physical activity, FHS functional health status, and DS depressive symptoms. 24

4.3 Comparison between observed means ($\bar{y}$) and expected values $\mathbb{E}(y)$ with their 99% confidence intervals of observed items of social activity, physical activity, functional health status and depressive symptoms. 28

CHAPTER 1

Introduction

The aging population in China is growing rapidly and it is estimated that the proportion of the population aged 60 or older will increase from 10% in 2000 to about 30% in 2050 (Banister, Bloom, & Rosenberg, 2012). As the burden of ischaemic heart disease, high systolic blood pressure and mental health disorders grows along with the economic development in China, the nature of health problem shifts from infectious to chronic illnesses (Zhou et al., 2019). The China Health and Retirement Longitudinal Study (CHARLS) (Zhao, Hu, Smith, Strauss, & Yang, 2012) is an ongoing longitudinal study of Chinese adults, who are 45 years of age or older and reside in administrative villages in rural areas and neighborhoods in urban areas across China, to assess their demographic characteristics, and socioeconomic and health change. Approximate 20,000 respondents from 150 counties in 20 provinces participated in the baseline and follow-up surveys. It provides a high-quality public micro-database for scientific and policy research on aging-related issues.

Late-life depression may be linked to the deterioration of many medical conditions, and impact the quality of life (Fiske, Wetherell, & Gatz, 2009). However, it is often neglected in the elderly (Xu et al., 2016). Social activity and physical activity may have a direct effect on the elderly's depressive symptoms, but the underlying mechanism by which multiple socio-behavioral factors simultaneously affect depressive symptoms over time can be complex (Fried et al., 2004). We and other groups have also demonstrated that the functional health status may act as a mediator in these pathways (Fried et al., 2004; Deng & Paul, 2018); a mediator is an intermediate factor through which another factor influences an outcome (depressive symptoms here) (MacKinnon, 2008). However, these existing studies do not investigate the dynamics of these interrelationships.

A variety of methods under the framework of the Structural Equation Modeling (SEM) have been developed to model the dynamics see (Little, 2013; McArdle & Nesselroade, 2014; Ferrer & McArdle, 2003). The standard approaches include the latent growth models, which use the latent variables of the intercept and slope to describe the growth trajectory of the outcome variable (Grimm, Ram, & Estabrook, 2016), and the autoregressive models, such as the simplex model that measures the temporal change of repeated measurements (Marsh, 1993). The autoregressive latent trajectory model developed by (Bollen & Curran, 2004) combines the desirable features of the two approaches above to describe individual growth trajectories while accounting for the temporal dependence, but it is still unable to account for factors. Similarly, although the dynamic structural equation models (Asparouhov, Hamaker, & Muthén, 2018) combines time-series models with SEM, it is designed for intensive longitudinal data with a large number of time points of a single variable, and cannot account for factors, either.

Therefore, multivariate extensions of the autoregressive models are useful for studying multidimensional structural relationships over time. A classical model is the cross-lagged panel model (Mayer, 1986), which estimates the regression coefficients between two repeatedly measured variables and their autoregressive coefficients over time. Extending this model, the Autoregressive Mediation Model (AMM) adds a third, time-lagged variable for longitudinal mediation analysis (Cole & Maxwell, 2003; MacKinnon, 2008). Using longitudinal data to test hypotheses on mediation allows us to control for the effects from earlier time points, hence minimizing bias in the estimated structural relationships (Selig & Preacher, 2009). However, the AMM can examine the dynamics or only one mediation relationship. Therefore, it is desirable to develop statistical models that simultaneously assess multiple structural relationships and mediations over time.

In this thesis, we generalize the AMM and propose the Autoregressive Structural Model (ASM) to capture the dynamics of a complex mechanism. The dynamics are described by an autoregressive model, and the mechanism is described by an SEM. The ASM can be visualized as a direct acyclic graph (DAG): the variables of interest (such as factors and outcomes) are the nodes and appear at multiple time points, and directed edges represent the mechanistic relationships within a time point as well as autoregressive dependence between time points. We assume the SEM to have an identical form over time, although the coefficient estimates may differ. We term this SEM the *template*, similar to the terminology used in the temporal graphical model developed by (Koller, Friedman, & Bach, 2009). In the following sections, the literature regarding the autoregressive model in SEM is summarized. We then provide the mathematical details of the ASM and present a real-world application of the ASM to fit the multivariate longitudinal data.

CHAPTER 2

Literature review

The wide availability of models using the SEM has provided researchers with an array of tools to analyze univariate and multivariate longitudinal data (Laursen, Little, & Card, 2011). Although carrying with the same features, these models are distinguished by explicit and subtle differences. In this chapter, I first present the general SEM approach and then discuss several longitudinal SEM models that bring alternatives together under the framework of the temporal relationships and autoregressive change.

2.1 Structural equation modeling

The SEM focuses on the modeling of variances and covariances of multivariate variables (Bollen, 1989). In the SEM framework, the relationships between latent and observed variables are expressed using a measurement equation and a structural equation. Consider p observed variables (factors and outcomes) and q latent variables. The measurement equation can be written as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y}$ is a random vector of the p observed variables, $\boldsymbol{\mu}$ an intercept vector, $\boldsymbol{\Lambda}$ a $p \times q$ factor loading matrix, $\boldsymbol{\eta}$ a vector of the q latent variables, and $\boldsymbol{\epsilon}$ a random vector of residuals.

The structural equation represents the relationships among latent variables as

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{2.2}$$

where $\boldsymbol{\Gamma}$ is a $q \times q$ coefficient matrix that indicates the structural relationships among latent variables, and $\boldsymbol{\zeta}$ also a random vector of residuals.

We can rewrite the structural equation (Equation 2.2) as

$$\boldsymbol{\eta} = (\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta},$$

and plug this expression in the measurement equation (Equation 2.1) to have

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta} + \boldsymbol{\epsilon}.$$

Then the covariance structure can be formulated as

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbb{E}\{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T\} = \mathbb{E}\{(\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta} + \boldsymbol{\epsilon})^T\} \\ &= \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\Psi}((\mathbf{I} - \boldsymbol{\Gamma})^{-1})^T\boldsymbol{\Lambda}^T + \boldsymbol{\Theta}.\end{aligned}$$

where $\boldsymbol{\Psi}$ is the covariance matrix of the residual vector $\boldsymbol{\zeta}$, $\boldsymbol{\Theta}$ is the covariance matrix of the residual vector $\boldsymbol{\epsilon}$, and $(\mathbf{I} - \boldsymbol{\Gamma})$ is a nonsingular matrix. The parameter vector $\boldsymbol{\theta}$ of the SEM is then

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Theta}\}.$$

2.2 The simplex model

The initial autoregressive model discussed here is the simplex model, which is utilized to examine the stability of individual differences in longitudinal data. The general concept of the simplex model was developed by Guttman (1954) and applied in social science. Variations of the simplex model are formulated to allow the current variable to be expressed as a function of the same variables measured at the previous time point. According to Jöreskog (1970), a quasi-simplex model fits data of one observed variable measured on multiple occasions and assumes the observations contain measurement errors. Subsequently, a stronger simplex model with multiple observed variables at each time point was developed (Marsh, 1993).

Under the SEM framework described in Section 2.1, the measurement model for the multi-item simplex model can be formulated based on Equation 2.1 and expressed as

$$y_{j,t} = \lambda_{j,t}\eta_t + \epsilon_{j,t}$$

where $y_{j,t}$ is the jth observed variable at time t, λs are the factor loadings, ηs are latent variables, and ϵs are the measurement errors for the corresponding observed variables. Subsequently, the structural model can be formulated based on Equation 2.2 and written as

$$X_t = \pi_{t-1}X_{t-1} + \zeta_t$$

where X_t is used to represent η_t as a single variable structure measured at time point t ($t = 1, 2, 3 \ldots, T$), π_{t-1} is the autoregressive coefficient that expresses the change of the variable of interest from time t to time $t-1$, ζ_t is random error at each time. The graphical representation is shown in Figure 2.1. Note that the autoregressive coefficients (e.g., πs) are named as the stability coefficients since they describe the degree to which there is a reshuffling of individuals' measures on that variable. A large score means that the change in individual differences was relatively small (Selig & Preacher, 2009).

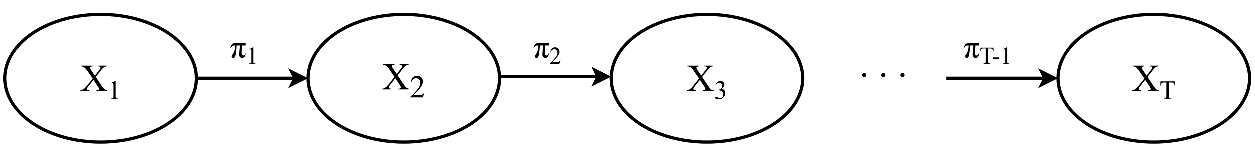

Figure 2.1: The simplex model ($t = 1, 2, 3 \ldots, T$).

2.3 The Cross-lagged Panel Model

The simplex models have the limitation of only measuring one construct across time. The cross-lagged panel model overcomes this limitation and allows us to investigate the causal relationship between two variables over time (Mayer, 1986). Similar to the simplex model, the autoregressive paths describe the stability of the latent variables from one occasion to the next. In addition, this model allows a cross-lagged influence of the other variable at a previous time. Since the measurement model of the cross-lagged panel model is same as the one for the simplex model, only the structural model was presented here, which can be expressed using two equations as

$$X_t = \pi_{1,t-1} X_{t-1} + \beta_{1,t-1} Z_{t-1} + \zeta_{X,t}$$

and

$$Z_t = \pi_{2,t-1} Z_{t-1} + \beta_{2,t-1} X_{t-1} + \zeta_{Z,t}$$

where X_t and Z_t are two different variables in the 2×1 vector η_t measured at time point t ($t = 1, 2, 3 \ldots, T$), $\pi_{1,t-1}$ and $\pi_{2,t-1}$ describe the autoregressive paths, $\beta_{1,t-1}$ and $\beta_{2,t-1}$ represent cross-lagged effects from a construct to another measured at a later occasion, $\zeta_{X,t}$ and $\zeta_{Z,t}$ are random errors that are different across time. The graph is presented in Figure 2.2.

