UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

STATE SPACE DYNAMIC MIXTURE MODELING:

FINDING PEOPLE WITH SIMILAR PATTERNS OF CHANGE

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MICHAEL DOUGLAS HUNTER
Norman, Oklahoma
2014

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/13885

STATE SPACE DYNAMIC MIXTURE MODELING:
FINDING PEOPLE WITH SIMILAR PATTERNS OF CHANGE

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

Dr. Joseph L. Rodgers, Co-Chair

Dr. Hairong Song, Co-Chair

Dr. Lara Mayeux

Dr. Robert Terry

Dr. Michael Richman

© Copyright by MICHAEL DOUGLAS HUNTER 2014
All rights reserved.

Acknowledgements

I wish to express my gratitude to my research advisor, Professor Joseph L. Rodgers, for teaching me a great deal about how to be a researcher and teacher, for encouraging to explore and work in quantitative psychology, for his guidance and unconditional support during all these years of graduate school, and for being a excellent mentor. I am grateful for having had the opportunity to work with him.

Contents

Acknowledgements **iv**

Table of Contents **vi**

List of Tables **vii**

List of Figures **xi**

Abstract **xii**

1 Introduction **1**
1.1 Motivation 5
1.1.1 The Problem of Pooling 6
1.1.2 Tests for Pooling 10
1.1.3 People as Groups 11
1.2 Modeling: Linear and Nonlinear, Static and Dynamic 12
1.3 Methods Used 19
1.3.1 State Space Models 19
1.3.2 Independent Mixture Distributions 28
1.4 Related and Alternative Methods 30
1.4.1 Cluster Analysis 30
1.4.2 Structural Equation Modeling 31
1.4.3 N-way Factor Analysis 38
1.4.4 Pooled Cross-Sectional Time Series Approaches 40
1.4.5 Multilevel Models 40
1.4.6 Latent Differential Equations 42
1.4.7 Relations and Combinations of Modeling Techniques . . . 43
1.5 Review 47

2 Simulation **49**
2.1 Introduction 49
2.2 Methods 51
2.2.1 Factors to be Varied 51

2.2.2 Data Generation 55
2.2.3 Sources of Eigenvalues 59
2.2.4 Sources of Error 64
2.2.5 Model Fitting 69
2.2.6 Data Analysis 71
2.3 Results 79
2.3.1 Simulation Completion 80
2.3.2 Parameter Estimation 84
2.3.3 Grouping Structure Recovery 95
2.4 Discussion 107

3 Cognitive Development of Youth in the NLSY 110
3.1 Introduction 110
3.2 Methods 120
3.2.1 Participants, Materials, and Procedures 120
3.2.2 Data Analysis 121
3.3 Results 121
3.4 Discussion 128

4 Discussion 130
4.1 Simulation 134
4.2 Genetics of Cognitive Growth 138
4.3 Future Work 141
4.4 Conclusion 143

References 145

List of Tables

1.1 Notations for Matrices and Vectors in State Space Models Without Inputs 22
1.2 Latent Process Model Under Different Notations 23
1.3 Measurement Model Under Different Notations 23

2.1 Marginal Mean Kappa and Accuracy Values Across Simulation Factors 101
2.2 Logistic Regression with Kappa as Outcome 102

3.1 Multisubject State Space ACE: Random Subsample 1 126
3.2 Multisubject State Space ACE: Random Subsample 2 126
3.3 Multisubject Mixture State Space ACE: Random Subsample 1 . . 127
3.4 Multisubject Mixture State Space ACE: Random Subsample 2 . . 128

List of Figures

1.1 Two Linear Maps 17
1.2 Two Nonlinear Maps 18
1.3 A robots path obtained by its odometry, overlaid with the given map. Uncorrected small errors accumulate here. Reproduced from Thrun (2002). 26

2.1 Eigenvalue Distribution for 10,000 Simulated 3-Dimensional Dynamics Matrices 60
2.2 Matrix Element Distribution for 10,000 Simulated 3×3 Dynamics Matrices 61
2.3 Latent values of a simulated 3-dimensional state space model with inputs 66
2.4 Observed values of a simulated 3-dimensional state space model with inputs 68
2.5 Number of Replications Completed for Each Simulation Condition 80
2.6 Median wall clock running time of models in each condition. Time is in log units. Breaks between different numbers of groups in the model and number of people per group are noted in the figure. . . 82
2.7 Proportion of convergence status codes given by the optimizer for each condition. Status Code 1 (Green) = "The final iterate satisfies the optimality conditions to the accuracy requested, but the sequence of iterates has not yet converged. NPSOL was terminated because no further improvement could be made in the merit function". Status Code 4 (Blue) = "The major iteration limit was reached". Status Code 6 (Red) = "The model does not satisfy the first-order optimality conditions to the required accuracy, and no improved point for the merit function could be found during the final linesearch". 83

2.8 Overall comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 85
2.9 For only the dynamics parameters, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 86
2.10 For only the dynamics parameters in the second group, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 87
2.11 For only the factor loadings, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 88
2.12 For only the residual variances, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 88
2.13 Effect of Number of Variables Per Factor on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 90

2.14 Effect of Number of Factors on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 91
2.15 Effect of Number of Occasions on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 92
2.16 Effect of Number of People Per Group on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 93
2.17 Effect of Number of Groups on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple. 94
2.18 Receiver Operating Characteristic Space 96
2.19 Group micro-average accuracy for all numbers of groups including non-mixture models with perfect accuracy. The blue line indicates the mean micro-accuracy. 98
2.20 Group micro-average accuracy for numbers of groups greater than one, that is, excluding non-mixture models with perfect accuracy. The blue line indicates the mean micro-accuracy 99
2.21 Cohen's Kapps for numbers of groups greater than one, that is, excluding non-mixture models with perfect accuracy. The blue line indicates the mean Kappa 100
2.22 Quantile-Quantile Plot of the z values of all the main effects and two-way interactions of the logistic regression with kappa as the outcome and the simulation factors as the predictors. Effects involving the number of factors are colored red. 104
2.23 Quantile-Quantile Plot of the z values of all the main effects, two-way interactions, and three-way interactions of the logistic regression with kappa as the outcome and the simulation factors as the predictors. Effects involving the number of factors are colored red. 105

3.1 A simulated between-person correlation of +0.80, but a within-person correlation of -0.75. 112
3.2 Latent values of a simulated 3-dimensional state space model with different inputs 116
3.3 Univariate ACE model 117
3.4 Multivariate ACE model 118
3.5 State Space Multivariate ACE model 119
3.6 Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). DIGIT=WISC Digit Span. 122
3.7 Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). MATH=PIAT Mathematical Achievement. 123
3.8 Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). RECOG=PIAT Reading Recognition. 123
3.9 Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). COMP=PIAT Reading Comprehension. 124
3.10 Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). PPVT=Peabody Picture Vocabulary Test. 124

Abstract

Increasingly, psychologists encounter data in which several individuals have been measured on multiple variables over numerous occasions. Many of the current methods for this situation combine the data, assuming everyone is a randomly equivalent to everyone else. The extreme alternative on the other side is to separately analyze each person's data, assuming no one is similar to anyone else. This dissertation proposes a method as a compromise between these two extremes. The goal of the method is to find people in the data that are undergoing similar change processes over time. Data were simulated under various conditions to explore what factors influenced the ability of the method to correctly estimate the change process and accurately find people with the same process. It was found that sample size had the greatest positive influence on parameter estimation and the dimension of the change process had the greatest positive impact on correctly grouping people together, likely due to the distinctiveness of their patterns of change. With some success in simulation, the method was applied to an archival data source reflecting cognitive growth in the National Longitudinal Survey of Youth Children data. This analysis suggested that the genetic effects operating between people on their cognitive development may be quite different from their within-person effects, but also revealed a limitation for the method on large sample sizes. Once software improvements are made to the

method, its applicability to large, real data should be reevaluated. State space mixture modeling, in its current form, offers one of the best-performing methods for simultaneously drawing conclusions about individual change processes while also analyzing multiple people.

Chapter 1

Introduction

People are different, and yet the same. There is a sense in which all people are the same; another in which all people are completely unique; and a third, perhaps, most frequent, in which people are partially both. So far, much of the history of psychological effort has emphasized the universal similarity and interchangeability of individuals. Kluckhohn and Murray (1948) echo these sentiments with their oft-cited quotation.

> Every man is in certain respects
>
> a. like all other men,
>
> b. like some other men,
>
> c. like no other man.

Other than the historical use of "man" instead of "human", the idea is quite modern and relevant today. There are globally uniform aspects to being human. These are often the raw, biological facts that distinguish humans from other species. The attributes that are identical over many species, by deductive reasoning, are also the same over individual people: every person must be born, learn to move in the environment, grow, mature, and die. Research that looks

for universal laws that apply to all individuals equally is often called nomothetic (e.g. Allport, 1937).

At the same time, there are many ways in which every person is completely individual, like no one else. No other person has had all the same experiences as any other person. Even twins raised together have unique environmental influences. In fact, identical twins do not have completely identical genomes: they differ in the number of copies of some genes (Bruder et al., 2008). Furthermore, as Molenaar (2004, p. 211) observed "Even if (counterfactually) all genetically stored information would be used to specify the intricate wiring of neurons in a developing brain, this would be far too little.". Research that emphasizes the ways in which people are inextricably singular is frequently called idiographic (e.g. Allport, 1937; Boker, Molenaar, & Nesselroade, 2009).

Nesselroade (2010) describes the idiographic perspective as a a "third discipline" in scientific psychology that extends and perhaps supplants the correlational and experimental disciplines of Cronbach (1957, 1975). The third discipline that Nesselroade portends considers the individual to be the central unit of analysis. By contrast, both the individual differences and experimental practices in psychology are more aligned with the nomothetic tradition and each other. In fact the statistical methods of Cronbach's two disciplines, regression for individual differences and ANOVA for experimental, are mathematically identical. Particularly, the underlying assumptions made in differential and experimental psychology are the same: namely, that individuals are homogeneous, randomly exchangeable units after controlling for a handful of covariates. This is the default assumption made by many nomothetic methods from t-tests to structural equation modeling. Mathematically stated, it says observations are independent and identically distributed (i.i.d.). The individual-level analyses of the idiographic

tradition alter this i.i.d. assumption in various ways to allow for intra-individual conclusions.

Of course, studying individuals has its problems and difficulties. One of the founding principles in science is repetition. It is difficult and risky to make gneral conclusions from the single case. Only when something happens repeatedly or consistently in some lawful way can it become the object of scientific inquiry. Psychology has often addressed the need for repetition by having multiple people in a given study. Considering all people to be equivalent provides the necessary repetition for scientific study. The idiographic tradition finds repetition not across multiple people, but rather through multiple observations on the same person. Thus, time series data, repeated observations, and their corresponding analytic methods are used by the idiographic psychologists.

The shift from between-person analyses to within-person analyses, like time series, creates at least three challenges. First, observations of the same person repeatedly can be prohibitively costly in terms of the time required to collect data. These costs can sometimes be addressed through the use of smart phones, neurophysiological measures, archival data, and automatic data collecting mechanisms. Second, when gathering data on the same person on multiple occasions, observations are likely to be non-independent. Lord and Novick (1968) facetiously suggested brain washing to create independent measurements from the same person, however methods of analyzing non-independent observations are preferable. Later we will review several methods for analyzing data when observations are not independent, for example when the same person is observed at nearby times. Third, when studying a single individual, many researchers accustomed to studying tens, hundreds, or even thousands of individuals will question the generalizability of the single-case researcher. This is a valid concern, but,

as will be discussed later, the concern is equally valid in the opposite direction. The question of when conclusions drawn from many people at a single occasion can be applied to the within-person variability of a particular person is of central importance to the work presented in this dissertation.

The gestalt tradition as exemplified by Lewin (1936) and Allport (1937) suggests that individuals are entirely unique. Hence, it would be fruitless to pursue generalized knowledge gathered from the faulty assumption that people are the same. However, the supposition that all people are entirely unique leads to the conclusion that generalizable knowledge does not exist for people. If, as the gestalt perspective suggests, all people are completely idiosyncratic, then no knowledge can be gained about people in general because there is no such thing as people-in-general. Rather than argue philosophically against this claim, suppose in accordance with Kluckhohn and Murray (1948) that it is only partially true. What then is a researcher to do when people might be different enough that they are not interchangeable but perhaps not so different that general knowledge is impossible? As alluded to by Nesselroade (2010), there is a trade-off between the idiographic and nomothetic traditions in psychology, a dialectic generated by the dual goals in psychology to understand individuals in all of their uniqueness while simultaneously seeking generalizable knowledge about people in general.

This dissertation seeks to to partially bridge the gap between people-in-general and the-completely-unique-individual by developing a data-analytic technique to find subpopulations undergoing similar processes over time.

1.1 Motivation

In clinical and developmental psychology, the notion of individual change is paramount. In clinical psychology, there is the notion that the pattern of change an individual displays during treatment may influence their long term outcome (Steinman, Hunter, & Teachman, 2013). Similarly, in developmental psychology the actual pattern of the acquisition of new skills among infants has recently come under investigation (Adolph, Robinson, Young, & Gill-Alvarez, 2008). More historically Vygotsky (1978) criticized cognitive developmental researchers for only examining stable endpoint behaviors, and Baltes, Reese, and Nesselroade (1977) provided a host of methods of analysis and alternative research design strategies.

One of the hallmarks of the lifespan developmental research methods discussed by Baltes et al. (1977) is multiple measurements of multiple people on multiple variables. As a rule of thumb, if the number of measurement occasions is less than 10 then these data are called *multivariate longitudinal data*; if there are more than 20 measurements per person then they are called *multiple multivariate time series*. We are concerned with both types in this dissertation. In either case, data of this kind are a rich resource for acquiring knowledge. They are so rich, in fact, that many statistical methods do not adequately handle these data. Transformations of these data are often used to reduce its complexity and make it analyzable by some preferred method. The major goal of this dissertation is to establish, test, and apply a method for analyzing multivariate longitudinal data and multiple multivariate time series.

1.1.1 The Problem of Pooling

When analyzing multiple multivariate time series, researchers often take the lead of mathematicians faced with a new problem: they reduce the new problem to an old problem that has already been solved. Hence when confronted with multiple multivariate time series, researchers often take various steps to turn multiple multivariate time series into one multivariate time series and then apply standard time series methods to this new constructed series. This reduction of the number of people from many to one is called "pooling people". Alternatively, researchers could ignore the time series nature of the data and reduce the effective number of time points from many to one: "pooling time points". For the present work, only pooling across people will be considered. I will discuss three methods of constructing a new, single time series from multiple time series: aggregation in the time domain, aggregation in the frequency domain, and concatenation. Subsequently, I will evaluate each method and discuss potential problems with pooling in general. In the next section, I will discuss a method for *selective* pooling of time series based on likelihood ratio test, and propose a new method of *selective* pooling based on clustering and mixture models.

Perhaps the easiest way to reduce multiple time series into a single time series is to simply add them up. Say you had 100 individuals with 200 time points each. Then create an aggregate time series 200 time points long by taking the mean across individuals at each time point. Now analyze the mean time series like any other multivariate time series. This obviously loses a lot of information, but if the different individuals are distinct from one another only in terms of random error then much of the lost information can be recovered by using the variability around each time point as an estimate of the error variance at that time. One

potential problem with time domain aggregation occurs when the individuals are not phase synchronized: a peak for person 1 at time 20 might coincide with a valley in person 2 at time 20 and these could have destructive interference canceling each other out, leading to the mean being an inaccurate description for time series. However, provided that there is not phase reseting within individuals across time, the time series can be phase synchronized by various methods (e.g. Bishop & Thompson, 1986) before pooling. If the length of time series differ, then sample-size-weighted averages at each time point could be used in aggregation.

The next method of pooling is a simple variation on the previous method. Now we will be pooling in the frequency domain. Perform a multivariate (spatial) Fourier transform on each individual time series taking it from the time domain to the frequency domain. Thus we have a power spectrum for each individual. Now we can take the mean power spectrum by taking the mean power across people for each frequency. Once this aggregate power spectrum is obtained, simply perform the inverse Fourier transform and analyze the new time series by standard methods. This method circumvents phase differences in the series by aggregating in a phase independent domain. By doing this, the phase information will be lost, but it is not often of much interest in time series methods.

Finally, the researcher could concatenate the individual time series into one very long series. When applying this pooling method with P-technique factor analysis Cattell (1966b) called this "chain P-technique". This method pastes the time series together end to end. A set of 100 individual time series with 200 time points each could construct a time series of length 100*200 = 20,000 time points. This method does not need to address issues of phase locking or individual time series of different lengths or loss of information due to aggregation. It does have to solve an artifact of the joining. Simple concatenation allows a sequence to pass

from one person to another. So the time series model would make predictions for person 2 and time 1 based on the observations of person 1 at time 200 (or whatever lag model you have). One way to solve this problem is to add missing data between people. If you have a VARMA(p, q) model then putting $k = \max(p, q)$ missing observations between people prevents the model from making predictions across people. Another way to solve this problem is to treat individuals as groups. If you have $N = 100$ groups (people), each with $T = 100$ time points then you can fit a 100 group time series model by constraining the free parameters in each model to be identical. The log likelihood of the multigroup model is just the sum of the log likelihoods of the individual models. This concatenation method of pooling would probably be the best method of pooling if it were absolutely necessary.

The key issues in pooling multiple time series are homogeneity and ergodicity. Homogeneity asks: Does every individual time series have the same underlying dynamics, that is, the same underlying parameters? Ergodicity, on the other hand, asks: Is the structure of within-person variability the same as the between-person variability at a single time? Homogeneity is necessary but not sufficient for ergodicity. Homogeneity is too weak a condition for ergodicity.

Although some analytic tests for ergodicity exist (e.g. Domowitz & El-Gamal, 1993, 1997, 2001), the empirically primary problem is actually homogeneity. This is to say, examples in psychology fall short on ergodicity because before ergodicity can be evaluated, multiple time series are shown to heterogeneous.

Peter Molenaar and colleagues have repeatedly put forth theoretical and empirical evidence that people differ from one another in their longitudinal dynamics and no one's longitudinal dynamics greatly resembles the cross sectional structure (Molenaar, 2004; Hamaker, Dolan, & Molenaar, 2005; Molenaar, Boomsma, &

Dolan, 1993; Molenaar & Campbell, 2009; Molenaar, Huizenga, & Nesselroade, 2003; Molenaar, Sinclair, Rovine, Ram, & Corneal, 2009). This provides a strong caution against pooling time series on the basis of heterogeneity. All pooling assumes at minimum homogeneity; moreover if generalizing across between- and within-person variability, then it also assumes ergodicity. Therefore, when homogeneity and/or ergodicity are not tenable, pooling is not advisable.

One rather fascinating consequence of ergodicity is that it works both ways. In a simple case in statistical mechanics, ergodicity states that the space average is equal to the time average. In psychological terms this becomes: the average across people at one time is equal to the average across time of one person. Typically, Molenaar and colleagues find that psychologists are looking at many people but at only one occasion and then making conclusions about one person at many occasions. This works fine if the process is ergodic. However, because ergodicity works both ways, it would also be fine to repeatedly sample one person and then make conclusions about many people. Psychology would be a very different science if this were common practice. I think many psychologists today would guffaw at a researcher who sampled one person 100 times and then tried to make conclusions about people in general, yet it is common practice to sample 100 people at one time point and then make conclusions about individual-level causes of behavior.

It seems that the problem of pooling time series is directly analogous to pooling data from nested designs (e.g. students within classrooms within schools). We have time points nested within people who are sampled from the population. The time domain and frequency domain aggregation mentioned early are both different kinds of aggregated regression analysis (Cohen, Cohen, West, & Aiken, 2003, Ch. 14), whereas the concatenation method is a disaggregated analysis.

Each method probably bares similar pitfalls to their analog in regression. Aggregated analyses throw away too much information by averaging, and disaggregated analyses claim more independent information than the really have.

Unfortunately, there are relatively few convenient representations of time series models that can be put into hierarchical linear model (HLM, also called mixed effects model, random coefficients model, multilevel model, etc.) form. Latent growth curve models can be put into HLM form and so can some ordinary differential equations (Boker & Ghisletta, 2001), but I do not know of easy ways to estimate multilevel VARMA(2, 4) models with existing software. Hidden Markov models (HMMs) are similar in many respects to VARMA models, but the only mixed effects HMMs that I am aware of take rather extreme amounts of time to run (Altman, 2007) because of the high dimensional numerical integration required.

1.1.2 Tests for Pooling

Nesselroade and Molenaar (1999) developed a method for empirically testing whether two time series can be pooled. The method is refreshingly direct. They fit two models and perform a likelihood ratio test to evaluate the relative fit of the two models. One model pools the data from two time series; the other treats the time series as independent groups. If the models fit about the same, then the simpler pooled model is preferred. Otherwise the unpooled model is chosen.

The work of Nesselroade and Molenaar (1999) thus changes the *assumption* of homogeneity into a *test* for it. The principle drawback is that with a large number of people, if all pairs are compared then the number of tests will be exceedingly large and false-positive (Type I) error rates will certainly be inflated.

For instance, if there were 10 to 100 each with a time series, and all pairwise comparisons are made then there would be 45 to 4,950 tests. By chance, 2 to 247 of these test would be statistically significant at the $\alpha = 0.05$ level.

A related problem with this test for pooling occurs when more than two time series are pooled. If N time series are pooled and the test yields a negative result, then no information is given about *which* time series cannot be pooled. It could be the case the all people could be pooled except for one; or it could be that no person could be pooled with any other; the test yields the same result. However, this is a general issue with multivariate tests (see Raykov & Marcoulides, 2008).

Overall a test for pooling could be an effective strategy, but some method of reducing the number of tests, or controlling the false positive error rate, should be employed. It is possible that theory could provide reason to believe that some people should be pooled and others should not. Genetic relatedness, as an example, could provide a strong argument for pooling individuals on processes that are highly heritable: pooling within a genetic category, but not between categories.

1.1.3 People as Groups

The primary technique here is to consider people as groups. Consequently, statistical methods often applied to groups will be used on individuals. People can be treated as groups when the person-level is composed of multiple observations, for instance multiple occasions of measurement. Instead of calling this situation multigroup modeling, some authors called this multisubject modeling (e.g. Gu, Preacher, Wu, & Yung, 2014). The single most important difference between conceiving of collections of times as groups and collections of people as groups

is the dynamic structure of the data. If occasions of measurement on a single person are independent then there is no intrinsic difference between the two situations. However, sequential observations on a single person often have important dynamics which can be found by any time series method.

1.2 Modeling: Linear and Nonlinear, Static and Dynamic

Modeling, as discussed here, can be thought of as a two-way ANOVA design. The applicable ANOVA is a 2 (static vs. dynamic) $\times$ 2 (linear vs. nonlinear) design. Dynamic models intrinsically involve time, whereas static models do not. Linear models can always be expressed in terms of matrix multiplication (Leon, 2006), whereas nonlinear models cannot. We will first discuss static and dynamic models. Data collected at a single time point (R and Q technique in Cattell's terminology, Cattell, 1952) can only utilize static models. So a question the researcher must ask himself or herself is: "Am I interested in processes and causal mechanisms that occur over time?". If not, then a static model will completely suffice. However, a related question may change the answer. "Am I interested in things that occur within individuals – things that make individuals act, think, and feel – or am I only concerned with what makes one person different from another?". The answers to these two questions must correspond because studying processes that occur over time is the only way to reliably understand things that occur within individuals (Molenaar, 2004; Boker et al., 2009; Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011; Molenaar & Campbell, 2009; Voelkle, Brose, Schmiedek, & Lindenberger, 2014; Nesselroade, 2001; Nesselroade & Molenaar,

2010b).

According to a recent undergraduate introductory psychology text book (Hockenbury & Hockenbury, 2010), the goal of psychology as a science is to understand behavior and mental processes. Behavior unfolds in time and mental processes are fundamentally time oriented, so dynamic models should be advocated by the very goals of psychology. Moreover, we know from the ergodic theorems from statistical mechanics (Birkhoff, 1931) that the structure of cross sectional data is different from that of longitudinal data unless ergodicity holds (see also Molenaar, 2004; Molenaar et al., 2003). So, in order to make conclusions about within-person processes and changes, we must have data and perform analyses within persons. That is, we must use dynamic models if we are interested in individual processes.

Now we are presented with the choice between linear and nonlinear models. There is a rather surprising method of making this decision to model linearly instead of nonlinearly, and John Tukey might be proud of it: look at the data. The distributions of the variables a researcher studies can inform decisions about choosing models. If the variables are distributed normally (i.e. with a Gaussian distribution), then linear models are completely sufficient. No nonlinear model will fit normally distributed data better than a linear model. To justify the previous statement I will need to discuss modeling in its own right.

A model is a compressed, or lower fidelity, description of data. That is, a model attempts to account for data with fewer pieces of information than the data already have and hence some of the information in the data is always lost by the model. Of course, if you allow models with zero degrees of freedom, then there is no loss of information but there is also no data compression. In this way a model is like a metaphor. Models simplify data in ways that can lead to insights

and hypotheses about data. They help us understand complex data by reducing it to e.g. simple correlations and variances. Just as with metaphors, truth is relatively orthogonal to the fidelity of a model (i.e. the model's fit). A poet and a physicist may have very different metatphors describing the Sun. Both of these metaphors will be useful for different purposes, but which one is more true? So it is with models. As George E. P. Box put it: "all models are wrong, but some are useful and some are importantly wrong" (Box & Draper, 1987, p. 424; Box, 1976, p. 792). Models are used to describe data and to yield insights that lead to higher fidelity, better fitting, models. Often, models are mathematical descriptions of data. Hence a model can usually be expressed in functional form: $g(x_1, x_2, \ldots, x_n) \approx f(x_1, x_2, \ldots, x_n)$, where $g(x_1, x_2, \ldots, x_n)$ is taken to be the data, $f(x_1, x_2, \ldots, x_n)$ is taken to be the model, and $\approx$ indicates the approximate as opposed to exact equality. For example, simple linear regression would be $g(x_1, x_2) = x_2$, and $f(x_1, x_2) = b_0 + b_1 x_1$ for some constants b_0 and b_1, so the model states $x_2 \approx b_0 + b_1 x_1$. Models are often made out of various kinds of convenient *ansatz*, but can also be inspired by domain specific knowledge, theory, and of course, previous models.

Now, to the crux of the matter for linear models. If $f(x_1, x_2, \ldots, x_n)$ and $g(x_1, x_2, \ldots, x_n)$ are linear functions, then the model is linear. We then have $f(x_1, x_2, \ldots, x_n) = F\vec{x}$ and $g(x_1, x_2, \ldots, x_n) = G\vec{x}$ for a appropriate matrices F and G. The statement made earlier that "No nonlinear model will fit normally distributed data better than a linear model." is true because only linear relationships exist between Gaussian distributed variables (Bertsekas & Tsitsiklis, 2008). This is why a lack of linear correlation is sufficient for independence in normally distributed data. So if a researcher looks at his or her data, and finds all the variables are normally distributed, then there is no need to try nonlinear models.

A nonlinear model will not fit the data any better than a linear model because no nonlinear relationships exist in the data.

If, however, the data are not all normally distributed, then a nonlinear model may be appropriate. Because we are psychologists, we should probably be using a dynamic model. Hence, when the data are not normally distributed we will be using nonlinear dynamic models as opposed to linear ones.

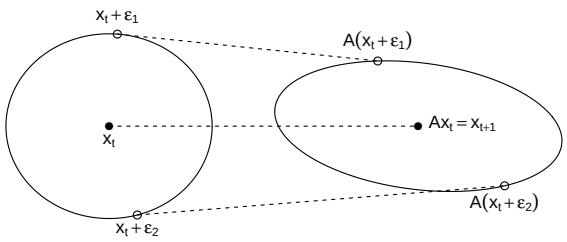
This is a good time to make clear that there are linear dynamic models and there are nonlinear dynamic models. Without making explicit reference, it is common practice in some social science circles to not distinguish between linear dynamics and nonlinear dynamics. This is a grave mistake because various phenomena of interest exist in nonlinear systems but not in linear ones. For instance, chaos, bifurcations, strange attractors, fractal dimension, and complex or emergent behavior do not exist in linear dynamical systems. These are purely nonlinear phenomena. If a research is looking for these, then they will not find them in linear systems.

In understanding linear dynamics, consider Figure 1.1. This Figure illustrates two example linear dynamic systems evolving in discrete time in two-dimensional spaces. The state space is shown at two times: t and $t+1$ with time t on the left. The initial time shows a point in two-dimensional space x_t along with a 2-ball (i.e. circle) around that point, highlighting two other points on that 2-ball. The subsequent time shows how the linear dynamics, A, modify the original 2-ball. Linear dynamics can only stretch, contract, rotate, or leave unchanged the original space. When linear dynamics alter a space, they can only do so along a given direction. For instance in two dimensions, imagine picking any two directions (not necessarily orthogonal) and stretching the space along one direction while contracting the space in another. Both directions could contract,

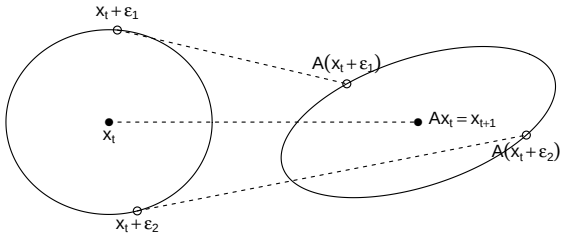
both could expand, one contracts while the other expands, or they could leave space unaltered. When the total (hyper-)volume of the space is increasing over time, the dynamics are called non-stationary. When the total (hyper-) volume of the space is decreasing over time, the dynamics are stationary. That is the effect of linear dynamics on a space.

By contrast for nonlinear dynamics, examine Figure 1.2. The same initial point and surrounding 2-ball are plotted for two examples of nonlinear dynamics. Figure 1.2a shows nonuniform contraction of the state space in both directions. The right and bottom sides are more heavily contracted than the left or top sides. Figure 1.2b actually shows twisting or folding of the state space. Broadly speaking, a combination of stretching, twisting, and folding of the state space is a recipe for chaotic maps (Smale, 1998, 1967; Strogatz, 2000).

The models considered here are exclusively linear. The state space model to be described in the next section is a linear model of discrete-time dynamics. With that being said, slight variations on the same model exist that are nonlinear, namely the state space model with either the extended or unscented Kalman filter instead of the the classical Kalman filter (Grewal & Andrews, 2008, 2001; Hartikainen, Solin, & Särkkä, 2011). Ultimately, nonlinear dynamics may be more realistic and more applicable to natural situations (e.g. Boker, 1996; Boker, Schreiber, Pompe, & Bertenthal, 1998; Wang et al., 2014; Molenaar & Newell, 2003). However, new methods need to be tested incrementally, so the restriction to linear dynamical systems is a permissible simplification, especially considering the relative ease of later switching to nonlinear dynamic models. Having discussed in generality static, dynamic, linear and nonlinear modeling, we now consider in greater detail, the methods used in this work.

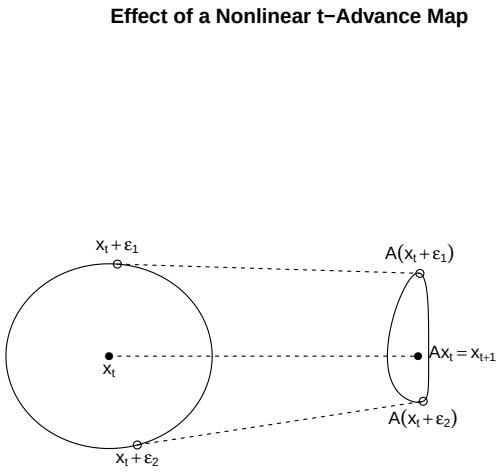


(a) Expansion and Contraction with Slight Rotation

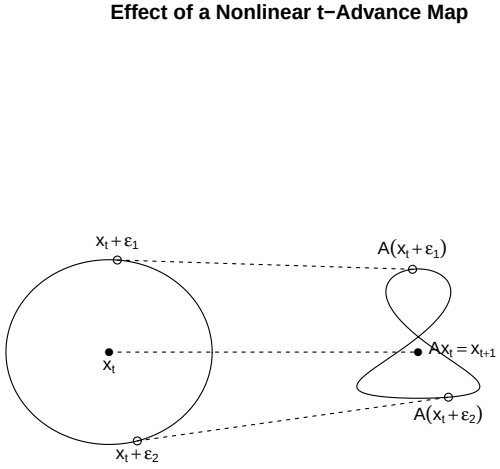


(b) Expansion and Contraction with Moderate Rotation

Figure 1.1: Two Linear Maps



(a) Nonuniform Expansion and Contraction



(b) Expansion and Contraction with Twisting

Figure 1.2: Two Nonlinear Maps

1.3 Methods Used

1.3.1 State Space Models

Simply put, state space models are how scientists and engineers (among others) represent changes over time in variables that are not measured perfectly. State space models (SSMs) in discrete time form recursive relationships where the latent variables at one time are related to the same variables later in time. These are called autoregressive dynamics. The latent variables in turn produce measured variables contemporaneously, like an ordinary factor analysis model.

Using the notation from engineering, a state space model is written

$$\vec{x}_{t+1} = A\vec{x}_t + B\vec{u}_t + \vec{q}_t \tag{1.1}$$

$$\vec{y}_t = C\vec{x}_t + D\vec{u}_t + \vec{r}_t \tag{1.2}$$

where

- $\vec{x}_t$ is a $l \times 1$ vector of the latent states
- $\vec{u}_t$ is a $m \times 1$ vector of observed inputs
- $\vec{q}_t$ is a $l \times 1$ vector of dynamic noise with covariance Q
- $\vec{y}_t$ is a $n \times 1$ vector of observed outputs
- $\vec{r}_t$ is a $n \times 1$ vector of observation noise with covariance R
- A is an $l \times l$ matrix of autoregressive dynamics
- B is an $l \times m$ matrix of covariate/input effects on the state
- C is an $n \times l$ matrix of factor loadings

- D is an $n \times m$ matrix of covariate/input effects on the observation

Equation 1.1 is called the state equation, and Equation 1.2 is called the output equation. The noise vectors $\vec{q}_t$ and $\vec{r}_t$ have zero mean, are assumed to be uncorrelated with each other, uncorrelated with themselves at other times, and uncorrelated with the observations $\vec{y}_t$.

Although the state space model in Equations 1.1 and 1.2 only represents lag-one dynamic relationships, higher-order lags can be incorporated through block matrices. A lag-p process without covariates would have a state equation like the following (Hamilton, 1994, p. 3043)

$$\underbrace{\begin{pmatrix} \vec{\xi}_{t+1} \\ \vec{\xi}_t \\ \vdots \\ \vec{\xi}_{t-p+2} \end{pmatrix}}_{\vec{x}_{t+1}} = \underbrace{\begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \vec{\xi}_t \\ \vec{\xi}_{t-1} \\ \vdots \\ \vec{\xi}_{t-p+1} \end{pmatrix}}_{\vec{x}_t} + \underbrace{\begin{pmatrix} \vec{\zeta}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\vec{q}_t} \tag{1.3}$$

where I is a $l \times l$ identity matrix. The output equation is similarly defined

$$\vec{y}_t = \underbrace{\begin{pmatrix} C_1 & 0 & \dots & 0 & 0 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} \vec{\xi}_t \\ \vec{\xi}_{t-1} \\ \vdots \\ \vec{\xi}_{t-p+1} \end{pmatrix}}_{\vec{x}_t} + \vec{r}_t \tag{1.4}$$

With the block matrix specification, the lag-one state space model is completely general to any arbitrary lag desired.

