Dyadic Structural Equation Modeling in R

John Kitchener Sakaluk

2024-10-26

Table of contents

Preface

This is a book about analyzing dyadic data with latent variables using the structural equation modeling (SEM) framework in *R*.

Academics often find reading their own writing to be a "cringe-worthy" experience, and reading that leading sentence is a particularly strange experience for me. I never once thought I'd write a book on something so statistics- and programming-oriented like dyadic SEM:

I dropped out of 11th grade programming course

I nearly failed my 12th grade introductory statistics course

My formal training in dyadic data analysis consists entirely of one weeklong workshop.

And yet…

This is a book about analyzing dyadic data with latent variables using the SEM framework in *R*, and I am writing it. And I hope it will prove to be a book worth reading, if you are interested in dyadic data analysis. Why? Because despite my aforementioned reasons to feel surprised that I am writing this book, I *do* know a thing or two about SEM and *R*. And maybe in spite of my earlier challenges with statistics and programming–actually, perhaps *because* of them–I will be able to provide you a different way of thinking about dyadic data analysis. These challenges have shaped the way I teach SEM-related content, and directly informed many of the design choices for how I have programmed the *R* package *dySEM*

Indeed, cast through the lens of SEM, the analysis of dyads affords unique opportunities to ask–and answer–some particularly interesting questions. Moreover, dyadic SEM is used so infrequently, as of the writing of this initial draft, so that I think the "market" to make an impact with this approach to data analysis is *wide open*. I hope, in time, through writing this book and developing open-source software like *dySEM*, that this "market" will become a bit more lively.

The good news, if I have your interest, is that I plan to write and maintain this book open-access. That means it will be legally free for you to use whenever you want. **The bad news** is that I plan to write and maintain this book open-access. That means that I will write it in bursts–not always in a straightforward linear way–and things will change. Don't be surprised to come back between versions and find I've totally reorganized a section, or a chapter, or even an "Act" of the book. And of course, all of my embarrassing mistakes and gaps in knowledge will be on display for all to see.

With that in mind, if you notice conceptual or code-based errors, or have requests for features or area of content converage, you'd be doing me a favour by submitting these to the "Issues" page of the GitHub repo, where I'll be maintaining the book, by using the **Report an issue** button in the top right of the screen.

Otherwise, I hope you enjoy this book and find something useful in it.

License and Supporting this Open-Access Book

License

Supporting this Open-Access Book

My hope is that this book becomes useful well *before* it is finished. As such, please feel free to peruse its chapters; if there are concepts, explanations, or coding examples that help you, while the rest of the book is in progress, all the better.

Citation

For citing the book, please use:

Sakaluk, J. K. (2024). *Dyadic structural equation modeling in R.* https://jsakaluk.github.io/Dyadic-Structural-Equation-Modeling-in-R/

Citing this book, even while–*especially* while–under development is one of the first, most direct and helpful ways you can support the development of this book. I am still deliberating whether to approach a traditional publisher for the creation of a hardcover of this book, and citations will help to index what level of demand there is for its content. And even if go with an exclusively online + open-access dissemination approach, citations help to signal to me (and my institution) that this is a useful allocation of my time. Likewise, you can "star" the book's GitHub repo here; this is the equivalent of a "like" in the GitHub space, and will help the book's visibility.

Sharing the Book

Relatedly, if you are talking about your own dyadic SEM work elsewhere (conferences, invited brownbags, etc.,), and this book was of help to you with the analyses you share, you'd be doing me a favour by mentioning the book. I don't plan to shout from the rooftops–at least not for some time–that this book is "a thing", but if you are with an audience who might be prospective readers or users, I'd welcome you sharing the book's existence with them.

Improving the Book

Submitting requests for additions, corrections, and clarifications at the Issues page would also be enormously helpful. After all, I am writing this book *for you*; if something about the organization, coverage, and/or explanation(s) in the book is not working for you, the sooner I know, the sooner I can consider an alternative.

Community Q&A

For inquiries that I don't think I can accomodate within the book, I would *strongly* encourage you to consider popping the question onto CrossValidated (google's stats-related Q&A site) under the "dyadic-data" tag (which is criminally underused). CrossValidated is a pretty amazing site for getting answers to stats-related questions, and the dyadic-data tag goes virtually unusued. But unlike asking questions over Twitter, or a FaceBook group, or within somebody's email inbox, questions (and answers) on CrossValidated are permanent and searchable, meaning someone else can learn from excellent quesitons and answers. There's also a "reputation" currency/incentive structure, and so good questions and answers float to the top of a given thread, and folks who take the time to ask thoughtful quesitons or provide helpful answers are recognized in some way/shape/form–it's really a win-win resource that keeps giving to folks in need of help, long after the original posters have moved on.

Swag

Lastly, if you're into such things, I periodically have dyadic SEM "swag" on me at conferences, including hex stickers and (more infrequently) t-shirts for the packages I maintain (dySEM and dySim). Feel free to hit me up if you see me in the wild, and I'll be happy for you to help rep these offerings.

About the Author

Hi! I am John Sakaluk (pronounced: “Sack-uh-luck”, like a bag of luck)!

Figure 1: A wild John appeared!

I use he/him prounouns, and I live/work in Canada on the traditional lands of the Anishinaabek, Haudenosaunee, Lūnaapéewak, and Chonnonton Nations

What's most relevant, for you, is that:

- I am an Assistant Professor in the Department of Psychology at Western University (formerly/occassionally still refered to as the University of Western Ontario) in London (not *that* London), in Southwestern Ontario, Canada.
- My research program spans the domains of sexuality, relationships, and psychological interventions. Most of what I'm known for would be considered applied quantitative psychology work, in which I use (or promote) emerging or improved analytic techniques, usually either for the purpose of psychometrics or research synthesis
- I was a postdoctoral fellow in Dr. Emily Impett's lab, in which she and her students often conduct dyadic data analyses
- Thanks to Emily's support, I took a week-long dyadic workshop from Dr. Dave Kenny and Dr. Tessa West. This workshop was mostly taught through the lens of MLM (and in SPSS), but was nonetheless was a useful training exercise that got me thinking a lot about dyadic data and why it is analyzed the way it is (and isn't)
- My "expertise" in dyadic data analysis is quite a recent phenomenon–emerging years after said workshop. During the submission of an article for a special issue in *Personal Relationships* (see John K. Sakaluk, Fisher, and Kilshaw (2021a)), the reviewers encouraged us to develop some *R* functionality to help with the computation of dyadic-versions of invariance testing effect size indexes (from Nye and Drasgow (2011)). Working through this request unlocked some programming barriers in my brain, and a few bursts of impromptu coding later, I had laid the foundations for the *dySEM* package for *R* (see here for link to CRAN site). After *dySEM* kicked around for awhile, and I gave a few smaller talks on dyadic SEM, I was invited to give the closing keynote at the 2023 Close Relationships Preconference of the Society for Personality and Social Psychology (REF OR SLIDES), during which I mostly evangilized abou the benefits of dyadic SEM. That talk–which was received with a strange amount of controversy–seems to have catalyzed things for this portion of my research program, as an invited review paper on dyadic SEM (REF) and a *dySEM* workshop (REF OR SLIDES) have quickly followed.
- Though my "expertise" in dyadic data analysis is a recent phemonenon, I have extensive training and experience with SEM and *R*. My first foray with factor analysis was as a 4th year undergrad working with Dr. Robin Milhausen; I borrowed her copy of *Multivariate Statistics* (Tabachnick and Fidell (2012)) and my mind has been lost to latent variable modeling ever since. Afterwards, as a graduate student at the University of Kansas, I took multiple courses on structural equation modeling with latent variables, including from Drs. Carol Wood, Mijke Rhemtulla, Alexander Schoemann, and Todd Little. This included learning how to apply SEM in contexts with dependent data [1], which is largely

[1] Todd was just finishing his now go-to book on Longitudinal SEM (Little (2013)) when I took his Longitudinal SEM class. Doing the proof-reading and code-testing for one of his book's chapters was a legitimate final project option for that course–a possibility I look back on with some sardonic amusement, now that I am

what has enabled me ot move into dyadic SEM even though I have not often analyzed primary dyadic data. It was around my 3rd year of grad school that I began routinely using the *psych* and *lavaan* packages for latent variable modeling in *R*.