In the cross-lagged panel model, the chain of autoregressive effects knits a stable structure for testing the cross-lagged effects. This is because the variance in Z_2 that can be explained by X_1 is residual variance conditioning on previous levels of Z_1. Thus, the inclusion of the autoregressive paths is imperative in order to minimize bias in the estimation of cross-lagged effects (Laursen et al., 2011). Covariates may also be included in cross-lagged panel models at any time point to minimize confounding bias. However, if additional covariates are included, the analyses become more complex and exploratory and less focused on a theoretically driven investigation (Newsom, 2015).

2.4 The Autoregressive Mediation Model

Mediation models are utilized to describe the mechanism by which one variable (e.g., X) has an effect on another variable (Z) through its influence on an intermediate variable (M).

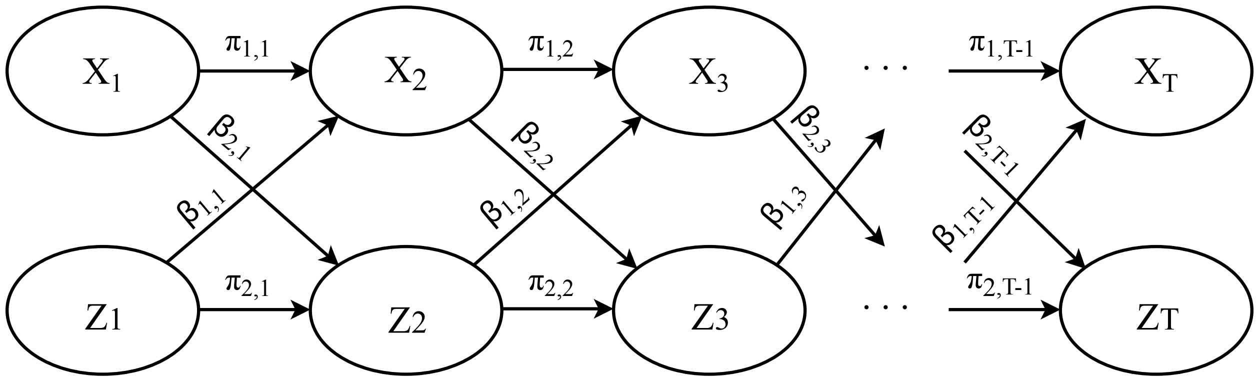

Figure 2.2: The cross-lagged panel model ($t = 1, 2, 3 \ldots, T$).

Although mediation models are often measured in a cross-sectional fashion, longitudinal data are to be preferred for the testing of mediation hypotheses. It is suggested that if we do not control for previous values of the variables, the mediation effects may be over- or underestimated relative to their true values (Selig & Preacher, 2009). In order to take the prior status of the mediation model into consideration, the Autoregressive Mediation Model was formulated as a multivariate extension of the univariate simplex model to examine longitudinal mediation (Cole & Maxwell, 2003; MacKinnon, 2008).

In the Autoregressive Mediation Model, three constructs (i.e., X M, and Z) are each measured at time point t ($t = 1, 2, 3 \ldots, T$). The structural model can be expressed by the following three equations:

$$X_t = \pi_{1,t-1} X_{t-1} + \zeta_{X,t}$$

$$M_t = \pi_{2,t-1} M_{t-1} + \beta_{1,t-1} X_{t-1} + \zeta_{M,t}$$

$$Z_t = \pi_{3,t-1} Z_{t-1} + \beta_{2,t-1} M_{t-1} + \beta_{3,t-2} X_{t-2} + \zeta_{Z,t}$$

where X_t, M_t and Z_t are three different constructs in the 3×1 vector η_t measured at time point t ($t = 1, 2, 3 \ldots, T$), $\pi_{1,t-1}$, $\pi_{2,t-1}$ and $\pi_{3,t-1}$ indicate the autoregressive paths, $\beta_{1,t-1}$, $\beta_{2,t-1}$ and $\beta_{3,t-2}$ represent cross-lagged mediational relationships, $\zeta_{X,t}$, $\zeta_{M,t}$ and $\zeta_{Z,t}$ are random errors that are different for each time. Variations of the Autoregressive Mediation Model are available. In the model described here (see Figure 2.3), the two-lag path is specified, such that Z_t is influenced by X_{t-2}. This specification fits to a particular kind of mediation hypothesis, the inclusion of alternative paths may be possible depending on the research context (Cole & Maxwell, 2003).

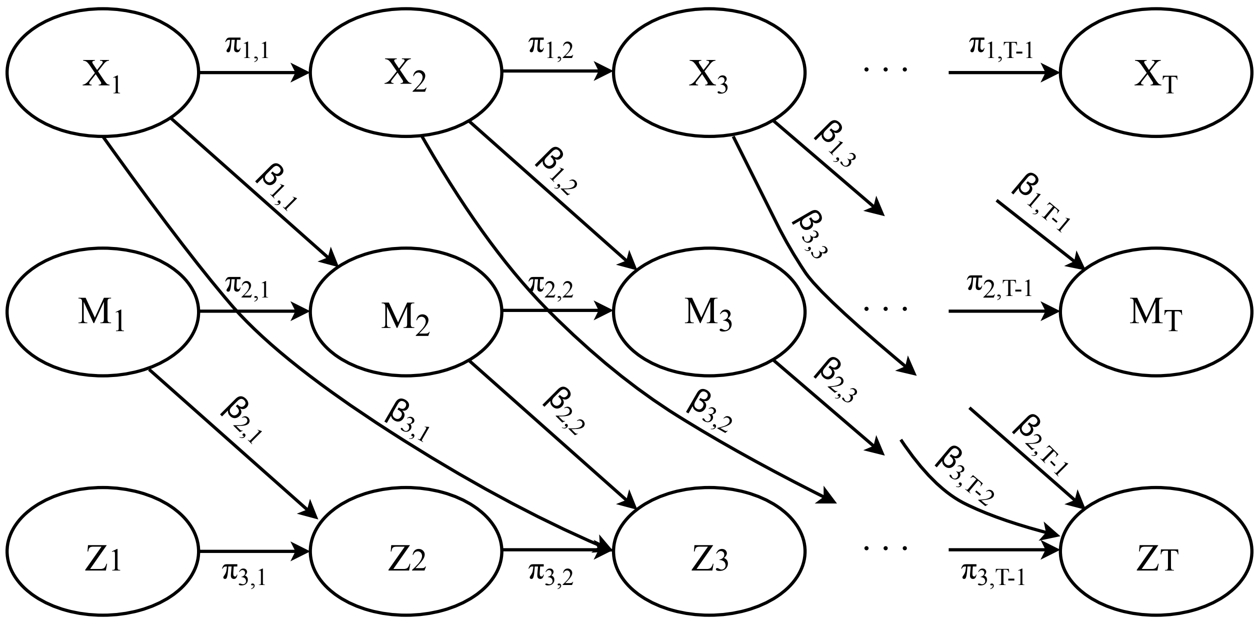

Figure 2.3: The autoregressive mediation model ($t = 1, 2, 3 \ldots, T$).

Despite the merits of the Autoregressive Mediation Model, such a model assumes that the mediation structure has only three unique constructs. It may be inappropriate to fit this model to data when we examine the temporal process of a more complicated structural mechanism delineated in certain theories. For example, the focal process of individuals is often subject to different contextual influences in human development research. These contextual variables may be located at different layers of the ecological system and influence the outcome variables directly and indirectly (Little, Bovaird, Card, et al., 2012). A structural model depicting this phenomenon may need to build a serial mediation, in which multiple mediators are included in a sequential manner. One may also construct a model using the parallel mediation, which allows one set of mediational relations conditioning on another set of mediational relations (Jones et al., 2015). Moreover, a combination of serial and parallel mediation may be appropriate in certain cases (Deng & Paul, 2018). In order to understand the temporal process of different complex mechanisms other than the three-variable mediation, a more general modeling approach that can fit a flexible structure into the process may be necessary. In the following chapter, an ASM is defined to offer an alternative tool to tackle this problem.

CHAPTER 3

Methods

In this chapter, we describe the mathematical details of the longitudinal measurement model and the ASM. Measurement invariance tests and covariates adjustment are discussed sequentially. Once the ASM is formulated using the SEM, we implement the SEM inference techniques to estimate the parameters of the model.

3.1 The longitudinal measurement model

For longitudinal data collected at T time points, the measurement equation at the tth time point is then

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Lambda}_t \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t. \tag{3.1}$$

The set of measurement equations across all time points can be stacked up to have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Lambda}_T \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix},$$

which we can concisely represent as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{3.2}$$

3.2 The Autoregressive Structural Model (ASM)

We propose the ASM that combines the two models above to capture the structural relationships behind longitudinal measurements. As the SEM, the ASM also contains a measurement equation and a structural equation. The structural equation of the ASM at the tth time can be formulated as

$$\boldsymbol{\eta}_t = \sum_{i=1}^{t-1} \mathbf{\Pi}_{i\to t}\boldsymbol{\eta}_i + \mathbf{B}_t\boldsymbol{\eta}_t + \boldsymbol{\zeta}_t, \tag{3.3}$$

where $\mathbf{\Pi}_{i\to t}$ is the $q \times q$ diagonal matrices containing the (higher-order) autoregressive coefficients from the ith time point, and $\mathbf{B}_t$ is the $q \times q$ coefficient matrix among the latent variables at the tth time. Across all time points, the autoregressive structural equations can also be stacked up and concisely written as follows:

$$\boldsymbol{\eta} \equiv \mathbf{\Gamma}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{3.4}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are both vectors of length qt, and $\mathbf{\Gamma}$ a $qt \times qt$ matrix of coefficients. We assume the same formulation of the structural equation at all the time points, although the coefficient estimates may be different at different time points. This identical structural equation is the template in our ASM.