State space models have been in existence and moderately heavy use since

their inception in the 1960s. Because of this history, a number of different notations have developed. Similarly, because of the relation of state space models to structural equation models, some structural equation modeling notation could also be adapted for state space models.

Table 1.1 shows 5 different notations for state space models. Durbin and Koopman (2001) provided a popular treatment on time series and state space models and have companion software that implements their methods. West and Harrison (1997) gave a more Bayesian perspective on dynamic models. Åström and Murray (2010) wrote an introductory book on feedback control systems engineering. Kalman (1960) and Kalman and Bucy (1961) put forth the original papers on state space models in mechanical and electrical engineering. Kalman's work, in general, is rather mathematical, and in particular was posed as an extension of the work of Norbert Wiener and Andrey Kolmogorov.

Of particular note in Table 1.1 is the fact that several of the same symbols appear as different conceptual units in different notations. For example, H_t is the factor loadings matrix (measurement model) in Kalman's notation, but it is the observation noise covariance (manifest error covariance) in the Durbin & Koopman model. Similarly, F_t appears in every notation except LISREL, and in each it represents a different part of the model.

Other formulations emphasize different portions of the models. Table 1.2 shows much of the same information as Table 1.1, but emphasizes the latent process model which is sometimes called the state equation. Table 1.3 compares the measurement models with varying notations.

Notice that the notations for state space models are numerous and inconsistent. The general form of the model is always the same, but the letters used vary. For this reason, the notation that seemed easiest to remember was picked. It is

Table 1.1: Notations for Matrices and Vectors in State Space Models Without Inputs

	Durbin & Koopman	LISREL Endogenous	West & Harrison `dlm` R Package	Åstrom & Murray	Kalman
Observed Outputs	$\vec{y}_t$	$\vec{y}_t$	$\vec{a}_t$	$\vec{y}_t$	$\vec{z}_t$
Latent States	$\vec{\alpha}_t$	$\vec{\eta}_t$	$\vec{\theta}_t$	$\vec{x}_t$	$\vec{x}_t$
Observation Noise	$\vec{\varepsilon}_t$	$\vec{\varepsilon}_t$	$\vec{v}_t$	$\vec{w}_t$	$\vec{v}_t$
Dynamic Noise	$R_t\vec{\eta}_t$	ζ_t	$\vec{w}_t$	$\vec{v}_t$	$\vec{w}_t$
Observation Noise Covariance	H_t	Θ_t	V_t	R_{wt}	R_t
Dynamic Noise Covariance	$R_tQ_tR_t^\mathsf{T}$	Ψ_t	W_t	$F_tR_{vt}F_t^\mathsf{T}$	Q_t
Factor Loadings	Z_t	Λ_t	F_t	C_t	H_t
Auto-regressive Dynamics	T_t	B_t	G_t	A_t	F_t, Φ_t
Prediction Error Estimate	$\vec{v}_t$	$\vec{y}_t - \Lambda_t\widehat{\vec{\eta}}_t$	$\vec{a}_t - F_t\widehat{\vec{\theta}}_t$	$\vec{y}_t - C_t\widehat{\vec{x}}_t$	$\widetilde{\vec{z}}_t$
Prediction Error Covariance	F_t	G_t	C_t	-	S_t

Table 1.2: Latent Process Model Under Different Notations

Durbin & Koopman	$\vec{\alpha}_{t+1}$	$=$	$T_t\vec{\alpha}_t + R_t\vec{\eta}_t$
LISREL	$\vec{\eta}_{t+1}$	$=$	$B_t\vec{\eta}_t + \Gamma_t\vec{\xi}_t + \vec{\zeta}_t$
West & Harrison	$\vec{\theta}_{t+1}$	$=$	$G_t\vec{\theta}_t + \vec{w}_t$
Åstrom & Murray	$\vec{x}_{t+1}$	$=$	$A_t\vec{x}_t + B_t\vec{u}_t + \vec{v}_t$
Kalman	$\vec{x}_{t+1}$	$=$	$F_t\vec{x}_t + B_t\vec{u}_t + \vec{w}_t$

Table 1.3: Measurement Model Under Different Notations

Durbin & Koopman	$\vec{y}_t$	$=$	$Z_t\vec{\alpha}_t + \vec{\varepsilon}_t$
LISREL	$\vec{y}_t$	$=$	$\Lambda_{yt}\vec{\eta}_t + \vec{\varepsilon}_t$
LISREL	$\vec{x}_t$	$=$	$\Lambda_{xt}\vec{\xi}_t + \vec{\delta}_t$
West & Harrison	$\vec{a}_t$	$=$	$F_t\vec{\theta}_t + \vec{v}_t$
Åstrom & Murray	$\vec{y}_t$	$=$	$C_t\vec{x}_t + D_t\vec{u}_t + \vec{w}_t$
Kalman	$\vec{z}_t$	$=$	$H_t\vec{x}_t + \vec{v}_t$

generally alphabetic: A, B, C, D, Q, R. In addition to these matrices, the model also requires the initial latent mean and variance to be specified ($\vec{x}_0$ and P_0), and the input/covariate variables ($\vec{u}_t$). The notational situation of state space models is in contrast to structural equation modeling in which different forms of the model are used that turn out to be equivalent. The unifying aspect of these state space models is their treatment of time.

The main benefit of state space modeling is the intrinsic special treatment of time. Modeling dynamics requires some form of treating time differently. State space models do that, and they can use latent variables. Other forms of modeling dynamics with latent variables (e.g. latent differential equations, dynamic factor analysis, P-technique factor analysis, N-way factor analysis, latent difference scores, latent growth curves, etc.) all exploit pre-existing techniques in various ways to force them to incorporate dynamic or temporal information. The exploitation works, but often has severe limitations. By contrast, state space models were designed from the ground up to incorporate latent variables and

non-stationary dynamics. With state space modeling there is no need to create a Chimera to allow for dynamic modeling or measurement error, these are native to the technique.

State space modeling, as implemented here, uses a type of recursive filter called a Kalman filter. The Kalman filter consists of alternating prediction and correction steps. The prediction step begins from some latent state vector at time t called $\vec{x}_{t|t}$ along with its error variance matrix called $P_{t|t}$, and creates a prediction or forecast for the state vector and covariance at the next time $t+1$. Thus, it uses the current estimates $\vec{x}_{t|t}$ and $P_{t|t}$ to create forecasts for the next estimates $\vec{x}_{t+1|t}$ and $P_{t+1|t}$.

As a set of equations, the Predict Step is

$$\vec{x}_{t+1|t} = A\vec{x}_{t|t} + B\vec{u}_t \tag{1.5}$$

$$P_{t+1|t} = AP_{t|t}A^{\mathsf{T}} + Q \tag{1.6}$$

The Predict Step consists merely of finding the model-predicted mean and variance matrix for the latent variables at the next time point conditional on their current estimates. Thought of in a probability sense, the Predict Step maps the current multivariate Gaussian probability distribution to a model-implied prediction of the next probability distribution. The Prediction Step depends only on the latent variable estimates, the model matrices (A, B, Q), and the measured covariates ($\vec{u}_t$).

On the other hand, the Correct Step, or measurement update, depends more on the measured data and its corresponding measurement model. The Update Step uses the observed data to correct the forecast from the Predict Step. Thus, it uses the forecast estimates $\vec{x}_{t+1|t}$ and $P_{t+1|t}$ along with the data at time $t+1$

to create updated estimates $\vec{x}_{t+1|t+1}$ and $P_{t+1|t+1}$.

As a set of equations, the Update Step is

$$\begin{aligned}
\vec{\widehat{y}}_{t+1} &= \widehat{Mean}(\vec{y}_{t+1}) = C\vec{x}_{t+1|t} + D\vec{u}_{t+1} && (1.7)\\
\vec{\widetilde{y}}_{t+1} &= \widehat{Residual}(\vec{y}_{t+1}) = \vec{y}_{t+1} - \vec{\widehat{y}}_{t+1} && (1.8)\\
\widehat{S}_{t+1} &= \widehat{Cov}(\vec{y}_{t+1}) = CP_{t+1|t}C^{\mathsf{T}} + R && (1.9)\\
K &= P_{t+1|t}C^{\mathsf{T}}\widehat{S}_{t+1}^{-1} && (1.10)\\
\vec{x}_{t+1|t+1} &= \vec{x}_{t+1|t} + K\ \vec{\widetilde{y}}_{t+1} && (1.11)\\
P_{t+1|t+1} &= P_{t+1|t} - KCP_{t+1|t} && (1.12)
\end{aligned}$$

The Update Step consists of adjusting the estimates from the Predict Step based on how well they correspond to the measured data, $\vec{y}_{t+1}$. The matrix K, called the Kalman gain, is set so that under certain regularity conditions the updated latent error variance matrix $P_{t+1|t+1}$ is as small as possible; that is, has minimum trace. This achieves the goal of having as precise (i.e. smallest variance) an estimate of the latent variables as possible with the given information. The Update Step is critical for preventing prediction errors from accumulating, especially for non-stationary time series. Thrun et al. (2002) shows the importance of the measurement update in preventing errors from accumulating in robotic mapping. One of his figures is reproduced as Figure 1.3 which shows the path of a robot navigating a maze obtained from its internal state variables without measurement updates.

The previous discussion of the Predict and Update Steps in the Kalman filter have illustrated the time-evolving distribution of the latent variables. It may not then be surprising that the state space model with a Kalman filter reduces to a

Figure 1.3: A robots path obtained by its odometry, overlaid with the given map. Uncorrected small errors accumulate here. Reproduced from Thrun (2002).



model of each data row as a multivariate Normal distribution.

$$\vec{y}_{t+1} \sim \mathcal{N}\left(\mu = C\vec{x}_{t+1|t} + D\vec{u}_{t+1},\ \Sigma = CP_{t+1|t}C^{\mathsf{T}} + R\right) \tag{1.13}$$

Hence the minus two log likelihood of row t given the model is

$$\begin{aligned} l_t &= \frac{-1}{2}\left(N_y \log(2\pi) + \log\left(\det\left(\widehat{S}_t\right)\right) + \vec{\tilde{y}}_t^{\mathsf{T}}\widehat{S}_t^{-1}\vec{\tilde{y}}_t\right) && (1.14)\\ -2\mathcal{L}_T &= -2\sum_{t=1}^{T} l_t && (1.15)\\ -2\mathcal{L}_T &= \sum_{t=1}^{T}\left(N_y \log(2\pi) + \log\left(\det\left(\widehat{S}_t\right)\right) + \vec{\tilde{y}}_t^{\mathsf{T}}\widehat{S}_t^{-1}\vec{\tilde{y}}_t\right) && (1.16)\\ -2\mathcal{L}_T &= TN_y \log(2\pi) + \sum_{t=1}^{T}\left(\log\left(\det\left(\widehat{S}_t\right)\right) + \vec{\tilde{y}}_t^{\mathsf{T}}\widehat{S}_t^{-1}\vec{\tilde{y}}_t\right) && (1.17) \end{aligned}$$

The likelihood of a state space model given above is called the prediction error decomposition.

To implement the state space model, OpenMx (Boker et al., 2011; Boker, Neale, et al., 2014) was used. OpenMx had structures to compute the above log likelihood in the special case where $S_t = S$ is the same for every row of data. Alternatively, OpenMx also allowed for a different covariance matrix for different rows when definition variables are used. However, the state space model log likelihood corresponds to a case where *every* variable is both a definition variable and a modeled variable. Moreover, S_t^{-1} must already be calculated for use within the state space model to generate the Kalman gain. Therefore, C/C++ code was written by the author to allow OpenMx to estimate state space models using a Kalman Filter.

1.3.2 Independent Mixture Distributions

Suppose there is a data set in which some observations are produced by one process and others are produced by another process. If the rows of data are independent, then a model like this is called an independent mixture distribution, alternatively a finite mixture model or mixture distribution (see McLachlan & Peel, 2000, for a more complete introduction). There might be c distinct processes, or models, producing a set of data. The data can be assigned a probability based on the model parameters, conditional on the model being true. The probability of the ith row of data given the jth model is written

$$\mathbb{P}(\vec{x}_i|C=j) \tag{1.18}$$

All such combinations of i and j can be arranged into a matrix and then multiplied by the probability of each model to produce the probability distribution for each row of data.

$$\begin{pmatrix} \mathbb{P}(\vec{x}_1) \\ \mathbb{P}(\vec{x}_2) \\ \mathbb{P}(\vec{x}_3) \\ \mathbb{P}(\vec{x}_4) \\ \vdots \\ \mathbb{P}(\vec{x}_N) \end{pmatrix} = \begin{pmatrix} \mathbb{P}(\vec{x}_1|C=1) & \mathbb{P}(\vec{x}_1|C=2) & \dots & \mathbb{P}(\vec{x}_1|C=c) \\ \mathbb{P}(\vec{x}_2|C=1) & \mathbb{P}(\vec{x}_2|C=2) & \dots & \mathbb{P}(\vec{x}_2|C=c) \\ \mathbb{P}(\vec{x}_3|C=1) & \mathbb{P}(\vec{x}_3|C=2) & \dots & \mathbb{P}(\vec{x}_3|C=c) \\ \mathbb{P}(\vec{x}_4|C=1) & \mathbb{P}(\vec{x}_4|C=2) & \dots & \mathbb{P}(\vec{x}_4|C=c) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}(\vec{x}_N|C=1) & \mathbb{P}(\vec{x}_N|C=2) & \dots & \mathbb{P}(\vec{x}_N|C=c) \end{pmatrix} \begin{pmatrix} \mathbb{P}(C=1) \\ \mathbb{P}(C=2) \\ \vdots \\ \mathbb{P}(C=c) \end{pmatrix} \tag{1.19}$$

It should be noted that even though $\mathbb{P}(\vec{x}_i)$ is written, it is frequently the probability *density* and not the actual probability. Also, there is an implicit conditioning

on the free parameters being estimated: $\mathbb{P}(\vec{x}_i)$ is intended to signify $\mathbb{P}(\vec{x}_i|\vec{\gamma})$ for the vector of all estimated parameters $\vec{\gamma}$.

In summation notation the ith row becomes

$$\sum_{j=1}^{c} \mathbb{P}(\vec{x}_i|C=j)\mathbb{P}(C=j) \tag{1.20}$$

where readers familiar with probability theory will recognize $\mathbb{P}(\vec{x}_i|C=j)\mathbb{P}(C=j)$ as the joint probability of $\vec{x}_i$ and $C=j$, and also that the summation over all possible values of C marginalizes this joint probability to create the following identity.

$$\mathbb{P}(\vec{x}_i) = \sum_{j=1}^{c} \mathbb{P}(\vec{x}_i|C=j)\mathbb{P}(C=j) \tag{1.21}$$

This is to say that a mixture distribution model describes the probability density function of data as a weighted combination of several density functions. The density functions each correspond to a discrete class and the combination weights are the marginal probabilities for each of the classes.

Finally, individual rows of data are assumed to be independently and identically distributed (i.i.d.) so the joint probability of all rows of data is simply the product of their individual probabilities.

$$\mathbb{P}(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = \mathbb{P}(\vec{x}_1)\mathbb{P}(\vec{x}_2)\ldots\mathbb{P}(\vec{x}_N) = \prod_{i=1}^{N} \mathbb{P}(\vec{x}_i) \tag{1.22}$$

However, because the product of a large number of values between 0 and 1 quickly approaches 0 and because computers have finite precision, the logarithm of the joint probability is often taken. The logarithm of a product is the sum of the logarithms of the elements of that product, so the following statement of the joint

probability of the data is possible.

$$\log\left(\mathbb{P}(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)\right) = \log\left(\mathbb{P}(\vec{x}_1)\mathbb{P}(\vec{x}_2)\ldots\mathbb{P}(\vec{x}_N)\right) \tag{1.23}$$

$$\log\left(\mathbb{P}(\vec{x}_1)\right) + \log\left(\mathbb{P}(\vec{x}_2)\right) + \ldots + \log\left(\mathbb{P}(\vec{x}_N)\right) = \sum_{i=1}^{N} \log\left(\mathbb{P}(\vec{x}_i)\right) \tag{1.24}$$

Returning to matrix notation yields

$$\log\left(\mathbb{P}(\vec{x}_1, \ldots, \vec{x}_N)\right) = \vec{1}^{\mathsf{T}} \log\left(\begin{pmatrix} \mathbb{P}(\vec{x}_1|C=1) & \ldots & \mathbb{P}(\vec{x}_1|C=c) \\ \mathbb{P}(\vec{x}_2|C=1) & \ldots & \mathbb{P}(\vec{x}_2|C=c) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(\vec{x}_N|C=1) & \ldots & \mathbb{P}(\vec{x}_N|C=c) \end{pmatrix} \begin{pmatrix} \mathbb{P}(C=1) \\ \vdots \\ \mathbb{P}(C=c) \end{pmatrix}\right) \tag{1.25}$$

The incredible versatility of mixture models is that literally any model that produces a probability can be used for the component classes. A mixture of several normal distributions is just as possible as a mixture of a binomial, a Poisson, a logistic, and a Wishart distribution. Moreover each distribution in the mixture can be generated by its own kind of model. Thus a mixture of three normal distributions can be composed of a structural equation model, a state space model, and a mutlilevel model.

1.4 Related and Alternative Methods

1.4.1 Cluster Analysis

Here, only k-means cluster analysis will be considered. The goal of k-means cluster analysis is to partition the data into k clusters, each of which is as homogeneous as possible (Everitt, Landau, Leese, & Stahl, 2011). Because of the large

number of possible solutions and the discrete solution space, cluster analysis can be a very time-intensive technique. The Stirling number of the second kind gives the number of partitions of N items into k groups, and hence the number of possible k-means solutions for N data rows. These numbers grow factorially and hence make brute force methods (i.e. try every possible solution) for cluster analysis computationally impossible. For instance, with 100 items placed into 3 groups, there are 8.59×10^{46} possible arrangements. If 1 trillion arrangements were tried every second, then trying all possible solutions would still take 2.72×10^{27} years. This is about one third of the way to the proposed heat death of the universe, and 38 quadrillion times longer than the Earth's sun is supposed to last before it burns out and engulfs nearby planets.

1.4.2 Structural Equation Modeling

Structural equations extend the regression/ANOVA modeling framework (e.g. Cohen, 1968; Cohen et al., 2003) to include latent variables. These latent, hidden, or unobserved variables are detected by their impact on the observed variables. Basic structural equation models (SEMs) assume the observed data and the latent variables are distributed as a multivariate Gaussian. Therefore, the mean and covariance completely specifies the distribution of both the observed and unobserved variables. Similarly, because the distribution is normal, only linear relationships exist among the variables. Consequently, linear models are completely sufficient to represent any relationship among the Gaussian variables. Linear models are isomorphic to linear functions which are in turn isomorphic to matrices, therefore SEMs involve relationships between matrices. Hence, basic SEMs contain parameters in various matrices that can be combined in particu-

lar ways to form the model-implied means vector and covariance matrix of the observed data. The model-implied moment matrices often depend on the parameters of the latent variables. Thus, the observed variables behave as if the latent variables existed to the extent that the model fits the data.

Over the years, several related sets of matrices have been used to specify SEMs. We will briefly describe the notations used for LISREL, Mplus/LISCOMP, and RAM because these will be relevant for the exposition of state space modeling later.

LISREL

The LISREL notation (LIneal Structural RELations; Jöreskog & Van Thillo, 1972; Jöreskog, 1973) describes SEMs in terms of endogenous and exogenous latent variables, each of which are measured by a set of observed variables. Endogenous variables are generated internally from the model; they are outcome/dependent variables, or y's. Exogenous variables are generated externally from the model; they are predictor/independent variables, or x's. The distinction between the "measurement" model and the "structural" model in SEMing originates in LISREL. The structural model in LISREL consists of an intercept ($\vec{\alpha}$), the regression effects of the endogenous variables on each other (B), the regression effects of the exogenous variables on the endogenous variables (Γ), and an additive residual ($\vec{\zeta_i}$) with zero mean and some covariance structure (Ψ).

$$\vec{\eta_i} = \vec{\alpha} + B\vec{\eta_i} + \Gamma\vec{\xi_i} + \vec{\zeta_i} \qquad \vec{\zeta_i} \sim \mathcal{N}\left(\vec{0}, \Psi\right) \qquad \vec{\xi_i} \sim \mathcal{N}\left(\vec{\kappa}, \Phi\right) \tag{1.26}$$

The measurement model in LISREL is entirely composed of two factor models. It consists of intercepts ($\vec{\tau}_x$ for exogenous variables and $\vec{\tau}_y$ for endogenous

variables), regressions of the observed variables on the latent variables (Λ_x for exogenous variables and Λ_y for endogenous variables), and additive residuals ($\vec{\delta}_i$ for exogenous variables and $\vec{\varepsilon}_i$ for endogenous variables) each with zero mean and some structured covariance matrix (Θ_δ, and Θ_ε, respectively).

$$\vec{x}_i = \vec{\tau}_x + \Lambda_x \vec{\xi}_i + \vec{\delta}_i \qquad \vec{\delta}_i \sim \mathcal{N}\left(\vec{0}, \Theta_\delta\right) \tag{1.27}$$

$$\vec{y}_i = \vec{\tau}_y + \Lambda_y \vec{\eta}_i + \vec{\varepsilon}_i \qquad \vec{\varepsilon}_i \sim \mathcal{N}\left(\vec{0}, \Theta_\epsilon\right) \tag{1.28}$$

Together, the structural and measurement models in LISREL imply the following mean and covariance matrices.

$$\mathscr{E}\left\{\begin{pmatrix} \vec{y} \\ \vec{x} \end{pmatrix}\right\} = mean\begin{pmatrix} \vec{y} \\ \vec{x} \end{pmatrix} = \begin{pmatrix} \tau_y + \Lambda_y(I-B)^{-1}(\alpha + \Gamma\kappa) \\ \tau_x + \Lambda_x\kappa \end{pmatrix} \tag{1.29}$$

$$\mathscr{E}\left\{\begin{pmatrix} \vec{y} \\ \vec{x} \end{pmatrix}\begin{pmatrix} \vec{y} & \vec{x} \end{pmatrix}\right\} = cov\begin{pmatrix} \vec{y} \\ \vec{x} \end{pmatrix} \tag{1.30}$$

$$cov\begin{pmatrix} \vec{y} \\ \vec{x} \end{pmatrix} = \begin{pmatrix} A(\Gamma\Phi\Gamma^\mathsf{T} + \Psi)A^\mathsf{T} + \Theta_\epsilon & A\Gamma\Phi\Lambda_x^\mathsf{T} \\ \Lambda_x\Phi\Gamma^\mathsf{T}A^\mathsf{T} & \Lambda_x\Phi\Lambda_x^\mathsf{T} + \Theta_\delta \end{pmatrix} \tag{1.31}$$

where $A = \Lambda_y(I-B)^{-1}$.

It should be noted that LISREL and other SEM notations are models of normally distributed data. SEMs imply the first and second moments of the data, the multivariate mean and variance, which completely specifies a Normal distribution. In a Normal distribution, only linear relationships exist between variables so the linear regression models are entirely sufficient conditional on

the multivariate Normality assumption. Moreover, the observations are assumed to be independently and identically distributed conditional on the model. So the expected mean and covariance no longer depend on the subscript i defining different rows of data. There is a single mean and covariance that describes all of the data rows at once. The highly related Mplus/LISCOMP notation for SEMs is discussed next.

Mplus/LISCOMP

The LISCOMP notation (LInear Structural COMPonents; Muthén, 1984; Muthén & Satorra, 1995), used in the Mplus software, is a slight variation on the LISREL model. The primary difference between LISCOMP and LISREL is that in LISCOMP the exogenous variables are manifest covariates instead of latent exogenous variables.

$$\vec{\eta}_i = \vec{\alpha} + B\vec{\eta}_i + \Gamma\vec{x}_i + \vec{\zeta}_i \qquad \vec{\zeta}_i \sim \mathcal{N}\left(\vec{0}, \Psi\right) \tag{1.32}$$

$$\vec{y}_i = \vec{\nu} + \Lambda\vec{\eta}_i + K\vec{x}_i + \vec{\varepsilon}_i \qquad \vec{\varepsilon}_i \sim \mathcal{N}\left(\vec{0}, \Theta\right) \tag{1.33}$$

As a consequence of treating exogenous variables as manifest, there is no measurement model for exogenous variables. Similarly, the expected multivariate mean and variance are conditional on the observed covariates ($\vec{x}$).

$$\mathscr{E}\left\{\vec{y} \mid \vec{x}\right\} = mean(\vec{y} \mid \vec{x}) = \vec{\nu} + \Lambda(I - B)^{-1}(\vec{\alpha} + \Gamma\vec{x}_i) \tag{1.34}$$

$$\mathscr{E}\left\{\vec{y}\vec{y}^{\mathsf{T}} \mid \vec{x}\right\} = cov(\vec{y} \mid \vec{x}) = \Lambda(I-B)^{-1}\Psi(I-B)^{-\mathsf{T}}\Lambda^{\mathsf{T}} + \Theta \tag{1.35}$$

The means vector in LISCOMP is identical to the means vector for the y variables in LISREL, with the exception that there is continued dependence on the subscript i in the LISCOMP means. Similarly, the covariance matrix in LISCOMP is equal to the covariance block of the y variables in LISREL when the exogenous matrices (Γ and Φ) are set to zero. The last SEM notation described is RAM.

RAM

The RAM specification (Recticular Action Model; McArdle & McDonald, 1984) is quite distinct from LISREL and LISCOMP. There is no differentiation between exogenous and endogenous variables; no separate measurement and structural models. The basic equation looks similar to that used in the structural models of LISREL and LISCOMP without covariates or exogenous variables.

$$\vec{v}_i = A\vec{v}_i + \vec{u}_i \qquad \vec{u}_i \sim \mathcal{N}\left(\vec{M}, S\right) \tag{1.36}$$

where $\vec{v}_i$ and $\vec{u}_i$ are vectors of manifest and latent variables. The $\vec{v}_i$ vector can be considered the manifest and latent variables vector, whereas the $\vec{u}_i$ vector can be thought of as the manifest and latent residuals corresponding to each element of $\vec{v}_i$. Equation 1.36 then implies

$$\vec{v}_i = (I-A)^{-1}\vec{u}_i \tag{1.37}$$

By introducing a matrix $F = (I; \mathbf{0})$, some part of which is the identity matrix and the other part being a zero matrix, the mix of latent and manifest variables in $\vec{v}$ can be selected to contain only the manifest variables $\vec{g}$.

$$\vec{g} = F\vec{v} = F(I - A)^{-1}\vec{u} \tag{1.38}$$

So, the observed mean and covariance matrices implied by the RAM specification are

$$\mathscr{E}\{\vec{g}\} = mean(\vec{g}) = F(I - A)^{-1}\vec{M} \tag{1.39}$$

and

$$\mathscr{E}\{\vec{g}\vec{g}^{\mathsf{T}}\} = cov(\vec{g}) = F(I - A)^{-1}S(I - A)^{-\mathsf{T}}F^{\mathsf{T}} \tag{1.40}$$

The main benefit of the RAM specification is its easy correspondence with SEM diagrams. McArdle and Boker (1990) and Boker, McArdle, and Neale (2002) describe the relation between the path tracing rules of Sewall Wright (1934, 1918, 1920) and modern SEM diagrams. All elements of the A matrix correspond to one-headed arrows from one variable (manifest or latent) to another: asymmetric relationships. All elements of the S matrix form two-headed arrows between variables: symmetric relationships of variances, covariances, and residual variances. The elements of the F matrix simply select which variables are manifest and which are latent: filtering latent (circles) from manifest (squares) variables. Finally, the M matrix contains all the means and intercepts of all the variables. The RAM notation is most notable for its economy of specification by using only four matrices: A, S, F, and M. By contrast LISREL uses 13 matrices and LISCOMP has 8.

Given the different model specifications, it is a valid question to ask "Can

any model in one notation also be specified in another?". The brief answer to this question is "yes". Given the correct block-constructed matrices any LISREL model can be put into RAM form (Horn & McArdle, 1980). Any model in RAM form can be put into a subset of the LISREL matrices. The LISREL and LISCOMP forms are likewise transformable into one another. The choice of notation is often a consequence of convenience concomitant to the choice of SEM software. The final item of SEMs examined here is their likelihood.

Likelihood

The distribution in structural equation modeling is multivariate Normal. Therefore, the model specification results in a vector of implied means and a matrix of implied covariances. The SEM then suggests that the data are distributed according to these implied means and covariances. If $\vec{y}_i$ is a vector of observations on some unit i then SEMs imply

$$\vec{y}_i \sim \mathcal{N}\left(\widehat{\vec{\mu}},\ \widehat{\Sigma}\right) \tag{1.41}$$

Hence the log likelihood of row i given the model is

$$l_i = \frac{-1}{2}\left(N_y \log(2\pi) + \log\left(\det\left(\widehat{\Sigma}\right)\right) + (\vec{y}_i - \widehat{\vec{\mu}})^{\mathsf{T}}\widehat{\Sigma}^{-1}(\vec{y}_i - \widehat{\vec{\mu}})\right) \tag{1.42}$$

or using $|*|$ for the determinant and $\vec{\tilde{y}}_i = \vec{y}_i - \widehat{\vec{\mu}}$

$$l_i = \frac{-1}{2}\left(N_y \log(2\pi) + \log|\widehat{\Sigma}| + \vec{\tilde{y}}_i^{\mathsf{T}}\widehat{\Sigma}^{-1}\vec{\tilde{y}}_i\right) \tag{1.43}$$

and the total minus two log likelihood is

$$-2\mathcal{L}_N = -2\sum_{i=1}^{N} l_i \tag{1.44}$$

The likelihood functions for SEMs and State Space Models are the same in many respects. The primary distinction is whether the distributions across rows of data are identical. In structural equation modeling, individual rows of data are assumed to be statistically independent and be governed by the same probability distribution/density function. Observations are said to be independent and identically distributed (i.i.d.). Rows of data that are i.i.d. are completely interchangeable. The joint distribution of i.i.d. observations is simply the product of the probabilities of the individual rows, or in log-space the sum of the log likelihoods as seen above. In state space models, the row residuals are independent but they are not identically distributed. Independence greatly reduces the potential complexity of the joint distribution function. Thus, independence is an extremely convenient property that is common to both SEMs and SSMs.

Before we elaborate on SSMs, we provide an overview of several other options that could be explored to investigate the central question of this dissertation: how do we find people who have similar patterns of change over time?

1.4.3 N-way Factor Analysis

The typical factor analysis contains two types of replicates. For example, Cattell's R-technique is performed on tests and persons. The factor structure of the tests is examined by using persons as replicated samples of the tests. Cattell's Q-technique seeks to find the factor structure of persons by using tests as replicated samples of the persons. All six possible two-way combinations of persons, occa-

sions, and tests form the basic factor analytic designs of Cattell (1952, 1966a), R/Q, P/O, S/T. These designs are various kinds of two-way factor analyses.

Tucker (1963, 1966) did foundational work that extended the two-way factor analyses to three "modes". Thus, 3-way factor analysis could handle multiple persons measured on several variables at many occasions. An N-way factor analysis to address the issues confronted by this dissertation would seek the person structure of multivariate time series. Extending further to N modes (e.g. Snyder, Law, & Hattie, 1984) directly addresses the issue of analyzing the full data box that Cattell (1966a) described as 10-dimensional.

Initially, N-way factor analysis seems like an ideal method for analyzing the kind of data considered here. However, the primary drawback of N-way factor analysis is that it is highly underdetermined. Snyder et al. (1984) provide not one N-way factor model, but 12 and each has its own assumptions, and unique, non-overlapping interpretations. There is not one N-way factor model but many. An additional difficulty in using N-way factor analysis to find the person structure of multivariate time series is there is no special treatment of time. Just as Cattell's (1947) P-technique was criticized for not handling the autocovariance dynamic structure of a time series (Anderson, 1963), an N-way factor analysis of persons, variables, and times would be subject to the same critique. Time in N-way factor analysis is treated like any dimension: without intrinsic order. Autoregressive, moving average models could be implemented with the method of time-delay embedding (Abarbanel, Brown, Sidorowich, & Tsimring, 1993, p. 1342, equation 18), but is not part of the method initially. Because of the underdeterminance and temporal non-ordering problems, N-way factor analysis will not be investigated in this dissertation.

1.4.4 Pooled Cross-Sectional Time Series Approaches

There are methods for analyzing pooled cross-sectional time series. For example, Dielman (1989) discussed models of multiple entities, each with a multivariate time series, primarily from the perspective of finance and comparative political ecology. Intuitively, this seems very relevant. However, in pooled cross-sectional time series analysis there is little or no consideration of latent variables and measurement error. It often uses OLS regression and no autoregressive effects (no dynamics). Repeated measurements of people almost certainly are temporally related to one another. Because of the limited treatment of latent variables and dynamics, pooled cross-sectional time series are not further considered here.

1.4.5 Multilevel Models

Multilevel modeling is another possible strategy for representing several multivariate time series at once. Level one units would be occasions of measurement which are nested within persons as level two units. The differences in dynamics between people would then be modeled as random effects with some mean and variance. However, several difficulties arise with this approach. First, autoregressive effects representing dynamics are possible but difficult to implement in multilevel models. They are also somewhat limited in the multivariate case. Second, multilevel models are usually made of directly observed variables without consideration of measurement error. Most popular multilevel methods work only on manifest variables, without latent variables. Most of these methods are multiple multilevel univariate regression. Multivariate multiple multilevel regression is possible, but it is difficult to specify a general covariance matrix for the level 1 or level 2 random effects. So the typical multilevel regression approach, having

only one outcome variable, would not adequately represent multiple multivariate time series.

However, there exists some software, for instance Mplus (Muthén & Muthén, 1998-2010), that is well-suited to model multivariate multilevel data with latent variables in a multilevel structural equation modeling framework. Mplus does not have built-in functionality for time series modeling, yet Song and Zhang (2014) recently showed that it can fit multilevel full information dynamic factor models by using lagged data columns. There are still several drawbacks to this strategy. First, the method of Song and Zhang (2014) does not incorporate a Kalman filter (Kalman, 1960; Kalman & Bucy, 1961; Grewal & Andrews, 2008). As such, the modeled time series must be stationary. This limits the applicability of the method to developmental processes, which are often non-stationary by having at least a time-varying mean. The second drawback to the Song and Zhang (2014) strategy regards the random effects. With this technique individual differences in temporal dynamics are conceived as qualitatively similar. They vary from one person to the next as random variation in a distribution. By contrast, with the state space mixture method there is some finite set of possible dynamics and the task is to select which people are members of each element in that set. The multilevel conception would have an infinitely large set of such possible dynamics, but the mixture model has a small finite set. This makes the mixture dynamics much simpler to understand for many applied researchers. Because of this ease of understanding and its ability to handle non-stationary dynamics, state space mixture models are preferred over state space multilevel models in this dissertation, but the latter are a fruitful avenue for further investigation.

