FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

Statistical Challenges In Modeling Longitudinal Dyadic Data

Fien Gistelinck

Supervisor: Prof. Dr. Tom Loeys

A dissertation submitted to Ghent University in partial
fulfillment of the requirements for the degree of
Doctor of Psychology

Academic year 2018–2019

GHENT
UNIVERSITY

FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

DEPARTMENT OF
DATA ANALYSIS

STATISTICAL CHALLENGES IN MODELING LONGITUDINAL DYADIC DATA

Fien Gistelinck

Supervisor: Prof. Dr. Tom Loeys

A dissertation submitted to Ghent University in partial
fulfillment of the requirements for the degree of
Doctor of Psychology

Academic year: 2018-2019

GHENT
UNIVERSITY

FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

DEPARTMENT OF
DATA ANALYSIS

Supervisor:	Prof. dr. Tom Loeys
Advisory committee:	Prof. dr. Marieke De Witte Prof. dr. Liesbet Goubert Prof. dr. Thierry Marchant Prof. dr. Axel Mayer
Examination committee:	Prof. dr. Mattias Desmet (chair) Prof. dr. Noémi Schuurman Prof. dr. Mariska Barendse Prof. dr. Jonas Lang Prof. dr. Jan De Neve

GHENT
UNIVERSITY

“To solve the human equation,
we need to add love,
subtract hate, multiply good,
and divide between truth and error.”

- Janet Coleman

Acknowledgments

As this dissertation proclaims, human life is defined by interactions. The combined effect of all these interactions led to this particular point in my life: the ending of my PhD, a journey filled with ups and downs. Needless to say, no journey is complete without acknowledging all the people involved.

First of all, I would like to thank my supervisor Prof. dr. Tom Loeys for the opportunity to lose myself in the statistics of psychology and behavioral science. Throughout the past four years, you managed to keep me on track, tempering my unrealistic suggestions. Your guidance, feedback and patience helped me grow, although I confess the stubbornness of my bombastic writing. I would like to thank each member of my advisory committee, Prof. dr. Marieke De Witte, Prof. dr. Liesbet Goubert, Prof. dr. Thierry Marchant and Prof. dr. Axel Mayer, for all the input and the interesting discussions we had during our annual meetings. I appreciate all the work and effort which helped me shape my PhD as it currently is. For the latter, I also thank the proofreaders of my dissertation.

Over the past four years, I spent a lot of my time at the office and none of that time would have been this enjoyable if I didn't have such a great colleagues. Justine, as insecure as we both were at the start, we encouraged each other along the wobbly road of our PhD. Lara, your enthusiasm gave me that little push to cease opportunities I otherwise would have passed. Ines, you made me feel welcome at work from the very first day. Your calm rationale helped me put things in more realistic perspectives. Isabelle, I admire your strength to keep things going for the entire department, while keeping an eye on every one of us. I greatly appreciate all the things you have done for me. Of course, I want to show my gratitude to all my outer-old-office colleagues as well: Han for your funny jokes and self-made music to brighten the mornings; Wouter for your advice and the after-work discussions; Heidelinde, Haeike, Jasper and Hannelore for making me feel at ease by just being there; Mariska and Wen Wei for making me appreciate the coffee breaks after dinner; etc. I want to thank the Krook-crew, Justine, Elissa and Freya for making the new location a comfortable working place. Thank you Elissa for indulging me at the lonely Krook-days, and Freya for helping me put my embarrassing self aside in order to meet new people who all helped me feel at home in the Krook: the pleasant people from IPEM, the friendly people from GUM and the nice people from the Krook-reception.

I also want to express my thanks to the MaStat-crew, Heidelinde, Arne, Tine and Robby: the sentence “What were we thinking” cannot describe the many late evenings before the deadlines; to the other MaStat students I met, Lisa, Annelies, Julien and Davor: I hope to meet you all again at one of the reunions; to my fellow-students of the SLO, Magali and Charlotte: you really helped me pull through the hard moments of the training; to my friends from dance: thank you for enduring my silliness during the Thursday evenings; and to Canteclear: I appreciate the harmony you radiate as a group. Next in line are those who were there even before the start of my PhD. Lisa, Lieselotte, Anneleen, Marjorie and Kelly, you kept supporting me over the many years and made sure to offer me a break every time I needed one. Thank you for all the laughs and tears that make life meaningful.

Last, but foremost, I would like to thank my family, my parents and all of my brothers, especially my youngest brother Koen. You all supported me and helped me during this PhD. I know I can be quite the annoying sister and the troublesome daughter, but you all had patience with me and loved me irregardless. I am certain I missed a lot of names, so for those who were there, through the good and the bad:

Thank you!

Fien,
August 2019

Table of Contents

List of Acronyms

A

APIM	Actor-Partner Interdependence Model
AR(1)	First-order AutoRegressive

B

BAY	Bayesian

I

IT	Indistinguishability Test

L

L-APIM	Longitudinal Actor-Partner Interdependence Model
LD-APIM	Lagged Dependent Actor-Partner Interdependence Model
LDD	Longitudinal Dyadic Data
LGC	Latent Growth Curve

M

ML	Maximum Likelihood
ML-AR(1)	MultiLevel First-order Autoregressive
MLM	MultiLevel Modeling

R

REML	REstricted or REsidual Maximum Likelihood
RM-APIM	Repeated Measurements Actor-Partner Interdependence Model

S

SEM	Structural Equation Modeling

V

VAR	Vector AutoRegression

1 Introduction

The goals of this dissertation are to investigate the different statistical challenges in modeling longitudinal dyadic data, to construct a model that deals with these issues, and to enable applied researchers to analyze this type of data by providing software tools. In this chapter, the definition of longitudinal dyadic data is explained and an overview of existing models to fit such data are introduced. The details of the subsequent chapters and their contribution towards the topic of longitudinal dyadic data are specified at the end of the chapter.

1.1 Dyadic data

Human life is defined by interactions. The way humans behave, work, communicate, experience life: it all occurs in the social context of life. Researchers investigating human behavior or emotions are often confronted with the fact that they have to take the social environment into account if they really want to understand what is going on. For example, developmental researchers often have to include the teachers, the parents and the peers from a child's social context when they study developmental problems in schoolchildren. As daily interactions with acquaintances or strangers often occur in pairs, the idea of examining pairs or *dyads* as a whole instead of independent individuals has become more and more conventional in social and behavioral science (DePaulo & Kashy, 1998). Relationship researchers will study married couples, family researchers will investigate parent-child dyads or sibling relationships, organizational psychologists study employer-employee negotiations, social psychologists engage in interracial interactions, etc.

Even though the dyad, consisting of only two individuals, is a simple social system, it is the most fundamental unit of interpersonal interactions and close relationships (Reis, Collins, & Bersheid, 2000). Love, conflict, person perception and aggression are only a few of the many theoretical concepts that are intrinsically dyadic in nature. The richness of research questions based on these dyadic interactions is stunning. Many of the standard methods in social science research are not able to address these questions correctly. This is mainly due to the fact that they are developed for the study of individuals. The discipline of psychology, which focuses on the individual, has dominated research in social and behavioral science for many generations (Bond & Kenny, 2002). More importantly, the traditional statistical methods exhibit an individualistic orientation by assuming independent sampling of the individuals. This *independence assumption* is violated in the case of dyadic data. For instance in the context of marital satisfaction, the possibility that the husband's responses are unrelated to the wife's responses is very slim as they inherently report on the same relationship (see Figure 1.1).

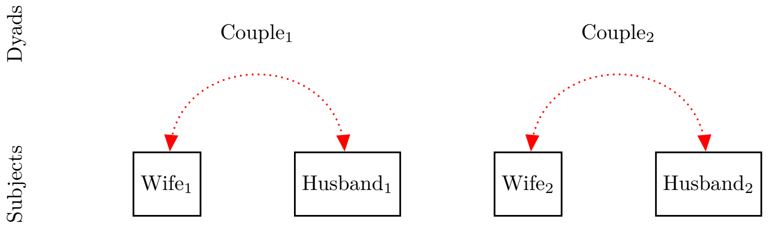


Figure 1.1 Two heterosexual couples from dyadic data. The red dotted lines illustrate the non-independence between the members of a dyad.

The idea that the two members of a dyad share something in common is one of the key features in dyadic data, and is referred to as *non-independence*, interdependence or linked scores. It means that the measurements of both members of a dyad are more (dis)similar than the measurements of two random people (Gonzalez & Griffin, 1999). There are different sources that can cause this heightened (dis)similarity. Compositional effects occur in married couples, as they are known to share similarities in education level, age, religion, etc. due to assortative mating (Epstein & Guttman, 1984). Mutual influence effects are illustrated by people who date: the less commitment is shown by a person's dating partner, the less committed that person will be in the relationship. Two roommates will show similar amounts of trust in their janitor because of common fate effects, namely, they share the same building (Kenny, 1996). From a statistical point of view, non-independence means that the two members of a dyad are correlated with one another, that is, there is a non-zero correlation between their scores. If one would ignore this correlation (i.e., assume it zero nonetheless), inference will be biased: the test statistics will have incorrect standard errors and the degrees of freedom will be inaccurate, affecting the Type I and Type II error rate (Kenny, Kashy, & Bolger, 1998). Formerly, researchers have dealt with this issue of non-independence by dropping one member's data or by aggregating the results on the dyad level (e.g., by taking the average of the two members scores). In the former, different conclusions are obtained as data from different members are discarded. In the latter, two very distinct scores are equated, leading to erroneous inference (Cook & Kenny, 2005). Alternatively, researchers have fitted two separate models, one for each role of dyad members, for example, one model for the mothers and one model for the children when investigating the bond between a child and her mother. This approach avoids the non-independence problem at a greater cost: it is no longer possible to answer questions with respect to dyadic effects (Planalp, Du, Braungart-Rieker, & Wang, 2017). Moreover, it presumes differences between both members of a dyad. In the case of mother-child bond, this sounds logical as they differ in age, life-experience, etc., but in the case of same-sex couples, this difference might not be as justifiable as in the previous case.

Conceptually, this non-independence also implies that the behavior of one member of a dyad is best understood if the researcher takes *the behavior of the other member* into account, aside from his or her own behavior. In other words, a second key feature of dyadic data incorporates the fact that members of the same dyad influence each other. For example, psychological control by the parent (i.e., manipulative and pressuring tactics that intrudes upon the child's psychological world) is highly related to internalizing problems and depressive feelings of the child (Barber & Harmon, 2002). This effect has typically been interpreted as a parenting effect, yet, one has to realize that the child's internalizing problems also affect the parental behavior. It may be possible that the adolescent maladjustment elicits psychological control of the parent over time (Soenens, Luyckx, Vansteenkiste, Duriez, & Goossens, 2008). In literature, the attribution error is often referred to as pseudo-unilaterality and the example illustrates the bidirectional influence fundamental to dyadic data. It is considered a deadly sin in dyadic data analysis to ignore one's own or one's partner's influence, that is, it is something a researcher should avoid doing when considering dyadic data (Kenny, Kashy, & Cook, 2006).

Treating the members of a dyad as (in)different from one another is a third key feature of dyadic data. A dyad is called *distinguishable* or nonexchangeable if there exists at least one variable that differentiates one member from the other. For example, in heterosexual couples, each dyad consist of a man and a woman (gender) and in industrial-organizational pairs, each dyad consist of a supervisor and an employee (status). It allows the researcher to introduce a systematic or meaningful order in the two scores of a dyad. However, there are situations in which no such natural distinction is available, for instance, in case of same-sex couples or coworkers. Those dyads are called *indistinguishable* or exchangeable dyads. Most data analytic techniques are more easily applied for distinguishable dyads. For this reason, researchers previously introduced an arbitrary variable into the data as a distinguishing variable. However, as this variable is not meaningful, either theoretically or empirically, it jeopardizes the validity of the results since it introduces unnecessarily complex models. In that case, the researcher may face convergence issues, power issues, and incorrect standard errors. Contrarily, if you treat dyads as indistinguishable while they were supposed to differ, then you are actually fitting the wrong model. Moreover, it is no longer possible to answer research questions that focus on distinguishability (e.g., questions about gender differences in social science) (Kenny et al., 2006).

1.1.1 Dyadic models

Over the last decade, researchers have developed several models that acknowledge these crucial aspects of dyadic data, mostly discriminated by the source that generates the non-independence within the dyads. The mutual influence model, also called the mutual-feedback model, allows for the two members of a dyad to directly affect each other's outcome scores (see Figure 1.2). These direct effects are called *reciprocal effects* and introduce a feedback loop in the model (Woody & Sadler, 2005). For instance, a person's commitment to the relationship (Y) is directly influenced by the amount of commitment shown by his or her partner. In this case, the independent variable is considered an *instrumental variable*: it predicts one of the outcome scores and is only related to the outcome score of the partner via the feedback loop. In the context of relationship commitment, such an instrumental variable could be the person's anxiety level of close relationships (X). The variability that cannot be explained by the predictor or the feedback loop is incorporated in the model by the *residuals* (ε). Note that these residuals are allowed to correlate due to the non-independence (see double headed arrow between the residuals).

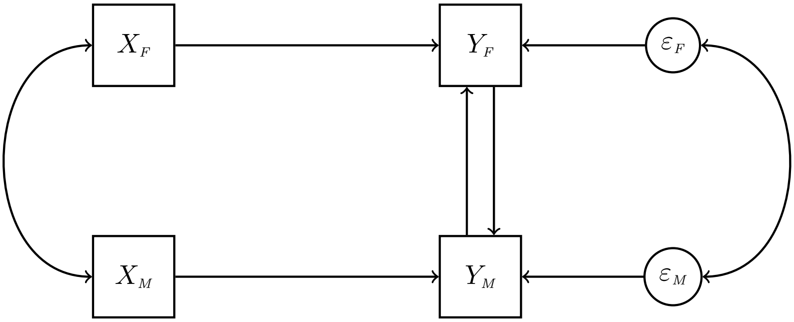


Figure 1.2 The mutual influence model for the effect of anxiety (X) on relationship commitment (Y) of heterosexual couples (M: male and F: female).

Common fate models, also called latent group models or dyadic-personality models, depicted by the path diagram in Figure 1.3, are used when there is a theoretical construct that affects both members (Ledermann & Kenny, 2015). Suppose you are interested in the effect of the trust that two siblings show in their mother (X) on the strength of the relation between both siblings (Y). The ratings for X (respectively Y) of the two siblings actually represent two facets of one and the same construct, namely trust in their mother (respectively relationship strength). These constructs are called *latent variables* as they represent variables that cannot be directly observed (the big circles in the path diagram). In this case, the residuals of the two members are allowed to correlate.

Combinations and modifications of dyadic models are possible too. For example, one could alter the mutual influence model so that the correlation between the residuals is replaced by direct effects from X_F to Y_M and from X_M to Y_F with 'F' the index for females and 'M' the index for males (Woody & Sadler, 2005). In Cheung (2009), a dyadic latent congruence model was introduced, which adapts the common fate model so that the latent variables represent the average and the difference of the dyad members' measurements, respectively.

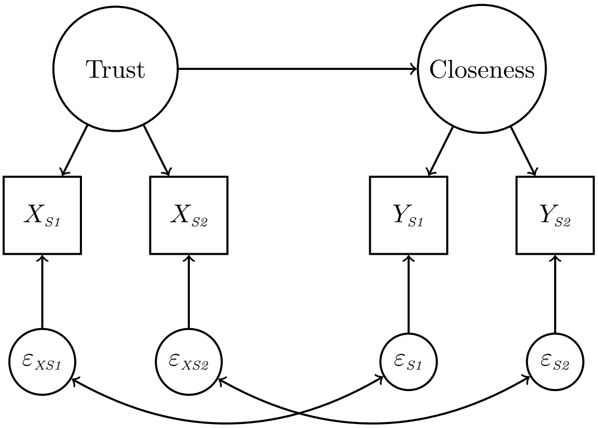


Figure 1.3 The common fate model for the effect of trust in their mother (X) on the relationship strength (Y) of two siblings (S1: oldest sibling and S2: youngest sibling).

The most dominant model in dyadic data analysis is the actor-partner interdependence model or APIM (Kenny et al., 2006). The APIM, illustrated by Figure 1.4, incorporates the fact that one has to acknowledge one's own (actor) and one's partner's (partner) characteristics, while keeping the correlation between both outcome scores caused by compositional effects into account (interdependence). To demonstrate the model, consider the study of Hinnekens, Stas, Gistelinck, and Verhofstadt (2019) investigating the association between actual, assumed and perceived understanding and the partner's levels of dyadic adjustment. The authors were interested, among other things, in whether the partner's levels of perceived understanding would be related to relationship satisfaction (in contrast to actual understanding). More specifically, does the way someone feels understood by his other partner have a high impact on dyadic adjustment? In this study, 152 cohabiting and married heterosexual couples were asked to rate the extent to which the respondents felt understood by their partner during a conflict discussion. Dyadic adjustment was measured using the Dutch questionnaire of the Dyadic Adjustment Scale (DAS). It reflects the degree to which the couple perceives that they agree on relationship aspects (dyadic consensus), that they are satisfied with their relationship (dyadic satisfaction), that they engage in common activities (dyadic cohesion), and that they express affections towards each other (affectional expression) (Buysse & Heene, 1997).

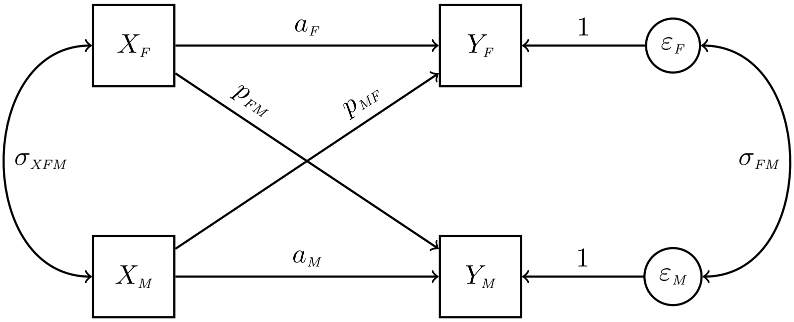


Figure 1.4 The actor-partner interdependence model for the effect of perceived understanding (X) on dyadic adjustment (Y) of heterosexual couples (M: male and F: female).

In the study of Hinnekens et al. (2019), dyads can theoretically be assumed distinguishable. As the authors interviewed heterosexual couples, gender is a meaningful variable that distinguishes both members from each other ('F' for females and 'M' for males). To explain the individual differences in dyadic adjustment, we explored the association between perceived understanding (PUN) and dyadic adjustment (DAS) using the APIM. In this case, the model can be expressed as the following multivariate regression,

$$\begin{cases} DAS_{Fj} = \mu_{YF} + a_F PUN_{Fj} + p_{MF} PUN_{Mj} + \varepsilon_{Fj} \\ DAS_{Mj} = \mu_{YM} + a_M PUN_{Mj} + p_{FM} PUN_{Fj} + \varepsilon_{Mj} \end{cases}, \tag{1.1}$$

with j referring to the dyad number ($j = 1, \ldots, 152$). The parameter estimates of the fitted APIM are shown in Table 1.1. The intercepts μ_F and μ_M represent the mean value of dyadic adjustment for females and males, respectively. The average dyadic adjustment for both dyad members is 118.65 for females, and 118.06 for males. Husbands and wives show a similarly high average level of satisfaction in their relationship. The *actor effects*, a_F and a_M, quantify the intrapersonal effect for the female member and the male member, respectively. Both actor effects are positively estimated ($a_F = 0.84$ and $a_M = 1.36$), which means that females and males report a higher relationship satisfaction when they perceive that their partner understands them. Moreover, the actor effect for males is higher than the one for females. If the husband believes that his wife fails to understand his feelings, then the impact on relationship satisfaction is more harmful than in case the wife feels misunderstood by her husband. The *partner effects*, p_{MF} and p_{FM}, quantify the interpersonal effects between both members of a dyad, ergo the double index. Following the convention for regression coefficients, the first index refers to the effect, while the second refers to the cause. So, as $p_{FM} = 0.26$ is not significant, it means the husband's relationship satisfaction is not significantly influenced by the fact that his wife feels (mis)understood by him. Conversely, $p_{MF} = 0.60$ indicates that there is a significant influence of the husband feeling (mis)understood by his wife on the dyadic adjustment score of the female partner. It is important for the wife's relationship satisfaction that the husband feels understood by her. Note that this importance is of a lesser extent than the wife herself feeling understood by her husband. The *residuals* or error terms, ε_F and ε_M, account for variation in relationship satisfaction that cannot be explained by perceived understanding. These residual variations are rather large ($\sigma_M^2 = 115.89$ for males and $\sigma_F^2 = 141.46$ for females). The disturbances are typically allowed to correlate in order to connote the compositional effects in the outcome variable. The correlation between the females and males adjustment score is $\frac{58.89}{\sqrt{115.89*141.46}} = 0.46$. It is obvious that if one would ignore this high, significant correlation, the above described results will change.

Table 1.1 Parameter estimates of the APIM for the effect of perceived understanding on dyadic adjustment.

	Estimate	(SE)	z-value	p-value	CI(95)
Effect on male outcome					
μ_M	118.65	0.90	132.58	<0.001	[116.89, 120.40]
a_M	1.36	0.21	6.43	<0.001	[0.95, 1.78]
p_{FM}	0.26	0.20	1.35	0.176	[-0.12, 0.65]
Effect on female outcome					
μ_F	118.06	0.99	119.91	<0.001	[116.13, 119.99]
a_F	0.84	0.21	3.93	<0.001	[0.42, 1.26]
p_{MF}	0.60	0.23	2.56	0.010	[0.14, 1.06]
Covariance parameters					
σ_M^2	115.89	13.22	8.76	<0.001	[89.97, 141.80]
σ_F^2	141.46	16.07	8.80	<0.001	[109.97, 172.95]
σ_{FM}	58.89	11.33	5.20	<0.001	[36.68, 81.09]

The APIM quantifies the interpersonal (partner effects) and intrapersonal effects (actor effects) for both members of a dyad, and it allows the researcher to compare them and test their significance. It is also possible to test other hypotheses, for instance, one might want to test whether this relationship context is either actor-oriented ($p_{MF} = 0 = p_{FM}$), partner-oriented ($a_F = 0 = a_M$), couple-oriented ($p_{MF} = a_F$ and $p_{FM} = a_M$) or social comparable ($p_{MF} + a_F = 0 = p_{FM} + a_M$) (Kenny & Cook, 1999). Symmetrical relations can be tested as well, referred to as *indistinguishability tests* or ITs. These tests empirically answer the question whether the distinguishing variable actually matters in the statistical analysis. For instance, as Table 1.1 suggests, the average score of the dyadic adjustment score is quite similar for males and females. In such a case, one might wonder whether gender induces distinguished average adjustment scores between both dyad members ($\mu_F = \mu_M$). One can evaluate such hypothesis by constraining the parameters in the model (i.e., restrict the values of these parameters) and by checking whether the model with or without the constraints fits the data better. In this case there is no empirical evidence of any gender difference between the mean dyadic adjustment value ($\chi^2_1 = 0.36$, p-value $= 0.55$).

The study of Hinnekens et al. (2019) uses a cross-sectional design: the two variables “perceived understanding” and “relationship satisfaction” are measured at one single time point. The APIM is also applicable when considering pre-post designs, when several variables are measured at two different time points, such as before and after an intervention. For example, one could measure the attachment style of therapist-client pairs and examine its effect on the working alliance reported after three weeks of therapy (Marmarosh et al., 2014). It is also possible that the same variable is measured for both dyad members at the two different time points. In Hunter, Fox, and Jones (2016), the authors investigated the influence of a best friends’ aggressive humor style to each others’ later use of the humor style. Here, the X in Figure 1.4 corresponds to aggressive humor at time point 1 and Y to aggressive humor at time point 2. The APIM applied on this type of pre-post design is often used to assess questions regarding stability (i.e., the actor effects) and reciprocity (i.e., the partner effects). It is equivalent to a bivariate autoregressive cross-lagged model for individuals with two time points. The latter fits two different outcome variables of the same person simultaneously and allows them to be correlated.

