This is to certify that the
thesis entitled

SPURIOUS PREDICTORS IN RANDOM COEFFICIENT
MODELING

presented by

Michael Thomas Braun

has been accepted towards fulfillment
of the requirements for the

M.A. degree in Psychology

Major Professor's Signature

12/9/2009

Date

SPURIOUS PREDICTORS IN RANDOM COEFFICIENT MODELING

By

Michael Thomas Braun

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Psychology

2009

ABSTRCT

SPURIOUS PREDICTORS IN RANDOM COEFFICIENT MODELING

By

Michael Thomas Braun

A model fit to a longitudinal process results in a trajectory or set of trajectories. When variability is present in the longitudinal growth process, researchers will typically use time-varying covariates to predict the observed heterogeneity. Random coefficient modeling such as Hierarchical Linear Modeling (HLM) is currently the dominant approach to the analysis of longitudinal data in psychology because of its ability to effectively deal with heterogeneous data while at the same time allowing researchers to insert predictors into the model. The application of random coefficient models to longitudinal data assumes that the psychological process under investigation results solely from a deterministic trend. However, if a process, at least partially, results from a stochastic trend, then, random coefficient regression results are spurious. Previous research on simple regression models and Monte Carlo simulations are used to demonstrate the spurious results across six commonly observed models. A data analytic strategy is proposed to help researchers identify potential stochastic processes to avoid making inaccurate statistical and scientific inferences. Finally, two statistical techniques are briefly explained that can effectively handle stochastic data and help researchers accurately model the longitudinal process under investigation.

ACKNOWLEDGEMENTS

I would like to extend my extreme gratitude to my advisor, Dr. Rick DeShon, for all of the time, effort, and support that he has given me throughout this research process. I am undoubtedly a better writer and researcher thanks to his guidance.

I would also like to thank my committee members, Neal Schmitt and Brent Donnellan, for providing excellent guidance and insight on improving my current project.

Finally, thanks to all of my friends and family for all of the loving support they have always provided me. A special thanks to my parents for their continued love, encouragement, patience, and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Longitudinal data structures are increasingly common as researchers focus attention on the dynamics of psychological processes. Along with the increase of longitudinal data structures in psychology, the number of time points being collected is increasing (Walls & Schafer, 2006). When utilizing longitudinal data to study psychological phenomena over time, two questions are frequently posed. First, how does the criterion of interest change or grow over time and second, what variables can help predict or explain the observed patterns in growth trajectories. When examining growth curves to answer the first question, it is quite common to encounter a great deal of heterogeneity across groups or individuals in the growth process (Collins & Sayer, 2001). Researchers then typically attempt to answer the second question by searching for predictors to explain the heterogeneity in growth. One common method to explain this heterogeneity is to examine the relationship between the outcome of interest and a time-varying covariate (predictor) over time. Random Coefficient Modeling (RCM), such as multilevel models or Hierarchical Linear Models, can effectively deal with any clustering (heterogeneity) that may exist, while at the same time allowing researchers to insert predictors into the model to account for unexplained within- or between-group variance (Gelman & Hill, 2006). As a result, RCM has become the most common analytical tool for psychologists when analyzing longitudinal data.

A model fit to a longitudinal process results in a trajectory or set of trajectories. Typically, in psychological research the observed trajectories are modeled as if they are the result of a noisy, purely deterministic process. Although not widely recognized in psychology, it is well known in other disciplines (e.g., economics, physics, and biology) that the observed trends in these trajectories may be due, at least in part, to a stochastic, or random, process. A problem can arise when using random coefficient modeling to model growth data because the regression model and generalizations of that model, such as RCM, require the assumption that all trends in the dependent variable result solely from a deterministic process. When the regression model is used to analyze data resulting at least partially from a stochastic process, a serious inflation of Type I error rates, called spurious regression, is often observed. Spurious results can occur when regressing one stochastic series on another, regressing a stochastic series on time, and when regressing a stochastic series onto time and other stochastic predictors (Granger & Newbold, 1974; Nelson & Kang, 1984). This issue is particularly problematic in psychological research because frequently the focus of longitudinal studies is on growth processes, where stochastic trends have been frequently discovered in other fields such as computer science (Tang, Jin, & Zhang, 2008), physics (Uhlenbeck & Ornstein, 1930), genetics (Wright, 1931), and economics (Nelson & Plosser, 1982). Since RCM is a generalization of the regression model, it is likely to encounter the spurious results that are known to occur when using regression to analyze stochastic processes. This is especially true when time or Level 1, time-varying covariates are in the model as predictors (Nelson & Kang, 1984).

This proposal is organized in the following manner: I will demonstrate the problem by outlining current trends in longitudinal data collection, discuss the issue of heterogeneity in longitudinal data, and explain various methods for analyzing longitudinal data concluding with a discussion of RCM, the most commonly used analytical method. I will then discuss random walks where spurious regression was first documented. In my discussion of random walks I will explain the concepts of deterministic and stochastic processes, conceptually and mathematically explain what a random walk is, and then outline multiple ways spurious regression is commonly observed with random walks. I will then conceptually and mathematically explain RCM and its common applications. Next, I will apply the case of spurious regression to RCM and discuss what can happen when researchers use RCM to evaluate the relationship between stochastic series. Monte Carlo simulations will then be performed to demonstrate the large inflation of Type I errors. Finally, I will discuss the implications of not accounting for stochastic processes and will conclude by suggesting possible ways to deal with this problem.

Longitudinal Data Collection

In recent years, many psychological methodologists have called for the use of time in models, research designs, and theoretical frameworks (e.g., Ancona, Okhuysen, & Perlow, 2001; George & Jones, 2000; McGrath & Rotchford, 1983; Mitchell & James, 2001). Along with the literature on methods and theory building, researchers in many substantive areas in psychology have also called for the inclusion of time in theory and models. Teams and multilevel theory are two of the most predominant areas in organizational psychology where this has taken place (e.g., Kozlowski & Ilgen, 2006; Kozlowski & Klein, 2000; Mohammed, Hamilton, & Lim, In Press).

Unfortunately, many challenges are encountered when trying to study groups or individuals over an extended period of time. Traditional longitudinal studies utilized parameters similar to those of most cross section designs, large sample sizes and very few time points. Table 1 showcases some examples of these longitudinal designs from the last fifteen years. These studies were randomly sampled from a large pool of journals using the PsycINFO database. On average these studies had very large samples ($\bar{N}$ = 1662) but relatively few time points ($\bar{T}$ = 6). The only exceptions in these traditional longitudinal designs are large national studies such as the Minnesota Twin Study and the National Longitudinal Survey of Youth. These two large scale national studies were chosen because they are representative of ongoing national surveys used to address psychological and other phenomena over time. Like most traditional longitudinal designs, these large national studies focused more heavily on sample size (in the tens of thousands), but they were also able to collect data over a relatively large number of time points ($\bar{T}$ = 24). With these types of parameters the focus of traditional longitudinal designs is largely on number of participants rather than number of time points. This is helpful for studying phenomena that are inherently between-person, however it still does not provide a good method for studying phenomena that are within-person. Therefore, to better understand intraindividual psychological phenomena over time, a different type of longitudinal design was needed.

Table 1. Traditional Longitudinal Study Statistics

Traditional Longitudinal Designs				
Author(s)	Year	N	T	Journal
Baldwin & Hoffman	2002	762	11	Journal of Youth and Adolescence
Grimm	2007	7078	7	International Journal of Behavioral Development
Harford & Munthen	2001	2465	3	Journal of Studies on Alcohol
Hoffman et al.	2000	651	4	Substance Use and Misuse
Jang et al.	2004	320	4	The Gerontologist
Johnson et al.	1997	765	3	Journal of Quantitative Criminology
Raudenbush & Chen	1993	1725	8	Journal of Consulting and Criminal Psychology
Welte et al.	2005	625	3	American Journal of Drug and Alcohol Abuse
Witkiewitz	2008	563	7	Psychology of Addictive Behaviors

Note. N = Sample Size; T = series length.

Psychological phenomena naturally occur and develop within individuals. It is therefore important to study and analyze these constructs at the individual level (Block, 1995). In fact, some psychological phenomena, for example self–efficacy, appear to behave differently when studied at the within-person level compared to the between-person level (Vancouver, Thompson, & Williams, 2001; Vancouver, More, & Yoder, 2008). These ideas have gained much support in recent years and there is now a push to return the individual back into scientific psychology (Curran, Wirth, Nesselroade, Rogosa, Thum, Tuerlinckx, & von Eye, 2004; Molenaar, 2004). Fortunately, along with the push from psychological scholars for more longitudinal designs and research questions focused on the individual came many technological advances to make the process of collecting and analyzing longitudinal data easier for researchers. Computer storage capacity has greatly increased over the last 20 years making it possible to organize and keep data from many individuals over long periods of time in a relatively cheap and easy way (Walls & Schafer, 2006). Statistical packages are now widely available that have the most cutting edge analytical techniques making it easy and convenient for researchers to run their analyses. Although the increased ability to store and analyze longitudinal data was helpful, possibly the most influential technological improvement on longitudinal data collection is the rise of cell phones and other wireless computer devices.

This advancement in personal computers led to a rapid increase in the use of event/experience sampling. Event sampling began in the form of diary studies and evolved into a technique that utilizes some type of personal data device (e.g., palm pilot, personal data assistant, cell phone, etc.) to randomly (at the researchers' control) collect information from participants multiple times a day for a given period of time. The data collected are typically stored directly on the device and then downloaded to a computer at the conclusion of the study. The rise of event sampling with personal computers makes collecting longitudinal data much easier and more reliable for researchers and much more convenient for participants. This allows researchers to focus on having fewer participants with much longer series of data. Table 2 shows a sample of event sampling studies published in APA journals in 2007 and 2008 gathered using PsycINFO. As seen in the table, the trend is to have relatively few participants ($\bar{N}$ = 123) but have a large number of data points for each person ($\bar{T}$ = 47). Focusing on longer data streams for each person allows researchers to get a clearer and more complete picture of the processes by which psychological phenomena unfold over time within individuals. The push to study psychological phenomena at the individual level and all of the benefits and insights gained from this type of design has caused the number of event sampling studies to greatly increase over the last few years (Walls and Schafer, 2006).

Table 2. Event/Experience Sampling Study Statistics

Experience/Event Sampling Studies				
Author(s)	Year	*N*	*T*	Journal
Bohnert et al.	2008	246	56	Journal of Research on Adolescence
Bono et al.	2008	57	40	Journal of Applied Psychology
Brown et al.	2007	245	56	Psychological Science
DeHart et al.	2007	100	30	Journal of Experimental Social Psychology
Fleeson	2007	26	56	Journal of Personality
Graham	2008	38	49	Journal of Personality and Social Psychology
Hogarth et al.	2007	74	60	Risk Analysis
Ilies et al.	2007	106	30	Journal of Applied Psychology
Impett et al.	2008	55	14	Journal of Personality and Social Psychology
Jones et al.	2007	420	29	Journal of Applied Psychology
Kane et al.	2007	124	56	Psychological Science
Kubiak et al.	2008	16	28	Appetite
Kuppens	2007	80	59	Journal of Research in Personality
Lucas et al.	2008	144	52	Journal of Personality
Moberly & Watkins	2008	108	56	Journal of Abnormal Psychology
Moghaddam & Ferguson	2007	70	31	Journal of Personality
Nezlek et al.	2008	36	126	Emotion
Oishi et al.	2007	332	21	Journal of Personality and Social Psychology
Piasecki et al.	2007	50	14	Journal of Addictive Behaviors

Table 2 (cont'd).

Schneiders et al.	2007	131	45	Journal of Research on Adolescence
Snir & Zohar	2008	65	28	Applied Psychology: An International Review
Song et al.	2008	230	41	Journal of Applied Psychology
Summerville & Roese	2007	34	98	Journal of Experimental Social Psychology
Thewissen et al.	2008	154	60	Journal of Abnormal Psychology

Note. N = Sample Size; T = series length.

Along with the many benefits gained from collecting longer series for each individual comes increased methodological complexity. With the increase in series length, longitudinal data structures are becoming intensive longitudinal designs (Walls & Schafer, 2006). As this happens psychological longitudinal data starts to mirror the structure and properties of time series that can be seen in biomedical and economic literature. In medicine, short time series are considered to have as few as eight time points (Ernst, Nau &, Bar-Joesph, 2005). In economics, short time series typically have between only five and nine time points (Bhargava & Sargan, 1983; Hsiao, Pesaran, & Tahmiscioglu, 2002). With event sampling, psychological data structures can have anywhere from thirteen to over one hundred time points (from Table 2), well within the range of being considered time series. It is important to note that with this increase in series length, comes an increase in the complexity of the components of the data that need to be addressed by statistical models (Walls & Schafer, 2006). It is easy to ignore this increased complexity since the statistical packages psychological researchers use, such as HLM, can easily handle the increased number of time points. However, by ignoring the methodological issues that arise from having longer series, researchers can get an inaccurate representation of the underlying psychological relationships. Therefore, it is important for psychological researchers to consider and account for the increased complexity that comes from dealing with longer time series. More specifically, it is important to understand, identify, and deal with stochastic processes and their effect on spurious regression.

Longitudinal Data Analysis

The primary issue dealt with in psychological longitudinal data analysis is the existence of heterogeneous response processes that result in clustered data. In longitudinal data structures it is common to observe heterogeneity across individuals on the psychological phenomena of interest (Collins & Horn, 1991; Collins & Sayer, 2001; Harris, 1963; Moskowitz & Hershberger, 2002). Heterogeneity that goes unnoticed or unmodeled can lead to biased results and faulty inferences (Barcikowski, 1981; Kreft & De Leeuw, 1998; Molenaar, 2004; Winer, Brown, & Michels, 1991). Therefore, it is important to account for the presence of clustering due to heterogeneous response processes in the data. The presence of clustering leads to three possible analytical techniques that psychological researchers can use: pooled, disaggregated, or partial pooling.

For most of the previous century, researchers utilized repeated-measures analysis of variance (RM-ANOVA) to analyze longitudinal data. This technique pools all of the data together and computes the average growth across people (Gelman & Hill, 2006; Winer, et al., 1991). This method has a number of flaws. First, by pooling the data all variability in the sample is lost. Thus, it becomes impossible to study or predict individual differences in growth processes. The analysis only describes the average growth across people and may not describe any one individual's actual growth process. The second flaw, and possibly the most important, is that pooling the data ignores any clustering that may exist in the data. Pooling the data treats every observation as independent. The presence of clustering indicates that all data points are not independent and thus should be grouped together. When treating clustered data as independent RM-ANOVA analyzes the data with significance tests that are too liberal due to the fact that the sample size is considered larger than it should be. This leads to the results being biased and an inflation of Type I error rates. In fact, even in situations where only a moderate degree of clustering exists, the Type I error rate can increase to as high as 70% when the number of individuals in each group is large (e.g., N = 100) (Barcikowski, 1981; Kreft & De Leeuw, 1998).

