The Pennsylvania State University

The Graduate School

College of Health and Human Development

TIME SERIES ANALYSIS AND PERSON-SPECIFIC PSYCHOLOGICAL DEVELOPMENT: STATE SPACE MODELING APPLICATIONS IN BEHAVIOR GENETIC AND NEUROCOGNTIVE DESIGNS

A Dissertation in

Human Development and Family Studies

by

Lawrence L. Lo

© 2016 Lawrence L. Lo

Submitted in Partial Fulfillment
of the Requirements
for the degree of

Doctor of Philosophy

August 2016

The dissertation of Lawrence L. Lo was reviewed and approved* by the following:

Michael Rovine
Professor of Human Development and Family Studies
Chair of committee

Peter Molenaar
Dissertation Adviser
Professor of Human Development and Family Studies

Sherry Corneal
Professor of Human Development and Family Studies

Jenae Neiderhiser
Professor of Psychology

Eva Lefkowitz
Professor of Human Development and Family Studies
Head of Human Development and Family Studies

*Signatures are on file in the Graduate School.

Abstract

Person-specific methodology puts a rigorous statistical focus on the individual. This dissertation gives an overview of this methodological perspective and contrasts various intraindividual and interindividual analytic techniques. The importance of time series statistical methods, with a special emphasis on state space models, is discussed within the context of person-specific methods. These methods are applied in two papers within the fields of behavior genetics and cognitive decision making. The first paper outlines recent criticisms of twin research within behavior genetics regarding epigenetics and somatic mosaicism. A novel model is proposed and tested within several simulation studies in order to examine estimation accuracy. The second paper involves an application of recently developed exploratory connectivity analysis to fMRI data from an incentivized decision making task. Both of these applications exemplify the capabilities of state space modeling within a person-specific perspective.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I owe my gratitude to the many people who have made this dissertation possible. I would first like to thank my co-advisors, Peter Molenaar and Michael Rovine. Peter has been my longtime intellectual hero and has taught me more in my academic life than any other person. Michael has been like a father to me during graduate school and has also passed on an enormous amount of knowledge to me. Both of these men have shaped and molded my mind the most during my graduate studies.

There are several other faculty that have made great contributions to this dissertation and my academic life. Jenae Neiderhiser has helped me develop my interests in behavior genetics as both a teacher and committee member. Charles Geier has similarly guided me along my pursuits in cognitive neuroscience and has been of invaluable help in obtaining the data used in this dissertation. I am also grateful to Sherry Corneal for her support and development of my teaching skills. The HDFS staff has also helped me countlessly throughout these years, and I would like to especially thank Christa Krepps and Mary-Jo Spicer.

Although many friends have supported me through these years, both Josue Preciado and Matt Syx have been like brothers to me. Their friendship has been invaluable to me.

None of this would have been possible without the love and patience of my family. My immediate family has been a constant source of support through these years. My father and mother, Alex and Bernadette, have been my foundation. I also couldn't ask for more supportive sisters than Sophie, Jenny, and Raphael.

Dissertation introduction: Person-specific methodology and time series analysis

Psychology has been traditionally dominated by an interindividual (between-persons) methodological perspective. This involves a focus on variation between people and statistics that produce features of populations. The inferences from these interindividual statistical methods are assumed to generalize to all the individuals that compose these populations. This introduction will address some of the problems of this perspective and summarize an alternative methodological perspective. This alternate perspective involves a focus on the individual and has been commonly referred to as person-specific analysis (Molenaar, 2004).

Conventional interindividual techniques will often make inferences about intraindividual (within-person) processes. Perhaps the most popular interindividual statistical method is mixed-effects modeling (Laird & Ware, 1982), also known as hierarchical linear modeling (Bryk & Raudenbush, 1992) and multi-level modeling (Singer & Willet, 2003). In this method, analysis from variation between people is used to make inferences on variation within a person. For example, a growth curve model can make several inferences on how an individual changes across time. There are several problems of inferring intraindividual processes from interindividual analyisis.

Molenaar (2004) addresses these problems within concepts based on ergodic theory, a branch of mathematics that provides conditions that are necessary for generalizing inferences from pooled aggregates to a single individual within this pooling procedure. These conditions (under Gaussian assumptions) are stationarity and homogeneity. Stationarity refers to the equivalence across time of statistical characteristics such as means, covariances, and

autocovariances. Homogeneity refers to the equivalence of statistical characteristics across individuals within the sample to be pooled. This latter condition of homogeneity is of primary concern when considering the mixed effects model. The insensitivity of these models to heterogeneity in individual processes has been demonstrated in both simulation studies (Molenaar, 1997; Molenaar, Huizenga, & Nesselroade, 2002) and formal mathematical proofs (Kelderman & Molenaar, 2007). Empirical demonstrations of heterogenous processes and the insensitivity of conventional interindividual models to detect these forms of heterogeneity have occurred in domains such as personality (Borkenau & Ostendorf, 1998) and affective processes (Rovine, Molenaar, & Corneal, 1999; Molenaar et al., 2009)

Person-specific analysis via individual time-series modeling avoids these problems of interindividual analysis and has been argued to be necessary for investigating any dynamic or developmental process (Molenaar, 2007; Molenaar and Campbell, 2009). The primary motivation for person-specific analysis is to avoid the strict condition of homogeneity across individuals. Rather than construct a model based on pooled data of aggregates that are assumed to be homogenous, the person-specific framework involves constructing individual time series models separately. The individual models are allowed to be different for each individual without any homogeneity assumption.

In an attempt to address heterogeneity within model features, several interindividual methodological techniques have been developed to model this heterogeneity within a population. The most popular example of this is the growth mixture model (Muthen and Muthen, 2000), where growth model parameters can differ across different subgroups. Although this modeling technique does make an attempt to address model heterogeneity, it still applies group level inferences on individual processes. The potential flaws of this perspective can be highlighted by

the popular American fiction writer, Jack Vance, in his book, *The Languages of Pao* (1958): "Any collocation of persons, no matter how numerous, how even their homogeneity, how firmly they profess common doctrine, will presently reveal themselves to consist of smaller groups espousing variant versions of the common creed; and these sub-groups will manifest sub-sub-groups, and so to the final limit of the single individual."

Nomothetic Inferences

It may appear as if the person-specific perspective may neglect studying or discovering any general scientific laws. Since the focus is on the individual, is there any room for making nomothetic inferences? While person-specific methodology may initially make idiographic influences, this perspective can also be an important part of nomothetic research.

A useful way of characterizing conventional interindividual methods from person-specific methods for making nomothetic inferences is to contrast them as "top-down" vs. "bottom-up" approaches. Figure 0.1 gives a visual depiction of this contrast in nomothetic inference. Conventional interindividual methods initially construct a group model and make inferences from this initial group model as nomothetic. This method may look for variation in this group model via mixed effects or mixtures, but still makes inferences on individuals from some initial group model in a "top-down" direction. In contrast, person-specific methodology begins by modeling the individual in order to properly capture within-person dynamics. After several individual models have been fit, parameters or other features of these individual models can be submitted to group-level analysis in order to see potential interindividual differences in these intraindividual processes. This process of forming nomothetic inferences from an initial individual-level focus can be seen to take a "bottom-up" direction. The person-specific

perspective can make nomothetic claims, but it incorporates the rich within-person data and dynamics in doing so. The remainder of this introduction highlights some specific person-specific methods that will be implemented in this dissertation.

The State Space Model

The state space model is one of several time series modeling approaches used in person-specific analysis. A state space model consists of a stochastic dynamic system model of state (or latent) variables with Gaussian noise and a measurement model relating these unobservable states to manifested observations. The most basic form of the state space model involves the discrete time links between states and observable variables. The model for state dynamics takes the form:

$$\eta_t = B\eta_{t-1} + \zeta_t \quad (0.1)$$

where η_t is a q-variate state vector at time *t*, B is a (q,q)-dimensional matrix containing transition (or autoregressive) coefficients, and ζ_t is a q-variate Gaussian process noise vector at time *t*. The measurement model takes the form:

$$y_t = \Lambda\eta_t + \varepsilon_t \quad (0.2)$$

where y_t is a p-variate observed variable vector at time *t*, Λ is a (p,q)-dimensional matrix containing measurement loading coefficients, and ε_t is a p-variate Gaussian measurement error vector at time *t*. Both of the stochastic components are assumed to have normal distributions,

with $\zeta_t \sim N(0, \Psi)$ and $\varepsilon_t \sim N(0, \Theta)$. An example diagram of a state-space model with 2 state variables and 4 observed variables is depicted in Figure 0.2.

A popular method for estimating state space model is the Kalman filter (KF; Kalman, 1960), which involves a recursive estimation scheme. Recursive methods in the KF framework involve a sequence of prediction and updating stages that can be summarized in a series of steps. First, the dynamic model predicts the states at the current time point based on initial or a priori values. Second, these predicted states are updated based on deviance of these predictions from the observed data at the current time as well as model components that contain uncertainty about the dynamic and measurement models. Next, these updated states are used to make predictions at the subsequent time point (first step repeated), which are updated based on data from the observed data at the subsequent time point (second step repeated). This prediction and update process repeats through each time point, starting at the beginning of the time series and ending at the final time point. In addition, predictions and updates of the state error covariance matrix occur alongside predictions and updates of the previously mentioned actual states. The KF recursion is based on recursive least-squares estimation; a clear demonstration of the derivation of the filter is provided in Simon (2006). Additional details regarding parameter estimation and maximum-likelihood (ML) approaches are given in Harvey (1990) and Gibbs (2011). The Appendix contains details on the recursive estimation procedure as well as the basic steps and equations for ML parameter estimation. Various extensions and types of the KF have been developed for different model types (i.e., linear vs. nonlinear) and temporal types (i.e., discrete time vs. continuous time); overviews of these different KF types can be found in Simon (2006) and Gibbs (2011).

Another popular method for fitting state space models is with time series structural equation modeling (SEM). Time series SEM involves fitting similar models to block-Toeplitz covariance matrices that incorporate the autocovariance structure of the data. Modeling these covariance structures involves pooling data across time; such procedures assume equal temporal spacing and neglect any data regarding the time intervals. In contrast, the recursive estimation schemes with the KF framework involve raw data procedures (analogous to full information maximum likelihood procedures in SEM). Further details of the block-Toeplitz estimation scheme can be seen in the Appendix. Other conceptual and practical comparisons of the KF and time series SEM have been addressed in the literature (Molenaar, 2003; Zhang, Hamaker, & Nesselroade, 2008; Chow et al., 2010).

Although initially developed in the engineering literature, the state space model has recently been applied in time series modeling of psychological processes. Some example applications include research on emotion-cognition linkages (Chow, Hamagani, & Nesselroade, 2007), dyadic relations (Chow, Ferrer, & Nesselroade, 2007), and affective processes (Molenaar et al., 2009). The following two papers within this dissertation give additional applications of the state space model to psychological research.

Figure 0.1: Nomothetic approaches

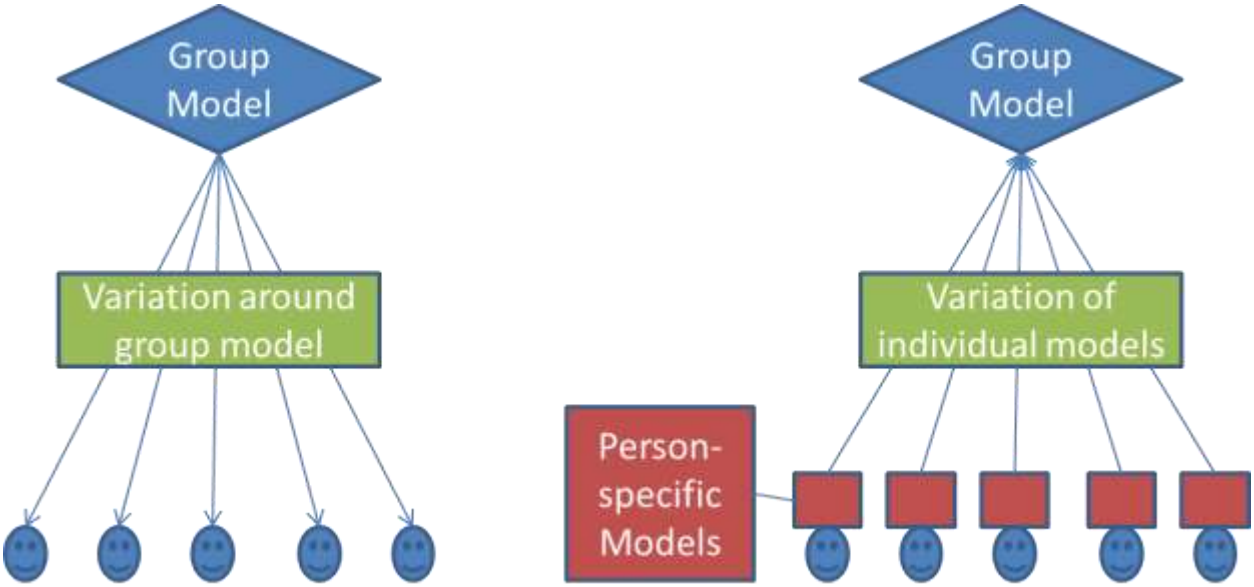

Fig. 0.1. This figure gives a heuristic contrast between conventional and person-specific approaches. The structure on the left displays a "top-down" approach while the structure on the right displays a "bottom-up" approach.