The APIM and the other above-mentioned dyadic models are used to explain individual differences in the outcome scores of a process caused by an antecedent. Yet, one might be interested in the process itself or in the way those associations between the predictor and the outcome unfold over time. For instance, instead of focusing on the characteristics of relationship satisfaction and the way it is affected by perceived understanding, focus may be on the way this effect changes over time. In this process approach or *process research*, questions are related to the underlying mechanisms. In order to reveal the multivariate nature of a process as well as its time dependency, dyads need to be measured repeatedly over time. If the interest lies in changes over time or in over-time differences between dyads, one needs to collect longitudinal dyadic data.

1.2 Longitudinal dyadic data

Consider the study of De Smet et al. (n.d.) which investigates different coping techniques to attenuate the effect of perceived intrusive partner behavior on individual well-being in the long run. More specifically, the authors interviewed 189 heterosexual couples at five different times points and measured, among other things, their perceived intrusive behavior and their stress level. The former was based on a questionnaire that rated how often they perceived their partner engaging in 11 types of privacy-invading behaviors (e.g., reading your text messages without consent, checking your pockets, or entering your room without knocking) (Vinkers, Finkenauer, & Hawk, 2011). The latter was based on the Perceived Stress Scale, measuring the degree to which situations in one's life are appraised as stressful (Cohen, Kamarck, & Mermelstein, 1984). In Figure 1.5, the standard design of the longitudinal dyadic data or LDD is presented. The key features of dyadic data are still extant in LDD: the upper part of the figure is the same as Figure 1.1. Two members of the same dyad are still connected due to their non-independence as the stress level of two married persons may be more (dis)similar than the one of two random persons. One still cannot ignore the impact of one's own and one's partner's characteristics. Thinking that your partner intrudes your privacy (partner) as well as being afraid to get busted after an act of intrusive behavior towards your partner (actor) will contribute to your level of stress. Lastly, (in)distinguishability is still a concept of interest.

Due to the longitudinal nature of the data, there are now two levels of variation: Level-2 or interindividual variation and Level-1 or intraindividual variation (see Figure 1.5). On one hand, the stress level might differ between members across different dyads as one can have a more stress resistant personality than the other (interindividual variation). On the other hand, a dyad member's stress level can show temporal fluctuations (intraindividual variation): a person may report more stress on a particularly hard day. From a statistical point of view, one can explicitly separate higher level sources of variation by fitting multilevel models, which uses random effect to do so (Hox, 2010). Unfortunately, the levels of variation are more complicated for LDD compared to a longitudinal individual case as they both inhibit non-independence. In fact, there are two kinds of non-independence to take into account: the non-independence that connects the two members of the same dyad and the non-independence between the repeated measures, also referred to as *autocorrelation* or temporal correlation. The scores of a variable measured at two different time points are associated to each other. For instance, the amount of stress a member experiences today is related to the amount of stress he or she experiences tomorrow, the day after tomorrow, the day thereafter, etc. Even though this association fades away as the time lag between the measurements increases, the autocorrelation cannot be ignored as this will lead to erroneous inference (Singer & Willett, 2003). This idea holds for each member of a dyad, but also across both members of the same dyad. As the time points are identical for each member of a dyad, these measurements of the LDD are correlated as well, as shown by the dotted arrows in Figure 1.5.

Formerly, researchers have dealt with LDD as if it originated from a three-level nested design. Indeed, there are three conceptual levels: time, individuals and dyads, but these levels do not correspond to the levels of analysis. Dyads are entirely defined by the two members it consists of. As a result, the individual level is saturated within the dyadic level and cannot generate any random variability (Diggle, Heagerty, Liang, & Zeger, 2002). Moreover, a three-level design constrains the correlation between both members of a dyad to be positive, which is not always the case in practice (e.g., one of two friends may compensate the unfriendly behavior of the companion by acting extra nice). Lastly, a three-level design assumes time and person to be nested: the correlation between two members of a dyad at each time point is expected to be zero. As mentioned above, this does not make any sense for LDD as it ignores time-specific sources of (dis)similarities within dyads (Kenny & Kashy, 2011).

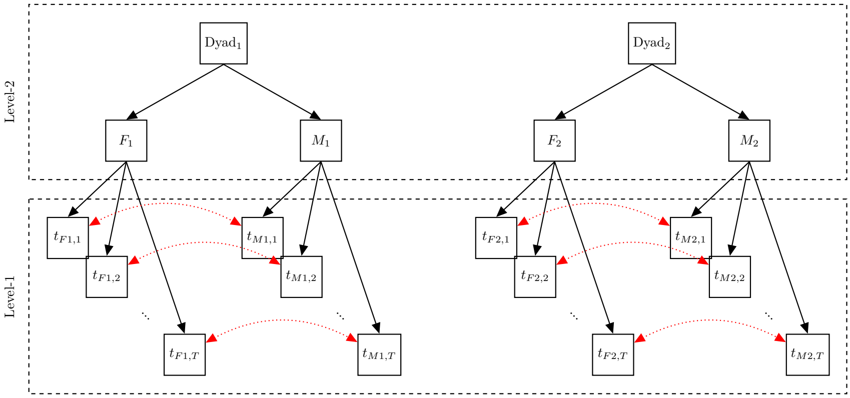


Figure 1.5 The two-level design of longitudinal dyadic data.

Thus, when analyzing LDD, one requires models that correctly acknowledges its two-level design by: (a) taking the interindividual variation into account due to the non-independence between the dyad members, (b) taking the intraindividual variation into account due to the non-independence between the time points, and (c) allowing the same time points across both members of a dyad to be correlated due to the non-independence between the dyad members.

1.2.1 Longitudinal dyadic models

The most logical way to solve a complex problem is to break it down. First consider a simple problem, solve it, and then extend the solution to the complex situation. Hence, several classes of models for longitudinal dyadic data already exist and originate from the longitudinal individual context. Indeed, it is possible to extend dynamic models from the individual to the dyadic context. Dynamic models use a set of equations to express the time dependency of a system. More specifically, it enables researchers to describe the present state of a system as a function of its previous state while presenting a rate and a direction of change. So, for the dyadic case, one has to implement a dynamic model that is able to simultaneously describe two processes, one for each dyad member, whilst allowing them to correlate at both Level-2 and Level-1.

Reconsider the study of De Smet et al. (n.d.) which focuses on the effect of the antecedent "intrusive behavior" on the outcome variable "stress" over time. If one is interested in questions such as "How does the husband's stress evolve over time?" or "What is the potential relation between the husband's trajectory and the wife's trajectory of stress?", latent growth curve (LGC) models for dyadic data could be used (Ferrer & McArdle, 2003). As Figure 1.6 shows, the dyadic LGC actually corresponds to two LGC models, one for females and one for males, which are allowed to correlate. The latent intercepts (I_F and I_M) represent the initial state of the process for each member, and the latent slopes (S_F and S_M) are used to represent the rate of change in stress, that is, its loadings λ define the shape of the trajectory. For example, if a constant increase in equally spaced time points is assumed, then the loadings are respectively fixed to 0,1,2, etc. The correlations between the latent intercepts and slopes incorporate the non-independence at Level-2, while the correlated residuals incorporates the non-independence at Level-1.

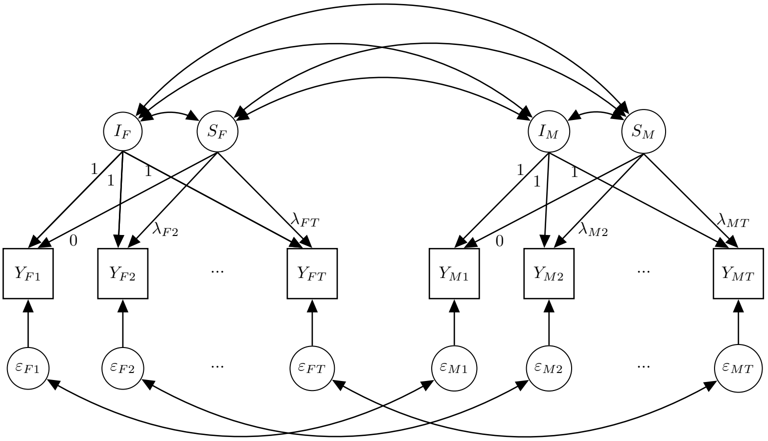


Figure 1.6 The dyadic latent growth curve model.

A class of models related to the LGCs is the latent change score model or latent difference score model. The latter provides information about the growth of each process and the way these processes are related, just like the LGC, but it also captures the order in which these processes influence each other by modeling the differences in change and by adding time-lagged effects between them (Estrada, Sbarra, & Ferrer, 2018).

The concept of directly implementing time-sequential dynamic associations is also reflected in the autoregressive cross-lagged models. Due to the bivariate nature of these models, the non-independence at Level-1 is already incorporated, see Figure 1.7. The models are usually referred to as the repeated measurement APIM or RM-APIM because they are equivalent to the pre-post design of the APIM, extended to more than two time points. These models are most often used to answer questions like “How is one’s own level of stress affected by one’s partner’s amount of stress experienced yesterday?” or “Is stress in both members of a dyad equally affected by yesterday’s amount of stress?”. Recently, Hamaker, Kuiper, and Grasman (2015) illustrated how this model can be extended with random intercepts to incorporate the non-independence at Level-2.

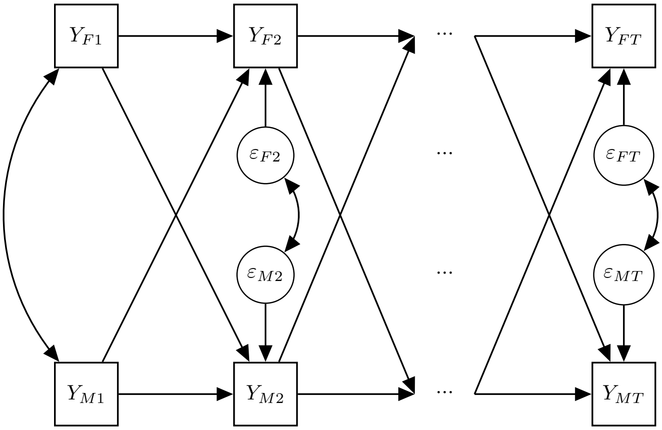


Figure 1.7 The repeated measurement actor-partner interdependence model.

The above-mentioned models for LDD succeed in capturing the temporal correlation in the outcome variable while acknowledging the non-independence due to the dyadic nature of the data. These models focus on questions related to the course and the shape of change. It allows researchers to examine when most of the change takes place, to compare the shape of change to an underlying theory, to examine how well an overall trajectory represents the individual trajectories, or to discover the differences in the trajectories between the two roles of the dyad members (Laurenceau, Hayes, & Feldman, 2007). However, these models do not allow to answer questions such as “How does the wife’s intrusive behavior affect the amount of stress experienced by the husband?” or “Does an increase in intrusive behavior of a person on a particular day (as compared to his or her average behavior) intensifies that person's stress level?". For these questions, the influence of a predictor on the trajectory is of higher importance than the trajectory itself. It requires the researcher to discover which variables clarify the shape of change.

1.2.2 A new class of longitudinal dyadic models?

To answer LDD questions about the influence of a predictor, one could argue that the above-mentioned models can be extended by allowing control variables. Bollen and Curran (2006) already discussed how to add explanatory variables to LGCs in the longitudinal individual setting. The same reasoning could be repeated when adding explanatory variables in the dyadic LGCs. However, transferring these extensions to the dyadic setting might not be as easy as it seems. In Nestler, Grimm, and Schönbrodt (2015), the authors demonstrate how to incorporate the effect of a single time-invariant dyadic variable into the dyadic LGC model, in other words, a variable measured on the dyad level (i.e., the score for both dyad members) that does not depend on time. But how do we handle time-varying dyadic predictors? And how do we deal with predictors measured at the member-level (i.e., the score is different between both dyad members)?

Above, models from the longitudinal individual setting are extended towards the dyadic setting. It involves facing the statistical challenges of extending cross-sectional individual models to dyadic models, together with new obstacles associated with the fact that these challenges are now situated in the longitudinal context. In this dissertation, we opted for a different approach. Here, dyadic models will be adapted towards the longitudinal setting. This way, the dyadic features are more easily preserved. The adaptation to the longitudinal setting can be based on approaches from the individual setting to handle the temporal characteristics. As mentioned before, the APIM is considered as one of the most popular models when considering dyadic data. Therefore, in this dissertation, the APIM was chosen to be extended towards the longitudinal setting.

1.3 Objectives

The goal of this project was to extend the APIM from the cross-sectional to the longitudinal setting, while dealing with the major issues associated with this improvement. The scope of this dissertation was limited to the extension of the APIM (and not the common fate model or the mutual influence model) because the APIM can be considered as one of the most prominent models within the dyadic research. As its latest bibliography update shows, more than 1,100 articles fitted the cross-sectional APIM on dyadic data as of April 2017, and its popularity is still exponentially increasing (Stas, Gistelinck, Ackerman, Garcia, & Kenny, 2017).

As one will notice in the next chapters, the extensions of the APIM for LDD are introduced in the context of distinguishable dyads. Of course, researchers first have to investigate whether or not these extensions should be altered for indistinguishable dyads. As mentioned before, this means that one has to perform indistinguishability tests before fitting any of these models. Indistinguishability might be considered as a theoretically driven decision (e.g., based on the design, previous research, or in case the research question involves assuming the dyads to be (in)distinguishable), but it can also be considered as an empirical question that allows researchers to increase power or to obtain simpler models, easier to interpret. Although indistinguishability tests are not novel for dyadic researchers, there are still some questions related to the subject that have not been covered yet. For example, indistinguishability tests are often performed within the structural equation modeling (SEM) framework. This framework is a multi-equation technique in which variables are allowed to be latent or unmeasured, and it easily implements constraints on the model parameters. Within SEM, model parameters are fitted using the maximum likelihood (ML) estimator in most software packages by default. Consequently, indistinguishability tests for (co)variance components of the model can become liberal due to the bias towards zero in the ML estimation of variance parameters. The restricted maximum likelihood (REML) estimator produces unbiased (co)variance estimates and might be preferred instead. Hence, the question arises how one could incorporate REML estimation in indistinguishability testing.

Once the researcher knows whether the dyads should be treated as (in)distinguishable, he or she can fit an appropriate model on the dyadic data. In case dyads are measured repeatedly over time, one needs longitudinal dyadic models to fit the data. As this dissertation focuses on research questions about the effect of an antecedent on the current behavior of each dyad member, we attempt to extend the cross-sectional APIM towards the longitudinal setting. Here, a major concern regards the covariance structure, which now has to incorporate two types of non-independence: one due to the dyadic nature of the data and one due to the repeated measures design of the data. Time-invariant sources of variability (i.e., Level-2 variation) are easily captured by a random effect/latent variable for each dyad member, which are allowed to be correlated. This is easily obtained in either the MLM or SEM framework. However, standard software packages will assume the residuals to be independent from one another. Hence, the challenge lies in modeling the time-specific covariance structure in combination with these random effects/latent variables.

In order to deal with temporal correlation, two major approaches can be distinguished in the individual setting. So, we adopted these progressions in order to expand the APIM towards LDD. The first progression involves the specification of a complex residual covariance structure that allows the residuals in the model to be correlated within dyad members as well as across dyad members. This progression yields marginal parameter estimates for the effect of an antecedent on an outcome variable, but requires a technical evolution in existing software packages in order to deal with this complex covariance structure. For instance, in social and behavioral science, researchers often assume a first-order autoregressive or AR(1) covariance structure within an individual. It defines the correlation between two outcome scores of a person to fade away as the time gap grows between both measurement occasions. This decrease in correlation is represented by the increasing exponent of a correlation parameter ρ $(-1 < \rho < 1)$. However, the covariance structure of the residuals for dyads is a lot more complex than the one for individuals. Most MLM software packages, often preferred when fitting longitudinal data, are not prepared to deal with such level of complexity. As a result, the SEM framework could be used instead, which is much more flexible than the MLM framework. Using standard SEM software packages, one can define any type of covariance structure. As shown later, this approach is computationally intensive, especially when the amount of time points becomes quite large. Hence, an alternative approach within the SEM framework was considered.

The second progression to incorporate temporal correlation works conditionally on the previous time point. In this case, the residuals can still be assumed independent from each other, hence, no (major) adaptations of the statistical software packages are expected. Yet, parameter estimates for the effect of an antecedent on an outcome variable are now interpreted conditionally on the previous outcome score. Within the dyadic context, this would imply extending the APIM with a *lagged dependent variable* in the mean structure of the model (i.e., adding the outcome score of the previous time point as a predictor for the current outcome score). Although this seems straightforward, it is not in combination with random effects in the model. Most MLM software packages follow the *exogeneity assumption*, that is, they expect all exogenous variables (including the lagged dependent outcome variables) to be independent from the random effects. This is a false assumption in the case of lagged dependent outcome variables as they still contribute to the random intercept of the model. Violating this assumption leads to biased inference, especially when the amount of time points is rather small. To resolve this bias, an alternative approach was investigated within the SEM framework.

Aside from the non-independence due to the dyadic design and the non-independence due to the repeated measures, there are other statistical challenges to consider. For instance, the effect of an overtime predictor can potentially have a different direction and/or magnitude when comparing its *time-averaged* (i.e., averaged over time) to its *time-specific* effect. In the study of De Smet et al. (n.d.), an increase in the average intrusive behavior of the partner (i.e., time-averaged effect) might only slightly affect one's average level of stress as he or she keeps experiencing this behavior as "normal" for the partner and acquiesces in the increase. However, if one's partner shows an increase in intrusive behavior on a particular day compared to the partner's average (i.e., time-specific effect), he or she might sense something is going on, leading to a significantly higher level of stress. Although this issue can easily be resolved, it harbors some additional technical adaptations when considering the APIM's extensions within the SEM framework.

As one notices, things become quite technical in this introduction. These statistical and technical complexities often form a barrier and discourage researchers to apply these more sophisticated models on their data. To counteract this, the longitudinal extensions of the APIM need to be more available and applicable for the user. Therefore, we made it our goal to develop a user-friendly free web application, called the *LDDinSEM*-application. It allows anyone to easily define the extensions of the APIM based on both progressions using the point-and-click interface. Afterwards, the application automatically fits the model on the uploaded data. Extra information is provided at each step of the application. Moreover, after fitting the model, a summary including the results can be downloaded. Furthermore, tutorials are provided containing a detailed example with interpreted results.

1.4 Outline

Chapter 2: ITs for the APIM

Before fitting dyadic models, the researcher has to decide whether or not the dyads should be assumed indistinguishable. This decision can be based on theory in case prior research showed evidence for differences between the two roles of the dyad members, or in case the design of the data provides a meaningful variable that distinguishes both members of the dyad. Alternatively, this decision can be answered empirically using indistinguishability tests. Treating indistinguishable dyads as distinguishable without checking whether this assumption holds, is considered one of the seven deadly sins according to Kenny et al. (2006). Although this sin is considered in the context of cross-sectional data, the same question arises when dyads are measured repeatedly over time. Hence, it is obvious that a dissertation considering models for longitudinal dyadic data can not avoid the topic of indistinguishability tests. This chapter discusses the different types of indistinguishability tests for the APIM in both the cross-sectional and a simple longitudinal setting. It addresses the differences in Type I error when REML is used instead of ML, and it shows the effect on the Type I error when splitting up an indistinguishability test in different sub-tests. Lastly, the chapter illustrates how to implement the indistinguishability tests in the SEM and MLM framework.

Chapter 3: L-APIM in SEM

As mentioned in the previous section, the goal of this dissertation is to define an APIM-based model to fit LDD. In this chapter, we extend the APIM according to the first progression: the longitudinal actor-partner interdependence model or L-APIM. The L-APIM (a) allows for non-independence between the outcomes of two members of a dyad, (b) takes the non-independence between the different time points within and between the two members of a dyad into account, (c) includes the effect of both the actor and partner characteristics on the behavior of a person, and (d) allows to separate the overtime predictors into a time-specific and time-averaged effect. Although the model can readily be fitted within the MLM framework using the software package SAS (SAS Institute, 2008), its implementation within the SEM framework is less straightforward. Moreover, the implementation based on crude coding takes a lot of time to fit when the amount of time points is large. So, an alternative approach is needed as well. In order to facilitate the use of the L-APIM by any applied researcher, we developed a Shiny-application, an application based on the R-package Shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2017), using this alternative implementation. As the chapter shows, the app allows the user to define and to fit the L-APIM in a more reasonable amount of time via an online point-and-click interface. Thanks to this user-friendly application, there is no need to download any statistical software or to buy any software licenses.

Chapter 4: Multilevel-AR(1) when T is small

Before extending the APIM towards the longitudinal setting based on the second progression, a more thorough investigation of the statistical challenges involving this progression is necessary. As a result, the individual setting is first considered, in the absence of any predictors. This way, the three major issues with respect to lagged dependent variables in the mean structure are identified: (a) the initial conditions problem, (b) the centering problem, and (c) the endogeneity problem. The first issue relates to the idea that “if one conditions on the previous outcome score, on what does one condition at the initial time point?”. The second issue relates to the fact that if one conditions on the previous outcome score, the interpretation of the intercept changes. A correct centering approach is required to regain the interpretation of the intercept as the overall mean value. The last issue corresponds to the bias introduced in the model as soon as the exogeneity assumption is violated. In this chapter, different models are investigated, each dealing with these issues in their own way and it is shown how these models are implemented in either the MLM or SEM framework. Alternatives within the Bayesian (BAY) modeling framework are considered as well, broadening the investigation beyond the frequentist approach. The BAY framework is often preferred over the frequentist approach when the sample size is small (Bolstad & Curran, 2016).

Chapter 5: LD-APIM in SEM

After examining the different issues involving the second progression, the newly gained knowledge is expanded to the longitudinal dyadic setting. We adopt the conclusions about best practices from the previous chapter in order to extend the APIM with lagged dependent variables. The extension is called the lagged dependent actor-partner interdependence model or LD-APIM. Again, the model incorporates all four statistical challenges of LDD, similar to the L-APIM (see chapter 3), but now the second progression is used to deal with the temporal correlation. To improve the accessibility of this advanced model, the Shiny-application *LDDinSEM* from the previous chapter is updated. Now, users can use the application to fit both the L-APIM and the LD-APIM. As most technicalities are already explained by either chapter 3 or chapter 4, this chapter focuses on the applicability of the LD-APIM. To be more specific, the *LDDinSEM*-application is promoted in this chapter: we illustrate how the LD-APIM can be defined using the app in only three steps, and show how the results obtained by the app are interpreted. This way, researchers are encouraged to fit the APIM extensions via the app, without having to be restricted by the computational and technical concepts.

Chapter 6: General discussion

The longitudinal dyadic models introduced in this dissertation are not perfect: they make assumptions that may be relaxed, their applicability may be limited to certain context, etc. Consequently, in this chapter, we dwell upon the contributions, limitations and the practical implications of the research executed in the previous chapters. There are also some unresolved issues related to the subject of this dissertation that need further exploration. Therefore, this chapter ends with recommendations for future research.

Note, as the chapters 2-5 are written as research papers and are supposed to stand on their own, some overlap between the content of the different chapters may appear. Consequently, notation is chapter-specific and should not be compared between the different chapters. Data storage fact sheets can be found in the last chapter of this dissertation.

References

Barber, B., & Harmon, E. (2002). Violating the self: Parental psychological control of children and adolescents. In B. Barber (Ed.), *Intrusive parenting: How psychological control affects children and adolescents* (p. 15-52). Washington, DC: American Psychological Association.

Bollen, K., & Curran, P. (2006). *Latent curve models: A structural equation perspective.* Hoboken, NJ: John Wiley & Sons, Inc.