A second possibility when analyzing heterogeneous longitudinal data is to completely disaggregate the data by analyzing each group or person separately and calculating individual growth curves. This method has flaws as well. Groups or individuals with few data points are weighted as strongly as those with many data points. This is ill-advised because with less information the reliability of the findings decreases. This problem could potentially be compounded if outliers exist in the groups with fewer data points. Outliers would influence groups with small numbers of data points more strongly than the overall sample or groups with more data points thus overemphasizing their impact. This could lead to inaccurate conclusions or inferences (Gelman & Hill, 2006).

The final method for analyzing longitudinal data is random coefficient modeling (RCM). RCM attenuates the problems of the prior two methods by compromising and partially pooling the data in an attempt to accurately represent any heterogeneity that exists between groups. The degree to which each groups' data gets pooled is a function of how much variability exists within the group and the amount of information available (number of data points) for that group. Groups that have a lot of within-group variance or very few data points get pooled in an attempt to bring the data closer to the mean. On the other hand, groups with strong within group agreement or large amount of information hardly get pooled at all (Gelman & Hill, 2006). This partial pooling allows for greater reliability for each group and, more importantly, eliminates the inflation of Type I errors due to clustering (Kreft & De Leeuw, 1998).

Along with the fact that RCM eliminates the problem of inflated Type I error rates due to clustering in the data, it has a number of additional advantages over RM-ANOVA and other analytical methods for longitudinal data. For example, RCM can easily handle missing data or data that is unevenly spaced between participants and can allow for the addition of continuous variables as predictors. Another advantage of RCM over previous analytical methods is that RCM allows for error structures that are both correlated and heteroscadastic. In longitudinal studies it is possible for the reliability of the data collected to change over time. Likewise it is possible that errors among individuals or groups across time points will become correlated (e.g., common method bias). Therefore, it is assumed in RCM that errors are correlated and heteroscadastic within groups or individuals but not between groups or individuals. This leads to a more complex error structure known as the block diagonal error structure (Gelman & Hill, 2006; Singer &

Willet, 2003). This is an advancement over RM-ANOVA where one of the primary assumptions is that errors are uncorrelated and have equal variance (sphericity) (Winer et al., 1991). The assumptions about the underlying error structure in RCM allow for a more accurate representation of the error processes that are commonly observed in longitudinal data than those of RM-ANOVA. All of these advantages resulted in RCM being the primary analytical tool used by psychological researchers when dealing with longitudinal data.

The Problem Presented

While RCM does provide a good solution for dealing with clustered data, it does not account for the methodological concern of spurious regression that can result from the presence of stochastic trends in the data that becomes more influential when analyzing longer time series. The regression model and generalizations of that model, such as RCM, require the assumption that the dependent variable is the result of a purely deterministic process. This is particularly problematic because frequently the focus of psychological longitudinal studies is on growth processes where stochastic processes have been observed in many other fields of study (e.g., computer science, physics, genetics, and economics). When the regression model is used to analyze stochastic data a serious inflation of Type I error rates, called spurious regression, is frequently observed. Spurious results are likely to be obtained anytime one stochastic series is regressed on another, a stochastic series is regressed on time, and when a stochastic series is regressed on time and other stochastic predictors (Granger & Newbold, 1974; Nelson & Kang, 1984).

To make matters worse, in the presence of stochastic trends as the number of time points increases, the significance tests used in regression analyses diverge, resulting in even greater Type I error rates (Durlauf & Phillips, 1988; Phillips, 1986, 1987). Therefore, the increased focus on the study of growth processes combined with the longer series used to study these processes leaves psychological researchers extremely susceptible to spurious regression when using RCM or other generalizations of the regression model. Two of the most dangerous cases are when *Time* or Level 1, time-varying covariates are in the model as predictors since both of these types of variables were shown to lead to spurious regression in the simple regression case (Granger & Newbold, 1974; Nelson & Kang, 1984). Unfortunately, these cases frequently exist in random coefficient modeling because a variable representing time is used in virtually all RCM analyses during the initial model testing of the unconditional growth model and since the inclusion of time-varying predictors is quite common and seen as a large advantage of using RCM as a method of analyzing longitudinal data (Raudenbush & Bryk, 2002; Singer & Willet, 2003).

Random Walks and Stochastic Processes

Random walks are one of the most commonly encountered and studied stochastic processes. They are prevalent in virtually every scientific discipline including computer science models of information search (Tang, Jin, & Zhang, 2008) , physics models of Brownian motion (Uhlenbeck & Ornstein, 1930), genetic models of genetic drift (Wright, 1931), ecological models of biodiffusion (Skellam, 1951) and population dynamics (Wang & Getz, 2007), and economic models of real GNP and employment (Nelson &

Plosser, 1982). In psychology, random walks are fundamental to the study of neuronal firing (Gerstein & Mandelbrot, 1964), speeded categorization (Nosofsky & Palmeri, 1997), diffusion models of decision processes (Busemeyer & Townsend, 1993), and consumer behavior such as new product adoption (Eliashberg & Chatterjee, 1986). The simplest form of a random walk is described by the equation:

$$Y_t = Y_{t-1} + \epsilon_t. \tag{1}$$

Where Y_t is the value at time t, Y_{t-1} is the value at time t-1, and ϵ_t is a random error term from a normal distribution that has a mean of zero and a constant variance, σ_ϵ^2. As seen in the equation, in random walks, each value is derived from the value of the data point directly preceding it in addition to some random error term (Enders, 1995). A random walk can be thought of as a series that is created by taking successive random steps. Therefore, any point in a random walk time series is the accumulation of random changes (i.e. error). Due to the fact that random walks are simply an accumulation of error terms, it is impossible to predict future values. This means that the best guess or expected value for any point in the future is the value directly preceding it. Therefore, the initial condition (i.e. intercept), Y_0, becomes the expected value for all future time points, as can be seen in the following equation:

$$\mathrm{E}(Y_t) = \mathrm{E}(Y_{t-1}) = \mathrm{E}(Y_0). \tag{2}$$

Since the intercept becomes the expected value for all time points, the mean of a random walk is time-invariant. The variance of a random walk, however, follows the equation:

$$\mathrm{Var}(Y_t) = \mathrm{var}(\epsilon_t + \epsilon_{t-1} + \ldots \epsilon_1) = \mathrm{t}\sigma_\varepsilon^2. \tag{3}$$

Therefore, as time increases so does the variance, meaning that the variance at one time point is not equal to the variance at any other time point. Figure 1 shows a random walk that appears to be a negatively directed growth process.

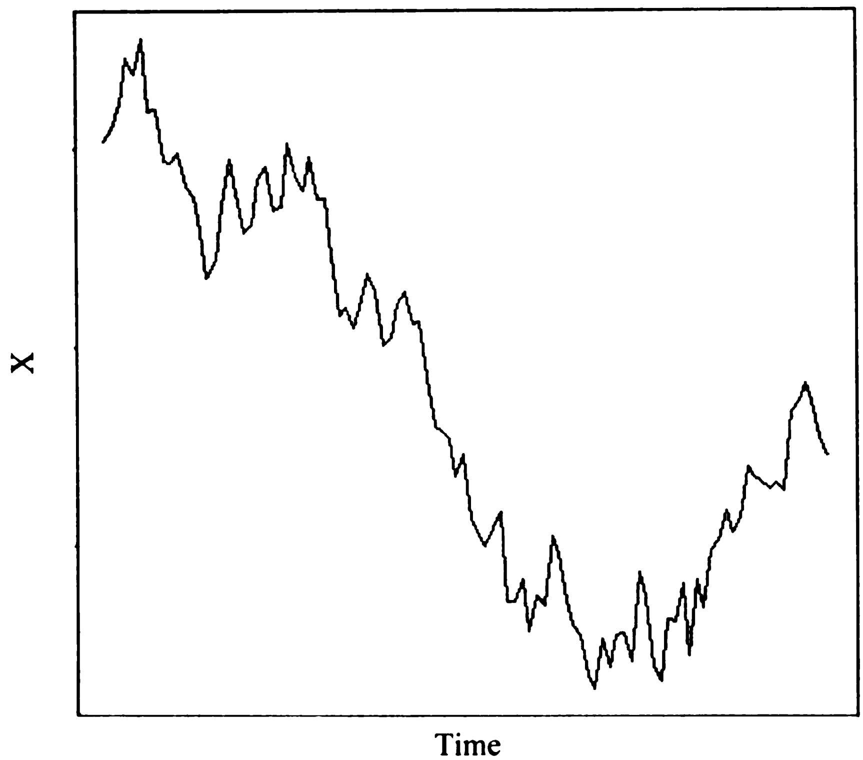

Figure 1. Random Walk

It is possible for random walks to be a bit more complicated and incorporate an environmental factor in the form of a drift or time trend. The distinction between stochastic and deterministic trends is seen most easily by examining the equation for a random walk with drift,

$$Y_t = Y_0 + \beta t + \sum_{i=1}^{t} \epsilon_t. \tag{4}$$

Y_t is the current value of a variable, Y_0 is the initial condition of a variable, βt is the deterministic component of the series, called a drift, and $\sum_{i=1}^{t} \epsilon_t$ is the stochastic component of the series representing an accumulation of random errors. This equation looks remarkably similar to the purely deterministic equation that is used in regression ($Y_t = \alpha + \beta t + \epsilon_t$). Often, it is very difficult to distinguish between series that are purely deterministic and ones that are purely stochastic. Figure 2 plots a regression line on data generated by either a purely deterministic or purely stochastic trend. Without the labels it would be near impossible to tell them apart, yet it is imperative that scientists are able to do so to make correct statistical and scientific inferences.

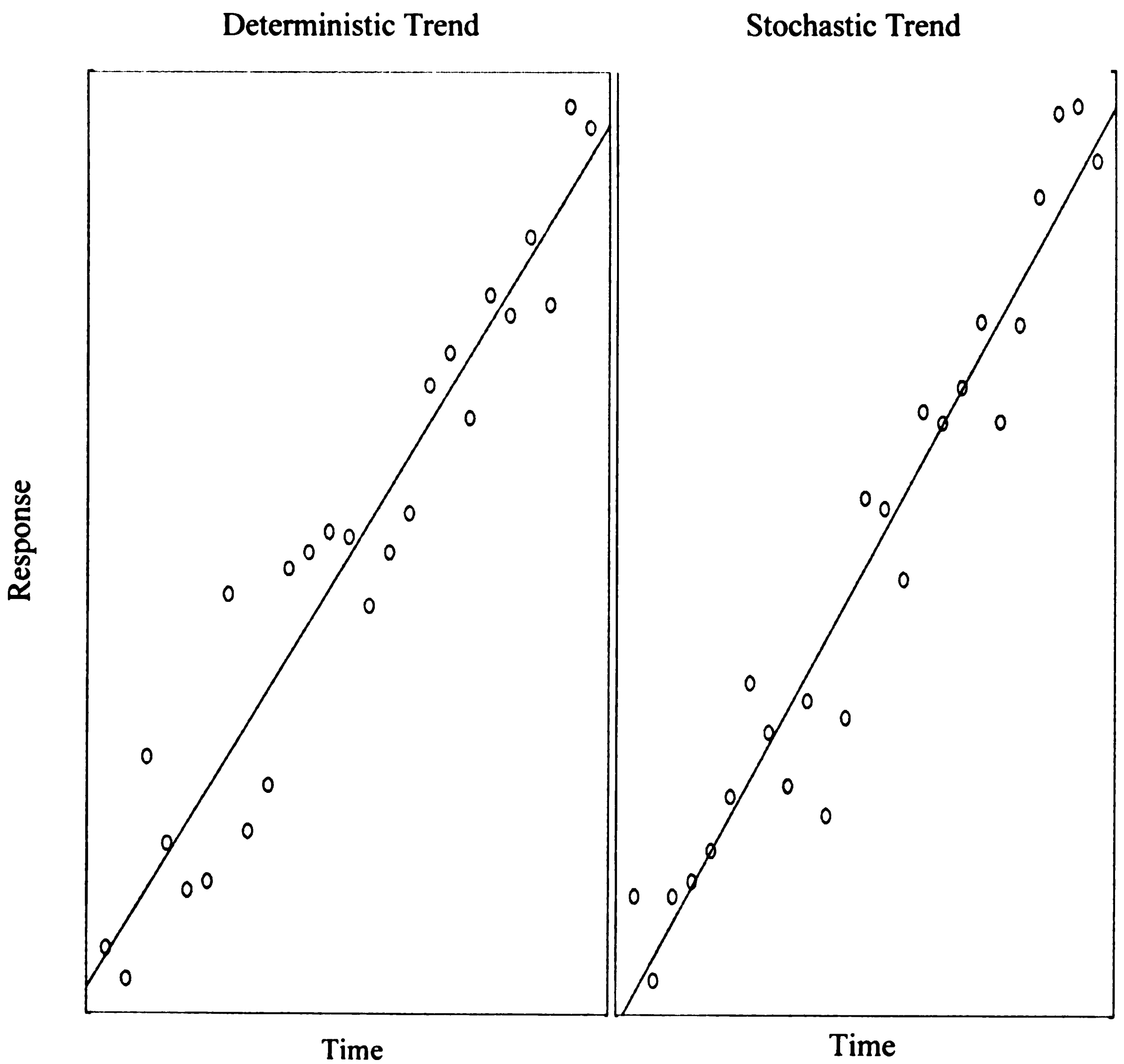

Figure 2. Deterministic and Stochastic Trends

For the purposes of this thesis, only the simplest case of random walks will be considered. Therefore, the effects of drifts and trends will be ignored. The addition of either a drift or trend would further complicate matters by making the random walk look even more like typical growth data than the stochastic trend demonstrated in Figure 2 and would result in a significant fixed effect for the slope of *Time* indicating that on average there is growth.

The prevalence of stochastic processes in the psychological sciences is still unknown. There are a number of potential areas of research that could incorporate stochastic trends in longitudinal data. One such area in organizational psychology is in the study of mood and emotions. Affective Events Theory is the dominant approach to the study of affect. It states that mood and emotions fluctuate over time and that events in the environment are proximal causes for affective reactions (Weiss & Cropanzano, 1996). Since environmental events occur in a somewhat random or stochastic fashion, it is possible that people's reactions to those events follow a somewhat random path (stochastic trend). Similarly, since almost all affect research relies on self report data, it is possible that over time measurement error is being compounded, which would lead to an accumulation of errors resulting in a stochastic process. Affect is just one area of research that has the potential to be impacted by the presence of stochastic trends in longitudinal data. Further investigation is needed to indentify the structure of affective data and to identify other potential stochastic processes in psychology.