For a first-order autoregressive structural equation, we have

$$\mathbf{\Pi}_{(t-1)\to t} = \begin{bmatrix} \pi_{1,(t-1)\to t} & 0 & 0 & 0 \\ 0 & \pi_{2,(t-1)\to t} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \pi_{q,(t-1)\to t} \end{bmatrix},$$

where $\pi_{k,(t-1)\to t}$, $k = 1, \ldots, q$, is the autoregressive coefficient of the kth latent variable in the template from the $(t-1)$th to tth time.

For an ASM of order 2, the coefficient matrix $\mathbf{\Gamma}$ is

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Pi}_{1\to 2} & \mathbf{B}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Pi}_{1\to 3} & \mathbf{\Pi}_{2\to 3} & \mathbf{B}_3 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}_{2\to 4} & \mathbf{\Pi}_{3\to 4} & \mathbf{B}_4 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Pi}_{(T-3)\to(T-1)} & \mathbf{\Pi}_{(T-2)\to(T-1)} & \mathbf{B}_{T-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Pi}_{(T-2)\to T} & \mathbf{\Pi}_{(T-1)\to T} & \mathbf{B}_T \end{bmatrix}.$$

3.3 Covariate adjustment

Demographic and socioeconomic covariates may bias parameter estimation. To control for these confounding variable, we include these covariates and rewrite Equation 3.3 as follows:

$$\boldsymbol{\eta}_t = \sum_{i=1}^{t-1} \mathbf{\Pi}_{i\to t}\boldsymbol{\eta}_i + \mathbf{B}_t\boldsymbol{\eta}_t + \mathbf{C}_t\boldsymbol{\eta}_c + \boldsymbol{\zeta}_t, \tag{3.5}$$

where the vector $\boldsymbol{\eta}_c$ contains the time-invariant covariates and the matrix of covariate coefficients $\mathbf{C}_t$ contains elements $c_{k,m,t}$, which is the coefficient of the mth covariates affecting the kth latent variable at the tth time.

3.4 Inference and model fit

Since the ASM can be formulated as an SEM (Equations 3.2 and 3.3, even accounting for covariates), we can use the inference methods developed for SEMs for parameter estimation here. Specifically, we will use the maximum likelihood method described in Chapter 4 of (Bollen, 1989). This method minimizes the differences between the sample covariance matrix (denoted as $\mathbf{S}$) and the estimated covariance matrix $\hat{\mathbf{\Sigma}}$ using a discrepancy function $F(\boldsymbol{S}, \hat{\mathbf{\Sigma}})$.

Note that the estimation of $\mathbf{\Sigma}$ involves estimation of the parameter vector $\boldsymbol{\theta}$. To emphasize such connection, we use the notation $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ in place of $\hat{\mathbf{\Sigma}}$, where $\hat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$. Under the assumption of multivariate normality of the observed variables, minimizing the discrepancy function can be obtained by maximizing the likelihood:

$$F_{\text{ML}}(\mathbf{S}, \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})) = \text{tr}(\mathbf{S}\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}) + \log|\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})| - \log|\mathbf{S}| - p, \tag{3.6}$$

where tr is trace and ML stands for maximum likelihood. The estimate $\hat{\boldsymbol{\theta}}$ is then the maximum likelihood estimate.

To assess the model fit for the ASM, we can also use the multiple metrics designed for SEMs (Hu & Bentler, 1999):

- The χ^2 statistic: This is the minimized discrepancy in Equation 3.6 and follows a χ^2 distribution with the degrees of freedom of the current model.

- The comparative fit index (CFI): It compares an invariance model to the null model which assumes zero covariances among the observed variables and is defined as

$$\text{CFI} = \frac{d_{\text{null}} - d_{\text{specified}}}{d_{\text{null}}},$$

  where $d_{\text{null}} = (\chi^2_{\text{null}} - df_{\text{null}})$ and $d_{\text{specified}} = (\chi^2_{\text{specified}} - df_{\text{specified}})$, which are the noncentrality parameters, and df indicates the degrees of freedom of the model.

- The standardized root mean square residual (SRMR):

$$\text{SRMR} = \sqrt{(\sum_{j=1}^{p}\sum_{k=1}^{p} r_{jk}^2)/e},$$

  where r_{jk} is the difference between the observed and estimated correlation between y_j and y_k, and $e = p(p+1)/2$, with p being the number of observed variables.

- The root mean square error of approximation (RMSEA) with its 90% confidence interval (CI). The RMSEA is defined as

$$\text{RMSEA} = \sqrt{\frac{(\chi^2_{\text{specified}}/df_{\text{specified}}) - 1}{n}}.$$

CFI, SRMR, and RMSEA with respective values of greater than .90, less than .08, less than .06 suggest a good model fit (Hu & Bentler, 1999). The confidence interval of RMSEA should be below 0.06.

3.5 Longitudinal measurement invariance

A key assumption of longitudinal SEM models (such as those discussed in the Introduction) is measurement invariance of the latent variables over different time points (Little, 2013); (Millsap & Cham, 2012). This assumption ensures that for any latent variable $\boldsymbol{\eta}$, the meodel measures the same effect over time. we consider three levels of invariance, namely the configural, weak and strong invariance, which give rise to three nested ASMs (Widaman, Ferrer, & Conger, 2010). The configural invariance requires the same structural re- lationships across time, and no constraint on parameters is added to the measurement equation (i.e., Equation 3.2). This is also our ASM without ad- ditional constraints, as we use the same template at all time points. The weak invariance requires equality in factor loadings over time:

$$\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \cdots = \boldsymbol{\Lambda}_T. \tag{3.7}$$

The strong invariance requires invariant loadings and invariant intercepts:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_T. \tag{3.8}$$

The strong invariance ASM is therefore nested in the weak invariance ASM, which in turn is nested in the configural invariance ASM. If the weak or strong invariance for all the elements is not supported by data, partial invariance may be imposed for a subset of factor loadings or intercepts.

Testing these different levels of invariance is effectively model selection. We use the model fit metrics described above to compare and select models. In particular, we can compute the differences in the CFI values (denoted as ΔCFI) between two models. When $\Delta\text{CFI} < .01$, the two models do not differ significantly.

CHAPTER 4

Application

4.1 The CHARLS study

The example uses the first three waves (2011, 2013 and 2015) of the CHARLS survey data to investigate the dynamic relationships between the social activity, physical activity, and functional health status, and their impact on depressive symptoms among Chinese adults of 45 years and older. In the 2011 national baseline study, a representative sample of 17,708 participants from 150 urban districts and rural counties in 28 provinces were recruited using a multistage probability sampling strategy. In 2013 and 2015 follow-up studies, 18,605 and 21,095 respondents participated respectively, including follow-up respondents and newly added ones. Only 8,959 participants responded to the physical activity survey at least once in the three waves of measurements. We focus on these participants in our analysis here. Among them, 4,739 (53%) are female, 7,837 (88%) are married and 5,759 (64%) reside in rural areas. The proportion of illiterate participants is 33% and 18% in rural and urban areas, respectively. Detailed sampling procedures and the cohort profile can be found in (Zhao et al., 2012).

Among the variables of interest, social activity measures the frequency for engaging in social activities (e.g., interacting with friends; going to community club; attending training course; caring for sick or disabled adult; taking part in the community-related organization, etc.) in the month prior to the survey. Physical activity consists of weekly durations of vigorous activity, moderate activity and walking. The functional health status utilizes the 5-item Instrumental Activities of Daily Living (IADLs) to assess the functional limitations in the engagement of essential skills for independent living. Depressive symptoms are measured using eight items of the Center for Epidemiologic Studies Depression Scale (CES-D). In total, fifteen observed variables constitute the four latent or observed variables of interest.

We consider four latent variables, one for each category, and construct an ASM of order-

2 to examine the relationships among these variables. Our ASM will also control for the time-invariant covariates, including sex, age, rural/urban residency (URB), marital status (MAR), and educational attainment (EDU).

The longitudinal measurement equation $\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Lambda}_t \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t$ at time t $(t = 1, 2, 3)$ can be expressed as follows:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ \vdots \\ y_{7,t} \\ y_{8,t} \\ \vdots \\ y_{15,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \\ \vdots \\ \mu_{7,t} \\ \mu_{8,t} \\ \vdots \\ \mu_{15,t} \end{bmatrix} + \begin{bmatrix} \lambda_{1,t} & 0 & 0 & 0 \\ 0 & \lambda_{2,t} & 0 & 0 \\ 0 & 0 & \lambda_{3,t} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \lambda_{7,t} & 0 \\ 0 & 0 & 0 & \lambda_{8,t} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_{15,t} \end{bmatrix} \begin{bmatrix} \eta_{SA,t} \\ \eta_{PA,t} \\ \eta_{FHS,t} \\ \eta_{DS,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \vdots \\ \epsilon_{7,t} \\ \epsilon_{8,t} \\ \vdots \\ \epsilon_{15,t} \end{bmatrix},$$

where $y_{1,t}$ is the observed variable for social activity, $y_{2,t}$ for physical activity, $y_{3,t}$ to $y_{7,t}$ for functional health status, and $y_{8,t}$ to $y_{15,t}$ for depressive symptoms at wave t. In addition, η_{SA}, η_{PA}, η_{FHS} and η_{DS} are the corresponding latent variables. The measurement equations for three time points are stacked to form $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda\eta} + \boldsymbol{\epsilon}$, which can be expressed as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \end{bmatrix}. \tag{4.1}$$