1.4.6 Latent Differential Equations

Latent differential equations use structural equations models to represent smoothed derivatives as latent variables and differential equations as regression relations between latent variables. Boker, Neale, and Rausch (2004) use a modified RAM notation (McArdle & McDonald, 1984) to derive the model-implied covariance matrix of a latent differential equation.

$$\hat{R} = L(I - A)^{-1}S(I - A)^{-\mathsf{T}}L^{\mathsf{T}} + U \tag{1.45}$$

where $\hat{R}$ is the model-implied covariance matrix, L is the factor loadings matrix that defines the latent variables as smoothed derivatives, A is the matrix of regression coefficients between latent variables that define a differential equation, S is the covariance matrix of the latent variables, and U is the diagonal matrix of residual variances for the manifest variables. The principle innovation of latent differential equation modeling is in the setting of the factor loadings matrix so the latent variables are defined as derivatives. It should be noted that the above implied covariance matrix is identical to that of an endogenous-only LISREL model, or an Mplus model without covariates. Using the notation of LISREL and Mplus, which are identical for endogenous-only models with no covariates provided the subscripts for the LISREL matrices are dropped, the same implied covariance matrix is

$$\hat{R} = \Lambda(I - B)^{-1}\Psi(I - B)^{\mathsf{T}}\Lambda^{\mathsf{T}} + \Theta \tag{1.46}$$

This then suggests that latent differential equation models can be written in

terms of LISREL/Mplus matrices.

$$\eta_i = B\eta_i + \zeta_i \tag{1.47}$$

$$y_i = \Lambda\eta_i + \varepsilon_i \tag{1.48}$$

In theory any differential equation could be represented as an LDE; however, as far as I know only a second-order linear harmonic oscillator and two coupled second-order harmonic oscillators have ever been published.

1.4.7 Relations and Combinations of Modeling Techniques

The Relation of Structural Equation Models and Linear Dynamical Systems

Recall that the Mplus model without means is stated as follows (Muthén, 2002):

$$\vec{\eta_i} = B\vec{\eta_i} + \Gamma\vec{x_i} + \vec{\zeta_i} \tag{1.49}$$

$$\vec{y_i} = \Lambda\vec{\eta_i} + K\vec{x_i} + \vec{\varepsilon_i} \tag{1.50}$$

It is composed of a structural model (Equation 1.49) and a measurement model (Equation 1.50).

The Linear Dynamical System Model (i.e. state space model from control theory), at first inspection, might not look very similar.

$$\vec{x}_{k+1} = A\vec{x}_k + B\vec{u}_k + \vec{q}_k \tag{1.51}$$

$$\vec{y}_k = C\vec{x}_k + D\vec{u}_k + \vec{r}_k \tag{1.52}$$

But, it is composed of a state equation (1.51) and an output equation (1.52). The state equation (1.51) is directly analogous to the structural model (1.49). Furthermore, the output equation (1.52) is *identical* to the measurement model (1.50). The relationship becomes evident when the same notation is used for both models.

$$\vec{\eta}_{i+1} = B\vec{\eta}_i + \Gamma\vec{x}_i + \vec{\zeta}_i \qquad \text{(State Space State Equation)} \tag{1.53}$$

$$\vec{\eta}_i = B\vec{\eta}_i + \Gamma\vec{x}_i + \vec{\zeta}_i \qquad \text{(Mplus Structural Model)} \tag{1.54}$$

$$\vec{y}_i = \Lambda\vec{\eta}_i + K\vec{x}_i + \vec{\varepsilon}_i \qquad \text{(State Space Output Equation)} \tag{1.55}$$

$$\vec{y}_i = \Lambda\vec{\eta}_i + K\vec{x}_i + \vec{\varepsilon}_i \qquad \text{(Mplus Measurement Model)} \tag{1.56}$$

The similarity of Equations 1.53 and 1.54, and the identical status of Equations 1.55 and 1.56 should be clear. There is also a deeper relationship between the SEM and SSM structural models.

If the dynamic system described by the state space model is at an equilibrium point (alternatively, a stationary point or a fixed point), then the latent variable values are unchanged by the increment of time: that is, when $\vec{x}_{k+1} = \vec{x}_k$, then $\vec{x}_k$ is called an equilibrium point. If the state space model is at an equilibrium point, then its state equation becomes

$$\vec{\eta}_i = B\vec{\eta}_i + \Gamma\vec{x}_i + \vec{\zeta}_i \qquad \text{(State Space State Equation at Equilibrium)} \tag{1.57}$$

where it should be noted that the measurement model (output equation) is left unchanged.

By inspection there is a one-to-one correspondence between every piece of the state equation for a linear dynamical system at an equilibrium point (Equation

1.57) and the structural part of the Mplus SEM (Equation 1.54). So, the state equation at an equilibrium point is identical to the structural part of the Mplus model.

The identity between the structural part of the Mplus model (Equation 1.49) and the state equation of the linear dynamical system at an equilibrium point (Equation 1.57) should be evident; as should the identity between their respective measurement models (Equations 1.50 and 1.52). In short, a structural equation model is a a linear dynamical system at an equilibrium point.

Moreover, recall that the LISREL structural equation model (e.g. Jöreskog & Sörbom, 1982) is stated as follows.

$$\vec{\eta}_i = B\vec{\eta}_i + \Gamma\vec{\xi}_i + \vec{\zeta}_i \tag{1.58}$$

$$\vec{y}_i = \Lambda_y\vec{\eta}_i + \vec{\varepsilon}_i \tag{1.59}$$

$$\vec{x}_i = \Lambda_x\vec{\xi}_i + \vec{\delta}_i \tag{1.60}$$

It is composed of a structural model (Equation 1.58) and a measurement model (Equations 1.59 and 1.60). The structural model is identical in form to that used in Mplus. So the same relationships hold between the LISREL and state space structural models as between the Mplus and state space models. It is also possible to show that the measurement models for LISREL and linear dynamical systems are identical when certain block matrices are utilized.

Alternatively, MacCallum and Ashby (1986) show that a linear dynamical system (called linear systems theory therein) can be represented as an SEM by utilizing block matrices. Furthermore, they argue that because linear systems theory (LST) was shown to be a special case of SEM, then SEM was more gen-

eral. However, because SEM is also a special case of LST neither can truly be considered more general. Rather, both models are extremely flexible and models of each type can be fit with whichever technique is most convenient at the time. Similarly, Molenaar (1985) used block matrices in a SEMing framework to define a white noise dynamic factor model. Later, Nesselroade, McArdle, Aggen, and Meyers (2002) developed an autoregressive dynamic factor model that is mathematically identical to a state space model. And then Browne and Zhang (2004) and Browne and Nesselroade (2005) discussed autoregressive moving average (ARMA) models at the latent level as new kinds of dynamic factor models. All of these can be specified as state space models (SSMs) and fit with SSM software. In general, when an SEM computer program is used, the model is called a dynamic factor model; when an SSM computer program is used, it is called a state space model.

Multilevel Models in State Space Form

In the time series literature, it has been shown that univariate and multivariate multilevel models (e.g. Raudenbush & Bryk, 2002; Lindstrom & Bates, 1990) can be written as state space models (Jones, 1993; Icaza & Jones, 1999). More recently, Gu et al. (2014) showed that several multilevel structural equation models (e.g. Muthén, 1991, 1994) also have a state space form, and when estimated as state space models these multilevel SEMs are more computationally efficient than when estimated by standard SEM methods.

The Combination of Structural Equation Models and Mixture Models

The flexibility of mixture models allows them to easily be combined with structural equation models. The conditional probability of each row of data given

each mixture class is calculated and combined using estimated class probability weights. A structural equation model (SEM) simply specifies a Gaussian distribution, so the probability of a data row is merely the probability density of that data row in Gaussian distribution implied by the model. Different classes correspond to different SEMs and hence distinct probability density functions. The likelihood of an entire mixture SEM is the same as that for a mixture model in which the probability distributions are Gaussian and given by SEMs.

The Combination of State Space Models and Mixture Models

The same malleability that allows mixtures of SEMs also allows mixtures of state space models. The method is identical, but to the author's knowledge has not previously been attempted or published. As evidence, several well-known books on the subject do not mention the possibility (see Durbin & Koopman, 2001; Shumway & Stoffer, 2010; Zucchini & MacDonald, 2009; Quenouille, 1957). The combination of mixture models and state space models is the primary innovation investigated in this dissertation.

1.5 Review

How should multiple multivariate time series be handled? Cattell's (1952, 1966a) data box is typically reduced to two dimensions before analysis. Multimode factor analysis is one way to actually analyze a multidimensional grid (Snyder et al., 1984), but the high degree of model indeterminacy and no special treatment of time makes it hard to justify for this problem. Similarly, multilevel models (Raudenbush & Bryk, 2002) could address the nesting of time points within persons, but the inability to efficiently handle autocorrelations and measurement

error make this strategy less realistic at the present. Because of their extreme flexibility and ease of implementation, mixture models can be combined with models of linear dynamical systems.

Two strategies are used to evaluate the combination of state space models and mixture distributions. The first is a simulation study aimed at finding the efficacy of the proposed technique for uncovering the correct grouping structure in a setting where the true answers are known. The second is an empirical data analysis whose purpose is to show how the method can be applied to real-world data which always unveils problems and complications that no simulation study is likely to find. The combination of mathematical derivation, simulation, and applied problem solution illustrates a threefold evaluation of the proposed method.

Chapter 2

Simulation

2.1 Introduction

A new method, or a new combination of methods, requires a mechanism of evaluation. Mathematical and theoretical proofs can demonstrate that a method will work in theory often under several assumptions that are never born out in real data, but a simulation can address practical factors that influence the success of a method in the safe environment where the answers are known. To evaluate the performance of state space mixture modeling as a method to identify individuals with similar patterns of change, data are simulated with an underlying grouping structure and state space mixture models are used to uncover the groups. The model knows that there are groups and how many there are, but has no information on group membership, only the data. Hence, the simulation is a test case in which the true data generation mechanism is known but not to the model as it is being estimated. Once the model is estimated, the free parameter estimates and model-selected grouping structure can be compared to their true, known values. Simulation provides a uniquely safe and efficacious environment to check the

validity of a method.

The research questions asked and answered by the simulation study are:

1. How well can state space mixture modeling recover the simulated grouping structure?

2. How well can state space mixture modeling recover the simulated parameter structure?

3. What model attributes influence the above two questions and how much?

4. What sizes of problems are reasonable to estimate with the current software in terms of computational time and memory limitations?

All of the above questions are conditional on the assumption that the factor structure (zero loadings), factor dimensionality, the dynamic structure (zero entries of the dynamics matrix), and the number of groups are known. These assumptions are made to simplify the estimation problem by addressing problems that are particular and new to state space mixture modeling while ignoring several problems that are common to many methods. Finding the factor structure in a set of data is not a new problem. Spearman (1904), Thurstone (1934), Tucker, Koopman, and Linn (1969), and Preacher, Zhang, Kim, and Mels (2013) among many others have been addressing this problem for over 100 years. It is common to any problem with a linear measurement model with unknown coefficients. The number of groups problem is similar to the number of factors problem. Roweis and Ghahramani (1999) discuss that these two problems are qualitatively similar. The dynamics matrix may be partially known from theory so its zero entries are fixed.

2.2 Methods

With simulation, factors that influence how well a method works can also be investigated. In this simulation, state space models are generated for multiple people measured at multiple occasions and some people come from the same model whereas others come from different models. The broad plan for the simulation was to generate data for a set of individuals according to various instances of state space models, set-up a state space mixture model matching the data generation mechanism but with random starting values, estimate the state space mixture model, store the results of the model estimation, and finally to analyze the stored model results in comparison to each other and to the known simulation values.

2.2.1 Factors to be Varied

Five factors were varied in this simulation: the number of groups or mixtures, the number of people belonging each group, the number of occasions measured for each person, the number latent factors, and the number of measured indicators for each factor. The simulation factors and their levels are summarized below, followed by a discussion of reasons for picking these factor and levels.

Simulation Conditions for data generation were as follows:

- Number of Groups/Clusters (4 levels): 1, 3, 5, 8
- Number of People Per Group (3 levels): 1, 10, 100
- Number of Occasions (4 levels): 5, 12, 50, 200
- Number of Factors (4 levels): 1, 3, 4, 8

- Number of Variables Per Factor (4 levels): 1, 3, 4, 6

Each of the $4 \times 3 \times 4 \times 4 \times 4 = 768$ combinations of varied factors creates a condition. The goal was to achieve 1000 replications of each of the 768 conditions, yielding a data set with $768,000$ rows.

The levels of the simulation factors were selected to represent typical examples of possible research designs for which state space mixture modeling might be applied. The underlying number of groups in real data are not known, but for this simulation it was assumed to be known, fixed, and correct. The difficulty of selecting the correct number of classes in a mixture model is well-known and has been studied previously (e.g. Tofighi & Enders, 2008). As Roweis and Ghahramani (1999) comment, the number of mixture classes problem is essentially the same as the number of factors problem or the number of clusters problem, so methods suggested by Preacher et al. (2013) for selecting the number of factors or Rousseeuw (1987) for clusters could be applied in the more realistic scenario of not knowing the true number of mixture classes. The most relevant questions addressed here concern only the parameter estimation quality and the recovery of the correct person grouping structure conditional on several aspects of the model being known.

The number of people belonging to each group was varied to allow for distinct sample sizes across different conditions. Combinations of the number of groups and the number of people per group create interesting classes of models. For example, if there are 8 groups and 1 person in each group then the sample is composed of 8 distinct individuals. The question addressed by this combination is "How well can state space mixture models correctly find that each person is unique?". At the other end of the spectrum, if there is only 1 group and 100

people per group then the sample has 100 total people, each of whom is randomly equivalent to the others. It is important to understand the the total number of people in the sample is the product of the number of groups and the number of people per group. Moreover, the sample size per group is fixed by the condition. It does not randomly vary within a condition.

The number of occasions per person adds another kind of sample size to this simulation study. In many social science research settings, "sample size" refers to only to the number of people. It may be the number of people assigned to a treatment condition, or at a particular research center, but the number of observations is equivalent the number of people. The number of measurements per person is generally assumed to be one. In idiographic settings, sample size would be the number of times an individual was measured. The number of individuals is assumed to be one. Here we combine both sample sizes. There is a sample size that refers to the number of people, and another that refers to the number of times a person was measured. Of course, we could also consider the number of variables measured on a person at each occasion to be a sample size. When considering a number of people measured at multiple occasions on several variables, Cattell's (1952, 1966a) data box becomes a useful visualization.

The Number of Occasions varies from 5 to 200. The time sample size varies from what would definitely be called "longitudinal" research at 5 time points, up to definitely "time series" research using 200 time points, with two points that have more moderated numbers of occasions. The spread in temporal sample size is intended to bridge the gap between longitudinal data and time series data by using the same method in both cases. This can be accomplished by trading off persons for time points (e.g. Boker, Horn, Meyer, & Turkheimer, 2014), or more generally trading off sampling units at different levels whether they be spacial

points and time points or persons and locations. Having multiple identical units adds statistical information on which models can be based. This trade off allows multiple homogeneous people measured at only a few times to be used to estimate a "time series" model, the state space model.

The number of factors was varied to create different kinds of dynamics. The number of factors could also be called the dimension of the state space. This is the dimension of the space in which the dynamics exist. The dimensionality of the state space is important because certain kinds of behavior have minimum requirements on their existence based on this dimensionality. For instance in continuous-time dynamical systems, chaos only exists in dimension three or more (Poincaré-Bendixson Theorem; Strogatz, 2000; Cvitanović, Artuso, Mainieri, Tanner, & Vattay, 2012), and when the space is finite-dimensional chaos only exists in non-linear systems (Hirsch, Smale, & Devaney, 2003). Hence, in the linear system modeled here, chaos is not possible. Likewise, oscillations require at least two dimensions in the state space (Arnold, 1973). Thought of intuitively and geometrically, oscillations in the state space consist of spirals, ellipses, and circles which only exist in two or more dimensions. Thus, by varying the state space dimension (i.e. the number of factors), the kinds of dynamic behavior possible are also varied. Similar to the Number of Variables Per Factor, the first level of the Number of Factors is a limiting case of a single underlying dimension. Many dynamic models in the literature limit themselves to this case either for conceptual or computational reasons. For example, Latent Differential Equations models most frequently occur as a single latent dimension (e.g. Boker et al., 2004; Boker, Montpetit, Hunter, & Bergeman, 2010; Steele & Ferrer, 2011), and only rarely occur as two dynamic dimensions. Dynamic factor analysis is the same (e.g. Browne & Nesselroade, 2005). The higher levels of Number of Factors are

a sample of what is seen in the state space modeling literature (e.g. Åström & Murray, 2010).

Lastly, the number of variables loading on each factor was varied from 1 to 6. The first level of Number of Variables Per Factor is a limiting case when each latent variable is identified with an observed variable. Thus it is a special case of essentially no latent variables, of using vector autoregressive modeling without the need for the stationarity assumption. When there is only one indicator for the factor then the factor is made identical to the indicator. In general the factor loadings between the latent and observed variables are generated at random from a uniform distribution between 0.4 and 1.0 to represent a range from modest factor loadings to very strong ones; however, when there is one variable per factor, the factor loading is set to 1.0 and is fixed in the model to identify it. The other levels of Number of Variables Per Factor span a range from a small number of latent variables to a moderate number. As the Number of Variables Per Factor increases the factor becomes more easily identified as the common information among the items. Note that according to Roweis and Ghahramani (1999, p. 309, footnote 5; Goodwin & Sin, 1984), a state space model can be identified if the latent state space dimension is less than or equal to the product of the observed variable dimension and the number of consecutive observations, $k \leq \tau p$, and in this simulation $k \leq p$ with additional constraints made to ensure identifiability of the models.

2.2.2 Data Generation

Aside from the varied simulation parameters, several other data generation decisions were made. First, without loss of generality the generating model was

always a state space model without covariates, inputs, or exogenous variables. Without input, we are considering only the free-running dynamics of the system. Generally, models are more easily estimated when under the influence of time-varying external variables (Gregson, 1983), so the lack of these actually creates a slightly more difficult estimation problem.

Individual time series were generated in measurement invariant groups. People within the same group were randomly sampled under identical generating models. People in distinct groups had the same factor loadings and residuals, but different autoregressive dynamics. In particular, the factor loadings matrices (C matrices) were generated with simple structure. When there was only one indicator per factor, the loading was fixed to unity. When there were more indicators for each factor, the nonzero loadings were randomly sampled from a uniform distribution between 0.4 and 1.0 ($\mathcal{U}(0.4, 1.0)$). In all cases, the factor variance was used to identify the scale of the latent variable by fixing its value to 0.2. As mentioned previously, the factor loadings were generated to be equal across people and mixtures.

Similar to the factor loadings, the unique variances were generated to be equal across people and mixtures. The unique variances (R matrices) were generated with diagonal structure. The nonzero variances were sampled from $\mathcal{U}(0.0001, 0.3)$. Because the factor loadings and uniquenesses were equal across people and groups, it is apt to describe these as measurement invariant groups. Only the dynamics differed between groups. This corresponds to a case in which the measurement process is identical across all individuals, but their patterns of change are distinct. An interesting alternative formulation is the converse: the dynamics are uniform across people, but they are measured in idiosyncratic ways. This alternative is a dynamic analog of Nesselroade and colleagues' idiographic filter (Nesselroade,

Gerstorf, Hardy, & Ram, 2007; Z. Zhang, Browne, & Nesselroade, 2011; Molenaar & Nesselroade, 2012). The idiographic filter design could be the subject of further simulation and applied investigation.

In the present study, only the dynamics matrices (A matrices) differed across groups. Because the qualitative behavior of a dynamical system is completely determined by its eigenvalues (e.g. Hirsch et al., 2003), the dynamics matrix for each group was generated based on its eigenvalues. First, the number of complex (as opposed to real) eigenvalues was randomly uniformly sampled from the maximum range possible for that simulation condition[1]. Then the remaining eigenvalues were assigned to be real. The complex eigenvalues were sampled uniformly from the unit square in the complex plane, whereas the real eigenvalues came from the unit line, $\mathcal{U}(-1, 1)$. Because some of the eigenvalues have modulus outside of the unit circle, some of the dynamics are non-stationary, but the vast majority are stationary. The Kalman filter method of estimation used here does not require stationary dynamics, so this does not violate any assumptions.

Once the real and complex eigenvalues were generated, they were then arranged into the block-diagonal Jordan canonical form, creating the canonical dynamics matrix for this randomly generated system. Finally, a unitary (complex orthogonal) rotation matrix, U, was generated by permuting the identity matrix, and the canonical dynamics were rotated into those used for the data generation.

[1]For conditions with one factor the possible number of complex eigenvalues is only zero; for 3 factors, the it is zero or two; for four factors, the it is zero, two, or four; and for eight factors, it is zero, two, four, six, or eight.

$$U = permute(I) \tag{2.1}$$

$$A_{use} = U^{-1} A_{canonical} U \tag{2.2}$$

Thought of in the dynamic sense, the $A_{canonical}$ matrix gives the canonical dynamics, the dynamics without regard to the actual directions of flow in the state space. The U matrix then gives those directions, the eigenvectors for the dynamics matrix which give the axes of the flow. The columns of the U matrix are the eigenvectors of the dynamics. Because the U matrix is simply a permutation of the identity matrix, the directions of flow are restricted to be orthogonal and are not completely general. With that being said, the qualitative dynamic behavior considered is completely general to all linear dynamic systems. The restricted eigenvectors were chosen to allow more zeros in the dynamics matrix than would exist if a more general rotation were used. The zeros were assumed to be known and fixed by the model. Hence, the zeros represent hypotheses by the researchers about which latent variables are not related to one another over time. Although early simulation results indicated that zero entries in the dynamics matrix could accurately be estimated when present, it was considered an additional challenge for state space mixture modeling and was deferred for later work.

To verify that the matrix elements and their eigenvalues had the intended distributions, a small-scale simulation was conducted on a 3-dimensional example. The simulation created 10,000 3-dimensional (i.e. 3×3) dynamics matrices. Figure 2.1 depicts the distribution of their eigenvalues. The upper section shows the eigenvalues in the complex plane along with the unit circle for reference. It should be clear that the eigenvalues are uniformly distributed bivariate uniformly

between -1 and +1 with the real line (imaginary part equal to zero) being over sampled. The lower section shows the frequency distribution of the modulus (magnitude) of the the same eigenvalues. The reference line at 1.0 coincides with the location of the unit circle. Figure 2.2 shows the frequency distribution of the the actual matrix elements of the dynamics matrices generated. Thus the histogram in the second row and the first column refers to the distribution of the analogous element of the 3×3 matrix. The diagonal elements are all uniformly distributed as expected. The off-diagonal elements are also uniformly distributed with the exception that zero is over-sampled, which corresponds to the real line also being over sampled as shown in the upper panel of Figure 2.1. Note that the range of the histograms is truncated because of the large spike in frequency at zero for the off-diagonal elements. An alternative rotation matrix U would induce a different matrix element distribution, but leave the eigenvalue distribution unchanged. Based on Figures 2.1 and 2.2, the simulation appears to be generating eigenvalues and matrix elements as expected.

With the discussion of the dynamics eigenvalues, there is an opportunity for confusion of the dynamics eigenvalues with the covariance eigenvalues typically encountered in principal components analysis (e.g. Leon, 2006; Raykov & Marcoulides, 2008). We next consider these differing sources of eigenvalues.

2.2.3 Sources of Eigenvalues

There are eigenvalues for different matrices. For example, the dynamics might be

Figure 2.1: Eigenvalue Distribution for 10,000 Simulated 3-Dimensional Dynamics Matrices

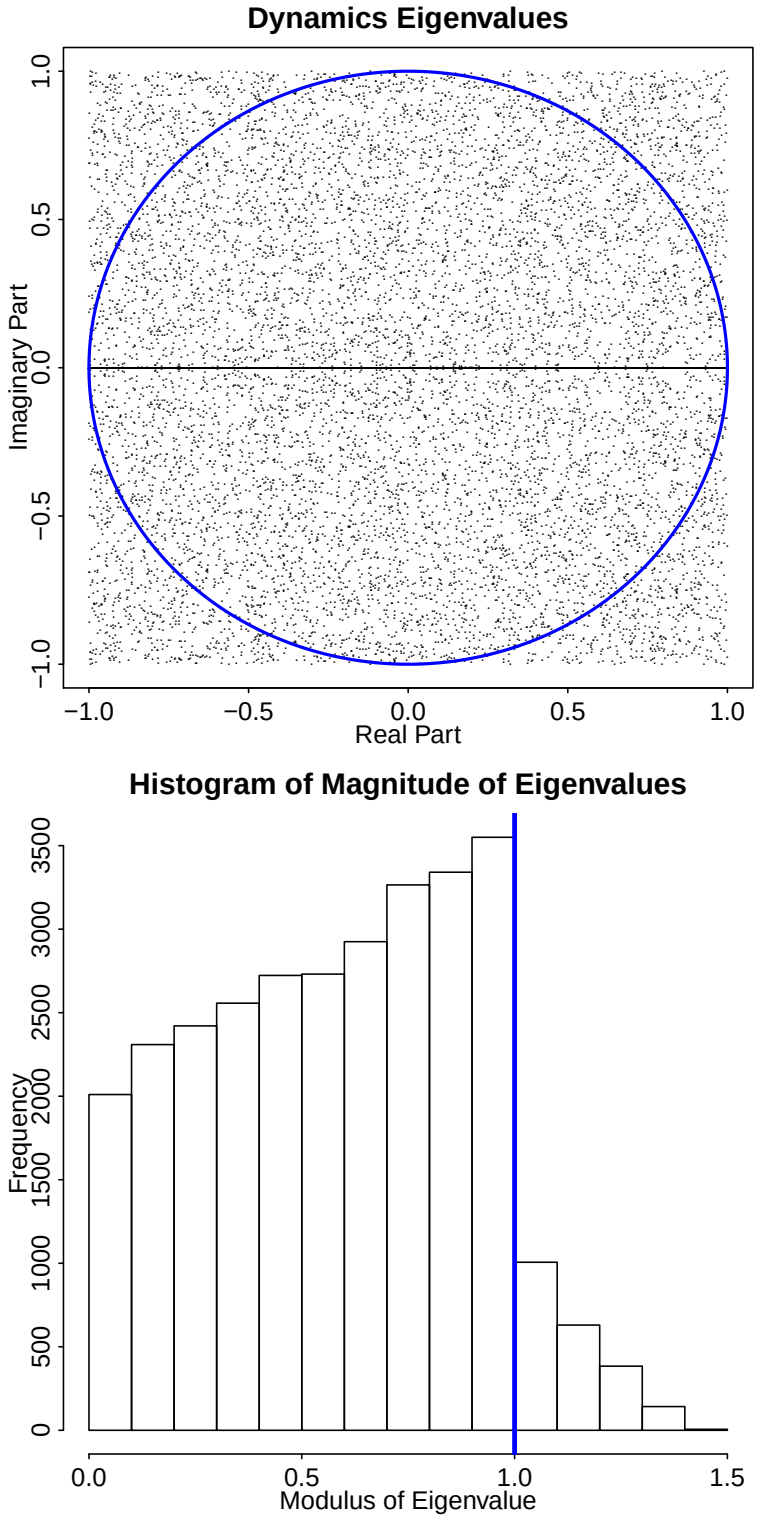
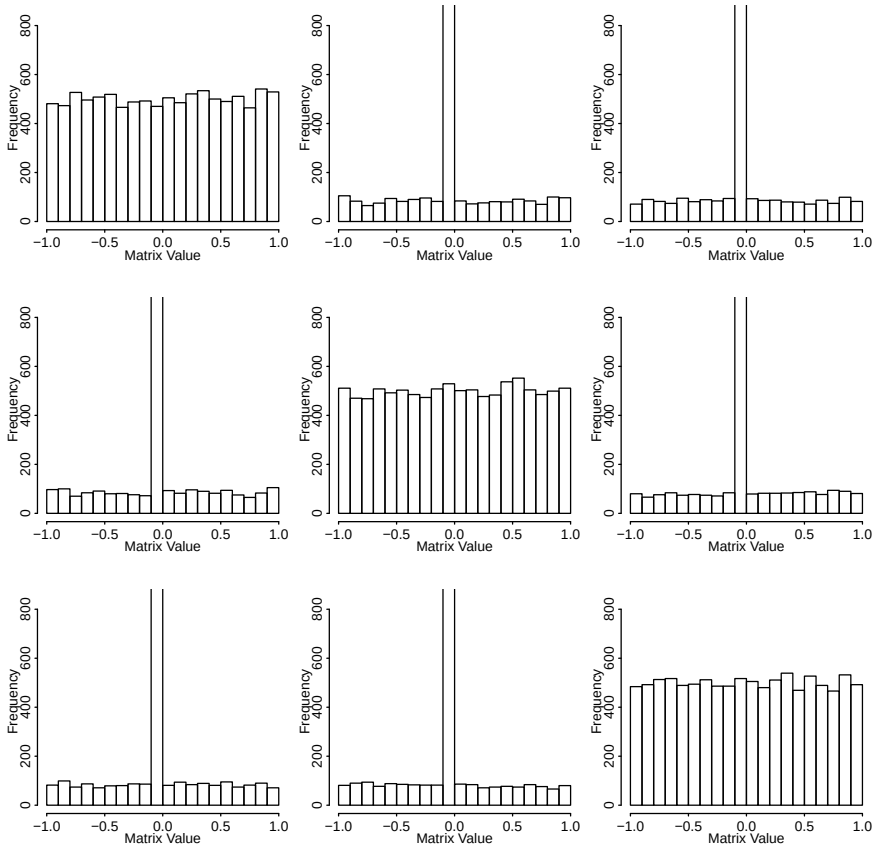


Figure 2.2: Matrix Element Distribution for 10,000 Simulated 3×3 Dynamics Matrices



$$A = \begin{pmatrix} 0.07 & -0.07 & 0.00 & 0.00 \\ 0.07 & 0.07 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.12 & 0.31 \\ 0.00 & 0.00 & -0.31 & -0.12 \end{pmatrix}$$

with eigenvalues $\lambda_k = -0.12 + 0.31i, -0.12 - 0.31i, 0.07 + 0.07i, 0.07 - 0.07i$. Notice that A is 4×4 and has 4 eigenvalues. These dynamics when paired with a scalar dynamic error covariance matrix $Q = 0.2I$ generate a contemporaneous (i.e. lag zero) covariance matrix as follows[2]

$$P \approx \begin{pmatrix} 0.20 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.20 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.22 & 0.01 \\ 0.00 & 0.00 & 0.01 & 1.22 \end{pmatrix}$$

with eigenvalues $\lambda_k = 0.24, 0.21, 0.20, 0.20$. Note that the eigenvalues of the dynamics matrix A are not equal to the eigenvalues of the covariance matrix P, and all of the covariance eigenvalues are real whereas the dynamics eigenvalues are sometimes complex. This is because the covariance matrix is necessarily symmetric and real, and hence Hermitian. All Hermitian matrices have all real eigenvalues (Magnus & Neudecker, 1988, p. 14; Leon, 2006, p. 347).

Continuing in this vein, generating data for a person with these dynamics yields a lag zero covariance matrix with eigenvalues

$$\lambda_k = 0.62, 0.55, 0.43, 0.37, 0.22, 0.21, 0.12, 0.10, 0.09, 0.08, 0.05, 0.02$$

[2]The formula used is $vec(P) = (I - A \otimes A)^{-1} vec(Q)$, the solution to the discrete Lyapunov equation from control theory (Kitagawa, 1977)

Notice that there are 12 eigenvalues, all of them real, not complex. Any real symmetric or complex Hermitian matrix has only real eigenvalues (Leon, 2006). The dynamics matrix A need not be symmetric, so its eigenvalues generally are complex. The lag zero covariance matrix, on the other hand, is always symmetric and hence has only real eigenvalues. The data in this model have 12 variables (three indicators per latent factor), thus the 12 eigenvalues. Generating data for another person with the same dynamics gives similar but not identical eigenvalues

$$\lambda_k = 0.72, 0.53, 0.43, 0.29, 0.21, 0.18, 0.15, 0.12, 0.09, 0.08, 0.05, 0.02$$

The eigenvalues of the correlation matrix of the data for person 1 are

$$2.47, 2.32, 1.88, 1.56, 0.87, 0.63, 0.56, 0.50, 0.42, 0.37, 0.25, 0.18$$

and person 2 with the same dynamics would have eigenvalues:

$$2.46, 2.36, 1.75, 1.53, 0.87, 0.81, 0.59, 0.54, 0.40, 0.35, 0.23, 0.11$$

Note that a scree plot of these covariance eigenvalues would select 4 factors.

Finally, the eigenvalues of latent 4-dimensional covariance space are

$$0.29, 0.22, 0.17, 0.13$$

for person 1, and

$$0.20, 0.20, 0.18, 0.17$$

for person 2. Both of these closely match the theoretically derived eigenvalues given above as $\lambda_k = 0.24, 0.21, 0.20, 0.20$.

To summarize, there is some covariance structure in the data but it is mostly autocovariance structure generated by the dynamics leading from one row to another rather than from one column to another at the same time. As mentioned in an earlier chapter, the dynamic error Q is essentially the factor covariance matrix in factor analysis: the Ψ or Φ matrix from LISREL notation. In this example, Q is a constant multiple of the identity matrix. More oblique factor structure in the state space model could be created through making Q non-diagonal. In addition to several distinct sources of eigenvalues in this simulation, there are also several source of error which are discussed next.

2.2.4 Sources of Error

There are two sources of error in this simulation. The first is dynamic error. Below is the state equation (structural model) for a state space model with no covariates/inputs.

$$\text{The State Equation:} \quad \vec{x}_{t+1} = A\vec{x}_t + \vec{q}_t \tag{2.3}$$

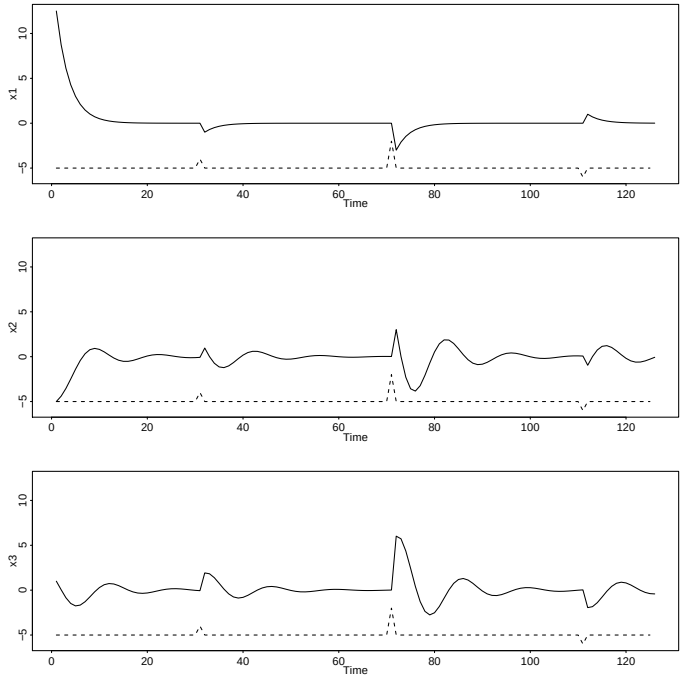
The true dynamics are in the A matrix, $\vec{x}_t$ is the vector of latent variables at time t, and q_t is a Gaussian-distributed random vector with mean zero and (often) diagonal covariance matrix. So the next latent state is not perfectly a function of A and the current state. Rather, it is a function of the true dynamics, the current state, and additive error that has a zero-mean and some covariance structure. Different covariance structures are possible: white noise (spherical), diagonal, zero, or some theory-determined alternative form. The kind of error in the state equation is sometimes called process noise, random shocks/disturbances, or dynamic error.