Bolstad, W., & Curran, J. (2016). *Introduction to bayesian statistics* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Bond, C., & Kenny, D. (2002). The triangle of interpersonal models. *Journal of Personality and Social Psychology*, *83*(2), 355-366. doi: 10.1037/0022-3514.83.2.355

Buysse, A., & Heene, E. (1997). *Dutch version of the dyadic adjustment scale (das, spanier, 1976).* (University of Ghent, Faculty of Psychology)

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). Shiny: Web application framework for R [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=shiny`

Cheung, G. (2009). Introducing the latent congruence model for improving the assessment of similarity, agreement, and fit in organizational research. *Organizational Research Methods*, *12*(1), 6-33. doi: 10.1177/1094428107308914

Cohen, S., Kamarck, T., & Mermelstein, R. (1984). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385-396. doi: 10.2307/2136404

Cook, W., & Kenny, D. (2005). The actor-partner interdependence model: A model of bidirectional effects in developmental studies. *International Jounal of Behavioral Development*, *29*(2), 101-109. doi: 10.1080/01650250444000405

DePaulo, B., & Kashy, D. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63-79. doi: 10.1037/0022-3514.74.1.63

De Smet, O., Buyukcan-Tetik, A., Gistelinck, F., Verhofstadt, L., Finkenauer, C., Loeys, T., & Buysse, A. (n.d.). *How to cope with intrusive behavior of the partner.* (University of Ghent, Faculty of Psychology)

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data.* Oxford, UK: Oxford University Press.

Epstein, E., & Guttman, R. (1984). Mate selection in man: Evidence, theory, and outcome. *Social Biology*, *31*(3-4), 243-278. doi: 10.1080/19485565.1984.9988579

Estrada, E., Sbarra, D., & Ferrer, E. (2018). Models for dyadic data. In A. Wright & M. Hallquist (Eds.), *Handbook of research methods in clinical psychology.* Cambridge, US: Cambridge University Press.

Ferrer, E., & McArdle, J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, *10*(4), 493-524. doi: 10.1207/S15328007SEM1004_1

Gonzalez, R., & Griffin, D. (1999). The correlation analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*(4), 449-469. doi: 10.1037/a0025651

Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. doi: 10.1037/a0038889

Hinnekens, C., Stas, L., Gistelinck, F., & Verhofstadt, L. (2019). “i think you understand me” Studying the associations between actual, assumed, and perceived understanding within couples. *European Journal of Social Psychology.* doi: 10.1002/ejsp.2614

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). London, UK: Routledge.

Hunter, S., Fox, C., & Jones, S. (2016). Humor style similarity and difference in friendship dyads. *Journal of Adolescence*, *46*, 30-37. doi: 10.1016/j.adolescence.2015.10.015

Kenny, D. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships*, *13*(2), 279-294. doi: 10.1177/0265407596132007

Kenny, D., & Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, *6*(4), 433-448. doi: 10.1111/j.1475-6811.1999.tb00202.x

Kenny, D., & Kashy, D. (2011). Dyadic data analysis using mutlilevel modeling. In J. Hox & J. Roberts (Eds.), *Handbook of advanced multilevel analysis* (p. 335-370). London, UK: Routledge.

Kenny, D., Kashy, D., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (p. 233-265). Boston, MA: McGraw-Hill.

Kenny, D., Kashy, D., & Cook, W. (2006). *Dyadic data analysis.* The

Guildford Press.

Laurenceau, J.-P., Hayes, A., & Feldman, G. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, *27*(6), 682-695. doi: 10.1016/j.cpr.2007.01.007

Ledermann, T., & Kenny, D. (2015). The common fate model for dyadic data: Variations of a theoretically important but underutilized model. *Journal of Family Psychology*, *26*(1), 140-148. doi: 10.1037/a0026624

Marmarosh, C., Kivlighan, D., Bieri, K., LaFauci Schutt, J., Barone, C., & Choi, J. (2014). The insecure psychotherapy base: Using client and therapist attachment styles to understand the early alliance. *Psychotherapy*, *51*(3), 404-412. doi: 10.1037/a0031989

Nestler, S., Grimm, K., & Schönbrodt, F. (2015). The social consequences and mechanisms of personality: How to analyse longitudinal data from individual, dyadic, round-robin and network designs. *European Journal of Personality*, *29*(2), 272-295. doi: 10.1002/per.1997

Planalp, E., Du, H., Braungart-Rieker, J., & Wang, L. (2017). Growth curve modeling to studying change: A comparison of approaches using longitudinal dyadic data with distinguishable dyads. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(1), 129-147. doi: 10.1080/10705511.2016.1224088

Reis, H., Collins, W., & Bersheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*(6), 844-872. doi: 10.1037/0033-2909.126.6.844

SAS Institute, I. (2008). *SAS/STAT 9.2 user's guide: The MIXED porcedure.* Cary, NC: SAS Institute Inc.

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

Soenens, B., Luyckx, K., Vansteenkiste, M., Duriez, B., & Goossens, L. (2008). Clarifying the link between parental psychological control and adolescents' depressive symptoms: Reciprocal versus unidirectional models. *Merrill-Palmer Quarterly*, *54*(4), 411-444. doi: jstor.org/stable/23096266

Stas, L., Gistelinck, F., Ackerman, R., Garcia, R., & Kenny, D. (2017). Giving dyadic data analysis away with DyadR: Free online shiny apps for the apim. In *Aps 29th annual convention.*

Vinkers, C., Finkenauer, C., & Hawk, S. (2011). why do close partners snoop? predictors of intrusive behavior in newlywed couples. *Personal Relationships*, *18*(1), 110-124. doi: 10.1111/j.1475-6811.2010.01314.x

Woody, W., & Sadler, P. (2005). Structural equation models for interchangeable dyads: being the same makes a difference. *Psychological Methods*, *10*(2), 139-158. doi: 10.1037/1082-989X.10.2.139

2

Indistinguishability tests in the actor–partner interdependence model

Abstract. When considering dyadic data, one of the questions is whether the roles of the two dyad members can be considered equal. This question may be answered empirically using indistinguishability tests in the actor–partner interdependence model. In this paper several issues related to such indistinguishability tests are discussed: the difference between maximum likelihood and restricted maximum likelihood based tests for equality in variance parameters; the choice between the structural equation modelling and multilevel modelling framework; and the use of sequential testing rather than one global test for a set of indistinguishability tests. Based on simulation studies, we provide guidelines for best practice. All different types of tests are illustrated with cross-sectional and longitudinal data, and corroborated with corresponding R code.

This chapter has been published in British Journal of Mathematical and Statistical Psychology. Gistelinck, F., Loeys, T., Decuyper, M. & Dewitte, M. (2018). Indistinguishability tests in the actor-partner interdependence model. *British Journal of Mathematical and Statistical Psychology, 71(3)*, 472-498.

2.1 Introduction

Historically, psychological data were mostly gathered on individuals, yet many of the phenomena studied by social and behavioural scientists are interpersonal by definition. Indeed, the daily lives of individuals are rarely spent in isolation from others, and arguably the most important context of daily life is that defined by close dyadic relationships (Reis, Collins, & Bersheid, 2000). The richness of dyadic data is enabling researchers to broaden their horizons and contemplate addressing research questions of increasing complexity. Hence, these are unprecedented times for the social sciences in terms of the availability of high-quality dyadic data.

When two people interact in a relationship, the outcome of each person can be affected by both his or her own inputs and his or her partner's inputs. The actor–partner interdependence model (APIM) offers an appealing approach to modelling such dyadic data (Kenny, Kashy, & Cook, 2006) when the same measurements are taken from both dyad members. An *actor* effect occurs when a person's score on a predictor variable affects that same person's score on an outcome; a *partner* effect occurs when a person's score on a predictor variable affects his or her partner's score. Consider for example a recent study by Decuyper, Gistelinck, Vergauwe, Pancorbo, and De Fruyt (2016) who explored the association between personality pathology and relationship satisfaction in 52 heterosexual couples. Relying on the APIM, these authors explored, *inter alia*, the effect of antagonism (i.e., showing hate, extreme unfriendliness, or active opposition to the partner) on one's own and one's partner's relationship satisfaction in males and females. The corresponding path model of this APIM is shown in Figure 2.1, with X representing the predictor variable (antagonism in our example) and Y the outcome (relationship satisfaction in our example).

An important question in dyadic research and data analysis is whether or not the two dyad members can be distinguished by some variable. Clearly, in our example with heterosexual couples, dyad members are theoretically distinguishable because of their gender. However, distinguishability is an empirical issue too, and the defining question is whether there are differences in the data for the two 'types' of partners (Gonzalez & Griffin, 1999). In other words, does gender really matter in our illustration? Olsen and Kenny (2006) describe in detail different types of distinguishability in the APIM: are the actor (and partner) effects the same between males and females, are the (residual) variances in the outcome

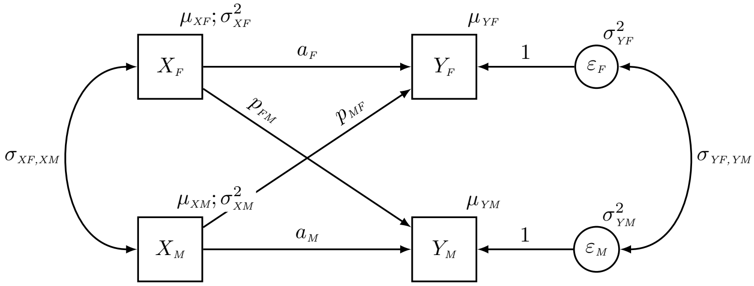


Figure 2.1 The Actor Partner Interdependence Model (APIM) used to analyse cross-sectional dyadic data. The index ‘M’ refers to the first dyad member (e.g., male), while ‘F’ refers to the second member (e.g., female).

the same across gender, are the mean and the variance of the predictors the same between males and females, etc. Assuming dyad members to be distinguishable without testing this assumption is called one of the ‘seven deadly sins’ in dyadic data analysis by Kenny et al. (2006). It is true that assuming distinguishability when it is not present in the data might jeopardize the validity of the results as it introduces unnecessary complex models. Consequently, the model might not converge, results will suffer from power issues, standard errors will deviate (especially if the covariance structure is misspecified) and the interpretation of the results will become more complicated.

Although accessible applications exist that allow relationship researchers to perform certain types of indistinguishability tests (Kenny, 2017), several issues deserve further attention. First, the choice of the estimation method may play a role. Estimation based on maximum likelihood (ML) is very prominent in many modelling frameworks. However, it is well known that ML estimates of variance components are biased towards zero, and that restricted or residual maximum likelihood (REML) estimation may be preferred instead (McNeish, 2017). Second, it should be noted that two different modelling frameworks are often used for modelling dyadic data (Ledermann & Kenny, 2017): the multilevel modelling (MLM) and the structural equation modelling (SEM) framework. Third, the test of indistinguishability may consist of several sub-tests. That is, one may be interested in testing equality in means and equality in variances for the outcome and/or the predictors of interest. The question then arises whether one should prefer a single global test, or one should rather go for multiple sub-tests (with or without multiplicity correction).

In this paper we take a closer look at how such indistinguishability tests can best be performed. The paper is organized as follows. We first describe the APIM in more detail in the cross-sectional setting, and explain the several types of indistinguishability that can empirically be tested in this model. We next discuss the three aforementioned issues (REML vs. ML, SEM vs.MLM, and global vs. sequential testing). We show that tests based on ML and REML may yield different results when one wants to test the equality of variances between males and females, for example. The pros and cons of both SEM and MLM approaches are discussed in light of indistinguishability tests, and clear practical guidance with best practice is provided in terms of global and sequential tests. Since an increasing number of publications consider longitudinal dyadic data, we investigate how our findings from the cross-sectional setting translate to the more complex longitudinal setting. To facilitate the implementation of the indistinguishability tests discussed in this paper, we provide two examples with R code available as Appendix 2.A. More specifically, we illustrate the different indistinguishability tests for a cross-sectional study and a longitudinal diary study in couples. We conclude with a discussion.

2.2 The cross-sectional APIM and tests for indistinguishability

Consider the APIM, depicted in a path diagram in Figure 2.1. For the sake of simplicity, throughout this paper we will use the subscripts M and F for males and females to denote two types of partners in a dyad, but other types of distinguishable dyads (e.g., brother and sister, oldest and youngest child) can obviously be considered as well. The parameters X_F and Y_F are the predictor and outcome variable for the females, and X_M and Y_M are the predictor and outcome variable for their male partner, respectively. The outcome disturbances are modelled as unobserved variables for both the females (ε_F) and the males (ε_M). These disturbances have zero means by assumption, and are allowed to correlate. The effect of the female's predictor on her own outcome (a_F) and the effect of the male's predictor on his own outcome (a_M) are referred to as *actor* effects. The effect of the male's predictor on his female partner's outcome (p_{MF}) and the effect of the female's predictor on her male partner's outcome (p_{FM}) are referred to as *partner* effects. Other parameters in the model include the male and female predictor means (μ_{XM}

and μ_{XF}), their variances (σ^2_{XM} and σ^2_{XF}), the covariance between these predictors ($\sigma_{XF,XM}$), the male and female outcome intercepts (μ_{YF} and μ_{YM}), the residual variances (σ^2_{YM} and σ^2_{YF}) and the residual covariance of their outcomes ($\sigma_{YF,YM}$). The model in Figure 2.1 can be expressed as a multivariate regression,

$$\begin{cases} Y_{Fj} = \mu_{YF} + a_F X_{Fj} + p_{MF} X_{Mj} + \varepsilon_{Fj} \\ Y_{Mj} = \mu_{YM} + a_M X_{Mj} + p_{FM} X_{Fj} + \varepsilon_{Mj} \end{cases}, \tag{2.1}$$

with j referring to the dyad number ($j = 0, \ldots, N$), and the residuals are assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_{Fj} \\ \varepsilon_{Mj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{YF} & \sigma_{YF,YM} \\ \sigma_{YF,YM} & \sigma^2_{YM} \end{pmatrix} \right).$$

In principle, several predictor variables may be considered in model (2.1), but for simplicity we will limit further discussions to one predictor variable X. Model (2.1) may be fitted within the MLM and SEM frameworks. Both frameworks yield identical ML estimators for the actor and partner effects and the residual (co)variances (Curran, 2003). Alternatively, instead of fitting a multivariate regression model, one may use a random intercept model to capture the correlation of outcomes within the dyad, for example relying on the model:

$$\begin{aligned} Y_{ij} = {} & \beta_{M0} + \beta_{M1} X_{act,ij} + \beta_{M2} X_{par,ij} \\ & + \beta_{D0} G_{ij} + \beta_{D1} X_{act,ij} G_{ij} + \beta_{D2} X_{par,ij} G_{ij} + \eta_j + \varepsilon_{ij}, \end{aligned} \tag{2.2}$$

with $i = M, F$, j the dyad number ($j = 0, \ldots, N$), $G_{ij} = 0$ for males and $G_{ij} = 1$ for females, $X_{act,ij}$ one's own predictor score and $X_{par,ij}$ one's partner's predictor score, and $\varepsilon_{ij} \sim N(0, \sigma^2_Y)$ and $\eta_j \sim N(0, \tau^2)$. In model (2.2), β_{M1} and β_{M2} represent the actor and partner effects in males, while β_{D1} and β_{D2} represent the difference between females and males in actor and partner effects, respectively. This approach is sometimes referred to as the ‘interaction’ approach (Kenny et al., 2006). Alternatively, one could use a two-intercept approach within the MLM framework (Kenny et al., 2006):

$$\begin{aligned} Y_{ij} = {} & \beta_{M0} M_{ij} + \beta_{M1} X_{act,ij} M_{ij} + \beta_{M2} X_{par,ij} M_{ij} \\ & + \beta_{F0} F_{ij} + \beta_{F1} X_{act,ij} F_{ij} + \beta_{F2} X_{par,ij} F_{ij} + \eta_j + \varepsilon_{ij}, \end{aligned} \tag{2.3}$$

with $i = M, F$, j the dyad number ($j = 0, \dots, N$), $M_{ij} = 1$ for males and 0 for females, and $F_{ij} = 1$ for females and 0 for males. In model (2.3), β_{M1} and β_{M2} now represent the actor and partner effects in males, while β_{F1} and β_{F2} represent the actor and partner effects in females, respectively. Hence, the effects in model (2.3) are directly linked with those in model (2.1), while model (2.2) allows for an immediate comparison of the actor and partner effects between both dyad members. As shown in Loeys and Molenberghs (2013)), model (2.2) and (2.3) make stronger assumptions than model (2.1) as they imply equal variability in males and females ($\sigma^2_{YM} = \sigma^2_{YF} = \sigma^2_Y$) and assume a positive correlation within dyads (i.e., the covariance is given by the variance of a random intercept η_j). Since we want to start from the most general model, we will further focus on model specification (2.1) instead of (2.2) or (2.3) in this paper.

In their paper on fitting APIMs with interchangeable dyads, Olsen and Kenny (2006) discussed different types of indistinguishability. Considering model (2.1), we distinguish the following indistinguishability tests (ITs):

- (IT1) equal residual variances for the outcomes: $\sigma^2_{YM} = \sigma^2_{YF}$
- (IT2) equal actor effects, $a_M = a_F$;
- (IT3) equal partner effects, $p_{MF} = p_{FM}$;
- (IT4) equal intercepts for the outcomes, $\mu_{YM} = \mu_{YF}$;
- (IT5) equal predictor variances, $\sigma^2_{XM} = \sigma^2_{XF}$;
- (IT6) equal predictor means, $\mu_{XM} = \mu_{XF}$.

We label IT1 as the (residual) Y-variance Indistinguishability Test (abbreviated as Y-var IT), IT2 and IT3 as Effect-IT and IT4 as Intercept-IT. We further refer to IT2-IT4 as the Y-mean IT, because the conditional (on X) means of Y are equal between roles when those three equalities hold. Combined with IT1, we label the first four equality constraints as Y-IT. This is because, under the normality assumption, the distribution of Y is completely characterized by its mean and variance. Similarly, assuming that the distribution of X is completely characterized by its mean and variance, we label IT5 as X-var IT, IT6 as X-mean IT, and IT5-IT6 as the X-IT. Finally, we refer to IT1-IT6 as complete indistinguishability (C-IT). A schematic overview of the different types of ITs is shown in Figure 2.2. Additionally, we will also label IT1 and IT5 as XY-var IT, and IT2-IT4 together with IT6 as XY-mean IT.

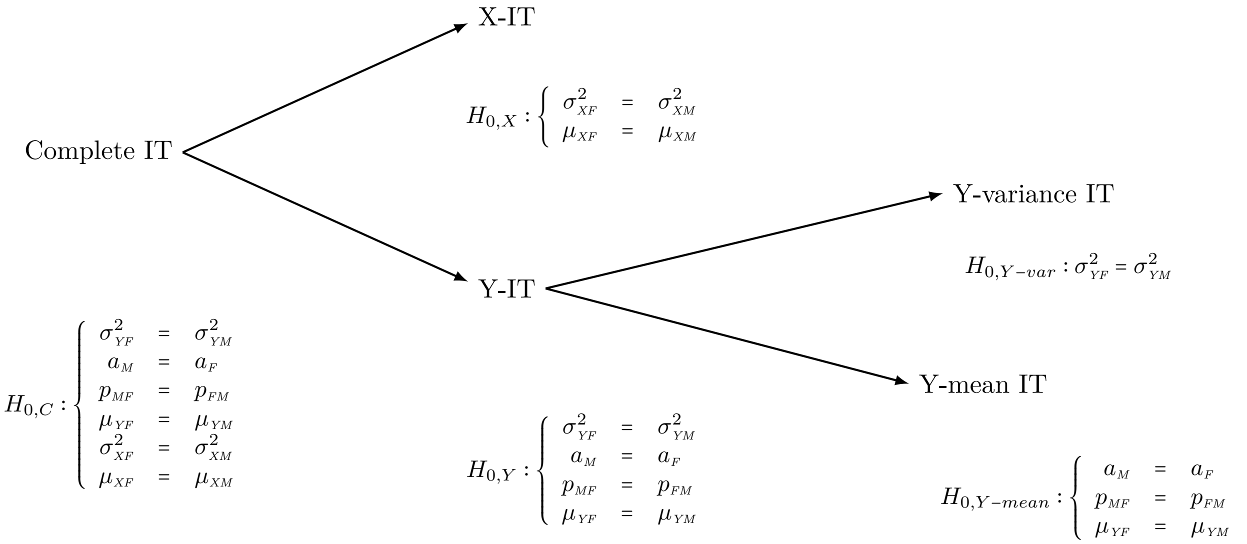


Figure 2.2 Schematic overview of the different types of indistinguishability in the APIM for cross-sectional dyadic data.

2.2.1 ML-estimation versus REML-estimation

Because exploring the (co)variance structure of residuals in the APIM (2.1) should be based on the most complete mean model, that is, different intercept, actor and partner effect assumed for men and women (Fitzmaurice, Laird, & Ware, 2011), we first focus on IT1. Indeed, a misspecified mean structure (e.g., erroneously assuming Y-mean indistinguishability) may lead to incorrect conclusions about the (co)variance structure of the model. Estimation of the parameters in model (2.1) is usually based on ML estimation. More specifically, it is based on maximizing

$$\mathrm{logLik}_{\mathrm{ML}} = -N\log(2\pi) - \frac{1}{2}\log(|V|) - \frac{1}{2}R^tV^{-1}R, \tag{2.4}$$

where N is the number of dyads, V the $(2N \times 2N)$ marginal covariance matrix of the outcome, and R the $(2N \times 1)$ residual vector. We refer to the Appendix for more specific expressions of V and R in this particular setting. It can easily be shown (see the section on Y-var distinguishability in the Appendix) that under model (2.1) the ML estimators of the residual (co)variances are

$$\hat{\sigma}^2_{\mathrm{ML},YM} = \frac{1}{N}\sum_{j=1}^{N} e^2_{Mj}, \qquad \hat{\sigma}_{\mathrm{ML},YMF} = \frac{1}{N}\sum_{j=1}^{N} e_{Mj}e_{Fj},$$
$$\hat{\sigma}^2_{\mathrm{ML},YF} = \frac{1}{N}\sum_{j=1}^{N} e^2_{Fj},$$

with e_{Mj} and e_{Fj} the estimated residuals in males and females, respectively, of the j^{th} dyad. However, ML estimation of the variance components is biased because it treats the fixed-effects estimates as known (Harville, 1977), and therefore REML is often recommended instead. The basic idea behind the REML estimation is to remove the fixed-effects parameters when estimating the variance components. A contrast matrix is chosen in such a way that the fixed effects parameters are not estimated. Using the log-likelihood function, one can obtain these unbiased variance components directly by maximizing

$$\begin{aligned}\mathrm{logLik}_{\mathrm{REML}} = -(N-p)\log(2\pi) - \frac{1}{2}\log(|V|) - \frac{1}{2}R^tV^{-1}R \\ - \frac{1}{2}\log\left(\left|X^tV^{-1}X\right|\right),\end{aligned} \tag{2.5}$$

where p is the number of mean model components including the intercept (i.e., $p = 3$ in model (2.1)), and X the $(2N \times 2p)$ design matrix of the model. The REML estimators of the (co)variances are

$$\hat{\sigma}^2_{\text{REML},YM} = \frac{1}{N-p}\sum_{j=1}^{N} e^2_{Mj}, \qquad \hat{\sigma}_{\text{REML},YMF} = \frac{1}{N-p}\sum_{j=1}^{N} e_{Mj}e_{Fj},$$
$$\hat{\sigma}^2_{\text{REML},YF} = \frac{1}{N-p}\sum_{j=1}^{N} e^2_{Fj},$$

(see the section on Y-var distinguishability in the Appendix). Hence, we have the following relationship between the ML and REML estimators of the (co)variances in model (2.1):

$$\begin{aligned} \hat{\sigma}^2_{\text{REML},YM} &= \frac{N}{N-p}\hat{\sigma}^2_{\text{ML},YM}, \\ \hat{\sigma}^2_{\text{REML},YF} &= \frac{N}{N-p}\hat{\sigma}^2_{\text{ML},YF}, \\ \hat{\sigma}_{\text{REML},YMF} &= \frac{N}{N-p}\hat{\sigma}_{\text{ML},YMF}. \end{aligned} \tag{2.6}$$

What implications does the choice between ML and REML estimation of the variance parameters have on the IT1? Typically, tests of hypotheses such as IT1 are performed based on a Likelihood Ratio Test (LRT), contrasting the full model with the reduced model (i.e., the model with the equality constraint $\sigma^2_{YM} = \sigma^2_{YF}$). The difference between the ‘deviances’ of these two specified models,

$$\chi^2 = -2\ \text{logLik}_{\text{red}} + 2\ \text{logLik}_{\text{full}},$$

is Chi-square distributed with degrees of freedom (*df*) equal to the difference in number of estimated model parameters (i.e., *df* = 1 for IT1).