Spurious Regression

Granger and Newbold (1974) first documented spurious regression in economics. They observed that frequently in the economic literature when regression was used to determine the relationship between time series, the regression coefficients were almost always significant and the amount of variance explained (R^2) was extremely high. They suggested this result was due to the presence of stochastic trends in the data, leading the regression models to inaccurately overestimate the true relationship between series. They demonstrated this problem by regressing one independent random walk on another. Since independent random walks are created only by an accumulation of errors, no true predictive relationship can exist between them. Therefore, any significant results beyond the nominal rate are due to problems with the estimation method. Granger and Newbold found that when regression is applied to random walks the variance of R^2 becomes too large, causing the distribution to no longer be unimodal around the origin. This results in R^2 being consistently overestimated, where the expected value goes from zero to 0.47 (47% of the variance explained). They also found that the standard error of the regression coefficient is grossly underestimated leading to a significance test that is too liberal. This results in the regression coefficient being significant approximately 76% of the time (for $\alpha = 0.05$), well beyond the nominal rate. To make matters worse, two unrelated series can often appear to covary, thus making their true relationship harder to detect. This has been documented many time using random walks; an example of such a case can be seen in Figure 3.

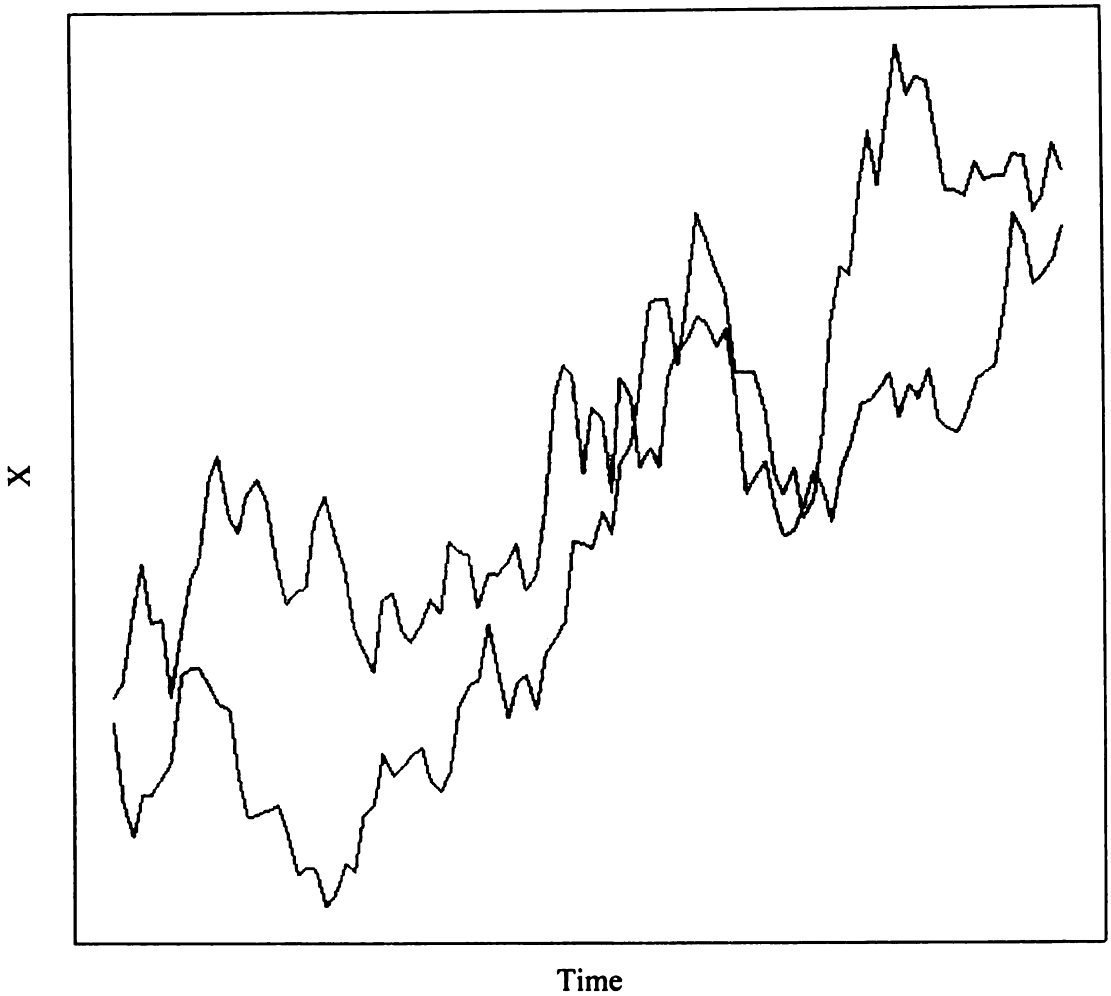

Figure 3. Independent Random Walks

Nelson and Kang (1984) expanded on the findings of Granger and Newbold (1974). They looked at how spurious regression manifested in the presence of a purely stochastic series (random walk) being regressed on to both an independent stochastic series (random walk) and a deterministic variable, *Time*. Just like in the previous case, the dependent random walk was created only from the accumulation of errors, so all significant results for any predictor beyond the nominal rate are due to faulty estimation. R^2 can never decrease as a result of adding predictors to the model so it is not surprising that the addition of *Time* as a predictor increased R^2 from 0.47 in the prior case to 0.50. The standard error of both *Time* and the predictor random walk were heavily underestimated resulting in very liberal significance tests. The addition of *Time* as a predictor decreases the percent of times that the slope parameter for the predictor random walk is significant from 76% in the prior case to 64% (for $\alpha = 0.05$). The slope of *Time*, on the other hand, is significant 83% of the time; meaning that once again both predictors are significant well above the nominal rate. To make matters worse, the effects of spurious regression are exacerbated as the number time points increases or the number of additional stochastic predictors increases (Granger & Newbold, 1974; Nelson & Kang, 1984; Phillips, 1986, 1987).

The effects of spurious regression are quite dramatic and could lead to many incorrect results and inferences if they are not accounted for. To show how spurious regression could play out in random coefficient models, a conceptual and mathematical overview of RCM will first be explained, followed by integrating the effects of spurious regression due to stochastic trends in the data with the models used in RCM analyses.

RCM Conceptual and Mathematical Overview

Random coefficient modeling (RCM) allows researchers to analyze longitudinal data using a two step process. Initially, the focus is on determining the shape and trajectory of growth curves and analyzing the degree to which there is heterogeneity in the response process. This initial step is done by partially pooling the data according to the within-group variance and amount of information available for each group. Then, RCM uses maximum likelihood estimation to estimate the hyperparameters for each variable (slope and intercept) in the model to maximize the fit between the observed and expected variance–covariance matrices. Hyperparameters are the mean and variance of each parameter across all groups or individuals in the sample. From these estimates the total within- and between-group variance is analyzed using the intraclass correlation coefficient (*ICC(1)*). This is where the second step in the data analysis process begins. If a significant amount of variance (either within or between) is left unexplained, RCM allows researchers to add predictor variables to the model in an attempt to explain additional variance. These predictors can take the form of either Level 1, within-person, predictors or Level 2, between-person, predictors. There are two distinct types of Level 1 predictors: time-invariant and time-varying. Time-invariant predictors vary between people but take on the same value for all time points within a given person. Time-varying predictors vary both between and within individuals over time. To evaluate the fit between models a log-likelihood ratio test or chi-squared difference test (called the deviance statistic) is used. The goal, as in all model comparisons, is to find a model that explains a maximum amount of variance while being as parsimonious as possible.

When psychologists utilize RCM to analyze their longitudinal data, the recommended practice is to begin by specifying two models as baseline models for future analyses. These two models are the multilevel unconditional means model and the multilevel unconditional growth model (Singer & Willet, 2003). These models are used to determine the trajectories of the growth curves and analyze the variance components.

The multilevel unconditional means model, as represented by Singer and Willet (2003), is specified as:

$$Y_{ij} = \pi_{0i} + \epsilon_{ij} \tag{5}$$

$$\pi_{0i} = \gamma_{00} + \zeta_{ij},$$

where it is assumed that:

$$\epsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ and } \zeta_{ij} \sim N(0, \sigma_0^2).$$

Y_{ij} is the dependent variable measured for each person i and each occasion j, π_{0i} is the mean of Y for individual i, γ_{00} is the mean of Y across everyone in the population, σ_ε^2 is the pooled variance of each individual's data around his/her mean, and σ_0^2 is the pooled variance of individual-specific means around the grand mean.

The unconditional means model splits the total variance into within- and between-person variance components. Using these variance components the intraclass correlation coefficient (*ICC(1)*) is computed, which indicates the proportion of total variance residing between individuals. From this, one can evaluate whether a substantial amount of variance resides within and between individuals. If it is determined that there is a significant amount of within- and between-person variance then it is suggested that the researcher search for predictors to explain both the within- and between-person variance. However, before adding substantive predictors it is recommended to run an additional baseline model, the multilevel unconditional growth model, to determine the shape and trajectory of the growth curves.

The multilevel unconditional growth model, as represented by Singer and Willet (2003), is specified as:

$$
\begin{aligned}
Y_{ij} &= \pi_{0i} + \pi_{1i} Time_{ij} + \epsilon_{ij} \\
\pi_{0i} &= \gamma_{00} + \zeta_{0i} \\
\pi_{1i} &= \gamma_{10} + \zeta_{1i},
\end{aligned} \tag{6}
$$

where it is assumed that:

$$\epsilon_{ij} \sim N(0, \sigma_\varepsilon^2)) \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right).$$

Y_{ij} is the dependent variable measured for each person *i* and each occasion *j*, π_{0i} is the initial status of Y for individual *i*, γ_{00} is the fixed effect for the intercepts and is the initial status of Y across everyone in the population, π_{1i} is the rate of change of Y for individual *i*, γ_{10} is the fixed effect for the slopes and is the rate of change of Y across everyone in the population, σ_ε^2 is the pooled variance of each individual's data around their linear change trajectory, σ_0^2 is the variance component of the random effect of the intercepts and is the unpredicted variability in initial status, σ_1^2 is variance component of the random effect of the slopes and is the unpredicted variability in rate of change, and σ_{10} is the population covariance between intercepts and slopes.

A log-likelihood ratio test allows the researcher to evaluate whether the unconditional growth model fits the data better than the unconditional means model. If this statistic is significant, the researcher would conclude that the unconditional growth model is a better representation of the data than the unconditional means model. If the resulting model still has a large amount of variance to be explained, in either the initial status (intercept) or the rate of change (slope), it is recommended that investigators search for additional predictors to explain the heterogeneity (Snijders & Bosker, 1999). Singer and Willet (2003) echo this notion by saying that researchers need to look at the variance components to "assess whether there is hope for future analyses" (p. 99).

If the researcher wishes to explain additional variance there are many possibilities that he/she can try and it is recommended that theory guide the choice of all subsequent predictors. When using RCM, additional predictors can take on two forms: Level 1 predictors and Level 2 predictors. Level 1 predictors are those that vary within individuals or groups. Some examples are salary, fatigue, intelligence, and mood. Level 2 predictors are those that vary between groups or individuals but are constant within groups or individuals. These are often grouping variable such as gender, race, religion, etc. Both Level 1 and Level 2 predictors can be entered into the model as a fixed effect, a random effect, or both. Predictors are usually added as fixed effects first and then the log-likelihood ratio test is used to determine whether additional variance is explained. Then, if there is still a significant amount of variance left unexplained, the predictor is allowed to vary as a random effect (Singer & Willet, 2003).

There are two distinct types of level 1 predictors: time-invariant and time-varying. Level 1, time-invariant predictors are things like salary, intelligence, personality, etc. where the variables can vary between people but are constant within a person over time. Level 1 time-varying predictors are things like fatigue, mood, etc. that can vary not only between people but also can vary within a person over time. Singer and Willet (2003) state that "because each predictor – whether time-invariant or time-varying – has its own value on each person's multiple records. A time-invariant predictor's values remain constant; a time-varying predictor's values vary. There is nothing more complicated to it than that" (pg. 160). Statistically speaking from an analytic point of view, this is absolutely true; however, inferentially there are great differences between time-invariant and time-varying predictors.

Spurious Regression in RCM

The current literature on the phenomenon of spurious regression only focuses on the simple regression case that looks at the aggregate relationship between variables across people. In random coefficient models panel data are collected and analyzed using separate series for each participant in the sample. To estimate how the effects of spurious regression would play out in growth curve analysis with panel data, one must consider the distribution theory for the parameter estimates of the simple regression model in the presence of stochastic trends in the data. Durlauf and Phillips (1988) derived the expected values and distributional convergence properties for each component of the regression model for the case of spurious regression due to stochastic data. With respect to RCM, the expected means of the distribution for each regression parameter represent the expected value for the fixed effects in RCM. Likewise, the expected variances of the distribution for each regression parameter represent the expected value for the variance components of the random effects in RCM. These insights allow for predictions to be made on how the various parameters in RCM will behave in the presence of stochastic data.

The application of the findings of Durlauf and Phillips (1988) and Nelson & Kang (1984) will be discussed for the fixed effects, the significance tests on the fixed effects, the variance components for the random effects, the significance tests on variance components for the random effects, and the deviance statistics.

Fixed Effects

Durlauf and Phillips (1988) mathematically proved that the average effect for both the intercept and slope of *Time* is zero. Nelson and Kang (1984) provide support for these findings by providing Monte Carlo simulations demonstrating that on average across replications the effect of the intercept and slope of *Time* is empirically zero. Therefore, since the fixed effects are estimated by calculating the average effects of each parameter across people, the fixed effects for both the intercept and slope of *Time* should be approximately zero for all models analyzed, regardless of the number of time points. When other stochastic predictors are added to the model they have been shown to converge weakly to a value which has empirically been demonstrated to be zero on average (Durlauf & Phillips, 1988; Nelson & Kang, 1984). Consequently, the fixed effects for all additional stochastic predictors should be approximately zero, regardless of the number of time points.

Significance Tests on the Fixed Effects

While the fixed effects across all models for all parameters are on average approximately zero, the behavior of their significance tests varies a great deal depending on what other parameters are in the model. When parameters are added to the model as only a fixed effect, it is essentially like adding them to the standard regression model. Therefore, all results obtained from previous research on simple spurious regression should be observed in the multilevel model as well. The significance tests on the parameter estimates for simple regression were proven to diverge with increased time points. This happens because as the number of time points increases, the regression model increasingly underestimates the standard errors of the parameters relative to their standard deviations, resulting in significance tests that are too liberal. This should cause the rejection rates for the parameters entered only as fixed effects to increase as the number of time points increases to well above the nominal rate (Durlauf & Phillips, 1988; Nelson & Kang, 1984).