- Since then, I have routinely used and taught *R* and SEM–10 years and counting–including to some leading dyadic data analysis folks in the field.
- Most recently, I have added the label of *R* developer to my set of experiences, having taken the plunge of *R* package creation and maintenance.

Outside of my work-related bona fides, I am known for:

- hailing from North Bay, ON
- making good food and drinks
- having a peculiar brand of academic-speak in both writing and talking[2]
- code-switching between sounding very American (I'm a very enthusiastic "y'all"-er), or very Canadian (think Letterkenny)
- using moderately? profane language
- playing basketball (I play a [very watered down] Shawn Marion-like game)
- having brutal handwriting
- typing very loudly (I snap my spacebar particularly hard, for some reason)
- having a very wonderful, cheerful partner, and a very large, goob-tastic dog

writing a book of my own and have my own graduate statistics classes.

[2]Emily's lab used to call this "Johnspeak"; you can credit (or blame) my late Grandmother, who was a powerfully idiomatic speaker

Acknowledgements

Part I

Act I: Getting Started

1 Introduction

This is a book about analyzing dyadic data with latent variables using the SEM framework in *R*. As such, it occupies a unique location within the market of available books across several domains. Kenny, Kashy, and Cook (2006), for example, is an excellent book focused exclusively on dyadic data analysis. It describes the challenges of managing dyadic data, as well as many of the most popular models of dyadic data. But though Kenny, Kashy, and Cook (2006) describe the SEM framework and its application to some models, they don't really mean "SEM" in the way that I do. That is, their consideration of SEM isn't particularly engaged with the prospect of modeling latent variables, while I am *exclusively* focused on the use of latent variables.

A book like Little (2013), meanwhile, describes the use of latent variable modeling for dependent data *like* dyadic data, albeit in the context of longitudinal research designs, where the source of dependency is repeated observation. As such, though this book possesses some wisdom for those wishing to use latent variables in their models of dyadic data, the book is not written with dyadic data or dyadic models in mind. And as it turns out, things get *weird* when you are analyzing dependent groups of two (as opposed to, say, individuals over the span of three, four, or five waves of repeated assessment).

There are many other excellent books, too, that offer additional guidance–to some (in)direct degree or another–for applying latent variable models to dyadic data, such as Brown (2015) and Kline (2023) and Bolger and Laurenceau (2013). By writing this book, I mean to take nothing away from their value propositions; they are great resources. But they either do not consider the unique features of dyadic data and/or do not substantively engage in the distinctive benefits to modeling dyadic data with latent variables.

This is a book about analyzing dyadic data with latent variables using the SEM framework in *R*. And so I will exclusively discuss the analysis of dyadic data, **and** I will exclusively discuss its analysis with latent variable models.

In **Act I** of the book, I will first lay out "The Big Picture" (Chapter 2) of the what's, why's, and how's of dyadic SEM with latent variables, as well as introducing some of the technical jargon I will use throughout the rest of the book. I then discuss some of the unique considerations of data management for dyadic data analysis (Chapter 3), before providing an overview of latent variable theory in the context of modeling dyadic data (Chapter 4).

In **Act II** of the book, I attempt to provide what I consider to be a "sufficient" overview of the conceptual and applied specifics of modeling latent variables, without yet engaging with how to extend this framework to the analysis of dyadic data. Chapter 5 is essential reading

if you are unaccustomed to the statistical features of latent variables models (e.g., the visual depictions, notation and interprestation for particular features, etc.,) and common analytic practices within them (e.g., fixing or constraining parameters). I then discuss two related problems that must be resolved in fitting SEMs (model identification and setting the scale of the latent variable(s)) (Chapter 6), before they can be estimated (Chapter 7). These topics then set the table for describing how we evaluate (Chapter 8) and compare (Chapter 9) SEMs. I then work through all of this (and more) in an applied non-dyadic example (Chapter 10).

I then pause in **Intermission 1** to provide some foreshadowing about the dyadic SEMs we will encounter in the subsequent chapters (Chapter 11). For those approaching this book while having some familiarity of the models described in Kenny, Kashy, and Cook (2006), this chapter will help to transition you to thinking about these models recast in latent space.

In **Act III**, we *finally* get into the specification of models for dyadic data with latent variables, beginning with the simplest cross-sectional models possible: those intended to capture only one *construct* (i.e., "uni-construct") shared somehow between dyad members. These include the dyadic one-factor model (Chapter 12), the correlated two-factor model (Chapter 13), the bifactor model (Chapter 14), and the hierarchical factor model (Chapter 15). These models have a surprisingly interesting (and complex) relationship to one another, which I discuss in the subsequent chapter (Chapter 16). I also describe how to use dyadic invariance testing within these models, in order to evaluate the generalizability of latent variable model parameters across partners (Chapter 17), which plays an important role in many other comparisons in dyadic data analysis, as well as being an (unappreciated) interesting phenomenon in its own right. I conclude this section with a discussion of an important but vastly underappreciated issue: how to choose among competing uni-construct models for a given set of dyadic data (Chapter 18).

In **Act IV**, we move to discussing dyadic SEMs that are latent embodiments of the kinds of models that may seem more prototypical in dyadic data analysis (i.e., those covered in Kenny, Kashy, and Cook (2006)). These models involve the prediction of one dyad-related construct by another (i.e., are bi-construct). I first describe bi-construct models where the predictor construct and outcome construct share the same uni-construct dyadic model, including the Couple Interdependence Model (Chapter 19), the Actor-Partner Interdependence Model (Chapter 20), the Bifactor Structural Model (Chapter 21), and the Common Fate Model (Chapter 22). I conclude this section with a discussion (and some encouragement) of how different uni-construct models could be combined in more boutique bi-construct models (Chapter 23).

We then pause once more, in **Intermission 2**, in order to discuss–with the knowledge of uni-construct and bi-construct dyadic SEMs under our belts–just how complicated the concept of "distinguishability" is, when cast through the SEM lens.

Finally, in **Act V**, we delve into even more complex applications of dyadic SEM, including some themes of practice that are not yet done, yet which I hope will be on the (near) horizon of analytic practice in our field. These include the modeling of so-called "third variable" processes (Chapter 25), testing dyadic SEMs across groups (Chapter 26), the modeling of both dyadic

and longitudinal dependency with latent variables (Chapter 27), and the deployment of data-driven exploratory models to provide a plausible dyadic measurement model (Chapter 28). I also discuss the application of (and need for more) Monte Carlo simulation studies (Chapter 29), to evaluate the performance of dyadic SEMs (and other modeling strategies); in this chapter, I also discuss how these simulations can help to inform sample size planning. I then end this section with some encouragement and guidance of how to contribute to open-source dyadic data modeling tools (Chapter 30), for those so inclined.

I'll also (eventually) write a Conclusion to this book (Chapter 31), and I'm sure it'll be very meaningful and impressive. But for now, I need to generate some content, before I can realize what it is I ought to conclude.