Figure 0.2: Depiction of an example state-space model.

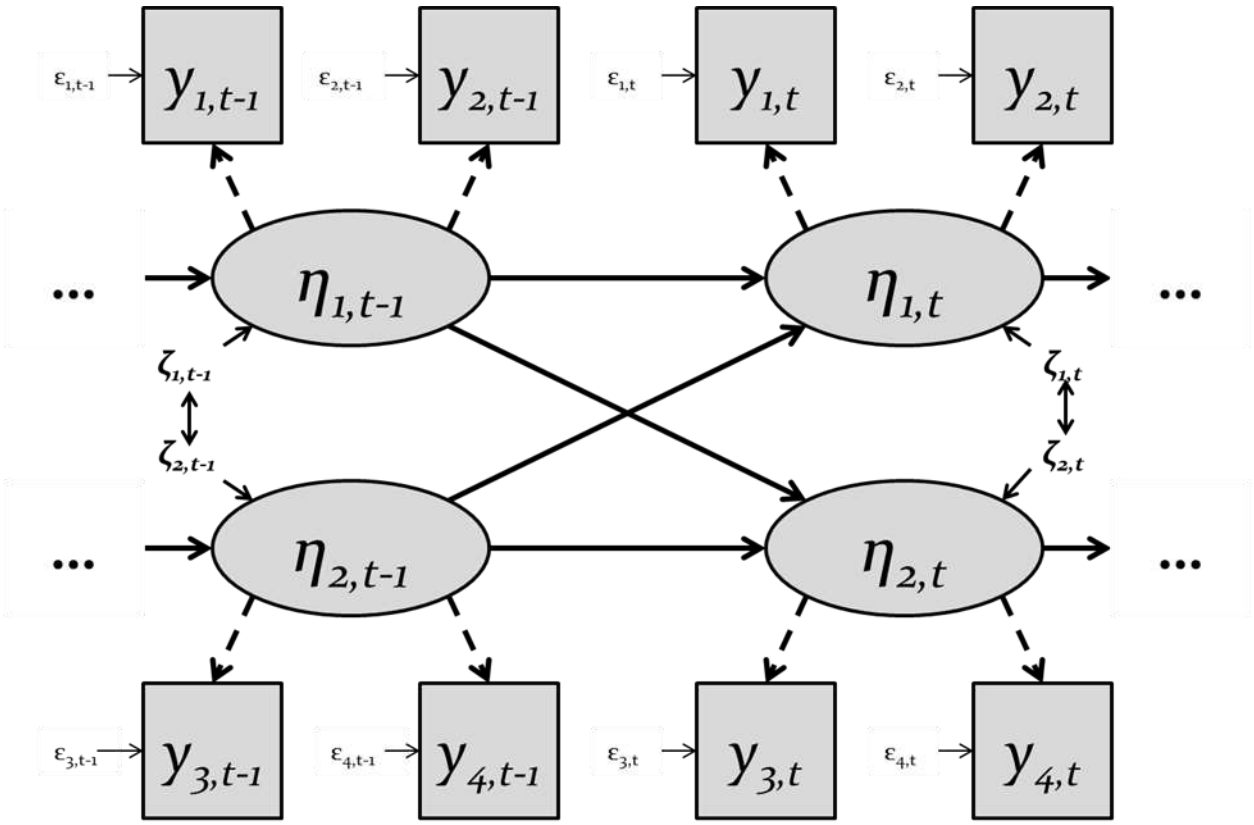

Fig. 0.1. The two state variables (η_1, η_2) at time *t-1* predict their values at the subsequent time point *t* with each state also being influenced by a Gaussian white noise component (ζ_1, ζ_2) that can covary concurrently. These unobserved state variables map onto observed variables (y_1, y_2, y_3, y_4) at concurrent time points and contain measurement errors (ε_1, ε_2, ε_3, ε_4).

References

Borkenau, P. & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Personality Research, 32,* 202–221.

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Chow, S-M., Ferrer, E. & Nesselroade J. R. (2007). An unscented Kalman filter approach to the estimation of nonlinear dynamical systems models. *Multivariate Behavioral Research, 42 (2),* 283-321.

Chow, S-M., Hamagami, F, & Nesselroade, J.R (2007). Age differences in dynamical emotion-cognition linkages. *Psychology and Aging, 22 (4),* 765-780.

Chow, S-M., Ho, M.H.R., Hamaker, E., & Dolan, C. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling, 17,* 303-332.

Clark D.M. & Teasdale J.D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of Abnoral. Psychology 91,* 87-95.

Gibbs, B.P. (2011). *Advanced Kalman filtering, least-squareds and modeling: a practical handbook.* Hoboken, N.J.: Wiley.

Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering, 82,* 35-45.

Laird, N.M. &Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics, 38 (4),* 963–974.

Kelderman, H. & Molenaar, P.C.M. (2007). The effect of individual differences in factor loadings on the standard factor model. *Multivariate Behavioral Research, 42*, 435-456.

Molenaar, P.C.M. (1997). Time series analysis and its relationship with longitudinal analysis. *International Journal of Sports Medicine, 18,* 232–237.

Molenaar, P.C.M., Huizenga, H.M., & Nesselroade, J.R. (2002). The relationship between the structure of inter-individual and intraindividual variability: A theoretical and empirical vindication of Developmental Systems Theory. In U.M. Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 339–360.). Dordrecht, The Netherlands: Kluwer.

Molenaar, P. C. M. (2003). *State space techniques in structural equation modeling: Transformation of latent variables in and out of latent variable models.* Unpublished manuscript, University of Amsterdam. Retrieved from http://www.hhdev.psu.edu/hdfs/faculty/pubs/StateSpaceTechniques.pdf

Molenaar, P.C.M. (2004). A manifesto on Psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2 (4),* 201-218.

Molenaar, P.C.M. (2007). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology, 50,* 60-69.

Molenaar, P.C.M., & Campbell, C.G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychology, 18 (2),* 112-117.

Molenaar, P.C.M., Sinclair, K.O., Rovine, M.J., Ram, N.,&Corneal, S.E. (2009). Analyzing developmental processes on a individual level using non-stationary time series modeling. *Developmental Psychology, 45*, 260–271.

Muthen, G., & Muthen, L. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, 24,* p. 882-891.

Rovine, M.J., Molenaar, P.C.M., & Corneal, S.E. (1999). Analysis of emotional response patterns for adolescent stepsons using P-technique factor analysis. In R. K. Silbereisen & A. von Eye (Eds.), *Growing up in times of social change* (pp. 261–285). Berlin: De Gruyter.

Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience.

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Yang, M-S. & Chow, S-M. (2010). Using state-space models with regime switching to represent the dynamics of facial electromyography (EMG) data. *Psychometrika, 74(4),* 744-771

Zhang, Z., Hamaker, E. L., & Nesselroade, J. R. (2008). Comparisons of four methods for estimating dynamic factor models. *Structural Equation Modeling, 15 (3),* 377-402.

Paper 1: Time-varying genetic factor models for modeling somatic mosaicism and epigenetic effects

1.1: Abstract

Genetic factor models such as the additive genetic, common environment, and unique environment (ACE) model are prevalent in twin designs within behavior genetic research. New perspectives in molecular genetics research, such as epigenetic development and somatic mosaicism, challenge some of the fundamental assumptions of this analytic framework. The current paper describes novel extensions of time series genetic factor models that include time-varying parameters that can address some of these developmental processes. The Kalman filter and its time-varying parameter extension are briefly introduced and demonstrated within a simulation design that reflects genetic development proposed by these molecular genetic theories. This novel modeling approach is expected to become invaluable in testing particular theories of development under genetic and environmental influence.

1.2: Introduction

Twin models were a major part of the earliest statistical developments in social science (e.g., Fisher 1918; Galton, 1889; Pearson, 1904). These statistical methods would develop into factor analytic models for examining twin covariance structures decomposed into additive genetic, common environment, and unique environment factors (ACE; Martin and Eaves, 1977). The ACE factor model has become one of the more prevalent statistical tools in behavior genetics using twin designs. Time series extensions of the ACE have allowed researchers to conduct twin-specific modeling (analogous to person-specific models). The current paper outlines developments in the ACE model and presents novel adaptations of the model to address new findings from molecular genetics research.

Standard Interindividual ACE Model

The ACE model contains several design requirements and statistical assumptions. The model requires at least two sibling types that produce different genetic intra-sibling covariance structures. Classic twin designs, which involve monozygotic (MZ) and dizygotic (DZ) twin pairs, are the primary research designs in which the ACE model is used. The model is essentially a multi-group constrained confirmatory factor analysis model with latent factors corresponding to ACE components. The multi-group scheme of the factor model corresponds to different twin types being represented as different groups. This factor model can be represented for univariate phenotypes by:

$$y_{i,j,k} = aA_{i,j,k} + cC_{i,j,k} + eE_{i,j,k} \quad (1.1)$$

where $y_{i,j,k}$ is the observed phenotype for the *j*th member (*j* = 1, 2) of the *i*th twin pair (*i* = 1, 2, …) of type *k* (*k* = 1for MZ and 2 for DZ). *A*, *C*, and *E* represent the individual values of the three components for the *j*th member of the *i*th twin pair of type *k*; each of these latent variables are usually assumed to have Gaussian distributions with fixed unit variances. The *a*, *c*, and *e* values represent the corresponding factor loadings of the three components on the phenotype and are assumed to be invariant across individuals. In the case of a phenotype with multiple indicators, a measurement error term can be incorporated into the model; otherwise the measurement error is absorbed into the non-shared latent factor, *E*. Extensions of this model, including factors associated to dominant gene effects, are elaborated in Neale and Cardon (1992).

A major feature of the model is the manner in which the factor covariance matrix is structured. This factor covariance matrix, usually denoted as Ψ in the factor analysis literature, contains a constrained pattern of values that require certain assumptions about the average genetic relatedness between twin pairs. In the model, the intra-twin genetic covariances take the pattern with $cor(A_{i,1,1}, A_{i,2,1})$ = 1.0 for the MZ twin correlation and $cor(A_{i,1,2}, A_{i,2,2})$ = 0.5 for the DZ twin correlation. This corresponds to the assumption that genes between MZ twins are identical and thus correlate perfectly while DZ twins share half their genes on average, and thus have a 0.5 correlation in their genetic factors. Likewise, $cor(C_{i,1,1}, C_{i,2,1}) = cor(C_{i,1,2}, C_{i2,,2})$ = 1.0, representing the shared-environment intra-twin covariance structure that is similar across twin types. A common representation of the genetic factor model follows a block matrix structure in the Linear Structural Relations framework (LISREL; Jőreskog & Sőrbom, 1993), where Ψ contains diagonal blocks which correspond to individual twin members and the intra-twin covariance constraints appear in the off-diagonal blocks; further details are given in Boomsma and Molenaar (1986).

Time-series ACE Models

Time-series extensions of the ACE model were developed in order to examine potential heterogeneous person-specific and twin-specific models. One of the first proposed genetic factor models for intra-twin analysis combined principles of the idiographic filter (IF; Nesselroade et al., 2007) and the ACE model into the IFACE (Nesselroade & Molenaar, 2010). This model could be implemented with standard structural equation modeling (SEM) software and required a time series of twin data. The main principles of the original IF framework are person-specific measurement loadings and common latent structural relations. In the IFACE hybrid, this corresponds to person-specific genetic influences (modeled by factor loadings) and common intra-twin genetic relatedness (modeled by latent variable correlations).

Molenaar (2010), building on the IFACE model, proposed an intra-twin time series model that relaxed the constraint of equivalent structural relations. This model allows most model components to be heterogeneous with the exception of the shared environment intra-twin genetic covariance being fixed at 1.0 as in standard genetic factor models. A key feature of this model is that the intra-twin genetic correlations are allowed to vary across twin pairs and are estimable parameters. Molenaar et al. (2012) provide justification for the identifiability of this model based on symbolic matrix calculus criteria (Bekker et al., 1994) and give an empirical demonstration with neurophysiological twin time series data. The model is comprised of (i) an observation model giving estimates of the influences of the traditional ACE factors and (ii) a dynamic model that explains both the time evolution of the ACE factors as well as the relatedness between twins. The observation model is similar to (1.1):

$$y_{i,j,k,t} = a_{i,j,k}A_{i,j,k,t} + c_{i,j,k}C_{i,j,k,t} + e_{i,j,k}E_{i,j,k,t} \qquad (1.2)$$

In (1.2), an additional index *t* for time is added to denote that the observed phenotype and latent factors follow a time series. The factor loadings *a*, *c*, and *e* are now allowed to vary across individuals. Again, a measurement error term may be added in situations where multiple phenotypic indicators are available. The dynamic model is:

$$\begin{aligned} A_{i,j,k,t} &= \alpha_{i,j,k} A_{i,j,k,t-1} + \delta_{i,j,k,t} \\ C_{i,j,k,t} &= \beta_{i,j,k} C_{i,j,k,t-1} + \zeta_{i,j,k,t} \\ E_{i,j,k,t} &= \gamma_{i,j,k} E_{i,j,k,t-1} + \kappa_{i,j,k,t} \end{aligned} \tag{1.3}$$

In (1.2), α, β, and γ are time-invariant, person-specific autoregressive relationships that describe the sequential dependency in the ACE factors while δ, ζ, and κ are the process noise components. All process noise components are uncorrelated except for the intra-twin shared environment correlation, $cor\left(C_{i,1,k,t}, C_{i,2,k,t}\right)$, which is fixed at 1.0, and the intra-twin genetic correlation, $cor\left(A_{i,1,k,t}, A_{i,2,k,t}\right)$, which is freely estimated and time-invariant. This model can be fit with the block-Toeplitz SEM procedure that is explained in detail in Molenaar (2012). Lo and Molenaar (in prep) investigated the estimation accuracy in recovering twin-specific genetic correlation in a simulation and found that the model provided satisfactory results in a variety of conditions.