When formulating the structural equation, we assume that social and physical activity may influence functional health status and depressive symptoms, and that functional health status may further influence depressive symptoms. We do not allow the reverse. We further assume that the covariates may influence all four latent variables. A diagram of this structural equation model is depicted in Figure 4.1. Specifically,

$$
\begin{aligned}
\begin{bmatrix} \eta_{SA,t} \\ \eta_{PA,t} \\ \eta_{FHS,t} \\ \eta_{DS,t} \end{bmatrix} &=
\begin{bmatrix} \pi_{1,(t-2)\to t} & 0 & 0 & 0 \\ 0 & \pi_{2,(t-2)\to t} & 0 & 0 \\ 0 & 0 & \pi_{3,(t-2)\to t} & 0 \\ 0 & 0 & 0 & \pi_{4,(t-2)\to t} \end{bmatrix}
\begin{bmatrix} \eta_{SA,(t-2)} \\ \eta_{PA,(t-2)} \\ \eta_{FHS,(t-2)} \\ \eta_{DS,(t-2)} \end{bmatrix} \\
&+ \begin{bmatrix} \pi_{1,(t-1)\to t} & 0 & 0 & 0 \\ 0 & \pi_{2,(t-1)\to t} & 0 & 0 \\ 0 & 0 & \pi_{3,(t-1)\to t} & 0 \\ 0 & 0 & 0 & \pi_{4,(t-1)\to t} \end{bmatrix}
\begin{bmatrix} \eta_{SA,(t-1)} \\ \eta_{PA,(t-1)} \\ \eta_{FHS,(t-1)} \\ \eta_{DS,(t-1)} \end{bmatrix} \\
&+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \beta_{13,t} & \beta_{23,t} & 0 & 0 \\ \beta_{14,t} & \beta_{24,t} & \beta_{34,t} & 0 \end{bmatrix}
\begin{bmatrix} \eta_{SA,t} \\ \eta_{PA,t} \\ \eta_{FHS,t} \\ \eta_{DS,t} \end{bmatrix} \\
&+ \begin{bmatrix} c_{1,1,t} & c_{1,2,t} & c_{1,3,t} & c_{1,4,t} & c_{1,5,t} \\ c_{2,1,t} & c_{2,2,t} & c_{2,3,t} & c_{2,4,t} & c_{2,5,t} \\ c_{3,1,t} & c_{3,2,t} & c_{3,3,t} & c_{3,4,t} & c_{3,5,t} \\ c_{4,1,t} & c_{4,2,t} & c_{4,3,t} & c_{4,4,t} & c_{4,5,t} \end{bmatrix}
\begin{bmatrix} \eta_{SEX} \\ \eta_{AGE} \\ \eta_{URB} \\ \eta_{MAR} \\ \eta_{EDU} \end{bmatrix}
+ \begin{bmatrix} \zeta_{SA,t} \\ \zeta_{PA,t} \\ \zeta_{FHS,t} \\ \zeta_{DS,t} \end{bmatrix}
\end{aligned}
\tag{4.2}
$$

In vector notation,

$$\boldsymbol{\eta}_t = \mathbf{\Pi}_{(t-2)\to t}\boldsymbol{\eta}_{(t-2)} + \mathbf{\Pi}_{(t-1)\to t}\boldsymbol{\eta}_{(t-1)} + \mathbf{B}_t\boldsymbol{\eta}_t + \mathbf{C}_t\boldsymbol{\eta}_c + \boldsymbol{\zeta}_t. \tag{4.3}$$

This structural model implies that the information on $\eta_{SA,t}$ and $\eta_{PA,t}$ comes only from the corresponding observed variables:

$$y_{1,t} = \mu_{1,t} + \eta_{SA,t}, \text{ and } y_{2,t} = \mu_{2,t} + \eta_{PA,t}. \tag{4.4}$$

In other words,

$$\lambda_1 = \lambda_2 = 1, \quad \text{and} \quad \epsilon_{1,t} = \epsilon_{2,t} = 0. \tag{4.5}$$

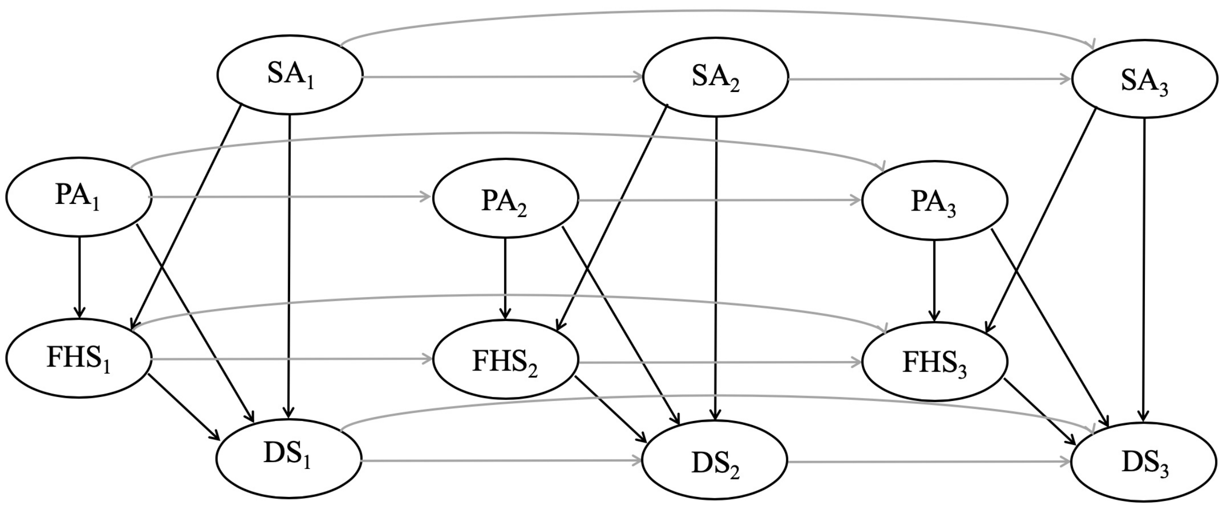

Figure 4.1: The graph representation the order-2 Autoregressive Structural Model for three time points. SA = social activity; PA = physical activity; FHS = functional health status; DS = depressive symptoms.

Similar to Section 3.5, we can further test models with different invariance constraints on the measurement equation. In the metric invariance test, we fix the factor loadings to be equal across three time points

$$\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \mathbf{\Lambda}_3. \tag{4.6}$$

In the scalar invariance test, we further constrain the intercepts

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3. \tag{4.7}$$

The analyses were conducted using Mplus 8.0 (Muthén & Muthén, 2017) and the Mplus codes are shown in Appendix A.

4.2 Results

The sample means of the observed variables from the three years are presented in Figure 4.3 and Table 4.2. Depressive symptoms showed a fluctuating pattern between 2011 and 2015, whereas average functional difficulties increased across time. Furthermore, the means of physical activity decreased from 2011 to 2015, while the means of social activity increased before the decline.

Table 4.1 reports the goodness-of-fit statistics for different invariance models described in Section 3.1. The configural invariance ASM yields a good fit. The weak invariance ASM ($\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \mathbf{\Lambda}_3$ does not different substantially from the configural model (ΔCFI = $0.943 - 0.940 = 0.003$). In the strong invariance ASM, we further constrain the intercepts of depressive symptoms to be equal across time ($\boldsymbol{\mu}_{i,1} = \boldsymbol{\mu}_{i,2} = \boldsymbol{\mu}_{i,3}$, where $i = 8, \cdots, 15$). We do not impose the invariance constraint on the intercepts of functional health status, as the sample mean scores of functional health status noticeably increased from 2011 to 2015. This (partially) strong invariance model also does not differ substantially from the configural model (ΔCFI = $0.943 - 0.938 = 0.005$). Therefore, the final model is the (partially) strong invariance model. We summarize the intercepts and factor loadings estimated in the final ASM in Table 4.2.

Table 4.1: Goodness-of-fit metrics of the ASM with different levels of invariance. The factor loadings of observed variables of functional health status and depressive symptoms were constrained to be invariant in the weak invariance model. The intercepts of observed variables of depressive symptoms were constrained to be invariant in the strong invariance model. df stands for degree of freedom, CFI comparative fit index, RMSEA root mean square error of approximation, and SRMR standardized root mean square residual.

	χ^2	*df*	CFI	RMSEA (CI)	SRMR
Configural Invariance ASM	8748.500	1044	0.943	0.029 (0.028,0.029)	0.033
Weak Invariance ASM	9135.501	1066	0.940	0.029 (0.029,0.030)	0.034
Strong Invariance ASM	9402.683	1082	0.938	0.029 (0.029,0.030)	0.035

We summarize the coefficient estimates in the structural model of our strong invariance ASM in Table 4.3 and Figure 4.2. Our ASM demonstrates a recurrent mediation relationship: social activity and physical activity independently and consistently affected depressive symptoms by mediating through functional health status (Table 4.3, Figure 4.2). The results show that participants who engaged in social activity and physical activity more frequently perceived lower levels of functional difficulties and fewer depressive symptoms. These structural relationships remained in 2015 even after conditioning on the same relationships measured in 2011 and 2013. In addition, we observe a positive relationship between physical activity and depressive symptoms consistently in three waves. The explained variances in depressive symptoms in the ASM gradually increased over time (wave1: $R^2 = 17.8\%$; wave2: $R^2 = 34.0\%$; and wave3: $R^2 = 43.2\%$).

Furthermore, most covariates have a significant influence on the variables in the ASM (see Table 4.4). Specifically, female participants engaged in a lower level of physical activity and a higher level of social activity, and reported depressive symptoms more frequently than male counterparts did. Older participants reported less physical and social activity and more functional difficulties, but perceived fewer depressive symptoms than younger ones did. Urban residents showed more participation in social activity and less in physical activity, and fewer functional difficulties and depressive symptoms than rural dwellers did. Non-married respondents were more likely to show depressive symptoms than married ones. Last but not least, more educated respondents reported less physical activity, more social activity, better functional health and fewer depressive symptoms. On the other hand, the covariates do not affect the estimates of the structural relationships.