What I will *not* write about at length, however, is the basics of using *R*–the open-source cross-platform statistical programming language that I use in my dyadic SEM work (and upon which this book and its applications currently rely). If you are entirely new to *R*, the good news is that most dyadi SEM modeling instantiations require precious little of fiddling around with basic data management in *R*. That is, you're often "good to go" soon after data importation. If you need additional scafolding for using *R*, however, I encourage you to check out "R for Data Science", or "R4DS" as it's sometimes known (**NEED REFERENCE**). It's a gold mine of useful information for *R* users of all levels of comfort.

And no: I will not provide analytic resources for other programming/statistics languages (e.g., SPSS/AMOS, SAS, MPlus). Though I have sometimes done this in the past (e.g., John K. Sakaluk and Short (2017), John Kitchener Sakaluk (2019)), and some of the models I describe herein are possible to specify in these other softwares, I have decided that I am done supporting proprietary software. This may lose me some readers; so be it. I want to create learning resources and tools that are avaiable to anyone and everyone, for free, and the increasing expense of these other software packages threatens what I see as a necessary mandate to democratize access to learning.

2 The Big Picture

This is a book about analyzing dyadic data with latent variables using the SEM framework in *R*, and I think your experience reading the remainder of the book will be aided by providing you with a rapid-fire tour of the bigger themes within it (and an introduction to some key jargon).

So, let us begin with that introductory phrase: analyzing *dyadic data* with *latent variables* using *the SEM framework*.

2.1 What is "dyadic data"?

Putting a finger on what dyadic data is, specifically, is surprisingly difficult when you begin to think of variety of ways in which the organisms we study (people, animals, nation-states, etc.) can, and do, "pair up". And so in some ways, it's easier to begin by stating what dyadic data *is not*.

Dyadic data is not data from organisms that have nothing in common. How much nothing? Total. Imagine you ran a goofy study where you experimentally assigned people to pairs and estimated the correlation between pairs on *Trait X*. In the long run, it should be $r = .00$, right? Okay: those data are *not dyadic*. You should be fine to structure these data in such a way that allows you to analyze them as coming from independent observations, and go on with your business.

Eliminating data from organisms that have nothing in common still leaves a lot of possible contexts on the table as potentially dyadic. At the opposite end of the spectrum of *dyadic-ness* are data that come from organisms that totally overlap. Though contexts like this are rare (rare enough that I can only think of: data from a sample of identical twins), they probably would satisfy our intuition-based criteria of what is dyadic data.

The tricky cases are those in-between: are data from research where organsism have something (but not everything) in common, *dyadic* data? For example, what if you collect data from a classroom ($n = 200$) and a couple of pairs of individuals (*pair 1* and *pair 2*, who otherwise do not interact) happen to share a hometown? Or in a sample of beetles ($n = 60) a few pairs (*pair 1*, *pair 2*, and *pair 3*), happen to be sibblings from the same brood?

Though examples like these contain *dyads*, I would not call the affiliated datasets *dyadic*. This is because the paired nature of some observations is inconsistent; there are a few dyads/pairs,

Figure 2.1: The first of many stats-related memes

but most of the data otherwise comes from individuals, and it's not clear how a researcher would systematically be able to identify the linked observations between the small number of dyads.

Instead, ***let us consider data to be "dyadic data" if it is dyadic by design***. That is, a researcher voluntarily deployed an approach to collecting data in such a way that all organisms in the sample are members of a dyad, known to the researcher. In this book, we will exclusively deal with an approach to analysis of quantitative dyadic data, though qualitative dyadic research designs are also possible (at least for human-based samples).

We will see in Chapter 3 that there are further ways we can conceptualize and categorize dyadic data–and the analytic techinques in this book are most useful to only one of these dyadic data designs–but that is a sufficient understanding of "dyadic data" for now.

2.2 What are "latent variables"?

What, then, are these so-called *latent variables*, and what is useful or interesting about them? I like Dictionary.com's definition of *latent* as a starting place:

> [leyt-nt] present but not visible, apparent, or activated; existing as potential

Typically, researchers depict latent variables (vs. "*observed*" or "*manifest*" or "*indicator*" variables) in a particular way when creating figures of their statistical models. Specifically, variables depicted as circles or ovals are latent variables, whereas variables depicted as squares or rectangles are observed variables (i.e., that are directly measureable). This can get a bit confusing, for a couple reasons, including:

- the same psychological construct (e.g., "self-esteem")–and even the same measure of the same construct (e.g., the Rosenberg Self-Esteem Scale)–can be specified as a latent variable in one model, and an observed variable (e.g., an average or sum of item scores) in another. That is, even if someone believes a construct to be latent, how it is modeled remains an analyti choice, and
- observed variables take the same shape in these kinds of diagrams (i.e., squares) but can have *very* different causal roles in the model. For example, one observed variable may be for the responses to a particular item for a given latent variable (modeled latently); another may be a predictor variable; another may be an outcome variable.

2.2.1 A Brief Bit of History

But what use is a variable that you cannot directly observe? Social scientists have been interested in latent variables since as far back as the late 1800's and early 1900's. A classic example is Spearman (1904) (of the "Spearman correlation coefficient" fame)(REF). Spearman

Figure 2.2: Dr. Srivastava is one of “the greats” when it comes to stats-related memes

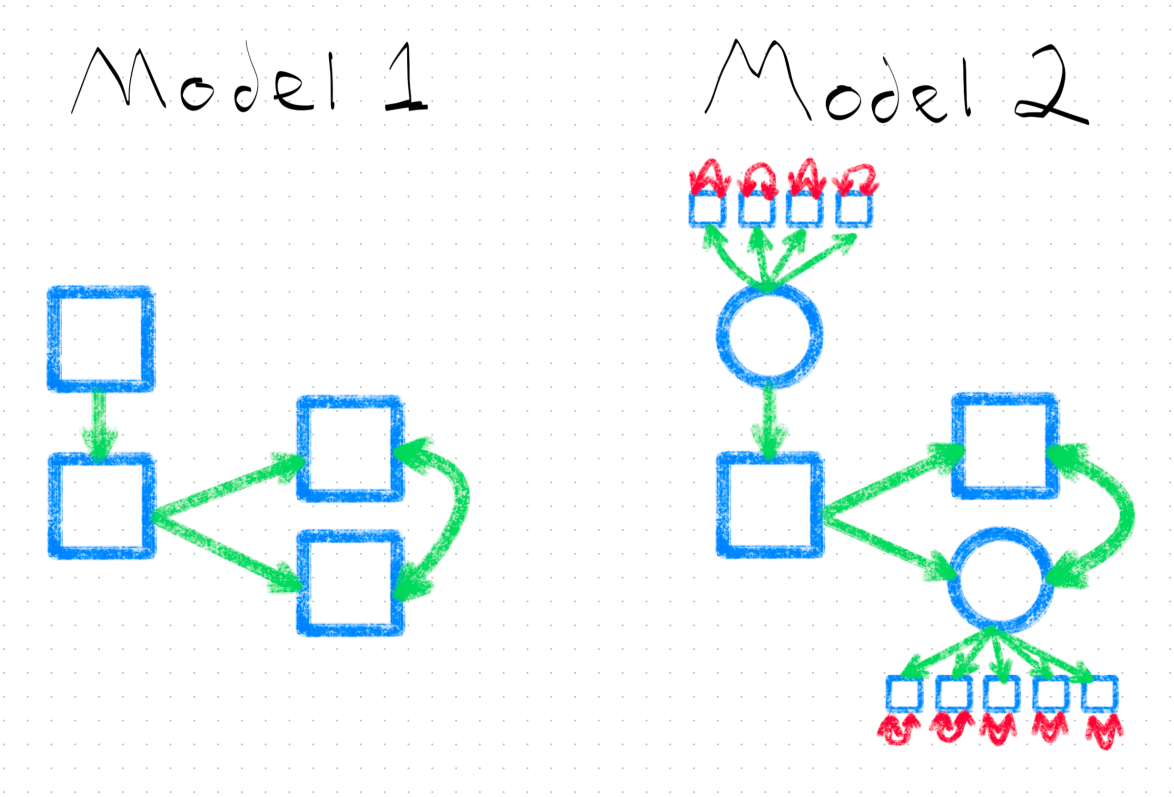

Figure 2.3: Model 1: a model of observed variables; Model 2: a model of latent variables and observed variables (in different causal roles)