Figure 1.1: Diagram of time series ACE model

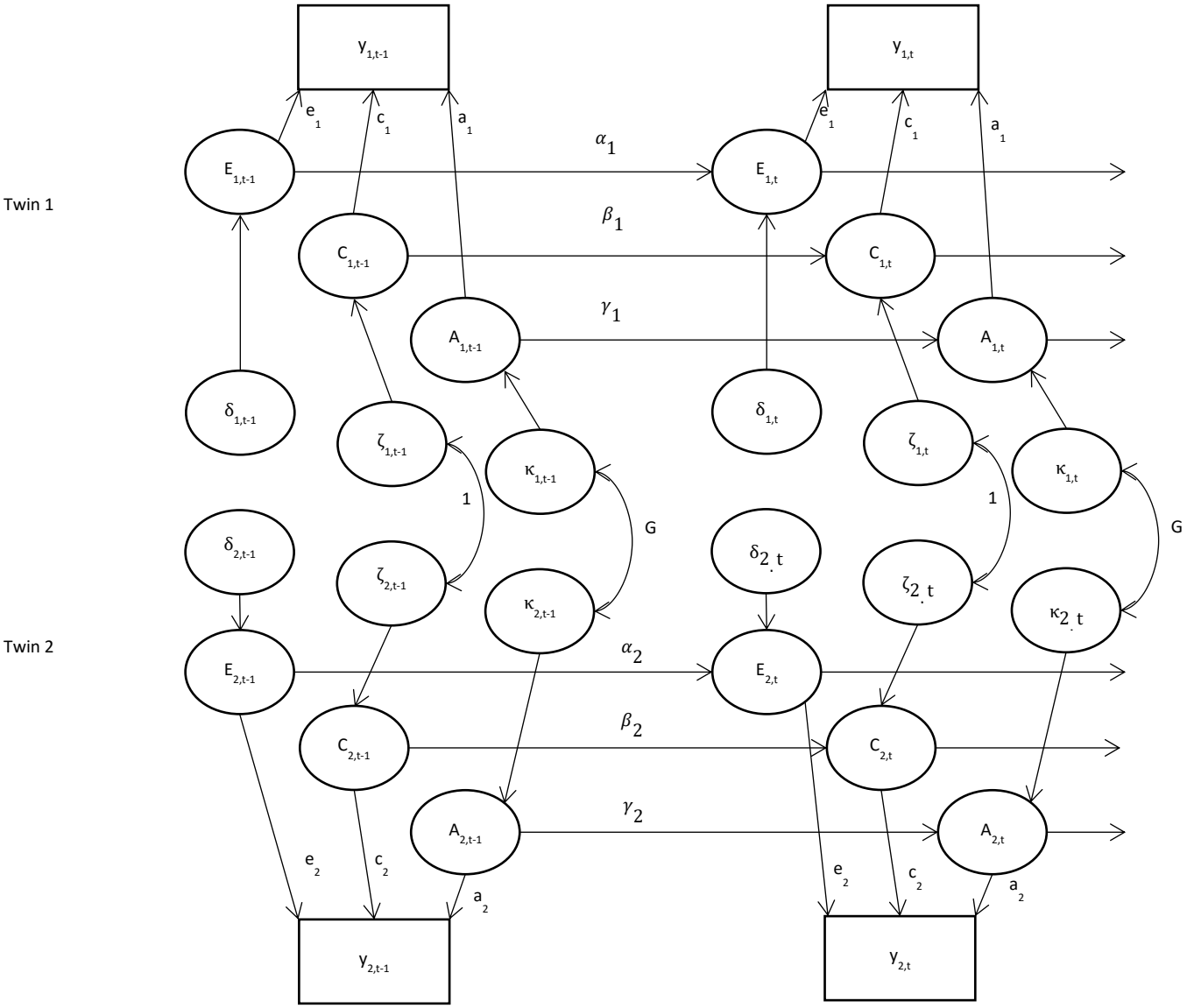

Fig. 1.1. An example structural diagram of a univariate time series ACE model for a single twin pair. Twin 1 appears on the top while twin 2 appears on the bottom. The parameters are described in 1.2 and 1.3 with pair and type indices omitted for simplicity (*i* and *k* indices).

Criticisms from Molecular Genetics

Recent developments in molecular genetics have presented challenges to behavior genetics research, specifically the ACE model. The first issue is epigenetics, which examines changes in gene expression that may be influenced by environmental factors. Second is somatic mosaicism, which concerns variance in the underlying genome that can increase developmentally. Both of these biological processes present unique challenges to behavior genetics and the ACE model.

Epigenetics is the study of heritable changes in gene expression that do not alter the genome. This term was initially given by Waddington (1942) in his research on cell differentiation. Molenaar et al. (1993) described epigenetics as a third source of developmental differences that is neither genetic nor environmental. Epigenetics involves non-linear dynamic processes that give rise to biological pattern formation and self-organization (Molenaar and Raijmakers, 2000). Although there are several epigenetic processes, the two most well-known types of epigenetic changes are DNA methylation and histone modification. DNA methylation involves addition of methyl groups to DNA, where highly methylated areas tend to be less transcriptionally active. In histone modification, the amino acids that make up histone proteins within DNA are changed and subsequently cause changes in DNA transcription. In both cases the epigenetic modification can persist from the germ line of a parent into the zygote, thus making this change heritable. A thorough overview of epigenetic processes is given in Jablonka and Lamb (2005). In general, epigenetic processes can promote or inhibit expression of particular genes and may exhibit strong reactivity to environmental influences.

Several simulation studies have examined how epigenetic influences may effect conventional ACE modeling. Eaves et al. (1999) and Molenaar and Raijmakers (1999) used

similar approaches by simulating models with the discrete logistic equation. In both cases, these nonlinear dynamic models produced chaotic behavior and resulted in decreasing heritability estimates that fell to zero. In contrast, Kan et al. (2010) simulated a nonlinear model that produced bifurcations instead of chaotic behavior; this resulted in increasing heritability and a developmental shift of shared to non-shared environmental influence. These features of increasing heritability and change from shared to non-shared environmental influence are characteristic of the findings within the behavior genetics literature on several phenotypes (Turkheimer, 2000). All of these simulation designs yielded time-varying heritability estimates based on standard genetic models despite the actual simulation models constraining genetic influence to be constant across time.

Although this simulation work has demonstrated the problems with ACE model when dealing with data containing epigenetic influence, no specific extensions of the ACE model to account for epigenetic influences have been proposed. It is difficult to imagine how the ACE model could incorporate an additional factor beyond the genetic and environmental components since the model relies on the contrasts between these factors in the factor correlation matrix to be estimable (although Wright, 1920, proposed models to estimate third source factors in animal inbreeding designs). An alternative approach would be to examine epigenetic influence following multiple twin time series ACE models on a particular phenotype. These multiple ACE models would provide a sample of individually obtained genetic factor loadings. The variance of these factor loadings or how these genetic loadings are predicted by some environmental factor could be an initial method of investigating the epigenetic influence related to a phenotype of interest. It is also possible that models with time varying genetic factor loadings may be incorporated into an epigenetic analysis. Such models would characterize how genetic influence

on a phenotype changes across time; these changes in genetic influence may be due to epigenetic influences. Both of these approaches would rely on twin and person specific time series ACE models as depicted in equations (1.2) and (1.3) since this would allow investigation of varying genetic influences on phenotypes (via varying genetic factor loadings) rather than homogenous genetic influence. Variation and change in genetic influence in these models could be an initial proxy for epigenetic influence.

Several genomic processes support the idea of somatic mosaicism, a description of the genetic diversity within an organism where different cells and tissues within an individual contain different genomes. These processes occur during development and produce genetic variation within an individual that can be driven by environmental, activity-dependent, or stochastic causes. Two of the main sources of somatic mosaicism are retrotransposons and copy number variations (CNVs). Retrotransposons (De, 2011), which are often referred to as "jumping genes", can carry out a process of (i) copying themselves to RNA and (ii) subsequent reverse-transcription into DNA; this effectively inserts a copied DNA sequence into a new location within the genome. These within-person changes in the genome will vary between persons (due to differential environmental exposure, activity, and chance) and will therefore produce genetic differences between genetically related individuals during development (Singer et al., 2010). CNVs are a source of DNA variation that involves a variety of structural variations in DNA. Unlike single nucleotide polymorphisms (SNPs), the most familiar form of DNA variation involving single nucleotides, CNVs are associated with complex multi-site variants that affect chromosomal structure (Redon et al., 2006). CNVs have been demonstrated to produce within-individual somatic mosaicism (Rodriquez-Santiago et al., 2010) and genetic differences between MZ twins (Bruder, 2008). Yet another form of somatic mosaicism can occur outside

the cell nucleus: mitochondrial DNA can also show substantial variation within an individual (Pfeiffer et al., 2004).

Charney (2012) discusses the potential impact of somatic mosaicism on ACE modeling. He alludes to two key assumptions of conventional ACE models that are challenged by somatic mosaicism. First, conventional ACE models assume that MZ twins will have identical genes while DZ twins will on average share half of their genes. As previously mentioned, somatic mosaicism can produce genetic differences between MZ twins and also cause further genetic differentiation between DZ twins. Second, group-based longitudinal ACE models assume that the genetic relatedness between twins is invariant across time. In contrast, somatic mosaicism may produce trajectories of genetic relatedness that decrease over time. In both cases, the increasing within-person genetic variance caused by somatic mosaicism will induce decreasing genetic similarity between twins. ACE models are incapable of capturing this potential developmental process of decreasing genetic similarity.

Time series ACE models are also susceptible to the challenges introduced by somatic mosaicism. The model depicted in equations (1.2) and (1.3) is capable of estimating twin-specific genetic correlation parameters and therefore avoids the first problematic assumption discussed by Charney. The model, however, still assumes time-invariant genetic correlation that is constant across the analyzed time series. As a result, the model is unable to capture the potential decreasing genetic similarity produced by somatic mosaicism.

The Current Study

In response to the challenges of epigenetics and somatic mosaicism on behavior genetic modeling, the current paper proposes extensions of the time series ACE model to accommodate

time-varying genetic factor loading and correlation parameters. The main addition is to make target parameters follow a scaled linear trend:

$$\theta_t = \theta_0 + \frac{\theta_1 t}{T} \quad (1.4)$$

where θ_t is the parameter of interest (e.g., genetic factor loading or genetic correlation), θ_0 is the intercept for the parameter, θ_1 is the linear trend coefficient, and *T* is the total number of time points (used for scaling purposes). This modeling approach can be handled with the Kalman filter (KF; Kalman, 1960) with time-varying parameter approaches detailed in Molenaar (1994). This may be the first model in behavior genetics twin research that can investigate possible somatic mosaicism since it is capable of estimating twin-specific genetic correlations that can vary across time. Although a simple linear trend model is used in this study, these same estimation principles should be able to apply to other functions such as logistic or polynomial trends in order to capture nonlinear trajectories of genetic correlation. A linear trend may be an initial approximation of these developmental changes that can be further improved with higher order polynomial terms. Time-varying factor loadings with a similar linear trend adjustment are also investigated in this study. The time-varying factor loading model, as previously mentioned, may be a potential approach to investigate epigenetic effects.

The current paper involved three simulations to answer separate questions related to genetic modeling. The first simulation examines the robustness of the conventional interindividual ACE model (equation 1.1) when assumptions of twin genetic correlations are violated. The second simulation assesses the performance of the Kalman filter for estimating the twin-specific ACE model (equation 1.3) with a linearly decreasing genetic correlation following

equation 1.4. Similar to the second simulation, the third simulation assesses the performance of recovering time-varying factor loadings. While the second simulation can determine how these models capture somatic mosaicism effects, the third simulation deals with the ability to capture epigenetic effects.