To the best of our knowledge, the relationship between the χ^2 statistic based on ML versus REML for testing indistinguishability in dyadic data has never been investigated in detail. In the Appendix we show that for model (2.1) the following simple relationship exists between the test statistics when testing IT1 using an LRT:

$$\chi^2_{\text{REML}} = \frac{N-p}{N}\chi^2_{\text{ML}}. \tag{2.7}$$

Hence, the LRT statistic for IT1 based on REML will systematically be smaller than the LRT statistic for IT1 based on ML. In small samples this clearly will have an impact. A literature review of studies using the APIM (Loeys & Molenberghs, 2013) revealed that sample sizes typically ranged from 30 to 300 dyads (first quartile=60, median=100, and third quartile=150). Thus testing IT1 may become an issue in a substantial amount of studies. As we will show in a simulation study later, the LRT based on ML for IT1 can be viewed as too liberal when the number of dyads is smaller than 60.

Secondly, we focus on the Y-mean IT (i.e., IT2-IT4). In this case, REML can not be applied. The REML-based deviances no longer contain any information about the mean components of the model. As a result, one should only use REML estimation to compare two models in case they have identical mean structures. This implies that one has to rely on ML-based LRT when testing IT2-IT4.

Thirdly, we consider IT5 and IT6 (i.e., X-IT). Because IT5 only contains variance parameters, one can similarly argue that one should consider the use of REML-based deviances in the LRT. It will reduce the bias on the (co)variance parameters estimates, especially when the sample size is rather small, as well as confine the liberal nature of the ML-based LRT. The IT6 considers equal predictor means and thus can also only be tested using ML-based LRTs.

2.2.2 SEM-framework versus MLM-framework

Popular SEM software such as Mplus (Muthén & Muthén, 2012), LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 2004), and the R package `lavaan` (Rosseel, 2012) provides several estimation methods, such as maximum likelihood, generalized least squares, and weighted least squares. In SEM, ML estimation is the most popular estimation method. To our knowledge, OpenMx (Boker et al., 2011) is the only SEM package that allows REML estimation for the variance parameters, and it requires the contrast matrix to be manually implemented (Cheung, 2013). MLM software such as HLM (Raudenbush, 2004), MLwiN (Goldstein, 2011), PROC MIXED in SAS (SAS Institute, 2008) and the R package `nlme` (Pinheiro & Bates, 2000), on the other hand, often set REML estimation as the default.

Curran (2003) investigated the similarity between the SEM and MLM frameworks and showed that parameter estimates obtained using ML will typically be (nearly) identical between both frameworks. This implies that the corresponding ITs using ML-based LRT will yield the same results in the MLM and SEM framework. Using the MLM framework can be advantageous when performing the IT1, because the researcher can more easily use REML estimation.

When it comes to testing X-IT in model (2.1), it is important to acknowledge that the variable X is considered as fixed in the MLM framework, while it may be treated as random in the SEM framework. The latter framework hence more naturally allows for testing IT5 and IT6. Alternatively, one may propose in the MLM framework an additional set of modelling equations with X as outcome, that is

$$\begin{cases} X_{Fj} = \mu_{XF} + \varepsilon_{XFj} \\ X_{Mj} = \mu_{XM} + \varepsilon_{XMj} \end{cases} \tag{2.8}$$

with

$$\begin{pmatrix} \varepsilon_{XFj} \\ \varepsilon_{XMj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{XF} & \sigma_{XF,XM} \\ \sigma_{XF,XM} & \sigma^2_{XM} \end{pmatrix} \right).$$

and perform an LRT contrasting the full model with the reduced model for testing IT5 and IT6.

2.2.3 Global testing versus sequential testing

A researcher may not be interested in one specific indistinguishability test, but rather in a simultaneous set of ITs such as Y indistinguishability or complete indistinguishability (see Figure 2.2).

Consider first the Y-IT (i.e., testing IT1-IT4). This type of indistinguishability includes a variance component as well as mean components of model (2.1). If we wished to perform this test in one go, we could use an LRT on 4 *df* comparing the full model to the model with the four equality constraints. We will refer to this test as the global Y-IT. Note that due to the presence of mean components in the null hypothesis of the test, we must use an ML-based LRT. Alternatively, one might consider a sequential test: first test IT1 with an LRT relying on REML, second test IT2-IT4 with an LRT relying on ML, that is, one could use the following procedure:

1. Y-var IT (IT1) using REML-based LRT
   → p–value p_1,

2. Y-mean IT (IT2-IT4) using ML-based LRT
   with covariance structure under $\begin{cases} H_{A,\,Y\text{-}var} & \text{if } p_1 < \alpha \\ H_{0,\,Y\text{-}var} & \text{otherwise} \end{cases}$
   $\rightarrow p-$value p_2,

and reject Y-IT if $p_1 < \frac{\alpha}{2}$ or $p_2 < \frac{\alpha}{2}$. Note that a Bonferroni correction is suggested in the above two-step approach to preserve the overall Type I error. As we will show later in the simulation study, ignoring such correction leads to seriously inflated Type I errors.

Next, we focus on the C-IT (i.e., testing IT1-IT6). Within the SEM framework this can be done simultaneously using an ML-based LRT with 6 *df* comparing the full model to the model with the six equality constraints. Again, one may alternatively separate the variance and mean components, and perform a sequential test:

1. XY-var IT (IT1 and IT5) using REML-based LRT
   $\rightarrow p-$value p_1,

2. XY-mean IT (IT2-IT4, and IT6) using ML-based LRT
   with covariance structure under $\begin{cases} H_{A,XY\text{-}var} & \text{if } p_1 < \alpha \\ H_{0,XY\text{-}var} & \text{otherwise} \end{cases}$
   $\rightarrow p-$value p_2,

in which complete indistinguishability is rejected if $p_1 < \frac{\alpha}{2}$ or $p_2 < \frac{\alpha}{2}$. The first step could easily be performed in the SEM framework if REML estimation were available. Within the MLM framework, this step can be achieved by considering models (2.1) and (2.8). Upon noting that the (residual) likelihood of the joint distribution of X and Y is equal to the product of the (residual) likelihood of the conditional distribution of Y given X and the distribution of X:

$$\text{logLik}_{XY} = \text{logLik}_X + \text{logLik}_{Y|X},$$

one can simply add up the $\chi^2(1)$ statistics for the X-var and Y-var IT (i.e., IT1 and IT5, respectively) and rely on the $\chi^2(2)$ distribution to perform the XY-var IT. Similarly, for the second step, the ML-based LRT for the XY-mean IT has a $\chi^2(4)$ distribution, obtained as the sum of a $\chi^2(1)$ and $\chi^2(3)$ test statistic in the MLM framework. Note that this second step would yield identical results in the SEM framework, but the latter does not require comparing the two models for X and Y separately.

2.2.4 Simulation study in cross-sectional setting

Using simulations, we now explore the impact of the three aforementioned choices (REML vs. ML, SEM vs. MLM and global vs. sequential testing). In the simulation study, we consider the following data-generating process (DGP):

$$\begin{cases} Y_{Fj} = 1.17 + 1.70\ X_{Fj} + 0.50\ X_{Mj} + \varepsilon_{Fj} \\ Y_{Mj} = 1.17 + 1.70\ X_{Mj} + 0.50\ X_{Fj} + \varepsilon_{Mj} \end{cases}$$

with

$$\begin{pmatrix} \varepsilon_{Mj} \\ \varepsilon_{Fj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.9 & 0.8 \\ 0.8 & 1.9 \end{pmatrix}\right)$$

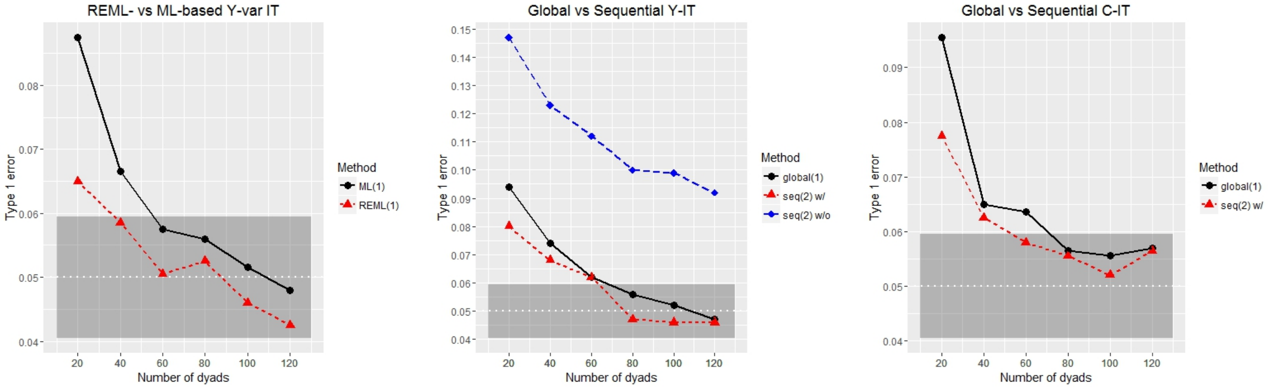
and

$$\begin{pmatrix} X_{Mj} \\ X_{Fj} \end{pmatrix} \sim N\left(\begin{pmatrix} 2.7 \\ 2.7 \end{pmatrix}, \begin{pmatrix} 2.52 & 0.3 \\ 0.3 & 2.52 \end{pmatrix}\right).$$

Under this DGP, complete indistinguishability holds. We vary the number of dyads from 20 to 120 in steps of 20, and consider 2,000 repetitions for each setting. All ITs have been performed in R using the `lavaan` package for SEM, and the `nlme` package for MLM.

We first focus on the Y-var IT (i.e., IT1) and contrast the type I error of the LRT based on ML with that based on REML. The upper panel of Figure 2.3 shows the empirical Type I error, that is, the percentage of times (out of 2,000 repetitions) that the null hypothesis is rejected, for both tests (performed at the 5% significance level) in each setting. With 2,000 repetitions the empirical Type I error should approximately lie between 4% and 6%. When the number of dyads is small, for example 20 or 40, the LRT based on ML is clearly too liberal. Note that the ratio of the LRT statistic based on REML versus ML exactly corresponds to equation (2.7).

Second, we focus on the Y-indistinguishability test (i.e., IT1-IT4) and contrast the Type I error of the global test based on ML using SEM, with the earlier proposed sequential test (i.e., first Y-var IT based on REML and then Y-mean IT based on ML, using MLM), with and without Bonferroni correction. The middle panel of Figure 2.3 clearly shows an inflated Type I error for the two-step sequential test in the absence of a Bonferroni correction (denoted *seq(2) w/o*) for all sample sizes. In small samples, the sequential test with Bonferroni correction (denoted *seq(2) w/*) is somewhat less liberal than the global Y-IT (denoted *global(1)*). Both test procedures approach the true nominal rate for increasing sample sizes.



(a) **(b)** **(c)**

Figure 2.3 The (averaged) empirical Type I error over 2,000 simulations, comparing the (sequential) REML-based LRTs with the (global) ML-based LRTs for (a) the Y-var IT, (b) the Y-IT, and (c) the C-IT. Numbers in parentheses represent the number of steps used to perform the IT.

Finally, we compare the global test versus the earlier proposed two-step sequential test for the C-IT (i.e., IT1-IT6) with Bonferroni correction. The lower panel of Figure 2.3 reveals a seriously inflated Type I error for the global test (denoted *global(1)*) in small samples, while again the sequential test (denoted *seq(2) w/*) is somewhat less liberal.

2.3 The over-time APIM and tests for indistinguishability

In the last decade the number of studies with longitudinal dyadic data has increased exponentially. This poses interesting statistical modelling challenges because now not only the correlation of measurements within dyads needs to be accounted for, but also the correlation of measurements within individuals over time. Kenny et al. (2006) describe three frequently used classes for modelling of longitudinal couple data: cross-lagged regressions, growth curves analyses, and over-time APIMs. Cross-lagged regressions allow stability to be examined by assessing the effect that the past behaviour of actor and partner have on the current behaviour (Cook & Kenny, 2005). Growth curve models, on the other hand, typically assume a linear relationship between the dyad members' behaviour and time. They allow exploration of, for example, whether dyad members have the same rate of growth in their behaviours (Newsom, 2002; Planalp, Du, Braungart-Rieker, & Wang, 2017). The over-time APIM extends the advantages of the APIM to the longitudinal setting (Bolger & Laurenceau, 2013; Laurenceau & Bolger, 2011). In this paper we will focus on over-time APIMs.

Figure 2.4 shows a graphical representation of the over-time APIM for three time points. Similar to the cross-sectional APIM, the outcome of a dyad member may depend on one's own and one's partner's predictor at a particular time point. We can allow those effects to be subject-specific, that is,

$$\begin{cases} Y_{Fij} = \mu_{YFj} + a_{Fj}X_{Fij} + p_{MFj}X_{Mij} + \varepsilon_{Fij} \\ Y_{Mij} = \mu_{YMj} + a_{Mj}X_{Mij} + p_{FMj}X_{Fij} + \varepsilon_{Mij} \end{cases}, \qquad (2.9)$$

with i referring to the time point ($i = 1, \dots, T$) and j to the dyad number ($j = 1, \dots, N$).

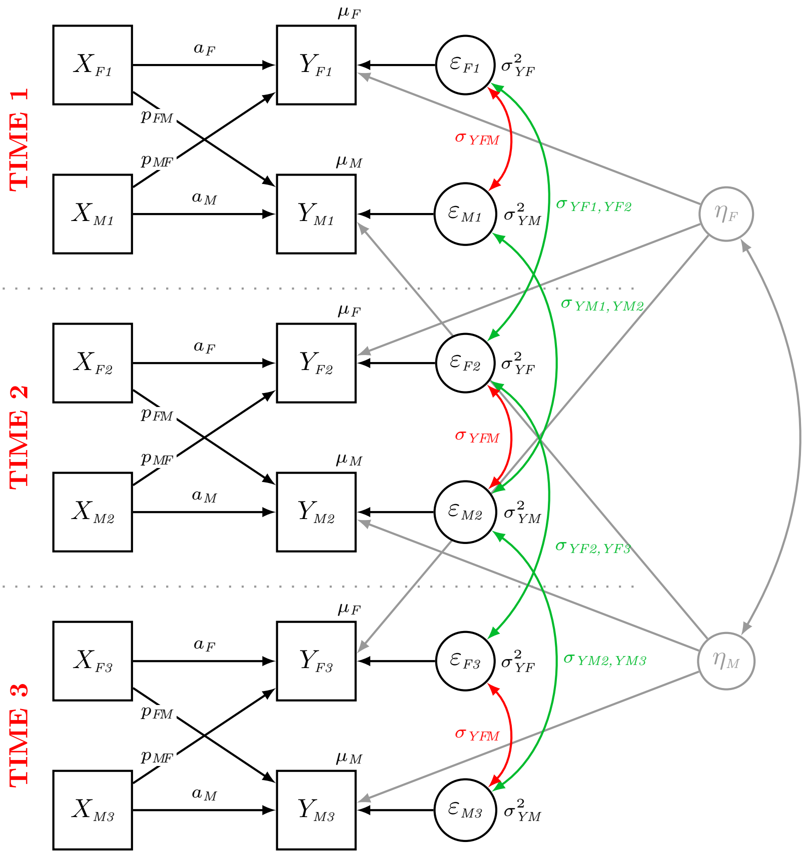


Figure 2.4 The over-time actor partner interdependence model measured over three time points. The index 'M' refers to the first dyad member (e.g., male), while 'F' refers to the second member (e.g., female).

In model (2.9), we have 6 parameters that both have a fixed and random part:

$$\begin{aligned}
\mu_{YFj} &= \mu_{YF} + \eta_{Fj} \\
\mu_{YMj} &= \mu_{YM} + \eta_{Mj} \\
a_{Fj} &= a_F + \eta_{aFj} \\
a_{Mj} &= a_M + \eta_{aMj} \\
p_{MFj} &= p_{MF} + \eta_{pFj} \\
p_{FMj} &= p_{FM} + \eta_{pMj}.
\end{aligned}$$

Sometimes the effect of time is included as a predictor in model (2.9) as well, but in practical applications this is often not relevant when outcomes randomly fluctuate over time. To keep things simple in Figure 2.4, the slopes were assumed to be the same between subjects. This is often done in practice when the amount of time points is relatively small. Indeed, in such setting one may often encounter convergence issues if many random effects are assumed. Interestingly, Baird and Maxwell (2016) recently showed that treating effects as fixed rather than random will not introduce bias in the fixed-effects estimator. For the remainder of this article, random effects will therefore be limited to random intercepts, that is, we assume $a_{Fj} = a_F, a_{Mj} = a_M, p_{MFj} = p_{MF}$ and $p_{FMj} = p_{FM}$.

It is important to discuss the covariance structure of the random effects and residual error terms in Figure 2.4. The random intercepts η_{Fj} and η_{Mj} are allowed to be correlated and are typically assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \eta_{Fj} \\ \eta_{Mj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_F^2 & \tau_{FM} \\ \tau_{FM} & \tau_M^2 \end{pmatrix} \right). \tag{2.10}$$

Not only the variances along the main diagonal (e.g., how much dyads differ from one another in terms of female intercepts), but also the covariances (e.g., whether male intercepts covary with female intercepts) are of interest. For a more detailed discussion, including the random slopes, see Bolger and Laurenceau (2013). The lower level residuals ε_{Fij} and ε_{Mij} are time-specific individual female and male error components with separate variances for each gender (i.e., σ_{YF}^2 and σ_{YM}^2 in Figure 2.4). These residuals are allowed to correlate at each time point to capture, for example, the daily residual covariation (i.e., σ_{FM} in Figure 2.4) in female and male outcomes. Furthermore, one may allow correlated errors across time points within individuals (e.g., $\sigma_{F1,F2}$ and $\sigma_{M1,M2}$, etc. in Figure 2.4). It should be noted, however, that additionally modelling autocorrelated errors for such multilevel model is impractical at this point in time for most software packages (Laurenceau & Bolger, 2011), except for SAS. The latter allows (specific) Kronecker products of different covariance structures to respectively represent the residual covariance between the dyad members at each time point, and the time correlation within a dyad member, for example, a 'UN@AR(1)' structure (Bolger & Laurenceau, 2013). In R packages such as `nlme`, the covariance structure is also limited and cannot properly allow for either daily residual covariance or over-time covariance. Since we will use `nlme` in simulation studies, we will therefore assume that the covariance matrix of the residuals can be restricted as follows:

$$\mathrm{Cov}\begin{pmatrix} \varepsilon_{F1j} \\ \varepsilon_{F2j} \\ \vdots \\ \varepsilon_{FTj} \\ \varepsilon_{M1j} \\ \varepsilon_{M2j} \\ \vdots \\ \varepsilon_{MTj} \end{pmatrix} = \begin{pmatrix} \sigma^2_{YF} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2_{YF} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{YF} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2_{YM} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \sigma^2_{YM} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \sigma^2_{YM} \end{pmatrix} \tag{2.11}$$

Note, in (2.11) and in Figure 2.4, that the variability in the outcomes does not change over time. Similarly, the actor and partner effects are constant over time, but the assumptions on time-constant means and variances can in principle both be relaxed.

When dealing with longitudinal dyadic data, it is important to revise the earlier definitions of the indistinguishability tests for cross-sectional dyadic data. Assuming multilevel model (2.9) with random effect covariance (2.10) and residual error covariance (2.11), we first note that the variability in the outcome is separated in between-subject variability and within-subject variability. When restricted to random effect models with random intercepts only, this implies that the IT1 can be divided into the following two sub-tests, referred to as I-var IT and Y-var IT, respectively:

(IT1A) equal random intercept variances: $\tau^2_M = \tau^2_F$;

(IT1B) equal residual variances for the outcomes: $\sigma^2_{YM} = \sigma^2_{YF}$.

For more complex random-effect models, additional equality constraints (i.e., equal random slope variances and equal covariances) would be required as well. Given that we assumed time-constant effects in model (2.9), the Y-mean (sub-)IT(s) remain(s) unchanged. When discussing X-IT, it is important to note that one may have within-subject predictors (varying over time) and between-subject predictors (constant over time). For the latter the X-IT is the same as before, but for the former, one may consider, for example, the within-subject variance and between-subject variance of the predictor and check whether these are equal for males and females.

2.3.1 Issues in testing indistinguishability in longitudinal dyadic data

Considering the three aforementioned issues related to indistinguishability tests in the cross-sectional setting (REML vs. ML, SEM vs. MLM, and global vs. sequential testing), we will focus on the first issue. More specifically, we will solely discuss the impact of the choice of REML versus ML for the IT1A and IT1B. Other issues are essentially the same in the longitudinal setting as in the cross-sectional setting.

Unfortunately, no analytic relationships between the ML and REML estimators of the variance components (i.e., both the random effect variance and the residual error variance) exist in more complex models such as model (2.9). Consider the model

$$\begin{cases} Y_{Fij} = (\mu_{YF} + \eta_{YFj}) + a_F X_{Fij} + p_{MF} X_{Mij} + \beta_{F3} Q_j + \varepsilon_{Fij} \\ Y_{Mij} = (\mu_{YM} + \eta_{YMj}) + a_M X_{Mij} + p_{FM} X_{Fij} + \beta_{M3} Q_j + \varepsilon_{Mij} \end{cases}, \tag{2.12}$$

where i refers to the time point ($i = 1, \ldots, T$), j refers to the dyad number ($j = 1, \ldots, N$), and we have the within-dyad predictor $\mathbf{X} \sim N(\boldsymbol{\mu}_{X,2T\times1}, \boldsymbol{\Sigma}_{X,2T\times2T})$, and the between-dyad predictor $Q \sim N(\mu_{Q,1\times1}, \sigma^2_{Q,1\times1})$. Upon considering the number of parameters at the lower and upper level, one may intuitively expect the following relationships:

$$\sigma^2_{\mathrm{REML},YF} = \frac{NT}{NT-p}\sigma^2_{\mathrm{ML},YF} \tag{2.13}$$

$$\sigma^2_{\mathrm{REML},YM} = \frac{NT}{NT-p}\sigma^2_{\mathrm{ML},YM} \tag{2.14}$$

and

$$\tau^2_{\text{REML},F} = \frac{N}{N-(1+q)}\tau^2_{\text{ML},F} \tag{2.15}$$

$$\tau^2_{\text{REML},M} = \frac{N}{N-(1+q)}\tau^2_{\text{ML},M}, \tag{2.16}$$

where N is the number of dyads, T the number of time points, p the number of parameters in the model (including the intercept), and q the number of between-dyad predictors. Simulation studies (see next section) provide empirical evidence that these expressions do not hold exactly, but yield good approximations on average. As a consequence, one may expect a similar impact on the Type I error of using REML-based versus ML-based LRT for testing equality in variance between males and females.

It should be noted though that for the residual variance component, the ratio of the REML-based and ML-based variance estimator will quickly converge to one as both T and N become larger. For the random intercept variance, on the other hand, the ratio only depends on N. For smaller samples, we thus hypothesize an inflated Type I error for the ML-based LRT for equality in random intercept variances, regardless of the number of time points. This hypothesis is verified by simulation study in the next subsection.