When predictors are added to the model as both fixed and random effects the significance tests for the fixed effects are much better behaved. As demonstrated in Appendix A, when variables are entered as both fixed and random effects the standard errors of the fixed effects closely approximate the standard deviations of those effects across the reasonable range of time points studied in psychology. Therefore, the t-ratios for the fixed effects are accurately estimated leading the rejection rates on those fixed effects to stay roughly nominal.

Variance Components for the Random Effects

The variance components for the random effects for each parameter in all models are equivalent to the variance components of the distributions derived by Durlauf and Phillips (1988). The variance of the intercept parameter increases linearly as a function of the number of time points defined by the equation $\frac{2T}{15}$. The variance of the parameter estimate on the slope of *Time*, however, converges toward its true value of zero as a function of the number of time points defined by the equation $\frac{6}{5T}$. Nelson and Kang (1984) once again support these findings by empirically demonstrating the effect of number of time points on the variance of the parameters of both *Time* and the intercept using Monte Carlo simulations. As the number of time points increases, the variance component of the random effect of the intercept should tend toward infinity while the variance component for the random effect for the slope of *Time* should converge toward its true value of zero. When additional predictors are added to the model as random effects the total variance of the system is divided up further. This will cause the above relationships between the number of time points and the variance components for the random effects to be weakened; however, their distributional convergence properties will remain the same. The variance of an additional predictor, *X*, was shown to converge weakly to a distribution, meaning that the variance of *X* will converge towards a value and remain at approximately that value for all additional numbers of time points. As a result, the variance component for the random effect for *X* should be approximately a constant for all reasonably large number of time points (Durlauf & Phillips, 1988).

Significance Tests on the Variance Components of the Random Effects

To determine how the significance tests for the variance components for the random effects will behave the equation used to estimate the χ^2 statistic must be analyzed. Raudenbush and Bryk (2002, Equation 3.103) define the equation as:

$$\frac{\sum\left(\beta_{qj}-\gamma_{q0}-\sum \gamma_{qs} \boldsymbol{W}_{sj}\right)^2}{\boldsymbol{V}_{qqj}}, \tag{7}$$

where $\boldsymbol{V}_{qqj}$ is a block diagonal covariance matrix of errors with each block defined as:

$$\boldsymbol{V}_{j=}\sigma^2\left(\boldsymbol{X}_j^T\boldsymbol{X}_j\right)^{-1}. \tag{8}$$

By combining Equations 7 and 8 we get:

$$\frac{\sum\left(\beta_{qj}-\gamma_{q0}-\sum \gamma_{qs} \boldsymbol{W}_{sj}\right)^2}{\left(\sigma^2\left(\boldsymbol{X}_j^T\boldsymbol{X}_j\right)^{-1}\right)_{jj}}. \tag{9}$$

The denominator of Equation 9 gets small incredibly fast as the number of time points increases because $\left(\boldsymbol{X}_j^T\boldsymbol{X}_j\right)^{-1}$ becomes exponentially larger with each increased time point. This occurs because each additional time point results in the $\boldsymbol{X}_j$ matrix increasing by a factor of *N* (because each block increases by one and there are *N* blocks). Also, not only does the matrix greatly increase in size with the addition of each time point, but the variance of *Time* increases as the number of time points increases as well. This leads to the value of $\boldsymbol{X}_j$ greatly increasing with every additional time point. The relationship between σ^2 and number of time points is linear and does increase with increased time points, but the magnitude of the increase is significantly smaller than that of the $\boldsymbol{X}_j$ matrix. Therefore, the denominator gets extremely small at a rapid pace. The numerator also gets smaller as the number of time points increases but at much slower rate than the denominator. This causes the value of the χ^2 statistic to go towards infinity as the number of time points increases which results in the tests on the random effects being significant. Adding more predictors to the model does not change the relationship described above. By adding in a stochastic predictor, *X*, the $\boldsymbol{X}_j$ matrix will increase even more with the addition of each time point because the matrix gets larger by a factor of *2N* (each block increases by 2 due to addition of another time point and another value of *X*). As a result, the significance tests for all the variance components for the random effects should be significant.

Deviance Statistic

The final statistic that is analyzed is that of the χ^2 test between models. The behavior of the χ^2 test is completely predictable because it is a combination of the significance of each additional predictor added to the model. Consequently, when the difference between models is only the addition of a fixed effect, the χ^2 test will be significant the same proportion of times that the significance test on the fixed effect is significant. Since when stochastic predictors are only added as fixed effects their significance tests become significant more frequently with increased number of time points, so will the χ^2 test. When predictors are added to the model as random effects the significance tests on the variance components of those effects will be significant. Therefore, the χ^2 test between the models will also be significant.

As mentioned earlier, RCM is typically used to analyze longitudinal data through a stepwise process of nested model testing (Kreft & De Leeuw, 1998; Raudenbush & Bryk, 2002; Singer & Willet, 2003; Snijders & Boskers, 1999). Six common models will be explained in this research. The first model presented will be the unconditional means model. The second model will be the unconditional growth model followed by model three which is the unconditional growth model with X included as a fixed effect. The fourth model will build from model three and have X as both a fixed and random effect. Some researchers choose not to include *Time* in the model after the initial model testing stage. As a result, a fifth model will be discussed that is the unconditional means model with X included as a fixed effect. The final model will build on model five by includeding X as both a fixed and random effect. Each model will be presented, the model parameters will be briefly explained, and then hypotheses regarding each parameter estimate and their corresponding significance test will be made using the above rationale. The model notation used will be taken from Singer & Willet (2003).

Model 1: Unconditional Means Model

The first recommended model is the unconditional means model:

$$\begin{aligned} Y_{ij} &= \pi_{0i} + \epsilon_{ij} \\ \pi_{0i} &= \gamma_{00} + \zeta_{ij}, \end{aligned} \tag{10}$$

where it is assumed that:

$$\epsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ and } \zeta_{ij} \sim N(0, \sigma_0^2).$$

This model includes a fixed effect for the intercept, γ_{00}, a random effect for the intercept, ζ_{0i}, a variance component for the random effect of the intercept, σ_0^2, and a variance component for the residual error termσ_ϵ^2. The primary purpose in running the unconditional means model is to calculate the *ICC(1)* to determine the proportion of variance that exists between groups or individuals. The *ICC(1)* is calculated by the amount of variance between individuals, σ_0^2, divided by the total variance $(\sigma_0^2 + \sigma_\epsilon^2)$. For a random walk, the within person (residual) variance is typically set to one. The between person variance or the variance between the means across random walks is equal to the variance of the intercepts. This occurs because the expected value or mean for a random walk is its initial value (see Equation 2). The variance at any time point is equal to σ^2t (Equation 3) and the intercept is calculated at t = 1 so the between-person variance is also equal to one. Therefore, the true *ICC(1)* for random walks is equal to $\frac{1}{(1+1)}$ or 0.5. When the regression model is used to analyze stochastic data (random walks) the estimated variance between individuals is approximately $\frac{T+1}{3}\sigma_\epsilon^2$, compared to the estimated variance within individuals which is approximately $\frac{T+1}{6}\sigma_\epsilon^2$ (Nelson & Kang, 1984). Using these values, the expected *ICC(1)* for random walk data is:

$$ICC(1) = \frac{\frac{T+1}{3}\sigma_\epsilon^2 *}{\frac{T+1}{3}\sigma_\epsilon^2 * + \frac{T+1}{6}\sigma_\epsilon^2 *} = \frac{2}{3}. \qquad (11)$$

Consequently, the *ICC(1)* should overestimate the true amount of variance that exists between individuals with a value of approximately $0.6\bar{6}$.

Hypothesis 1: The ICC(1) obtained from the unconditional means model will be overestimated and indicate approximately 66% of the variance exists between individuals.

Model 2: Unconditional Growth Model

After determining that a significant amount of variance exists between individuals the next recommended step in RCM is to run the unconditional growth model specified as:

$$Y_{ij} = [\gamma_{00} + \gamma_{10} Time_{ij}] + [\zeta_{0i} + \zeta_{1i} Time_{ij} + \epsilon_{ij}]. \quad (12)$$

This model includes fixed effects for the intercept and slope of *Time*, γ_{00} and γ_{10} respectively, as well as random effects for the intercept and *Time*, ζ_{0i} and ζ_{1i} respectively, along with variance components for the random effects for the intercept and *Time*, σ_0^2 and σ_1^2 respectively, and a variance component for the residual error term, σ_ϵ^2. Following the logic and findings presented above, the hypotheses for the parameter estimates and significance tests for Model 2 are as follows:

Hypothesis 2a: The fixed effect of the intercept will be approximately zero regardless of the number of time points.

Hypothesis 2b: The fixed effect of the slope of Time will be approximately zero regardless of the number of time points.

Hypothesis 3a: The significant test for the fixed effect of the intercept will stay approximately at the nominal rate with increased time points.

Hypothesis 3b: The significant test for the fixed effect of the slope of Time will stay approximately at the nominal rate with increased time points.

Hypothesis 4a: The variance component for the random effect of the intercept will diverge toward infinity as the number of time points increase. The empirical value for any given number of time points will be approximately $\frac{2T}{15}$.

Hypothesis 4b: The variance component for the random effect of the slope of Time will converge toward zero as the number of time points increase. The empirical value for any given number of time points will be approximately $\frac{6}{5T}$.

Hypothesis 5a: The significance test on the variance component for the random effect for the intercept will be significant regardless of the number of time points.

Hypothesis 5b: The significance test on the variance component for the random effect for the slope of Time will be significant regardless of the number of time points.

Hypothesis 6: The χ^2 *significance test (also called the deviance statistic) will indicate that the unconditional growth model fits the data better than the unconditional means model.*

Model 3: Unconditional Growth Model with X as a Fixed Effect

After determining that the unconditional growth model fits the data better it is common for psychological researchers to add in predictors to explain any remaining unexplained within- or between-person variance. It is suggested by methodologists to first input the predictor variables into the model as only fixed effects (Singer & Willet, 2003). The predictor, *X*, in this model is entered as a Level 1, time-varying covariate. This model has the form:

$$Y_{ij} = [\gamma_{00} + \gamma_{10}Time_{ij} + \gamma_{20}X_{ij}] + [\zeta_{0i} + \zeta_{1i}Time_{ij} + \epsilon_{ij}]. \tag{13}$$

This model includes fixed effects for the intercept, slope of *Time*, and *X*, γ_{00}, γ_{10}, and γ_{20} respectively, as well as random effects for the intercept and *Time*, ζ_{0i} and ζ_{1i} respectively, along with variance components for the random effects for the intercept and *Time*, σ_0^2 and σ_1^2 respectively, and a variance component for the residual error term, σ_ϵ^2. The hypotheses for Model 3 resulting from the expectations derived above are:

Hypothesis 7: The parameter estimates for the fixed effects and the variance components of the random effects for the slope of Time and the intercept will be approximately the same as for the unconditional growth model.

Hypothesis 8: The significance tests for the fixed effects and the variance components of the random effects for the slope of Time and the intercept will follow the same patterns as for the unconditional growth model.

Hypothesis 9: The fixed effect for X will be approximately zero, regardless of the number of time points.

Hypothesis 10: The significance test for the fixed effect of X will diverge as the number of time points increases resulting in X being significant more frequently for higher numbers of time points.

Hypothesis 11: The χ^2 significance test between the unconditional growth model and the model with X as a fixed effect will be equivalent to the significance test on the fixed effect parameter for X and thus diverge as the number of time points increases resulting in the test being significant more frequently for higher numbers of time points.

Model 4: Unconditional Growth Model with X as a Fixed and Random Effect

After adding in *X* as only a fixed effect, the next recommended step is to allow *X* to vary and enter it into the model as a random effect as well. The model is:

$$Y_{ij} = [\gamma_{00} + \gamma_{10}Time_{ij} + \gamma_{20}X_{ij}] + [\zeta_{0i} + \zeta_{1i}Time_{ij} + \zeta_{2i}X_{ij} + \epsilon_{ij}]. \tag{14}$$

This model includes fixed effects for the intercept, slope of *Time*, and *X*, γ_{00}, γ_{10}, and γ_{20} respectively, as well as random effects for the intercept, *Time*, and *X*, ζ_{0i}, ζ_{1i}, and ζ_{2i}, respectively, along with variance components for the random effects for the intercept and *Time*, σ_0^2, σ_1^2, and σ_2^2 respectively, and a variance component for the residual error term, σ_ϵ^2.

Hypothesis 12: The parameter estimates for the fixed effects for the slope of Time and the intercept will be approximately the same as for the unconditional growth model and the model with X as only a fixed effect.

Hypothesis 13: The significance tests for the fixed effects for the slope of Time and the intercept will follow the same patterns as for the unconditional growth model and the model with X as only a fixed effect.

Hypothesis 14a: The variance component for the random effect for the intercept will diverge toward infinity as the number of time points increases. The empirical value for any given number of time points will be slightly less than $\frac{2T}{15}$.

Hypothesis 14b: The variance component for the random effect for the slope of Time will converge toward zero as the number of time points increases. The empirical value for any given number of time points will be slightly greater than $\frac{6}{5T}$.

Hypothesis 15: The fixed effect of X will be approximately zero, regardless of number of time points.

Hypothesis 16: The significant test for the fixed effect of X will stay approximately at the nominal rate with increased time points.

Hypothesis 17: The variance component for the random effect of X will converge to a value and approximately stay at that value for all additional number of time points.

Hypothesis 18: The significance tests on the variance components for the random effects for the intercept, slope of Time, and X will be significant.

Hypothesis 19: The χ^2 significance test between the model with X as a random effect and the model with X only as a fixed effect will be significant.

Model 5: Unconditional Means Model with X as a Fixed Effect

When testing the impact of Level 1 time-varying covariates some researchers opt to remove *Time* from the model after testing the unconditional growth model. While this decision may be theoretically valid, it does not remove the effects of spurious regression. Typically, what researchers who choose this route do is analyze the unconditional means model to assess the *ICC(1)*, then test the unconditional growth model to determine the average effects and trajectory of the growth processes. Researchers would then generally remove *Time* from the model and insert the predictor as only a fixed effect first, as was done in the previous models. The model typically takes the form:

$$Y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij}] + [\zeta_{0i} + \epsilon_{ij}]. \tag{15}$$

This model includes fixed effects for the intercept and *X*, γ_{00}, and γ_{10}, respectively, as well as a random effect for the intercept, ζ_{0i}, along with a variance component for the random effect for the intercept, σ_0^2, and a variance component for the residual error term,

σ_ϵ^2. This model is roughly equivalent to the third model presented that had *Time* as well as X predicting the dependent variable. Therefore, the hypotheses follow closely in line and are:

Hypothesis 20: The fixed effect of the intercept will be approximately zero regardless of the number of time points.