1.3: Methods

Simulation 1

The first simulation was conducted in order to see how ACE model parameters are affected when the standard ACE genetic correlation assumptions are violated. In order to focus on this specific question, a simple univariate phenotype ACE model without measurement errors was implemented. There were two main simulation factors that were manipulated in the simulation. First, the ACE factor loadings had various conditions representing different proportions of variance attributed to the different factors. The conditions consisted of various permutations of [.2, .5, .8] resulting in different proportional contribution of the ACE factors; the conditions are explicitly given in Table 1.1. The second simulation design factor was the degree of violation of the genetic correlation assumption; this was manipulated by simulating data with decreased genetic correlation between twins while estimating the models with the standard MZ=1.0/DZ=0.5 genetic fixed constraints. There were three conditions where the simulated genetic correlations were decreased by 0 (no change baseline condition), .1, and .2 for both MZ and DZ twins. For simplicity, the sample sizes of both MZ and DZ twins were held constant at 100 each.

These two design factors were fully crossed with 1,000 replications within each simulation design cell with the main criteria being bias and the root means square deviation (RMSD) in the factor loading estimates. Here, bias refers to the average discrepancy between the estimated and true parameter across simulation replications:

$$Bias(\theta) = \sum_{i=1}^{R} \frac{\hat{\theta}_i - \theta_i}{R} \tag{1.5}$$

where $\hat{\theta}_i$ is the parameter estimate for replication *i*, θ_i is the true value for replication *i*, and *R* is the total amount of simulation replications. In interpreting bias, positive values indicate a tendency to overestimate the parameter, negative values indicate underestimation tendency, and values close to zero indicate accuracy. The RMSD is an accuracy statistic that represents the square root of the average squared deviation:

$$RMSD(\theta) = \sqrt{\sum_{i=1}^{R} \frac{(\hat{\theta}_i - \theta_i)^2}{R}} \quad (1.6)$$

The RMSD produces a positive value that represents the average inaccuracy of estimation regardless of direction (i.e., overestimating or underestimating).

Simulation 2

The second simulation examined the ability of the Kalman filter to capture a linearly changing genetic correlation in a twin-specific ACE model. There were three main simulation factors that were manipulated in this simulation. First, the starting point (or intercept, represented as θ_0 in equation 1.4) of the genetic correlation was set to either 1.0 or 0.5, representing scenarios where the twins would start off as either MZ or DZ twins. Second, the degree of change across the time (represented as θ_0 in equation 1.4) was manipulated to be either -.25 or -.5, representing different degrees of genetic correlation decrease due to somatic mosaicism. Third, three time series lengths of 100, 500, and 1000 were used. These three design factors were fully crossed with 100 replications per design cell. The computational demands and

lengthy estimation times per model required a relatively limited amount of simulation replications.

There were several other features that were constant across the manipulated simulation design factors. There were three phenotypic variables per twin and measurement error variances were incorporated into the simulation model. The pattern of factor loadings was constant and is depicted in Table 1.2. All measurement error variances were set at 0.1 and were estimated with these parameters constrained to be equal. The process noise variances were set at 1.0 and fixed at those values during estimation. The autoregressive parameters for all ACE factors were set at 0.7 and were estimated without equality constraints.

Simulation 3

The third simulation study assessed the recovery of time varying factor loadings. This simulation was nearly identical to simulation 2 except for this change in focus. In this simulation the genetic correlation was simulated at a constant 0.5 and freely estimated during model fitting. The factor loading patter was similar to the one used in simulation 2, except the (1,1) element was simulated as a linearly changing parameter (the element indicated with a * in Table 1.2). There were similarly three manipulated simulation factors, with the sample size conditions being similar to simulation 2. The other simulation design factors were the starting point of the factor loading and the degree of linear change. The combination of starting points and degrees of change yielded 4 scenarios: (i) $\theta_0 = .7$, $\theta_1 = -.6$, (ii) $\theta_0 = .7$, $\theta_1 = -.3$, (iii) $\theta_0 = .1$, $\theta_1 = .6$, and (iv) $\theta_0 = .4$, $\theta_1 = .3$. The first two scenarios represent different magnitudes of degrees from the same higher starting point while the third and fourth scenarios represent an increase to the same value (.7) from two different change magnitudes and starting points. The measurement noise variances,

process noise variances, autoregressive parameters, and number of simulation replications were similar to simulation 2.

Table 1.1

Simulation 1 loading value conditions

Condition	Loading values [a,c,e]
1	[.8, .5, .2]
2	[.8, .2, .5]
3	[.5, .8, .2]
4	[.5, .2, .8]
5	[.2, .8, .5]
6	[.2, .5, .8]

These loading values were simulated in a larger factor loading with dimensions 2x6. These values appear in the first 3 elements in row 1 and the last 3 elements in row 2.

Table 1.2

Simulation 2 factor loading pattern

Observed	Twin 1 ACE factors			Twin 2 ACE factors		
variables	$A_{1,t}$	$C_{1,t}$	$E_{1,t}$	$A_{2,t}$	$C_{2,t}$	$E_{2,t}$
$y1_{1,t}$	.7*	.5	.5	0	0	0
$y2_{1,t}$	.5	.5	.7	0	0	0
$y3_{1,t}$	.4	0	.9	0	0	0
$y1_{2,t}$	0	0	0	.5	.7	.5
$y2_{2,t}$	0	0	0	.5	.5	.7
$y3_{2,t}$	0	0	0	0	.4	.9

y1-y3 represent the different observed phenotypes. The first subscript indicates the twin within the dyad while the second subscript indicates the time point. The first element (indicated with a *) was manipulated in simulation 3.

1.4: Results

Simulation 1

There appear to be trends in estimation bias associated with both simulation design conditions in simulation 1. The results of this simulation are displayed in Table 1.3. First, it appears as if the direction of bias is consistent based on which parameter is being examined. Both the additive genetic (*a)* and common environment (c) loadings appear to be consistently negative in bias conditions while the unique environment (e) loadings tend to be consistently positive; this suggests that both the *a* and *c* loadings tend to be underestimated while the *e* loadings tends to be overestimated. An exception to this is the last factor loading pattern (pattern 6 corresponds to *a*=.2, *c*=.5, *e*=.8), where the *a* loadings are slightly positive biased and the *e* loadings are slightly negatively biased.

There appear to be consistent effects related to the genetic correlation bias conditions that occur across all of the loading patterns. Increasing the simulated bias in the genetic correlation appears to increase the negative bias in both the *a* and *c* loadings and increase the positive bias in the *e* loadings. Since these three types of loadings contribute to the total variance of an observed phenotype variable, it appears as if the underestimation of both the *a* and *c* loadings are absorbed by an inflated *e* loading.

The differing loading patterns also seem to produce various effects on estimation bias of the loadings. For the *a* loading, loading pattern conditions with larger simulated *a* values (i.e., patterns 1 and 2) produced larger estimation bias of the *a* loading parameter while the loading patterns with smaller simulated *a* values (i.e., patterns 5 and 6) produced small negative bias and even positive bias in pattern 6 as previously mentioned. A similar trend is seen with the *c* loading, where patterns with smaller simulated *c* values (i.e., patterns 2 and 4) had smaller

negative bias than the patterns with larger simulated *c* values (i.e., patterns 3 and 5). For the *e* loading, it appears as if the positive bias for this parameter is larger in conditions where this loading is simulated to be smaller (i.e., patterns 1 and 3).

Simulation 2

Simulation 2 assesses the accuracy of the Kalman filter in recovering the intercept and slope of a linearly changing genetic correlation within a twin time series ACE model. Here the slope parameter, $\hat{\theta}_1$, is scaled according to equation 1.4 and actually corresponds to the total decrease of the genetic correlation across the time series. Across simulation conditions, it appears as if the intercept parameters, $\hat{\theta}_0$, are consistently negatively biased, thus tend to be underestimated. An exception to this is some of the conditions where the simulated intercept is 0.5. In interpreting the bias of the slope parameter, the simulated values are always negative (i.e., either -.25 or -.5) so positive bias values indicate that the magnitude of decrease is being underestimated. It appears as if the slope bias is consistently positive across conditions, indicating that the magnitude of genetic correlation decrease is consistently underestimated.

There are some clear trends related to both the simulated intercept and bias conditions. Conditions where the simulated intercept is set to 1.0 produces larger negative bias in estimating the intercept parameter. In conditions where the simulated intercept is 0.5 and the sample sizes are larger, this parameter is slightly positively biased, as previously indicated. The intercept conditions also appear to affect recovery of the slope parameter; the slope parameter has stronger positive bias in conditions where the intercept is simulated at the larger value. The simulated slope seems to have little effect on bias of either parameter except for the case of the slope

parameter in the smallest sample size condition; in this sample size, larger simulated decreases produced larger positive bias in the slopes.

The sample size conditions appear to have a drastic effect on accuracy of both types of parameters. Bias for both parameter types seem to decrease sharply as the length of the time series increases. This suggests that relatively long time series lengths are needed in order to accurately recover a changing genetic correlation.

Simulation 3

Simulation 3 assesses the accuracy of the Kalman filter in recovering the intercept and slope of a linearly changing genetic factor loading within a twin time series ACE model. Unlike the first two simulations, it does not appear as if there is consistent positive or negative bias in results, although most of the bias appears to be positive. Again, the slope parameter, $\hat{\theta}_1$, is scaled according to equation 1.4 and actually corresponds to the total decrease of the factor loading across the time series. Unlike simulation 2, the factor loading change design factor included conditions where the parameter would increase. When interpreting the bias, positive or negative bias correspond to either overestimation or underestimation depending on whether the factor loading; for example, positive bias in a decreasing loading condition would correspond to underestimating the magnitude of decrease while positive bias in an increasing loading condition would correspond to an overestimation of the magnitude of increase.

There are several notable trends in bias related to the simulated intercept and slope conditions. It appears as if the larger intercept conditions produced smaller bias and RMSD for estimating the corresponding intercept parameter. The recovery of the slope parameter also appears to be better in these larger simulated intercept conditions, although this effect does not

seem as pronounced as it is for the recovery of the intercept parameter. The magnitude of change in the simulated slope appears to also have an effect on estimation accuracy of both parameter types. Larger simulated slopes, regardless of direction, appear to decrease the estimation accuracy of both the intercept and slope.

Similar to what was found in simulation 2, sample size seems to be an important design factor. Again, larger time series length produced better accuracy in terms of bias and accuracy across simulated intercept and slope conditions. This improvement does not seem as drastic for recovering the factor loading, however, as estimation is fairly accurate at even lower sample sizes. In general, the bias and RMSD for these parameters appear to be smaller than those found in simulation 2 with the linear changing genetic correlation; this suggests that time-varying factor loadings are much easier to recover than time varying factor loadings.

Table 1.3

Simulation 1 results

Loading pattern	ΔG	Bias *a*	Bias *c*	Bias *e*
1	0	-.005	-.014	.003
1	.1	-.026	-.075	.123
1	.2	-.104	-.130	.245
2	0	-.047	-.007	-.002
2	.1	-.096	.011	.074
2	.2	-.107	-.084	.123
3	0	-.011	-.010	.002
3	.1	-.012	-.092	.156
3	.2	.003	-.291	.197
4	0	-.097	.002	.005
4	.1	-.081	-.040	.017
4	.2	-.125	-.054	.025
5	0	-.080	-.012	.007
5	.1	-.070	-.100	.113
5	.2	-.081	-.136	.151
6	0	.042	-.064	-.020
6	.1	.047	-.086	-.015
6	.2	.006	-.086	-.001

The loading patterns 1-6 are depicted in Table 1.1. ΔG reflects the simulated bias (the constrained value – simulated value) in the genetic correlation.

Table 1.4

Simulation 2 results

Conditions						
T	θ_0	θ_1	$\hat{\theta}_0$ Bias	$\hat{\theta}_0$ RMSD	$\hat{\theta}_1$ Bias	$\hat{\theta}_1$ RMSD
100	1.0	-.25	-.22	.31	.19	.29
100	1.0	-.5	-.32	.31	.34	.31
100	.5	-.25	-.08	.28	.10	.19
100	.5	-.5	-.09	.30	.23	.19
500	1.0	-.25	-.18	.27	.12	.22
500	1.0	-.5	-.19	.27	.12	.30
500	.5	-.25	.13	.20	-.01	.16
500	.5	-.5	.05	.29	.01	.15
1000	1.0	-.25	-.13	.23	.09	.19
1000	1.0	-.5	-.15	.24	.10	.29
1000	.5	-.25	.10	.17	.00	.11
1000	.5	-.5	.04	.25	.01	.11

Table 1.5

Simulation 3 results

Conditions						
T	θ_0	θ_1	$\hat{\theta}_0$ Bias	$\hat{\theta}_0$ RMSD	$\hat{\theta}_1$ Bias	$\hat{\theta}_1$ RMSD
100	.7	-.6	.004	.054	.125	.205
100	.7	-.3	.003	.025	.053	.131
100	.1	.6	.094	.199	-.080	.177
100	.4	.3	-.007	.064	.014	.014
500	.7	-.6	.002	.051	.119	.198
500	.7	-.3	.001	.023	.046	.130
500	.1	.6	.092	.189	-.069	.178
500	.4	.3	.002	.060	.013	.013
1000	.7	-.6	.001	.049	.115	.193
1000	.7	-.3	.001	.024	.045	.133
1000	.1	.6	.090	.186	-.068	.175
1000	.4	.3	-.001	.058	.013	.013

1.5: Discussion

Conventional interindividual ACE modeling in twin research has been a popular methodological technique for examining possible contributions of genetics and environment on psychological phenotypes, but there are several critiques from molecular genetics and person-specific methodology addressed in this dissertation. While the first simulation examines how violations of standard ACE model assumptions can affect their estimation, the second and third simulation test the accuracy of a novel model that attempts to address several of the concerns of the standard ACE model.