was psychologist who was interested in, among other things, mental abilities. At the time, there were a number of competing theories of ways to categorize "types" or "kinds" of mental abilities (e.g., visual, memory-based, spatial, etc.,), and how to organize them. Spearman's contributition to this literature was noteworthy, as he provided evidence for a *latent* kind of *generalized intelligence* (which he dubbed "g")–not a form of mental ability you could directly see or appraise, but which shaped performance on all tests of various mental abilities to one extent or another. The impact of this work was tremendous, both in terms of its theoretical value and the way it contributed to subsequent statistical methodology. With respect to the former, Spearman helped to establish the plausibility of latent variables (of which generalized intelligence may just be one) as important determinants of social and psychological processes. Methodologically, meanwhile, through is work, Spearman "casualy" provided one of the first exemplars of factor analysis (and a very specific kind of it, at that)–an analytic approach that would come to play a central role in a staggering amount of future research. Indeed, factor analysis is the beating heart of many of the models of dyadic data we will consider in this book.

The notion of latent generalized intelligence may seem reasonably intuitive, but what of latent variables in dyadic data? Here, I must profess, it's easy for me to provide exemplars that work in the discipline of psychology, and much more difficult to provide examples that work in disciplines farther afield (e.g., ecology). Still, let me try.

For my money, the game-changing moment in Psychology to open the door to the study of latent variables was the Cognitive Revolution. The theoretical coin of the realm before the Cognitive Revolution was, for many decades, Behaviorism. And under Behaviorism–specifically, the edicts of "Methodological Behaviorism" (WATSON 1924 REF)–the experience of thought and feeling was thought to be bereft of scientific value. Only that which could be seen, directly–behavior–was amenable to scientific inqiury. The Cognitive Revolution was largely responsible for upending this ban on studying thoughts and feelings, and with that, it became open-season to scientifically study intrapsychic phenomena like attitudes, beliefs, emotions, motives, and values. These concepts represent the kind of psychological terrain where wild latent variables may be found, and studied.

2.2.2 Typologies of Latent Variable Models

The kind of latent variables we will focus on modeling, throughout this book, have a few notable features that we will expand upon later, but which must be at least "gestured" at now.

First, and most importantly, they are ***reflective*** latent variables, which means the direction of causality flows from unobservable entity, to measurable entity. For example, latent depression–which cannot be directly seen, or assessed–*causes* someone to lose sleep, have low mood, etc. (and thereby indicate as much in their questionnaire responses). Not the other

way around. *Those* kinds of latent variables are referred to as ***formative*** (Bollen and Diamantopoulos (2017), Rhemtulla, Van Bork, and Borsboom (2020)) –socioeconomic status is a classic example–and we shan't speak of their kind any further.

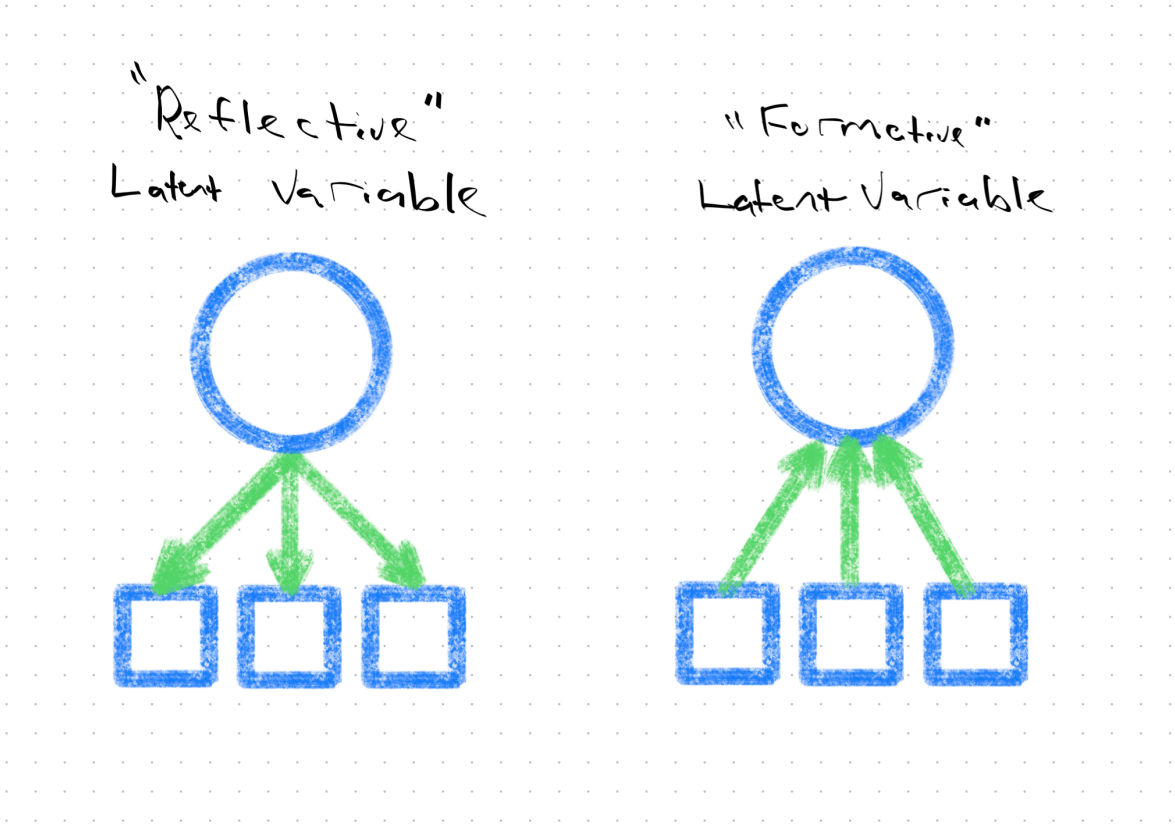

Figure 2.4: A noble, reflective latent variable model on the left; an evil, formative latent variable model on the right

Another important property is that we will focus exclusively on ***continuous/dimensional*** latent variables. That is, latent variables that exist on some sort of numeric scale (think ratio scale of measurement, STEVENS REF), where there can be smaller and larger precise amounts of the latent stuff. Latent variables can, however, be ***categorical/discrete*** instead–representing an underlying set of unobserverable categories (often referred to as *mixtures*), which may be ordered or unordered. Though mixture modeling is possible, with *latent class analysis*(Collins and Lanza (2009)) and *latent profile analysis* (Pastor et al. (2007), Rosenberg et al. (2018)) its classic instantiations, extending these frameworks to dyadic data is not without its complications[1]. Trust me that learning how to deploy latent dimensional models to dyadic data will be task enough. Besides, the vast majority of psychological entities

[1]and all due respect, but most who have applied mixture models to dyadic data have all made similar model specification errors

are dimensional, not categorical (Haslam, Holland, and Kuppens (2012), OTHER HASLAM REF).