To check whether the parameter estimates are sensible, we calculated the expected values for functional health status and depressive symptoms in 2011, 2013 and 2015, using the estimated coefficients, and compared them with the sample means (Figure 4.3). The comparison shows that the predicted means are close to the observed mean values with minor deviances.

Our findings suggest that social activity and physical activity is simultaneously and consistently associated with depressive symptoms by mediating through functional health status during the aging period. However, the concurrence of higher levels of physical activity and elevated depressive symptoms may be explained by negative impacts of domestic and occupational physical activity, because they are major sources of activity among Chinese populations, especially among those with lower socioeconomic status (Chen, Stevinson, Ku, Chang, & Chu, 2012; Deng & Paul, 2018). Our findings indicate the need for long-term monitoring of the socio-behavioral factors for depressive symptoms among Chinese elderly. An important implication from our analysis is that regular participation of both social and physical activities may be beneficial for functional and mental health in elderly. Future research may design interventions to increase the participation of social and leisure-time physical activities among Chinese older adults. Strategies may include emphasizing the role of senior social organizations, encouraging social interactions and improving infrastructure for physical exercises in rural and urban communities. Since the structural relationships we inferred here are recurrent, the interventions suggested above will be meaningful for middle-aged and older adults for the future.

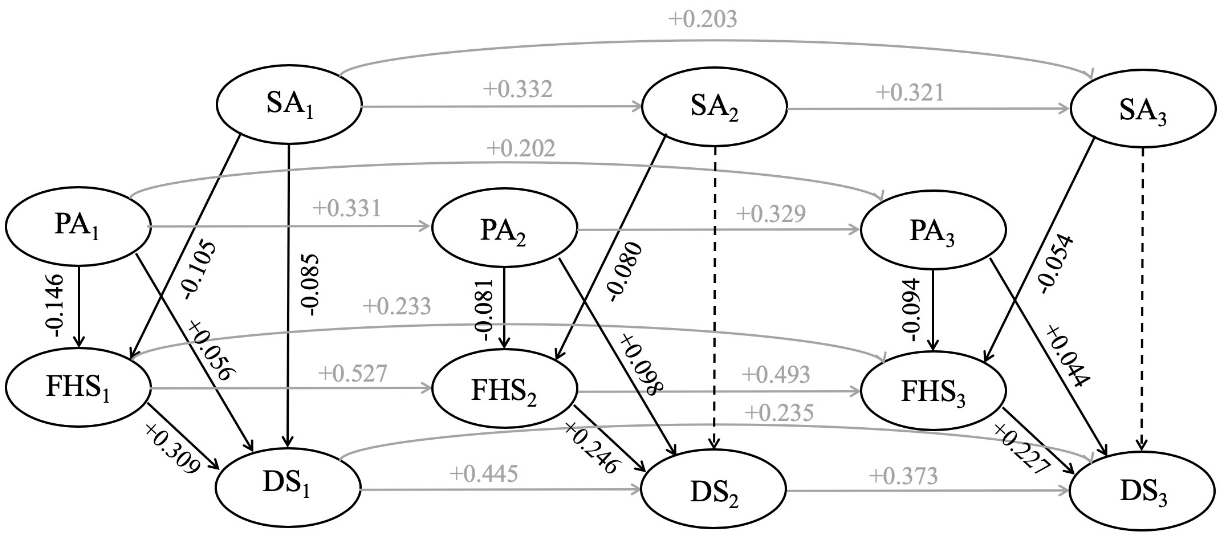

Figure 4.2: Coefficient estimates in the structural model of our order-2 strong invariance ASM over three time points. Black estimates are standardized β coefficients, which are the effects among factors, and gray estimates are standardized π coefficients, which are the autoregressive effects (see Equation 4.2). Solid lines indicate statistically significant coefficient estimates, whereas dashed lines indicate insignificant ones. SA stands for social activity, PA physical activity, FHS functional health status, and DS depressive symptoms.

Table 4.2: Observed means ($\bar{y}$), estimated intercepts (μ) and factor loadings (λ) in the measurement model of the strong invariance ASM.

Variable	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	μ_1	μ_2	μ_3	λ
Social activity							
1. Social activity	1.272	1.470	1.337	-0.287	0.051	0.388	1.000
Physical activity							
2. Physical activity	145.829	125.573	122.270	489.123	272.545	147.046	1.000
Functional health status							
3. Doing household chores	1.130	1.175	1.265	0.873	0.858	0.857	0.348
4. Preparing hot meals	1.134	1.168	1.218	0.868	0.840	0.795	0.361
5. Shopping for groceries	1.144	1.163	1.217	0.843	0.793	0.739	0.407
6. Managing money	1.243	1.238	1.267	0.968	0.899	0.830	0.372
7. Taking medications	1.091	1.081	1.097	0.975	0.938	0.912	0.158
Depressive symptoms							
8. Bothered by things that usually did not bother me	2.041	1.765	1.906	1.718	-	-	0.660
9. Trouble keeping mind on tasks	1.926	1.739	1.888	1.682	-	-	0.600
10. Felt depressed	1.987	1.759	1.896	1.673	-	-	0.737
11. Felt that everything was an effort	2.022	1.822	1.896	1.721	-	-	0.692
12. Felt fearful	1.354	1.275	1.904	1.224	-	-	0.335
13. Restless sleep	2.038	2.043	1.323	1.919	-	-	0.455
14. Felt lonely	1.523	1.438	2.069	1.375	-	-	0.447
15. Could not get going	1.369	1.297	1.370	1.236	-	-	0.389

Table 4.3: Coefficient estimates (βs and πs) in the structural model of our strong invariance ASM. 99% bootstrap confidence intervals are obtained for standardized estimates. SA stands for social activity, PA physical activity, FHS functional health status, and DS depressive symptoms.

Coefficient	Path	Estimate	Standardized Estimate	99% CI for Std Est	p-value
$\beta_{13,1}$	$SA_1 \to FHS_1$	-0.064	-0.105	(-0.128, -0.081)	< 0.001
$\beta_{23,1}$	$PA_1 \to FHS_1$	-0.001	-0.146	(-0.172, -0.118)	< 0.001
$\beta_{14,1}$	$SA_1 \to DS_1$	-0.053	-0.085	(-0.110, -0.058)	< 0.001
$\beta_{24,1}$	$PA_1 \to DS_1$	0.001	0.056	(0.034, 0.079)	< 0.001
$\beta_{34,1}$	$FHS_1 \to DS_1$	0.322	0.309	(0.269, 0.349)	< 0.001
$\beta_{13,2}$	$SA_2 \to FHS_2$	-0.050	-0.080	(-0.100, -0.059)	< 0.001
$\beta_{23,2}$	$PA_2 \to FHS_2$	-0.001	-0.081	(-0.104, -0.058)	< 0.001
$\beta_{14,2}$	$SA_2 \to DS_2$	-0.008	-0.015	(-0.039, 0.010)	0.119
$\beta_{24,2}$	$PA_2 \to DS_2$	0.001	0.098	(0.071, 0.125)	< 0.001
$\beta_{34,2}$	$FHS_2 \to DS_2$	0.213	0.246	(0.205, 0.287)	< 0.001
$\beta_{13,3}$	$SA_3 \to FHS_3$	-0.041	-0.054	(-0.072, -0.035)	< 0.001
$\beta_{23,3}$	$PA_3 \to FHS_3$	-0.001	-0.094	(-0.116, -0.071)	< 0.001
$\beta_{14,3}$	$SA_3 \to DS_3$	-0.009	-0.015	(-0.038, 0.009)	0.106
$\beta_{24,3}$	$PA_3 \to DS_3$	0.000	0.044	(0.019, 0.070)	< 0.001
$\beta_{34,3}$	$FHS_3 \to DS_3$	0.190	0.227	(0.188, 0.265)	< 0.001
$\pi_{1,1\to 2}$	$SA_1 \to SA_2$	0.362	0.332	(0.302, 0.362)	< 0.001
$\pi_{1,1\to 3}$	$SA_1 \to SA_3$	0.216	0.203	(0.173, 0.233)	< 0.001
$\pi_{1,2\to 3}$	$SA_2 \to SA_3$	0.313	0.321	(0.290, 0.352)	< 0.001
$\pi_{2,1\to 2}$	$PA_1 \to PA_2$	0.304	0.331	(0.303, 0.358)	< 0.001
$\pi_{2,1\to 3}$	$PA_1 \to PA_3$	0.180	0.202	(0.173, 0.232)	< 0.001
$\pi_{2,2\to 3}$	$PA_2 \to PA_3$	0.320	0.329	(0.298, 0.359)	< 0.001
$\pi_{3,1\to 2}$	$FHS_1 \to FHS_2$	0.596	0.527	(0.463, 0.591)	< 0.001
$\pi_{3,1\to 3}$	$FHS_1 \to FHS_3$	0.309	0.233	(0.163, 0.300)	< 0.001
$\pi_{3,2\to 3}$	$FHS_2 \to FHS_3$	0.576	0.493	(0.420, 0.563)	< 0.001
$\pi_{4,1\to 2}$	$DS_1 \to DS_2$	0.417	0.445	(0.411, 0.477)	< 0.001
$\pi_{4,1\to 3}$	$DS_1 \to DS_3$	0.248	0.235	(0.197, 0.271)	< 0.001
$\pi_{4,2\to 3}$	$DS_2 \to DS_3$	0.420	0.373	(0.335, 0.412)	< 0.001

Table 4.4: Standardized coefficient estimates for the covariates in our strong invariance ASM (denoted by c in Equation 4.2). Italic coefficients are significant. URB stands for rural/urban, MAR marital status, and EDU educational attainment.