Although it has yet to be determined if somatic mosaicism is expressed within the brain, we would expect its effects in decreasing the genetic correlation of twins regardless of zygosity type; this would cause violations of the standard 1.0/0.5 genetic correlation assumptions of standard ACE models. Simulation 1 demonstrates that violations of this assumption can cause several problems. First, both the additive genetic and common environmental sources of variance may be underestimated. This effect appears to be more consistent for common environment factor loadings, as conditions with small simulated genetic factor loadings produced trivial negative bias and even positive bias. Second, the contribution of the unique environment may be overestimated during these situations of somatic mosaicism. It appears as if this variance component may be absorbing some of the effects of underestimating the other two source types. Although it is unknown if the simulated bias corresponds to the actual decrease in genetic correlation between twins, the relationship of increasing estimation bias over increasing genetic correlation change is quite informative. If molecular genetics advances and finds genetic evidence for mosaicism in human brains, the degree of change observed in this genetic evidence can give researchers an idea of the bias in their ACE model estimates.

Simulations 2 and 3 examine the accuracy of the novel model proposed in this study. As the novel feature of these models is a time-varying parameter that follows a linear trajectory, estimation recovery of the intercept and slope of these time-varying parameters were of primary interest. In attempting to recover a time-varying genetic correlation in simulation 2, it appears as if a large amount of observations is required for accurate estimation of this model. Recovering a time-varying factor loading within simulation 3 proved to be much easier, as estimation bias appeared to be small even in short time length conditions. As previously mentioned, since decreasing genetic factor loadings correspond to somatic mosaicism and changing genetic factor loadings may be a proxy for epigenetic changes, it appears as if this model is much better at detecting epigenetic effects than somatic mosaicism effects.

There were several limitations in the simulations of this study. Simulation 1 had few design factors and conditions per factor and provides limited information on the effects of somatic mosaicism on ACE estimates. Also, this simulation only examined a univariate phenotype model and may not generalize to the scenario of a multivariate phenotype model with more complicated loading pattern situations. As previously mentioned, it appeared as if mosaicism caused inflated unique environment components; it is quite possible that a multivariate phenotype model with measurement errors may behave differently and have inflated measurement errors instead of unique environment loadings. The degree of simulated bias in the genetic correlation was equally applied to both MZ and DZ twins, further research could attempt to examine differences in degree of change between twin types. Another limitation that future research could investigate is the role of sample size on these effects; perhaps larger sample sizes may attenuate the bias caused by this assumption violation.

Simulations 2 and 3 had several common limitations that could be addressed in future research. There was no variation in the factor loading pattern (with the exception of the one target time-varying parameter in simulation 3) and it is expected that the pattern of loadings will have a drastic effects on the model. Several candidate patterns were attempted in simulation pilot work before these larger simulations were conducted, and many of these candidate patterns would produce data sets that would cause extreme model fitting problems and non-convergence in estimation. Future work will have to investigate the properties of factor loading matrices that may make them suitable or problematic for this new modeling method. In addition to patterns within a particular loading matrices, future work should investigate the patterns of loadings across twins to see if homogeneity/heterogeneity in loadings across sibling loadings will affect these models. The simulated measurement errors and autoregressive parameters were constant across conditions and may need to be varied in future simulation work investigating this model; this could give a clearer picture of how the model performs in different signal-to-noise ratio conditions and also different degrees of factor stability.

An additional major limitation of the proposed model is that it is limited to only linear change in the time-varying parameter. Although this linear trend is a simple model for a time-varying parameter, it is important to assess the performance of this model in the most basic case as a starting point. Future work could examine more complicated trend structures like higher order polynomial trends. Modeling this time-varying parameter with non-trend models such as a random walk or an autoregressive parameter are also possible avenues for future research.

The long time series lengths needed for estimation accuracy of the time-varying genetic correlation demonstrated in simulation 2 may appear to be a daunting task for many twin researchers. Although these data demands may be easily met within physiological measures

such as fMRI or EEG studies, the process of decreasing genetic correlation between siblings will likely take a relatively longer span (i.e., years as opposed to minutes). Therefore, a practical next step for future research will be to develop a sample pooling method that can incorporate analysis of multiple twins in order to reduce the time series length demands. This will likely cause a major trade-off in the ability to estimate person or twin specific models, but it may be an important step to being able to empirically demonstrate somatic mosaicism from a behavior genetics perspective. Although this would examine group-level mosaicism with many of the potentially problematic assumptions that occur in interindividual analysis, this would still be potentially important discovery.

1.6: References

Boomsma, D.I., & Molenaar, P.C.M. (1986). Using LISREL to analyze genetic and environmental covariance structure. *Behavior Genetics, 16,* 237-250

Bruder, C.E.G. (2008). Phenotypically concordant and discordant monozygotic twins displaydifferent DNA copy-number-variation profiles." *American Journal of Human Genetics, 82*, 1-9

Charney, E. (2012). Behavior genetics and postgenomics. *Behavioral and Brain Sciences, 35,* 331-358.

De, S. (2011). Somatic mosaicism in healthy human tissues. *Trends in Genetics, 27*, 217–223.

Eaves, L.J., Kirk, K.M., Martin, N.G., & Russell, R.J. (1999). Some implications of chaos theory for the genetic analysis of human development and variation. *Twin Research, 2 (1),* 43–48.

Fisher, R.A. (1918). On the correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52,* 399-433.

Galton, F. (1889). *Natural Inheritance.* Macmillan, London.

Jablonka, E., &Lamb, M.J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral and symbolic variation in the history of life.* Cambridge, MA: MIT Press.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, 82,* 35–45.

Kan, K.J., Ploeger, A. , Raijmakers M.E.J., Dolan, C.V., & van der Maas, H.L.J. (2010). Nonlinear Epigenetic Variance: Review and Simulations. *Developmental Science, 13,* 11-27.

Martin, N.G., & Eaves, L.J. (1977). The genetic analysis of covariance structures. *Heredity, 38,* 79-93.

Martin, Sandra L. 2009. "Developmental biology: Jumping-gene roulette." *Nature, 460,* 1087-1088

Molenaar, P.C.M., Boomsma, D., & Dolan, C. (1993). A third source of developmental differences, *Behavior Genetics, 23,* 519-524.

Molenaar, P.C.M. (2010). On the limits of standard quantitative genetic modeling of inter-individual variation: extensions, ergodic conditions and a new genetic factor model of intra-individual variation. In: Hood KE, Halpern CT, Greenberg G, Lerner RM (eds) *Handbook of developmental science, behavior, and genetics.* Blackwell Publishing, Malden, MA

Molneaar, P.C.M., Boomsma, D.I., Smit, D., & Nesselroade, J.R. (2012). Estimation of subject-specific heritabilities form intra-individual variation: iFACE. *Behavioral and Brain Sciences, 15,* 393-400.

Neale, M.C. & Cardon, L.R. (1992) *Methodology for Genetic Studies of Twins and Families.* Dordrecht, NL: Kluwer Academic Publishers

Pearson, K. (1904). On the laws of inheritance of man. II. On the inheritance of the mental and moral characters in man, and its comparison with the inheritance of the physical characters. *Biometrika, 3,* 131-190.

Pfeiffer, H., Lutz-Bonengel, S., Pollak, S., Fimmers, R., Baur, M., & Brinkmann. B. (2004). Mitochondrial DNA control region diversity in hairs and body fluids of monozygotic triplets. *International Journal of Legal Medicine, 118,* 71-4.

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews D. et al. (2006). Global variation in copy number in the human genome. *Nature, 444,* 444-54.

Rodriguez-Santiago, B., Malats, M., Rothman, N., Armengol, L., Garcia-Closas, M., Kogevinas, M, et al. (2010). Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *The American Journal of Human Genetics, 87,* 129-38.

Singer, T., McConnell, M., Marchetto, M., Coufal, N., & Gage, F. (2010). LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences, 33,* 345-354

Taylor, T. H.; Gitlin, S. A.; Patrick, J. L.; Crain, J. L.; Wilson, J. M.; Griffin, D. K. (2014). "The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans". *Human Reproduction Update,* 20 (4): 571–581.

Turkheimer, E. (2000). Three laws of behaviour genetics and what they mean. *Current Directions in Psychological Science, 9 (5),* 160–164

Turkheimer, E., Waldron, M. (2000). Nonshared environment: a theoretical, methodological, and quantitative review, *Psychological Bulletin, 126,* 78-108

Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. *Nature, 150,* 563-565.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proc. Natl.. Acad. Sci. USA, 6,* 320-332.

Paper 2: Developmental change in effective connectivity networks for decision-making processes

2.1: Abstract

Adaptive decision making involves both cognitive control and reward processes, yet few studies have examined how these processes integrate within a connectivity network. Furthermore, while the differential developmental trajectories of these processes have been well studied, there may be additional information gleaned from investigating the development of the connectivity structure for these processes. The current study examines age-related differences in in individual connectivity networks using person-specific exploratory effective connectivity modeling. These modeling techniques were applied to fMRI measures from an incentivized antisaccade task.

2.2: Introduction

Integration of cognitive control and reward processes are essential in decision making (Ernst and Paulus, 2005). Cognitive control includes working memory for maintaining and updating internal representations of choices and goals as well as response inhibition for resisting responses goal-irrelevant information or stimuli. Reward processes enable an individual to evaluate available choices or actions based on preference or need in order to efficiently prioritize behavior. These processes undergo differential maturation trajectories through adolescence that may contribute to emergence of risk taking during this developmental period (Sommerville et al., 2010; Steinberg, 2010); these models describing asynchronous development of these control and reward process have been referred to as 'dual process models'. The current paper examines how the connectivity pattern of regions associated to these processes develops across several age groups.

Dual process models

In the dual process model, adolescents experience heightened reward seeking due to earlier maturation of associated brain regions while simultaneously lacking compensatory cognitive control systems due to underdeveloped brain regions that have yet to fully develop. Brain regions of interest (ROIs) typically associated with reward seeking are the limbic and paralimbic areas of the brain, including the amygdala, ventral striatum, orbitofrontal cortex, medial prefrontal cortex while ROIs typically associated with cognitive control are the lateral prefrontal and parietal cortices as well as the anterior cingulate cortex. The reward seeking system is also referred to as the socioemotional system (Steinberg, 2008) and provides necessary incentivizing and rewarding processes related to dopaminergic activity that are usually

controlled, inhibited, or modulated by the cognitive control system. According to this model, reward seeking ROIs associated with dopaminergic activity have pronounced development following puberty and develop earlier relative to cognitive control system ROIs. The brain areas associated with cognitive control will develop later and at a slower rate due to synaptic pruning and the continued myelination of prefrontal brain regions. Therefore, adolescent risk-taking is due to a reward-control imbalance resulting from early pubertal onset of increased dopaminergic activity within the reward seeking system that precedes the structural maturation of the cognitive control system.

A primary source of evidence for the dual process model comes from functional magnetic resonance imaging (fMRI) studies that look at trajectories of ROI activation associated with reward seeking and cognitive control systems (van Leijenhorst et al., 2006; Steinberg, 2008). These types of designs can generally be considered localization methods that mainly look at differences across ROIs in blood oxygen level dependent (BOLD) activity. Typically, the general linear model (GLM) is implemented in order to provide analysis of variance (ANOVA) or other related mean-comparison models that examine differences in BOLD activity due to ROI, age, stimulus, or some other relevant covariate. In the case of the dual process model, changes in BOLD response across different age groups (i.e., childhood, adolescence, and adulthood) are used to infer differing levels of maturity of ROIs associated with reward seeking and cognitive control. In general, an emphasis is placed on imbalance of activation of reward seeking and cognitive control ROIs rather than the interdependence or connections between these systems. And as previously mentioned, emphasis is placed on contrasts between age groups and thus uses group based modeling procedures rather than individual models which can accommodate potential within-group heterogeneity.

The Role of Connectivity Research

Research on dual process models has focused on an imbalance of reward seeking and cognitive control during adolescence, but typically has paid little attention to the integration of these ROIs during development. It is possible that risk taking during adolescence is also due to underdeveloped connection and integration between the reward and control systems. While localization research focuses on mean differences of ROI activity across contrasts, connectivity research aims to investigate how ROIs may co-vary or exhibit a causal network. Connectivity can be broadly divided into 3 types: (i) structural or anatomical, (ii) functional, and (iii) effective. Structural connectivity pertains to investigating the actual anatomical connections between ROIs, such as diffusion tensor imaging (DTI), which maps the orientation of with matter tracts. Functional connectivity is simply the covariance of ROI BOLD signals and gives inferences on non-directional ROI connections. Effective connectivity provides a causal network of directed paths between ROIs and also operates on ROI BOLD signals. Friston et al. (2014) make a distinction between directed functional connectivity and effective connectivity, where the former involves statistical dependencies that preclude causality while the latter involves causal relations.