Finally, the examples of dyadic SEM we will traverse in this book will all use continuous (or continu-ish) ***indicator*** or ***manifest*** variables (i.e., in which the underlying latent variable is reflected), like responses to self-report questionnaire items with some sort of likert-like rating scale. Discretely scaled indicators can be used with dyadic SEM, but require the use of different estimators (Chapter 7), and there parameter estimates often have a different corresponding interpretation. We'll see if I make time/space, when writing this book, to discuss these sorts of models, but they won't be the prototypical case.

2.2.3 Analtyic Use-Cases of Latent Variables

Is all of this sounding somewhat complicated? It is.

Are latent variable models still worth using? I think so.

What are latent variable models good for, then? Excellent question.

Latent variable models of the kind we'll be indulging have, in my view, at least two primary, and sometimes overlapping, use-case. We'll cover both in more detail later, but for now, in brief:

1. ***Latent variable models can improve the estimation accuracy of your effects*** If you analyze your desired pattern of linear relations between your constructs of interest using (i) linear regression using composite scores (e.g., average or sum scores) as stand-ins for your constructs, and again using (ii) SEM with latent variables, *and the gospel population truth was that your constructs came from a universe in which they were indeed latent variables* (a "big, if true" proposition, see Rhemtulla, Van Bork, and Borsboom (2020)), then methodological scholarship strongly suggests the latter will return closer estimates to the population values of your effects of interest (e.g., slopes, correlations, etc.,) than the former (Cole and Preacher (2014), Ledgerwood and Shrout (2011)), including for dyadic models like the *actor-partner interdependence model (APIM)* (Kim and Kim (2022)). You may have heard of this whole 'crisis' of relicability/credibility in the social (and medical) sciences (Open Science Collaboration (2015), Vazire, Schiavone, and Bottesini (2022))? Many have suggested (e.g., Flake and Fried (2020), Hussey and Hughes (2020)) that measurement problems play a focal role in this crisis, and therefore a statistical framework that renders more trustworthy estimates of effects could be very useful indeed! And,

2. ***Latent variable models are distinctively useful for interogating differences in bias in measurement and/or latent conceptualization*** A tacit assumption of most quantiative analyses is that assessment procedures work more or less than same across peoples, and that (when studying psychological entities) people have the same psychological concept in mind when responding to questionnaire prompts. However, there's

rich historical evidence of psychological assessments working better and worse for some groups of people than others (EXAMPLES/REFS), and increasing evidence that different groups of people have different notions in mind when you ask them questions about squishy constructs, like political identity, sexualit, etc., (EXAMPLES/REFS). Latent variable modeling techniques offer unmatched capacity to evaluate whether assessments and conceptualizations are shared–or differ–across groups and other design features like time (Little (2013)), and/or dyadic partner membership (John K. Sakaluk, Fisher, and Kilshaw (2021b)).

2.2.4 In Short

In brief, latent variables have been of interest in the field for a long, long time. They come in different varieties (only some of which we'll discuss in this book), and modeling them imposes some additional analytic complexity, but this can be worth it for the gains you make in estimation accuracy, and assurances around the generalizability of your assessment procedure and/or your participants' conceptualizations[2]

2.3 What is "structural equation modeling (SEM)" ?

You likely have a workflow by which you determine what analysis is appropriate for you to pursue, given the data that you have. Have a binary outcome variable that you want to predict from a slew of predictor variables? A binary logistic regression model is probably the way to go. Have 20 or more days worth of questionnaire responses from a sample of individuals across a "daily diary" design? You're probably in line for some kind of multilevel model.

Likewise, SEM is an analytic framework that accomodates the modeling of certain kinds of data. The main distinctive feature is that unlike linear regression, logistic regression, or mulitlevel modeling–which only enable the analysis of one outcome variable at a time (i.e., they are ***univariate*** models)–SEM is a ***multivariate*** framework for data analysis, which means, *it faciltates the fitting of linear models to multiple outcome variables simultaneously*[3].

The multivariate-friendly feature of the SEM framework is key for dyadic SEM with latent variables, as each latent variable we want to model will require us to have multiple indicators that are causal manifestations (i.e., outcomes) of it.

What kind of outcome variables? Like I said, we'll focus on continuous outcome variables in this book for the most part[4]. And like multilevel modeling, SEM has some capacity to

[2]though note, I think differences in conceptualization–indexed through different psychometric measurement models–are one of the most interesting and under-used analytic outcomes in the field. Expect me to return to this soapbox later.