2.3.2 Simulation study in longitudinal setting

As a DGP in the simulations we consider modelling equations (2.12). The residual covariance matrix equals (2.11) with $\sigma^2_{YF} = \sigma^2_{YM}$, while the random intercept covariance satisfies structure (2.10) with $\tau^2_M = \tau^2_F$. Hence, the data is generated under the assumption of the null hypothesis of both the IT1A and IT1B (i.e., IT1 holds). The values of the covariance matrices were randomly generated via the package `clusterGeneration` in R, while the mean values and the fixed effects were randomly generated from a standard uniform distribution. In the simulation study, we varied the number of dyads from 20 to 120 in steps of 20, and the number of time points from 4 to 20 in steps of 2. Considering less than four time points may lead to convergence issues when fitting such complex models (Hesser, 2015). We considered three different estimation models:

(M1) Model (2.12) with random effect covariance matrix (2.10) and residual covariance matrix (2.11), that is, IT1A and IT1B are assumed;

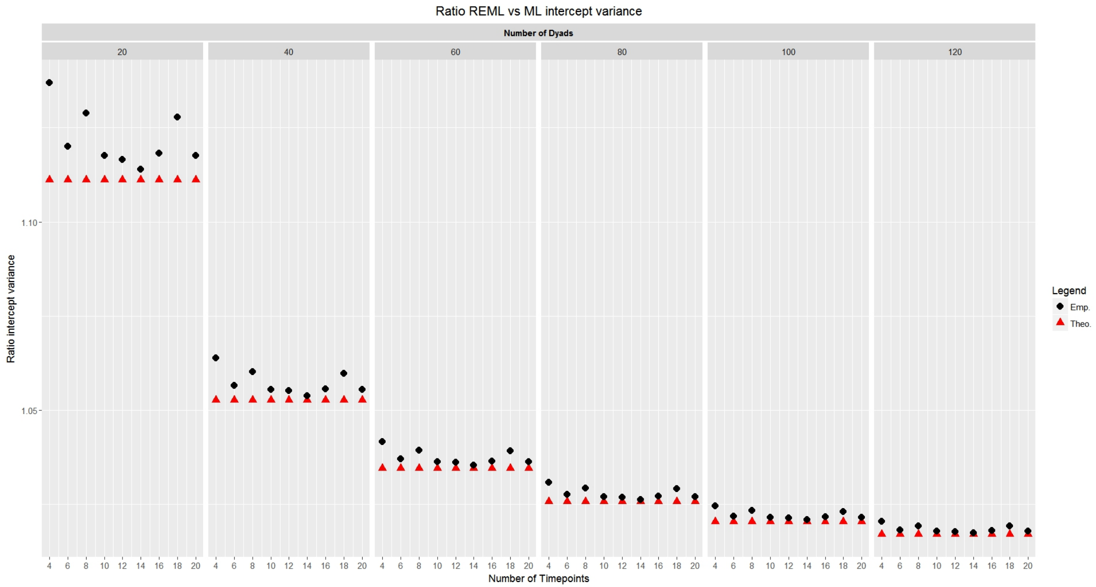
(M2) Model (2.12) with random effect covariance matrix (2.10) constrained to $\tau^2_{YF} = \tau^2_{YM}$ and residual covariance matrix (2.11), that is, the IT1A is assumed;

(M3) Model (2.12) with random effect covariance matrix (2.10) and residual covariance matrix (2.11) constrained to $\sigma^2_{YF} = \sigma^2_{YM}$, that is, the IT1B is assumed.

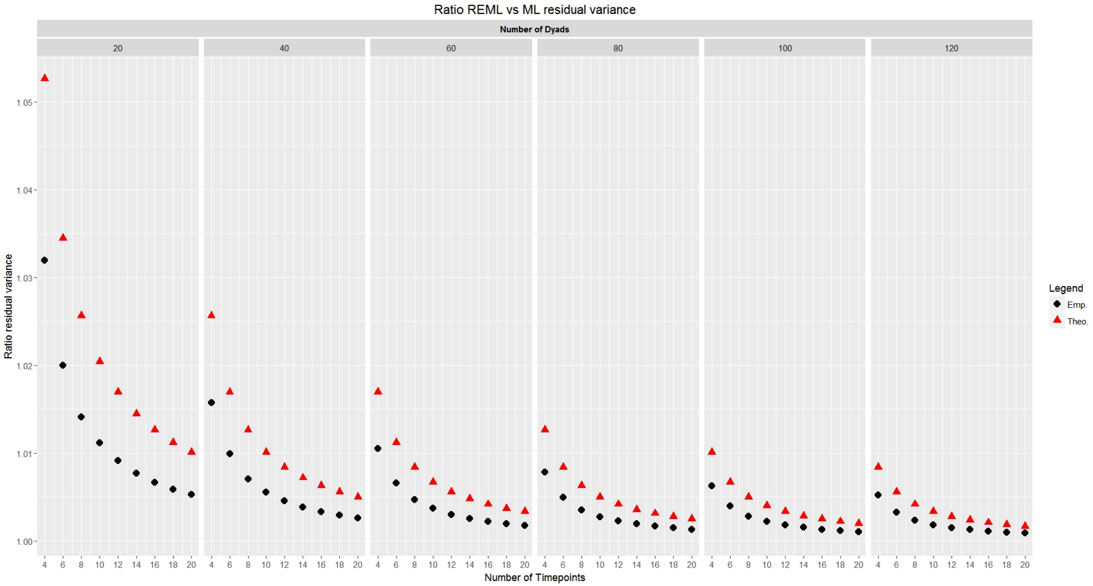
All three models were fitted in the R package `nlme` relying on ML and REML. Likelihood ratio tests comparing M2 with M1, and M3 with M1 were performed to test equality in random intercept variances and in residual variances between males and females, respectively.

From Figure 2.5a, one can see that expression (2.16) yields a slight underestimation of the ratio of the REML-based and ML-based random intercept variance. This confirms our expectation that the ML-based estimator is smaller than the REML-based estimator of the random intercept variance, especially in smaller samples (regardless of the number of time points). Consequently, the LRT for I-var IT will be larger when based on ML instead of REML. From Figure 2.6a, we see that the ML-based IT for random intercept variance is more liberal than the REML-based IT in samples with a small number of dyads, independent from the number of time points in the sample.

Next, the ratio of the REML-based and ML-based residual variance tends to be slightly overestimated by expression (2.14) (see Figure 2.5b). However, given that the ratio depends on both the number of dyads and the number of time points, the ratio quickly converges to one. A similar conclusion can be made for the LRT statistics based on REML and ML. From Figure 2.6b, one can see that there is hardly any inflation of the Type I error for the IT for the residual variances based on ML compared to REML, even for a small number of dyads and a small number of time points.

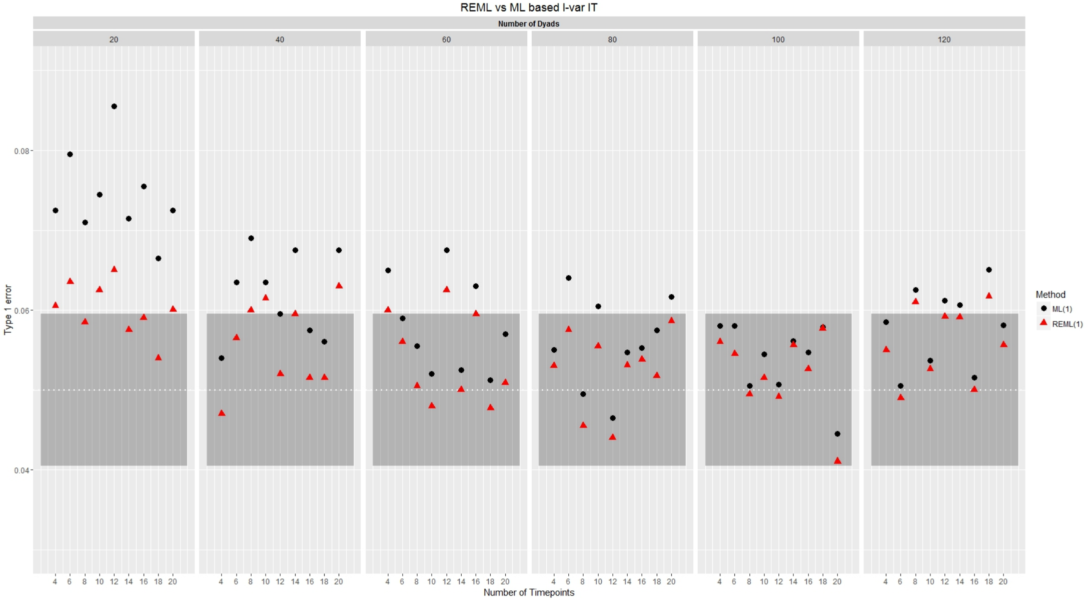


(a) The empirical ratio (Emp.) of the REML-based and ML-based variance estimator of the random intercept variance based on 2,000 simulations, compared to the theoretical ratio (Theo.) according to equation (2.16).

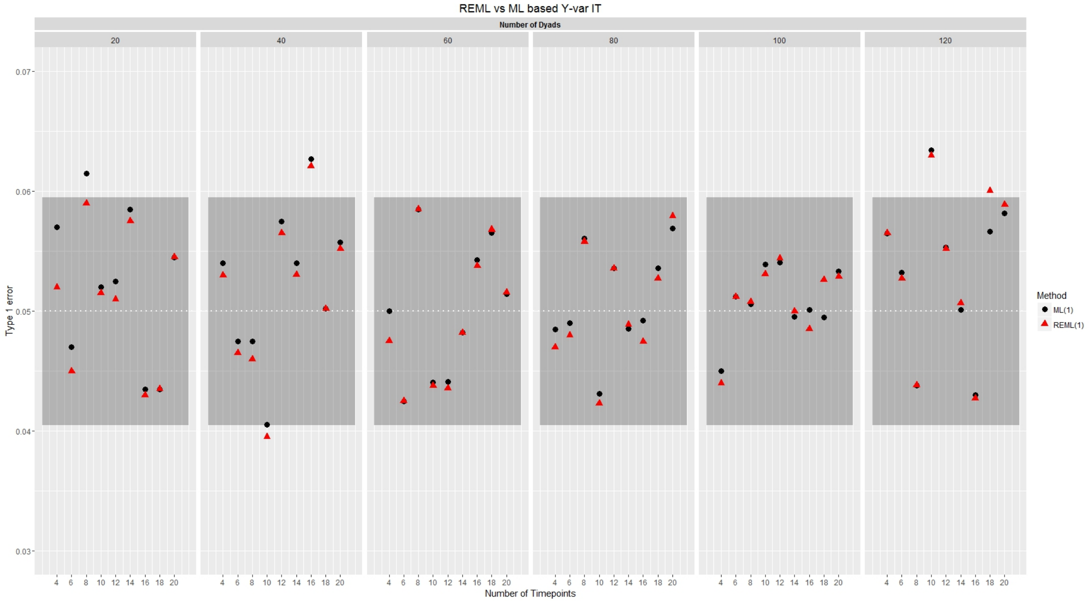


(b) The empirical ratio (Emp.) of the REML-based and ML-based variance estimator of the residual variance based on 2,000 simulations, compared to the theoretical ratio (Theo.) according to equation (2.14).

Figure 2.5 The empirical ratio (Emp.) of the REML-based and ML-based variance estimator compared to the theoretical ratio (Theo.).



(a) The (averaged) empirical Type I error over 2,000 simulations, comparing the (sequential) REML-based ITs with the (global) ML-based ITs for the I-var IT.



(b) The (averaged) empirical Type I error over 2,000 simulations, comparing the (sequential) REML-based ITs with the (global) ML-based ITs for the Y-var IT.

Figure 2.6 The (averaged) empirical Type I error over 2,000 simulations, comparing the (sequential) REML-based ITs with the (global) ML-based ITs. The grey band represents the confidence interval for the nominal significance level of 5%.

2.4 Illustrative examples

We now illustrate the practical implementation of the different ITs (i.e., IT1-IT6 and their combinations) on real cross-sectional and longitudinal dyadic data. Both the SEM and MLM approach are used, and global as well as sequential tests are performed. In the Appendix 2.A, we provide R code for those ITs. This should enable the reader to perform all tests described in this paper. Note that for the MLM framework we rely on the *lme* function from the R package `nlme`, while for the SEM- ramework analyses are performed with the *sem* function from the R package `lavaan`.

2.4.1 Cross-sectional example

Decuyper et al. (2016) recently explored the association between personality pathology and relationship satisfaction in 52 heterosexual couples. More specifically, they investigated the effect of five different ratings of personality pathology on three types of relationship satisfaction scales. Here we focus on the effect in males and females of one's own and one's partner's antagonism on relationship satisfaction, as measured by the revised dyadic adjustment scale. The authors raised the question whether there was any distinguishability present in the data. In other words, can the roles of males and females be considered exchangeable in this setting?

To address this question, we started from the most complete model (2.1) with an unstructured residual covariance. Table 2.1 shows the p-values of the different ITs (and combinations thereof) under both frameworks. Any difference between both frameworks for IT1 and IT6 can be attributed to the use of REML instead of ML for testing variance components. If ML estimation were used in both frameworks, results would be identical. When considering C-IT, the global ML-based test in the SEM framework points towards distinguishability. This is in contrast to the sequential test (based on REML for the variance components) which does not yield evidence for distinguishability. This difference may be explained by the liberal nature of the C-IT in the SEM framework.

Table 2.1 The *p*-values of the different ITs (i.e., IT1-IT6, and their combinations) for the data of Decuyper et al. (2016).

IT	SEM	MLM
IT1	.049	.056
IT2[a]	.525	.525
IT3	.909	.909
IT4	.142	.142
IT5[a]	.454	.459
IT6	.013	.013
Y-mean IT	.399[b]	.399[b]
Y-IT[a]	.150[b]	.056/.413[c]
C-IT[a]	.035[b]	.122/.059[d]

Notes. [a] *p*-values of the REML-based LRT for testing the equality of variances in the MLM framework. All other *p*-values are from ML-based LRTs.
[b] *p*-values from the global test.
[c] *p*-values from the sequential test for variance and mean of Y, respectively.
[d] *p*-values from the sequential test for variance and mean of X and Y, respectively.

Table 2.2 The *p*-values of the different ITs (i.e., IT1-IT6, and their combinations) for the data of Dewitte et al. (2015).

IT	SEM	MLM
IT1	.063	.065
IT2[a]	.765	.765
IT3	.393	.393
IT4	.017	.017
IT5[a]	.785	.788
IT6	.670	.670
Y-mean IT	.089[b]	.089[b]
Y-IT[a]	.034[b]	.065/.087[c]
C-IT[a]	.126[b]	.203/.167[d]

Notes. [a] *p*-values of the REML-based LRT for testing the equality of variances in the MLM framework. All other *p*-values are from ML-based LRTs.
[b] *p*-values from the global test.
[c] *p*-values from the sequential test for variance and mean of Y, respectively.
[d] *p*-values from the sequential test for variance and mean of X and Y, respectively.

2.4.2 Longitudinal diary study in heterosexual couples

Dewitte et al. (2015) performed a daily diary study on sexual behaviour in 66 heterosexual couples in Flanders. Every morning during 3 weeks, participants were asked about their sexual and intimate behaviour since the last time they had filled out their morning diary (i.e., sexual behaviour over the past 24 hr). Every evening, the participants were asked to report on their individual, relational, and partner-related feelings and behaviour, as experienced during that day. Here, we focus on the extent to which they report intimate acts with their partner (described as the amount of kissing, cuddling and caressing on a 7-point scale from ‘not at all’ to ‘very frequent’) and on the daily evening reports of positive relational feelings. The latter were obtained by averaging the scores (on a 7-point scale) on nine items (the extent to which they felt happy, satisfied, understood, supported, accepted, loved, in love, connected, and close). The research question discussed here considers the contribution of one’s own and one’s partner’s reported amount of intimacy to next-day positive relationship feelings, and whether men and women are indistinguishable from that perspective.

Following Bolger and Laurenceau (2013), Dewitte et al. (2015) used an over-time APIM like model (2.12), but separated within effects from between effects of intimacy:

$$\begin{cases} Y_{Fij} = (\beta_{0YF} + \eta_{YFj}) + \beta_{1WF}(X_{Fij} - \bar{X}_{F.j}) + \beta_{1BF}\bar{X}_{F.j} \\ \qquad\qquad\qquad + \beta_{2WF}(X_{Mij} - \bar{X}_{M.j}) + \beta_{2BF}\bar{X}_{M.j} + \varepsilon_{Fij} \\ Y_{Mij} = (\beta_{0YM} + \eta_{YMj}) + \beta_{1WM}(X_{Mij} - \bar{X}_{M.j}) + \beta_{1BM}\bar{X}_{M.j} \\ \qquad\qquad\qquad + \beta_{2WM}(X_{Fij} - \bar{X}_{F.j}) + \beta_{2BM}\bar{X}_{F.j} + \varepsilon_{Mij} \end{cases} \tag{2.17}$$

with random effect variance (2.10) and residual variance (2.11).
In model (2.17), $\bar{X}_{F.j}$ and $\bar{X}_{M.j}$ are the average score of intimacy over 3 weeks for the female and male partner, respectively. The parameters β_{1WF} and β_{1WM} represent the within-subject actor effects in males and females, β_{2WF} and β_{2WM} the within-subject partner effects, while β_{1BF}, β_{1BM}, β_{2BF} and β_{2BM} represent the between-subject actor and partner effects. If one does not separate within and between effects as in model (2.17), the estimated parameters in model (2.9) are a mixture of those effects (Enders & Tofighi, 2007). Based on model (2.17), we performed all ITs (including the usual combinations) both in the MLM and SEM framework. More specifically, we considered the following tests:

- (IT1) equal residual variances and random intercept variances for the outcomes, $\sigma^2_{YM} = \sigma^2_{YF}$ and $\tau^2_M = \tau^2_F$;
- (IT2) equal actor effects (both within and between): $\beta_{1WM} = \beta_{1WF}$ and $\beta_{1BM} = \beta_{1BF}$;
- (IT3) equal partner effects (both within and between): $\beta_{2WM} = \beta_{2WF}$ and $\beta_{2BM} = \beta_{2BF}$;
- (IT4) equal intercepts for the outcomes: $\mu_{YM} = \mu_{YF}$;
- (IT5) equal predictor variances: $\sigma^2_{XM} = \sigma^2_{XF}$;
- (IT6) equal predictor means: $\mu_{YM} = \mu_{YF}$;

The R code to perform these tests in the SEM and MLM framework with `lavaan` and `nlme`, respectively, can be found in Appendix 2.A. The corresponding *p*-values for each of these tests can be found in Table 2.2. Any difference between SEM and MLM in the individual ITs can be attributed to the use of REML instead of ML. There is some indication of different partner effects between males and females, but this finding disappears after applying a Bonferonni correction. The conclusions on the Y-IT differ between SEM and MLM, with again the former being more liberal, because of the use of ML.

2.5 Discussion

In this paper we discussed indistinguishability tests in the cross-sectional and longitudinal setting. We illustrated how both the SEM and MLM framework can be used to perform ITs and yield the same results when ML estimation is used. In order to preserve the Type I error in small samples, REML estimation should be preferred over ML estimation to test Y-var and X-var indistinguishability. In contrast to its wide availability in MLM packages, REML is not implemented in most SEM software. It should be noted that global indistinguishability tests are incorporated in some recent APIM tools that were developed for dyadic researchers (e.g., Kenny (2017)): *APIM_MM*, *APIM_SEM*, *Dingy*, etc. None of these, however, discuss the possibility of incorporating REML estimation, nor the possibility of performing complete indistinguishability in case of MLM framework, nor do they allow for sequential tests. Finally, we also stressed the risk of inflated Type I errors when examining each aspect of indistinguishability separately without any multiplicity correction. When testing complete indistinguishability is the primary research question of interest, this may be a serious concern. In this case, we would recommend the researcher to use the sequential ITs, especially when the amount of dyads is rather small, and starting with the Y(I)-var IT, next the Y-mean IT and finally the C-IT. But one may also take a more pragmatic approach and view the testing of indistinguishability as an intermediate step towards more parsimonious models rather than a true empirical question. In that view, an inflated Type I error may be more acceptable and a global indistinguishability test, as implemented in more recent APIM tools mentioned above, might be used.

Not only the Type I error deserves attention here. It is important to acknowledge that the lack of evidence for distinguishable dyad members does not necessarily imply indistinguishability. That is, failure to reject the null hypothesis does not prove the null hypothesis. Indeed, the power to detect distinguishability may be small in dyadic studies. As an example, we reconsidered our illustrative cross-sectional study and supposed all parameters (i.e., intercepts, actor and partner effects and variances) 10% larger in men than in women. A similar study of 52 couples has only about 18% power to reject complete indistinguishability (based on a global test in the SEM-framework) at the 5% significance level.

We focused in this paper on APIMs with outcomes measured at the interval level, but APIMs with categorical outcomes could be considered as well (Loeys & Molenberghs, 2013). For the latter type of outcomes, estimation within the generalized estimating equations (GEE) framework has been suggested (Loeys, Cook, De Smet, Wietzker, & Buysse, 2014). When using GEE and logistic regression for Bernoulli outcomes or Poisson regression for count outcomes, it is important to realize that the distribution is completely characterized by the mean, given the mean-variance relationship for Bernoulli and Poisson outcomes. It thus suffices to test equality in intercepts, actor and partner effects for complete indistinguishability in those models. However, given that GEE is not a likelihood-based method, one can not rely on the LRT to compare models. Instead, one may use a single multivariate Wald test for simultaneous testing of those equalities or several univariate Wald tests with Bonferroni correction for each equality separately.

It should also be noted that although SEM and MLM framework yield identical results in complete case settings (Kashy & Donnellan, 2008), this might no longer be true in the presence of missing values. MLM packages such as `nlme` apply listwise deletion, while SEM packages such as `lavaan` will use all available data if either X- and/or Y-values are missing (Ledermann & Kenny, 2017).

Finally, we would like to stress the need for better software tools for complex longitudinal dyadic modelling. In our simulation study, the covariance structure was very restricted. For example, the 'UN@AR(1)' residual covariance structure is, as far as we know, only available in the PROC MIXED statement of SAS. Although it is possible in theory to define such complex structure manually in SEM packages such as `lavaan`, the code easily becomes very cumbersome and computation time increases dramatically.

References

Baird, R., & Maxwell, S. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *12*(2), 175-188. doi: 10.1037/met0000070

Bentler, P. (2004). *EQS 6 structural equations program manual* (6th ed.). Encino, CA: Multivariate Software, Inc.

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

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: The Guildford Press.

Cheung, M. W.-L. (2013). Implementing restricted maximum likelihood estimation in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(1), 157-167. doi: 10.1080/10705511.2013.742404

Cook, W., & Kenny, D. (2005). The actor-partner interdependence model: A model of bidirectional effects in developmental studies. *International Jounal of Behavioral Development*, *29*(2), 101-109. doi: 10.1080/01650250444000405

Curran, P. (2003). Have multilevel models been structural equations models all along? *Multivariate Behavioral Research*, *38*(4), 529-569. doi: 10.1207/s15327906mbr3804_5

Decuyper, M., Gistelinck, F., Vergauwe, J., Pancorbo, G., & De Fruyt, F. (2016). Personality pathology and relationship satisfaction in dating and married couples. *Personality Disorders: Theory, Research, and Treatment*, *9*(1), 81-92. doi: 10.1037/per0000219

Dewitte, M., Van Lankveld, J., Vandenberghe, S., & Loeys, T. (2015). Sex in its daily relational context. *Journal of sexual medicine*, *12*(12), 2436-2450. doi: 10.1111/jsm.13050

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121

Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Gonzalez, R., & Griffin, D. (1999). The correlation analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*(4), 449-469. doi: 10.1037/a0025651

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320-338. doi: 10.1080/01621459.1977.10480998

Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions*, *2*(2), 110 - 120. doi: 10.1016/j.invent.2015.02.003

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide* (2nd ed.). Lincolnwood, IL: Scientific Software International, Inc.

Kashy, D., & Donnellan, M. (2008). *Comparing MLM and SEM approaches to analyzing developmental dyadic data: Growth curve models of hostility in families.* New York, NY: Routledge/Taylor & Francis Group.

Kenny, D. (2017, January 5). *DyadR: Web programs for dyadic data analysis.* Retrieved from `http://davidakenny.net/DyadR/DyadRweb.htm`

Kenny, D., Kashy, D., & Cook, W. (2006). *Dyadic data analysis.* New York, NY: The Guilford Press.