There have been several promising findings regarding the connectivity between reward and control systems. These findings from structural, functional, and effective connectivity suggest underdeveloped connections between reward and control systems during adolescence. Asato et al. (1999) in a DTI study found that adolescents had immature white matter connectivity relative to adults. This differential structural connectivity between adolescents and adults hinted that functional and effective connectivity analogues could also exist. A series of studies applying graph theory models to functional connectivity networks from resting state fMRI

provided evidence of development from local to distributed network organization (Fair et al., 2007; Fair et al., 2009). This developmental process is characterized by a greater number of connections between proximal ROIs during childhood that segregate in adulthood while distal but functionally associated ROIs, such as fronto-striatal connections, increase by adulthood. These findings are consistent with generally established neurodevelopmental processes of synaptic pruning, which may correspond to the segregation of early formed proximal connectivity, and myelination, which likewise may correspond to the formation of distal connectivity between associated ROIs. There has also been some evidence for the development of effective connectivity networks between reward and control systems. Hwang et al., (2010) applied vector autoregressive (VAR) modeling to different age groups and found that control to reward ROI paths increased in magnitude across age. These findings from a connectivity perspective suggest that risk taking in adolescence may in part be due to underdeveloped connectivity or "inefficient communication" from control to reward systems rather than just an imbalance of activity in these systems.

A limitation within the current effective connectivity research on adolescent risk taking is the lack of possible contemporaneous directed paths between reward and control ROIs. The VAR model utilized in Hwang et al. (2010) only includes lagged or autoregressive relationships between ROIs. The unified SEM (USEM; Kim et al., 2007) and extended unified SEM (EUSEM; Gates et al., 2011) are effective connectivity models that simultaneously estimate lagged and contemporaneous paths. An example lag-1 USEM can be represented as

$$\eta_t = A\eta_t + B\eta_{t-1} + \zeta_t \qquad (2.1)$$

where η_t is a q-variate vector of ROIs at time *t*, A is a (q,q)-dimensional matrix containing contemporaneous connectivity coefficients, B is a (q,q)-dimensional matrix containing (or autoregressive or lagged connectivity coefficients, and ζ_t is a q-variate Gaussian process noise vector at time *t* with a diagonal covariance matrix, Ψ. Similarly, an example lag-1 EUSEM for a univariate external input series can be represented as

$$\eta_t = A\eta_t + B\eta_{t-1} + \gamma_0 u_t + \gamma_1 u_{t-1} + \tau_{11}\eta_{t-1}u_{t-1} + \zeta_t \qquad (2.2)$$

where u_t is an external input series at time *t*, γ_0 is the contemporaneous direct effect of the external input, $\gamma_1$is the lagged direct effect of the external input, and τ_{11} is a (q,q)-dimensional matrix containing the modulating or bilinear effects. Here, the direct effects give the relationship of the external stimuli to the ROI while the modulating effects describe how the external stimuli affect the connectivity coefficients.

While the USEM is generally useful in block designs, the EUSEM adds the ability to model how external stimuli can modulate various connection parameters in event related designs. Both of these models demonstrated via simulation work the potential bias of strictly lagged (VAR) and strictly contemporaneous (SEM) effective connectivity models. The EUSEM could be potentially useful in decision making research because of the flexibility of this model to handle external stimuli; decision making designs often employ stimuli with varying incentives that could be appropriately handled by the EUSEM.

Figure 2.1: Example EUSEM model diagram

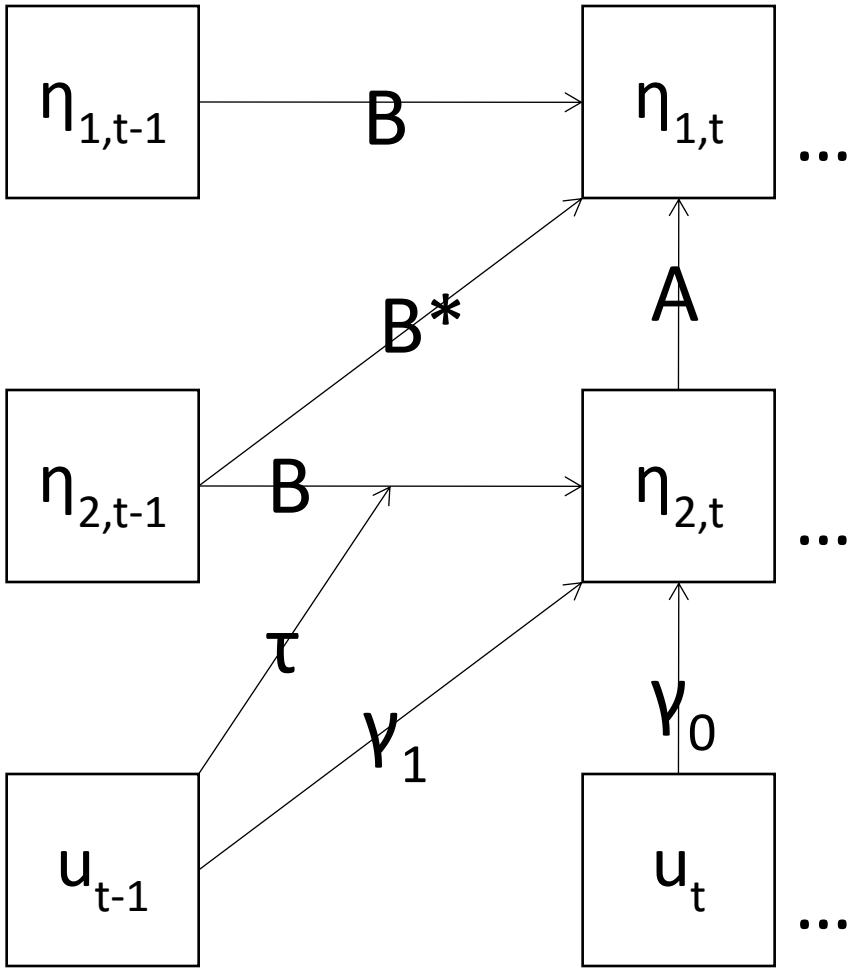

Fig. 2.1. An example diagram of a EUSEM model with 2 ROIs. In this diagram, η_1 and η_2 represent two 2 ROIs and *u* represents a univariate external input or stimulus. The parameters corresponding to the arrows in the diagram are described in equation 2.2. The *B* parameters refer to autoregressive relationships while the B^* parameter represents a cross-lagged parameter. The elipses indicate that this structure would continue on to the subsequent time point, *t+1*.

The evidence for underdeveloped control-reward connectivity as a cause for adolescent risk taking appears promising, but these studies have relied on conventional group modeling procedures. Covariance structure models such as the functional and effective connectivity models applied in the previously mentioned studies have problems with the ecological fallacy (inferring the individual from the aggregate) and fail to capture underlying heterogeneity in connectivity structure (Molenaar, 2004). Many of the connectivity models in fMRI data analysis, such as the VAR model, can be characterized as state space models and can be applied either to a group with homogeneity assumptions or separately to individual fMRI time series. In recent methodological developments, statistical tools have been constructed in order to build individual models while simultaneously examining shared or group based connections (Gates and Molenaar, 2012). This research found in simulation studies that conventional group-based connectivity modeling is incapable of detecting heterogeneity of connectivity patterns within the analyzed group and that individual based modeling methods are essential to detecting heterogeneity in connectivity.

Exploratory modeling techniques

Exploratory connectivity modeling in the framework of the EUSEM could be helpful to individual difference inferences in decision making research. Gates et al. (2010) proposed data-driven exploratory connectivity methods based on Lagrange multiplier (or modification index) tests (Sorbom, 1989). These methods are able to fit models that are qualitatively different across individuals rather than varying on some numeric quantity or magnitude. The Hwang et al. (2010) effective connectivity study on risk taking focused on confirmatory VAR models that test particular paths of interest, e.g., fronto-striatal paths. Even if such a confirmatory model was

allowed to be individually fit, it would only produce differences in magnitude of paths contained within the confirmatory model. Full exploratory models could allow for models where the pattern of paths, not just path magnitude or strength, could vary across individuals. It is possible that qualitatively different connectivity models could correspond to qualitatively different decision making strategies that relate to risk taking.

The Current Paper

The current paper applies exploratory connectivity modeling using the EUSEM to fMRI measures from an incentivized antisaccade task. Person-specific models were fit to data from different age groups and then compared across age to examine potential age-related differences in development of these decision making ROI networks. There were two main hypotheses of the current study. First, connections associated with cognitive control areas will be greater in adults in contrast to adolescents and children. Second, connections associated with reward and loss regions will be greater in adolescents compared to children and adults.

2.3: Method

Participants, divided into age groups, were given an incentivized antisaccade task. The sample consisted of 42 participants with 13 children (9 female, ages 10-12, mean age = 10.9), 14 adolescents (3 female, ages 13-17, mean age = 14.9), and 15 young adults (6 female, all age 18). In the task design, each trial began with the presentation of possible incentive cues displaying possible rewards or losses. A response preparation period would occur next, followed by a visual stimulus that the participants were instructed to avoid and instead look at a mirror location. Imaging data were acquired and processed into time series ROI data for subsequent exploratory connectivity modeling. A brief schematic of the incentivized antisaccade task is given in Figure 2.2. Further details of the incentivized antisaccade task and imaging data acquisition and processing are given in Padmanabhan et al. (2011).

Exploratory connectivity modeling will be conducted using EUSEM model using modification index automatic search procedures (Gates et al., 2010). The exploratory in the current study used empirically evaluated χ^2 differences between candidate models as opposed to estimated χ^2 differences in the method given by Sorbom (1989); this method involves brute-force fitting of all candidate parameters within separate models and calculating actual likelihood differences. Input contrasts were constructed by convolving task epochs with a canonical gamma function to model hemodynamic response functions (Sarty, 2007). Epochs of interest were onset times related to incentive (reward/ loss cue) presentation with separate input series for each event type.

There were 7 ROIs that were used for exploratory modeling. The amygdala (AMY) and nucleus accumbens (NACC) were the primary loss and reward ROIs. Activation of the AMY is associated with avoidance of negative emotional valence, while activation of the NACC is

related to approaching positive emotional valence. The anterior cingulate cortex (ACC) was a region chosen for its function in cognitive control in attention tasks. The frontal eye fields (FEF; both the left and right were used) are ROIs important for visual awareness and saccadic eye movements. The posterior parietal cortex (PPC; both the left and right regions were used) is important for attention and control of eye movement.

Figure 2.2: Incentivized antisaccade task

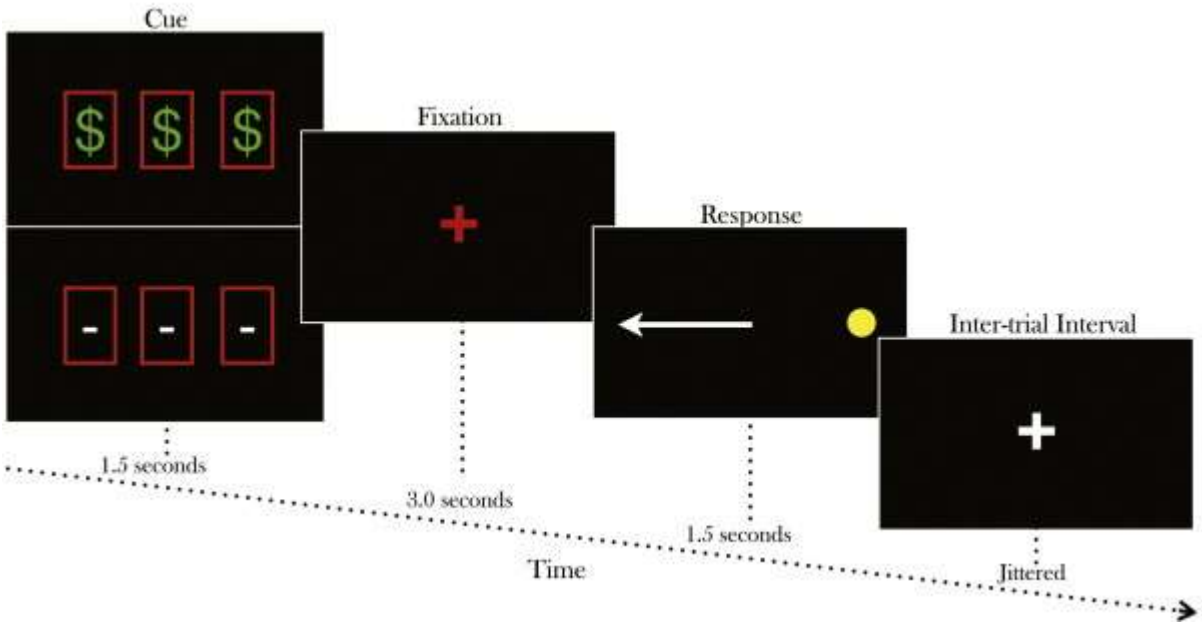

Fig. 2.2. From Padmanabhan et al. (2011). A row of three dollar signs indicated that participant could win money contingent on their performance in the upcoming trial (reward trial). A row of three horizontal dashed lines indicated that no money could be won in the upcoming trial (neutral trial). Incentive cues were presented for 1.5 s. Following that, a red fixation cross appeared for 3 s. A peripheral light appeared for 1.5 s during which, participants were instructed to generate a saccade to its mirror location.