[3]though SEM easily accomodates the analysis of just one outcome variable, too

[4]though SEM easily accomodates categorical data too

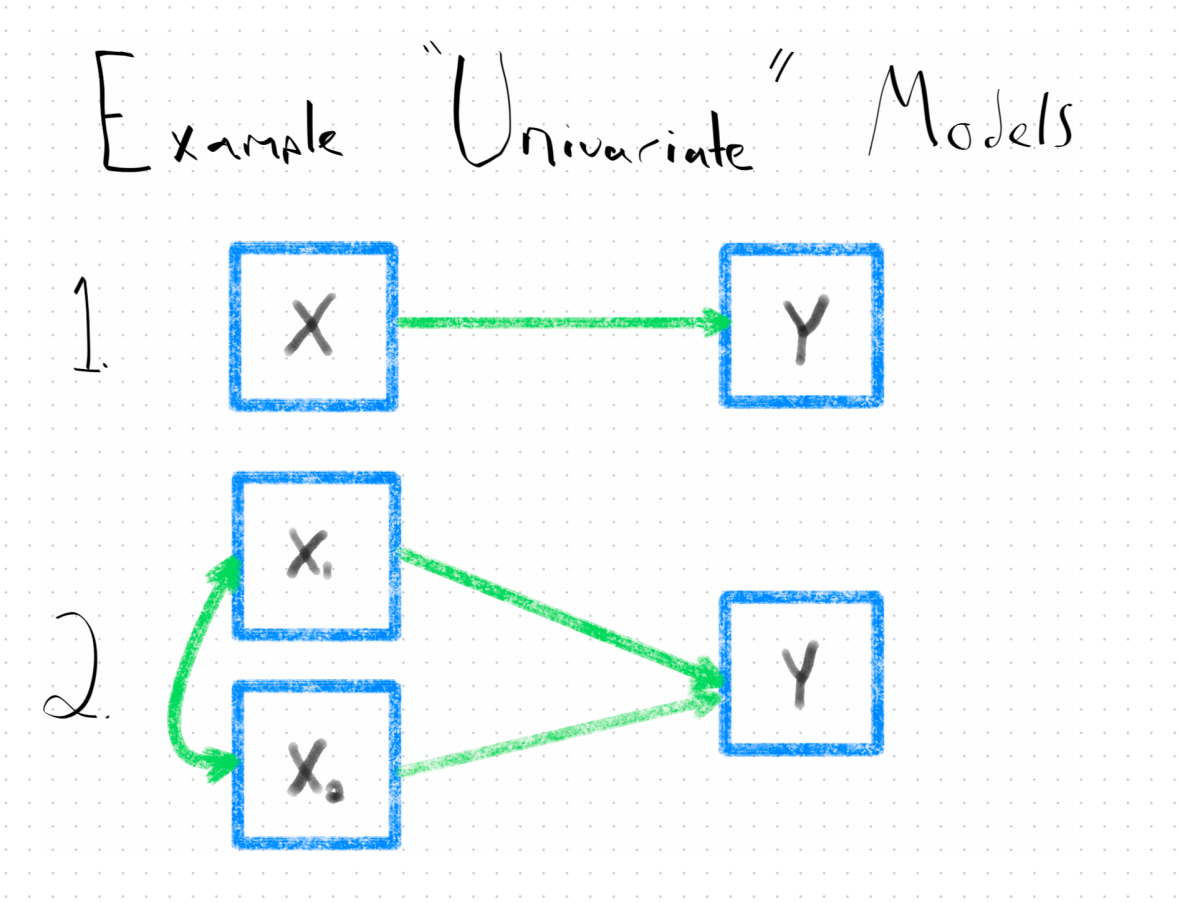

Figure 2.5: Some simple exemplar univariate models

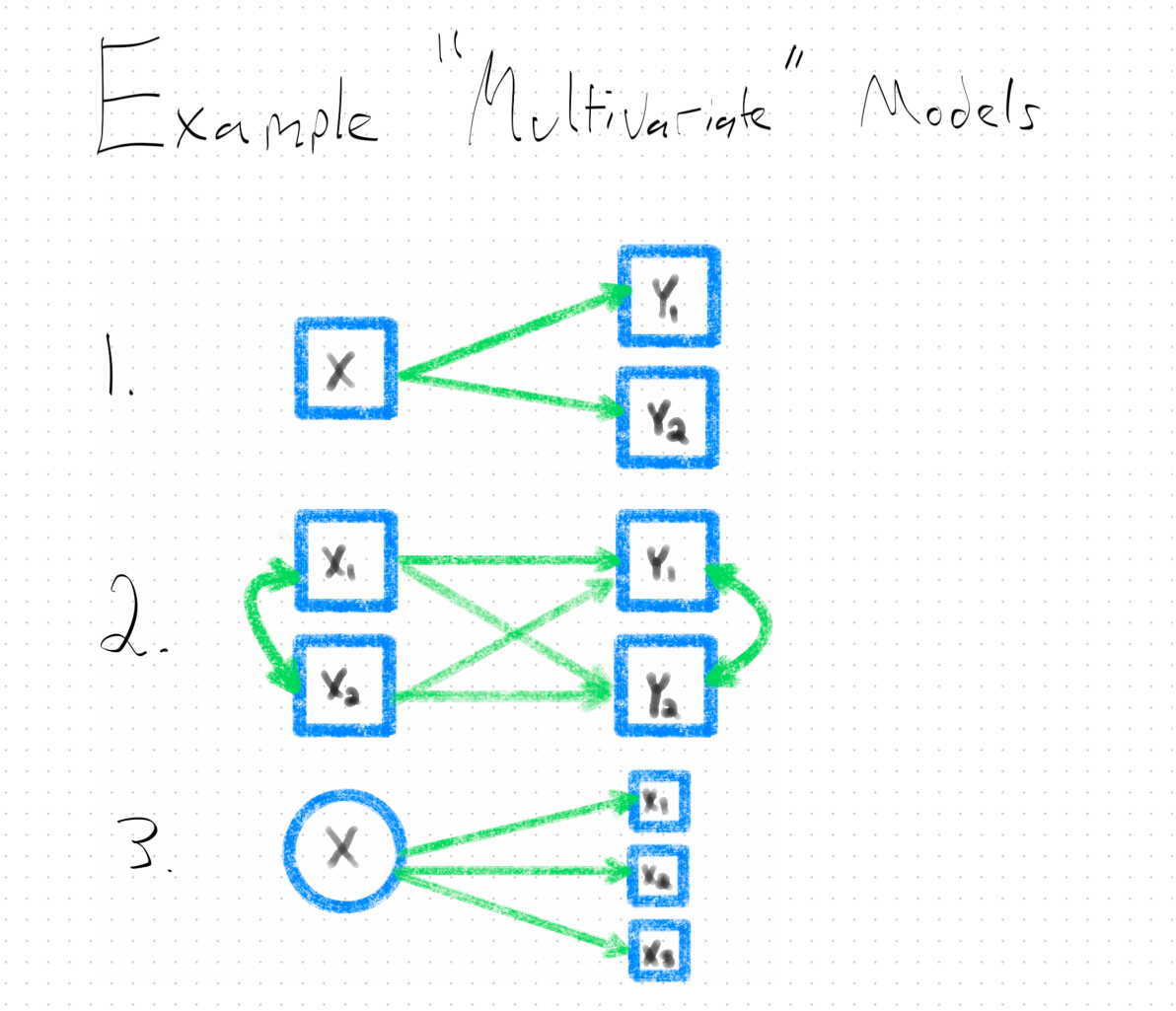

Figure 2.6: Some simple exemplar multivariate models

accomodate dependent data (like from dyadic designs, and/or repeated assessments), though SEM can only handle so much dependency (~10 waves of data?) before models get unduly complex (we'll return to this in Chapter 27).

Sounds like SEM is kind of an "all in one" analysis framework, doesn't it? That's because it is–and this is one of the reasons I'm such a fanatic about it! But this flexibility means that it's worth clarifying some language up about how SEM is used. SEM can clearly be used for a variety of purposes–including the modeling of non-latent variables (e.g., observed sums, averages, and single-item responses). SEM has a rich history of being used to analyze this kind of data, from dyadic designs. Though this is, technically, "dyadic SEM", it's an application of dyadic SEM that makes little use of the unique analytic capacities of the framework. I therefore refer to this application as ***dyadic path analysis***.

I use the term ***dyadic SEM***, therefore, to refer more specifically to dyadic models that primarily (but perhaps not exclusively) feature latent variables. And if one wants to analyze latent variables, the multivariate-friendly functionality of SEM is what's needed.

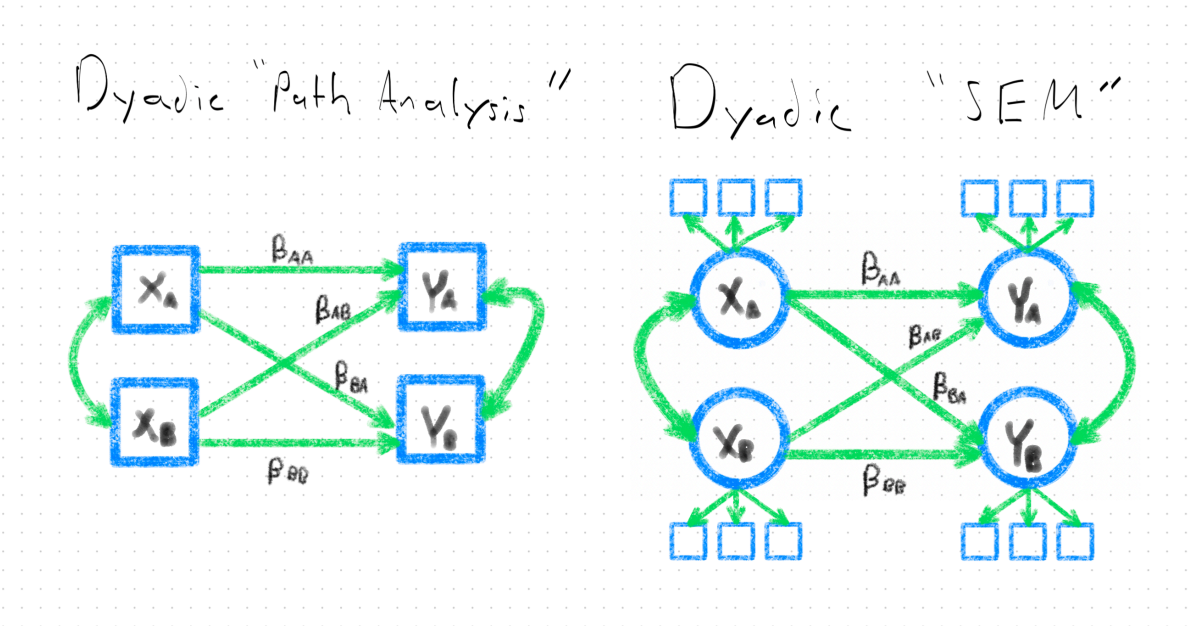

Figure 2.7: The same pattern of dyadic relations between constructs (hypothetical X and Y) represented using observed variables in what I refer to as a *dyadic path analysis* model, versus represented using latent variables in what I refer to as bona fide *dyadic SEM*

But what SEM does tells you very little how it does it. The high-level summary is that the parameters (Chapter 5) estimated in an SEM (Chapter 7) are used in matrix algebra to solve for a set of guesses of what the variances and covariances of the variables in your model ought to be; these are then compared against what your variables' variances and covariances actually

are (Chapter 8) in order to determine whether your model does a half-decent job representing the data. If it does, then you're off to the races interpreting the estimated effects in your SEM. If not, then it's back to the drawing board to select another model that performs better.