Laurenceau, J.-P., & Bolger, N. (2011). Analyzing diary and intensive longitudinal data from dyads. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 407-422). New York: Guilford Press.

Ledermann, T., & Kenny, D. (2017). Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. *Journal of Family Psychology*, *31*(4), 442-452. doi: 10.1037/fam0000290

Loeys, T., Cook, W., De Smet, O., Wietzker, A., & Buysse, A. (2014). The actor-partner interdependence model for categorical dyadic data: a user-friendly guide to GEE. *Personal relationships*, *21*(2), 225-241. doi: 10.111/pere.12028

Loeys, T., & Molenberghs, G. (2013). Modeling actor and partner effects in dyadic data when outcomes are categorical. *Psychological Methods*, *18*(2), 220-236. doi: 10.1037/a0030640

McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of reml and the kenward-roger correction. *Multivariate Behavioral Research*, *52*(5), 661-670. doi: 10.1080/00273171.2017.1344538

Muthén, L., & Muthén, B. (2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Newsom, J. (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(3), 431-447. doi: 10.1207/S15328007SEM0903_7

Olsen, J., & Kenny, D. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*(2), 127-141. doi: 10.1037/1082-989X.11.2.127

Pinheiro, J., & Bates, M. (2000). *Mixed-effects models in S and S-PLUS.* New York, NY: Springer.

Planalp, E., Du, H., Braungart-Rieker, J., & Wang, L. (2017). Growth curve modeling to studying change: A comparison of approaches using longitudinal dyadic data with distinguishable dyads. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(1), 129-147. doi: 10.1080/10705511.2016.1224088

Raudenbush, S. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Reis, H., Collins, W., & Bersheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*(6), 844-872. doi: 10.1037/0033-2909.126.6.844

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. Retrieved from `http://www.jstatsoft.org/v48/i02/` doi: 10.18637/jss.v048.i02

SAS Institute, I. (2008). *SAS/STAT 9.2 user's guide: The MIXED porcedure.* Cary, NC: SAS Institute Inc.

Appendix

2.A Corresponding R code of the illustrative examples

We will only display the R code for the C-IT for cross-sectional and longitudinal dyadic data. Moreover, for the latter case, truncated code will be presented for simplicity. The complete code and the R code for all other IT's can be found in the original supplementary materials at *wileyonlinelibrary.com.*

2.A.1 R code C-IT for the cross-sectional dyadic data

```
library(nlme) # for fitting GLS
library(lavaan) # for fitting SEM

# data_wide: your data set in the wide/dyad format
# (one line for each dyad)
# i.e., colnames(data_wide) should correspond to
# c(dyad, Y.male, Y.female, X.male, X.female)

# data_long: your data set in the long/individual
# format (one column for each variable)
# i.e., colnames(data_long) should correspond to
# c(dyad, gender, Y, X_actor, X_partner)
```

Based on ML using `lavaan`

```
model_dis <- '
 # mean structure
  Y.male ~ mean.M*1 + a.M*X.male + p.FM*X.female
  Y.female ~ mean.F*1 + a.F*X.female + p.MF*X.male
 # residual structure
  Y.male ~~ varM*Y.male + varMF*Y.female
  Y.female ~~ varF*Y.female
```

```
 # predictor mean structure
  X.male ~ meanX.M*1
  X.female ~ meanX.F*1
 # predictor covariance structure
  X.male ~~ varX.M*X.male + varX.MF*X.female
  X.female ~~ varX.F*X.female'
fit_dis <- sem(model_dis, data = data_wide,
  fixed.x = FALSE)

model_indis <- '
 # mean structure
  Y.male ~ mean*1 + a*X.male + p*X.female
  Y.female ~ mean*1 + a*X.female + p*X.male
 # residual structure
  Y.male ~~ var*Y.male + varMF*Y.female
  Y.female ~~ var*Y.female
 # predictor mean structure
  X.male ~ meanX*1
  X.female ~ meanX*1
 # predictor covariance structure
  X.male ~~ varX*X.male + varX.MF*X.female
  X.female ~~ varX*X.female'
fit_indis <- sem(model_indis, data = data_wide,
  fixed.x = FALSE)

anova(fit_dis,fit_indis)
```

Based on REML using `nlme`

```
# step 1: XYvar-IT
fit_dis <- gls(X_actor ~ gender -1,
 data = data_long,
 correlation = corSymm(form =~ 1 | dyad),
 weights = varIdent(form =~ 1 | gender))
fit_indis <- gls(X_actor ~ gender -1,
 data = data_long,
 correlation = corSymm(form =~ 1 | dyad))
stat1 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]
```

```
fit_dis <- gls(Y ~ gender + gender:X_actor
   + gender:X_partner -1,
 data = data_long,
 correlation = corSymm(form =~ 1 | dyad),
 weights = varIdent(form =~ 1 | gender))
fit_indis <- gls(Y ~ gender + gender:X_actor
   + gender:X_partner -1,
 data = data_long,
 correlation = corSymm(form =~ 1 | dyad))
stat2 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]

p1 <- 1-pchisq(stat1+stat2,df=2)


# step 2: XYmean-IT
if(p1 < 0.05){

 # Xmean
 fit_dis <- gls(X_actor ~ gender -1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad),
  weights = varIdent(form =~ 1 | gender))
 fit_indis <- gls(X_actor ~ 1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad),
  weights = varIdent(form =~ 1 | gender))
 stat1 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]

 # Ymean
 fit_dis <- gls(Y ~ gender + gender:X_actor
    + gender:X_partner -1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad),
  weights = varIdent(form =~ 1 | gender))
 fit_indis <- gls(Y ~ X_actor + X_partner,
  data = data_long,
```

```
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad),
  weights = varIdent(form =~ 1 | gender))
 stat2 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]

 p2 <- 1-pchisq(stat1+stat2,df=4)

}else{

 # Xmean
 fit_dis <- gls(X_actor ~ gender -1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad))
 fit_indis <- gls(X_actor ~ 1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad))
 stat1 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]

 # Ymean
 fit_dis <- gls(Y ~ gender + gender:X_actor
   + gender:X_partner -1,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad))
 fit_indis <- gls(Y ~ X_actor + X_partner,
  data = data_long,
  method = "ML",
  correlation = corSymm(form =~ 1 | dyad))
 stat2 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]

 p2 <- 1-pchisq(stat1+stat2,df=4)
}


# Conclusion:
# reject Y-indistinguishability as soon
# as p1 < 0.05/2 or p2 < 0.05/2
```

2.A.2 R code C-IT for the longitudinal dyadic data

```
library(nlme) # for fitting GLS
library(lavaan) # for fitting SEM

# data_wide: your data set in the wide/dyad format
# (one line for each dyad)
# i.e., colnames(data_wide) should correspond to
# c(dyad, YM.1-21, YF.1-21, XM.1-21, XF.1-21,
# QM.1, QF.1)

# data_long: your data set in the long/individual
# format (one column for each variable)
# i.e., colnames(data_long) should correspond to
# c(dyad, gender, seq, Y, X_actor, X_partner,
# Q_actor, Q_partner)
```

Based on ML using `lavaan`

```
model_dis <- '
 # random intercepts
  i.M =~ 1*YM.1 + [...] + 1*YM.21
  i.F =~ 1*YF.1 + [...] + 1*YF.21
 # random covariance structure
  i.M ~~ ivarA*i.M + iA*i.F
  i.F ~~ ivarB*i.F
 # mean structure
  YM.1 ~ mean.M*1 + a.M*XM.1 + p.FM*XF.1
    + b.MM*QM.1 + b.FM*QF.1
  [...]
  YM.21 ~ mean.M*1 + a.M*XM.21 + p.FM*XF.21
    + b.MM*QM.1 + b.FM*QF.1
  YF.1 ~ mean.F*1 + a.F*XF.1 + p.MF*XM.1
    + b.FF*QF.1 + b.MF*QM.1
  [...]
  YF.21 ~ mean.F*1 + a.F*XF.21 + p.MF*XM.21
    + b.FF*QF.1 + b.MF*QM.1
```

```
 # residual structure
  YM.1 ~~ varM*YM.1
  [...]
  YM.21 ~~ varM*YM.21
  YF.1 ~~ varF*YF.1
  [...]
  YF.21 ~~ varF*YF.21
 # predictor mean structure
  XM.1 ~ meanX.M*1
  [...]
  XM.21 ~ meanX.M*1
  XF.1 ~ meanX.F*1
  [...]
  XF.21 ~ meanX.F*1
 # predictor covariance structure
  XM.1 ~~ varX.M*XM.1
  [...]
  XM.21 ~~ varX.M*XM.21
  XF.1 ~~ varX.F*XF.1
  [...]
  XF.21 ~~ varX.F*XF.21
 # predictor mean structure
  QM.1 ~ meanQ.M*1
  QF.1 ~ meanQ.F*1
 # predictor covariance structure
  QM.1 ~~ varQ.M*QM.1
  QF.1 ~~ varQ.F*QF.1'
fit_dis <- sem(model_dis,
 data = data_wide,
 missing = "fiml",
 fixed.x = FALSE)

model_indis <- '
 # random intercepts
  i.M =~ 1*YM.1 + [...] + 1*YM.21
  i.F =~ 1*YF.1 + [...] + 1*YF.21
 # random covariance structure
  i.M ~~ ivar*i.M + iA*i.F
  i.F ~~ ivar*i.F
```

```
# mean structure
 YM.1 ~ mean*1 + a*XM.1 + p*XF.1
   + b1*QM.1 + b2*QF.1
 [...]
 YM.21 ~ mean*1 + a*XM.21 + p*XF.21
   + b1*QM.1 + b2*QF.1
 YF.1 ~ mean*1 + a*XF.1 + p*XM.1
   + b1*QF.1 + b2*QM.1
 [...]
 YF.21 ~ mean*1 + a*XF.21 + p*XM.21
   + b1*QF.1 + b2*QM.1
# residual structure
 YM.1 ~~ var*YM.1
 [...]
 YM.21 ~~ var*YM.21
 YF.1 ~~ var*YF.1
 [...]
 YF.21 ~~ var*YF.21
# predictor mean structure
 XM.1 ~ meanX*1
 [...]
 XM.21 ~ meanX*1
 XF.1 ~ meanX*1
 [...]
 XF.21 ~ meanX*1
# predictor covariance structure
 XM.1 ~~ varX*XM.1
 [...]
 XM.21 ~~ varX*XM.21
 XF.1 ~~ varX*XF.1
 [...]
 XF.21 ~~ varX*XF.21
# predictor mean structure
 QM.1 ~ meanQ*1
 QF.1 ~ meanQ*1
# predictor covariance structure
 QM.1 ~~ varQ*QM.1
 QF.1 ~~ varQ*QF.1'
```

```
fit_indis <- sem(model_indis,
 data = data_wide,
 missing = "fiml",
 fixed.x = FALSE)

anova(fit_dis, fit_indis)
```

Based on REML using `nlme`

```
# step 1: XYIvar-IT
fit_dis <- gls(X_actor ~ gender -1,
 data = data_long,
 weights = varIdent(form =~ 1 | gender),
 na.action = na.omit)
fit_indis <- gls(X_actor ~ gender -1,
 data = data_long,
 na.action = na.omit)
stat1 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

fit_dis <- gls(Q_actor ~ gender -1,
 data = data_long[which(data_long$seq == 1),],
 weights = varIdent(form =~ 1 | gender))
fit_indis <- gls(Q_actor ~ gender -1,
 data = data_long[which(data_long$seq == 1),])
stat2 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

fit_dis <- lme(Y ~ gender -1
  + gender:X_actor + gender:X_partner
  + gender:Q_actor + gender:Q_partner,
 data = data_long,
 random = list(dyad = pdSymm(~ gender -1)),
 weights = varIdent(form =~ 1|gender),
 na.action = na.omit)
fit_indis <- lme(Y ~ gender -1
  + gender:X_actor + gender:X_partner
  + gender:Q_actor + gender:Q_partner,
 data = data_long,
 random = list(dyad = pdCompSymm(~ gender -1)),
 na.action = na.omit)
```

```
stat3 <- anova(fit_dis, fit_indis)$"L.Ratio"[2]
p1 <- 1-pchisq(stat1+stat2+stat3,df=4)


# step 2: XYmean-IT
if(p1 < 0.05){

 # Xmean
 fit_dis <- gls(X_actor ~ gender -1,
  data = data_long,
  method = "ML",
  weights = varIdent(form =~ 1 | gender),
  na.action = na.omit)
 fit_indis <- gls(X_actor ~ 1,
  data = data_long,
  method = "ML",
  weights = varIdent(form =~ 1 | gender),
  na.action = na.omit)
 stat1 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

 fit_dis <- gls(Q_actor ~ gender -1,
  data = data_long[which(data_long$seq == 1),],
  method = "ML",
  weights = varIdent(form =~ 1 | gender))
 fit_indis <- gls(Q_actor ~ 1,
  data = data_long[which(data_long$seq == 1),],
  method = "ML",
  weights = varIdent(form =~ 1 | gender))
 stat2 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

 # Ymean
 fit_dis1 <- lme(Y ~ gender -1
   + gender:X_actor + gender:X_partner
   + gender:Q_actor + gender:Q_partner,
  data = data_long,
  method = "ML",
  random = list(dyad = pdSymm(~ gender -1)),
  weights = varIdent(form =~ 1|gender),
  na.action = na.omit)
```

```
 fit_indis1 <- lme(Y ~ X_actor + X_partner
   + Q_actor + Q_partner,
  data = data_long,
  method = "ML",
  random = list(dyad = pdSymm(~ gender -1)),
  weights = varIdent(form =~ 1|gender),
  na.action = na.omit)
 stat3 <- anova(fit_dis1,fit_indis1)$"L.Ratio"[2]

 p2 <- 1-pchisq(stat1+stat2+stat3,df=7)

}else{

 fit_dis <- gls(X_actor ~ gender -1,
  data = data_long,
  method = "ML",
  na.action = na.omit)
 fit_indis <- gls(X_actor ~ 1,
  data = data_long,
  method = "ML",
  na.action = na.omit)
 stat1 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

 fit_dis <- gls(Q_actor ~ gender -1,
  data = data_long[which(data_long$seq == 1),],
  method = "ML")
 fit_indis <- gls(Q_actor ~ 1,
  data = data_long[which(data_long$seq == 1),],
  method = "ML")
 stat2 <- anova(fit_dis,fit_indis)$"L.Ratio"[2]

 fit_dis1 <- lme(Y ~ gender -1
   + gender:X_actor + gender:X_partner
   + gender:Q_actor + gender:Q_partner,
  data = data_long,
  method = "ML",
  random = list(dyad = pdCompSymm(~ gender -1)),
  na.action = na.omit)
```

```
 fit_indis1 <- lme(Y ~ X_actor + X_partner
   + Q_actor + Q_partner,
  data = data_long,
  method = "ML",
  random = list(dyad = pdCompSymm(~ gender -1)),
  na.action = na.omit)
 stat3 <- anova(fit_dis1,fit_indis1)$"L.Ratio"[2]

 p2 <- 1-pchisq(stat1+stat2+stat3,df=7)
}

# Conclusion
# reject Y-indistinguishability as soon
# as p1 < 0.05/2 or p2 < 0.05/2
```

2.B The analytic derivation of the relationship between ML and REML

In this appendix we demonstrate the analytic connection between the REML and ML (co)variance estimates (see equation (2.6)), as well as the connection between the REML- and ML-based LRT statistics in the Y-var indistinguishability test for cross-sectional dyadic data (i.e., equation (2.7)) when using the APIM.

2.B.1 Y-var distinguishability

As mentioned in the body of the text, the ML and REML estimation correspond to the maximization of equation (2.4) and (2.5), respectively. First, suppose the dyads are Y-var distinguishable. For model (2.1), this would imply the different matrices R, V and X in equations (2.4) and (2.5) to be as follows:

$$V = \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_N \end{pmatrix}, \quad \text{with} \quad V_j = \begin{pmatrix} \sigma^2_{YM} & \sigma_{YMF} \\ \sigma_{YMF} & \sigma^2_{YF} \end{pmatrix},$$

$$R = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix}, \quad \text{with} \quad R_j = \begin{pmatrix} e_{Mj} \\ e_{Fj} \end{pmatrix},$$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \quad \text{with} \quad X_j = \begin{pmatrix} 1 & A_{Mj} & P_{Mj} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & A_{Fj} & P_{Fj} \end{pmatrix},$$

in which A refers to the matrix of actor effects and P to the matrix of partner effects. We can now rewrite each of the elements of the log-likelihood function from equation (2.7):
first,

$$\begin{aligned} |V| &= |V_1| \cdot \dots \cdot |V_N| \\ &= |V_j|^N \\ &= \left\{\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YMF}\right\}^N ; \end{aligned}$$

second,

$$\begin{aligned} R^t V^{-1} R &= \frac{1}{|V_j|} \sum_{j=1}^{N} R_j^t V_j^{-1} R_j \\ &= \frac{1}{|V_j|} \left\{ \sigma_{YF}^2 \sum_{j=1}^{N} e_{Mj}^2 - 2\sigma_{YFM} \sum_{j=1}^{N} e_{Mj} e_{Fj} + \sigma_{YM}^2 \sum_{j=1}^{N} e_{Fj}^2 \right\}; \end{aligned}$$

and third,

$$\begin{aligned} X^t V^{-1} X &= \frac{1}{|V_j|} \sum_{j=1}^{N} X_j^t V_j^{-1} X_j \\ &= \frac{1}{|V_j|} \begin{pmatrix} \sigma_{YF}^2 A & -\sigma_{YFM} B \\ -\sigma_{YFM} B^t & \sigma_{YM}^2 C \end{pmatrix}, \end{aligned}$$

with

$$A = \begin{pmatrix} N & \sum_{j=1}^{N} A_{Mj} & \sum_{j=1}^{N} P_{Mj} \\ \sum_{j=1}^{N} A_{Mj} & \sum_{j=1}^{N} A_{Mj}^2 & \sum_{j=1}^{N} A_{Mj} P_{Mj} \\ \sum_{j=1}^{N} P_{Mj} & \sum_{j=1}^{N} A_{Mj} P_{Mj} & \sum_{j=1}^{N} P_{Mj}^2 \end{pmatrix},$$

$$B = \begin{pmatrix} N & \sum_{j=1}^{N} A_{Fj} & \sum_{j=1}^{N} P_{Fj} \\ \sum_{j=1}^{N} A_{Mj} & \sum_{j=1}^{N} A_{Mj} A_{Fj} & \sum_{j=1}^{N} A_{Mj} P_{Fj} \\ \sum_{j=1}^{N} P_{Mj} & \sum_{j=1}^{N} A_{Fj} P_{Mj} & \sum_{j=1}^{N} P_{Mj} P_{Fj} \end{pmatrix},$$

$$C = \begin{pmatrix} N & \sum_{j=1}^{N} A_{Fj} & \sum_{j=1}^{N} P_{Fj} \\ \sum_{j=1}^{N} A_{Fj} & \sum_{j=1}^{N} A_{Fj}^2 & \sum_{j=1}^{N} A_{Fj} P_{Fj} \\ \sum_{j=1}^{N} P_{Fj} & \sum_{j=1}^{N} A_{Fj} P_{Fj} & \sum_{j=1}^{N} P_{Fj}^2 \end{pmatrix}.$$

So, the extra term in the REML-based log-likelihood function can be reformulated as

$$\begin{aligned}\log\left|X^tV^{-1}X\right| &= \log\left(\left(\frac{1}{|V_j|}\right)^{2p}\begin{vmatrix}\sigma^2_{YF}A & -\sigma_{YFM}B\\ -\sigma_{YFM}B^t & \sigma^2_{YM}C\end{vmatrix}\right)\\ &= -2p\log|V_j| + \log\left|\sigma^2_{YM}\sigma^2_{YF}AC - \sigma^2_{YFM}B^tB\right|\\ &= -2p\log|V_j| + \log\left|(\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YFM})AC\right|\\ &= -p\log|V_j| + \log|AC|.\end{aligned}$$

Note that the simplifications are possible due to the APIM characteristics: $A_{Mj} = P_{Fj}$ and $A_{Fj} = P_{Mj}$. Now, putting all the pieces together, one can find the following expression for the log-likelihood function:

$$\begin{aligned}-2\ \text{logLik}_{\text{ML}} &= 2N\log(2\pi) + N\log\left(\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YFM}\right)\\ &\quad + \frac{1}{\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YFM}}\left\{\sigma^2_{YF}\sum_{j=1}^{N}e^2_{Mj}\right.\\ &\quad\left. -2\sigma_{YFM}\sum_{j=1}^{N}e_{Mj}e_{Fj} + \sigma^2_{YM}\sum_{j=1}^{N}e^2_{Fj}\right\},\end{aligned}$$

$$\begin{aligned}-2\ \text{logLik}_{\text{REML}} &= 2(N-p)\log(2\pi) + (N-p)\log\left(\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YFM}\right)\\ &\quad + \log|AC| + \frac{1}{\sigma^2_{YM}\sigma^2_{YF} - \sigma^2_{YFM}}\left\{\sigma^2_{YF}\sum_{j=1}^{N}e^2_{Mj}\right.\\ &\quad\left. -2\sigma_{YFM}\sum_{j=1}^{N}e_{Mj}e_{Fj} + \sigma^2_{YM}\sum_{j=1}^{N}e^2_{Fj}\right\}.\end{aligned} \tag{2.18}$$

We will take the derivative of (2.18) to obtain the formulas for the ML and REML estimators for the (co)variance parameters. The resulting formulas are as follows:

$$\begin{aligned}\hat\sigma^2_{\text{ML},YM} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}e^2_{Mj}, & \hat\sigma^2_{\text{REML},YM} &= \tfrac{1}{N-p}\textstyle\sum_{j=1}^{N}e^2_{Mj},\\ \hat\sigma^2_{\text{ML},YF} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}e^2_{Fj}, & \hat\sigma^2_{\text{REML},YF} &= \tfrac{1}{N-p}\textstyle\sum_{j=1}^{N}e^2_{Fj},\\ \hat\sigma_{\text{ML},YMF} &= \tfrac{1}{N}\textstyle\sum_{j=1}^{N}e_{Mj}e_{Fj}, & \hat\sigma_{\text{REML},YMF} &= \tfrac{1}{N-p}\textstyle\sum_{j=1}^{N}e_{Mj}e_{Fj}.\end{aligned} \tag{2.19}$$

Hence, (minus twice) the log-likelihood function equals:

$$
\begin{aligned}
-2\ \text{logLik}_{\text{ML}} &= 2N\log(2\pi) - 2N\log(N) + 2N \\
&\qquad + N\log\left\{\sum_{j=1}^{N} e_{Mj}^2 \sum_{j=1}^{N} e_{Fj}^2 - \left(\sum_{j=1}^{N} e_{Mj}e_{Fj}\right)^2\right\},
\end{aligned}
$$

$$
\begin{aligned}
-2\ \text{logLik}_{\text{REML}} &= 2(N-p)\log(2\pi) - 2(N-p)\log(N-p) + 2(N-p) \\
&\qquad + \log|AC| + (N-p)\log\left\{\sum_{j=1}^{N} e_{Mj}^2 \sum_{j=1}^{N} e_{Fj}^2 - \left(\sum_{j=1}^{N} e_{Mj}e_{Fj}\right)^2\right\}.
\end{aligned}
\tag{2.20}
$$

2.B.2 Y-var indistinguishability

Now consider the dyads to be Y-var indistinguishable. The residual (co)-variance matrix should be adapted to:

$$
V = \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_N \end{pmatrix} \quad \text{with} \quad V_j = \begin{pmatrix} \sigma_Y^2 & \sigma_Y^2\rho \\ \sigma_Y^2\rho & \sigma_Y^2 \end{pmatrix}.
$$