The second source of error is in measurement. Below is the output equation (measurement model) for a state space model with no covariates/inputs.

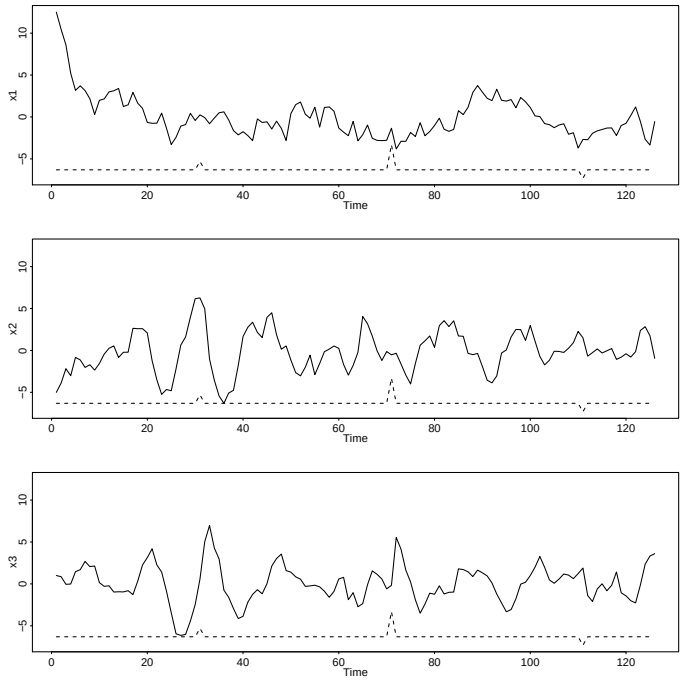
$$\text{The Output Equation:} \quad \vec{y}_t = C\vec{x}_t + \vec{r}_t \tag{2.4}$$

The true output rules are in the C matrix, $\vec{y}_t$ is the vector of observed variables at time t, and $\vec{r}_t$ is a Gaussian-distributed random vector with mean zero and (often) diagonal covariance matrix. This is directly analogous to factor analysis. In fact, the factor model can be represented as a state space model by setting the dynamics matrix, A, to a zero matrix. Priestley and Subba Rao (1975) showed that regression factor scores are identical to Kalman filter state estimates. The matrix C is isomorphic to a linear measurement function that determines how the latent variables produce actual observations. This measurement process is not perfect, but rather is subject to additive noise with zero mean and some covariance structure.

The relationship between dynamic and measurement noise can be seen in Figures 2.3 and 2.4. Figure 2.3a shows a three-dimensional latent process without dynamic error. The process unfolds smoothly over time with clear jump discontinuities in response to the occasional measured shock inputs, shown by the dashed lines at the bottom of each panel. In Figure 2.3b, the same dynamics are governing the behavior of the system, but now there is dynamic noise contributing to process. Importantly, the dynamic noise process is not simply the result of adding noise to the noise-free process. Rather, the noise influences the dynamics as they proceed. The noise acts as a random, unmeasured shock or input to the system at every time point, just as the inputs/covariates act as systematic, measured shocks at times 30, 70, and 110.



(a) Without dynamic noise



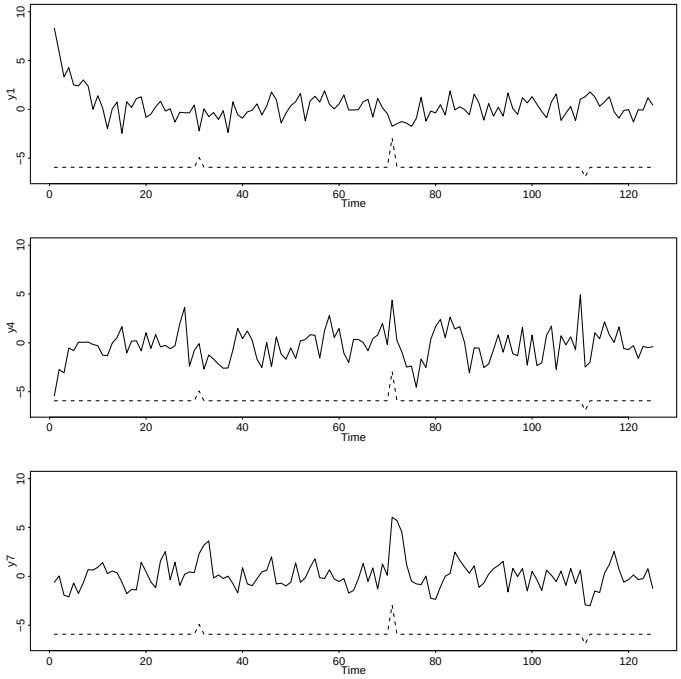
(b) With dynamic noise

Figure 2.3: Latent values of a simulated 3-dimensional state space model with inputs

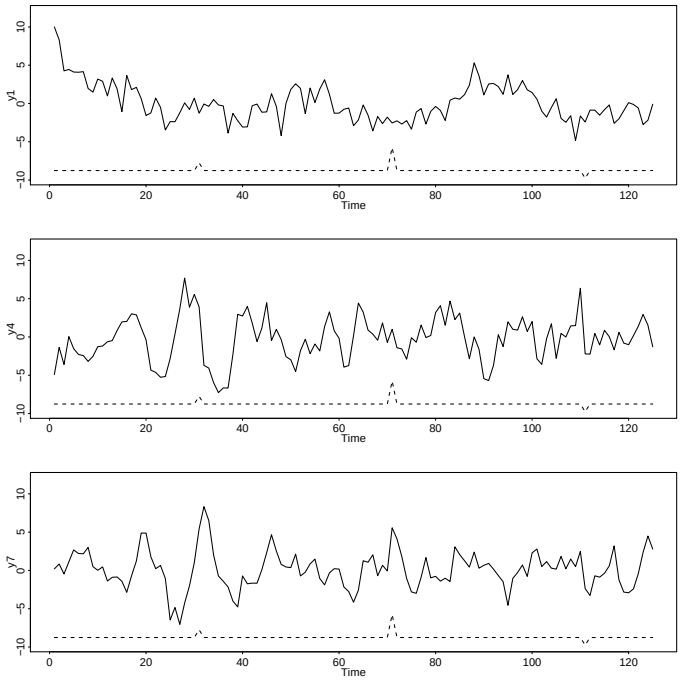
Figure 2.4 acts in parallel with Figure 2.3. Where Figure 2.3 shows the *latent* variables over time, Figure 2.4 shows the *observed* variables over time. One indicator is shown for each factor, rather than all nine indicators. Note that both observed time series appear noisy. This is due to the measurement error operating in both conditions. Likewise, notice that the observed processes do not appear identical when their only difference is the dynamic error. The dynamic rules are the same across both processes but their observed patterns are distinct. Thus, applying a dynamic model allows a degree of abstraction away from the observed values toward the rules that determine the underlying process.

For this simulation study, data are generated according to the state space model. The simulated data incorporate both process noise and measurement noise. Because both of these sources of noise are part of the model, the simulated data represent a random sample from a population in which the generating model is correct. In contrast to the approach here, Tucker et al. (1969) simulated data from a factor model by using a factor model with a small number of true factors with large variance and large number of noise factors with small variance. The Tucker-Koopman-Linn procedure is intended to create data that have an underlying factor model, but where the factor model is only approximately true (MacCallum & Tucker, 1991; MacCallum, 2003). The difference between the generating model and the estimated model is called model error (MacCallum & Tucker, 1991). In this simulation, there is no model error. Both sources of error (process and measurement) are represented by the estimation routine. Put another way, the model is true in the population; there is no model error.

If the purpose of the simulation study were to fit a single model to each person's multivariate time series, then generating time series with model error would be trivial. However, our purpose is not so simple. The purpose is to



(a) Without dynamic noise



(b) With dynamic noise

Figure 2.4: Observed values of a simulated 3-dimensional state space model with inputs

take several people, each with their own time series and some of whom share the same dynamics, and then to fit a model to all of these people and determine which people are governed by the same dynamics. Each person has a state space model that is true in the population of time points. The state space model for each person is not approximately true, but exactly true in this simulation. The question of interest lies in how well people can be grouped by similar dynamics when the model is true. Future work should certainly address the more difficult and applicable question of how well this procedure works when the model is only partially correct.

2.2.5 Model Fitting

A state space mixture model was fit to each multigroup data set. The number of mixtures and the number of factors were assumed known and correct. Additionally, the zero factor loadings were assumed known. However, the nonzero loadings, mixing parameters, nonzero dynamics matrix elements, and residual variances were freely estimated. With the exception of the dynamics matrix, starting values were randomly generated as another call to the same functions that randomly generated the true values for the parameters. The dynamics starting values were generated by taking the true dynamics and adding uniform random values sampled from $\mathcal{U}(-.5, .5)$ to the nonzero values. The different procedure for the dynamics matrix was used to keep the number and placement of nonzero entries constant from the true values to the starting values. Because of the rotation used on the dynamics eigenvalues, this procedure also keeps the number of real and complex eigenvalues consistent between true and starting values by adjusting only their magnitudes.

Models were estimated in R (R Development Core Team, 2014) using the OpenMx package (Neale et al., in press) and the `mxExpectationStateSpace` function (Hunter, 2014b) which implements state space models fit with full information maximum likelihood and prediction error decomposition via the Kalman filter. The state space mixture was created in OpenMx by having a submodel for each person in each class, and making a weighted combination of these submodel likelihoods following a mixture model. Critically, the mixing parameters apply to all people. Although it is possible to estimate a mixing parameter for person-group combination, it creates an inconsistent estimator similar to the problem of estimating person and item parameters simultaneously in item response theory: as the number of people increases, the estimation does not improve asymptotically because for each new person there is another free parameter that must be estimated. Consequently, the mixing parameters apply to each group identically across all people. Individual likelihoods of each person within each group can then be calculated *post hoc* by weighting their individual model likelihoods for each model with the mixing parameters via Bayes' theorem, treating the mixing parameters as prior probabilities.

Matching the two-fold goals of this simulation study, two data sets were created to store the results of the model fitting. In the first data set, the parameter estimation was recorded. All starting values, true values, and estimated values were written to this data file. Additionally, the number of iterations used by the optimizer, the convergence of the optimizer, and the amount of estimation time were also recorded. In the second data file, the recovered grouping structure was tracked by keeping the mixing parameters and the likelihood of each person-group combination. These two files were then analyzed to assess the feasibility of state space mixture modeling to recover individuals with similar patterns of change

over time from a sample of multiple people measured at multiple times.

2.2.6 Data Analysis

The data analyzed are the results of the model fitting, not the data generated and used in the model fitting. Thus the analysis uses the modeling results as data.

Parameter Estimation

A reasonable stipulation is that the parameters of a state space mixture model can be accurately estimated, at least when the sample sizes of persons and occasions per group are sufficiently large. In fact, accurate parameter recovery was shown by Hunter (2014a) with a single simulated time series of 1000 observations spread over two mixture classes. However, there remain several interesting questions to address. First, how well can dynamics parameters be estimated on longitudinal data (e.g. five observations) when the number of persons is large? That is, what is the trade-off between additional time points and additional people with regard to parameter estimation? And second, what other factors influence parameter recovery? Put another way, what are sufficient sample sizes to adequately estimate parameters for these models?

The primary tools used to evaluate parameter recovery are correlation and the root mean square (RMS) difference. The correlation coefficient provides a relative measure of agreement, whereas the RMS indicates absolute agreement. Both are used because of their complementary strengths and weaknesses. Note that because starting values, true values, and estimated values were all stored there are actually three comparisons that can be made about the parameters: starting-true,

starting-estimated, true-estimated. The starting-true comparison is a reasonable null against which to compare the true-estimated recovery. The starting values were all randomly generated, so the degree of agreement (as measured by the correlation or RMS) between the starting and true parameters is only what you would expect by chance. Similarly, the starting and estimated parameters should be dissimilar, unless an estimation problem occurred in which the parameters were not able to move far enough away from their initial values. Finally, the true-estimated comparison provides the traditional view of parameter recovery. The estimated parameters should be unbiased estimates of the true parameters with estimation errors that approach zero as the sample sizes increase. We next discuss the analysis of the grouping structure.

Grouping Structure

The key goal of this simulation study was to explore how well the underlying grouping structure could be recovered in state space mixture models under varying conditions. The only way this goal was achievable was in the simulation context where the true grouping identities were known. When the true grouping assignment is known the problem of evaluating state space mixture models' assignments can be considered in several ways, each with their own metrics. First, grouping recovery can be conceptualized as a multiclass classification problem. The state space mixture model is treated as a classifier where its job is to assign each person to one of several distinct classes. Second, because the true and estimated grouping can be arranged into a confusion matrix, a two-way frequency table, the independence of the two categorical variables can be evaluated with a χ^2 test of independence, the G likelihood ratio test, and/or the mutual information. Third and finally, the true and estimated groups can be thought of as

two raters and the measure of rating quality is the inter-rater reliability. The known groups form one rater and the estimated groups constitute the other. A coefficient of inter-rater agreement beyond what would be expected by chance, Cohen's kappa (κ; Cohen, 1960) is the primary measure in this context. Because κ uses agreement as its goal, rather than independence, and for other reasons that will become clear as we proceed, κ was the main tool used to evaluate grouping recovery. However, because the other approaches have much to offer, they were also considered but in a more limited scope. All of the approaches listed previously will now be discussed in greater detail.

Multiclass Classification

Thought of in one way, the recovery of the person grouping structure is an issue of multiclass classification. The broad approach used for multiclass classification is to reduce the problem to a set of binary classifiers: either one-versus-rest, in which an N-way classification is transformed into N 2-way classifications, or one-versus-one, in which an N-way classification is turned into the $N(N-1)/2$ unique pairs of 2-way classifiers (see Fawcett, 2006, for an introduction). Once the multiclass problem is reduced to a set of binary classifications, standard binary classification can be applied. The goal with binary classification evaluation is to obtain metrics of the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) rates and to use combinations of these metrics to assess classifier performance.

As an example of multiclass classification with three classes, one might obtain the 3×3 confusion matrix below which lists the frequencies of various events. The number of times class A was chosen when class B was true is given by d, and so on. The set of equations below shows the transformation of the 3×3 confusion

matrix into 3 2×2 confusion matrices using the one-versus-rest approach to N-way classification.

$$\text{Picked} \left\{ \overbrace{\left(\begin{array}{c|ccc} & A & B & C \\ \hline A & a & d & g \\ B & b & e & h \\ C & c & f & i \end{array} \right)}^{\text{True}} \right. \xrightarrow{\text{Category A vs Not}} \begin{pmatrix} a & d+g \\ b+c & e+f+h+i \end{pmatrix} \quad (2.5)$$

$$\xrightarrow{\text{Category B vs Not}} \begin{pmatrix} e & b+h \\ d+f & a+c+g+i \end{pmatrix} \quad (2.6)$$

$$\xrightarrow{\text{Category C vs Not}} \begin{pmatrix} i & c+f \\ g+h & a+b+d+e \end{pmatrix} \quad (2.7)$$

In the first matrix (Equation 2.5), the contrast A-versus-Not-A was created. The number of True Positives is a; the number of False Positives is $d+g$; False Negatives $b+c$; and True Negatives $e+f+h+i$. In the second matrix (Equation 2.6), the contrast B-versus-Not-B was used with correspondingly different True Positives (e), False Positives ($b+h$), False Negatives ($d+f$), and True Negatives ($a+c+g+i$). The final matrix (Equation 2.7) is defined similarly as the C-versus-Not-C contrast. Hence, reduction of the 3×3 matrix to 3 2×2 matrices has created not one, but three True Positive counts. So, The next challenge is to collapse each 2×2 matrix down to one, global confusion matrix that describes the overall classifier performance.

Two basic methods can be taken to summarize the N 2-way classifications, micro-averaging and macro-averaging (e.g. Özgür, Özgür, & Gügör, 2005, p. 611). In micro-averaging, the N 2×2 tables are summed elementwise to create

a single 2×2 table which then gives the average counts for the True Positives, False Positives, False Negatives, and True Negatives. These, in turn, are used to create the various metrics of binary classifier performance. By contrast, in macro-averaging, one performance metric is computed for each of the N 2×2 tables, and then the mean is taken over the N performance metrics.

Micro- and macro-averaging can be illustrated with the 3×3 example and Equations 2.5 through 2.7. The micro-average for Accuracy would proceed as follows.

$$\begin{aligned} TP_{micro} &= \text{sum}(a,\ e,\ i) = a + e + i && (2.8) \\ FP_{micro} &= \text{sum}(d+g,\ b+h,\ c+f) = b + c + d + f + g + h && (2.9) \\ FN_{micro} &= \text{sum}(b+c,\ d+f,\ g+h) = FP_{micro} && (2.10) \\ TN_{micro} &= \text{sum}(e+f+h+i,\ a+c+g+i,\ a+b+d+e) && (2.11) \\ &= 2(a+e+i) + b + c + d + f + g + h && (2.12) \\ &= 2TP_{micro} + FP_{micro} && (2.13) \end{aligned}$$

Note that the micro-average False Positives across all classes must equal the micro-average of the False Negatives, creating a symmetric micro-average confusion matrix. The micro-average Accuracy is

$$Accuracy_{micro} = \frac{TP_{micro} + TN_{micro}}{TP_{micro} + FN_{micro} + FP_{micro} + TN_{micro}} \quad (2.14)$$

By contrast, the macro-average Accuracy is

$$Accuracy_{macro} = \frac{1}{N} \sum_i Accuracy_i = \frac{1}{N} \sum_i \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i} \quad (2.15)$$

In general, the micro-Accuracy and the macro-Accuracy are not equal. The micro-average performance metrics operate in a highly constrained space because the micro-average confusion matrix must be symmetric. This limits the utility of the micro-average metrics. Broadly speaking, the multiclass classification approach is somewhat limited by the lack of performance metrics based on the general N-class case, instead reducing the N-class case to N or $N(N-1)$ 2-class cases and summarizing. If only a single instance of a classifier were being evaluated, then the averaging may not be needed and each of the 2-way classifiers could be examined in detail. However, in the simulation study there was hundreds of thousands of N-way classifications, all of which must be evaluated. As such, the multiclass classification metrics were examined in general but not exclusively relied upon. Other approaches will now be discussed.

Contingency Table

Thought of in another way, the $N \times N$ contingency table (confusion matrix) representing the cross tabulation of the true and selected models can be tested for independence as any other contingency table with the Pearson χ^2 test of independence.

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{2.16}$$

where O_i is the observed count for cell i in the contingency table and E_i is the expected count based on the assumption of independence. Similarly, metrics with the baseline assumption of independence could also be used. For example, the mutual information of the "Picked" and "True" categorical random variables

could be assessed.

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \tag{2.17}$$

where the entropy, $H(X)$, of a discrete random variable X is defined as

$$H(X) = \sum_i p(x_i) \ln \left(p(x_i)\right) \tag{2.18}$$

in which $p(x_i)$ is the probability mass function of X. The related G-test likelihood ratio could also be computed to assess the independence of the "Picked" and "True" categorical random variables.

$$G = 2NMI(X,Y) = 2\sum_i O_i \ln \frac{O_i}{E_i} \tag{2.19}$$

These are all highly inter-related tests. The original χ^2 statistic (Pearson, 1900) is an approximation to the G-test likelihood ratio which, in turn, is related to the mutual information. The Pearson χ^2 formula (Pearson, 1900) can be derived as a second-order Taylor series approximation to the G test (Hoey, 2012). Because of the inter-relations of the χ^2, the mutual information, and the G-test, they all share the same weakness. In particular, the baseline for statistical comparison is independence. In evaluating classification performance, more is desired than the classifier that is not completely independent of the true underlying structure. Non-independence is too weak a criteria to evaluate classifier performance. Thus, although non-independence provides some evidence of good grouping structure recovery, another classification measure must be considered.

Inter-rater Reliability

A final, and perhaps most useful, conceptualization of the grouping recovery is as the inter-rater reliability of the True rater and the model selection rater. Considering the grouping recovery problem as one of inter-rater reliability would suggest the use of a metric such as Cohen's kappa (κ; Cohen, 1960).

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \tag{2.20}$$

where $P(a)$ is the probability of agreement between raters based on the contingency table, and $P(e)$ is the probability of agreement by chance. When the probability of agreement is equal to the probability of agreement by chance, κ is equal to zero. When agreement is perfect (and the probability of agreement by chance is not equal to one), κ is one. Thus, κ reflects a standardized inter-rater agreement beyond agreement by chance. The binary classification metrics (e.g. accuracy) are often on a similar scale, but are difficult to generalize to multiple classes. The χ^2, G, and mutual information indices are scaled so that complete agreement is anchored at one end with statistical significance indicating a departure from independence on the other end. However, they all have differing scales depending on the degrees of freedom and other factors, making them somewhat difficult to reconcile across simulation conditions. Coefficient κ extends nicely to multiple categories and is anchored by perfect agreement and chance agreement, thus it was the primary outcome by which group recovery structure was assessed. Because of its bounded scale between zero and one, logistic regression was used to predict Cohen's κ as a function of the simulation condition and their interactions.

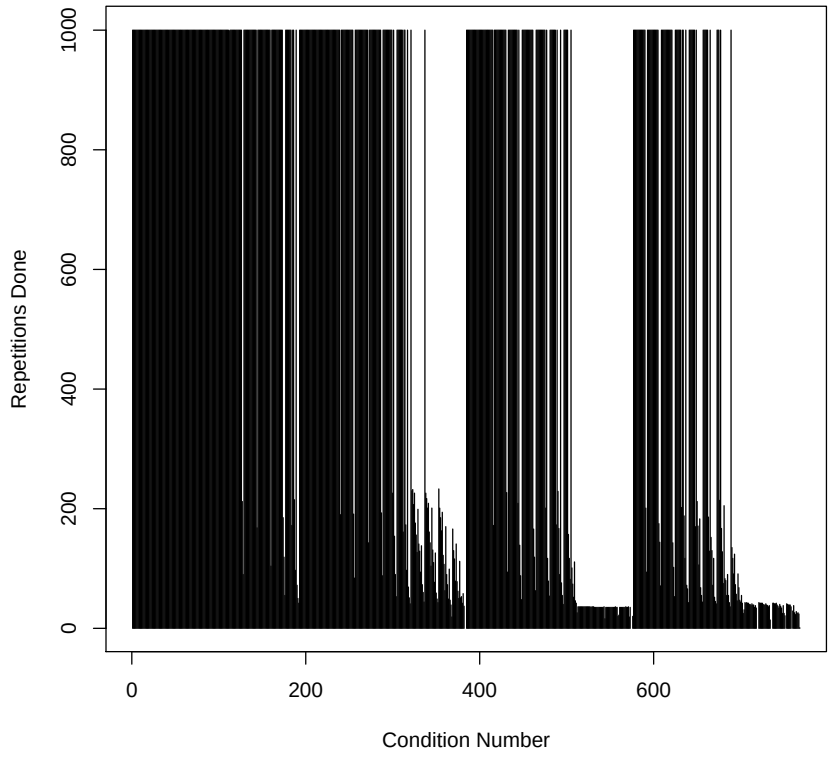
Dealing with Label Switching

The parameters of a mixture model are only uniquely determined up to a permutation of the classes: the *estimated* classes might be a permutation of the *true* classes. For example, if there are three classes (A, B, and C) in a univariate Gaussian mixture model with true means 10, 15, and 20, respectively, all with variance 1, then the estimated classes A, B, and C could have means 15.1, 19.4, and 10.7, respectively. Thus, the estimated classes A, B, and C align with the true classes B, C, and A. This permute-ability of the classes and parameters of a mixture model is the problem of label switching. An approach akin to that of Stephens (2000) was used to resolve the problem of label switching. The Bayesian posterior likelihood of each person being in each group conditional on each group's prior probability was used to determine group membership. Subsequently, group identity was determined by maximizing the true positive rate (sum of the diagonal of the confusion matrix) over the space of all possible permutations of group labeling.

2.3 Results

The results of the simulation study will now be described. The findings can be placed into three general categories. First, the findings regarding the simulation completion, model convergence, and estimation time are discussed. Second, the parameter estimation quality and some factors that influence it are described. Finally, the recovery of the underlying grouping structure, the primary outcome of interest, and influential factors for it are characterized.

Figure 2.5: Number of Replications Completed for Each Simulation Condition



2.3.1 Simulation Completion

Figure 2.5 shows the actual number of replications finished for each condition. The goal number of replications was not achieved for all conditions because of computational time to estimate certain kinds of models and software limitations in generating some sufficiently large models. The mean number of replications completed was 630.7, median was 1000. Ten conditions had zero replications. Thus the simulation generated 484,394 rows of data instead of 768,000 (61.6%). As mentioned previously, computational time was a major limiting factor. The simulation as conducted required 10.9 years of computer time just for model estimation, not including data generation or model building. This computer time

was distributed over a maximum of 256 nodes on the OU OSCER supercomputing system for several weeks. A conservative estimate of the amount of time required to fill in the missing 283,606 replications suggests 388 additional years of computer time would be needed because the remaining conditions are large and computationally intensive.

Figure 2.6 depicts the median model run times across all conditions. The run times are shown in log units. There is a clear effect on the run time around Condition 200, 400, and 600. This is the effect of number of groups. For the first 192 conditions, there is only 1 group. For conditions 193 to 384, there are 3 groups; and so on. There is also a relatively clear effect of the number of people per group (1, 10, or 100). The break points for 1, 10, and 100 people per group are noted on the graph when the number of groups is 5. The lower bound for the median runtime among the conditions in this simulation is about .14 seconds (e^{-2}), and the upper bound was about 48 hours ($log(48 * 60 * 60) \approx 12$). The upper bound was determined by the limitations of the supercomputer which only allowed a job to be run for at most 48 hours.

The convergence of models across conditions is illustrated by Figure 2.7. A status code of zero, indicated by a black line in the figure, means that model estimation finished normally: the optimizer found a flat spot in the parameter space and sequence of parameter estimates from one iteration to the next steadily converged on the solution. For almost all conditions, the most frequent status code was zero, indicating that model estimation was generally successful. The next most frequent status code was one, indicated by a green line in the figure. A status code of one means that the optimizer found a flat spot in the likelihood space, but it happened rather suddenly so the sequence of parameter estimated from one iteration to the next did not steadily converge on the solution. Sta-

Figure 2.6: Median wall clock running time of models in each condition. Time is in log units. Breaks between different numbers of groups in the model and number of people per group are noted in the figure.

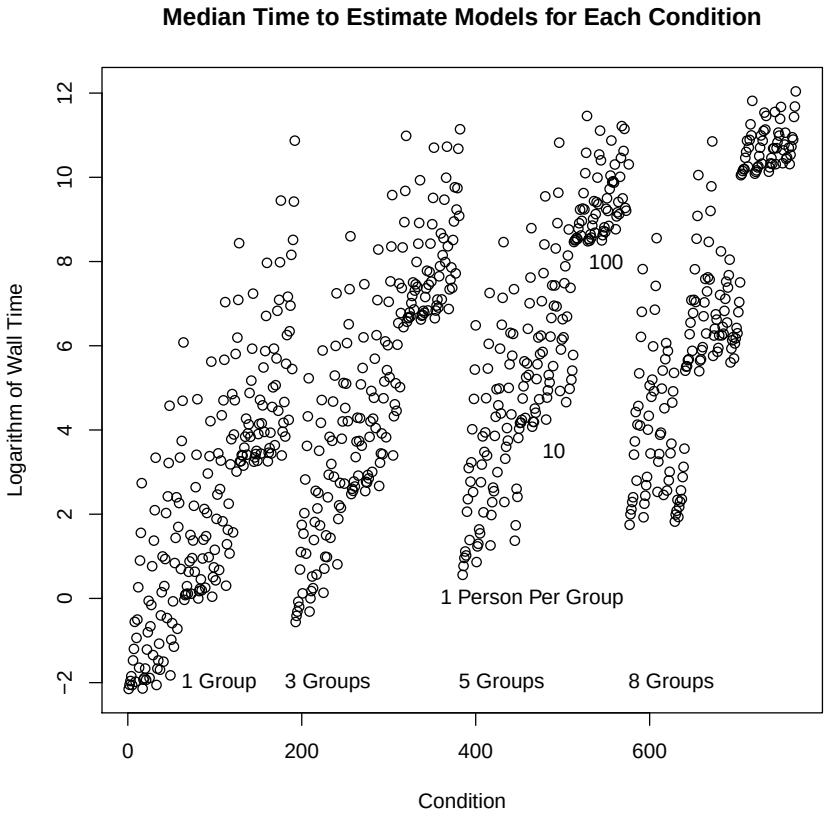
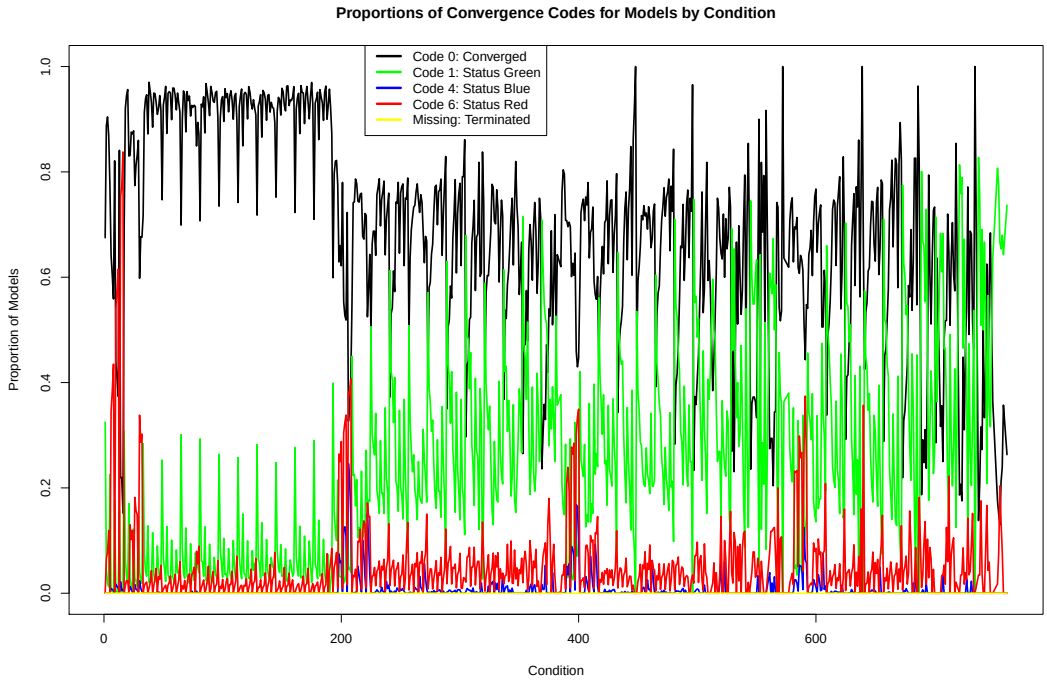


Figure 2.7: Proportion of convergence status codes given by the optimizer for each condition. Status Code 1 (Green) = "The final iterate satisfies the optimality conditions to the accuracy requested, but the sequence of iterates has not yet converged. NPSOL was terminated because no further improvement could be made in the merit function". Status Code 4 (Blue) = "The major iteration limit was reached". Status Code 6 (Red) = "The model does not satisfy the first-order optimality conditions to the required accuracy, and no improved point for the merit function could be found during the final linesearch".



tus code one generally does not indicate an estimation problem. On the other hand, status code six, shown as a red line in the figure, does indicate a serious optimization problem. Status code six means that a flat spot could not be found in the likelihood space and the optimizer could not find a better solution. Most often, status code six is resolved by picking different and better starting values. The blue line shows status code 4: the iteration limit was reached. Finally, the yellow line shows that virtually no models were terminated.

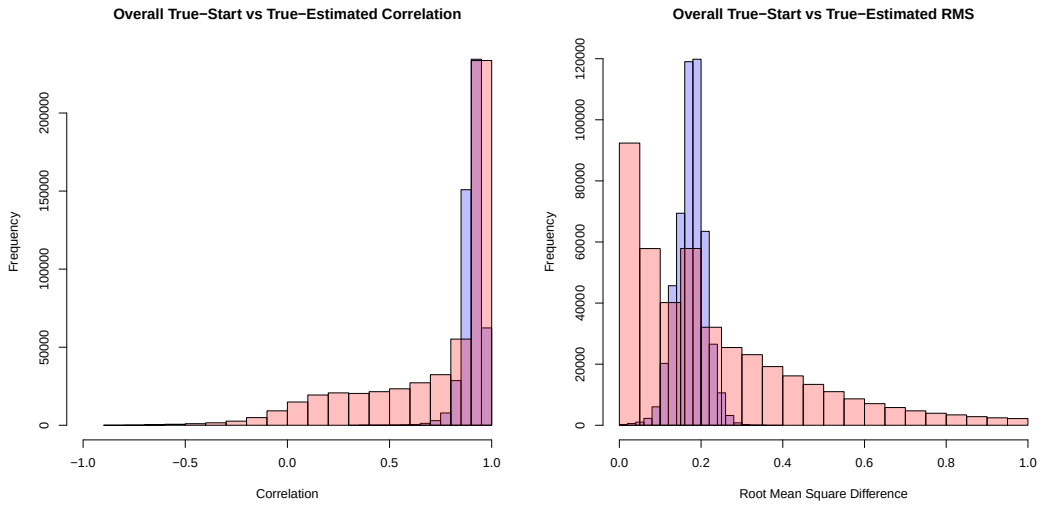
Figure 2.7 has several features that are worth discussing. First, note that

conditions 1 through 192 contain only one group, and hence are not state space mixture models at all; they are just state space models. The non-mixture models are used simply as a comparison for the mixture models. It is relatively evident that mixture models converge less frequently than non-mixture state space models. However, the highest proportion of non-convergent models have status code 1 (Green) which generally speaking does not indicate a problem. Second, the status code 6's appear to spike around condition 1, condition 200, condition 400, and condition 600. These conditions correspond to the number of people per group being one and the number of occasions being five. The status code 4's (iteration limit reached) spike for the same conditions. Thus, the smaller sample conditions have more estimation problems. Be that as it may, eighty to ninety percent of models across all conditions had status zero or one. So, overall it appears that model estimation and convergence was not terribly problematic. We now consider the parameter estimation.

2.3.2 Parameter Estimation

As mentioned previously, for each repetition of each condition the starting values, estimated values, and true values of the parameters were recorded. Moreover, the true values vary from one repetition to the next because they were randomly generated. Therefore, the true and estimated values for the parameters are compared within each repetition, creating one comparison for each row of simulation data. Hence, there is a distribution of True-Estimated comparisons. Similarly, there is a distribution of True-Starting comparisons that provides the null relation between the true parameter values and the randomly generated starting values. Two tools were used to make comparisons between parameter sets: the correlation and the

Figure 2.8: Overall comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.



root mean square (RMS) difference.