2.4 What is *R*?

R is an open-source cross-platform programming language for statistical modeling and visualization. One line into this subsection, and these few descriptors already put *R* in stark contrast against other statistical analysis program offerings. That is, unlike other SEM-friendly software like AMOS and Mplus, *R* is legitimately and legally free, and will work on virtually any kind of computing environment you have access to. With costs of living being so high, and investments in science and education dwindling, that you can get quality statistical modeling software for free is an incredible boon.

There are, however, other open-source statistical programming languages in town, namely python. And if you're looking for a statistical language skill that pays the bills, by all metrics, you'd be hard pressed to do better than python. Python is great. But for SEM, it's got a ways to go. For example, the *semopy* package (REF)–what appears to be python's most popular SEM package–has only received 93 citations since 2020.

In contrast, *R*'s most popular SEM offering, the *lavaan* package (Rosseel (2012)) has been cited more than 15,000 times in the same timeframe. And while *R* doesn't dazzle as much as python in the programming language studies by tiobe and IEEE, it's held its own over the years. For example, for the better part of 10 years, *R* was considered a top-10 programming language according to IEEE (and this included languages like SQL and C+ that are not merely statistics-oriented). And based on the TIOBE index, *R*'s performance–ranked as highly as the #8 programming language, in August of 2020–has been relatively stable.

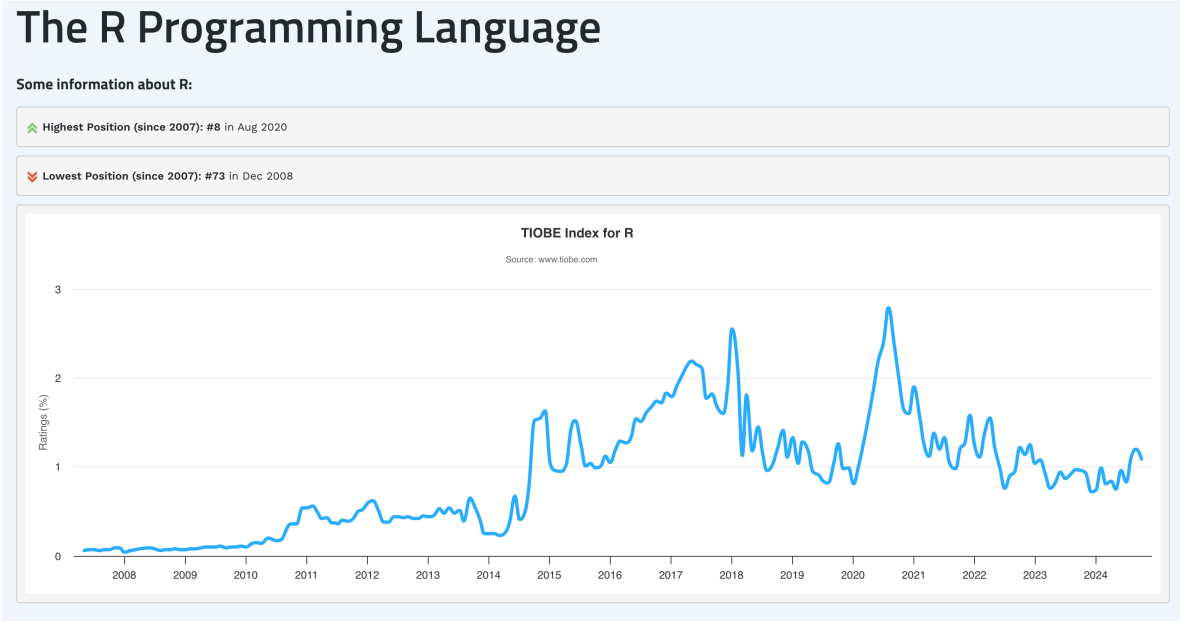

Figure 2.8: R's TIOBE index over time, from www.tiobe.com

I am fond of the explanation of *R*'s staying power offered by IEEE in their 2023 rankings write-up:[5]

> But don't let Python and SQL's rankings fool you: Programming is still far from becoming a monoculture. Java and the various C-like languages outweigh Python in their combined popularity, especially for high-performance or resource-sensitive tasks where that interpreter overhead of Python's is still too costly (although there are a number of attempts to make Python more competitive on that front). And there are software ecologies that are resistant to being absorbed into Python for other reasons.
>
> For example, R, a language used for statistical analysis and visualization, came to prominence with the rise of big data several years ago. Although powerful, it's not easy to learn, with enigmatic syntax and functions typically being performed on entire vectors, lists, and other high-level data structures. But although there are Python libraries that provide similar analytic and graphical functionality, R has remained popular, likely precisely because of its peculiarities. They make R scripts hard to port, a significant issue given the enormous body of statistical analysis and academic research built on R. Entire fields of researchers and analysts would have to learn a new language and rebuild their work. (Side note: We use R to crunch the numbers for the TPL.)

[5]The original link is broken, redirecting to the 2024 results, while all other previous survey result sites are available. I therefore quote liberally from the page, having found it with the Wayback Machine

SEM is one of those analysis types that seems to remains most easily done in *R*, with packages like *lavaan*, perhaps owing to *R*'s pecularities. And don't let this quote scare you about R's usability. Though it used to be difficult to learn and use *R*, especially for data management and visualization task, great strides having been made here. Indeed, the *tidyverse* conglomerate of packages, including the famed *dplyr* and *ggplot2* packages, offer simple and powerful tool for data management and visualization. So much so, in fact, that it appears python users have a bit of envy for these offerings.

Whatever your comfort level with *R*, let me assure you: the programming required for SEM (including dyadic SEM) is suprisingly light in terms of data management. Usually once you've imported a data set, you are ready to roll (though see Appendix A if you need a boost on some of the basics). *lavaan*, meanwhile, offers commercial-grade SEM software for free–it's incredible what it can do. And while I wouldn't call *lavaan*'s manner of specifying SEMs "simple", it's basically on the level with its main competitors (e.g., there's many a similarity between *lavaan* and Mplus syntax). Further, while it's important to understand *lavaan* syntax well enough to navigate it on your own, I've developed *dySEM* to simplify and expedite using lavaan to fit dyadic SEMs.

All to say, it's the perfect time to join the dyadic SEM party. Come on into the *R* pool; the water's fine!

2.5 Bringing it all together

This is a book about analyzing dyadic data with latent variables using the SEM framework in R. Having reached the end of this chapter (sure to be one of the larger ones), you now hopefully understand this means the book is about:

- analyzing data that comes from stable dyads, *by design*–that is, researchers deliberately sampled pairs of friends, romantic partners/mates, siblings, etc.
- using techniques that allow us to model the influence of unobservable latent variables, which provide increased estimation accuracy and the opportunity to evalaute the generalizability of our assessment procedues and participants' conceptualizations of intrapsychic constructs
- doing all of this using SEM, a statistical framework that accomodates the linear modeling of (very) many outcome variabels simultaneously (an essential ingredient for latent variable modeling)
- using the *R* programming langauge, which provides open-source cross-platform commericial-grade SEM functionality to you, via packages like *lavaan* and *dySEM*

If that sounds good to you, then let's get on with it, starting with a discussion of the data structure needs of dyadic SEM.