2.4: Results

After fitting individual exploratory connectivity networks to each individual of the sample, there appeared to be several connectivity paths that were chosen by the exploratory procedure at a relatively high proportion while many paths were not chosen for an individual. As expected there were common paths from the reward ROI, NACC, and the loss ROI, to ROIs associated with attention and saccades (i.e., FEF and PPC ROIs). In addition, the connections involving the NACC had common modulating paths associated with reward task cues. Similarly, connections involving the AMY had common modulating paths associated with loss task cues. A particularly troubling result is that the main ROI associated with cognitive control, the ACC, had very few connections with other ROIs in the present sample.

From examining commonly chosen paths, two networks were constructed in order to concisely present the network findings. Connections that were present in at least 10% of the sample are shown in these networks. The loss network, seen in Figure 2.3, shows a network of consisting of the AMY and the four task-related ROIs: FEF.L, FEF.R, PPC.L, and PPC.R (where L = left and R = right). The reward network, seen in Figure 2.4, shows a similar network with the NACC and the same four task-related ROIs. Table 2.1 gives more detailed results for the loss network, including the proportion of each age group having the chosen parameter, and the mean parameter value across those individuals for whom the parameter was chosen. Table 2.2 gives the same information for the reward network, while Table 2.3 gives the results for the commonly chosen connectivity parameters that were not associated with the AMY or NACC. All of these networks focus on the contemporaneous connections between ROIs, as these were the primary common parameters chosen by the exploratory algorithm. Unfortunately, there were a few scattered paths involving the ACC that did not meet the >10% threshold; there were

actually none of these paths that exceeded a stricter >5% threshold, which would be equal to the false positive rate set for the exploratory algorithm. Therefore, the ACC was not included in any of these networks.

There are several notable trends observed in the loss network results. First, the base connectivity parameters were of larger magnitude (range= .14 - .68) compared to the modulating parameters (range= .00 - .03). This could indicate that either (i) the loss cues were not impactful to the decision making networks or (ii) the salience of the loss cues lasted longer than response function modeled by the canonical gamma function. Second, there were only two paths where the adolescents had a greater magnitude relative to the other age groups (i.e., the base paths for b and c). Although these paths were higher for adolescents, the differences in age group on these paths were not statistically significant in follow-up ANOVA analysis. In fact, there were no statistically significant differences in parameter means (via ANOVA) or proportions (via contingency table χ^2 tests); this failure to find statistically significant group differences is likely due to the low power and small sample size of the study.

There were similar findings in the reward network. Again, the base connectivity parameters were of larger magnitude (range= .02 - .32) compared to the modulating parameters (range= .00 - .04). This may be due to similar reasons that were previously mentioned for the loss network. Because of the small sample size and lower power of the study, there were no statistically significant group differences in parameter means or proportions. Although no statistically significant, adolescents were higher on base connections of paths e and h within this reward network.

The non-modulated network connectivity had smaller proportions of connections and smaller parameter means relative to both the loss and reward networks. As seen in both Figure

2.3 and Figure 2.4, these paths involve anatomically similar ROIs (i.e., right to left FEF and left to right PPC). These connections are likely related to the attention and motor processing of the task rather than the decision making processes involving inhibiting a response.

Figure 2.3: Loss network

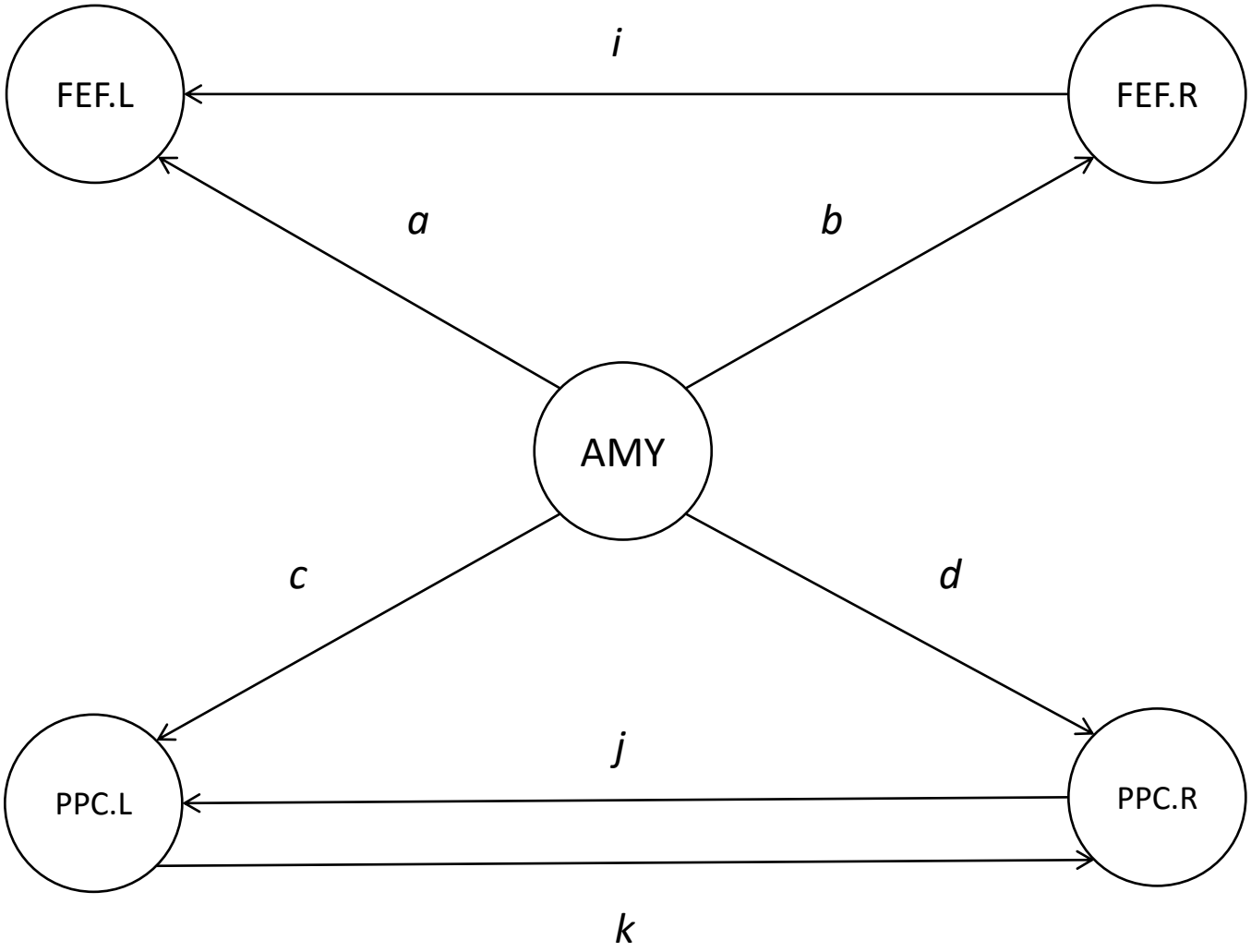

Fig. 2.2. The effective connectivity network associated with loss cues. The arrow labels correspond to the results displayed in Table 2.1.

Figure 2.4: Reward network

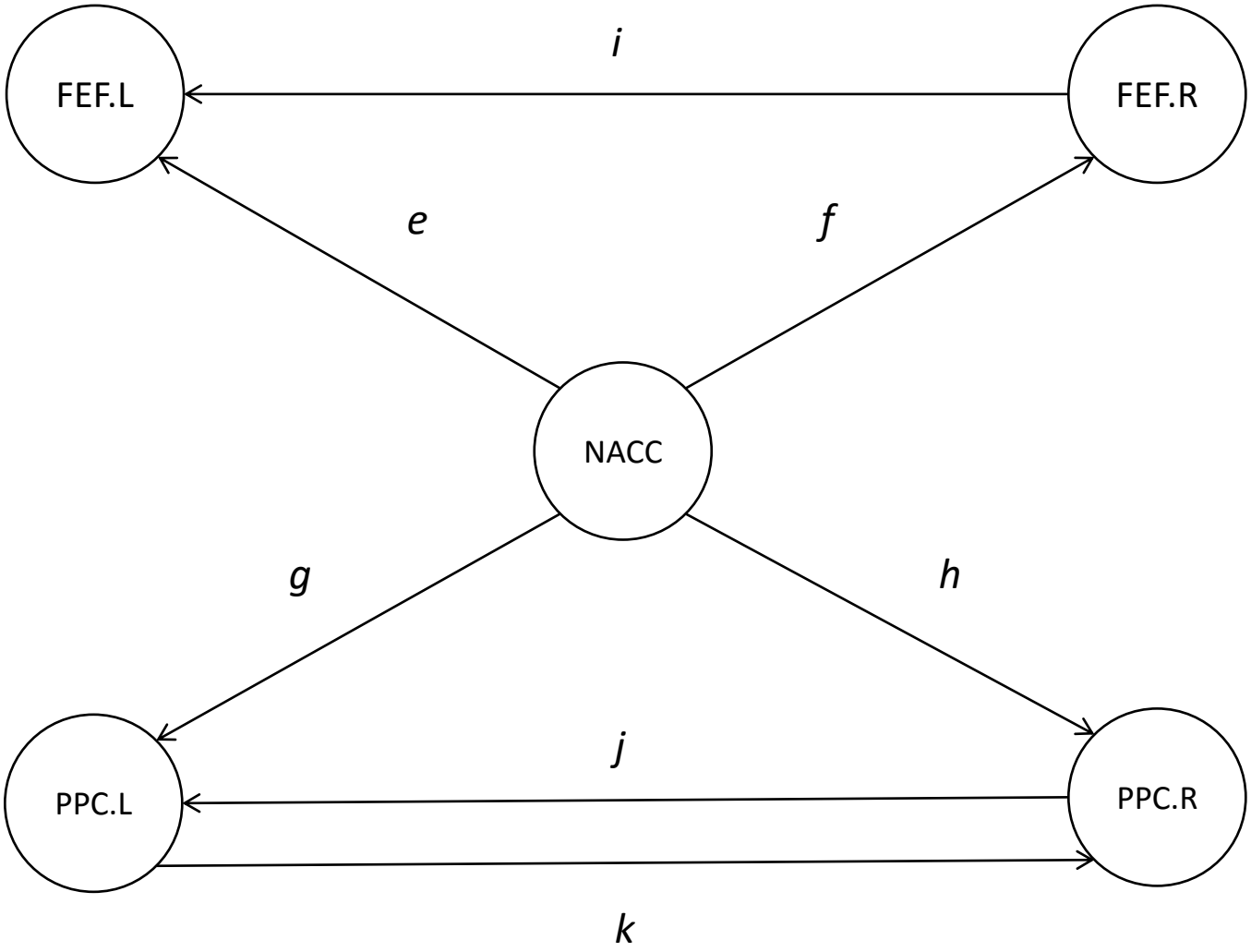

Fig. 2.2. The effective connectivity network associated with reward cues. The arrow labels correspond to the results displayed in Table 2.1.

Table 2.1

Loss network connectivity results

Path	Type	Children		Adolescents		Adults	
		Mean	Proportion	Mean	Proportion	Mean	Proportion
a	Base	.33	.38	.28	.57	.14	.27
	Mod.	.03	.62	.02	.57	.02	.53
b	Base	.31	.62	.68	.21	.22	.36
	Mod.	.02	.46	.02	.57	.03	.57
c	Base	.15	.62	.33	.36	.31	.40
	Mod.	.03	.46	.03	.43	.01	.53
d	Base	.31	.54	.21	.36	.19	.40
	Mod.	.03	.46	.03	.57	.03	.53

Connectivity parameter estimates. The paths correspond to Figure 2.2. The base connection refers to the connection without input while the mod. connection refers to how the connection was modulated by the loss incentive input contrast. The proportion gives the proportion of the group which the connection was chosen within the exploratory method, and the mean gives the mean parameter value of those that proportion with the connection chosen.

Table 2.2

Reward network connectivity results

Path	Type	Children		Adolescents		Adults	
		Mean	Proportion	Mean	Proportion	Mean	Proportion
e	Base	.05	.46	.29	.36	.16	.33
	Mod.	.02	.62	.03	.50	.03	.53
f	Base	.02	.23	.05	.07	.13	.20
	Mod.	.03	.62	.00	.21	.02	.53
g	Base	.32	.08	.08	.14	.20	.47
	Mod.	.01	.54	.02	.50	.02	.53
h	Base	.13	.15	.39	.29	.13	.13
	Mod.	.02	.62	.02	.43	.04	.47

Connectivity parameter estimates. The paths correspond to Figure 2.3. The base connection refers to the connection without input while the mod. connection refers to how the connection was modulated by the reward incentive input contrast. The proportion gives the proportion of the group which the connection was chosen within the exploratory method, and the mean gives the mean parameter value of those that proportion with the connection chosen.