Figures 2.8 through 2.12 examine distribution of the True-Estimated correlations and RMS values in different parameter subsets using the True-Start distribution as a null for comparison. Figure 2.8 makes the overall comparison using all of the parameters (i.e. the dynamics parameters, factor loadings, and residual variances). The correlation and RMS both indicate that the most frequent True-Estimated values provide better fit than many of the True-Start values. There are, however many instances in which the estimated parameters are farther away from the true values than their starting values. This unfortunate circumstance might be driven by the smaller sample size conditions, where the sampling error makes the best estimates of the parameters for the data somewhat far from their population values because the random sample is not particularly close to the population.

Figure 2.9: For only the dynamics parameters, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

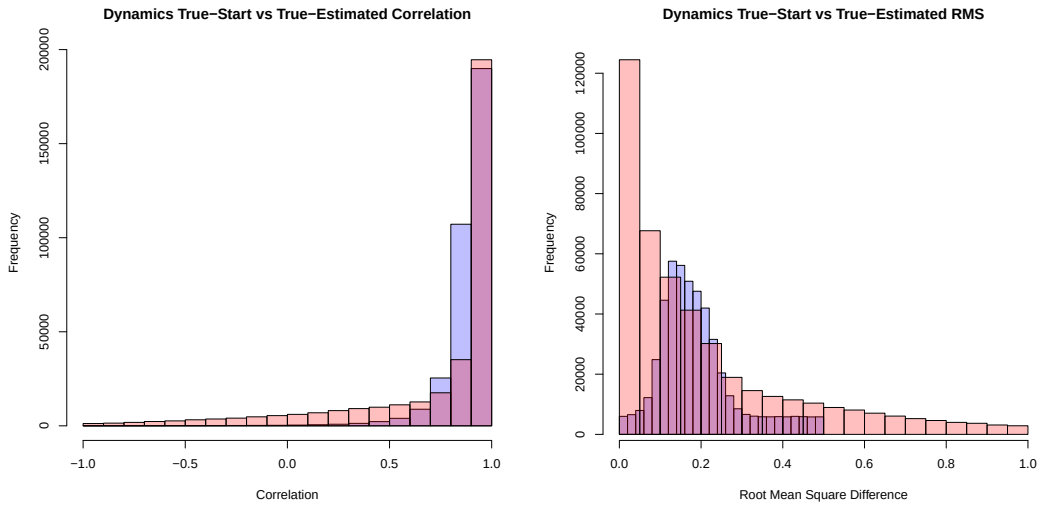
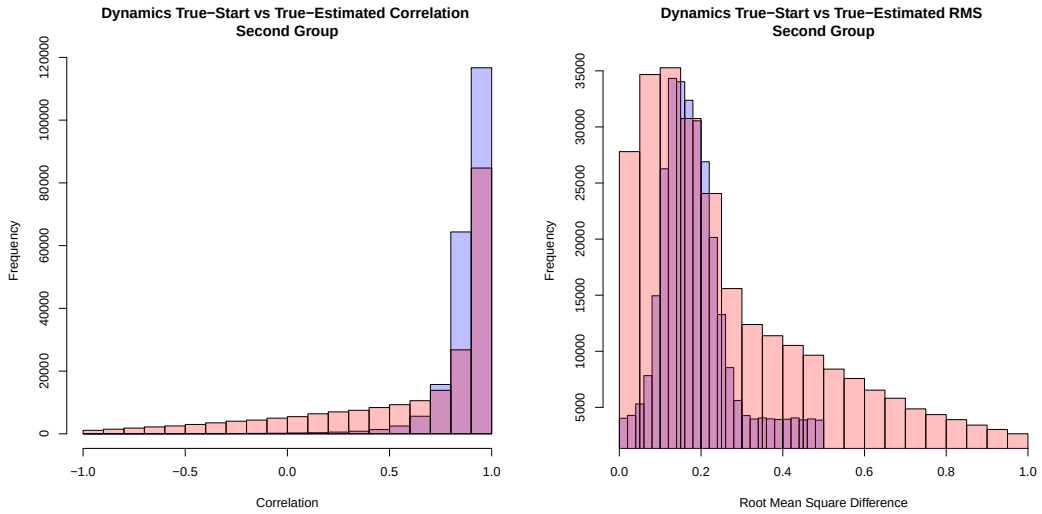


Figure 2.9 makes a similar comparison to Figure 2.8, but restricts the comparison to only the dynamics parameters. The dynamics parameters determine how each person operates over time. In general, the dynamics parameters resemble the overall parameters. Because the dynamics vary from one group to the next, there are sets of dynamics parameters for each group. Figure 2.9 showed the first group dynamics parameters that includes the case in which there was only one group, and hence a non-mixture state space model. By contrast, Figure 2.10 shows the second group dynamics matrix. Only cases of state space mixture models are included here. The comparisons for groups three through eight are similar to group two. It appears that the dynamics parameters were easier to recover when there was no mixture model. This replicates the findings of Figure 2.7 with the optimizer convergence and status codes.

Figure 2.10: For only the dynamics parameters in the second group, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.



Figures 2.11 and 2.12 restrict the parameter comparison to the factor loadings and residual variances, respectively. These parameters are recovered much better than the dynamics parameter. Also, the null distributions of the True-Start correlations for the factor loadings and residual variances look qualitatively different from those of the dynamics parameters. But the distribution of the True-Estimated correlations is relatively similar for the factor loading and dynamics parameters. Recall that the starting values for the dynamics were generated by taking the true values and adding a uniform random variable distributed as $\mathcal{U}(-.5, .5)$. It is possible that the less favorable True-Start versus True-Estimated comparison for the dynamics parameters is due to this different procedure for generating starting values for the dynamics matrices.

None of Figures 2.8 through 2.12 consider the differential estimation quality

Figure 2.11: For only the factor loadings, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

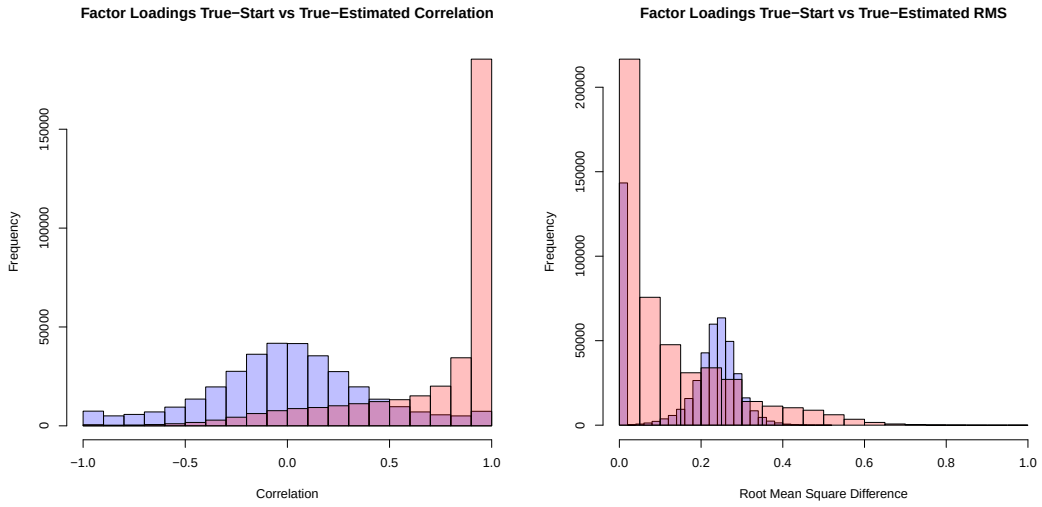
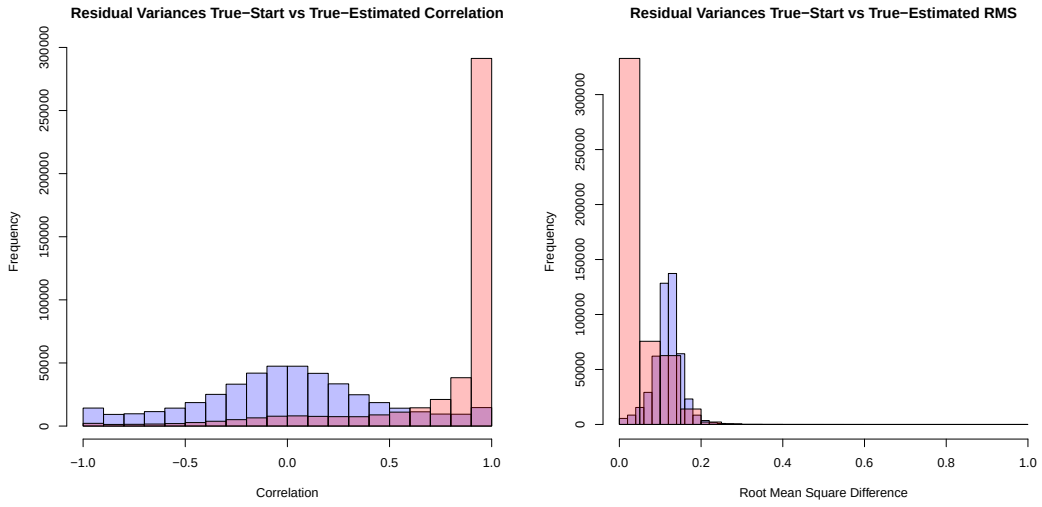


Figure 2.12: For only the residual variances, comparison of the distributions of True-Start parameter correlations and the True-Estimated parameter correlations. Similarly, the Root Mean Square (RMS) comparison is made. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.



for the various simulation conditions. The marginal effect of each simulation condition is considered in Figures 2.13 through 2.17. From these figures, the factors that influence parameter estimation can be seen. In Figure 2.13, the overall parameter estimation quality is shown in terms for root mean square difference for each level of "Number of Variables Per Factor": 1, 3, 4, and 6. There is not an obvious, strong influence of the number of variables per factor other than for one variable versus more than one. Dissimilarly, Figure 2.14 shows that as the number of factors increases the overall parameter estimation improves. This may be due to the added constraints in the higher dimensional dynamics matrices. All zero entries in the dynamics matrices were assumed fixed and known, so higher dimensional spaces became easier to estimate because of the added information about the zero parameters.

Figure 2.15 characterizes the effect of the number of occasions on parameter recovery. This is the first sample size effect considered. As the number of occasions increases there is a clear and strong improvement in the overall parameter estimation. The persistent spike around the null distribution RMS (about 0.2 in the graph) even for cases of 200 occasions of measurement indicates that other factors are also critical for quality parameter recovery. One of those other factors is the second sample size considered, the number of people within each group. The effect of number of people per group is detailed in Figure 2.16. Again, as the number of homogeneous replicates of the time series increases, the model estimation improves. Finally, as the number of groups increases the parameter estimation quality declines (Figure 2.17). The one-group case is a non-mixture state space model and has the best parameter estimation. Each higher level of number of groups decrements the parameter estimation.

Broadly speaking, the parameters for state space mixture models can be esti-

Figure 2.13: Effect of Number of Variables Per Factor on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

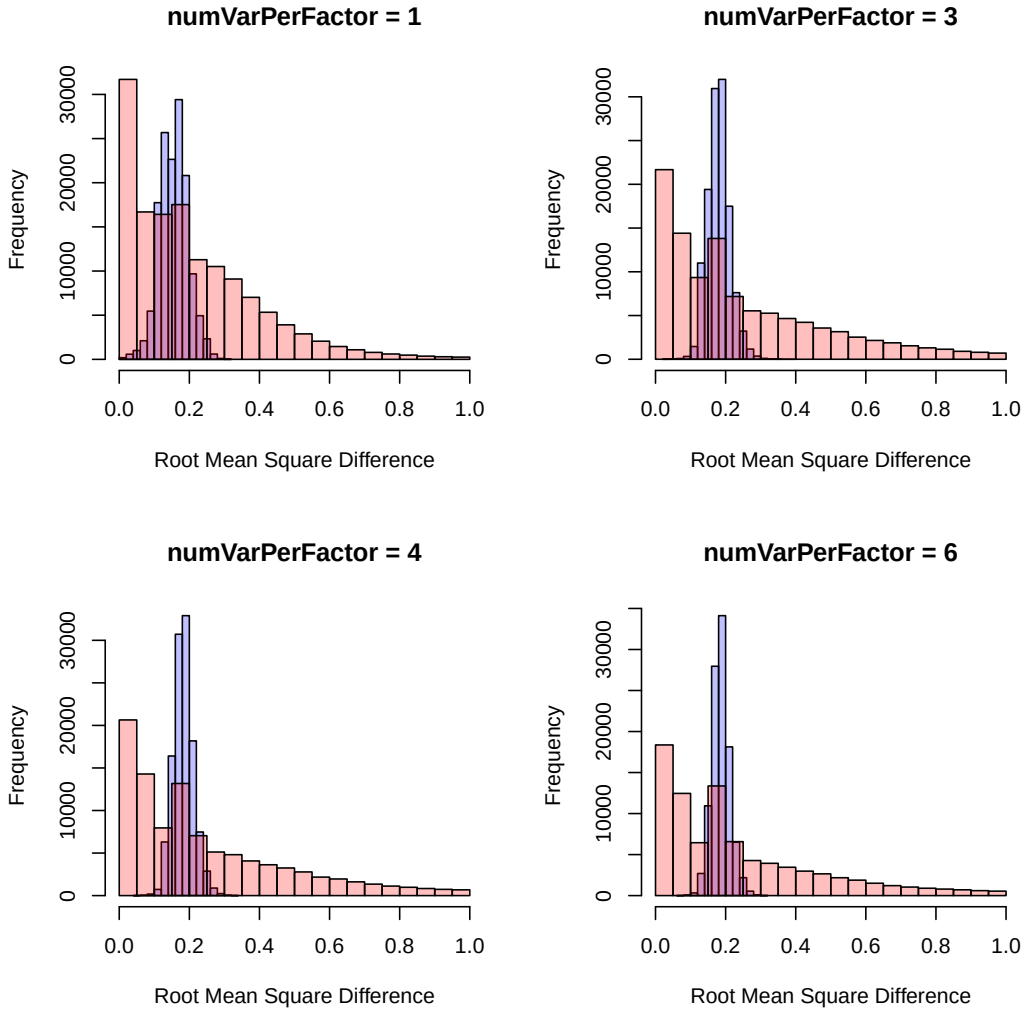


Figure 2.14: Effect of Number of Factors on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

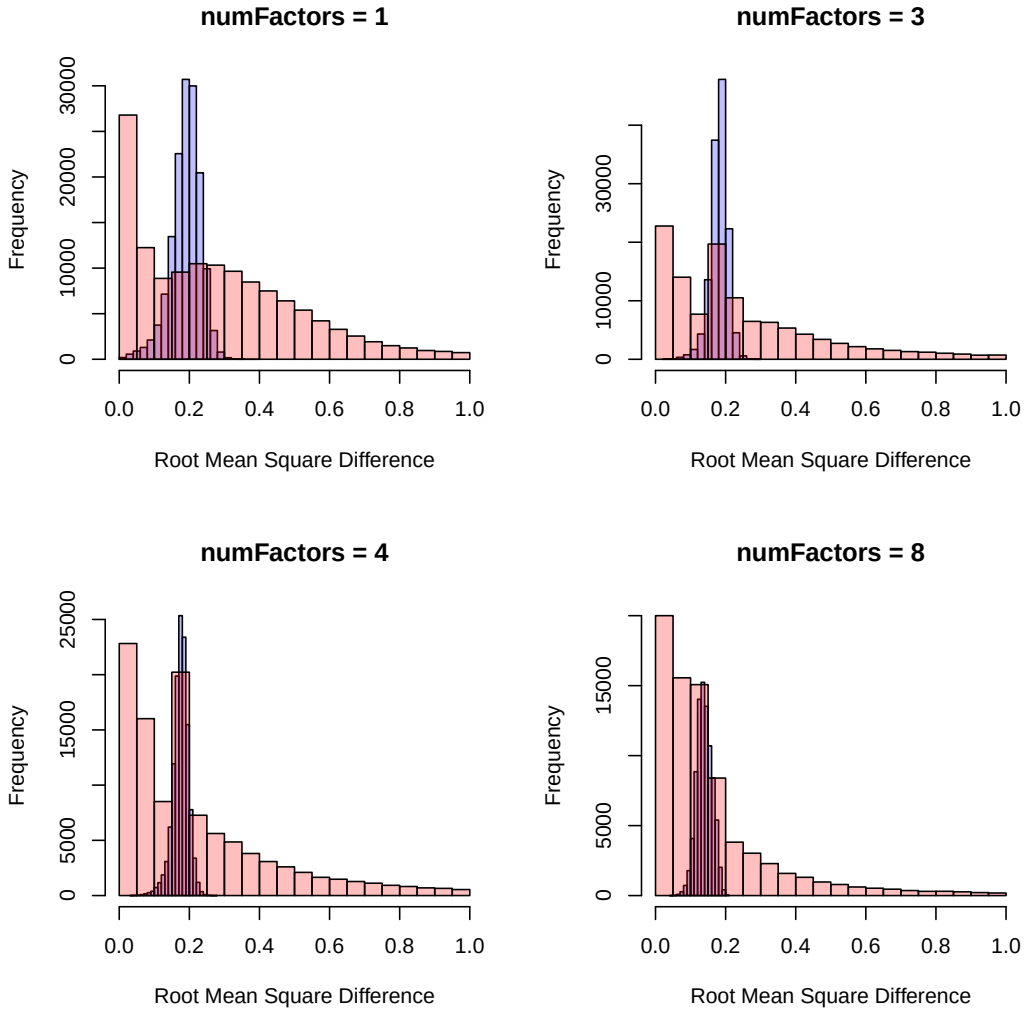


Figure 2.15: Effect of Number of Occasions on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

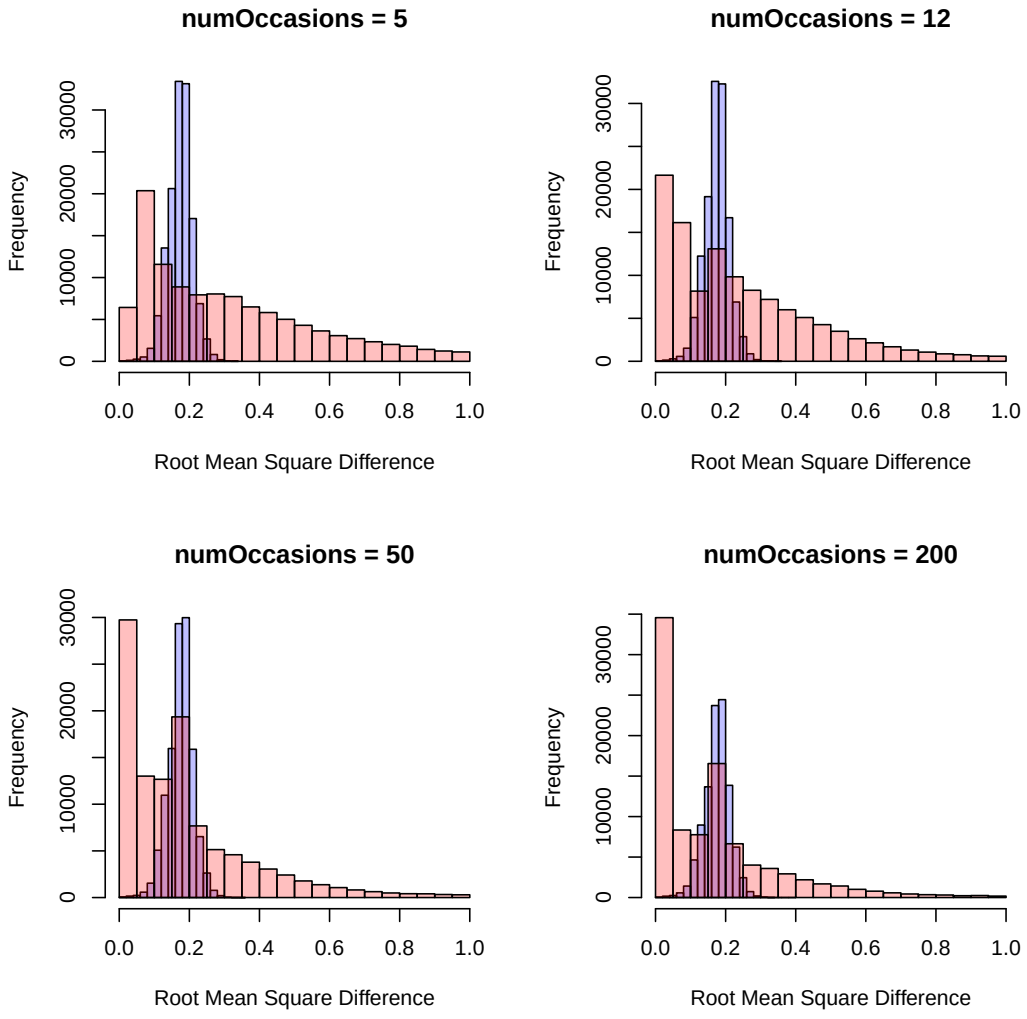


Figure 2.16: Effect of Number of People Per Group on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.

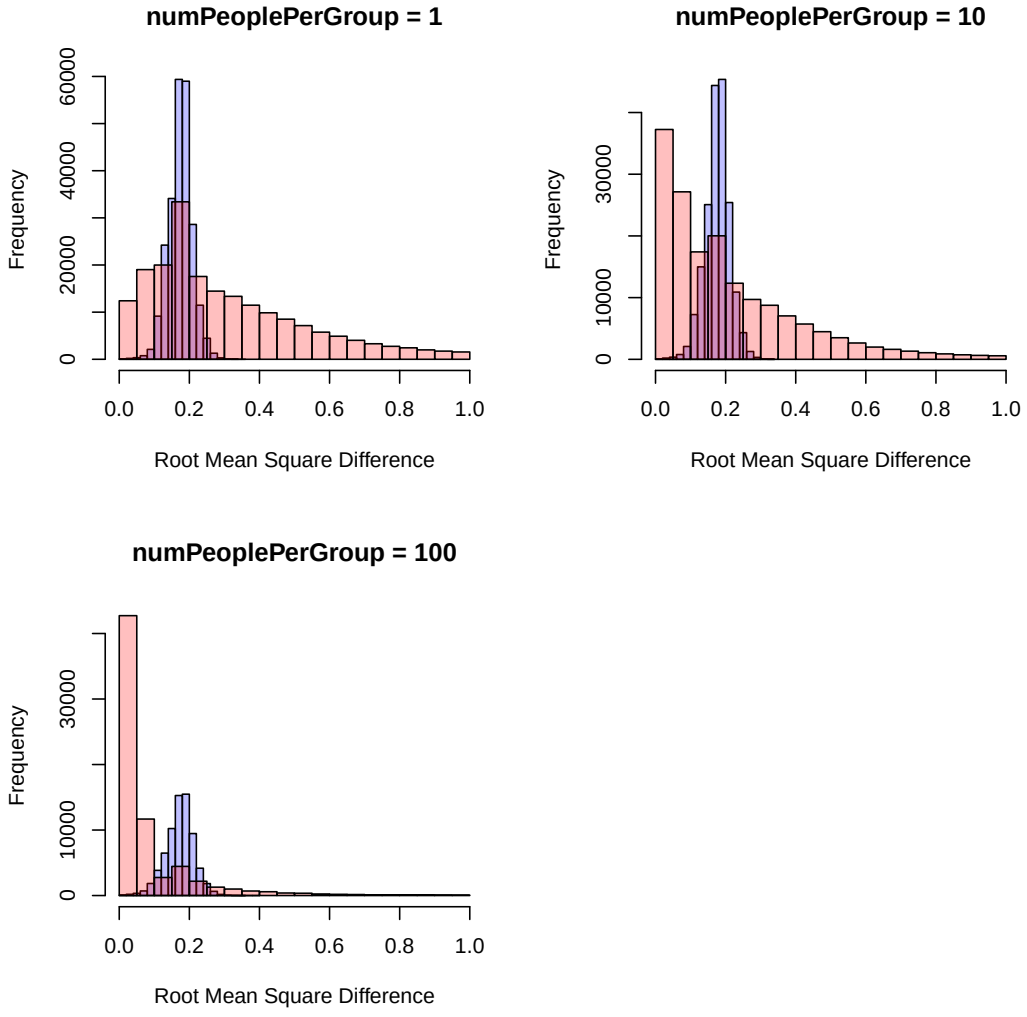
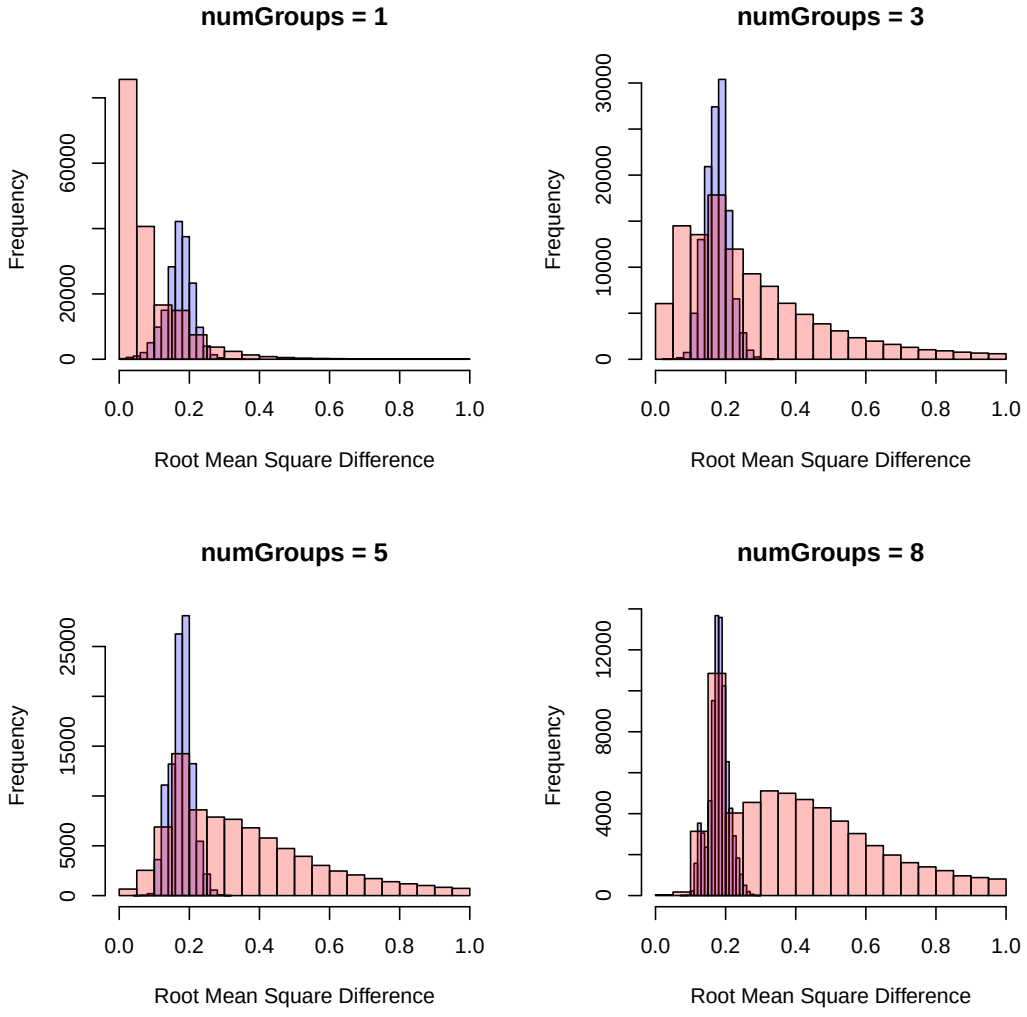


Figure 2.17: Effect of Number of Groups on parameter recovery: comparison of the distributions of True-Start parameter Root Mean Square (RMS) difference and the True-Estimated parameter RMS. The True-Start relationship distribution is in transparent Blue. The True-Estimated relationship distribution is in transparent Red. The regions of overlap become Purple.



mated reasonably accurately. The measurement model parameters (factor loadings and residual variances) seem easier to recover than the dynamics parameters. Based on the simulation findings, a recipe for accurate parameter estimation would include at least three measured variables for each factor, as many constraints on the dynamics as possible, measuring people as many times as you can, getting as many homogeneous people as you have available, and using as few mixture classes as can be justified. With that in mind, we now turn to the issue of grouping structure recovery: the proper assignment, based on the imperfectly estimated model, of people to their generating groups.

2.3.3 Grouping Structure Recovery

A key outcome of the simulation study is the proper assignment of people to their generating model. Given imperfectly estimated parameters and data with sampling error, it is a non-trivial task to correctly classify individuals into groups. As discussed previously, the problem can be evaluated as a multiclass classification using standard metrics from that area of research. Figure 2.18 shows the group recovery in terms of the micro-averaged True Positive and False Positive rates. The space of True Positive and False Positive rates is often called Receiver Operating Characteristic (ROC) space. Completely chance performance of the classifier occurs when the true positive rate is equal to the false positive rate: that is, the chance of making a correct assignment to a generating model is equal to the chance of making an incorrect assignment. Perfect classification occurs in the upper left-hand corner where the True Positive rate is 1.0 and the False Positive rate is 0.0. Figure 2.18 shows the effect of number of groups on classification in ROC space. Critically, it shows every individual row of the simulation data with

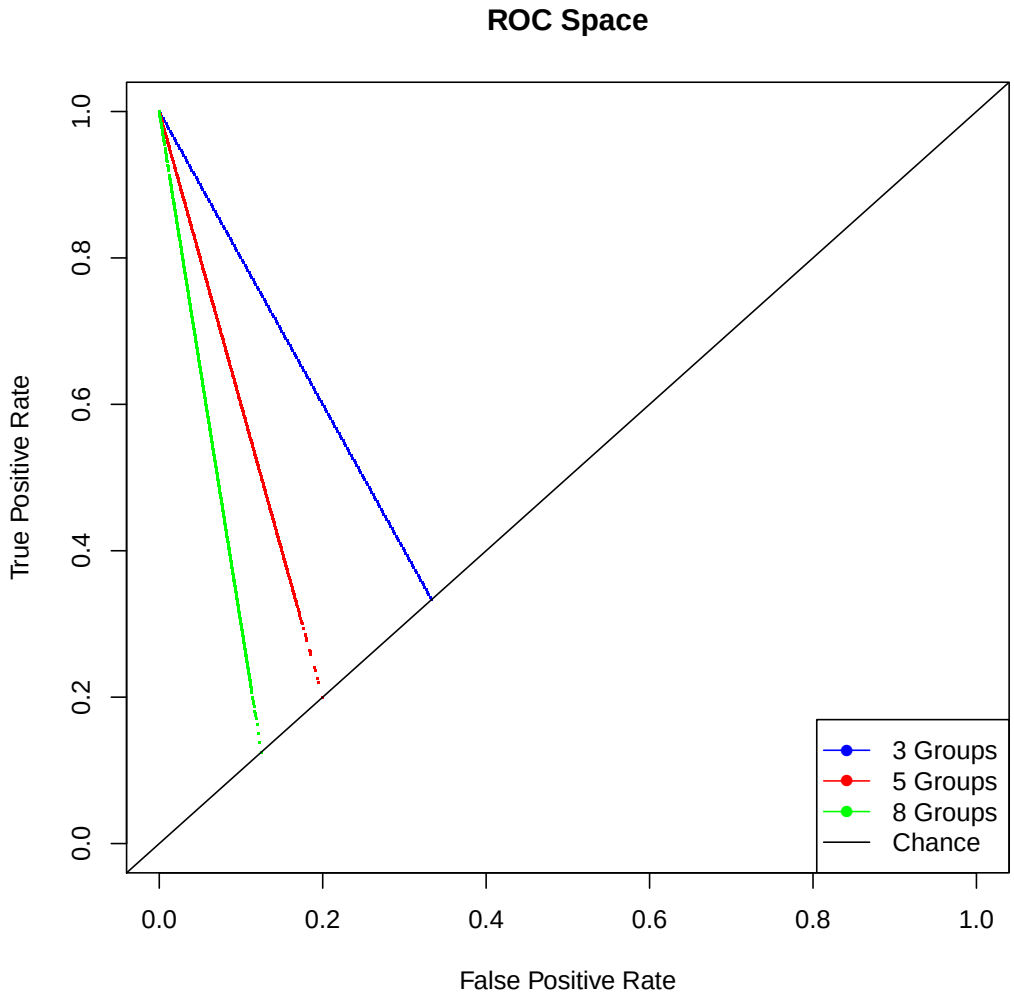


Figure 2.18: Receiver Operating Characteristic Space

no averaging. The obvious separation between the number of groups is due to the micro-averaging that reduces $N \times N$ classification into N 2×2 classifications. The induced pattern is a limitation of these findings.

The micro-accuracy, shown in Figure 2.19, does not have the same limitation as the ROC space. The overall mean accuracy across all conditions was 0.91. However, this includes cases of trivially correct classification. When there is only one group, classification is necessarily correct and accuracy is 1.0. Figure 2.20

corrects for this by excluding all cases of where the number of groups was one. There are still over 50,000 cases in which 100 percent accuracy was achieved. The mean dropped from 0.91 to 0.86. Overall, this implies high accuracy in classification. A similar picture is painted by Figure 2.21. Importantly, the scale of kappa differs slightly from that of accuracy. The accuracy metric has no demarcation of what chance accuracy is for a given classification problem, whereas kappa does. Chance agreement between raters, in terms of kappa, is indicated by zero. So, the mean kappa of 0.58 does not mean that agreement was near chance by this metric. On the contrary, the mean kappa says that although agreement was not perfect, it was far above chance.