	Covariates				
	SEX	AGE	URB	MAR	EDU
Social Activity					
$t = 1$	*0.041*	*0.046*	*0.077*	0.005	*0.133*
$t = 2$	*0.028*	-0.020	*0.078*	*0.039*	*0.111*
$t = 3$	*0.029*	*-0.043*	*0.025*	*0.024*	*0.088*
Physical Activity					
$t = 1$	*-0.174*	*-0.220*	*-0.173*	*-0.038*	*-0.121*
$t = 2$	*-0.069*	*-0.180*	*-0.065*	-0.009	*-0.059*
$t = 3$	-0.008	*-0.077*	*-0.079*	*-0.022*	-0.007
Functional Health Status					
$t = 1$	*0.033*	*0.173*	*-0.038*	0.008	*-0.108*
$t = 2$	0.003	*0.098*	0.009	-0.010	*-0.051*
$t = 3$	*0.030*	*0.066*	*-0.035*	0.010	-0.006
Depressive Symptoms					
$t = 1$	*0.141*	-0.026	*-0.076*	*0.071*	*-0.082*
$t = 2$	*0.076*	*-0.090*	-0.022	*0.040*	*-0.029*
$t = 3$	*0.064*	*-0.028*	*-0.024*	0.000	*-0.027*

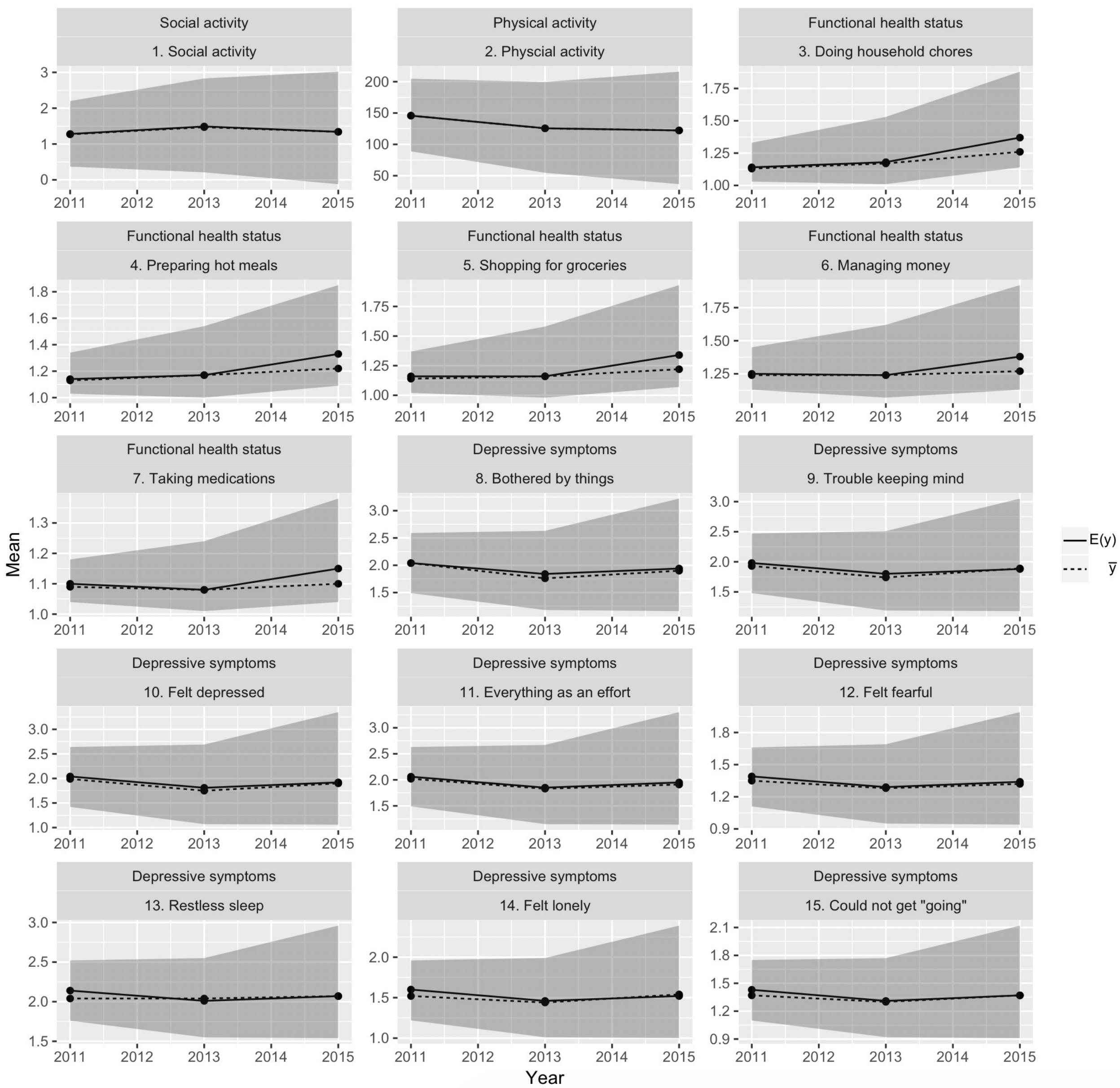

Figure 4.3: Comparison between observed means ($\bar{y}$) and expected values $\mathbb{E}(y)$ with their 99% confidence intervals of observed items of social activity, physical activity, functional health status and depressive symptoms.

CHAPTER 5

Discussion

In this thesis, we propose the ASM as a novel model for investigating the dynamics of a system using multivariate longitudinal data. The ASM extends the previous work by modeling the complex relationships within the template structure over time while accounting for autoregressive dependency. This model has several main features: (i) the structural model that captures complex relationships among multiple variables and the structure can be preserved over multiple time points; (ii) the autoregressive component that accounts for dependency over time; (iii) accounting for covariates and thus reducing bias in parameter estimates; and (iv) allowing for measurement invariance test under the complete model with the three features above. Applying our ASM to the CHARLS data, we examine how complex structural relationships evolve over time and show that social activity and physical activity are simultaneously and consistently associated with depressive symptoms by mediating through functional health status over the course of five years.

On the other hand, our ASM has several limitations. First, we assume a multivariate normal distribution for the observed variables and linear relationships. Alternatively, one can consider nonparametric structural equations (Pearl, 2009) and probabilistic graphical model (Koller et al., 2009). Second, accounting for covariates leads to many parameters to be included in the model, which may reduce the stability in estimated parameters. Additionally, we assume that the impact of covariates is independent of time. Allowing for time-dependent covariates will include even more parameters. Techniques, such as the inverse probability of treatment weights (Robins, Hernán, & Brumback, 2000), can help address the challenges of high dimensional covariates and time-dependent confounding. Last but not least, when we investigate a temporal mechanism with extensive repetitions of measurements, the inference techniques of classical SEM may not be optimal to estimate a large number of parameters. Alternative inference methods are needed to handle high-dimensional models.

In our ASM and its application, if one variable influences another, we assume that the impact can be measured at the same point without delay. However, some behavioral mechanisms may take effect after a time lag and the optimal time for measuring the effect depends on the underlying process (Selig & Preacher, 2009). For example, the treatment for a chronic medical condition may take time for the condition to improve. Hence, there is considerable interest in the timing of measurements and how the time lags impact the inference of the structural relationships (Cole & Maxwell, 2003; Dormann & Griffin, 2015). Future work will extend the ASM to account for such time lags. This extended model will provide additional flexibility.

References

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388.

Banister, J., Bloom, D. E., & Rosenberg, L. (2012). Population aging and economic growth in china. In *the chinese economy* (pp. 114–149). Springer.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York New York John Wiley & Sons.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (alt) models a synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383.

Chen, L.-J., Stevinson, C., Ku, P.-W., Chang, Y.-K., & Chu, D.-C. (2012). Relationships of leisure-time and non-leisure-time physical activity with depressive symptoms: a population-based study of taiwanese older adults. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 28.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558.

Deng, Y., & Paul, D. R. (2018). The relationships between depressive symptoms, functional health status, physical activity, and the availability of recreational facilities: a rural-urban comparison in middle-aged and older chinese adults. *International Journal of Behavioral Medicine*, *25*(3), 322–330.

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods*, *20*(4), 489.

Ferrer, E., & McArdle, J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(4), 493–524.

Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in older adults. *Annual Review of Clinical Psychology*, *5*, 363–389.

Fried, L. P., Carlson, M. C., Freedman, M., Frick, K. D., Glass, T. A., Hill, J., ... others (2004). A social model for health promotion for an aging population: Initial evidence on the experience corps model. *journal of Urban Health: Bulletin of the New York Academy of Medicine*, *81*(1), 64.

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches.* Guilford Publications.

Guttman, L. (1954). A new approach to factor analysis: The radix. In P. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, IL: free Press.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1–55.

Jones, C. L., Jensen, J. D., Scherr, C. L., Brown, N. R., Christy, K., & Weaver, J. (2015). The health belief model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. *Health Communication*, *30*(6), 566–576.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *ETS Research Bulletin Series*, *1970*(2), i–45.

Koller, D., Friedman, N., & Bach, F. (2009). *Probabilistic graphical models: principles and techniques.* MIT press.

Laursen, B. P., Little, T. D., & Card, N. A. (2011). *Handbook of developmental research methods.* Guilford Press.

Little, T. D. (2013). *Longitudinal structural equation modeling.* Guilford press.

Little, T. D., Bovaird, J. A., Card, N. A., et al. (2012). *Modeling contextual effects in longitudinal studies.* Routledge.

MacKinnon, D. (2008). *Introduction to statistical mediation analysis.* Routledge.

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, *30*(2), 157–183.

Mayer, L. S. (1986). On cross-lagged panel models with serially correlated errors. *Journal of Business & Economic Statistics*, *4*(3), 347–357.

McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models.* American Psychological Association.

Millsap, R. E., & Cham, H. (2012). *Investigating factorial invariance in longitudinal data.* Guilford Press.

Muthén, L., & Muthén, B. (2017). *Mplus User's Guide, 8th ed.* Muthén & Muthén.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction.* Routledge.

Pearl, J. (2009). *Causality.* Cambridge university press.

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 551.

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2-3), 144–164.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18.

Xu, Y., Yang, J., Gao, J., Zhou, Z., Zhang, T., Ren, J., . . . Chen, G. (2016). Decomposing socioeconomic inequalities in depressive symptoms among the elderly in china. *BMC Public Health*, *16*(1), 1214.

Zhao, Y., Hu, Y., Smith, J. P., Strauss, J., & Yang, G. (2012). Cohort profile: The china health and retirement longitudinal study (charls). *International Journal of Epidemiology*, *43*(1), 61–68.

Zhou, M., Wang, H., Zeng, X., Yin, P., Zhu, J., Chen, W., . . . others (2019). Mortality, morbidity, and risk factors in china and its provinces, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*.

Appendix A: Mplus code for the Autoregressive Structural Model using the CHARLS data

Appendix A.1. The ASM - Configural invariance

```
TITLE:
    The ASM - configural invariance


DATA:
    FILE IS mydataImpNewCov.dat;


VARIABLE:
    NAMES ARE ID hhID comID
              SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
              SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
              SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
              SEX AGE URBAN MARRIAGE EDU
              PA1 PA2 PA3;
    !Note: SAiTj  i=time j=indicator index


    USEVARIABLES ARE
              SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
              SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
              SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
              SEX AGE URBAN MARRIAGE EDU
              PA1 PA2 PA3;
```

```
ANALYSIS:

    ESTIMATOR IS ML;


MODEL:


    IADL1 BY IADL1T1*

             IADL1T2

             IADL1T3

             IADL1T4

             IADL1T5;


    IADL2 BY IADL2T1*

             IADL2T2

             IADL2T3

             IADL2T4

             IADL2T5;


    IADL3 BY IADL3T1*

             IADL3T2

             IADL3T3

             IADL3T4

             IADL3T5;


    DEP1 BY DEP1T1*

            DEP1T2

            DEP1T3

            DEP1T4
```

```
        DEP1T5
        DEP1T6
        DEP1T7
        DEP1T8;

DEP2 BY DEP2T1*
        DEP2T2
        DEP2T3
        DEP2T4
        DEP2T5
        DEP2T6
        DEP2T7
        DEP2T8;

DEP3 BY DEP3T1*
        DEP3T2
        DEP3T3
        DEP3T4
        DEP3T5
        DEP3T6
        DEP3T7
        DEP3T8;

!allow correlated residuals across time for IADL
IADL1T1 with IADL2T1 IADL3T1;
IADL2T1 with IADL3T1;
IADL1T2 with IADL2T2 IADL3T2;
```

```
    IADL2T2 with IADL3T2;
    IADL1T3 with IADL2T3 IADL3T3;
    IADL2T3 with IADL3T3;
    IADL1T4 with IADL2T4 IADL3T4;
    IADL2T4 with IADL3T4;
    IADL1T5 with IADL2T5 IADL3T5;
    IADL2T5 with IADL3T5;

    !allow correlated residuals across time for DEP
    DEP1T1 with DEP2T1 DEP3T1;
    DEP2T1 with DEP3T1;
    DEP1T2 with DEP2T2 DEP3T2;
    DEP2T2 with DEP3T2;
    DEP1T3 with DEP2T3 DEP3T3;
    DEP2T3 with DEP3T3;
    DEP1T4 with DEP2T4 DEP3T4;
    DEP2T4 with DEP3T4;
    DEP1T5 with DEP2T5 DEP3T5;
    DEP2T5 with DEP3T5;
    DEP1T6 with DEP2T6 DEP3T6;
    DEP2T6 with DEP3T6;
    DEP1T7 with DEP2T7 DEP3T7;
    DEP2T7 with DEP3T7;
    DEP1T8 with DEP2T8 DEP3T8;
    DEP2T8 with DEP3T8;

!modif indices suggest correlated residuals
```

```
    IADL1T1 with IADL1T2;
    IADL2T1 with IADL2T2;
    IADL3T1 with IADL3T2;
    DEP1T7 with DEP1T8;
    DEP2T7 with DEP2T8;
    DEP3T7 with DEP3T8;

!factor variance fixed to 1 for identification

    IADL1@1 IADL2@1 IADL3@1;
    DEP1@1 DEP2@1 DEP3@1;

!latent factor means fixed to 0 for identification
    [IADL1@0 IADL2@0 IADL3@0];
    [DEP1@0 DEP2@0 DEP3@0];

!structural paths
     !cross sectional structural paths
     !time 1
     IADL1 ON PA1;
     IADL1 ON SA1;
     DEP1 ON PA1;
     DEP1 ON SA1;
     DEP1 ON IADL1;

     !time 2
     IADL2 ON PA2;
```

```
IADL2 ON SA2;
DEP2 ON PA2;
DEP2 ON SA2;
DEP2 ON IADL2;

!time 3
IADL3 ON PA3;
IADL3 ON SA3;
DEP3 ON PA3;
DEP3 ON SA3;
DEP3 ON IADL3;

!autoregression paths across time AR(2)
PA2 ON PA1;
PA3 ON PA2;
PA3 ON PA1;
SA2 ON SA1;
SA3 ON SA2;
SA3 ON SA1;
IADL2 ON IADL1;
IADL3 ON IADL2;
IADL3 ON IADL1;
DEP2 ON DEP1;
DEP3 ON DEP2;
DEP3 ON DEP1;

!covariates paths
```

```
 PA1 ON SEX AGE URBAN MARRIAGE EDU;
 PA2 ON SEX AGE URBAN MARRIAGE EDU;
 PA3 ON SEX AGE URBAN MARRIAGE EDU;
 SA1 ON SEX AGE URBAN MARRIAGE EDU;
 SA2 ON SEX AGE URBAN MARRIAGE EDU;
 SA3 ON SEX AGE URBAN MARRIAGE EDU;
 IADL1 ON SEX AGE URBAN MARRIAGE EDU;
 IADL2 ON SEX AGE URBAN MARRIAGE EDU;
 IADL3 ON SEX AGE URBAN MARRIAGE EDU;
 DEP1 ON SEX AGE URBAN MARRIAGE EDU;
 DEP2 ON SEX AGE URBAN MARRIAGE EDU;
 DEP3 ON SEX AGE URBAN MARRIAGE EDU;

OUTPUT:
    TECH1 TECH4
    STANDARDIZED
    MODINDICES;
```

Appendix A.2. The ASM - weak invariance

```
TITLE:
    The ASM - weak (factor loading) invariance

DATA:
    FILE IS mydataImpNewCov.dat;

VARIABLE:
    NAMES ARE ID hhID comID
```

```
            SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
            SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
            SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
            SEX AGE URBAN MARRIAGE EDU
            PA1 PA2 PA3;
    !Note: SAiTj  i=time j=indicator index

    USEVARIABLES ARE
            SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
            SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
            SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
            SEX AGE URBAN MARRIAGE EDU
            PA1 PA2 PA3;

ANALYSIS:
    ESTIMATOR IS ML;

MODEL:

! Label for constraints
! The loadings of indicators are same across time

    IADL1 BY IADL1T1* (L7)
             IADL1T2  (L8)
             IADL1T3  (L9)
             IADL1T4  (L10)
             IADL1T5  (L11);
```

```
IADL2 BY IADL2T1* (L7)
         IADL2T2  (L8)
         IADL2T3  (L9)
         IADL2T4  (L10)
         IADL2T5  (L11);

IADL3 BY IADL3T1* (L7)
         IADL3T2  (L8)
         IADL3T3  (L9)
         IADL3T4  (L10)
         IADL3T5  (L11);

DEP1 BY DEP1T1* (L12)
        DEP1T2  (L13)
        DEP1T3  (L14)
        DEP1T4  (L15)
        DEP1T5  (L16)
        DEP1T6  (L17)
        DEP1T7  (L18)
        DEP1T8  (L19);

DEP2 BY DEP2T1* (L12)
        DEP2T2  (L13)
        DEP2T3  (L14)
        DEP2T4  (L15)
        DEP2T5  (L16)
```

```
        DEP2T6  (L17)
        DEP2T7  (L18)
        DEP2T8  (L19);

DEP3 BY DEP3T1* (L12)
        DEP3T2  (L13)
        DEP3T3  (L14)
        DEP3T4  (L15)
        DEP3T5  (L16)
        DEP3T6  (L17)
        DEP3T7  (L18)
        DEP3T8  (L19);

!allow correlated residuals across time for IADL
IADL1T1 with IADL2T1 IADL3T1;
IADL2T1 with IADL3T1;
IADL1T2 with IADL2T2 IADL3T2;
IADL2T2 with IADL3T2;
IADL1T3 with IADL2T3 IADL3T3;
IADL2T3 with IADL3T3;
IADL1T4 with IADL2T4 IADL3T4;
IADL2T4 with IADL3T4;
IADL1T5 with IADL2T5 IADL3T5;
IADL2T5 with IADL3T5;

!allow correlated residuals across time for DEP
DEP1T1 with DEP2T1 DEP3T1;
```

```
    DEP2T1 with DEP3T1;
    DEP1T2 with DEP2T2 DEP3T2;
    DEP2T2 with DEP3T2;
    DEP1T3 with DEP2T3 DEP3T3;
    DEP2T3 with DEP3T3;
    DEP1T4 with DEP2T4 DEP3T4;
    DEP2T4 with DEP3T4;
    DEP1T5 with DEP2T5 DEP3T5;
    DEP2T5 with DEP3T5;
    DEP1T6 with DEP2T6 DEP3T6;
    DEP2T6 with DEP3T6;
    DEP1T7 with DEP2T7 DEP3T7;
    DEP2T7 with DEP3T7;
    DEP1T8 with DEP2T8 DEP3T8;
    DEP2T8 with DEP3T8;

!modification indices suggest correlated residuals
    IADL1T1 with IADL1T2;
    IADL2T1 with IADL2T2;
    IADL3T1 with IADL3T2;
    DEP1T7 with DEP1T8;
    DEP2T7 with DEP2T8;
    DEP3T7 with DEP3T8;

!factor variance fixed to 1 for identification
    IADL1@1 ;
    DEP1@1 ;
```

```
!latent factor means fixed to 0 for identification
    [IADL1@0 IADL2@0 IADL3@0];
    [DEP1@0 DEP2@0 DEP3@0];

!structural paths
     !cross sectional structural paths
     !time 1
     IADL1 ON PA1;
     IADL1 ON SA1;
     DEP1 ON PA1;
     DEP1 ON SA1;
     DEP1 ON IADL1;

     !time 2
     IADL2 ON PA2;
     IADL2 ON SA2;
     DEP2 ON PA2;
     DEP2 ON SA2;
     DEP2 ON IADL2;

     !time 3
     IADL3 ON PA3;
     IADL3 ON SA3;
     DEP3 ON PA3;
     DEP3 ON SA3;
     DEP3 ON IADL3;
```

```
!autoregression paths across time AR(2)
PA2 ON PA1;
PA3 ON PA2;
PA3 ON PA1;
SA2 ON SA1;
SA3 ON SA2;
SA3 ON SA1;
IADL2 ON IADL1;
IADL3 ON IADL2;
IADL3 ON IADL1;
DEP2 ON DEP1;
DEP3 ON DEP2;
DEP3 ON DEP1;

!covariates paths
PA1 ON SEX AGE URBAN MARRIAGE EDU;
PA2 ON SEX AGE URBAN MARRIAGE EDU;
PA3 ON SEX AGE URBAN MARRIAGE EDU;
SA1 ON SEX AGE URBAN MARRIAGE EDU;
SA2 ON SEX AGE URBAN MARRIAGE EDU;
SA3 ON SEX AGE URBAN MARRIAGE EDU;
IADL1 ON SEX AGE URBAN MARRIAGE EDU;
IADL2 ON SEX AGE URBAN MARRIAGE EDU;
IADL3 ON SEX AGE URBAN MARRIAGE EDU;
DEP1 ON SEX AGE URBAN MARRIAGE EDU;
DEP2 ON SEX AGE URBAN MARRIAGE EDU;
```

```
 DEP3 ON SEX AGE URBAN MARRIAGE EDU;


OUTPUT:
    TECH1 TECH4
    STANDARDIZED
    MODINDICES;
```