Hypothesis 21: The significant test for the fixed effect of the intercept will stay approximately at the nominal rate with increased time points.

Hypothesis 22: The variance component for the random effect for the intercept will diverge toward infinity as the number of time points increases. The empirical value for any given number of time points will be approximately $\frac{2T}{15}$.

Hypothesis 23: The fixed effect of X will be approximately zero, regardless of number of time points.

Hypothesis 24: The significance test for the fixed effect of X will diverge as the number of time points increases resulting in X being significant more frequently for higher numbers of time points.

Hypothesis 25: The χ^2 *significance test between the unconditional means model and the model with X as a fixed effect will be equivalent to the significance test on the fixed effect parameter for X and thus diverge as the number of time points increases resulting in the test being significant more frequently for higher numbers of time points.*

Model 6: Unconditional Means Model with X as a Fixed and Random Effect

After testing the addition of a predictor, *X*, as only a fixed effect it is common to allow the predictor to vary as a random effect as well. This model takes the form:

$$Y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij}] + [\zeta_{0i} + \zeta_{1i}X_{ij} + \epsilon_{ij}]. \quad (16)$$

This model includes fixed effects for the intercept and *X*, γ_{00} and γ_{10} respectively, as well as random effects for the intercept and *X*, ζ_{0i} and ζ_{1i} respectively, along with variance components for the random effects for the intercept and *X*, σ_0^2 and σ_1^2 respectively, and a variance component for the residual error term, σ_ϵ^2. The model is very similar in structure to that of the unconditional growth model and thus the hypotheses are:

Hypothesis 26: The fixed effect of the intercept will be approximately zero regardless of the number of time points.

Hypothesis 27: The significant test for the fixed effect of the intercept will stay approximately at the nominal rate with increased time points.

Hypothesis 28: The variance component for the random effect for the intercept will tend toward infinity as the number of time points increases. The empirical value for any given number of time points will be $\frac{2T}{15}$.

Hypothesis 29: The fixed effect of X will be approximately zero, regardless of number of time points.

Hypothesis 30: The significance test for the fixed effect of X will stay approximately at the nominal rate with increased time points.

Hypothesis 31: The variance component for the random effect for X will stay approximately constant as the number of time points increases.

Hypothesis 32: The significance tests on the variance components for random effects for the intercept and X will be significant.

Hypothesis 33: The χ^2 significance test between the model with X as a random effect and the model with X only as a fixed effect will be significant.

METHOD

To demonstrate the effects of stochastic trends on random coefficient modeling, Monte Carlo simulations with random walks was used. Since event/experience sampling methodology is becoming increasingly common and is the domain where most longer time series in psychology are observed, the simulation parameters mirrored those commonly found in event sampling. The two parameters that were manipulated are the sample size and the number of time points. Based on the reported review of event sampling studies (Table 2) the average number of participants is approximately 123 with a relatively large range from 26 up to 420 participants. While sample size is not hypothesized to affect the results due to stochastic trends it is important to account for the possibility that it could have an impact via statistical power. To simulate the range of possible sample sizes two values were analyzed. To evaluate effects of smaller samples the lower value for sample size was set at 30. To simulate larger samples that are more commonly found in the literature, the larger value for sample size was set at 150. For each sample size four different series lengths were simulated. The studies in Table 2 indicate that on average the number of time points observed is 47 with a range from 14 to 126 time points. To simulate this large range of potential lengths four values were analyzed for each sample size. These values were 5, 10, 40, and 70 time points. These values cover the range of commonly occurring series lengths in the literature as well as allow trends in the data over time to be observed. The overall design has two sample size conditions with four series length conditions for each sample size for a total of eight conditions (2 x 4).

Data for each condition was generated and analyzed in the statistical program, R. The dependent variable (*Y*) was be created by generating an independent random walk for each "participant" in the sample. The values that make up the dependent random walk were randomly sampled from a normal distribution with a mean of zero and a variance of one. The *Time* variable was created by having a linear increase from one to the maximum number of time points (5, 10, 40, or 70). To make the intercept value more interpretable, the *Time* variable had one subtracted from each value so it began at zero. For models 3 through 6 the predictor variable (*X*) was also created by generating independent random walks for each "participant" in the sample. The values that make up the predictor random walk were also randomly sampled from a normal distribution with a mean of zero and a variance of one. Each model was analyzed using the lmer function in the LME4 package in R. All simulation conditions were replicated 1000 times. The results presented for the fixed and random effects are the average values across the 1000 replications. The results presented for the significance tests for the fixed and random effects as well as the deviance statistics are the proportion of times across the 1000 replications that the null hypothesis was rejected. Since the random walks are independent and generated by taking successive random steps, no true predictive relationships can exist. Therefore, any relationship above the nominal rate ($\alpha = 0.05$) between the variables is solely due to faulty estimation caused by spurious regression.

RESULTS

Model 1: Unconditional Means Model

When researchers analyze the unconditional means model it is common to evaluate the fixed effect of the intercept to determine the grand mean. As seen in Table 3, the estimate for the fixed effect for the intercept was approximately zero in all eight conditions, thus being well behaved. Similarly, the rejection rate on the intercept fixed effect was also well behaved and was approximately nominal across all conditions. The unconditional means model is also often used to determine whether a significant amount of variance resides between means, commonly computed via a χ^2 test. Regardless of sample size, the between group variance increased as the number of time points increases. Across all conditions, the χ^2 statistic was significant in every replication indicating significant between mean variance. The most common statistic that is analyzed from the unconditional means model is the *ICC(1)*, which determines the percentage of variance that resides between groups or individuals. Consistent with hypothesis 1, the *ICC(1)* was systematically overestimated at 0.66 in all conditions. These results would lead researchers to incorrectly believe that 66% of the total variance resides between individuals and that the between group variance was significant, thus encouraging additional model testing.

Table 3. Model 1 - Unconditional Means Model

N	30				150			
T	5	10	40	70	5	10	40	70
	Fixed Effect Statistics							
$\bar{\gamma}_{00}$ $(S_{\gamma_{00}})$	-0.01 (0.28)	0.01 (0.36)	0.01 (0.67)	0.00 (0.89)	-0.01 (0.12)	0.00 (0.17)	0.01 (0.31)	-0.02 (0.39)
$\overline{SE}_{\gamma_{00}}$	0.06	0.05	0.05	0.05	0.04	0.05	0.05	0.04
$P_{\gamma_{00}}$	0.27	0.35	0.66	0.87	0.12	0.16	0.30	0.40
	Random Effect Statistics							
$\bar{\sigma}_0^2$	1.95	3.55	13.21	22.99	1.99	3.65	13.51	23.55
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_\epsilon^2$	1.01	1.83	6.82	11.87	1.00	1.83	6.82	11.86
ICC(1)	0.65	0.65	0.65	0.65	0.66	0.66	0.66	0.66

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of grand mean; $S_{\gamma_{00}}$ = standard deviation of grand mean; $\overline{SE}_{\gamma_{00}}$ = standard error of grand mean; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on grand mean; $\bar{\sigma}_0^2$ = average variance of individual means; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of individual means; $\bar{\sigma}_\epsilon^2$ = average within-person variance; *ICC(1)* = intraclass correlation coefficient.

Model 2: Unconditional Growth Model

When the unconditional growth model is fit to longitudinal data, a number of parameters are evaluated. As shown in Table 4, the standard errors for the fixed effects of both time and intercepts were accurately estimated relative to their respective standard deviations. Therefore, both fixed effects were accurately estimated to be approximately zero, with nominal rejection rates for all conditions, thus supporting hypotheses 2 and 3.

Table 4. Model 2 - Unconditional Growth Model - Fixed Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\gamma}_{00}$ $(S_{\gamma_{00}})$	-0.01 (0.20)	0.01 (0.24)	0.01 (0.44)	0.00 (0.56)	-0.00 (0.09)	-0.00 (0.11)	-0.00 (0.19)	-0.01 (0.25)
$\overline{SE}_{\gamma_{00}}$	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05
$P_{\gamma_{00}}$	0.20	0.24	0.43	0.55	0.09	0.11	0.19	0.25
$\bar{\gamma}_{10}$ $(S_{\gamma_{10}})$	0.00 (0.09)	0.00 (0.06)	0.00 (0.03)	0.00 (0.02)	0.00 (0.04)	0.00 (0.03)	0.00 (0.01)	0.00 (0.01)
$\overline{SE}_{\gamma_{10}}$	0.05	0.06	0.05	0.05	0.05	0.06	0.05	0.05
$P_{\gamma_{10}}$	0.09	0.06	0.03	0.02	0.04	0.03	0.01	0.01

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of population intercept; $S_{\gamma_{00}}$ = standard deviation of population intercept; $\overline{SE}_{\gamma_{00}}$ = standard error of population intercept; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on population intercept; $\bar{\gamma}_{10}$ = average estimate of population slope; $S_{\gamma_{10}}$ = standard deviation of population slope; $\overline{SE}_{\gamma_{10}}$ = standard error of population slope; $P_{\gamma_{10}}$ = rejection rate of hypothesis test on population slope.

Consistent with hypothesis 4a, the variance component of the random effect for the intercept increased as the number of time points increases (Table 5). Regardless of sample size, for any given number of time points the variance component of the intercept was approximately $\frac{2T}{15}$. The variance component for the random effect of *Time*, however, decreased with increased number of time points at a rate of approximately $\frac{6}{5T}$, thus supporting hypothesis 4b. For both the intercept and *Time*, the variance components for the random effects were significant in every replication across all eight conditions, consistent with hypothesis 5. Not surprisingly, since the random effects were significant across all conditions, the deviance statistic between the unconditional means model and unconditional growth model was also significant in every replication of every condition, thus confirming hypothesis 6. These results would mislead researchers to conclude that there was significance heterogeneity in the growth process and ultimately to search for predictors of the unexplained heterogeneity. Since in reality all data came from the same data generating mechanism, no true heterogeneity exists and all significant predictors would solely be due to Type I error.

Table 5. Model 2 - Unconditional Growth Model - Random Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\sigma}_0^2$	0.92	1.50	5.33	9.03	0.95	1.54	5.45	9.41
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_1^2$	0.21	0.11	0.03	0.02	0.21	0.11	0.03	0.02
$P_{\sigma_1^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_{01}^2$	0.20	0.00	-0.18	-0.20	0.17	-0.01	-0.19	-0.21
$P_{\sigma_{01}^2}$	0.14	0.06	0.19	0.21	0.34	0.06	0.64	0.75
$\bar{\sigma}_\epsilon^2$	0.47	0.81	2.80	4.82	0.47	0.80	2.80	4.80
$P_{\Delta_{\chi^2}}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note. *N* = Sample Size; *T* = series length; $\bar{\sigma}_0^2$ = average variance of intercepts; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of intercepts; $\bar{\sigma}_1^2$ = average variance of slopes; $P_{\sigma_1^2}$ = rejection rate of hypothesis test on variance of slopes; $\bar{\sigma}_{01}^2$ = average correlation between intercepts and slopes; $P_{\sigma_{01}^2}$ = rejection rate of hypothesis test on correlation between intercepts and slopes; $\bar{\sigma}_\epsilon^2$ = average within-person variance around

linear trajectories; $P_{\Delta_{\chi^2}}$ = rejection rate of the hypothesis test on the difference between model fit of unconditional means and unconditional growth models.

Model 3: Unconditional Growth Model with X as a Fixed Effect

If researchers choose to continue with additional model testing the next recommended step is to insert predictors as only fixed effects. Doing so did not affect the fixed effects or the variance components of the random effects for the intercept and *Time*. Likewise, the corresponding significance tests were approximately identical to those in model 2. These findings support hypotheses 7 and 8.

As shown in Table 6, the fixed effect for the time-varying predictor, *X*, was well behaved and was approximately zero, supporting hypothesis 9. Unlike the fixed effects of *Time* and the intercept, the standard error for the fixed effect of *X* is underestimated relative to its standard deviation. This results in the significance tests for the fixed effect of *X* exceeding the nominal rate (at $\alpha = 0.05$). Regardless of sample size, as the number of time points increased the standard errors were increasingly underestimated. This caused the significance test for the fixed effect of *X* to diverge, resulting in extremely inflated Type I error rates, consistent with hypothesis 10. Since the deviance test between this model and the unconditional growth model is a one degree of freedom test, it is approximately equal to the rejection rate of the fixed effect of *X*. Therefore, as the number of time points increases, the deviance statistic becomes significant a greater proportion of times (Table 7). Since no true predictive relationship exists between the criterion and the time-varying predictor, *X*, researchers are increasingly likely to make incorrect inferences as the number of time points collected increases. This is counter intuitive to most cases in psychology where having a greater number of time points is preferred because it provides researchers with more information and gives greater statistical power to detect true relationships. These findings would once again encourage researchers to continue on to test additional models despite the fact that all findings, current and future, are completely spurious.

Table 6. Model 3 - UGM with X as a Fixed Effect - Fixed Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\gamma}_{00}$ ($S_{\gamma_{00}}$)	-0.01 (0.20)	0.01 (0.24)	0.01 (0.44)	0.00 (0.56)	0.00 (0.09)	0.00 (0.11)	0.00 (0.19)	0.00 (0.25)
$\overline{SE}_{\gamma_{00}}$	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05
$P_{\gamma_{00}}$	0.20	0.24	0.43	0.55	0.09	0.11	0.19	0.25
$\bar{\gamma}_{10}$ ($S_{\gamma_{10}}$)	0.00 (0.09)	0.00 (0.06)	0.00 (0.03)	0.00 (0.02)	0.00 (0.04)	0.00 (0.03)	0.00 (0.01)	0.00 (0.01)
$\overline{SE}_{\gamma_{10}}$	0.06	0.05	0.05	0.05	0.04	0.06	0.05	0.05
$P_{\gamma_{10}}$	0.09	0.06	0.03	0.02	0.04	0.03	0.01	0.01
$\bar{\gamma}_{20}$ ($S_{\gamma_{20}}$)	0.00 (0.09)	0.00 (0.08)	0.00 (0.08)	0.00 (0.08)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)	0.00 (0.04)
$\overline{SE}_{\gamma_{20}}$	0.08	0.16	0.45	0.61	0.08	0.15	0.46	0.61
$P_{\gamma_{20}}$	0.08	0.06	0.03	0.02	0.04	0.03	0.01	0.01

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of population intercept; $S_{\gamma_{00}}$ = standard deviation of population intercept; $\overline{SE}_{\gamma_{00}}$ = standard error of population intercept; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on population intercept; $\bar{\gamma}_{10}$ = average estimate of population slope; $S_{\gamma_{10}}$ = standard deviation of population slope; $\overline{SE}_{\gamma_{10}}$ =

standard error of population slope; $P_{\gamma_{10}}$ = rejection rate of hypothesis test on population slope; ; $\bar{\gamma}_{20}$ = average estimate of *X*; $S_{\gamma_{20}}$ = standard deviation of *X*; $\overline{SE}_{\gamma_{20}}$ = standard error of *X*; $P_{\gamma_{20}}$ = rejection rate of hypothesis test on *X*.