Hence, we rewrite

$$
\begin{aligned}
\log\{|V|\} &= N\log\left\{\sigma_Y^4\left(1-\rho^2\right)\right\}, \\
R^t V^{-1} R &= \frac{1}{|V_j|}\sigma_Y^2\left\{\sum_{j=1}^{N}\left(e_{Mj}^2 + e_{Fj}^2\right) - 2\rho\sum_{j=1}^{N} e_{Mj}e_{Fj}\right\}.
\end{aligned}
$$

As a result, the formulas for (minus twice) the log-likelihood function with respect to the (co)variance functions become

$$
\begin{aligned}
-2\ \text{logLik}_{\text{ML}} &= 2N\log(2\pi) + N\log\left\{\sigma_Y^4(1-\rho^2)\right\} \\
&\qquad + \frac{1}{\sigma_Y^2(1-\rho^2)}\left\{\sum_{j=1}^{N}\left(e_{Mj}^2 + e_{Fj}^2\right) - 2\rho\sum_{j=1}^{N} e_{Mj}e_{Fj}\right\},
\end{aligned}
$$

$$
\begin{aligned}
-2\ \text{logLik}_{\text{REML}} &= 2(N-p)\ log(2\pi) + (N-p)\log\left\{\sigma_Y^4(1-\rho^2\right\} \\
&\qquad + \log|AC| + \frac{1}{\sigma_Y^2(1-\rho^2)}\left\{\sum_{j=1}^{N}\left(e_{Mj}^2 + e_{Fj}^2\right) - 2\rho\sum_{j=1}^{N} e_{Mj}e_{Fj}\right\}.
\end{aligned}
\tag{2.21}
$$

The derivative is taken from the expressions in (2.21) to obtain the ML and REML, formulas for the (co)variance parameters:

$$\hat{\sigma}^2_{\mathrm{ML},Y} = \tfrac{1}{2N}\textstyle\sum_{j=1}^N \left(e^2_{Mj} + e^2_{Fj}\right), \quad \hat{\sigma}^2_{\mathrm{REML},Y} = \tfrac{1}{2(N-p)}\textstyle\sum_{j=1}^N \left(e^2_{Mj} + e^2_{Fj}\right),$$

$$\hat{\rho}_{\mathrm{ML}} = 2\tfrac{\sum_{j=1}^N e_{Mj}e_{Fj}}{\sum_{j=1}^N \left(e^2_{Mj}+e^2_{Fj}\right)}, \quad \hat{\rho}_{\mathrm{REML}} = \hat{\rho}_{\mathrm{ML}}. \tag{2.22}$$

Again, just as in the distinguishable case, the following equations hold:

$$\hat{\sigma}^2_{\mathrm{REML},Y} = \tfrac{N}{N-p}\hat{\sigma}^2_{\mathrm{ML},Y}, \qquad \hat{\sigma}_{\mathrm{REML},YMF} = \tfrac{N}{N-p}\hat{\sigma}_{\mathrm{ML},YMF}. \tag{2.23}$$

Substituting these expressions into equations (2.4) and (2.5), one obtains the log-likelihood as a function of the residuals only:

$$\begin{aligned}
-2\ \mathrm{logLik}_{\mathrm{ML}} &= 2N\log(2\pi) - 2N\log(N) + 2N \\
&\quad + N\log\left\{\frac{1}{4}\left[\sum_{j=1}^N e^2_{Mj} + \sum_{j=1}^N e^2_{Fj}\right]^2 - \left[\sum_{j=1}^N e_{Mj}e_{Fj}\right]^2\right\},
\end{aligned} \tag{2.24}$$

$$\begin{aligned}
-2\ \mathrm{logLik}_{\mathrm{REML}} &= 2(N-p)\log(2\pi) - 2(N-p)\log(N-p) \\
&\quad + 2(N-p) + \log|AC| \\
&\quad + (N-p)\log\left\{\frac{1}{4}\left[\sum_{j=1}^N e^2_{Mj} + \sum_{j=1}^N e^2_{Fj}\right]^2 - \left[\sum_{j=1}^N e_{Mj}e_{Fj}\right]^2\right\}.
\end{aligned}$$

Taking the difference between the corresponding equations (2.24) and (2.20), one obtains the connection between the likelihood ratio test statistic using ML and REML for the Y-var indistinguishability test:

$$\chi^2_{\mathrm{ML}} = N\log\left\{\frac{1}{4}\left[\sum_{j=1}^{N} e^2_{Mj} + \sum_{j=1}^{N} e^2_{Fj}\right]^2 - \left[\sum_{j=1}^{N} e_{Mj}e_{Fj}\right]^2\right\} \\ - N\log\left\{\sum_{j=1}^{N} e^2_{Mj}\sum_{j=1}^{N} e^2_{Fj} - \left(\sum_{j=1}^{N} e_{Mj}e_{Fj}\right)^2\right\},$$

$$\chi^2_{\mathrm{REML}} = (N-p)\log\left\{\frac{1}{4}\left[\sum_{j=1}^{N} e^2_{Mj} + \sum_{j=1}^{N} e^2_{Fj}\right]^2 - \left[\sum_{j=1}^{N} e_{Mj}e_{Fj}\right]^2\right\} \\ - (N-p)\log\left\{\sum_{j=1}^{N} e^2_{Mj}\sum_{j=1}^{N} e^2_{Fj} - \left(\sum_{j=1}^{N} e_{Mj}e_{Fj}\right)^2\right\}. \tag{2.25}$$

Hence, the connection can be written as:

$$\chi^2_{\mathrm{REML}} = \frac{N-p}{N}\chi^2_{\mathrm{ML}} \tag{2.26}$$

3 The actor-partner interdependence model for longitudinal dyadic data: an implementation in the SEM framework

Abstract. In dyadic research, the actor-partner interdependence model (APIM) is widely used to model the effect of a predictor measured across dyad members on one's own and one's partner outcome. When such dyadic data are measured repeatedly over time, both the non-independence within couples and the non-independence over time need to be accounted for. In this paper, we present a longitudinal extension of the APIM, the L-APIM, that allows for both stable and time-varying sources of non-independence. Its implementation is readily available in multilevel software, such as proc mixed in SAS, but is lacking in the structural equation modeling (SEM) framework. We tackle the computational challenges associated with its SEM-implementation and propose a user-friendly free application for the L-APIM, which can be found at "`http://fgisteli.shinyapps.io/Shiny_LDD2/`". As an illustration, we explore the actor and partner effects of positive relationship feelings on next day's intimacy using 3-week diary data of 66 heterosexual couples.

This chapter has been published in Structural Equation Modeling: A Multidisciplinary Journal. Gistelinck, F. & Loeys, T. (2019). The Actor–Partner Interdependence Model for Longitudinal Dyadic Data: An Implementation in the SEM Framework. *Structural Equation Modeling: A Multidisciplinary Journal, 26(3)*, 329-347.

3.1 Introduction

Over the last decade, the popularity of dyadic data has been increasing exponentially. Most studied phenomena in behavioral and social science are interpersonal by definition. Indeed, daily lives are rarely spent in isolation from each other and contexts for human beings involve close dyadic relationships (Reis, Collins, & Bersheid, 2000). These dyadic interactions inhibit complex research questions. The richness of these questions becomes even more profound when the dyads are studied over time.

Consider for example the study of Dewitte, Van Lankveld, Vandenberghe, and Loeys (2015), in which 66 heterosexual couples in Flanders performed a daily diary study on sexual behavior over a period of three weeks using a standard dyadic design (Kashy & Kenny, 2000). Each person corresponded to one and only one dyad, and both men and women were asked about their sexual and intimate behavior, as well as their individual, relational and partner-related feelings and behavior of the past day. Suppose interest lies in exploring the association between the positive feelings about the relationship and next-day's perception of intimacy in those couples. Addressing such research question with longitudinal dyadic data (LDD) entails different statistical challenges. First, one needs to acknowledge the presence of non-independence between the two members of a dyad (Gonzalez & Griffin, 1999). For instance, the amount of perceived intimacy on a specific day between two couple members will show more similarities compared to two random people. Ignoring this non-independence will lead to incorrect inference as it violates the regression postulate of independent observations (Cook & Kenny, 2005). Second, one must also contend with non-independence of the observations within a dyad member due to the repeated-measurement design. This type of non-independence is sometimes referred to as *autocorrelation* and depicts the association between a variable measured at one point in time and the same variable measured at another point in time. Today's perception of intimacy is likely to be positively correlated with tomorrow's perception of intimacy (Cranford et al., 2006). Ignoring this temporal correlation will also result in incorrect inference (Fitzmaurice, Laird, & Ware, 2011; Hox, 2010). Third, one needs to account for the effect of one person's behavior or emotions on his/her own score (i.e., *actor effect*), as well as for the effect of one partner's behavior or emotions on that person's score (i.e., *partner effect*). For example, a person might perceive more intimacy if he or she feels more positively affected by the relationship, but also if his or her partner experiences more positive relationship feelings. Fourth, one should acknowledge within- and between-dyad (member) variation (Bolger & Laurenceau, 2013). The effect of overtime-predictors can be disentangled into an average effect over time (i.e., *time-averaged effect*) and a deviant effect towards this average at a particular moment in time (i.e., *time-specific effect*). For example, the average level of positive relationship feelings over time may have a different effect between persons on the perception of intimacy than a sudden drop or increase in positive relationship feelings within a person on a certain day. If one does not separate those effects, the estimated effect will be a mixture of both the time-specific and time-averaged effect (Enders & Tofighi, 2007).

Older approaches for dealing with LDD often fail to incorporate all four issues simultaneously. Formerly, researchers have been dealing with dyadic data by fitting two separate models for each dyad member (with or without reciprocal effects). This approach makes it impossible to answer research questions about non-independence between the members of a dyad (Planalp, Du, Braungart-Rieker, & Wang, 2017). Alternatively, researchers took the average of the two member's scores. The modified data lose information as one is no longer able to answer research questions about between-member effects of a variable. This often leads to misleading conclusions as two potentially very distinct scores are equated (Cook & Kenny, 2004).

More recently, researchers have been developing several classes of models for LDD, each model answering different types of research questions. The first class assumes the outcomes from dyad members to follow a particular trajectory over time (Kenny & Kashy, 2011). For example, these models focus on research questions such as "How does the intimate behavior evolve over time for each dyad member, and how do these trajectories covary?". A (different) time trend for each dyad member is then fitted simultaneously (Lyons & Sayer, 2005; Raudenbush, R.T., & Barnett, 1995), for example assuming (latent) growth curve models (Atkins, 2005; Kenny & Kashy, 2011; Newsom, 2002; Peugh, DiLillo, & Panuzio, 2013). A second class models a dyad member's score as a function of his/her own and his/her partner's past scores. Such models focus on questions like "How does one's own and one's partner yesterday's intimacy affect today's amount of intimacy?". Cross-lagged autoregressive regression models for LDD describe such reciprocal relationships or directional influences over time in a single variable for both dyad members simultaneously (Ackerman, Donnellan, Kashy, & Conger, 2012; Cook & Kenny, 2005; Nestler, Grimm, & Schönbrodt, 2015). A third class explores the effect of an antecedent on current behavior of each dyad member (Bolger & Laurenceau, 2013; Laurenceau & Bolger, 2005). This longitudinal actor-partner interdependence model enables researcher to answer questions such as "How does the positive relationship feelings of a partner affect the next-day perception of intimacy of the actor?" (Kenny, Kashy, & Cook, 2006).

In this article, we will focus on the latter model, further abbreviated as the L-APIM. While Kenny et al. (2006) only briefly mentioned this model in the multilevel modeling (MLM) framework, we will elaborate on this model as it is the most natural extension of the popular cross-sectional actor-partner interdependence model (APIM) to the longitudinal setting. Since hardly any documentation is available on its implementation (e.g., Kenny et al. (2006) only implemented the model using SAS software), a more flexible implementation of this model is presented here. More specifically, we will show how this model can also be fitted within the structural equation modeling (SEM) framework.

The outline of the paper is as follows. First, we briefly review the cross-sectional APIM. Next, we introduce its extension to the longitudinal setting: the L-APIM. We show how the L-APIM can incorporate the four aforementioned issues: (1) the correlation between dyad members, (2) the autocorrelation, (3) the effect of actor and partner characteristics and (4) the partitioning into time-specific and time-averaged effects. Thereafter, we discuss its implementation within the MLM and SEM framework. The challenges within the SEM framework are especially highlighted. Finally, we present an application written within the statistical software R using the package `shiny` (Chang, Cheng, Allaire, Xie, & McPherson, 2017; RStudio Inc., 2017), that enables researchers to fit the L-APIM within the SEM framework, hereby relying on `lavaan`. We illustrate this application with empirical data, and end with a discussion.

3.2 The Cross-sectional APIM and its implementation

One of the most widely used models to analyze cross-sectional dyadic data is the actor-partner interdependence model (Kashy & Kenny, 2000). The APIM is generally formulated as follows:

$$\begin{cases} Y_{Mj} = \mu_M + a_{M(X)}X_{Mj} + p_{FM(X)}X_{Fj} + \varepsilon_{Mj} \\ Y_{Fj} = \mu_F + a_{F(X)}X_{Fj} + p_{MF(X)}X_{Mj} + \varepsilon_{Fj} \end{cases}, \tag{3.1}$$

with j the dyad number and the residuals ε, disturbance terms which are correlated within a dyad. A path diagram of the model can be found in Figure 3.2.1. Here, the two members of the dyad are represented by the index “M” and “F” for males and females respectively, but other types of distinguishable dyads can be used as well (e.g., brother and sister, oldest and youngest child, etc.). Reconsidering the study of Dewitte et al. (2015) from the introduction, the predictors X_F and X_M may represent, for example, the positive relationship feelings of the females and males, respectively, while Y_F and Y_M correspond to the perceived intimacy of the females and males, respectively. The model can also be used for indistinguishable dyads (e.g., twin brothers, same-sex couples, etc.) in which the roles of the dyad members do not differ and the dyad member indexes are treated as arbitrary (Olsen & Kenny, 2005). If dyads are indistinguishable, constraints can be added to the model in order to ensure equality in intercepts, actor and partner effects, and variances (Gistelinck, Loeys, Decuyper, & Dewitte, 2018). In the remainder of the paper, we focus on distinguishable dyads for the ease of exposition, although the application presented later allows for indistinguishable dyads as well.

The cross-sectional APIM in Figure 3.2.1 is very popular among dyadic researcher. First, it allows the residuals of the outcome variable to covary. This way, the non-independence between the two members of a dyad is taken into account (cfr. issue 1 from the introduction). Furthermore, the model naturally incorporates the effect of one’s own characteristics on his/her outcome values (i.e., the actor effects $a_{M(X)}$ and $a_{F(X)}$), as well as the effect of one’s partner characteristics on his/her own outcome values (i.e., the partner effects $p_{FM(X)}$ and $p_{MF(X)}$). Due to the presence of these reciprocal effects, each dyad member influences the other member, and vice versa (cfr. issue 3 from the introduction). Note that in model (3.1), the predictor X is a covariate measured at the level of the dyad member (i.e. *member variables*). In the study of Dewitte et al. (2015), the positive relationship feelings are member variables as they may be different for each person within the dyad. One can easily extend the model, however, with predictors measured at the dyad level (i.e., *dyad variables*), such as the duration of a relationship. Dyad variables do not have a separate actor and partner partner effect though.

The cross-sectional APIM can be fitted within different modeling frameworks. The two most prominent ones are the multilevel modeling (MLM) and structural equation modeling (SEM) framework (Ledermann & Kenny, 2017). The MLM approach, also called hierarchical modeling or mixed effect modeling, extends regression models to dependent data structures using random effects to capture the variation on a particular hierarchical level (Hox, 2010). Here, the non-independence is usually captured by a random intercept for each dyad. This implies that the covariance between dyad members is modeled as a variance and thus assumed to be positive. This poses some restriction because outcomes within dyads may also be negatively correlated. Alternatively, one may explicitly model the residual covariance structure in a marginal approach (Kenny & Kashy, 2011). Both ways of modeling the covariance structure are also possible in the SEM framework. In the latter framework, random intercepts are viewed as latent variables. Although these two frameworks have totally different origins, the SEM and MLM approach have been compared in several papers in the individual setting (Curran, 2003), as well as in the dyadic setting (Wendorf, 2002), providing empirical evidence that both frameworks yield (nearly) identical results.

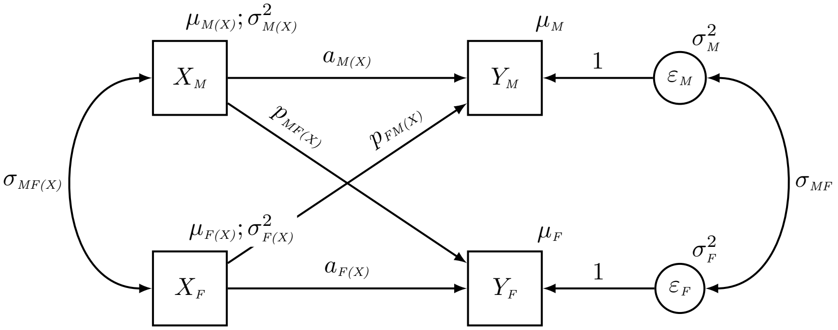


Figure 3.2.1 A schematic overview of the actor partner interdependence model (APIM), a model used to analyze cross-sectional dyadic data. The index 'M' refers to the first dyad member (e.g., male), while 'F' refers to the second member (e.g., female).

3.3 The L-APIM for longitudinal dyadic data

If predictors and outcomes in dyads are measured repeatedly over time, one obtains longitudinal dyad data. In order to fit LDD, one could use the longitudinal APIM or L-APIM. The L-APIM can generally be formulated as follows

$$\begin{cases} Y_{Fij} = (\mu_F + \eta_{Fj}) + a_{F(X)} X_{Fij} + p_{MF(X)} X_{Mij} + \varepsilon_{Fij} \\ Y_{Mij} = (\mu_M + \eta_{Mj}) + a_{M(X)} X_{Mij} + p_{FM(X)} X_{Fij} + \varepsilon_{Mij} \end{cases}, \tag{3.2}$$

with i referring to the time point ($i = 1, \ldots, T$) and j to the dyad number ($j = 1, \ldots, N$). A graphical representation of the model can be found in Figure 3.3.1. For ease of explanation, we will again refer to the first dyad member as males ('M') and second dyad member as females ('F') for the remainder of our exposition. However, similar to the cross-sectional case, other types of distinguishable dyads can be considered. In model (3.2), the predictor X is again considered to be a member variable. Yet, similar as before, one can easily extend this model with predictors measured at the dyad level. The L-APIM naturally extends the cross-sectional APIM as it simply looks like fitting the model at each time point. However, things are a bit more complicated. We will first discuss the mean structure and the covariance structure of the proposed L-APIM thereafter, since the latter is more complex than the former.

3.3.1 The mean structure of the L-APIM

First, it is important to incorporate the effect of one's own and one's partner characteristics (cfr. issue 3 from the introduction). Similar to the cross-sectional setting, this is possible by introducing actor and partner effects (i.e., $a_{M(X)}$ and $a_{F(X)}$, respectively $p_{FM(X)}$ and $p_{MF(X)}$) for the member variables. However, in the longitudinal case, a dyad or member variable X can be further categorized into an *overtime* or *time-constant* variable, depending on its variation over time. For instance, positive relational feelings is an overtime-member variable, while personality traits can be considered a time-constant-member variable in the course of the study.

As mentioned in the introduction, it is important to realize that overtime predictors may have a time-specific and time-averaged effect (cfr. issue 4 from the introduction). The time-averaged effect captures variation in the means between persons, while the time-specific effect captures

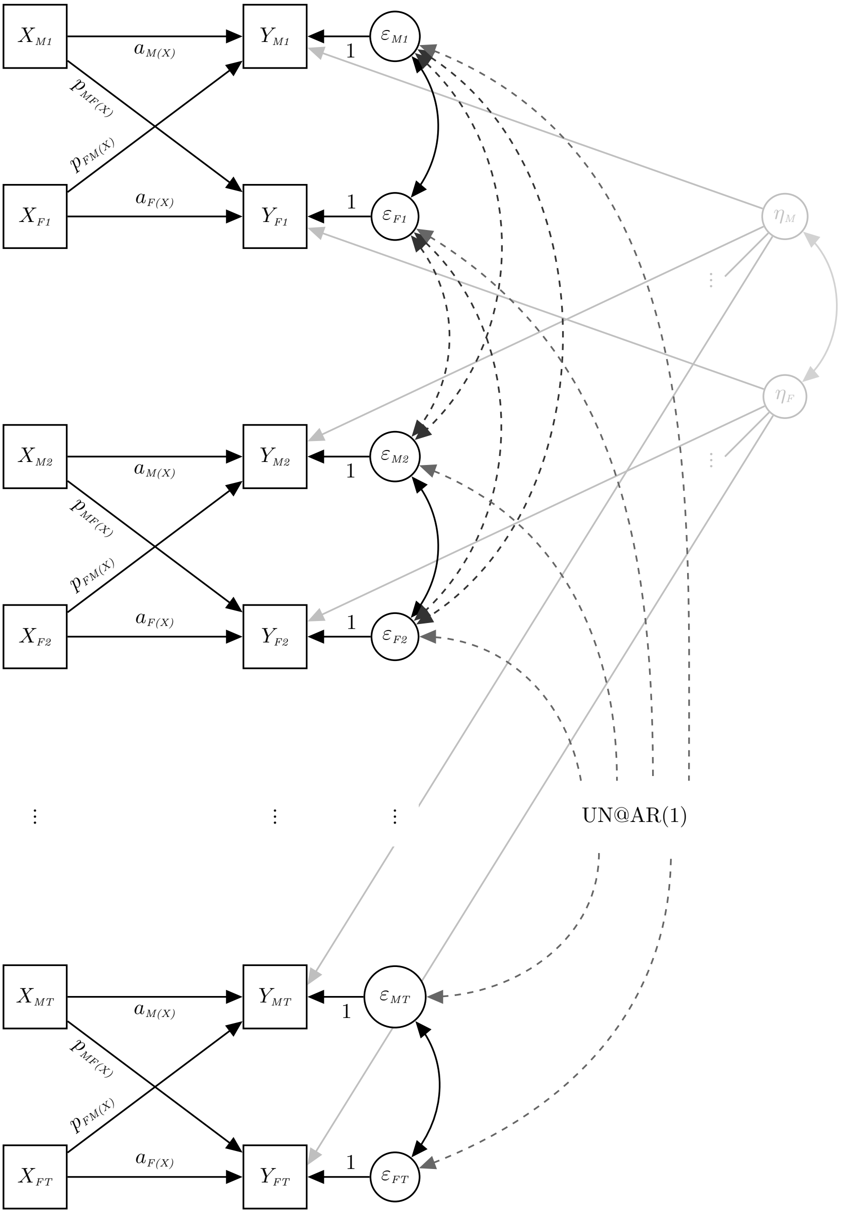


Figure 3.3.1 A schematic overview of the longitudinal actor-partner interdependence model (L-APIM), a model used to analyze longitudinal dyadic data, measured over T time points. The index 'M' refers to the first dyad member (e.g., male), while 'F' refers to the second member (e.g., female).

the variation within each person in the deviations of each score from the member-specific average (Bolger & Laurenceau, 2013). Suppose, for example, X_{Fij} corresponds to an overtime-member predictor in females. The time-averaged effect is defined as the effect of the average of the overtime predictor over all time points for each women separately, denoted as $X_{F.j}$. Note that the mean $X_{F.j}$ now becomes a time-constant variable as it does no longer vary over time. The time-specific effect at time i is then computed as the effect of the difference between X_{Fij} and its time-averaged component $X_{F.j}$. Hence, rather than model (3.2), we will consider the following model:

$$\begin{cases} Y_{Fij} = (\mu_F + \eta_{Fj}) + a_{F(XA)}X_{F.j} + a_{F(XS)}(X_{Fij} - X_{F.j}) \\ \qquad\qquad + p_{MF(XA)}X_{M.j} + p_{MF(XS)}(X_{Mij} - X_{M.j}) + \varepsilon_{Fij} \\ \\ Y_{Mij} = (\mu_M + \eta_{Mj}) + a_{M(XA)}X_{M.j} + a_{M(XS)}(X_{Mij} - X_{M.j}) \\ \qquad\qquad + p_{FM(XA)}X_{F.j} + p_{FM(XS)}(X_{Fij} - X_{F.j}) + \varepsilon_{Mij} \end{cases}, \quad (3.3)$$

with the indeces XA and XS in the parameters referring to the time-averaged and time-specific effect of the predictor X, respectively. In contrast to model (3.2), model (3.3) will allow us to estimate the time-averaged and time-specific effects. If one does not perform this split-up, the estimated effect of X in model (3.2) will be a mixture of both time-averaged and time-specific effect (Enders & Tofighi, 2007). A similar procedure can be performed for the overtime-dyad variables.