To see the effect that the simulation conditions had on kappa and accuracy, Table 2.1 shows the marginal means for these variables on each level of the the simulation factors. All the simulation factors appear to have some effect on kappa and accuracy, however, these are not always in the same direction. The number of groups seems to hurt kappa, but improve accuracy. This could be due to the difference in how kappa and accuracy treat chance agreement. The number of people per group has little effect on either outcome, but recall this factor had a large influence on parameter recovery (see Figure 2.16). The number of occasions and number of variables per factor have influences on classification that agree across kappa and accuracy. The number of occasions improves both, whereas the number of variables per factor decreases both. The single largest simulation parameter that influence grouping structure recovery was the number of factors. When the number of factors was 1, the mean kappa was only 0.43, but when the number of factors was 8, the mean kappa was 0.73, the highest mean kappa across any of the marginal conditions investigated. As discussed in the section on parameter estimation, the increase in number of factors bears with

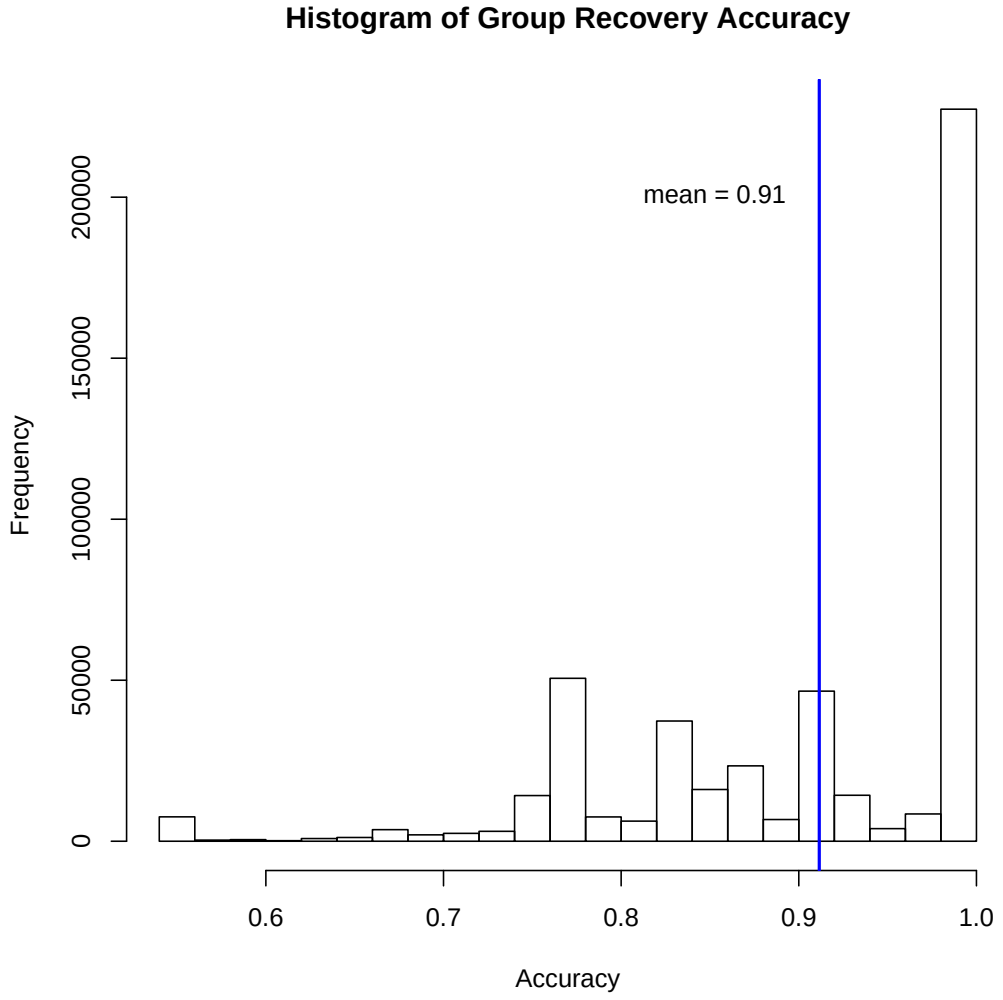


Figure 2.19: Group micro-average accuracy for all numbers of groups including non-mixture models with perfect accuracy. The blue line indicates the mean micro-accuracy.

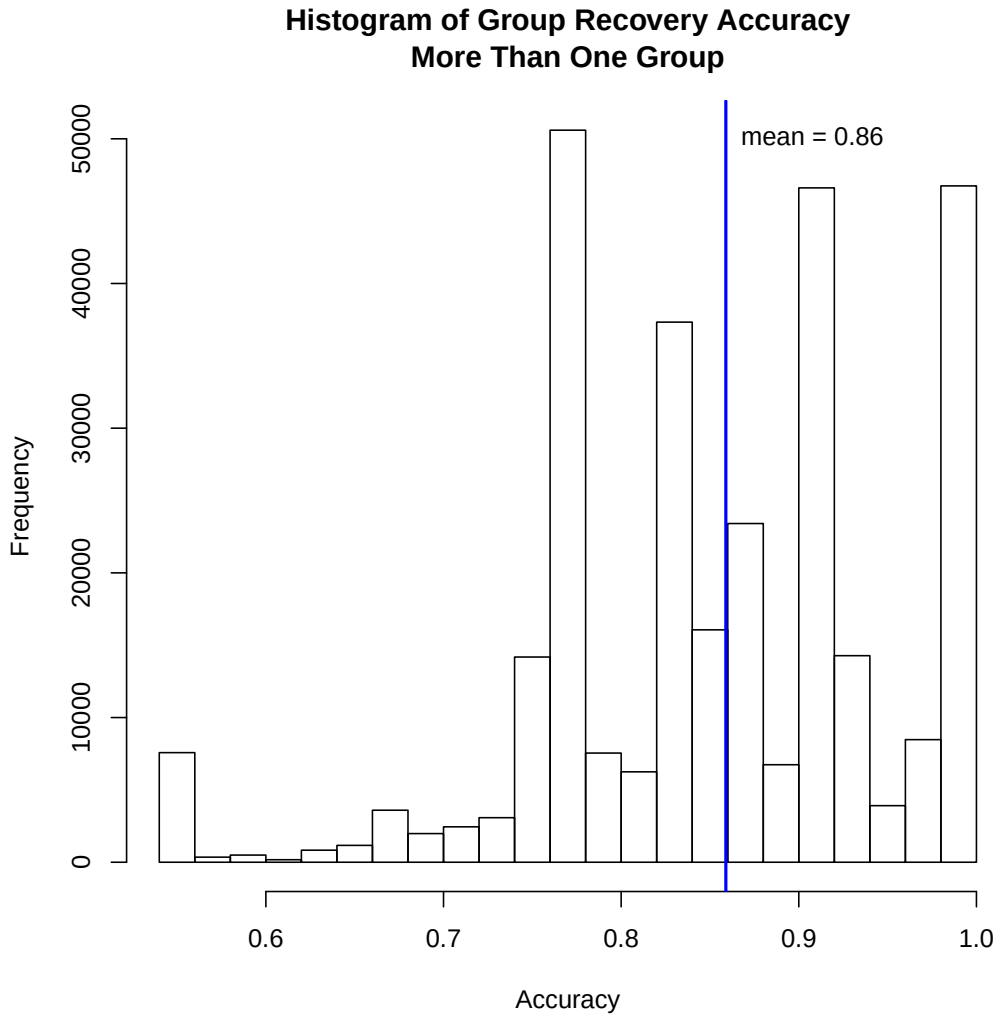


Figure 2.20: Group micro-average accuracy for numbers of groups greater than one, that is, excluding non-mixture models with perfect accuracy. The blue line indicates the mean micro-accuracy

Figure 2.21: Cohen's Kapps for numbers of groups greater than one, that is, excluding non-mixture models with perfect accuracy. The blue line indicates the mean Kappa

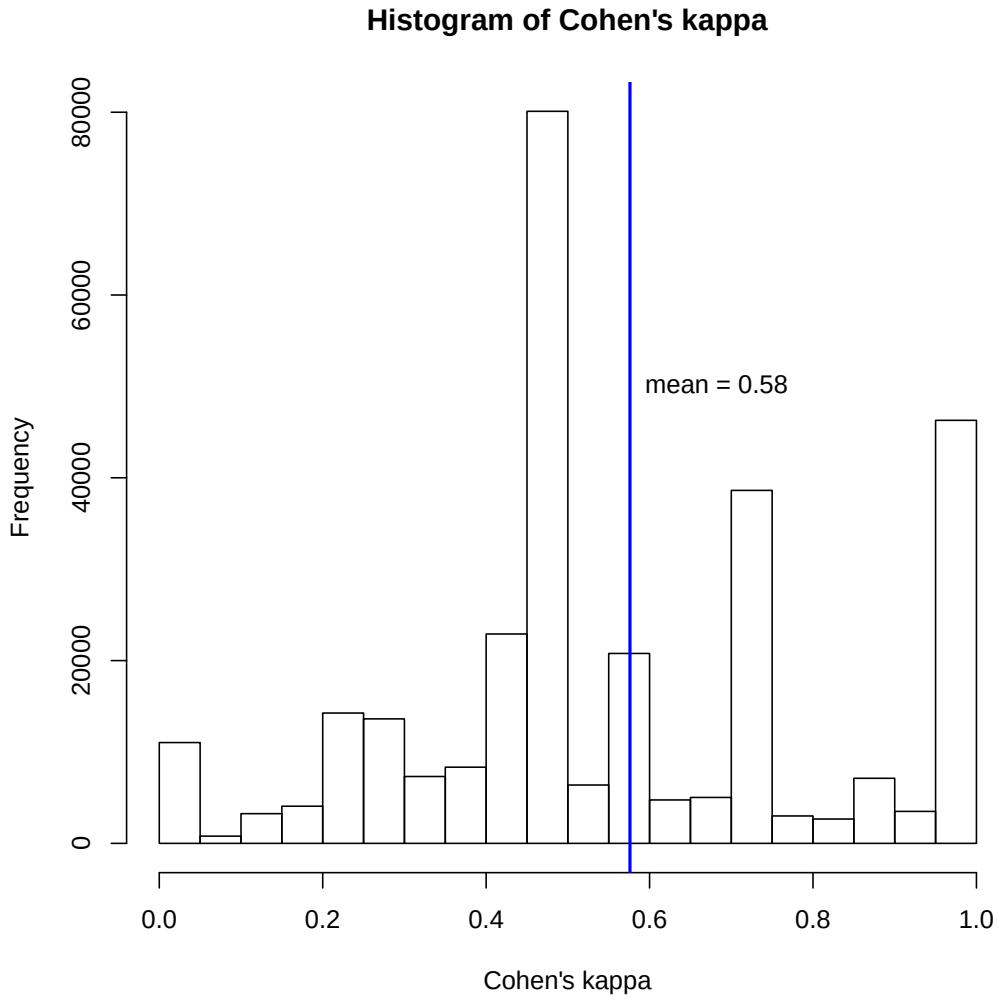


Table 2.1: Marginal Mean Kappa and Accuracy Values Across Simulation Factors

Simulation Factor	Level	Kappa	Accuracy
Number of Groups	3	0.63	0.84
	5	0.57	0.86
	8	0.50	0.89
Number of People Per Group	1	0.58	0.86
	10	0.57	0.86
	100	0.59	0.84
Number of Occasions	5	0.52	0.84
	12	0.56	0.86
	50	0.61	0.87
	200	0.63	0.87
Number of Factors	1	0.43	0.81
	3	0.59	0.87
	4	0.65	0.88
	8	0.73	0.90
Number of Variables Per Factor	1	0.63	0.88
	3	0.57	0.86
	4	0.56	0.85
	6	0.52	0.84

it an increase in the distinctness of the dynamics. In 1-dimensional space, there are few possibilities for how dynamical systems can behave, but in 8-dimensional space many patterns are possible that are easily separable. The finding that the number of factors so greatly influences groups recovery is preliminary evidence that people are easier to classify correctly when their dynamics are distinct.

To further investigate and quantify how the the simulation factors impacted group recovery, logistic regressions were conducted with the inter-rater reliability Cohen's kappa as the outcome. A summary of the main effects model is presented in Table 2.2. Similar to the results found in Table 2.1, the logistic regression suggests that the number of groups negatively impacts classification quality, as does the number of variables per factor to a lesser extent. The sample size factors, number of people per group and number of occasions, aid in determining group

Table 2.2: Logistic Regression with Kappa as Outcome

	Estimate	Std. Error	z value	Odds Ratio
(Intercept)	-0.58	0.0030	-194.5	0.56
numGroups5	-0.34	0.0016	-208.2	0.71
numGroups8	-0.62	0.0015	-405.9	0.54
numPeoplePerGroup10	0.05	0.0025	20.7	1.05
numPeoplePerGroup100	0.12	0.0025	47.4	1.13
numOccasions12	0.30	0.0016	187.9	1.35
numOccasions50	0.54	0.0017	312.4	1.71
numOccasions200	0.69	0.0019	357.3	2.00
numFactors3	1.05	0.0015	688.5	2.85
numFactors4	1.42	0.0017	845.5	4.14
numFactors8	2.02	0.0024	826.4	7.53
numVarPerFactor3	-0.09	0.0016	-56.8	0.91
numVarPerFactor4	-0.10	0.0017	-58.5	0.91
numVarPerFactor6	-0.17	0.0018	-96.0	0.84

Note: Estimate is in log odds units. Odds ratio is $e^{Estimate}$. All p-values are less than 1×10^{-10}. The Intercept term has numGroups=3, numPeoplePerGroup=1, numOccasions=5, numFactors=1, numVarPerFactor=1. All coefficients are contrasted against this Intercept condition. The numGroups=1 condition is omitted because it always has perfect classification.

membership. Again, the largest influence is found for the number of factors. Examination of the z values and odds ratios is particularly informative. For example, when the number of factors is eight and leaving all other simulation parameters constant, perfect classification is 7.53 times more likely than when the number of factors is one. The odds ratio for the lowest factor contrast (one factor versus three factors) is still higher than the next largest effect (5 occasions versus 200 occasions). This implies that if group recovery is the goal, then measuring an additional latent dimension is far more beneficial than measuring addition time points, at least when some relations are known between the additional latent variable and the other latent variables.

Further models were tested that included two-way and three-way interactions with similar results. Given the sample size and the number of effects tested, sta-

tistical significant is not the most useful criterion. All of the effects are significant and there are a large number of them. Consequently, it makes sense to consider the distribution of effects as compared to a similarly constructed normal distribution of effects. Parameters that diverge from normality can then be found and considered separately. A plot of the z-values of the two-way interaction model versus the z-values of a standard normal distribution is shown in Figure 2.22. Departures from normality can be seen as departures from the diagonal line. It is evident from this figure that several effects diverge from normality. The departure is such that the tails of the two-way interaction effects distribution are too heavy. That is, there are more very large positive and very large negative effects than a normal distribution implies. The coefficients that involve the number of factors are highlighted in red. The vast majority of effects that depart from the line involve the number of factors. The two, large, positive effects in Figure 2.22 that do not involve number of factors are for the main effects of number of occasions being 50 or 200 with the effect of 50 being the lower of the two.

A similar figure for three-way interactions was constructed (Figure 2.23) with similar result. The larger effects tend to involve the number of factors, and to a lesser extent the number of occasions.

To summarize, the results of the simulation study can be placed into three categories: model estimation time/convergence, parameter estimation, and group membership identification. With regard to model estimation, the number of groups and number of people per group were the largest influences. Models with more groups and/or more people per group took much longer to estimate. Furthermore, these longer running models had more convergence issues, yet most models, even in the larger conditions, did not have overly problematic convergence. For the parameter estimation, the measurement model parameters seemed

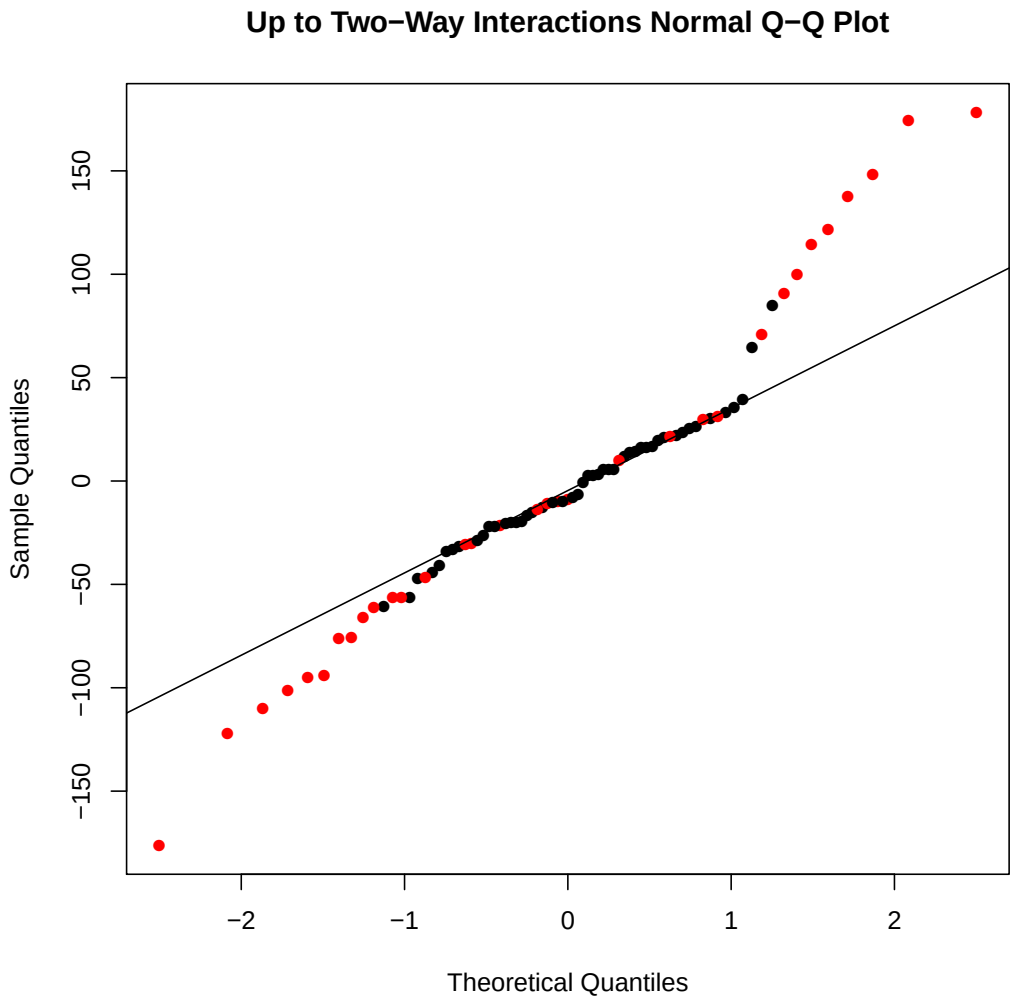


Figure 2.22: Quantile-Quantile Plot of the z values of all the main effects and two-way interactions of the logistic regression with kappa as the outcome and the simulation factors as the predictors. Effects involving the number of factors are colored red.

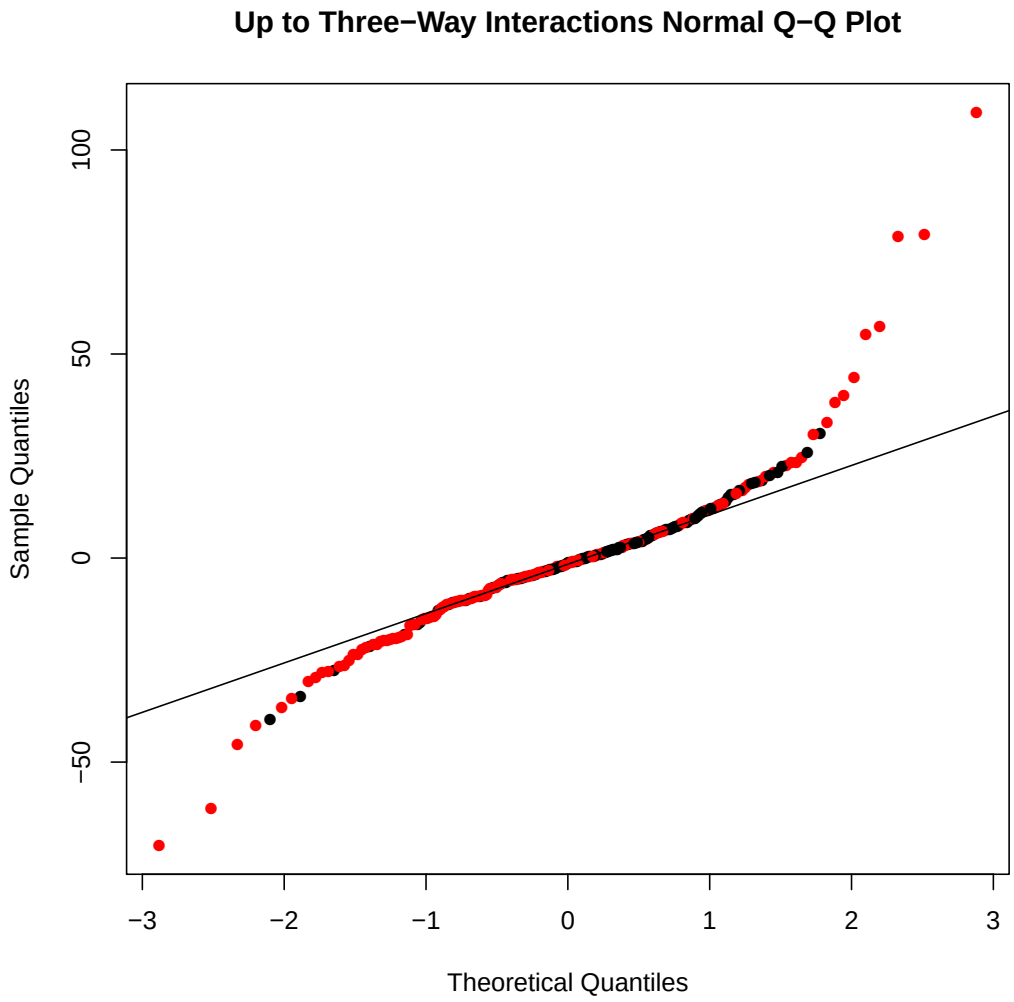


Figure 2.23: Quantile-Quantile Plot of the z values of all the main effects, two-way interactions, and three-way interactions of the logistic regression with kappa as the outcome and the simulation factors as the predictors. Effects involving the number of factors are colored red.

easier to recover than the dynamics parameters. This could be due to the starting value distribution for the dynamics parameters being comparatively closer to the true values than the starting values for the measurement model parameters. The dynamics parameters appeared to be estimated adequately, but the estimated values were not as much of an improvement over the starting values as was the case for the measurement parameters. The sample size conditions, number of people per group and the number of occasions generally improved parameter estimation. By contrast, the number of groups (i.e. number of mixtures) made parameter estimation less accurate. Finally for the group membership identification, state space mixture modeling was able to correctly classify individuals with similar measurement processes but different dynamics. Classification had 86 percent accuracy across all conditions and a mean inter-rater reliability of 0.58. Critically, it appeared that the dynamics were the key factor in correctly classifying individuals. The dimension of the state space (i.e. the number of factors) had the single largest impact on correctly identifying the correct generating model for each individual. Inter-rater reliability improved greatly as the number of factors increased from one to eight. Second to the number of factors, the number of occasions had the next largest effect. More occasions of measurement provide more temporal examples of the dynamics, allowing them to be more easily distinguished from dissimilar dynamics patterns. Thus, the dimension of the space in which dynamics unfold and the number temporal observations made in that space are the most important factors for correctly classifying individuals into their generating dynamic models.

2.4 Discussion

This chapter presented a simulation study of state space mixture modeling. The purpose of state space mixture modeling is to bridge the gap between the completely idiographic modeling of each individual separately and the completely nomothetic modeling of all individuals as a single homogeneous group. The simulation study generated 384,394 instances of state space models, 303,814 of which where state space mixture models. The state space mixture models had identical measurement models across groups, but differed in their dynamics. Thus, the primary task use-case for the state space mixture models evaluated was the separation of unknown group who had similar measurement processes but different underlying dynamics. Five factors were varied, creating 768 conditions, to see how the factors influenced parameter estimation and the recovery of the underlying grouping structure.

It was found that all simulation factors had some influence, but for parameter estimation the largest effects where found for the sample size factors: number of people per group and number of occasions. It may not be surprising, but it is encouraging that parameter estimation improved as the sample sizes increased. This was not the case with earlier versions of state space mixture modeling. In an earlier version of state space mixture modeling, a free parameter was estimated that determined the individual probability of group membership for each person in each group. This parameter greatly increased estimation time, and created an inconsistent estimator: as the sample size increased, the parameter estimates did not asymptotically approach their true values. A similar inconsistent estimator is found for non-Rasch item response theory models (Baker & Kim, 2004) using joint maximum likelihood which simultaneously estimates person-level trait scores

and item-level parameters. The small modification of estimating the group-level mixing proportions and as a second step calculating empirical Bayes probabilities of group membership for each individual manifests the consistent estimator reported here.

Similar to the findings for parameter estimation, all factors had nonzero influence for the recovery of grouping structure. However, the number of factors was found to be the single largest effect. At least two mechanisms exist that could account for the impact of the number of factors. First, the dimensionality of the state space places limits on the kinds of behavior that can be displayed in linear systems. For example, oscillations are not possible in one-dimensional space. All stationary dynamic behavior in one dimension is qualitatively similar. Hence, the different dynamics generated in one dimension are not often qualitatively different. In three dimensions, it is not possible to have two, coupled oscillators. The entire range of possibilities is either (a) three independent systems or (b) one oscillator pair and one independent system. As the dimension of the state space (i.e. number of factors) increases, more kinds of qualitatively distinguishable behavior are possible. A second mechanism for the influence of the number of factors is more a product of the simulation design than the nature of the dynamics. To aid in parameter and group recovery, the zero elements of the dynamics matrices were assumed fixed and known. More to the point, as the number of factors increases the number of zero elements in the dynamics matrices increases. Because of how the dynamics were generated, the number and placement of zeros in the dynamics matrices are not likely to be the same for higher-dimensional state spaces. Thus, group membership becomes related to the test of a parameter being zero or nonzero for that individual because in one group the parameter is fixed and known to be zero whereas in another candidate

group the same parameter is far from zero. Deciding to which group a person's data belong becomes easier when the groups differ in their pattern of fixed, zero elements. This makes higher-dimensional state spaces easier to estimate, and easier to find the correct grouping structure as an artifact of the simulation process. On the other hand, researchers who wish to use state space mixture modeling may have hypotheses about particular patterns of parameters being zero among some individuals and nonzero among others. So, this aspect of the simulation design directly mimics a viable research situation. One such research application is considered next with the biometric growth structure of cognitive ability in the National Longitudinal Survey of Youth.

Chapter 3

Cognitive Development of Youth in the NLSY

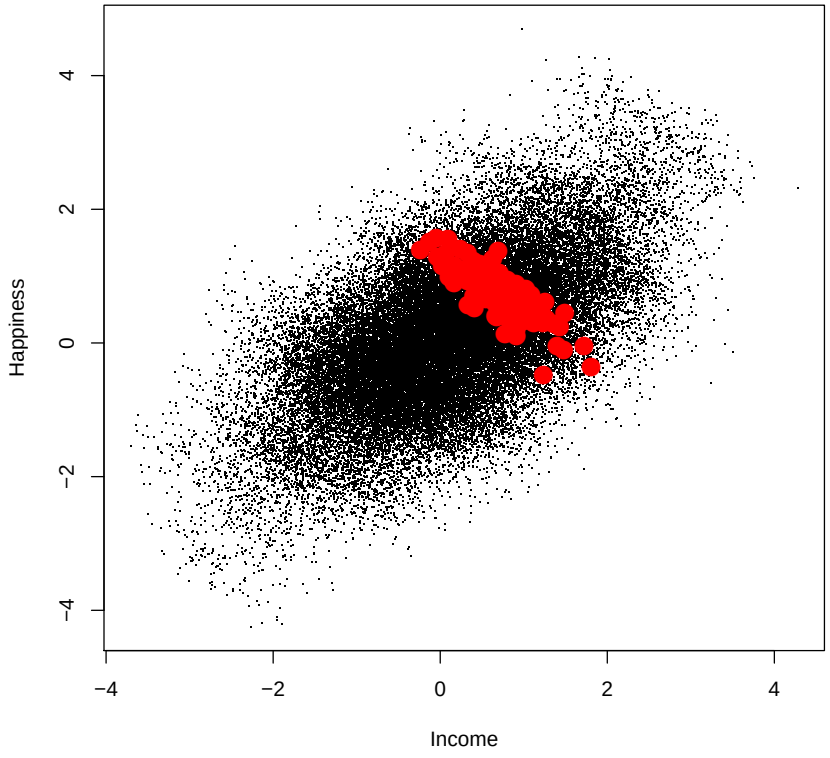
3.1 Introduction

Many authors have written that the fundamental unit of analysis in behavioral sciences is rightly the individual. Behaviors are emitted only from individuals, and thus they should form the centerpiece of any science founded on the study of behavior. The study of the individual, however, is often found in direct contradiction to scientific methods, many of which require observations to be replicated. Science is conducted on repeated instances of events, not single events. Thus, the individual is frequently rejected as a possible unit of analysis because of the need to replicate observations. As an example, Lord and Novick (1968) begin their famous tome on the measurement of mental abilities with a measurement theory founded on the individual. They begin to develop a test theory of the individual and then switch to that concerned exclusively with between-person variability: "The true and error scores defined above are not those primarily considered in

test theory .... They are, however, those that would be of interest to a theory that deals with individuals rather than with groups (counseling rather than selection)." (Lord & Novick, 1968, p. 32). In a similar vein, Bereiter argues that the study of the individual is the logical emphasis of behavioral sciences, but that the study of interindividual differences is "an expedient substitute" Bereiter (1963, p. 15).

More recently Molenaar and colleagues (Molenaar et al., 1993; Nesselroade & Molenaar, 2010a; Molenaar, 2010, 2004) have strongly argued against the plausibility of substituting the study of interindividual differences for intraindvidual differences. Perhaps the clearest way to argue against the substitution of many people at a single time for a single person at many times is to undercut the argument that the two forms of variability have anything to do with one another. Figure 3.1 shows within- and between-person variability in a simulated situation. Each point is the score of a single person at a single time measured on two variables, income and happiness. The red points are the scores for a single individual across several occasions of measurement. The between-person structure of Figure 3.1 is vividly a strong positive relationship. People who are happy tend to have high income, and people who are unhappy tend to have low income. This structure might suggest an intervention in which we raise the income of a sample of people and expect their happiness to correspondingly rise. In this simulated situation, the opposite effect would occur. The within-person structure of income and happiness, as highlighted by the red points, is a strong negative relationship. When any individual receives more income, their happiness decreases. Alternatively, whenever a person become happier, their income decreases. Hence Figure 3.1 shows that the within-person relationship between any pair of variables is completely independent of the between-person relationship.

Figure 3.1: A simulated between-person correlation of +0.80, but a within-person correlation of -0.75.



Given the independence of the within-person and between-person variabilities just described, the generally accepted validity of between-person models may be called into question. However, for some purposes between-person models are precisely the correct models. Between-person models remain valid, but under a slightly more restrictive set of conditions than may have been previously thought. Between-person models are valid, but only in describing differences between people unless further conditions are satisfied. In particular, for between-person models to apply within-persons the population must be homogeneous *and* the process of interest must be ergodic. Homogeneity is not enough. In the example in Figure 3.1, every individual has the same population within-person correlation (-0.75). The within-person correlation is homogeneous, yet the between-person correlation is completely different (+0.80). This can occur precisely because the process in not ergodic. A full discussion of ergodicity is beyond the present scope, but two brief points can be made about it. First, ergodicity, not homogeneity, is the key property that makes within-person and between-person variability interchangeable. When the process is ergodic, a sample of many individuals measured at just one time will give the same results as a sample of a single individual measured at many times: the within-person across time process behaves the same as the between-person single-time process. Second, the concept of ergodicity is related to the process of mixing. Ergodic processes tend to "forget" their initial conditions. As in the mixture of rum and cola in a beverage (e.g. Arnold & Avez, 1968, p. 19-20), the initial distribution of the two liquids is lost once the mixing has occurred. So, the lack of ergodicity in the example can be seen as a consequence of the non-mixing nature of the process: because the means do not drift and mix, the within-person and between-person structures are preserved separately and independently.

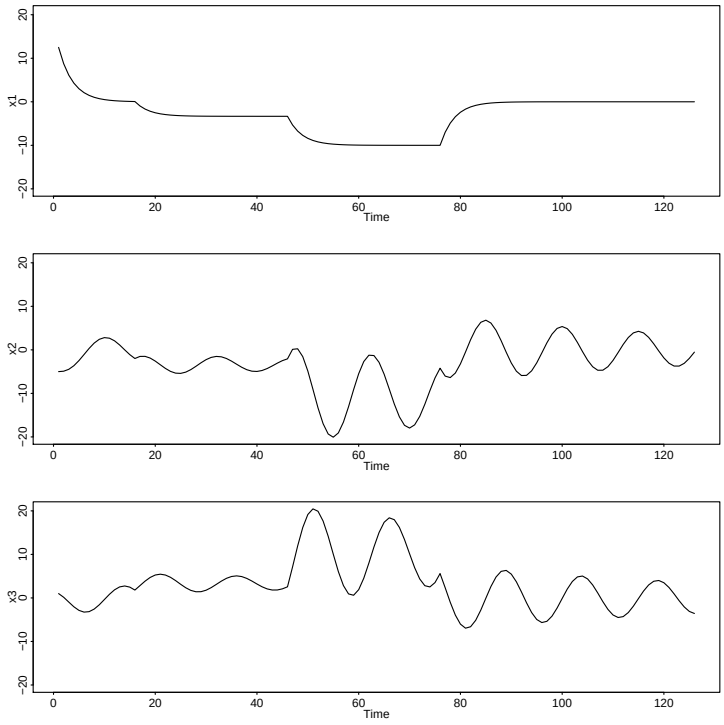
When ergodicity and homogeneity have been investigated in empirical data in psychology, it generally has not been found (e.g. Borkenau & Ostendorf, 1998; Gonzales & Ferrer, 2014; Hamaker et al., 2005). Likewise, standard statistical techniques do not indicate problems of model fit even when the generating model for individuals is extremely heterogeneous (Kelderman & Molenaar, 2007). So, when the goal is to understand individual-level processes the default strategy should be to measure individuals multiple times and then test whether homogeneity and/or ergodicity hold for this particular process. The balance then must be struck between the idiographic goal of modeling and understanding individuals on the one hand, and the the nomothetic goal of attaining generalizable knowledge about people on the other hand. Assuming complete heterogeneity forces an abandonment of the nomothetic goal, and implies modeling each individual separately and independently. Assuming complete homogeneity forces an abandonment of the idiographic goal, but allows modeling all individuals together. A middle path between these two extremes is taken here. Processes are modeled somewhat individually and partially homogeneously. A more detailed description of the approach taken will be provided after some further discussion of the dynamic perspective as it relates to this application.

Individual dynamics are represented by a state space model. The benefits of the dynamic perspective are described next. Figure 3.2 shows the latent values of two 3-dimensional time series. Both time series follow the same dynamics; they have the same rules governing their behavior. However, in Figure 3.2a the series of external inputs to the system are on/off as if in a set of treatments administered in a particular order, whereas in Figure 3.2b the series of external inputs to the system are continuous as if varying in some longitudinal pattern. The key to the dynamic perspective in this case is that the same rules (i.e. dynamics)

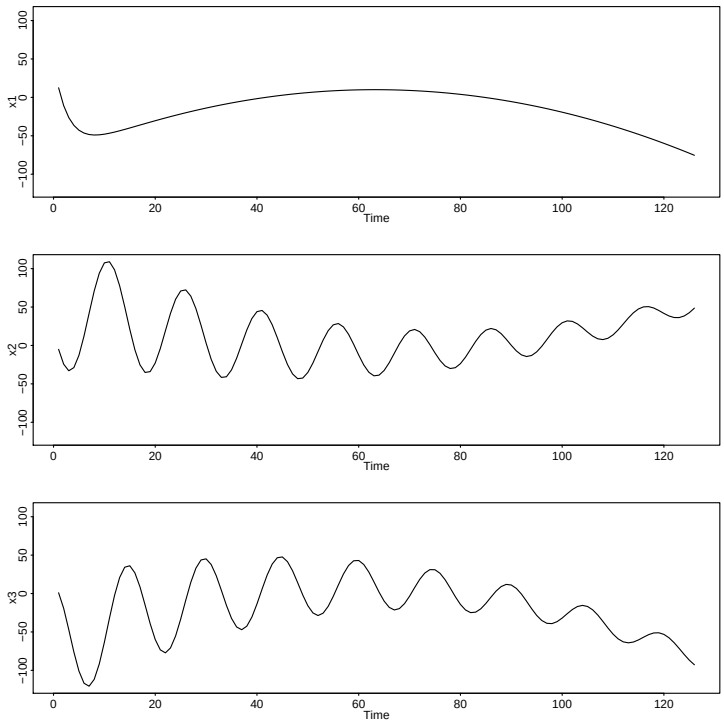
can produce very different looking behavior when subjected to different external forces (i.e. inputs). Much of psychology has emphasized the observed patterns of behavior. According to that perspective, Figures 3.2a and 3.2b are completely different, and one might hypothesize two distinct patterns of change: one that makes jumpy transitions and the other that smoothly varies. But both of these systems are identical. They merely have unequal external forces acting upon them. This is the primary advantage of the dynamical perspective. Dynamcis allow us to abstract away from the direct form of the observed data to examine the rules that govern the generation of that behavior.