Table 7. Model 3 - UGM with X as a Fixed Effect - Random Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\sigma}_0^2$	0.91	1.49	5.30	8.97	0.95	1.54	5.44	9.40
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_1^2$	0.21	0.11	0.03	0.02	0.21	0.11	0.03	0.02
$P_{\sigma_1^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_{01}^2$	0.20	0.00	-0.18	-0.20	0.17	-0.01	-0.19	-0.21
$P_{\sigma_{01}^2}$	0.14	0.06	0.18	0.21	0.33	0.07	0.63	0.74
$\bar{\sigma}_\epsilon^2$	0.46	0.80	2.78	4.79	0.47	0.80	2.80	4.79
$P_{\Delta_{\chi^2}}$	0.08	0.19	0.46	0.62	0.08	0.16	0.46	0.61

Note. *N* = Sample Size; *T* = series length; $\bar{\sigma}_0^2$ = average variance of intercepts; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of intercepts; $\bar{\sigma}_1^2$ = average variance of slopes; $P_{\sigma_1^2}$ = rejection rate of hypothesis test on variance of slopes; $\bar{\sigma}_{01}^2$ = average correlation between intercepts and slopes; $P_{\sigma_{01}^2}$ = rejection rate of hypothesis test on correlation between intercepts and slopes; $\bar{\sigma}_\epsilon^2$ = average within-person variance around linear trajectories; $P_{\Delta_{\chi^2}}$ = rejection rate of the hypothesis test on the difference between model fit of Model 2 and Model 3.

Model 4: Unconditional Growth Model with X as a Fixed and Random Effect

The addition of a time-varying predictor to the model as both a fixed and random effect had no effect on the fixed effect estimates of the intercept and *Time*. Table 8 shows that both parameters were still approximately zero in all conditions with approximately nominal rejection rates, confirming hypotheses 12 and 13. Likewise, the fixed effect of X stayed approximately zero for all conditions, supporting hypothesis 15. However, by allowing X to vary, the covariance structure of the model more closely approximates the covariance structure of the data, resulting in the accurate approximation of the standard error relative to the standard deviation for the fixed effect of X. Therefore, hypothesis 16 is supported because the rejection rates of the fixed effect of X are no longer inflated and are now approximately nominal regardless of the number of time points.

By adding an additional random effect to the model, the variance of the system is divided up further. Therefore, the patterns of the variance components of the random effects for the intercept and *Time* were slightly ameliorated (see Table 9). This caused the variance component of the random effect for the intercept to be slightly less than $\frac{2T}{15}$ and the variance component of the random effect for *Time* to be slightly greater than $\frac{6}{5T}$, supporting hypothesis 14. Despite the mathematical proof that showed that the variance component of the random effect of X should remain constant at some value (Durlauf & Phillips, 1988), the variance increased as the number of time points increases. Therefore, hypothesis 17 is not supported. It is possible that a greater number of time points would need to be observed before the variance stabilized.

The significance test for the variance components of the random effect for *Time* was significant for every replication, across all eight conditions. However, the significance tests for the variance components of the random effects for the intercept and *X* were dependent on both sample size and number of time points. The rejection rate for the variance component of the random effect for the intercept was one hundred percent in every condition except the one with the least information ($N = 30$; $T = 5$). Even in this condition the rejection rate was very high at 99.8%. However, for the variance component of the random effect for *X* the rejection rate varied considerably based on both sample size and number of time points. By increasing either sample size or number of time points the rejection rate became significant 100% of the time. However, in conditions of low information ($N = 30$; $T = 5$) the rejection rate was as low as 9.4%. Therefore, hypothesis 18 was only partially supported.

Similarly, the deviance statistic between models 3 and 4 was affected by the amount of information and followed the same pattern of results as the rejection rate of the variance component of the random effect for *X*, thus only partially supporting hypothesis 19. This result is not overly surprising, given the behavior of the rejection rates of the variance components of the random effects of *X*. The deviance statistic between these models is a three degree of freedom test to determine if any of the three new parameters are significantly different from zero. The two new correlations (between the intercept and *X* and between *X* and *Time*) had rejection rates slightly above the nominal rate (~0.08) so the test essentially became equivalent to the χ^2 test for the parameter estimate of the variance component for the random effect for *X*. The results here are worrisome because they would lead researchers to conclude that a relationship exists between the time-varying predictor and the criterion when in fact, no such relationship exists in the data generating mechanism. Also, these results are once again somewhat counter intuitive because as researchers gain greater amounts of information through both sample size and time points (generally thought of as a good thing), researchers are much more susceptible to spurious results and inferences.

Table 8. Model 4 - UGM with X as a Fixed and Random Effect - Fixed Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\gamma}_{00}$ $(S_{\gamma_{00}})$	-0.01 (0.20)	0.01 (0.24)	0.01 (0.41)	0.00 (0.52)	0.00 (0.09)	0.00 (0.10)	0.00 (0.18)	0.00 (0.24)
$\overline{SE}_{\gamma_{00}}$	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.06
$P_{\gamma_{00}}$	0.20	0.23	0.40	0.51	0.09	0.11	0.18	0.24
$\bar{\gamma}_{10}$ $(S_{\gamma_{10}})$	0.00 (0.10)	0.00 (0.07)	0.00 (0.03)	0.00 (0.02)	0.00 (0.04)	0.00 (0.03)	0.00 (0.01)	0.00 (0.01)
$\overline{SE}_{\gamma_{10}}$	0.06	0.05	0.04	0.05	0.04	0.06	0.06	0.05
$P_{\gamma_{10}}$	0.09	0.06	0.03	0.02	0.04	0.03	0.01	0.01
$\bar{\gamma}_{20}$ $(S_{\gamma_{20}})$	0.00 (0.09)	0.00 (0.08)	0.00 (0.08)	0.00 (0.08)	0.00 (0.04)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
$\overline{SE}_{\gamma_{20}}$	0.05	0.07	0.05	0.05	0.06	0.05	0.05	0.04
$P_{\gamma_{20}}$	0.09	0.08	0.07	0.08	0.04	0.04	0.03	0.03

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of population intercept; $S_{\gamma_{00}}$ = standard deviation of population intercept; $\overline{SE}_{\gamma_{00}}$ = standard error of population intercept; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on population intercept; $\bar{\gamma}_{10}$ = average estimate of population slope; $S_{\gamma_{10}}$ = standard deviation of population slope; $\overline{SE}_{\gamma_{10}}$ = standard error of population slope; $P_{\gamma_{10}}$ = rejection rate of hypothesis test on population slope; ; $\bar{\gamma}_{20}$ = average estimate of *X*; $S_{\gamma_{20}}$ = standard deviation of *X*; $\overline{SE}_{\gamma_{20}}$ = standard error of *X*; $P_{\gamma_{20}}$ = rejection rate of hypothesis test on *X*.

Table 9. Model 4 - UGM with X as a Fixed and Random Effect - Random Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\sigma}_0^2$	0.86	1.36	4.53	7.64	0.91	1.40	4.68	7.98
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_1^2$	0.20	0.11	0.03	0.02	0.21	0.11	0.03	0.02
$P_{\sigma_1^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_2^2$	0.05	0.08	0.14	0.16	0.04	0.08	0.14	0.16
$P_{\sigma_2^2}$	0.09	0.61	1.00	1.00	0.33	1.00	1.00	1.00
$\bar{\sigma}_{01}^2$	0.20	0.00	-0.17	-0.20	0.17	0.00	-0.18	-0.20
$P_{\sigma_{01}^2}$	0.13	0.06	0.18	0.20	0.32	0.07	0.54	0.69
$\bar{\sigma}_{02}^2$	-0.01	0.00	0.00	0.00	0.01	0.01	-0.01	0.00
$P_{\sigma_{02}^2}$	0.06	0.09	0.09	0.09	0.06	0.06	0.08	0.09
$\bar{\sigma}_{12}^2$	0.02	0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00
$P_{\sigma_{12}^2}$	0.06	0.07	0.09	0.08	0.06	0.06	0.08	0.08
$\bar{\sigma}_\epsilon^2$	0.44	0.73	2.40	4.08	0.45	0.73	2.41	4.07
$P_{\Delta_{\chi^2}}$	0.08	0.52	1.00	1.00	0.22	0.99	1.00	1.00

Note. *N* = Sample Size; *T* = series length; $\bar{\sigma}_0^2$ = average variance of intercepts; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of intercepts; $\bar{\sigma}_1^2$ = average variance of slopes; $P_{\sigma_1^2}$ = rejection rate of hypothesis test on variance of slopes; $\bar{\sigma}_2^2$ = average variance of X; $P_{\sigma_2^2}$ = rejection rate of hypothesis test on variance of X; $\bar{\sigma}_{01}^2$ = average correlation between intercepts and slopes; $P_{\sigma_{01}^2}$ = rejection rate of hypothesis test on correlation between intercepts and slopes; $\bar{\sigma}_{02}^2$ = average correlation between intercepts and X; $P_{\sigma_{02}^2}$ = rejection rate of hypothesis test on correlation between intercepts and X; $\bar{\sigma}_{12}^2$ = average correlation between slopes and X; $P_{\sigma_{12}^2}$ = rejection rate of hypothesis test on correlation between slopes and X; $\bar{\sigma}_{\epsilon}^2$ = average within-person variance around linear trajectories; $P_{\Delta_{\chi^2}}$ = rejection rate of the hypothesis test on the difference between model fit of Model 3 and Model 4.

Model 5: Unconditional Means Model with X as a Fixed Effect

It is common for researchers to exclude the variable *Time* from RCM models when using time-varying covariates as predictors. The addition of a time-varying predictor did not alter the estimates of the fixed effect or variance component of the random effect for the intercept. Likewise, the significance tests on these components behaved identically to those in the unconditional means model. Therefore, hypotheses 20 through 22 were supported. The primary focus of this model is the fixed effect estimate for the time-varying covariate, X. As has been the case in all of the models, the fixed effect is accurately estimated to be approximately zero in all experimental conditions, supporting hypothesis 23 (Table 10). Consistent with the rationale given earlier, the significance test on the fixed effect of X diverged as the number of time points increased. This caused the rejection rate to steadily increase as the number of time points increased. Therefore, hypothesis 24 was supported. Only one parameter was added beyond the unconditional means model, causing the deviance statistic to become a significance test on that parameter. Consistent with the single parameter significance test, the deviance statistic diverged as the number of time points increased leading the rejection rate to increase with an increased number of time points, supporting hypothesis 25.

Table 10. Model 5 - Unconditional Means Model with X as a Fixed Effect

N	30				150			
T	5	10	40	70	5	10	40	70
	Fixed Effect Statistics							
$\bar{\gamma}_{00}$ $(S_{\gamma_{00}})$	-0.01 (0.28)	0.01 (0.36)	0.01 (0.67)	0.00 (0.89)	-0.01 (0.12)	0.00 (0.17)	0.01 (0.31)	-0.02 (0.39)
$\overline{SE}_{\gamma_{00}}$	0.06	0.05	0.05	0.05	0.04	0.06	0.05	0.04
$P_{\gamma_{00}}$	0.26	0.35	0.66	0.87	0.12	0.16	0.30	0.40
$\bar{\gamma}_{10}$ $(S_{\gamma_{10}})$	0.00 (0.12)	0.00 (0.12)	-0.01 (0.11)	0.00 (0.11)	0.00 (0.06)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
$\overline{SE}_{\gamma_{10}}$	0.17	0.32	0.57	0.69	0.19	0.30	0.60	0.73
$P_{\gamma_{10}}$	0.08	0.06	0.03	0.02	0.04	0.03	0.01	0.01
	Random Effect Statistics							
$\bar{\sigma}_0^2$	1.92	3.51	13.06	22.73	1.98	3.64	13.48	23.49
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_\epsilon^2$	0.99	1.81	6.74	11.72	1.00	1.83	6.80	11.83
$P_{\Delta\chi^2}$	0.18	0.33	0.59	0.71	0.19	0.31	0.60	0.73

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of population intercept; $S_{\gamma_{00}}$ = standard deviation of population intercept; $\overline{SE}_{\gamma_{00}}$ = standard error of population intercept; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on population intercept; $\bar{\gamma}_{10}$ = average estimate of X; $S_{\gamma_{10}}$ = standard deviation of X; $\overline{SE}_{\gamma_{10}}$ = standard error of X; $P_{\gamma_{10}}$ = rejection rate of hypothesis test on X; $\bar{\sigma}_0^2$ = average variance of intercepts; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of intercepts; $\bar{\sigma}_\epsilon^2$ = average within-person

variance around linear trajectories; $P_{\Delta_{\chi^2}}$ = rejection rate of the hypothesis test on the difference between model fit of unconditional means model and Model 5.

Model 6: Unconditional Growth Model with X as a Fixed and Random Effect

The final model that was tested is a model where a time-varying predictor, *X*, is inserted as both a fixed and random effect and the variable Time is excluded. This model takes on the functional form of the unconditional growth model. Table 11 shows that the fixed effects for both the intercept and *X* are well behaved and estimated to be approximately zero, supporting hypotheses 26 and 29. Similarly, hypotheses 27 and 30 were supported because the standard errors of those fixed effects closely approximate the corresponding standard deviations leading the significance tests to remain approximately nominal.

As seen in Table 12, the variance component of the random effect of the intercept increased as the number of time points increases. However, the value did not converge to the function $\frac{2T}{15}$. Instead, the value was estimated to be greater in every model. Therefore, hypothesis 28 was only partially supported. The significance test for the variance component of the random effect of the intercept was significant in every replication of every condition in support of hypothesis 32. The variance component of the random effect for *X* also increased as the number of time points increased. This is contrary to the expectations laid out from Durlauf and Phillips (1988). This finding is consistent however, with the results for Model 4 where X is included as a fixed and random effect in addition to the variable *Time*. Therefore, hypothesis 31 was not supported. The significance test for the variance component of the random effect of *X* was significant the majority of the time. In the conditions with few people and few time points (N = 30; T = 5, 10) the rejection rate were not quite at one hundred percent (72% and 99% respectively). Therefore, hypothesis 32 was only partially supported.

The deviance statistic between Models 5 and 6 tests whether the addition of the variance component of the random effect for X and the correlation between X and intercepts is significant. The correlation is significant only slightly more than the nominal rate, so the primary statistic being analyzed is the variance component of the random effect. Not surprisingly then, the rejection rate of the deviance statistic follows the same pattern as the rejection rate of the variance component of the random effect for X. For conditions with both few people and few time points (N = 30; T = 5, 10) the rejection rate is slightly less than one hundred percent (68% and 99% respectively), partially supporting hypothesis 33.