3 Dyadic Data Structure

4 Latent Variables in SEM

Part II

Act II: SEM Fundamentals

5 Models (and Their Elements)

6 Model Identification and Scale-Setting

7 Estimating Dyadic SEMs: Common Practices, Problems, and Solutions

8 Model Fit Indexes

9 Model Comparison Strategies

10 A Simple (Non-Dyadic) Example in {lavaan}

Part III

Intermission 1: Meet Your Dyadic Models

11 Design, Scope, and Form of Dyadic Models

Part IV

Act III: Uni-Construct Cross-Sectional Dyadic Models

12 The One-Factor (Surprinslgy Dyadic) Model

13 The Correlated Two-Factor Model

14 The Bifactor Model

15 The Hierarchical Factor Model

16 The Causal Relations Between Univariate Cross-Sectional Dyadic SEMs

17 Dyadic Invariance Testing in Univariate Cross-Sectional Dyadic SEMs

18 Choosing Between Uni-Construct Cross-Sectional Dyadic Models

Part V

Act IV: Bi-Construct Cross-Sectional Dyadic Models

19 The Couple Interdependence Model

20 The Actor-Partner Interdependence Model

21 The Bifactor Structural Model

22 The Common Fate Model

23 Mixing Univariate Structures in Bivariate Models

Part VI

Intermission 2: What Is '(In)distinguishability'?

24 (In)distinguishability Is More Complicated than You Think

Here is an example of an equation where different parts are color-coded:

$$E = mc^2 \quad \text{where} \quad \color{red}{E} = \color{blue}{m} \cdot \color{green}{c^2}$$

Part VII

Act V: Advanced Applications of Dyadic SEM

25 “Third Variable” Cross-Sectional Dyadic Models

26 Multi-Group Cross-Sectional Dyadic Models

27 Longitudinal Dyadic Models

28 Dyadic Exploratory Factor Analysis

29 Monte Carlo Simulations of Dyadic Modeling Strategies

30 Contributing to Open-Source Software Development for Dyadic SEM (and Beyond)

Part VIII

Fin

31 Conclusion

References

Bolger, Niall, and Jean-Philippe Laurenceau. 2013. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York, NY: Guilford Press.

Bollen, Kenneth A., and Adamantios Diamantopoulos. 2017. "In Defense of Causal-Formative Indicators: A Minority Report." *Psychological Methods* 22 (3): 581–96. https://doi.org/10.1037/met0000056.

Brown, Timothy A. 2015. *Confirmatory Factor Analysis for Applied Research*. 2nd Ed. New York, NY: Guilford Press.

Cole, David A., and Kristopher J. Preacher. 2014. "Manifest Variable Path Analysis: Potentially Serious and Misleading Consequences Due to Uncorrected Measurement Error." *Psychological Methods* 19 (2): 300–315. https://doi.org/10.1037/a0033805.

Collins, Linda M., and Stephanie T. Lanza. 2009. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences.* Hoboken, NJ: Wiley.

Flake, Jessica Kay, and Eiko I Fried. 2020. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *Advances in Methods and Practices in Psychological Science* 3: 456–65.

Haslam, N., E. Holland, and P. Kuppens. 2012. "Categories Versus Dimensions in Personality and Psychopathology: A Quantitative Review of Taxometric Research." *Psychological Medicine* 42 (05): 903–20. https://doi.org/10.1017/S0033291711001966.

Hussey, Ian, and Sean Hughes. 2020. "Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology." *Advances in Methods and Practices in Psychological Science* 3 (2): 166–84. https://doi.org/10.1177/2515245919882903.

Kenny, David A., Deborah A. Kashy, and William L. Cook. 2006. *Dyadic Data Analysis*. New York, NY: Guilford Press.

Kim, Hanna, and Jee-Seon Kim. 2022. "Extending the Actor-Partner Interdependence Model to Accommodate Multivariate Dyadic Data Using Latent Variables." *Psychological Methods*, October. https://doi.org/10.1037/met0000531.

Kline, Rex B. 2023. *Principles and Practice of Structural Equation Modeling*. 5th Ed. New York, NY: Guilford Press.

Ledgerwood, Alison, and Patrick E. Shrout. 2011. "The Trade-Off Between Accuracy and Precision in Latent Variable Models of Mediation Processes." *Journal of Personality and Social Psychology* 101 (6): 1174–88. https://doi.org/10.1037/a0024776.

Little, Todd D. 2013. *Longitudinal Structural Equation Modeling*. New York, NY: Guilford Press.

Nye, Christopher D., and Fritz Drasgow. 2011. "Effect Size Indices for Analyses of Measurement Equivalence: Understanding the Practical Importance of Differences Between

Groups." *Journal of Applied Psychology* 96 (5): 966–80. https://doi.org/10.1037/a0022955.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716–16. https://doi.org/10.1126/science.aac4716.

Pastor, Dena A., Kenneth E. Barron, B. J. Miller, and Susan L. Davis. 2007. "A Latent Profile Analysis of College Students' Achievement Goal Orientation." *Contemporary Educational Psychology* 32 (1): 8–47. https://doi.org/10.1016/j.cedpsych.2006.10.003.

Rhemtulla, Mijke, Riet Van Bork, and Denny Borsboom. 2020. "Worse Than Measurement Error: Consequences of Inappropriate Latent Variable Measurement Models." *Psychological Methods* 25 (1): 30–45. https://doi.org/10.1037/met0000220.

Rosenberg, Joshua, Patrick Beymer, Daniel Anderson, C.j. Van Lissa, and Jennifer Schmidt. 2018. "tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software." *Journal of Open Source Software* 3 (30): 978. https://doi.org/10.21105/joss.00978.

Rosseel, Yves. 2012. "**Lavaan** : An *r* Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2). https://doi.org/10.18637/jss.v048.i02.

Sakaluk, John K., Alexandra N. Fisher, and Robyn E. Kilshaw. 2021a. "Dyadic Measurement Invariance and Its Importance for Replicability in Romantic Relationship Science." *Personal Relationships* 28 (1): 190–226. https://doi.org/10.1111/pere.12341.

———. 2021b. "Dyadic Measurement Invariance and Its Importance for Replicability in Romantic Relationship Science." *Personal Relationships* 28 (1): 190–226. https://doi.org/10.1111/pere.12341.

Sakaluk, John Kitchener. 2019. "Expanding Statistical Frontiers in Sexual Science: Taxometric, Invariance, and Equivalence Testing." *The Journal of Sex Research* 56 (4-5): 475–510. https://doi.org/10.1080/00224499.2019.1568377.

Sakaluk, John K., and Stephen D. Short. 2017. "A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources." *The Journal of Sex Research* 54 (1): 1–9. https://doi.org/10.1080/00224499.2015.1137538.

Spearman, C. 1904. ""General Intelligence," Objectively Determined and Measured." *The American Journal of Psychology* 15 (2): 201–92. https://doi.org/10.2307/1412107.

Tabachnick, Barbara G., and Linda S. Fidell. 2012. *Using Multivariate Statistics*. 6th ed. New York, NY: Pearson.

Vazire, Simine, Sarah R. Schiavone, and Julia G. Bottesini. 2022. "Credibility Beyond Replicability: Improving the Four Validities in Psychological Science." *Current Directions in Psychological Science* 31 (2): 162–68. https://doi.org/10.1177/09637214211067779.

A A Review of Sufficient *R* for Dyadic SEM