Table 2.3

Non-modulated network connectivity results

	Children		Adolescents		Adults	
Path	Mean	Proportion	Mean	Proportion	Mean	Proportion
i	.15	.38	.07	.14	.04	.07
j	.09	.23	.12	.36	.10	.27
k	.17	.38	.04	.14	.02	.07

Connectivity parameter estimates. The paths correspond to Figure 2.2/2.3. The proportion gives the proportion of the group which the connection was chosen within the exploratory method, and the mean gives the mean parameter value of those that proportion with the connection chosen. No modulating paths were chosen by the exploratory algorithm for these parameters, so only base values are given.

2.5: Discussion

Both of the study hypotheses were unsupported by the results; however, there were several promising findings. For the first hypothesis, there was no evidence of connections associated with cognitive control areas being greater in adults in contrast to adolescents and children. Since there were too few connections chosen by the exploratory algorithm involving the main cognitive control ROI, the ACC, it was not possible to examine age group differences in connections associated with this ROI. For the second hypothesis, there were trending but not statistically significant findings that connections associated with reward and loss regions will be greater in adolescents compared to children and adults. This was likely due to low power and small sample size.

A promising finding of the current study is the emergence of the common loss and reward networks. The exploratory algorithm found common paths that involved the main reward and loss ROIs (i.e., AMY and NACC) influencing ROIs associated with the tasks (i.e., FEF and PPC) and also found modulating paths associated with either loss or reward cues linked to the corresponding connectivity paths. While Hwang et al. (2012) analyzed similar network structures from a confirmatory modeling perspective, the current study demonstrates how these networks can emerge from the exploratory algorithm being driven by the data rather than a-priori confirmatory structures.

There were several limitations of the current study that will need to be addressed in future research expanding on this topic. Most notably, the sample sizes were insufficient to have adequately powered between-group analysis. Although the within-person time series lengths were sufficient to do the individual time series modeling, post-analysis of group differences in these individual model parameters will require more participants. Other possible post-analysis

group modeling will also require larger sample sizes, such as cluster analysis of the individual connectivity network structures to examine qualitatively different groups of connectivity networks.

Another limitation of this study is how the reward and loss cues in the incentivized antisaccade task were modeled. As seen in both the loss and reward network results, the modulating paths associated with these cues were small in magnitude relative to the base connectivity parameters. This is likely due to salience of these loss and reward cues lasting longer than the gamma hemodynamic response linked to the presentation of the incentive stimulus. While other possible hemodynamic response functions could be attempted, a potentially promising future direction would be to implement a block-design rather than the event-related design conducted within the current study. A block design where separate blocks would be designated for either reward, loss, or neutral trials would still allow for examining contrasts between incentive types while avoiding within-trial timing problems associated with particular hemodynamic response functions.

Expanding the range of ROIs in future research could allow for several important possibilities. First, since there were few ACC connections found in the current study, it may be helpful to include other potential ROIs that are associated with cognitive control, such as regions within the prefrontal cortex. It may also be important to include other ROIs that are not related to decision-making or antisaccade tasks, but that are anatomically proximal to these regions. Exploratory modeling of these anatomically proximal regions may also be particularly relevant for theories from Fair et al. (2009) on the transition from local to distributed network organization. According to this theory, one would expect exploratory models for younger age groups to produce connections between proximal ROIs while adults would lack these proximal

connections in exchange for functionally related distal ROIs. In this situation, a data-driven exploratory approach would be ideal since this would allow younger age groups to have connections between functionally unrelated ROIs that would typically not be contained in a confirmatory model.

Although both of the hypotheses of the study were not supported, there are several straightforward steps to alleviate limitations of the current study that led to these inconclusive results. Regardless of the study hypotheses, it was still promising to find loss and reward connectivity networks emerge from an exploratory data driven approach that supports the existing theory and literature on these networks.

2.6 References

Bunge, S.A., Wallis, J.D., Parker, A., Brass, M., Crone, E.A., Hoshi, E., & Sakai, K. (2005). Neural circuitry underlying rule use in humans and nonhuman primates. *The Journal of Neuroscience, 25,* 10347–10350.

Casey BJ, Tottenham N, Liston C, & Durston S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitve Science, 9,* 104–110.

Chung, T., Geier, C., Luna, B., Pajtek, S., Terwilliger, R., Thatcher, D., & Clark, D. (2011). Enhancing respone ingivition by incentive: comparison of adolescents with and without substance use disorder. *Drug and Alchohol Dependence, 115,* 43-50.

Camara, E., Rodriguez-Fornells, A., & Munte, T.F. (2008). Functional connectivity of reward processing in the brain. *Frontiers in Human Neuroscience, 2:19.*

Cohen, M.X., Elger, C.E., & Weber, B. (2007). Amygdala, tractography predicts functional connectivity and learning during feedback-guided decision-making. *Neuroimage, 39,* 1396-1407.

Ernst, M., Nelson, E.E., Jazbec, S., McClure, E.B., Monk,C.S., Leibenluft, E., et al. (2005). Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage, 25,* 1279–1291.

Ernst, M., & Paulus, M. P. (2005). Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry, 58,* 597–604.

Fair, DA, Cohen, AL, Power, JD, Dosenbach, NUF, Church, JA, S, Miezin, FM, Schlaggar, BL, & Petersen, SE. (2009). Functional Brain Networks Develop from a "Local to Distributed" Organization. *PLoS Comput Biol, 5(5),* e1000381. Doi:10.1371/journal.pcbi.1000381

Gates, K. M., Molenaar, P. C. M., Hillary, F., Ram, N., & Rovine, M. (2010). Automatic search in fMRI connectivity mapping: An alternative to Granger causality using formal equivalences between SEM path modeling, VAR, and unified SEM. *NeuroImage, 53,* 1118-1125

Gates, K. M., Molenaar, P. C. M., Hillary, F., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage, 54,* 1151-1158

Gazzaley, A., Rissman, J., & D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective, & Behavioral Neuroscience, 4,* 580-599

Geier, C.F. & Luna, B. (2011). Developmental differences in the effects of incentives on response inhibition. *Child Development, 83,* 1262-1274

Friston, K., Bastos, A., Oswal, A., van Wijk, B., Richter, C., & Litvak, V. (2014). Granger causality revisited. *Neuroimage, 101,* 796-808

Hwang, K., Velanova, K., & Luna, B. (2010). Strengthening of top-down frontal cognitive control networks underlying the development of inhibitory control: A functional magnetic resonance imaging effective connectivity study. *The Journal of Neuroscience, 30,* 15535-15545

Kim, J., Zhu, W., Chang, L., Bentler, P.M., Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Hum. Brain Mapp. 28,* 85–93.

Padmanabhan, A., Geier, C., Ordaz, S., Teslovich, T., & Luna, B. (2011). Developmnetal changes in brain function underlying the influence of reward processing on inhibitory control. *Developmental Cognitivie Neuroscience, 1,* 517-529.

Power, J.D., Fair, D.A., Schlaggar, B.L., & Petersen, S.E. (2010). The development of human functional brain networks. *Neuron, 67(5),* 735-748

Sarty, G.E. (2007). *Computational brain activity maps from fMRI time-series images*. Cambridge University Press, New York

Sommerville, L., Jones, R., & Casey, B.J. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues, *Brain and Cognition, 72,* 124-133.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28,* 78-106

Steinberg, L. (2010). A dual systems model of adolescent risk-taking, *Developmental Psychobiology, 52,* 216-224.

Appendix

Kalman Filter Recursion

The recursion for the DKF consists of the prediction equations:

$$\begin{aligned} \eta_{t|t-1} &= \mathrm{B}\eta_{t-1|t-1} \\ P_{t|t-1} &= \mathrm{B}P_{t-1|t-1}\mathrm{B}^T + \Psi \end{aligned} \tag{A1}$$

where $\eta_{t|t-1}$ is the predicted state estimated at time current time t while $\eta_{t-1|t-1}$ is the updated state estimated at the previous time point $t-1$, and likewise $P_{t|t-1}$ is the predicted state covariance matrix estimated at time current time t while $P_{t-1|t-1}$ is the updated state covariance matrix estimated at the previous time point $t-1$. the . The updating equations for the DKF recursion are:

$$\begin{aligned} \varepsilon_t &= y_t - \Lambda\eta_{t|t-1} \\ S_t &= \Lambda P_{t|t-1}\Lambda^T + \Theta \\ K_t &= P_{t|t-1}\Lambda^T S_t^{-1} \\ \eta_{t|t} &= \eta_{t|t-1} + K_t\varepsilon_t \\ P_{t|t} &= (I - K_t\Lambda)P_{t|t-1}(I - K_t\Lambda)^T + K_t\Theta K_t^T \end{aligned} \tag{A2}$$

where ε_t is the residual at time t, y_t is the observed data at time t, and S_t is the residual covariance matrix at time t, and K_t is the Kalman gain matrix at time t.

Kalman Filter Maximum likelihood estimation

The recursion equations given in (A1-A4) only give state estimates with given model parameters $(\mathrm{B}, \Psi, \Lambda, \Theta)$. The recursion can be embedded in a likelihood function in order to yield maximum likelihood estimates by some numerical optimization scheme. The log-likelihood function for the KF takes the form:

$$LL(\theta) = \frac{1}{2}\sum_{t=1}^{N}[pln(2\pi) - ln|S_t| - \varepsilon_t^T S_t^{-1}\varepsilon_t] \quad \text{(A5)}$$

where p is the number of observed variables and N is the number of time points. Wald tests for the parameters can be obtained by the Hessian matrix evaluated at the ML parameter values. Standard errors for the Wald tests can be obtained from the following transformation of the Hessian:

$$SE(\theta) = \{diag[H(\theta)^{-1}]\}^{1/2} \quad \text{(A5)}$$

where $H(\theta)$ is the Hessian matrix evaluated at the ML parameter values. The Hessian matrix can be obtained numerically alongside most optimization routines; however, this numerical approximation may be unreliable and provide inadequate standard errors.

Block-Toeplitz SEM

The block-Toeplitz SEM method for fitting state space models involves modelling a covariance matrix that incorporates lags. An example lag-1block-Toeplitz matrix can take the form:

Block-toeplitz covariance matrix (Σ)

	x_{t-1}	x_t	u_{t-1}	u_t	$x_{t-1}u_{t-1}$
x_{t-1}					
x_t					
u_{t-1}					
u_t					
$x_{t-1}u_{t-1}$					

where x_t is a set of ROIs at time t, x_{t-1} is the set of ROIs with a lag, u is the input series with the corresponding time index, and the multiplicative terms $x_{t-1}u_{t-1}$ will be used for modeling modulation effects.

The structural relations matrix, commonly referred to as the Beta matrix in LISREL, will take the following form:

Structural relations matrix (β)

	x_{t-1}	x_t	u_{t-1}	u_t	$x_{t-1}u_{t-1}$
x_{t-1}					
x_t	**B**	**A**	$\mathbf{g_1}$	$\mathbf{g_0}$	**G**
u_{t-1}					
u_t					
$x_{t-1}u_{t-1}$					

Here the B matrix gives the autoregressive and cross-lag parameters, A contains the contemporaneous connections (diagonals fixed to zero), g_1 gives the lagged input effects, g_0 gives the contemporaneous input effects, and G gives the modulating effects. The unshaded portions of this matrix are fixed to zero.

The last relevant matrix is the residual covariance matrix, commonly referred to as Psi in LIREL. This matrix will take the following form:

Residual covariance matrix (Ψ)

	x_{t-1}	x_t	u_{t-1}	u_t	$x_{t-1}u_{t-1}$
x_{t-1}					
x_t					
u_{t-1}					
u_t					
$x_{t-1}u_{t-1}$					

Here the shaded regions of the matrix are estimated and the unshaded regions are fixed to zero. The exception to this is the diagonal components of the (2, 2) submatrix consisting of the residual covariances of the (x_t, x_t) components: these diagonal elements are freely estimated.

The block-Toeplitz covariance matrix can be modeled by the form:

$$\Sigma = (I-\beta)^{-1}\Psi\ (I-\beta)^{-T} \qquad \text{(A6)}$$

Curriculum Vitae

Lawrence L. Lo

Email: ollecnerwal180@gmail.com Phone: 763-354-0161

419 Biobehavioral Health Building, Pennsylvania State University, PA 16802

Research Interests

I am interested in time series analysis and its applications to human development. My primary methodological and statistical interests are in Kalman filtering and time series extensions of structural equation models. The motivation for these methodological pursuits is to address new research inquiries for my substantive interests of cognitive neuroscience and behavior genetics.

Education

2014-2016	The Pennsylvania State University
	Ph.D. in Human Development and Family Studies
2009-2014	The Pennsylvania State University
	M.S. in Human Development and Family Studies
	Thesis: *Kalman filter models for ecological momentary assessment designs*
2005-2009	The University of Minnesota
	B.A. in Psychology
	Topic: Item response theory and computerized adaptive testing
	B.A. in Philosophy
	Topic: Philosophy of science with a focus on psychological measurement