Table 11. Model 6 - UMM with X as a Fixed and Random Effect - Fixed Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\gamma}_{00}$ $(S_{\gamma_{00}})$	0.00 (0.27)	0.01 (0.32)	0.01 (0.60)	0.00 (0.74)	-0.01 (0.11)	0.00 (0.15)	0.00 (0.26)	-0.02 (0.34)
$\overline{SE}_{\gamma_{00}}$	0.06	0.05	0.05	0.04	0.04	0.05	0.05	0.05
$P_{\gamma_{00}}$	0.25	0.31	0.56	0.74	0.11	0.14	0.26	0.33
$\bar{\gamma}_{10}$ $(S_{\gamma_{10}})$	0.00 (0.11)	0.00 (0.11)	0.00 (0.11)	0.00 (0.11)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
$\overline{SE}_{\gamma_{10}}$	0.06	0.05	0.05	0.05	0.04	0.03	0.05	0.06
$P_{\gamma_{10}}$	0.12	0.11	0.11	0.11	0.05	0.05	0.05	0.05

Note. N = Sample Size; T = series length; $\bar{\gamma}_{00}$ = average estimate of population intercept; $S_{\gamma_{00}}$ = standard deviation of population intercept; $\overline{SE}_{\gamma_{00}}$ = standard error of population intercept; $P_{\gamma_{00}}$ = rejection rate of hypothesis test on population intercept; $\bar{\gamma}_{10}$ = average estimate of X; $S_{\gamma_{10}}$ = standard deviation of X; $\overline{SE}_{\gamma_{10}}$ = standard error of X; $P_{\gamma_{10}}$ = rejection rate of hypothesis test on X.

Table 12. Model 6 - UMM with X as a Fixed and Random Effect - Random Effect Statistics

N	30				150			
T	5	10	40	70	5	10	40	70
$\bar{\sigma}_0^2$	1.58	2.69	9.38	16.28	1.62	2.77	9.66	16.51
$P_{\sigma_0^2}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\bar{\sigma}_1^2$	0.20	0.27	0.35	0.36	0.20	0.28	0.35	0.37
$P_{\sigma_1^2}$	0.82	1.40	4.87	8.42	0.82	1.39	4.87	8.34
$\bar{\sigma}_{01}^2$	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$P_{\sigma_{01}^2}$	0.10	0.11	0.14	0.10	0.10	0.10	0.10	0.11
$\bar{\sigma}_\epsilon^2$	0.72	0.99	1.00	1.00	1.00	1.00	1.00	1.00
$P_{\Delta_{\chi^2}}$	0.68	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Note. N = Sample Size; T = series length; $\bar{\sigma}_0^2$ = average variance of intercepts; $P_{\sigma_0^2}$ = rejection rate of hypothesis test on variance of intercepts; $\bar{\sigma}_1^2$ = average variance of X; $P_{\sigma_1^2}$ = rejection rate of hypothesis test on variance of X; $\bar{\sigma}_{01}^2$ = average correlation between intercepts and X; $P_{\sigma_{01}^2}$ = rejection rate of hypothesis test on correlation between intercepts and X; $\bar{\sigma}_\epsilon^2$ = average within-person variance around linear trajectories; $P_{\Delta_{\chi^2}}$ = rejection rate of the hypothesis test on the difference between model fit of Model 5 and Model 6.

General Findings

Across the six models and eight conditions a number of distinct patterns emerged. The fixed effects estimates were well behaved and accurately estimated to be approximately zero in all cases. When a corresponding random effect was in the model, the fixed effect significance tests were also well behaved and stayed approximately at the nominal rate. If a corresponding random effect was not in the model, however, the fixed effects significance tests diverged with increased number of time points, leading to a serious inflation of Type I error rates and rejection rates well beyond the nominal level (upwards of 70%).

The variance components of the random effects also followed very distinct patterns across all the models and conditions. The variance component of the intercept increased as time points increased for every model. Similarly, the variance component of the random effect for *Time* decreased as the number of time points increased for every model. The variance component of the random effect of X slightly increased as the number of time points increased, contrary to the rationale derived from the distributional theory of normal regression parameters. When a large amount of information was present (i.e. large number of people or time points) the variance components were always significant. For a few models, when information was low (small N and T), the rejection rates of the variance components dropped slightly below one hundred percent but were still much higher than the nominal level ($\alpha = 0.05$).

The deviance statistics closely followed the rationale presented earlier. When the only difference between the models was the addition of a fixed effect, the deviance statistic became another significance test on that parameter; as the number of time points increased, the rejection rate between the models increased toward one hundred percent. When the difference between models was multiple parameters resulting from the addition of random effects, the deviance statistics closely mirrored the rejection rates of the inserted random effect. The correlations were often small and non-significant so once again, the deviance statistic essentially became a single parameter test for the variance component of the random effect. As such, in some cases when low amounts of information were present (small *N* or *T*) the rejection rate dropped slightly below one hundred percent. When information was high, however, (large *N* or *T*) the rejection rates between models were always one hundred percent.

DISCUSSION

Recommendations

The effects of spurious regression in RCM due to stochastic data are quite dramatic and could easily lead psychological researchers to make faulty statistical and scientific inferences. It is still unclear as to the prevalence of stochastic processes that exist in psychology. As mentioned earlier, random walks and stochastic processes are quite common in other fields (e.g., economics, computer sciences, and physics). Therefore, it is likely that similar processes exist in the psychological sciences as well. In a presentation by Kujanin, Braun, and DeShon (2009) two processes, one cognitive and one behavioral, were identified as being stochastic. They found that relative team performance (studied through NBA basketball teams) was found to be indistinguishable from a random walk with drift. Similarly, they found that cognitive perceptions of confidence ratings over time followed a random walk. It is imperative that scientists discover other processes that could be stochastic to avoid making the statistical and scientific inferences that result from analyzing stochastic data with RCM or other regression models.

The regression model and generalizations of that model make a number of assumptions about the dependent variable. One of those assumptions is that all trends present in the dependent variable are purely deterministic. If the dependent variable contains stochastic trends then researchers are extremely susceptible to the statistical and inferential mistakes caused by spurious regression as highlighted by this paper. Alternatively, if only deterministic trends are present in the dependent variable, then

regardless of the trends present in the predictors, regression models are well behaved. It is therefore the recommendation of this thesis that psychologists test the dependent variable of their longitudinal data for the presence of stochastic trends before applying random coefficient models or other regression models to the data.

Fortunately, methods have been created to distinguish between trajectories with purely deterministic trends (e.g., $Y_t = \alpha + \beta t + \epsilon_t$) and trajectories that are caused at least in part by stochastic trends (e.g., $Y_t = Y_0 + \beta t + \sum_{i=1}^{t} \epsilon_t$). The most commonly used test to distinguish between deterministic and stochastic trends is the Augmented Dickey Fuller (ADF) test (Dickey & Fuller, 1979; Said & Dickey, 1984). For a single time series, the ADF is

$$\Delta Y_t = \alpha + \gamma Y_{t-1} + \sum_{i=1}^{p} \delta_p \Delta Y_{t-p} + \epsilon_t, \tag{17}$$

where Y_t is the series, $\Delta Y_t = Y_t - Y_{t-1}$, α is the drift, p is the lag order of the autoregressive process, δ_p are the structural autoregressive effects, and ϵ_t is the error term. The null hypothesis associated with this test is, $\gamma = 0$, says that a series is stochastic and indistinguishable from a random walk. If the null hypothesis is rejected, then the series is distinguishable from a random walk and thus deterministic. To run the test, the analyst needs to determine the lag structure of the time series and if there is a drift. The ADF is not a high power test and estimating unnecessary parameters for long lags and drift wastes degrees of freedom. The lag structure of a time series is investigated by looking at the autocorrelation and partial autocorrelation functions while the existence of drift in the series is generally assessed visually. The free statistical software R includes the A

autoc

Psych

is nec

from

autoc

comm

the m

(e.g.,

devel

each

com

(i.e.

seri

util

if th

any

spuri

the con

with *stoc*

the ADF in its set of analytical techniques as well as the autocorrelation and partial autocorrelation functions.

The standard ADF test is used to evaluate the trends present in a single trajectory. Psychologists generally gather data on several individuals and a panel version of the ADF is needed to determine whether the sample of trajectories as a whole is distinguishable from multiple random walks. To use the test it is necessary to examine the autocorrelation and partial autocorrelation functions of each series to determine the most common lag structure across the sample and to determine whether drift exists in at least the majority of the series. This process is described in most introductory time series texts (e.g., Enders, 2004). Once these decisions are made, the panel version of the ADF test developed by Im, Pesaran, and Shin (2003) is computed by applying the standard ADF on each series and then taking the average value of $\hat{\gamma}$ in Equation 17. This average value is compared to a percentile (e.g., 90th or 95th) from the distribution of estimated unit roots (i.e., γ) on random walks for the specified lag order, drift, length, and number of time series (see Im, Pesaran, & Shin, 2003).

If the null hypothesis of the ADF is not rejected for the dependent variable, then utilizing RCM or other generalizations of the regression model is appropriate. However, if the null hypothesis is rejected then researchers are encouraged not to use regression or any generalizations of the regression model, such as RCM, because the estimates will be spurious.

A number of alternative models have been developed that can handle data with the complex properties of stochastic trends. The two most common approaches to dealing with stochastic data are the use of autoregressive-integrated-moving average (ARIMA) models and structural time series models (Harvey, 1989, 1997). ARIMA models difference the data until the resulting trends are solely deterministic. Then, they allow researchers to input variables to predict the deterministic component. This method is often criticized because by differencing the data many times the resulting deterministic trend may be nothing like the original data. This leaves the door open for researchers to make inferential mistakes by determining relationships between types of data that do not exist in the real world. Structural time series models on the other hand, allow researchers to insert a stochastic component into the model to filter the stochastic trend out while simultaneously allowing for predictors of the deterministic trend to be added. This allows researchers to more accurately model the true underlying process by accounting for the stochastic trend and then predicting any remaining deterministic trends.

Application to Latent Growth Models

All of the examples and simulations in this thesis were done using random coefficient modeling. Another common method used to analyze longitudinal data and answer questions about growth processes is to use a structural equation modeling approach called latent growth modeling. To the extent that latent growth models mirror the structure and properties of random coefficient models, they suffer from the same estimation problems of spurious regression due to the presence of stochastic data. More specifically, the fixed and random effect estimates are identical and the corresponding significance tests exhibit the same patterns as documented in RCM. Also, the chi-squared tests on the latent growth models become exponentially larger with increased number of time points leading to 100% rejection rates for significance tests between nested models and on individual parameters (e.g., variance components).

The use of latent growth models is often seen as having one large benefit over RCM in that global fit indices may be used to evaluate overall model fit. Three of the most common global fit indices that are used are the model chi-squared, the Comparative Fit Index (CFI), and the Root Mean Squared Error of Approximation (RMSEA) (Hu & Bentler, 1999). Unfortunately, these fit indices are also biased and behave in systematic ways when analyzing longitudinal data with stochastic trends. The model chi-squared is heavily biased by the presence of stochastic data and increases toward infinity with increased time points, thus always indicating poor model fit. This occurs for the same reason the chi-squared tests on the variance components for the random effects in RCM diverged and always indicate significance. Both the CFI and RMSEA are functions of the model chi-squared and thus exhibit similar distinct patterns as the number of time points increases. The CFI is given by the equation (Kline, 2005):

$$\mathrm{CFI} = 1 - \frac{\hat{\delta}_M}{\hat{\delta}_B}, \tag{18}$$

where

$$\hat{\delta}_M = \max\left(\chi^2_M - df_M\right).$$

Therefore, as the model chi-squared goes toward infinity, so does $\hat{\delta}_M$. As this happens the numerator in the second part of Equation 18 rapidly dominates the denominator, leading to the second part of the equation getting large with increased numbers of time points. This causes the estimate of CFI to go toward zero as the number of time points increases, thus indicating poor model fit. Similarly, the equation of RMSEA is (Kline, 2005):

$$\text{RMSEA} = \sqrt{\frac{\hat{\delta}_M}{df_M(N-1)}}, \tag{19}$$

meaning with increased time points $\hat{\delta}_M$ will increase, leading the RMSEA to also increase, indicating worse model fit. As a result, as the number of time points increase, the model chi-squared, CFI, and RMSEA will always indicate poor model fit. However, all of the variance components will indicate a significance amount of variance yet to be explained, encouraging researchers to search for predictors.

While the fact that the global fit indices indicate poor model fit may be seen as an advantage, it does not actually alleviate any of the problems caused by analyzing stochastic data. If researchers attempt to explain additional variance or improve model fit by adding predictors they will be engaging in the proverbial snipe hunt because no true predictive relationships exist, and any significant predictors that are found would be due solely to Type I error. Also, the indication of poor model fit does not address the larger issue of identifying and understanding the true underlying structure of the data. If the true data generating mechanism has a stochastic trend, then it should be modeled accordingly using one of the methods described above. To understand the phenomenon of interest and make correct statistical and scientific inferences, it is important to test and determine the structure of the data before deciding on an appropriate analytical tool.

Limitations and Future Directions

Despite the shocking nature of the results of this paper, a number of limitations exist. The Augmented Dickey-Fuller (ADF) test is low power test. Therefore, for the ADF to properly distinguish between deterministic and stochastic trends, a relatively large number of time points is needed. If researchers do not have a sufficient number of time points, the test will always fail to reject the null hypothesis, thus indicating that all series are stochastic. Similarly, the previously mentioned methods to handle stochastic data, such as ARIMA, require a large number of time points to function properly. Therefore, the recommendations of this paper most directly apply to event/experience sampling methodologies where a large number of time points is frequently collected. Unfortunately, since the presence of stochastic trends can impact statistical and scientific inferences with as few as five time points, all longitudinal studies need to worry about the results. More research is needed to determine how many time points is sufficient to run the ADF and to apply ARIMA or structural time series models to longitudinal data. More research is also needed to determine the exact benefits and limitations of both ARIMA and structural time series models compared to RCM and other commonly used analytical techniques in psychology.

Just as more research needs to be done on the proposed data analysis methods, more research is needed to identify stochastic processes in psychology. One possible method of discovering stochastic series is by analyzing previously published event/experience sampling studies and testing the data for the presence of stochastic trends. Finally, more work also needs to be done to determine how much relative influence stochastic processes need to have before their presence is problematic.

Although more work is needed to determine the exact scope of the problem and what the best solution is, this paper does provide a good start for recognizing the potential presence of stochastic trends and the problems that can arise from their presence.