Appendix A.3. The ASM - strong invariance

```
TITLE:
    The Autoregressive Structural Model - strong (intercept) invariance


DATA:
    FILE IS mydataImpNewCov.dat;


VARIABLE:
    NAMES ARE ID hhID comID
              SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
              SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
              SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
              SEX AGE URBAN MARRIAGE EDU
              PA1 PA2 PA3;
    !Note: SAiTj  i=time j=indicator index


    USEVARIABLES ARE
              SA1 IADL1T1-IADL1T5 DEP1T1-DEP1T8
              SA2 IADL2T1-IADL2T5 DEP2T1-DEP2T8
              SA3 IADL3T1-IADL3T5 DEP3T1-DEP3T8
```

```
              SEX AGE URBAN MARRIAGE EDU
              PA1 PA2 PA3;

ANALYSIS:
    ESTIMATOR IS ML;
    BOOTSTRAP IS 20000;
    PROCESSORS = 4;

MODEL:

! Label for constraints
! The loadings of indicators are same across time

    IADL1 BY IADL1T1* (L7)
             IADL1T2  (L8)
             IADL1T3  (L9)
             IADL1T4  (L10)
             IADL1T5  (L11);

    IADL2 BY IADL2T1* (L7)
             IADL2T2  (L8)
             IADL2T3  (L9)
             IADL2T4  (L10)
             IADL2T5  (L11);

    IADL3 BY IADL3T1* (L7)
             IADL3T2  (L8)
```

```
          IADL3T3  (L9)
          IADL3T4  (L10)
          IADL3T5  (L11);

DEP1 BY DEP1T1* (L12)
        DEP1T2  (L13)
        DEP1T3  (L14)
        DEP1T4  (L15)
        DEP1T5  (L16)
        DEP1T6  (L17)
        DEP1T7  (L18)
        DEP1T8  (L19);

DEP2 BY DEP2T1* (L12)
        DEP2T2  (L13)
        DEP2T3  (L14)
        DEP2T4  (L15)
        DEP2T5  (L16)
        DEP2T6  (L17)
        DEP2T7  (L18)
        DEP2T8  (L19);

DEP3 BY DEP3T1* (L12)
        DEP3T2  (L13)
        DEP3T3  (L14)
        DEP3T4  (L15)
        DEP3T5  (L16)
```

```
        DEP3T6  (L17)
        DEP3T7  (L18)
        DEP3T8  (L19);

!allow correlated residuals across time for IADL
IADL1T1 with IADL2T1 IADL3T1;
IADL2T1 with IADL3T1;
IADL1T2 with IADL2T2 IADL3T2;
IADL2T2 with IADL3T2;
IADL1T3 with IADL2T3 IADL3T3;
IADL2T3 with IADL3T3;
IADL1T4 with IADL2T4 IADL3T4;
IADL2T4 with IADL3T4;
IADL1T5 with IADL2T5 IADL3T5;
IADL2T5 with IADL3T5;

!allow correlated residuals across time for DEP
DEP1T1 with DEP2T1 DEP3T1;
DEP2T1 with DEP3T1;
DEP1T2 with DEP2T2 DEP3T2;
DEP2T2 with DEP3T2;
DEP1T3 with DEP2T3 DEP3T3;
DEP2T3 with DEP3T3;
DEP1T4 with DEP2T4 DEP3T4;
DEP2T4 with DEP3T4;
DEP1T5 with DEP2T5 DEP3T5;
DEP2T5 with DEP3T5;
```

```
    DEP1T6 with DEP2T6 DEP3T6;
    DEP2T6 with DEP3T6;
    DEP1T7 with DEP2T7 DEP3T7;
    DEP2T7 with DEP3T7;
    DEP1T8 with DEP2T8 DEP3T8;
    DEP2T8 with DEP3T8;

!modif indices suggest correlated residuals
    IADL1T1 with IADL1T2;
    IADL2T1 with IADL2T2;
    IADL3T1 with IADL3T2;
    DEP1T7 with DEP1T8;
    DEP2T7 with DEP2T8;
    DEP3T7 with DEP3T8;

!factor variance fixed to 1 for identification
    IADL1@1 ;
    DEP1@1 ;

!latent factor means fixed to 0 for identification
    [IADL1@0 IADL2@0 IADL3@0];
    [DEP1@0 DEP2@0 DEP3@0];

!intercept constraints across time
     [IADL1T1-IADL1T5] (I7-I11);
     [IADL2T1-IADL2T5] (K7-K11);
     [IADL3T1-IADL3T5] (P7-P11);
```

```
    [DEP1T1-DEP1T8] (I12-I19);
    [DEP2T1-DEP2T8] (I12-I19);
    [DEP3T1-DEP3T8] (I12-I19);

!structural paths
    !cross sectional structural paths
    !time 1
    IADL1 ON PA1;
    IADL1 ON SA1;
    DEP1 ON PA1;
    DEP1 ON SA1;
    DEP1 ON IADL1;

    !time 2
    IADL2 ON PA2;
    IADL2 ON SA2;
    DEP2 ON PA2;
    DEP2 ON SA2;
    DEP2 ON IADL2;

    !time 3
    IADL3 ON PA3;
    IADL3 ON SA3;
    DEP3 ON PA3;
    DEP3 ON SA3;
    DEP3 ON IADL3;
```

```
!autoregression paths across time AR(2)
PA2 ON PA1;
PA3 ON PA2;
PA3 ON PA1;
SA2 ON SA1;
SA3 ON SA2;
SA3 ON SA1;
IADL2 ON IADL1;
IADL3 ON IADL2;
IADL3 ON IADL1;
DEP2 ON DEP1;
DEP3 ON DEP2;
DEP3 ON DEP1;

!covariates paths
PA1 ON SEX AGE URBAN MARRIAGE EDU;
PA2 ON SEX AGE URBAN MARRIAGE EDU;
PA3 ON SEX AGE URBAN MARRIAGE EDU;
SA1 ON SEX AGE URBAN MARRIAGE EDU;
SA2 ON SEX AGE URBAN MARRIAGE EDU;
SA3 ON SEX AGE URBAN MARRIAGE EDU;
IADL1 ON SEX AGE URBAN MARRIAGE EDU;
IADL2 ON SEX AGE URBAN MARRIAGE EDU;
IADL3 ON SEX AGE URBAN MARRIAGE EDU;
DEP1 ON SEX AGE URBAN MARRIAGE EDU;
DEP2 ON SEX AGE URBAN MARRIAGE EDU;
DEP3 ON SEX AGE URBAN MARRIAGE EDU;
```

```
Model indirect:
    DEP1 IND PA1;
    DEP1 IND SA1;
    DEP2 IND PA2;
    DEP2 IND SA2;
    DEP3 IND PA3;
    DEP3 IND SA3;
    DEP2 IND PA1;
    DEP2 IND SA1;
    DEP3 IND PA2;
    DEP3 IND SA2;
    DEP3 IND PA1;
    DEP3 IND SA1;

OUTPUT:
    TECH1 TECH4
    STANDARDIZED
    MODINDICES
    CINTERVAL (BCBOOTSTRAP);
```