The application presented in this chapter is on the cognitive development of children in the National Longitudinal Survey of Youth Children (NLSYC) data. A standard inquiry about cognitive development in these data would likely capitalize on the genetic relatedness of some of the individuals within it. Behavior genetics models could readily be applied. One such model is shown in Figure 3.3. In this situation, a single variable is modeled across two genetically related individuals. Their coefficient of relatedness is given by R. The variance of the single observed variable is partitioned into three pieces: additive genetics (A), common environments (C), and unique environments (E). Such a model is usually referred to as a univariate ACE model (Neale & Maes, 2004). The key to the interpretation of the latent factors is that their correlation structure is known. The additive genetics factors should correlate by R across individuals whose genetic relatedness is R. For identical (monozygotic) twins $R = 1$ because all[1] of their genes are the same. For fraternal (dizygotic) twins and other full siblings $R = 0.5$ on average because these individuals share half of their segregating

[1]Identical twins are not completely identical (Bruder et al., 2008; Molenaar et al., 1993), but for the purposes of ACE modeling it is a useful fiction.



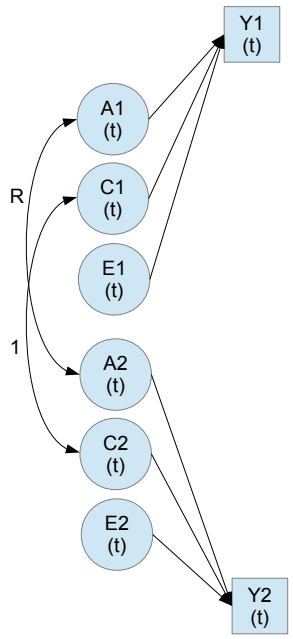
(a) On/Off Treatment-Style Inputs



(b) Continuous Covariate-Style Inputs

Figure 3.2: Latent values of a simulated 3-dimensional state space model with different inputs

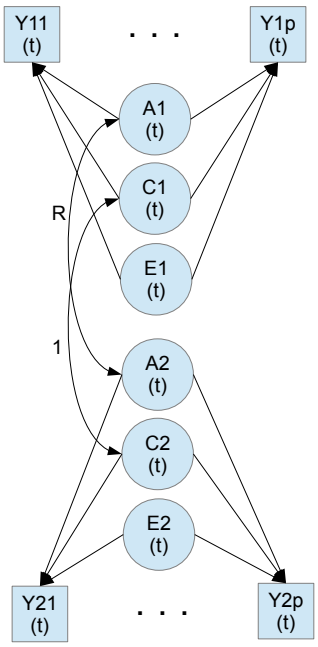
Figure 3.3: Univariate ACE model



genes on average. Other relations can be described similarly. For siblings living together, their common environments are assumed identical and hence their C factors are correlated one. This is called the equal environments assumption. The unique environments are always assumed to be correlated zero across twins to allow for completely idiosyncratic sources of variability. From a factor modeling perspective the multivariate extension of the ACE model is straightforward: add more indicators to the factors. An example of a multivariate ACE model is shown in Figure 3.4.

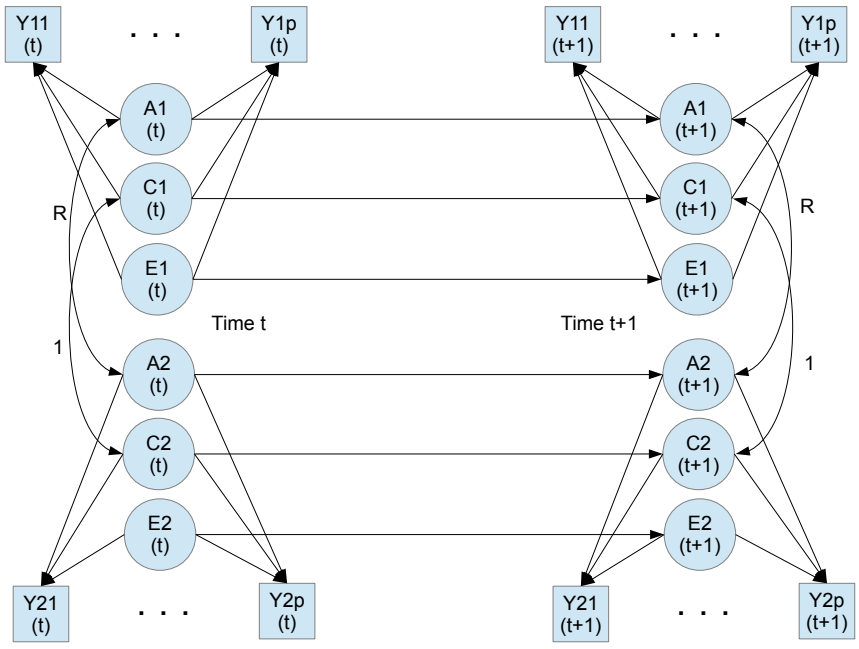
The univariate and multivariate ACE models described above are models of between-person variability. They elucidate the relative proportions of between-person variability that can be attributed to additive genetics, common environments, and unique environments. As described above, these between-person models generically diverge from within-person processes. Nesselroade and Molenaar

Figure 3.4: Multivariate ACE model



(2010a) alluded to several possible approaches to model within-person variability in the behavior genetics context. One such method is diagrammed in Figure 3.5. This model will be called the state space ACE model. It consists of a multivariate ACE model as the measurement model with autoregressive dynamics from one time to the next. Importantly, the rows of data in the state space ACE model correspond to different time points, whereas in the multivariate ACE model the rows of data are different pairs of individuals. Hence, the state space ACE model is a within-person, dynamic, behavior genetics model. Because this is a within-person model, the coefficient of relatedness may not be exactly equal to its population average: full siblings may share more or less than half of their genes. Consequently, the R coefficient can be estimated for an individual sibling pair. Similarly, the proportion of variance in the phenotype that is due to additive genetics may be allowed to differ across members of the sibling pair: one sibling's

Figure 3.5: State Space Multivariate ACE model



mathematical ability may be primarily inherited whereas the other's may be due to unique environments. When the coefficient of relatedness is freely estimated and the heritabilities of the sibling pairs are allowed to differ, the state space ACE model is called the idiographic filter ACE (iFACE) model (Molenaar, 2010; Molenaar, Smit, Boomsma, & Nesselroade, 2012) because of how it incorporates the idiographic filter (Nesselroade et al., 2007; Molenaar & Nesselroade, 2012), namely a fixed latent variable correlation structure with a measurement structure that varies across people. The goal of this chapter is to apply state space mixture modeling to the five of the NLSY cognitive variables where the state space model used is a state space ACE model.

3.2 Methods

3.2.1 Participants, Materials, and Procedures

The original NLSY sample (NLSY79 or NLSY79-Gen1) is a nationally representative household probability sample of 14- to 22-year-olds in the United States in 1979. It consists of 12,686 young men and women. The data analyzed here are from the NLSY children (NLSYC or NLSY79-Gen2), the children of the NLSY79 females. There are 3,276 different mothers for the NLSYC with 11,075 NLSYC kinship links. The R package NlsyLinks (Beasley, Bard, Hunter, Meredith, & Rodgers, 2013) was used for the kinship links. The NlsyLinks package uses responses by children and parents to many questions about parentage, relatedness, cohabitation, *et cetera* to make a deterministic, accurate classification of siblings (full, half), cousins, aunt/niece, parent/child, and so on. One important result in the NlsyLinks pacakge is the Links79Pair data set which provides, among other things, a coefficient of relatedness for sibling pairs in the NLSYC.

Five longitudinally measured cognitive variables were chosen for modeling in this application. The modeled variables were the Peabody Individual Achievement Test (PIAT) Reading Recognition, PIAT Reading Comprehension, PIAT Math, Peabody Picture Vocabulary Test (PPVT), and the Digit Span subscale from the Wechsler Intelligence Scales for Children-Revised. The PIAT reading and math were assessed every two years when the child was between ages 5 and 14. PPVT was administered biennially between ages 3 and 14. Digit span was assessed biennially from ages 7 to 11. The modal number of occasions of measurement for these variables ranged between two and four.

3.2.2 Data Analysis

The goal of the present analysis was to examine longitudinal variation in cognitive ability using biometrically informed models. As a preliminary step, cross-sectional ACE models were constructed for each age between 5 and 14. This replicates the between-person finding that cognitive abilities are, in general, highly heritable with the proportion of additive genetic variance increasing over different ages. Next, a multisubject state space ACE model was fitted to a random subsample of the NLSYC data, and cross-validated with another random subsample of the same size. Finally, a mixture multisubject state space ACE model was fitted to random subsamples of the NLSYC to explore the possible existence of unknown homogeneous groups.

3.3 Results

Cross-sectional analyses were conducted on the NLSYC for the five cognitive variables under investigation. Analyses were conducted on each age as a separate sample. The proportions of variance attributable to additive genetics (A), common environments (C), and unique environments (E) at each age are shown in Figure 3.6 through 3.10. Digit span, mathematical achievement, and PPVT show the popular cross-sectional finding that heritability (i.e. proportion of additive genetic variance) increases at higher ages. Reading recognition and reading comprehension do not shown these same patterns. The interpretation of these cross-sectional findings can be misleading. Each age represents a snapshot of the between-person structure (ensemble average). This snapshot may or may not reflect within-person changes in the heritability of cognitive ability. The simple addition of an age variable does not make the analyses longitudinal or represen-

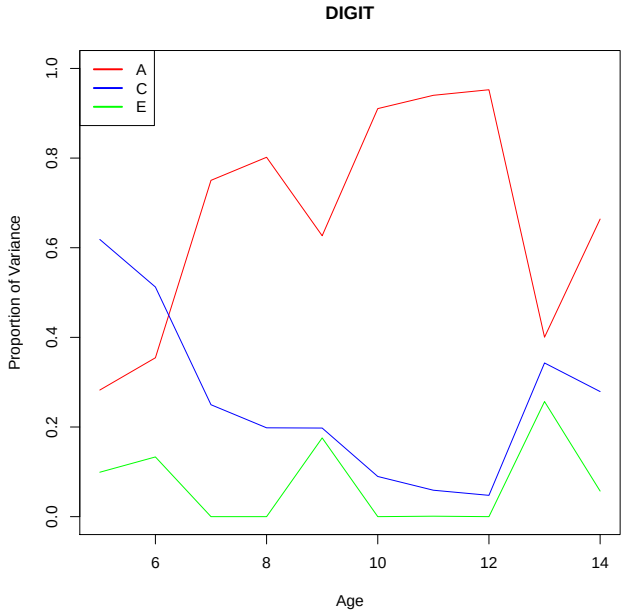


Figure 3.6: Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). DIGIT=WISC Digit Span.

tative of within-person changes. Recall the example of income and happiness in Figure 3.1 in which the between-person structure at every time point was constant but completely different from the within-person structure. This means that within-person analyses are critical for understanding individual processes.

Therefore, within-person analyses were conducted. The results of a multisubject state space model are shown in Table 3.1. The model was a multisubject extension of the state space ACE model from Figure 3.5. The model was run on a randomly selected subsample of the NLSYC with $N = 300$ kinship pairs. Heritabilities were constrained to be equal across kinship pairs. Moreover, homogeneity was assumed for the entire sample. Thus, the same parameters were fixed across the whole subsample. It is worth noting that the cross-sectional models reported previously also assume homogeneity in the sample, and they also rep-

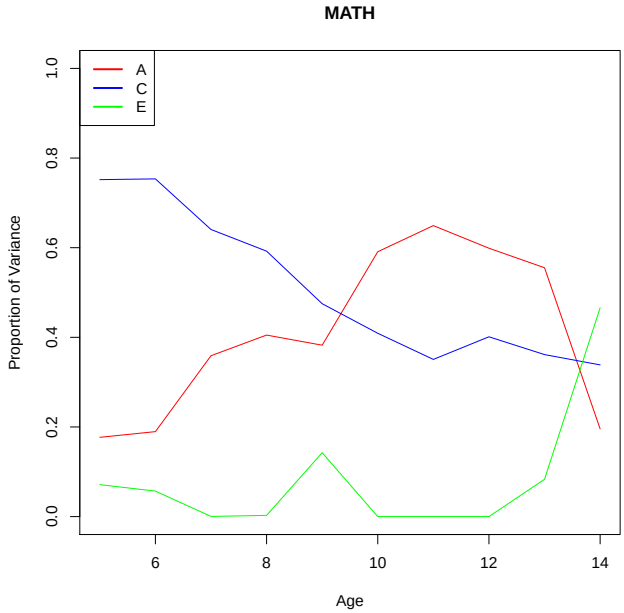


Figure 3.7: Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). MATH=PIAT Mathematical Achievement.

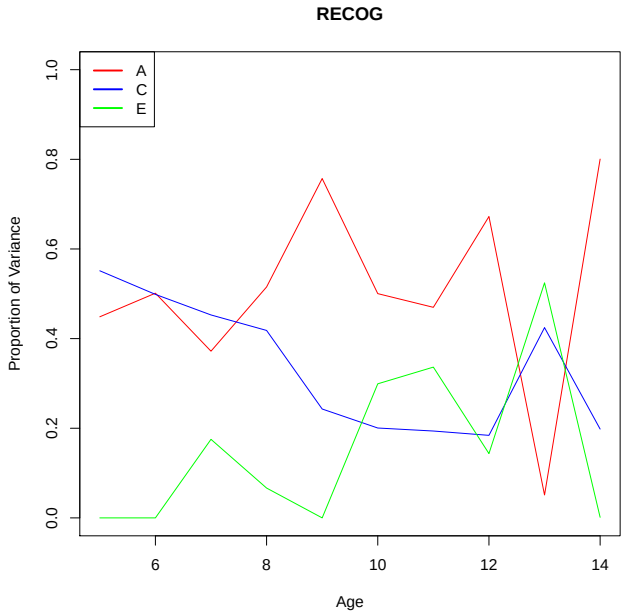


Figure 3.8: Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). RECOG=PIAT Reading Recognition.

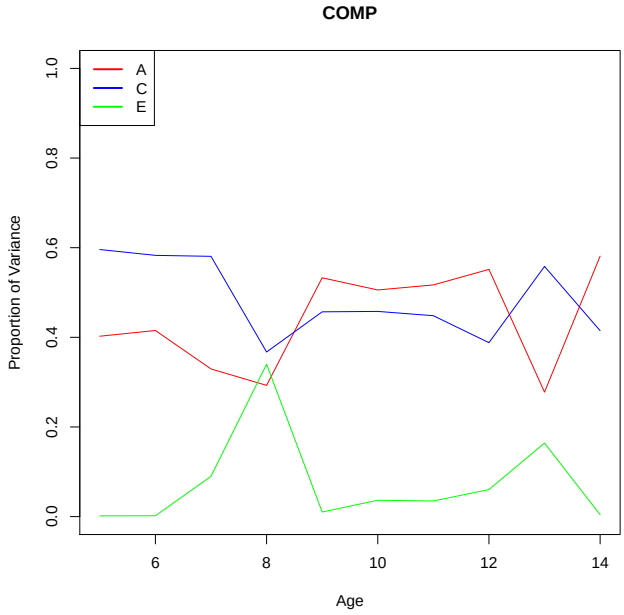


Figure 3.9: Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). COMP=PIAT Reading Comprehension.

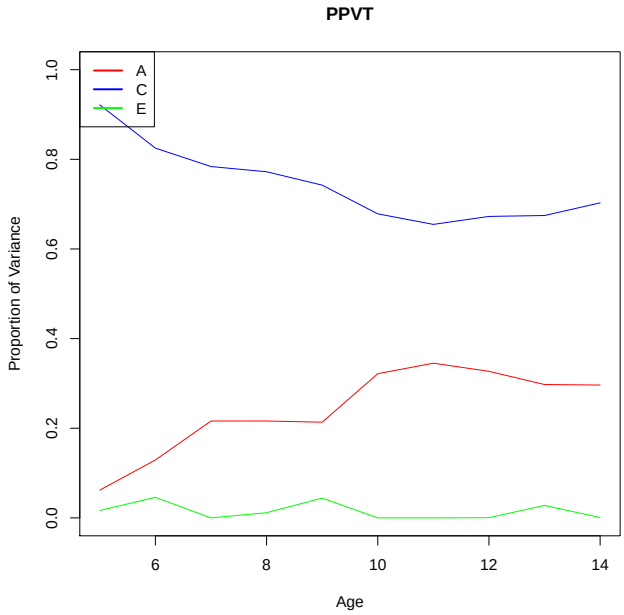


Figure 3.10: Cross-sectional changes in proportion of variance explained by additive genetics (A), common environments (C), and unique environments (E). PPVT=Peabody Picture Vocabulary Test.

resent between-person differences instead of within-person differences. The modeling procedure was repeated for cross-validation on a second random sample of the same size with its results shown in Table 3.2. The picture painted by the within-person results differs rather drastically from the cross-sectional results. The A, C, and E factors are found to be autoregressive with large coefficients between 0.8 and 0.9, indicating a highly correlated process from one time to the next. The highest contribution of additive genetics is found for PIAT math in the first sample (0.27), indicating that 27 percent of the within-person variability in PIAT math is attributable to additive genetics. In the second subsample, however, this estimate is somewhat lower (0.15). Broadly speaking, the A components are much lower in the multisubject state space model than in the cross sectional models. For many of the cross sectional models (Figures 3.6 to 3.10), the A component is between 0.20 and 0.60. By contrast, the within-person model suggests the A contribution is generally between 0.0 and 0.2. Much more variance is assigned to the common environments in these state space ACE models.

The next step was to search for underlying subgroups in the NLSYC. A state space mixture model with two mixture classes was fit to a random sample of 150 kinship pairs in the NLSYC[2]. Results for the first subsample are shown in Table 3.3. The analyses on this subsample suggest a large subgroup (89 percent) that behaves very similarly to the homogeneous case. They show moderate within-person heritability for PIAT math, but the vast majority of variance is portioned into the common environments (about 70 to 100 percent) with minimal contributions by additive genetics and unique environments. The smaller subpopulation found in the first subsample differs in that reading recognition and comprehen-

[2]More mixture classes and higher sample sizes were planned, but software limitations only allowed this as the maximum size due to a "protection pointer stack overflow".

Table 3.1: Multisubject State Space ACE: Random Subsample 1

	A	C	E
AREG	0.82	0.89	0.85
DIGIT	0.14	0.66	0.19
MATH	0.27	0.71	0.02
RECOG	0.03	0.74	0.24
COMP	0.07	0.79	0.15
PPVT	0.00	1.00	0.00

Note. Sample size is 300. AREG=autoregressive parameters. DIGIT=digit span, MATH=PIAT mathematical achievement, RECOG=PIAT reading comprehension, COMP=PIAT reading comprehension, PPVT=Peabody Picture Vocabulary Test. PIAT=Peabody Individual Achievement Test.

Table 3.2: Multisubject State Space ACE: Random Subsample 2

	A	C	E
AREG	0.87	0.88	0.84
DIGIT	0.19	0.57	0.24
MATH	0.15	0.80	0.05
RECOG	0.00	0.68	0.32
COMP	0.05	0.78	0.17
PPVT	0.00	1.00	0.00

Note. Sample size is 300. AREG=autoregressive parameters. DIGIT=digit span, MATH=PIAT mathematical achievement, RECOG=PIAT reading comprehension, COMP=PIAT reading comprehension, PPVT=Peabody Picture Vocabulary Test. PIAT=Peabody Individual Achievement Test.

Table 3.3: Multisubject Mixture State Space ACE: Random Subsample 1

	Subgroup 1 $\pi = .11$			Subgroup 2 $\pi = .89$		
	A1	C1	E1	A2	C2	E2
AREG	0.80	0.89	0.84	-	-	-
DIGIT	0.00	0.16	0.84	0.03	0.83	0.14
MATH	0.02	0.98	0.00	0.28	0.69	0.03
RECOG	0.26	0.39	0.35	0.04	0.76	0.14
COMP	0.63	0.37	0.00	0.07	0.79	0.14
PPVT	0.00	1.00	0.00	0.00	1.00	0.00

Note. Sample size is 150. AREG=autoregressive parameters. DIGIT=digit span, MATH=PIAT mathematical achievement, RECOG=PIAT reading comprehension, COMP=PIAT reading comprehension, PPVT=Peabody Picture Vocabulary Test. PIAT=Peabody Individual Achievement Test. '-' indicates parameter was fixed across models.

sion have heritabilities somewhat similar to those found cross-sectionally, but the other three cognitive variables (digit span, math, and PPVT) continue to display little genetic influence.

The same analyses were conducted on a separate subsample of the same size. Some of the results differ in this subsample. The two mixture classes are similarly distributed to the first subsample with one being very large (99.9 percent) and the other very small (0.1 percent). In this subsample, the smaller mixture class has moderate heritabilities, about 0.30, for all the cognitive variables except math. The larger mixture class has moderate heritabilities for digit span and math, but near zero for the reading variables and PPVT. The larger mixture class in the second subsample resembles the larger mixutre class from the first subsample. The exception is the heritability for digit span is near zero in the first subsample, but moderate in the second subsample. The within-person findings do not parallel the between-person results in either of the larger mixture classes; but some of them are similar for the smaller mixture classes.

Table 3.4: Multisubject Mixture State Space ACE: Random Subsample 2

	Subgroup 1 $\pi = .001$			Subgroup 2 $\pi = .999$		
	A1	C1	E1	A2	C2	E2
AREG	0.84	0.89	0.85	-	-	-
DIGIT	0.28	0.01	0.71	0.33	0.54	0.13
MATH	0.04	0.25	0.72	0.27	0.71	0.02
RECOG	0.25	0.15	0.60	0.04	0.70	0.26
COMP	0.30	0.21	0.49	0.08	0.77	0.15
PPVT	0.30	0.37	0.33	0.00	1.00	0.00

Note. Sample size is 150. AREG=autoregressive parameters. DIGIT=digit span, MATH=PIAT mathematical achievement, RECOG=PIAT reading comprehension, COMP=PIAT reading comprehension, PPVT=Peabody Picture Vocabulary Test. PIAT=Peabody Individual Achievement Test. '-' indicates parameter was fixed across models

3.4 Discussion

As a field, behavior genetics has historically emphasized between-person variability and population-level differences. This is a valid approach when results are only used to explain between-person differences and no attempt is made to account for individuals, individual-level process/mechanisms, or within-person differences. Unless the population is homogeneous and the process is ergodic, between- and within-person phenomena bare no necessary resemblance.

The standard finding that cognitive ability is highly heritable between individuals was replicated in several cross-sectional subsets of the NLSY. However, the within-person model over time showed minimal contribution from additive genetic variance across the five cognitive variables. The within- and between-person structure of cognitive development were found to be very different when homogeneity was assumed in both cases. But, the population might not be homogeneous. When the homogeneity assumption was relaxed in the within-person

model, a state space mixture model with two groups found a small subgroup in which cognitive ability was heritable within persons, but for the majority of individuals studied the intraindividual variance was dominated by common and specific environmental factors, principally the former. The structure of intraindividual heritabilty of cognitive ability thus appears quite different from that found in standard between-person biometric modeling.

Chapter 4

Discussion

A trade-off exists between the two goals of psychology. On one hand, psychology is a science of individuals, seeking to understand mechanisms and processes that occur within persons. On the other hand, psychology as a science must endeavor to find knowledge about people in general. These two goals are sometimes at cross purposes. Case studies, although interesting and potentially illustrative especially in situations of traumatic brain injury that elucidate cognitive functions, are not representative of scientific psychology. Indeed clinical psychology, which deals almost exclusively with individuals, has been found to be rather separate from the rest of psychology and this observation has been the basis for the formation of the Association for Psychological Science as a splinter group from the American Psychological Association (Rodgers, 1988).

This dissertation proposed, evaluated, and applied a method for analyzing data in a middle ground between between complete heterogeneity and complete homogeneity. State space mixture modeling uses an individual-level model that allows for the representation of change over time. Simultaneously, it capitalizes on the potential for people to share similar change processes, while still allow-

ing others to be quite different. Instead of assuming homogeneity of all persons and pursuing a nomothetic purpose of knowledge about people in general, even though that general knowledge might not apply to any particular individual studied, that assumption is relaxed to allow for some heterogeneity among people. Alternatively, instead of assuming complete heterogeneity and pursuing an idiographic purpose of knowledge about a single person, even though that individual knowledge might not apply to any other individual in the world, that assumption is relaxed to allow for some homogeneity among people. The goal of this dissertation was to combine idiographic and nomothetic approaches with a method that finds subpopulations with similar dynamic processes.

The inspiration for this avenue of research was the problem of pooling time series, more to the point it was the analysis of multivariate time series from multiple individuals. When multiple observations are made on multiple variables it creates multiple time series (Hannan, 1970; Quenouille, 1957). In more modern terms, multiple time series could be called multivariate or vector time series (Shumway & Stoffer, 2010; and for a historical review see Spanos, 2003). The problem of pooling regards the analysis of several vector time series. An example of several vector time series would be 20 individuals, each measured 100 times on 10 variables. Each of the 20 people has a 10-variate time series of length 100. Analysis of these data present several challenges, but multiple approaches exist. First, a mean across people could be taken in the time-domain, creating a single 10-variate mean time series of length 100. Second, the mean across people could be taken in the frequency domain by Fourier transforming each of the time series and then taking the mean, creating a 1-variate mean frequency power spectrum. The mean time-domain series is often found in the studies using functional magnetic resonance imaging (fMRI), and the mean frequency-

domain series is similarly applied in studies using electroencephalography (EEG; Molenaar & Gates, 2012).

A third option for analyses is concatenating the time series, creating single 10-variate time series of length $20 * 100 = 2,000$. This is sometimes done in both fMRI and EEG studies. Gates and Molenaar (2012) tried concatenating multivariate time series, along with their own individual analysis and pooling method, on simulated fMRI data from connectivity mapping generated by Smith et al. (2011), but their individual-level analyses worked far better.

A fourth and final method for handling multiple multivariate time series is to construct a multilevel time series model. Song and Zhang (2014) take this multilevel approach by specifying a multilevel structural equation model on lagged variables. It is the raw data version of dynamic factor analysis (Geweke & Singleton, 1981; Engle & Watson, 1981; Molenaar, 1985; Nesselroade et al., 2002) with the addition of multilevel modeling. The analytic problem encountered by Song and Zhang (2014) was the desired number of random slopes, coefficients allowed to vary randomly across people, was too large to actually estimate. There is a difference in perspective between the multilevel method and the mixture method applied here. In the multilevel method, differences across people are quantitative. Parameters vary according to a unimodal, exponential family distribution, often a normal distribution with some mean and variance. Opposed to this is the mixture method in which individuals are qualitatively different. Parameters vary according to a multimodal distribution, often a mixture of normal distributions. An example of an alternative to multilevel approach is the GIMME algorithm for fMRI data by Gates and Molenaar (2012). It regards each individual separately and as part of the group, allowing for discrete differences without any assumed distribution. The method is primarily data driven, with free use of modification

indices (Sörbom, 1989) to automate model selection. Whether the multilevel or the mixture method is selected is a matter of how the interindividual differences are conceived by the researcher. Discrete, qualitative differences suggest a mixture model; continuous, quantitative differences suggest a multilevel model.

When qualitative differences between people are hypothesized, a mixture method is not the only approach possible. Part of the impetus for the state space mixture method was work by Nesselroade and Molenaar (1999) in which a direct likelihood ratio test for pooling time series was suggested. Gonzales and Ferrer (2014) presented a relatively direct evaluation of the method suggested by Nesselroade and Molenaar (1999). They used the block-Toeplitz dynamic factor analysis program DyFa (Browne & Zhang, 2010) to empirically test for the ability to pool lagged covariances using simulated and real data. Of their simulated data they said "the test of covariance heterogeneity is not sufficiently sensitive to differences among multivariate covariance matrices." and "even when pooling of an entire sample is reasonably expected, such pooling did not occur." (Gonzales & Ferrer, 2014, p. 256). Thus, their results were not particularly encouraging.

Worth mentioning is the fact that state space mixture models are not a direct test of pooling. Rather, they are a combination of multisubject state space models with mixture models. Previously, single-subject dynamic models have been combined with mixture models (Hunter, 2014a) and dependent mixture models (Chow & Zhang, 2013; Yang & Chow, 2010; Chow, Grimm, Filteau, Dolan, & McArdle, 2013), but the application to finding subpopulations with matching dynamics is a novel aspect of the present work.

The novelty is made possible largely by software. The OpenMx (Neale et al., in press) implementation of state space models (Hunter, 2014b) allowed for the the direct combination of multisubject, state space, and mixture modeling. Such

a combination was simply not possible with other software like MKFM (Dolan, 2005), DyFa (Browne & Zhang, 2010), MATLAB (MATLAB, 2014), or the R package dlm (Petris, 2010). A burgeoning theme in statistical modeling is the unity between state space model and structural equation models (SEMs). Song and Ferrer (2009) and Chow, Ho, Hamaker, and Dolan (2010) both provided evidence for this unity. Much earlier, MacCallum and Ashby (1986) furnished a block matrix equivalence across state space models (called linear systems theory therein) and SEMs. Earlier still, Priestley and Subba Rao (1975) showed the identity between regression factor scores and Kalman updated state estimates. The implementation of the Kalman filter state space model in OpenMx made the unity of SEMs with state space models more tangible and easier to explore. To the author's knowledge OpenMx is the first conventional SEM program that also estimates state space models with a Kalman filter. It is hoped that the SEM and state space unity will become more clear with the aid of common software specification of both techniques. We next turn to further discussion of the simulation study that evaluated state space mixture modeling as a viable technique to uncover groups with similar dynamics.

4.1 Simulation

The simulation study generated almost 500,000 multigroup, multisubject, multivariate, multi-occasion data sets and fitted state space mixture models to those data sets. Although this represented only about 60 percent of the planned simulation study, the questions about parameter estimation and group membership recovery could still be addressed. Parameter estimation appeared adequate overall. As expected, increasing sample size (number of occasions and number of people

per group) had strong positive effects on the parameter estimation. Increasing the number of variables per factor and the number of factors also appeared to aid in parameter estimation, but less so than the sample size conditions. The only negative effect observed was for the number of groups, or mixture classes. Perhaps not surprisingly, increasing the number of mixtures in the model decreased the quality of parameter estimation.

The second question addressed by the simulation was the ability of state space mixture models to correctly assign people to their generating dynamic model. This, too, generally seemed to succeed. The mean accuracy of group assignment for models with more than one group was 86 percent. Treating the true and estimated group membership as two raters of a categorical variable, the mean Cohen's kappa for inter-rater reliability was 0.58, indicating strong agreement between raters far above chance but also notably far from perfect. The simulation factors were used to predict Cohen's kappa in a logistic regression. From this, it was found that the number of factors had the single largest influence on kappa with more factors corresponding to higher inter-rater agreement. The number of occasions had the second largest affect, again with more occasions leading to better agreement. The number of factors result is likely due to the increase dynamic dissimilarity at higher dimensions which was an artifact of the simulation parameters. An interesting question could still be answered with the simulation data that attempts to separate the similarity of the dynamics from the dimension in which they exist. In particular, continuous time dynamics are qualitatively identical (topologically conjugate) based on the number of zero, positive, and negative real parts of their eigenvalues (Hirsch et al., 2003). Discrete time dynamics might be distinguishable based on the number of positive, negative, real, and complex eigenvalues. If the number of factors effect disappears

after controlling for the effect of the eigenvalues, then the increasing classification quality would be explained by the actual dynamic properties of the individuals rather than the dimension of the space in which their dynamics exist.

There were several limitations of this simulation study. First, the planned simulation study was not completed. Almost 40 percent of the repetitions were not run. This was due to the unanticipated time some of the conditions took to run. Many of the larger conditions took almost 48 hours to run a single model. With that being said, all but 10 conditions had at least some repetitions, and the total number of simulation repetition was sufficiently large to make statistical significance virtually irrelevant.

Second, many aspects of the dynamic model were assumed known prior to fitting the model. The number of factors, the factor structure (i.e. which loadings were zero), the number of groups, and the dynamic structure (i.e. which dynamics parameters were zero) were all assumed known. This replicates a confirmatory modeling procedure. More information about the efficacy of exploratory state space mixture modeling is needed. Some initial analytic work has been conducted on exploratory dynamic factor analysis (G. Zhang, 2014; G. Zhang, Browne, Ong, & Chow, 2014), but little empirical investigation using exploratory methods. The number of factors and number of mixture components problems can likely be solved similar to non-dynamic settings. The factor structure and dynamic structure problems are related to the factor rotation problem and may have a similar solution (G. Zhang et al., 2014). These assumptions were necessary for an initial investigation of state space mixture modeling, but should be relaxed in later studies.

A third limitation of the simulation was the exclusion of model error. The data generated contain only sampling error; the model was true in the population.

Because all models to varying extents are wrong (MacCallum, 2003; MacCallum & Tucker, 1991; Rodgers, 2010), making the model only an approximation to the data generating mechanism would be a major step toward realism in the simulation study. Analogous to the structural assumptions mentioned previously, the correct, data-generating model was used in this simulation as an initial simplification to give state space mixtures an optimal chance of success. To the extent that this simulation study has established a degree of success, later work can add the additional challenge of fitting an imperfect model.

The fourth limitation was an additional structural assumption made, namely that measurement was assumed to be invariant across the entire sample of both occasions and people. Measurement invariance (see Meredith, 1964a, 1964b, 1993, for the classical approach) must be established in empirical investigations to ensure that the constructs being measured are uniform across the time and person sampling domains. In modeling longitudinal data, measurement invariance can be particularly nuanced and difficult (Meredith & Horn, 2001). As a matter of simplification, no attempt was made in this dissertation to address the potential issue of measurement invariance. Later studies and applications of state space mixture modeling should routinely evaluate this simplification.