Conclusion

The underlying structure of longitudinal data can have drastic impacts on the accuracy and quality of statistical estimates and scientific inferences. It is imperative that researchers determine the nature and structure of their longitudinal data before determining what analytic method to use. The first step in running any longitudinal data analysis should be using the Augmented Dickey Fuller (ADF) or some other statistical test to identify whether stochastic trends are present in the dependent variable. Then, after determining the underlying structure of the data, an appropriate analytic technique may be chosen. For data that contain only deterministic trends, random coefficient models or any other generalization of the regression model is appropriate. However, if stochastic trends are present in the data, then more complex analytical methods (e.g., ARIMA or structural time series) must be used to avoid making the statistical and inferential mistakes that are caused by spurious regression.

APPENDIX

APPENDIX A

Standard Errors as Approximations of Standard Deviations of the Sampling Distributions of the Fixed Effects

Durlauf and Phillips (1988) derive the theoretical sampling distribution of intercepts and slopes for a simple regression (i.e., $Y_t = \alpha + \beta t + \epsilon_t$) applied to data generated from an underlying random walk process. In the unconditional growth model, the standard deviations of the sampling distributions for the fixed effects (i.e., γ_{00} and γ_{10}) are approximately equal to the standard deviations from Theorem 2.1 of Durlauf and Phillips (1988) divided by the sample size, N. Thus, the standard deviation of the intercept (i.e., γ_{00}) sampling distribution is:

$$\sigma_0 = \sqrt{\frac{2T}{15N}\sigma_\epsilon^2}, \tag{20}$$

where T is number of time points for each series. The standard deviation of the slope (i.e., γ_{10}) sampling distribution is:

$$\sigma_0 = \sqrt{\frac{6}{5TN}\sigma_\epsilon^2}. \tag{21}$$

The standard errors of the fixed effects computed in the unconditional growth model closely approximate these standard deviations. The standard errors are given by Raudenbush and Bryk (2002) Equations 3.32 and 3.33

$$\boldsymbol{SE} = \sqrt{\sum_{j=1}^{N}(\boldsymbol{W'}_j \boldsymbol{\Delta}_j^{-1} \boldsymbol{W}_j)^{-1}}, \tag{22}$$

where j represents individuals and $\boldsymbol{W}_j$ is a matrix of level 2 predictors for each individual., $\boldsymbol{\Delta}_j = \boldsymbol{S} + \sigma_\varepsilon^2(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}$, is the variance-covariance matrix of the random effects (i.e., intercepts and slopes), $\boldsymbol{X}_j$ is the matrix of level 1 predictors, and σ_ε^2 is the error variance.. The diagonal elements of Equation 22 are needed to compute the standard errors of the fixed effects. Durlauf and Phillips (1988; Theorem 2.1) provide an approximation of the diagonal elements (i.e., variance components) of **S**, and the diagonal elements of the inverse of $\boldsymbol{X}_j'\boldsymbol{X}_j$ are derivable when *Time* is the only predictor. Nelson and Kang (1984; Equation 2.9) provide the needed approximation of the error variance.

Using the previously mentioned approximations, the standard error of γ_{00} is

$$SE_{\gamma_{00}} = \sqrt{\left(\left(\frac{2T}{15} + \left(\left(\frac{T+1}{15}\right)\left(\frac{4T+2}{T(T-1)}\right)\right)\right)^{-1} N\right)^{-1}}. \tag{23}$$

As T increases, the second term in Equation 23 is rapidly dominated by the first term. Therefore, Equation 23 closely approximates Equation 20.

Using the previously mentioned approximations, the standard error of the slope fixed effect is,

$$SE_{\gamma_{10}} = \sqrt{\left(\left(\frac{6}{5T} + \left(\left(\frac{T+1}{15}\right)\left(\frac{12}{T^3-T}\right)\right)\right)^{-1} N\right)^{-1}}. \tag{24}$$

As T increases, the second term in Equation 24 is rapidly dominated by the first term.

Therefore, Equation 24 closely approximates Equation 21.

REFERENCES

REFERENCES

Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review, 26*, 512-529.

Baldwin, S. A., & Hoffman, J. P. (2002). The dynamics of self-esteem: A growth curve analysis. *Journal of Youth and Adolescence, 31*, 101-113.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics, 6*, 267-285.

Bhargava, A., & Sargan, J. D. (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica, 51*, 1635-1659.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117*, 187-215.

Bohnert, A. M., Richards, M. H., Kolmodin, K. E., & Lakin, B. L. (2008). Young urban african american adolescents experience of discretionary time activities. *Journal of Research on Adolescence, 18*, 517539.

Bono, J. E., Foldes, H. J., Vinson, G., & Muros, P. (2007). Workplace emotions: The role of supervision and leadership. *Journal of Applied Psychology, 92*, 1357-1367.

Brown, L. H., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). When the need to belong goes wrong the expression of social anhedonia and social anxiety in daily life. *Psychological Science, 18*, 778-782.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review, 100*, 432-459.

Collins, L. M., & Horn, J. L. (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, D.C: American Psychological Association.

Collins, L. M., & Sayer, A. G. (2001). New *methods for the analysis of change*. Washington, D.C.: American Psychological Association.

Curran, P. J., Wirth, R. J., Nesselroade, J. R., Rogosa, D., Thum, Y. M., Tuerlinckx, F., & von Eye, A. (2004). Interindividual differences in intraindividual variation: balancing internal and external validity. *Measurement: Interdisciplinary Research & Perspective, 2*, 219-247.

DeHart, T., Tennen, H., Armeli, S., Todd, M., & Affleck, G. (2008). Drinking to regulate negative romantic relationship interactions: The moderating role of self-esteem. *Journal of Experimental Social Psychology, 44*, 527-538.

Dickey, D. A., & Fuller, W. A. (1979). Distribution for the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 44*, 427-431.

Durlauf, S. N., & Phillips, P. C. . B. (1988). Trends versus random walks in time series analysis. *Econometrica, 56*, 1333-1354.

Eliashberg, J., & Chatterjee, R. (1986). Stochastic issues in innovation diffusion models. In V. Mahajan & Y. Wind (Eds.), *Innovation diffusion models of new product acceptance* (p. 151- 199). Massachusetts: Nallinger Publishing Company.

Enders, W. (1995). *Applied econometric time series*. New York: John Wiley & Sons.

Enders, W. (2004). *Applied econometric time series* (2nd ed.). Hoboken, NJ: Wiley.

Ernst, J., Nau, G. J., & Bar-Joseph, Z. (2005). Clustering short time series gene expression data. *Bioinformatics, 21*, 159-168.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. Journal *of Personality, 75*, 825-862.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: University Press.

George, J. M., & Jones, G. R. (2000). The role of time in theory and theory building. *Journal of Management, 26*, 657-684.

Gerstein, G., & Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal, 4*, 41-68.

Graham, J. M. (2008). Self-expansion and flow in couples momentary experiences: An experience sampling study. *Journal of Personality and Social Psychology, 95*, 679-694.

Granger, C. W. J., & Newbold, P. (1974). Spurious regression in econometrics. *Journal of Econometrics, 2*, 111-120.

Grimm, K. J. (2007). Multivariate longitudinal methods for studying developmental relationships between depression and academic achievement. *International Journal of Behavioral Development, 31*, 328-339.

Harford, T. C., & Muthen, B. O. (2001). Alcohol use among college students: The effects of prior problem behaviors and change of residence. *Journal of the Study of Alcohol, 62*, 306-312.

Harris, C. W. (1963). *Problems in measuring change.* Madison, WI: University of Wisconsin Press.

Harvey, A. (1989). *Forecasting, structural time series models, and the kalman filter.* Cambridge, UK: Cambridge University Press.

Harvey, A. C. (1997). Trends, cylces, and autoregressions. *The Economic Journal, 107*, 192-201.

Hoffmann, J. P., Cerbone, F. G., & Su, S. S. (2000). A growth curve analysis of stress and adolescent drug use. *Substance Use & Misuse, 35*, 687-716.

Hogarth, R. M., Portell, M., & Cuxart, A. (2007). What risks do people perceive in everyday life? a perspective gained from the experience sampling method (esm). *Risk Analysis, 27*, 1427-1439.

Hsaio, C., Pesaran, M. H., & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics, 109*, 107-150.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis:conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Ilies, R., Schwind, K. M., Wagner, D. T., Johnson, M. D., DeRue, D. S., & Ilgen, D. R. (2007). When can employees have a family life? The effects of daily workload and affect on work family conflict and social behaviors at home. *Journal of Applied Psychology, 92*, 1368-1379.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*, 53-74.

Impett, E. A., Strachman, A., Finkel, E. J., & Gable, S. L. (2008). Maintaining sexual desire in intimate relationships: The importance of approach goals. *Journal of Personality and Social Psychology, 94*, 808-823.

Jang, Y., Clay, O. J., Roth, D. L., Haley, W. E., & Mittelman, M. S. (2004). Neuroticism and longitudinal change in caregiver depression: Impact of a spouse-caregiver intervention program. *The Gerontologist, 44*, 311-317.

Johnson, R. A., Hoffmann, J. P., Su, S. S., & Gerstin, D. R. (1997). Growth curves of deviant behavior in early adolescence: A multilevel analysis. *Journal of Quantitative Criminology, 13*, 429-467.

Jones, F., OConnor, D. B., Conner, M., McMillan, B., & Ferguson, E. (2007). Impact of daily mood, work hours, and iso-strain variables on self-reported health behaviors. *Journal of Applied Psychology, 92*, 1731-1740.

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when an experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*, 614-621.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest, 7*, 77-124.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (p. 3-90). San Francisco: Jossey-Bass.

Kreft, I., & De Leeuw, J. (1988). *Introducing multilevel modeling*. London: Sage Publications.

Kubiak, T., Vošgele, C., Siering, M., Schiel, R., & Weber, H. (2008). Daily hassles and emotional eating in obese adolescents under restricted dietary conditions - the role of ruminative thinking. *Appetite, 51*, 206-209.

Kuljanin, G., Braun, M. T., & DeShon, R. P. (2009, April). A cautionary note on applying multilevel models to longitudinal data. New Orleans, LA.

Kuppens, P. (2007). Individual differences in the relationship between pleasure and arousal. *Journal of Research in Personality, 42*, 10531059.

Lucas, R. E., Le, K., & Dyrenforth, P. S. (2008). Explaining the extraversion/positive affect relation: Sociability cannot account for extraverts greater happiness. *Journal of Personality, 76*, 385-413.

McGrath, J. E., & Rotchford, N. L. (1983). Time and behavior in organizations. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 5, p. 57-101). Greenwich, CT: JAI Press.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review, 26*, 530-547.

Moberly, N. J., & Watkins, E. R. (2008). Ruminative self-focus and negative affect: An experience sampling study. *Journal of Abnormal Psychology, 114*, 314-323.

Moghaddam, N. G., & Ferguson, E. (2007). Smoking, mood regulation, and personality: An event-sampling exploration of potential models and moderation. *Journal of Personality, 75*, 451-478.

Mohammed, S., Hamilton, K., & Lim, A. (In Press). The incorporation of time in team research: Past, current, and future. In E. Salas, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches*. Mahwah, NJ: LEA.

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.

Moskowitz, D. S., & Hershberger, S. L. (2002). *Modeling intraindividual variability with repeated measures data: Methods and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Nelson, C. R., & Kang, H. (1984). Pitfalls in the use of time as an explanatory variable in regression. *Journal of Business and Economic Statistics, 2*, 73-82.

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics, 10*, 139-162.

Nezlek, J. B., Vansteelandt, K., Mechelen, I. V., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion, 8*, 145-150.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*, 266-300.

Oishi, S., Diener, E., Choi, D., Kim-Prieto, C., & Choi, I. (2007). The dynamics of daily events and well-being across cultures: When less is more. *Journal of Personality and Social Psychology, 93*, 685-698.

Phillips, P. (1987). Time series regression with a unit root. *Econometrica, 55*, 277-301.

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. Journal of *Econometrics, 33*, 311-340.

Piasecki, T. M., Richardson, A. E., & Smith, S. M. (2007). Self-monitored motives for smoking among college students. *Psychology of Addictive Behaviors, 3*, 328-337.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and *data analysis methods*. Newbury Park, CA: Sage Publications.

Raudenbush, S. W., & Chan, W. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Counseling and Clinical Psychology, 61*, 941-951.

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika, 3*, 599-607.

Schneiders, J., Nicolson, N. A., & Berkhof, J. (2007). Mood in daily contexts: Relationship with risk in early adolescence. *Journal of Research on Adolescence, 17*, 697-722.

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

Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika, 38*, 196-218.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis. an introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage Publications.

Snir, R., & Zohar, D. (2008). Workaholism as discretionary time investment at work: An experience-sampling study. *Applied Psychology: An International Review, 57*, 109-127.

Song, Z., Foo, M., & Uy, M. A. (2008). Mood spillover and crossover among dual-earner couples: A cell phone event sampling study. *Journal of Applied Psychology, 93*, 443-452.

Summerville, A., & Roese, N. J. (2007). Dare to compare: Fact-based versus simulation-based comparison in daily life. *Journal of Experimental Social Psychology, 44*, 664-671.

Tang, J., Jin, R., & Zhang, J. (2008). A topic modeling approach and its integration into the random walk framework for academic search. In *Proceedings of the 2008 eighth ieee international conference on data mining* (p. 1055-1060).

Thewissen, V., Bentall, R. P., Lecomte, T., Os, J. van, & Myin-Germeys, I. (2008). Fluctuations in self-esteem and paranoia in the context of daily life. *Journal of Abnormal Psychology, 117*, 143-153.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of brownian motion. *Physics Review, 36*, 823-841.

Vancouver, J. B., More, K. M., & Yoder, R. J. (2008). Self-efficacy and resource allocation: Support for a nonmonotonic, discontinuous model. *Journal of Applied Psychology, 93*, 35-47.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*, 605-620.

Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data.* New York: Oxford University Press.

Wang, G., & Getz, L. L. (2007). State-space models for stochastic and seasonal fluctuations of voles and shrews in east central illinois. *Ecological Modelling, 207*, 189-196.

Weiss, H.M., & Cropanzano, R. (1996). Affective Events Theory: A Theoretical Discussion of the Structure, Causes and Consequences of Affective Experiences at Work. In B. M. Staw & L.L. Cummings (Eds). *Research in Organizational Behavior* (p. 1-74). Greenwich, Connecticut: JAI Press.

Welte, J. W., Barnes, G. M., Hoffman, J. H., Wieczorek, W. F., & Zhang, L. (2005). Substance involvement and the trajectory of criminal offending in young males. *American Journal of Drug and Alcohol Abuse, 31*, 267-285.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.

Witkiewitz, K., & Masyn, K. E. (2008). Drinking trajectories following an initial lapse. *Psychology of Addictive Behavior, 22*, 157-167.

Wright, S. (1931). Evolution in mendelian populations. *Genetics, 16*, 97-159.