Software has the capacity to surmount some of the limitations mentioned in the previous paragraphs. A few relatively simple software improvements could be made that may greatly increase the ability to estimate these large and complicated models. First, the specification of the state space mixture model involves making a "model" for each person-group combination. This creates a structure that scales multiplicatively with the number of people and the number of groups. A method that scales additively is currently under development. A second refinement of the software would utilize multiple processors in parallel. The

typical way that OpenMx parallelizes model estimation is by partitioning the data, conducting calculations on those partitions in parallel, and then gathering the results (a scatter-gather method of parallelization). The same method cannot be applied blindly with state space models because of the serial dependence of the rows of data. However, the same method could be applied on the person level: partitioning the data for each person as a whole, conducting calculations for each person, and then gathering the person results. This parallelization would improve the likelihood estimation time by almost a factor of P, where P is the number of cores, processors, or threads used. We next turn to a further discussion of intraindividual behavior genetics as an application of state space mixture modeling.

4.2 Genetics of Cognitive Growth

The empirical study applied state space mixture modeling to cognitive growth in the National Longitudinal Survey of Youth Children (NLSYC) using biometrically informed data from the NlsyLinks (Beasley et al., 2013). The cross-sectional findings that cognitive ability is quite heritable between people and tends to increase from childhood to adolescence were replicated in this sample. However, when examined intraindividually assuming homogeneity, cognitive ability was found to have very nearly zero heritability with most of the within-person variance being accounted for by common environments instead of additive genetics. When the assumption of homogeneity was relaxed slightly to included two homogeneous groups, one group was found that was very similar to the completely homogeneous model and a second smaller group was found in which some of the cognitive variables appeared moderately heritable. These results are preliminary

are certainly require further investigation.

This application has a number of shortcomings. First, the current implementation of the software to estimate state space mixture models could not estimate a homogeneous model on the entire sample, let alone a mixture model. Cross-validation on an additional random subsample of the data had results that were widely similar to those on the initial random sample, but a mixture model on the whole dataset still was not achieved. The software improvements outlined in the previous section should make the complete data model estimable. Until then, the subsample findings can be taken at least as preliminary evidence of a within-person genetic structure that may differ substantially from the between-person structure more commonly reported.

A second difficulty with the application was the interpretation of intraindividual heritability. Heritability is generally defined between individuals. The use of the term "intraindividual heritability" may, in reality, be inaccurate. Time-varying common and specific environments are uncontroversial, but it is generally accepted that the DNA sequence itself does not vary over time. Epigenetic influences may create heritable changes in gene expression, but not directly modify the genome of an individual (see Barber & Rastegar, 2010, for a review of the epigenetic control of Hox genes as an example). Thus, the proportion of variance accounted for by additive genetics in a within-person model is likely to be tapping into an alternative source of variance from the more typical between-person models. The additive genetic component in the within-person model might for example represent a gene-environment interaction, creating a time-varying element of variability to model. Hence, the small variability due to the additive genetics factor might be caused by no differences in additive genetics, but rather the limited variability due to genetic-environmental interactions. The term "intraindividual

heritability" is intended to have meaning by reference to its between-person analogue. After further within-person behavior genetic models are fit, an alternative term might prudently replace "intraindividual heritability". The nature of the within-person processes that lead to genetic or common environmental effects is yet not well-understood (Turkheimer, 2004).

Third, the small number of time points for each child makes the application of a time series method like state space modeling seem inappropriate. However, the simulation study suggested that time series parameters could be adequately estimated even when the number of occasions was 5, provided that the number of people per group was large. Part of the purpose of the application with such a small number of measurement times was to bridge the gap between longitudinal modeling and time series modeling.

Fourth, the model estimated only had autoregressive effects from each latent variable to itself at a later time. This certainly does not capture the potential for gene-environment interactions. Beam and Turkheimer (2013) operationalized gene-environment interactions as autoregressive effects from A at time t to C at time $t+1$. Their simulations that suggested a feedback loop between environments and phenotype that changed the gene-environment correlation within families. There may be additional feedback between the environment and the phenotype. Further autoregressive relationships between ACE components could be tested to look for these feedback processes.

Fifth and finally, it is known that socioeconomic status (SES) often moderates cognitive development. The goal of later applications will be to better understand the role that SES plays within individuals in their development (e.g. Bard, Hunter, Beasley, Rodgers, & Meredith, 2013). Incorporation of SES as a time-varying covariate or a moderator of the genetic effect may be necessary to

fully explain the within-person process. Next, we discuss plans for future work.

4.3 Future Work

The nature of science seems to be that every answered question leads to at least three more unanswered questions. And the present work is no exception. Several extensions have previously been mentioned and will not be recapitulated here. One that was not previously discussed is to conduct the inverse of the simulation study already reported. In the simulation performed, the measurement model was invariant across groups and the dynamics differed. An alternative would be to simulate invariant dynamics with diverse underlying measurement models, creating a dynamic idiographic filter (Z. Zhang et al., 2011).

A second line to pursue is the use of an expectation maximization (EM) algorithm for optimization of the mixture model. Because the state space mixture models used here are just instances of Gaussian mixture models, an EM algorithm exists that might aid in solution quality or speed in finding the solution (e.g. Roweis & Ghahramani, 1999). Along a related line, Gaussian mixtures can be described as a "soft" k-means cluster analysis (see section 14.3.7 of Hastie, Tibshirani, & Friedman, 2009) because class membership is determined probabilistically, without complete and deterministic assignment of records to clusters. So, the E-step from the EM algorithm for Gaussian mixture models could be used to overcome the limitations of k-means cluster analysis, namely that a brute force search of all possible group membership assignments even for a small problem would take prohibitively long. The E-step would be used to guide the search to make it far more efficient, into the realm of computational feasibility.

Many small changes to simulation parameters could be explored. Missing

data could be added to see if similar rules for missing data apply for state space models. The current implementation allows for completely and partially missing observations, and early explorations indicate similar performance to structural equation models with missing data. The inter-observation lag could be varied. Time-varying parameters could be allowed (Chow, Zu, Shifren, & Zhang, 2011). Continuous time models could be implemented using the hybrid Kalman filter (Kalman & Bucy, 1961). Nonlinear models could likewise be estimated with the extended Kalman filter. Prototype hybrid and extended Kalman filters exist, but have not been thoroughly tested.

Along a separate line of work, state space models could and should continue to be related to other methods. So far, much of this work has focused on structural equation models. One obvious family of methods that seems to have been unjustly left out of many investigations are the epidemic models of the onset of social activities (EMOSA; Rodgers & Rowe, 1993; Rodgers, Rowe, & Buster, 1998b; Stoolmiller, 1998; Rodgers, Rowe, & Buster, 1998a; Rodgers, 2007). These models bear striking resemblance to nonlinear state space models, often taking a form not unlike predator-prey models of competing populations (e.g. Gard & Kannan, 1976).

A final road to probe is provided by some recent work in neuroimaging. Unified structural equation modeling (uSEM; Kim, Zhu, Chang, Bentler, & Ernst, 2007) was developed to combine time-lagged and contemporaneous covariances in a single model for the purpose of analyzing fMRI data. Extended unified SEM (euSEM; Gates, Molenaar, Hillary, & Slobounov, 2011) was designed to further add the modulating effects found in event-related fMRI, as opposed to block trial fMRI. The state space modeling implementation in OpenMx (Hunter, 2014b; Neale et al., in press) may also allow uSEM and euSEM. Moreover, a

persistent issue in neuroimaging has been the heterogeneity of people (Molenaar, 2006, p. 47-55). Therefore, state space mixture modeling could be applied in this setting with the potential for some positive results. Next, we close with some brief concluding remarks.

4.4 Conclusion

More and more, the problem of measuring and modeling multiple individuals at many occasions on several variables is one of the growing fundamental challenges in the field of psychology. Several recent papers have presented attempts at a solution (Gonzales & Ferrer, 2014; Voelkle et al., 2014; Q. J. Zhang & Wang, 2014; Voelkle & Oud, 2014; Steele, Ferrer, & Nesselroade, 2014; Song & Zhang, 2014; McArdle, Hamagami, Chang, & Hishinuma, 2014). This dissertation presented another solution: state space mixture modeling. In simulation, the method appeared to be quite successful by showing good convergence, parameter estimation, and group recovery. From the simulation, it appeared that sample size was the primary factor determining accurate parameter estimation; whereas the dynamics were most critical in properly assigning people to groups. This was essentially an ideal result. State space mixture modeling correctly estimated parameters, and found people with similar dynamics in simulation. In application, the method showed the benefit of a within-person dynamic modeling perspective (see also Butner, Gagnon, Geuss, Lessard, & Story, 2014), but major limitations with regard to large sample sizes. Software improvements have already begun to remedy this situation. The method presented here will not be the final solution to the persistent problem of simultaneously respecting the communality and the uniqueness of individuals, but it is a good start.

With that being said, the broader impacts of state space mixture modeling have the potential to be far-reaching. The method correctly identified individuals with the same dynamics. The novel combination of state space models with mixture models allowed the discovery of underlying populations that were undergoing the same change processes. No other method has been published showing this much success at analyzing multiple multivariate time series. With this tool now readily available, applied studies that simultaneously allow for individual differences in change processes and generalizable knowledge across multiple individuals can begin.

References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., & Tsimring, L. S. (1993, October). The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, *65*(4), 1331-1392. doi:10.1103/RevModPhys.65.1331

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*(3), 527-543. doi:10.1037/0033-295X.115.3.527

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Henry Holt and Company.

Altman, R. M. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, *102*, 201-210. doi:10.1198/016214506000001086

Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, *28*(1), 1-25. doi:10.1007/BF02289543

Arnold, V. I. (1973). *Ordinary differential equations*. Cambridge, MA: MIT Press. (Translated from the Russian by R. A. Silverman)

Arnold, V. I., & Avez, A. (1968). *Ergodic problems of classical mechanics* (Vol. 9). Benjamin.

Åström, K. J., & Murray, R. M. (2010). *Feedback systems: An introduction for scientists and engineers* (Version v2.10c ed.). Princeton, NJ: Princeton University Press.

Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Belmont, CA: Wadsworth.

Barber, B. A., & Rastegar, M. (2010, Sep). Epigenetic control of Hox genes during neurogenesis, development, and disease. *Annals of Anatomy - Anatomischer Anzeiger*, *192*(5), 261-274. doi:10.1016/j.aanat.2010.07.009

Bard, D. E., Hunter, M. D., Beasley, W. H., Rodgers, J. L., & Meredith, K. (2013). Biometric nonlinear growth curves for cognitive development among nlsy children and youth [abstract]. *Behavior Genetics*, *43*, 507. doi:10.1007/s10519-013-9623-9

Beam, C. R., & Turkheimer, E. (2013, Feb). Phenotype-environment correlations in longitudinal twin models. *Development and Psychopathology*, *25*(01), 7–16. doi:10.1017/s0954579412000867

Beasley, W. H., Bard, D. E., Hunter, M. D., Meredith, K., & Rodgers, J. L. (2013). NLSY kinship links: Creating biometrical design structures from cross-generational data [abstract]. *Behavior Genetics*, *43*, 507-508. doi:10.1007/s10519-013-9623-9

Bereiter, C. (1963). Persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (p. 3-20). Madison, WI: University of Wisconsin Press.

Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to probability* (Second ed.). Nashua, NH: Athena Scientific.

Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America*, *17*(12), 656-660.

Bishop, M. J., & Thompson, E. A. (1986). Maximum likelihood alignment of DNA sequences. *Journal of Molecular Biology*, *190*, 159-165. doi:10.1016/0022-2836(86)90289-5

Boker, S. M. (1996). *Linear and nonlinear dynamical systems data analytic techniques and an application to developmental data.* Unpublished doctoral dissertation, University of Virginia.

Boker, S. M., & Ghisletta, P. (2001). Random coefficients models for control parameters in dynamical systems. *Multilevel Modeling Newsletter*, *13*(1), 10-17.

Boker, S. M., Horn, E., Meyer, M., & Turkheimer, E. (2014, June). A method for ACE decomposition of parameters of dynamical systems. In *Annual meeting of the behavior genetics association.* Charottesville, VA.

Boker, S. M., McArdle, J. J., & Neale, M. C. (2002). An algorithm for the hierarchical organization of path diagrams and calculation of componemts of expected covariance. *Structural Equation Modeling*, *9*, 174-194. doi:10.1207/S15328007SEM0902_2

Boker, S. M., Molenaar, P. C. M., & Nesselroade, J. R. (2009). Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. *Psychology and Aging*, *24*(4), 858-862. doi:10.1037/a0017912

Boker, S. M., Montpetit, M. A., Hunter, M. D., & Bergeman, C. S. (2010). Modeling resilience with differential equations. In P. C. M. Molenaar & K. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (p. 183-206). Washington, D.C.: American Psychological Association. doi:10.1037/12140-011

Boker, S. M., Neale, M. C., Maes, H., Wilde, M., Spiegel, M., Brick, T. R., ... Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., ... Team OpenMx (2014). Openmx 2.0 user guide [Computer software manual].

Boker, S. M., Neale, M. C., & Rausch, J. (2004). Latent differential equation modeling with multivariate multi-occassion indicators. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications* (p. 151-174). Dordrecht, Netherlands: Kluwer Academic Publishers. doi:10.1007/978-1-4020-1958-6_9

Boker, S. M., Schreiber, T., Pompe, B., & Bertenthal, B. I. (1998). Nonlinear analysis of perceptual-motor coupling in the development of postural control. *Nonlinear Analysis of Physiological Data*, 251-269. doi:10.1007/978-3-642-71949-3_15

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 Research in Personality*, *32*, 202-221. doi:10.1006/jrpe.1997.2206

Box, G. E. P. (1976, December). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791-799. doi:10.1080/01621459.1976.10480949

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* John Wiley & Sons.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (p. 415-452). Mahwah, NJ: Lawrence Erlbaum.

Browne, M. W., & Zhang, G. (2004). Developments in the factor analysis of individual time series. In *Factor analysis at 100: Historical developments and future directions.* Mahwah, NJ: Lawrence Erlbaum.

Browne, M. W., & Zhang, G. (2010). DyFA 3.00 user guide. Retrieved from `http://quantrm2.psy.ohiostate.edu/browne/`

Bruder, C. E., Piotrowski, A., Gijsbers, A. A., Andersson, R., Erickson, S., Diaz de Ståhl, T., ... others (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *The American Journal of Human Genetics*, *82*(3), 763–771. doi:10.1016/j.ajhg.2007.12.011

Butner, J. E., Gagnon, K. T., Geuss, M. N., Lessard, D. A., & Story, T. N. (2014). Utilizing topology to generate and test theories of change. *Psychological Methods*. doi:10.1037/a0037802

Cattell, R. B. (1952). The three basic factor-analytic research designs–their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499-520. doi:10.1037/h0054245

Cattell, R. B. (1966a). The data box: Its order of total resources in terms of possible relation systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (p. 67-128). Chicago, IL: Rand McNally.

Cattell, R. B. (1966b). Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology.* Chicago, IL: Rand McNally.

Cattell, R. B., Cattell, K. S., & Rhymer, R. M. (1947, December). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*(4), 267-288. doi:10.1007/BF02288941

Chow, S., Grimm, K. J., Filteau, G., Dolan, C. V., & McArdle, J. J. (2013). Regime-switching bivariate dual change score model. *Multivariate Behavioral Research*, *48*(4), 463–502. doi:10.1080/00273171.2013.787870

Chow, S., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, *17*, 303-332. doi:10.1080/10705511003661553

Chow, S., & Zhang, G. (2013). Nonlinear regime-switching state-space (rsss) models. *Psychometrika*, *78*(4), 740768. doi:10.1007/S11336-013-9330-8

Chow, S., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, *46*(2), 303-339. doi:10.1080/00273171.2011.563697

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. doi:10.1177/001316446002000104

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*(6), 426-443. doi:10.1037/h0026714

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavorial sciences* (3rd ed.). Routledge.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684. doi:10.1037/h0043943

Cronbach, L. J. (1975, February). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127. doi:10.1037/h0076829

Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., & Vattay, G. (2012). *Chaos: Classical and quantum*. Copenhagen: Niels Bohr Institute. Retrieved from `http://ChaosBook.org`

Dielman, T. E. (1989). *Pooled cross-sectional and time series data analysis*. New York: Marcel Dekker.

Dolan, C. V. (2005). MKFM6: Multi-group, multi-subject stationary time series modeling based on the kalman filter. Retrieved from `http://users/fmg.uva.nl/cdolan/`

Domowitz, I., & El-Gamal, M. A. (1993, December). A consistent test of stationary-ergodicity. *Econometric Theory*, *9*(4), 589-601. doi:10.1017/S0266466600007994

Domowitz, I., & El-Gamal, M. A. (1997). *A consistent nonparametric test of ergodicity for time series with applications* (Working Papers). Wisconsin Madison - Social Systems.

Domowitz, I., & El-Gamal, M. A. (2001). A consistent nonparametric test of ergodicity for time series with applications. *Journal of Econometrics*, *102*, 365-398. doi:10.1016/S0304-4076(01)00058-6

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.

Engle, R., & Watson, M. (1981, December). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, *76*(376), 774-781. doi:10.1080/01621459.1981.10477720

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Kings College London, UK: Wiley.

Fawcett, T. (2006, Jun). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874. doi:10.1016/j.patrec.2005.10.010

Gard, T. C., & Kannan, D. (1976, September). On a stochastic differential equation modeling of prey-predator evolution. *Journal of Applied Probability*, *13*(3), 429-443. doi:10.2307/3212463

Gates, K. M., & Molenaar, P. C. M. (2012, Oct). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310-319. doi:10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011, Jan). Extended unified sem approach for modeling event-related fmri data. *NeuroImage*, *54*(2), 1151-1158. doi:10.1016/j.neuroimage.2010.08.051

Geweke, J. F., & Singleton, K. J. (1981, February). Maximum likelihood "confirmatory" factor analysis of economic time series. *International Economic Review*, *22*(1), 37-54. doi:10.2307/2526134

Gonzales, J. E., & Ferrer, E. (2014, May). Individual pooling for group-based modeling under the assumption of ergodicity. *Multivariate Behavioral Research*, *49*(3), 245-260. doi:10.1080/00273171.2014.902298

Goodwin, G., & Sin, K. S. (1984). *Adaptive prediction, filtering and control*.

Prentice hall.

Gregson, R. A. M. (1983). *Time series in psychology*. Hillsdale, NJ: Lawrence Erlbaum.

Grewal, M. S., & Andrews, A. P. (2001). *Kalman filtering: Theory and practice using MATLAB* (2nd ed.). Wiley.

Grewal, M. S., & Andrews, A. P. (2008). *Kalman filtering: Theory and practice using MATLAB* (3rd ed.). Wiley.

Gu, F., Preacher, K. J., Wu, W., & Yung, Y.-F. (2014, Mar). A computationally efficient state space approach to estimating multilevel regression models and multilevel confirmatory factor models. *Multivariate Behavioral Research*, *49*(2), 119–129. doi:10.1080/00273171.2013.866537

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate time series analysis. *Multivariate Behavioral Research*, *40*, 207-233. doi:10.1207/s15327906mbr4002_3

Hamilton, J. D. (1994). State-space models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, p. 3039-3080). Elsevier. Retrieved from `http://www-stat.wharton.upenn.edu/~stine/stat910/`

Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.

Hartikainen, J., Solin, A., & Särkkä, S. (2011). Optimal filtering with Kalman filters and smoothers. *Department of Biomedica Engineering and Computational Sciences, Aalto University School of Science, 16th August*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.

Hirsch, M. W., Smale, S., & Devaney, R. (2003). *Differential equations, dynamical systems, and an introduction to chaos* (2nd ed.). Burlington, MA: Academic Press.

Hoey, J. (2012). The two-way likelihood ratio (g) test and comparison to two-way chi-squared test. Retrieved from `http://arxiv.org/abs/1206.4881v2`

Horn, J. L., & McArdle, J. J. (1980, October). Perspectives on mathemati-

cal/statistical model building (MASMOB) in reseach on aging. In L. W. Poon (Ed.), *Aging in the 1980's: Selected contemporary issues in the psychology of aging* (p. 503-541). Washington, DC: American Psychological Association. doi:10.1037/10050-037

Hunter, M. D. (2014a, May). Abstract: Dynamic mixture modeling of a single simulated case. *Multivariate Behavioral Research*, *49*(3), 286–287. doi:10.1080/00273171.2014.912890

Hunter, M. D. (2014b, May 22-25). Extended structural equations and state space models when data are missing at random. In *Annual meeting of the association for psychological science.* San Francisco, CA.

Icaza, G., & Jones, R. (1999, Mar). A state-space EM algorithm for longitudinal data. *Journal of Time Series Analysis*, *20*(5), 537–550. doi:10.1111/1467-9892.00155

Jones, R. H. (1993). *Longitudinal data with serial correlation: a state-space approach* (Vol. 47) (No. 2). CRC Press. doi:10.1080/00273171.2013.866537

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system.

Jöreskog, K. G., & Sörbom, D. (1982, November). Recent developments on structural equation modeling. *Journal of Marketing Research*, *19*(4), 404-416. doi:10.2307/3151714

Jöreskog, K. G., & Van Thillo, M. (1972). LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Basic Engineering*, *82*, 35-45. doi:10.1115/1.3662552

Kalman, R. E., & Bucy, R. S. (1961, March). New results in linear filtering and prediction theory. *Transactions of the ASME, Series D, Journal of Basic Engineering*, *83*, 95-108. doi:10.1115/1.3658902

Kelderman, H., & Molenaar, P. C. M. (2007, Oct). The effect of individual differences in factor loadings on the standard factor model. *Multivariate Behavioral Research*, *42*(3), 435–456. doi:10.1080/00273170701382997

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. *Human Brain Mapping*, *28*(2), 85-93. doi:10.1002/hbm.20259

Kitagawa, G. (1977, May). An algorithm for solving the matrix equation $X = FXF^T + S$. *International Journal of Control*, *25*(5), 745-753. doi:10.1080/00207177708922266

Kluckhohn, C., & Murray, H. A. (1948). *Personality in nature, society, and culture*. AA Knopf.

Leon, S. J. (2006). *Linear algebra with applications* (Seventh ed.). Upper Saddle River, NJ: Prentice Hall.

Lewin, K. (1936). *Principles of topological psychology*. York, PA: Maple Press.

Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging*, *26*(1), 34-47. doi:10.1037/a0020525

Lindstrom, M., & Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, *46*(3), 673–687. doi:10.2307/2532087

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley. (including material by Allan Birnbaum)

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research*, *38*, 113-139. doi:10.1207/S15327906MBR3801_5

MacCallum, R. C., & Ashby, F. G. (1986). Relationships between linear systems theory and covariance structure modeling. *Journal of Mathematical Psychology*, *30*, 1-27. doi:10.1207/S15327906MBR3801_5

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502-511. doi:10.1037/0033-2909.109.3.502

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Wiley.

MATLAB. (2014). *version 8.3 (r2014a)*. Natick, Massachusetts: The MathWorks

Inc.

McArdle, J., & Boker, S. (1990). *RAMpath: A computer program for automatic path diagrams.* Hillsdale, NJ: Erlbaum.

McArdle, J. J., Hamagami, F., Chang, J. Y., & Hishinuma, E. S. (2014, Jul). Longitudinal dynamic analyses of depression and academic achievement in the hawaiian high schools health survey using contemporary latent variable change models. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 608-629. doi:10.1080/10705511.2014.919824

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 234-251. doi:10.1111/j.2044-8317.1984.tb00802.x

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: John Wiley.

Meredith, W. (1964a). Notes on factorial invariance. *Psychometrika*, *29*, 177-185.

Meredith, W. (1964b). Rotation to achieve factorial invariance. *Psychometrika*, *29*, 187-206.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. *New methods for the analysis of change.*, 203-240. doi:10.1037/10409-007

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181-202. doi:10.1007/BF02294246

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201-218. doi:10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2006). Psychophysical dualism from the point of view of a working psychologist. *Erkenntnis*, *65*(1), 47–69. doi:10.1007/s10670-006-9014-1

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 K. Hood, C. Halpern, G. Greenberg, & R. Lerner (Eds.), *Handbook of developmental science, behavior, and genetics* (p. 626-648). Malden, MA: Blackwell Publishing Ltd. doi:10.1002/9781444327632.ch19

Molenaar, P. C. M., Boomsma, D. I., & Dolan, C. V. (1993). A third source of developmental differences. *Behavior Genetics*, *23*(6), 519-524. doi:10.1007/BF01068142

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112-117. doi:10.1111/j.1467-8721.2009.01619.x

Molenaar, P. C. M., & Gates, K. M. (2012, May). Aspects of psychophysiological data analysis: EEG coherency and fMRI connectivity mapping. *Monographs of the Society for Research in Child Development*, *77*(2), 129-138. doi:10.1111/j.1540-5834.2011.00671.x

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (p. 339-360). Springer. doi:10.1007/978-1-4615-0357-6_15

Molenaar, P. C. M., & Nesselroade, J. R. (2012). Merging the idiographic filter with dynamic factor analysis to model process. *Applied Developmental Science*, *16*(4), 210-219. doi:10.1080/10888691.2012.722884

Molenaar, P. C. M., & Newell, K. M. (2003). Direct fit of theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical and Statistical Psychology*, *56*, 199-214. doi:10.1348/000711003770480002

Molenaar, P. C. M., Sinclair, K. O., Rovine, M. J., Ram, N., & Corneal, S. E. (2009). Analyzing developmental processes on an individual level using nonstationary time series modeling. *Developmental Psychology*, *45*(1), 260-271. doi:10.1037/a0014170

Molenaar, P. C. M., Smit, D. J. A., Boomsma, D. I., & Nesselroade, J. R. (2012, Jun). Estimation of subject-specific heritabilities from intra-individual variation: iFACE. *Twin Research and Human Genetics*, *15*(3), 393-400.

doi:10.1017/thg.2012.9

Muthén, B. O. (1984, March). A general structural equation model with dichotomous ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115-132. doi:10.1007/BF02294210

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338-354. doi:10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O. (1994, February). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*(3), 376-398. doi:10.1177/0049124194022003006

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*(1), 81–118. doi:10.2333/bhmk.29.81

Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*(4), 489–503. doi:10.1007/BF02294325

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (Sixth ed.). Los Angeles, CA: Muthén & Muthén.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (in press). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*. doi:10.1007/s11336-014-9435-8

Neale, M. C., & Maes, H. H. (2004). *Methodology for genetic studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic Publishers B.V.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist*, *6*, 187-193. doi:10.1027//1016-9040.6.3.187

Nesselroade, J. R. (2010). On an emerging third discipline in scientific psychology. In P. C. M. Molenaar & K. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (p. 209-218). Washington, D.C.: American Psychological Association. doi:10.1037/12140-012

Nesselroade, J. R., Gerstorf, D., Hardy, S. A., & Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement*, *5*, 217-235.

doi:10.1080/15366360701741807

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (p. 235-265). Mahwah, NJ: Lawrence Erlbaum.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (p. 223-250). Thousand Oaks, CA: Sage.

Nesselroade, J. R., & Molenaar, P. C. M. (2010a). Analyzing intra-person variation: Hybridizing the ACE model with p-technique factor analysis and the idiographic filter. *Behavior genetics*, *40*(6), 776–783. doi:10.1007/s10519-010-9373-x

Nesselroade, J. R., & Molenaar, P. C. M. (2010b). Emphasizing intraindividual variability in the study of development over the life span. In W. F. Overton & R. M. Lerner (Eds.), *Handbook of life-span development* (p. 30-54). Hoboken, NJ: Wiley. doi:10.1002/9780470880166.hlsd001002

Özgür, A., Özgür, L., & Güngör, T. (2005). Text categorization with class-based and corpus-based keyword selection. *Lecture Notes in Computer Science*, 606-615. doi:10.1007/11569596_63

Pearson, K. (1900, Jul). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, *50*(302), 157-175. doi:10.1080/14786440009463897

Petris, G. (2010). A R package for dynamic linear models. *Journal of Statistical Software*, *36*(12), 1-16.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: a model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56. doi:10.1080/00273171.2012.710386

Priestley, M., & Subba Rao, T. (1975). The estimation of factor scores and kalman filtering for discrete parameter stationary processes. *International Journal of*

Control, *21*(6), 971–975. doi:10.1080/00207177508922050

Quenouille, M. H. (1957). *The analysis of multiple time series* (A. Stuart, Ed.) (No. 1). New York: Hafner.

R Development Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. New York: Routledge.

Rodgers, J. L. (1988). Structural models of the american psychological association in 1986: A taxonomy for reorganization. *American Psychologist*, *43*(5), 372. doi:10.1037/0003-066X.43.5.372

Rodgers, J. L. (2007). The shape of things to come: Diagnosing social contagion from adolescent smoking and drinking curves. In *Modeling contextual effects in longitudinal studies* (p. 343-362). Mahwah, NJ: La.

Rodgers, J. L. (2010, January). The epistemology of mathematical and statistical models: A quiet methodology revolution. *American Psychologist*, *65*(1), 1-12. doi:10.1037/a0018326

Rodgers, J. L., & Rowe, D. C. (1993). Social contagion and adolescent sexual behavior: A developmental EMOSA model. *Psychological Review*, *100*(3), 479-510. doi:10.1037/0033-295X.100.3.479

Rodgers, J. L., Rowe, D. C., & Buster, M. (1998a). Nonlinear dynamic modeling and social contagion: Reply to Stoolmiller (1998). *Developmental Psychology*, *34*(5), 1117-1118. doi:10.1037/0012-1649.34.5.1117

Rodgers, J. L., Rowe, D. C., & Buster, M. (1998b). Social contagion, adolescent sexual behavior, and pregnancy: A nonlinear dynamic EMOSA model. *Developmental Psychology*, *34*(5), 1096-1113. doi:10.1037/0012-1649.34.5.1096

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(5), 53 - 65. doi:10.1016/0377-0427(87)90125-7

Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Computation*, *11*, 305-345. doi:10.1162/089976699300016674

Shumway, R. H., & Stoffer, D. S. (2010). *Time series analysis and its applications with R examples* (Third ed.). New York: Springer.

Smale, S. (1967). Differentiable dynamical systems. *Bulletin of the American mathematical Society*, *73*(6), 747–817. doi:10.1090/S0002-9904-1967-11798-1

Smale, S. (1998). Finding a horseshoe on the beaches of Rio. *The Mathematical Intelligencer*, *20*(1), 39–44. doi:10.1007/BF03024399

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011, Jan). Network modelling methods for fMRI. *NeuroImage*, *54*(2), 875-891. doi:10.1016/j.neuroimage.2010.08.063

Snyder, C. W., Jr., Law, H. G., & Hattie, J. A. (1984). Overview of multimode analytic methods. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (p. 2-35). New York: Praeger.

Song, H., & Ferrer, E. (2009). State-space modeling of dynamic psychological processes via the Kalman smoother algorithm: Rationale, finite sample properties, and applications. *Structural Equation Modeling*, *16*, 338-363. doi:10.1080/10705510902751432

Song, H., & Zhang, Z. (2014, Jan). Analyzing multiple multivariate time series data using multilevel dynamic factor models. *Multivariate Behavioral Research*, *49*(1), 67-77. doi:10.1080/00273171.2013.851018

Sörbom, D. (1989, Sep). Model modification. *Psychometrika*, *54*(3), 371-384. doi:10.1007/bf02294623

Spanos, A. (2003, January). Time series and dynamic models. *A Companion to Theoretical Econometrics*, 585-609. doi:10.1002/9780470996249.ch29

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. doi:10.2307/1412107

Steele, J. S., & Ferrer, E. (2011). Latent differential equation modeling of self-regulatory and coregulatory affective processes. *Multivariate Behavioral Re-*

search, *46*(6), 956-984. doi:10.1080/00273171.2011.625305

Steele, J. S., Ferrer, E., & Nesselroade, J. R. (2014, Oct). An idiographic approach to estimating models of dyadic interactions with differential equations. *Psychometrika*, *79*(4). doi:10.1007/s11336-013-9366-9

Steinman, S. A., Hunter, M. D., & Teachman, B. A. (2013). Do patterns of change during treatment for panic disorder predict future panic symptoms? *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 150-157. doi:10.1016/j.jbtep.2012.09.003

Stephens, M. (2000, Nov). Dealing with label switching in mixture models. *J Royal Statistical Soc B*, *62*(4), 795-809. doi:10.1111/1467-9868.00265

Stoolmiller, M. (1998). Comment on "social contagion, adolescent sexual behavior, and pregnancy: A nonlinear dynamic EMOSA model. *Developmental Psychology*, *34*(5), 1114-1116. doi:10.1037/0012-1649.34.5.1114

Strogatz, S. H. (2000). *Nonlinear dynamics and chaos*. Cambridge, MA: Perseus Books.

Thrun, S., et al. (2002). Robotic mapping: A survey. In *Exploring artificial intelligence in the new millennium* (pp. 1–35). San Mateo, CA: Morgan Kaufmann.

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*(1), 1-32. doi:10.1037/h0075959

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (p. 317-341). Charlotte, NC: Information Age Publishing.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (p. 122-137). Madison, WI: University of Wisconsin Press.

Tucker, L. R. (1966, September). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279-311. doi:10.1007/BF02289464

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969, December). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421-459. doi:10.1007/BF02290601

Turkheimer, E. (2004). Spinach and ice cream: Why social science is so difficult. In L. DiLalla (Ed.), *Behavior genetics principles: Perspectives in development, personality, and psychopathology* (p. 161-189). Washington, DC: American Psychological Association.

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014, May). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*(3), 193-213. doi:10.1080/00273171.2014.889593

Voelkle, M. C., & Oud, J. H. L. (2014). Relating latent change score and continuous time models. *Structural Equation Modeling*. doi:10.1080/10705511.2014.935918

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Teiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard university press.

Wang, Q., Molenaar, P. C. M., Harsh, S., Freeman, K., Xie, J., Zhou, J., ... Ulbrecht, J. (2014). Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose and meal intake: An extended kalman filter approach. *Journal of Diabetes Science and Technology*. doi:10.1177/1932296814524080

West, M., & Harrison, P. (1997). *Bayesian forecasting and dynamic models*. New York: Springer.

Wright, S. (1918). On the nature of size factors. *Genetics*, *3*, 367-374.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, *6*, 320-332.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161-215.

Yang, M., & Chow, S. (2010). Using state-space model with regime switching to represent the dynamics of facial electromyography (EMG) data. *Psychometrika*, *75*, 744-771. doi:10.1007/s11336-010-9176-2

Zhang, G. (2014, Jul). Estimating standard errors in exploratory factor analysis. *Multivariate Behavioral Research*, *49*(4), 339–353.

doi:10.1080/00273171.2014.908271

Zhang, G., Browne, M. W., Ong, A. D., & Chow, S. (2014, Feb). Analytic standard errors for exploratory process factor analysis. *Psychometrika*, *79*(3), 444–469. doi:10.1007/s11336-013-9365-x

Zhang, Q. J., & Wang, L. P. (2014, Mar). Aggregating and testing intra-individual correlations: Methods and comparisons. *Multivariate Behavioral Research*, *49*(2), 130-148. doi:10.1080/00273171.2013.870877

Zhang, Z., Browne, M. W., & Nesselroade, J. R. (2011). Higher-order factor invariance and idiographic mapping of constructs to observables. *Applied Developmental Science*, *15*(4), 186-200. doi:10.1080/10888691.2011.618099

Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov models for time series: An introduction using R.* Boca Raton, FL: Chapman & Hall/CRC Press.