Note that this split-up in time-specific and time-averaged components of an overtime variable should not be confused with a grand-mean centering procedure. When grand-mean centering a variable, whether time-invariant or not, one takes the average over all scores, thus over all time points, dyad members and dyads (i.e., $X_{...} =: \bar{X}$). This overall average $\bar{X}$ is then subtracted from all the scores in order to obtain the grand-mean centered values of the variable.

It is important for the reader to note that model (3.3) implies some important assumptions. First, all actor and partner effects are assumed to be time-independent, and time is not included into the model. This may be reasonable in diary studies such as in the study of Dewitte et al. (2015), where the daily measurements are not collected around a specific event. These assumptions can be relaxed, but for ease of exposition, we assume them to hold here. Second, only the intercept is considered to be member-specific. This imposes the assumption that no unmeasured slope heterogeneity is present in the data. One could, of course, allow predictor effects to be member-specific (Bolger & Laurenceau, 2013). Yet, this might unnecessary complicate the model and often induces convergence problems. Treating predictor effects as fixed rather than random does not introduce bias in the fixed effects estimator, but may distort the standard errors (Baird & Maxwell, 2016).

3.3.2 The covariance structure of the L-APIM

Next, we discuss the covariance structure. In model (3.3), the member-specific intercepts η_{Mj} and η_{Fj} describe random variation around the mean outcome (μ_F and μ_M) and are assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \eta_{Fj} \\ \eta_{Mj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_F^2 & \tau_{FM} \\ \tau_{FM} & \tau_M^2 \end{pmatrix}\right). \tag{3.4}$$

In MLM terminology, the covariance structure of the random effects corresponds to the G-side of the covariance structure. The covariance of the outcome disturbances (ε_{Mij} and ε_{Fij}) defines the so-called R-side in MLM terminology. Similar to the cross-sectional APIM, the outcomes within a dyad are typically correlated at each time point (cfr. solid arrows in Figure 3.3.1) in order to capture the non-independence within dyads (cfr. issue 1 from the introduction). Additionally, time specific correlations (see dashed arrows in Figure 3.3.1) need to be taken into account as well (cfr. issue 2 from the introduction). If one would assume a completely general pattern (i.e., unstructured covariance matrix), the $2T \times 2T$ symmetric covariance matrix would imply the estimation of $\frac{2T(2T+1)}{2}$ covariance parameters (i.e., 903 residual covariance in the 21 day diary study of Dewitte et al. (2015)). This can be greatly simplified by making two reasonable assumptions:

1. The dyad members (co)variances are assumed to be stable over time.
2. The adjacent time points are correlated by a first-order autoregressive process.

Note that such an autoregressive process only makes sense if timing of measurements is equally spaced. These two assumptions reduce the number of residual covariance parameters to exactly four, independent on the amount of time points in the data set: a residual variance for the males, a residual variance for the females, a residual covariance between males and females and the autoregressive correlation parameter. The proposed covariance structure can be represented by a Kronecker Product of standard covariance structures (Bock, 1997; Bolger & Laurenceau, 2013). More specifically, the residual covariance structure corresponds to

$$\mathrm{Cov}\begin{pmatrix}\varepsilon_{F1j}\\ \varepsilon_{F2j}\\ \vdots\\ \varepsilon_{FTj}\\ \varepsilon_{M1j}\\ \varepsilon_{M2j}\\ \vdots\\ \varepsilon_{MTj}\end{pmatrix}=\begin{pmatrix}\sigma_F^2 & \sigma_{FM}\\ \sigma_{FM} & \sigma_M^2\end{pmatrix}\otimes\begin{pmatrix}1 & \rho & \cdots & \rho^{T-2} & \rho^{T-1}\\ \rho & 1 & \cdots & \rho^{T-3} & \rho^{T-2}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \rho^{T-2} & \rho^{T-3} & \cdots & 1 & \rho\\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1\end{pmatrix}. \quad (3.5)$$

In expression (3.5), σ_M^2 and σ_F^2 represent the residual variance for males and females, respectively. The covariance between males and females at each time point is captured by σ_{FM}. Correlation between different time points within a dyad member or between the two dyad members are both assumed to be identical and can be characterized by the first-order correlation parameter ρ. In other words, the non-independence between time points decreases as they are further away from each other in time. We will show that, although the residual covariance structure looks very simple, its implementation is not that easy, especially in the SEM framework.

3.4 The implementation of the L-APIM

Similar to the cross-sectional APIM (Ledermann & Kenny, 2017), it is possible to fit the L-APIM in both the MLM and SEM framework. We will discuss each of these frameworks separately.

3.4.1 The MLM framework

The MLM framework allows to analyze longitudinal individual data, as well as grouped data. Due to its hierarchical nature, researchers can introduce random effects to capture the variation at the higher levels. Researchers favoring the MLM framework often mistakenly considered longitudinal dyadic data as a three-level nested design: time nested within person and person nested within dyad, because it contains three factors that define the data structure (dyad, person and time). However, using this nested design (a) assumes random variability present at the second level, while the model is saturated at this level once the role of the dyad member is included (Diggle, Heagerty, Liang, & Zeger, 2002), (b) constrains the correlation between the two partners to be positive as it is estimated as the variance of a random effect, and (c) assumes the correlation between the two members of a dyad at each time point to be zero, precluding the examination of time-specific sources on (dis)similarities (Kenny & Kashy, 2011). Therefore, researchers should consider this type of data design as two-level where time and person are not nested, but crossed: the time point is the same for the two persons at each point of measurement (Laurenceau & Bolger, 2011). Then, the inclusion of a random intercept for each dyad member allows for stable sources of non-independence. As pointed out by Bolger and Shrout (2007), time-varying sources of non-independence are accounted for by allowing the residuals to correlate by the covariance matrix in expression (3.5).

The implementation of model (3.3) with covariances (3.4) and (3.5) has been discussed by Bolger and Laurenceau (2013) in the MLM framework, albeit in the absence of partner effects. Most MLM software such as HLM (Raudenbush, 2004), SAS proc mixed (SAS Institute, 2008) or R-packages like `nlme` (Pinheiro & Bates, 2000) and `lme4` (Bates, Mächler, Bolker, & Walker, 2015), easily allow for random effects with covariance structure (3.4). The main issue, however, is the residual covariance structure (3.5). As already mentioned by Kenny et al. (2006), most MLM software does not support this particular structure, except for SAS (SAS Institute, 2008). In SAS, one can define a repeated measurement statement in the proc mixed procedure that allows a Kronecker Product of some standard covariance structures, including structure (3.5), which is referred to as the “UN@AR(1)” structure within SAS. The letters before the “@” correspond to the first matrix of the Kronecker Product, in this case, an unstructured covariance matrix (“UN”, i.e., different residual variances for males and females and a non-zero residual covariance between males and females). The letters after the “@” correspond to the second matrix of the Kronecker Product. In this case it represents the first-order autoregressive covariance structure that defines the correlation between the different time points (“AR(1)”). Unfortunately, SAS only provides a limited amount of covariance structures based on the Kronecker Products. In R-packages such as `nlme` and `lme4` (Bates et al., 2015; Pinheiro & Bates, 2000), the possible covariance structures are limited to standard structures, excluding any structures based on Kronecker Products. These packages do not offer proper alternatives for covariance structure like (3.5) as they either ignore the correlation between the two dyad members at any time point, or stack the scores of the two dyad members and treat them as consecutive measures of the same entity.

3.4.2 The SEM framework

The implementation of the L-APIM within the SEM framework has hardly been discussed. Although Bolger and Shrout (2007) provide some hints and present a reasonable approximation for the construction of the residual covariance structure (3.5), they limited their exposition to estimating dyadic covariances, excluding directed actor and partner effects extant in the L-APIM. We will show that such approximation can be avoided and that the residual covariance structure can now be implemented by the researcher in any SEM software, may it be Mplus (Muthén & Muthén, 2012), LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 2004), or R-packages like `lavaan` (Rosseel, 2012) or OpenMx (Boker et al., 2011). Being able to fit the L-APIM within the SEM framework will offer much more flexibility. As opposed to the MLM-implementation, the SEM-implementation can easily be adapted to allow for indistinguishability (by putting equality constraints), latent variables, etc.

Below, we will discuss the implementation using the R-package `lavaan`, but other packages would have been possible too. The two main challenges to implement the proposed L-APIM in the SEM software are:

1. the combination of the covariance structure of the random effects and the residual covariance structure,
2. the specification of the constraints for the complex “UN@AR(1)” structure in this setting.

For the first, it is important whether the data are stored in a long or a wide format. For multilevel SEM (MSEM), data are typically in a long format (Bovaird, 2012). Current implementations of MSEM typically assume the residuals to be independent from each other and do not allow any heterogeneity (Mehta & Neale, 2005; Muthén & Asparouhov, 2011). Hence, we will limit our discussion to the standard SEM framework, which assumes data in the wide format. This implies that data are arranged as separate records for each couple, preserving multiple columns for each variable. The random intercepts can then be specified as latent factors with factor loadings equal to one for all its manifest variables. Furthermore, using the wide format, one can easily specify which residual errors should be correlated at which time point and place constraints if required.

This brings us to the second challenge, the specification of the covariance structure defined by (3.5). A first naive approach would simply code all constraints. In essence, the Kronecker Product from equation (3.5) can be viewed as one big matrix, depicted in Figure 3.4.1 (for three repetitions). Corresponding constraints are then defined with respect to each matrix element. More specifically, the covariance structure (3.5) could be defined as follows:

$$\mathrm{Cov}\begin{pmatrix} \varepsilon_{F1j} \\ \varepsilon_{F2j} \\ \vdots \\ \varepsilon_{FTj} \\ \varepsilon_{M1j} \\ \varepsilon_{M2j} \\ \vdots \\ \varepsilon_{MTj} \end{pmatrix} = \begin{pmatrix} \sigma_F^2 & A & \cdots & C & \sigma_{FM} & D & \cdots & F \\ A & \sigma_F^2 & \cdots & B & D & \sigma_{FM} & \cdots & E \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C & B & \cdots & \sigma_F^2 & F & E & \cdots & \sigma_{FM} \\ \sigma_{FM} & D & \cdots & F & \sigma_M^2 & G & \cdots & I \\ D & \sigma_{FM} & \cdots & E & G & \sigma_M^2 & \cdots & H \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ F & E & \cdots & \sigma_{FM} & H & G & \cdots & \sigma_M^2 \end{pmatrix}. \tag{3.6}$$

In expression (3.6), the autocorrelation parameter ρ corresponds to $\frac{A}{\sigma_F^2}$, $\frac{G}{\sigma_M^2}$ and $\frac{D}{\sigma_{FM}}$. Suppose we set $\rho = \frac{A}{\sigma_F^2}$, then the remaining constraints in function of A correspond to

$$\begin{aligned} &C = \left(\tfrac{A}{\sigma_F^2}\right)^{T-1} \sigma_F^2, \quad && B = \left(\tfrac{A}{\sigma_F^2}\right)^{T-2} \sigma_F^2, \quad && \ldots, \\ &F = \left(\tfrac{A}{\sigma_F^2}\right)^{T-1} \sigma_{FM}, \quad && E = \left(\tfrac{A}{\sigma_F^2}\right)^{T-2} \sigma_{FM}, \quad && \ldots, \quad D = \left(\tfrac{A}{\sigma_F^2}\right) \sigma_{FM}, \\ &I = \left(\tfrac{A}{\sigma_F^2}\right)^{T-1} \sigma_M^2, \quad && H = \left(\tfrac{A}{\sigma_F^2}\right)^{T-2} \sigma_M^2, \quad && \ldots, \quad G = \left(\tfrac{A}{\sigma_F^2}\right) \sigma_M^2. \end{aligned} \tag{3.7}$$

The amount of constraints becomes, however, quite excessive with increasing number of time points. As a consequence, model fits may not converge. Furthermore, even in case of convergence, the computational time increases exponentially with increasing number of time points. Using equality constraints like (3.7), the CPU time to fit the L-APIM with five to ten time points easily takes from two minutes to an hour.

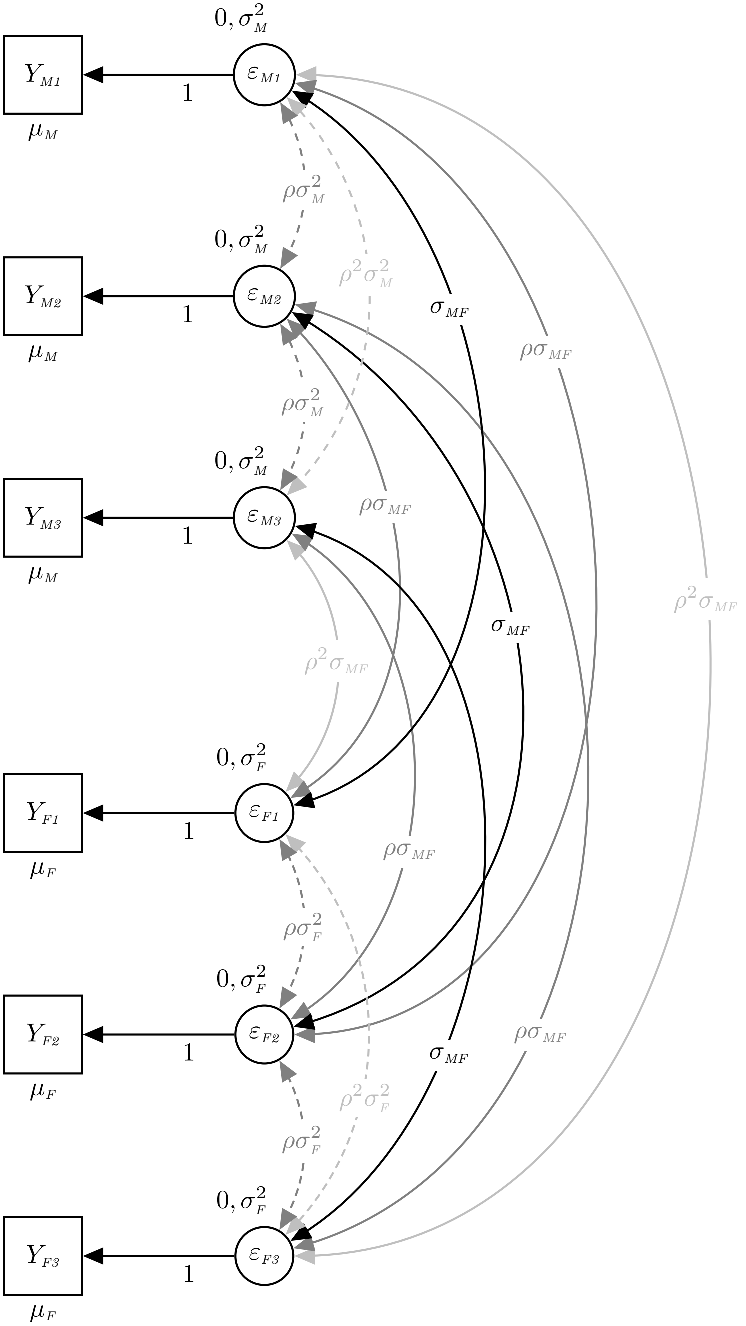


Figure 3.4.1 The standard representation of the residual UN@AR(1) covariance structure for LDD with three repeated measurements.

However, an alternative approach was suggested by Bolger and Shrout (2007), which is less demanding, and is inspired by the Trait-State-Error model (Kenny & Zautra, 2001). The latter model decomposes the relationship between the measured outcome scores of a person into a stable trait (ST), an autoregressive trait (ART), and a state (S). In our setting, the stable trait would correspond to the random intercept of a dyad member and the autoregressive trait would correspond to the first-order autoregressive process between the residuals. In case of a first-order autoregressive process, this makes perfectly sense since the process can be expressed by a moving average process (Beck, 1992; Hamilton, 1994). In Figure 3.4.2, one can find the path diagram of this alternative implementation (for three repetitions). First, a "latent" variable Z_{kij} is specified for every outcome measurement Y_{kij} for each time point i, dyad j, and dyad member k (k = M or F). Next, the original covariance structure is "neutralized", meaning that the original errors ε_{kij} have a mean and a variance of zero. The first-order autoregressive process is then reconstructed with the aid of these latent variables. More specifically, a simple linear regression model is assumed between the latent variables for each of the dyad members:

$$\begin{aligned} Z_{kij} &= \rho Z_{k(i-1)j} + \upsilon_{kij} \\ &= \sum_{r=0}^{\infty} \rho^r \ \upsilon_{k(i-s)j}, \end{aligned} \tag{3.8}$$

with i referring to the time point ($i = 2, \ldots, T$), j to the dyad number ($j = 1, \ldots, N$), and k to the dyad member (k = M or F). The parameters υ_{kij} are independent white noise, that is, $\upsilon_{kij} \sim N(0, \nu_{ki}^2)$. In equation (3.8), the latent variable at the first time point for each dyad member is considered an exogenous variable. As a result, $Var(Z_{k1j}) = Var(\upsilon_{k1j}) = \nu_{k1}^2 =: \sigma_k^2$, while the variances of all other latent variables are defined by:

$$Var(Z_{kij}) = \rho^2 \ Var(Z_{k(i-1)j}) + Var(\upsilon_{kij}),$$

with $i = 2, \ldots, T$, $j = 1, \ldots, N$ and k = M or F. So, for $i = 2$, we have that $Var(Z_{k2j}) = (1-\rho^2)\sigma_k^2$. In order to have a constant residual variance over time, the following constraint is added:

$$\begin{aligned} \nu_{kij}^2 &= \nu_{k1}^2 \left(1 - \rho^2\right) \\ &= \sigma_k^2 \left(1 - \rho^2\right). \end{aligned} \tag{3.9}$$

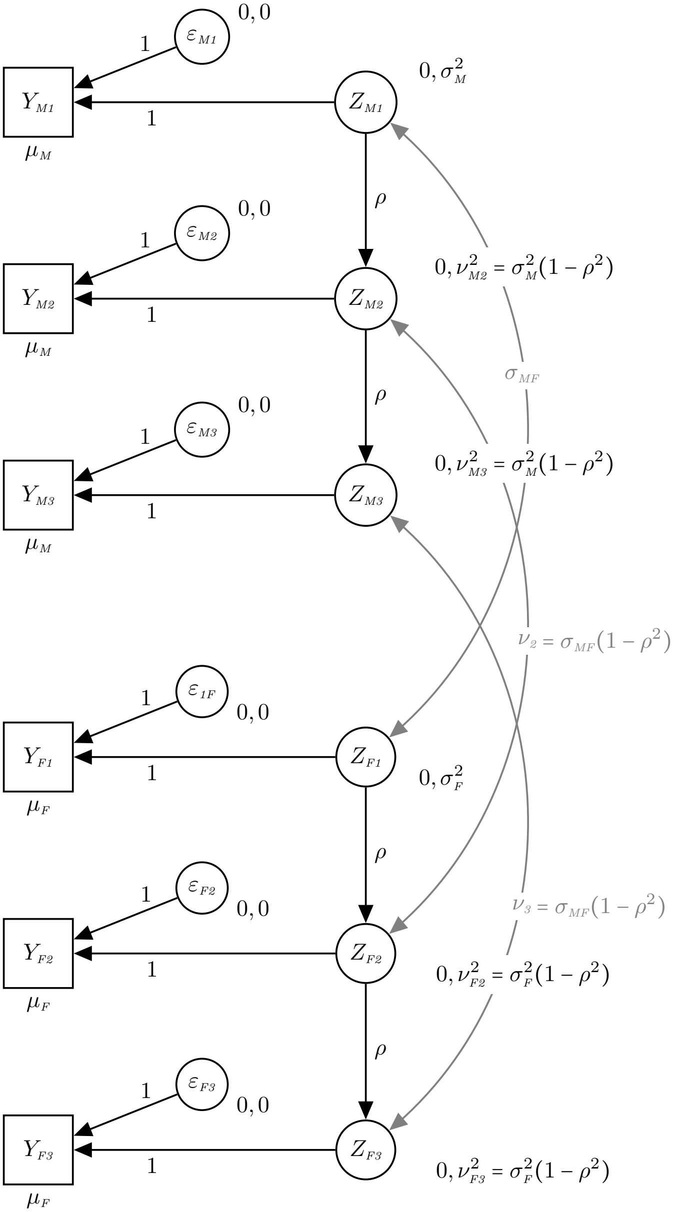


Figure 3.4.2 The alternative representation of the residual UN@AR(1) covariance structure for LDD with three repeated measurements. Note, the residuals ε are inherent to the model implementation and cannot be removed. Hence, we eliminated them by fixing their means and variances to zero.

This way, we end up with a first-order autoregressive correlation structure within each dyad member:

$$\begin{aligned}
Cov(Z_{kij}, Z_{ki^\star j}) &= Cov(\sum_{r=0}^{\infty} \rho^r \ \upsilon_{k(i-r)j}, \sum_{s=0}^{\infty} \rho^s \ \upsilon_{k(i^\star-s)j}) \\
&= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \rho^r \ \rho^s \ Cov(\upsilon_{k(i-r)j}, \upsilon_{k(i^\star-s)j}) \\
&= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \rho^r \ \rho^s \ \nu_{kij}^2 \qquad \text{for } s = r + |i - i^\star| \\
&= \rho^{|i-i^\star|} \ \nu_{kij}^2 \ \sum_{r=0}^{\infty} \rho^{2r} \\
&= \rho^{|i-i^\star|} \ \left(\frac{\nu_{kij}^2}{1-\rho^2} \right) \\
&= \rho^{|i-i^\star|} \ \sigma_k^2,
\end{aligned} \tag{3.10}$$

for i and $i^\star = 2, \ldots, T$. So far, we have only recreated a first-order autoregressive process within each dyad member. Hence, we still need to establish a correlation structure between the dyad members. To this end, we allow the disturbance terms υ_{kij} to correlate at each time point for each dyad: $Cov(\upsilon_{kij}, \upsilon_{k^\star ij}) = \nu_i$, for all $j = 1, \ldots, N$. Due to the fact that the latent variables at the first time point for each dyad member are exogenous, we have that $Cov(Z_{k1j}, Z_{k^\star 1j}) = Cov(\upsilon_{k1j}, \upsilon_{k^\star 1j}) = \nu_1 =: \sigma_{MF}$, and similar as before:

$$\begin{aligned}
Cov(Z_{kij}, Z_{k^\star i^\star j}) &= \rho^{|i-i^\star|} \ \left(\frac{\nu_{ij}}{1-\rho^2} \right) \\
&= \rho^{|i-i^\star|} \ \sigma_{MF},
\end{aligned}$$

where the latter equality holds if one constrains $\nu_{ij} = \sigma_{MF}(1-\rho^2)$. Note that these new latent variables are not allowed to correlate with the latent intercepts η_F and η_M. The residual covariance of the newly defined latent variables now is:

$$\mathrm{Cov}\begin{pmatrix} \upsilon_{F1j} \\ \upsilon_{F2j} \\ \vdots \\ \upsilon_{FTj} \\ \upsilon_{M1j} \\ \upsilon_{M2j} \\ \vdots \\ \upsilon_{MTj} \end{pmatrix} = \begin{pmatrix} \nu_{F1}^2 & 0 & \cdots & 0 & \nu_{FM1} & 0 & \cdots & 0 \\ 0 & \nu_{F2}^2 & \cdots & 0 & 0 & \nu_{FM2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \nu_{FT}^2 & 0 & 0 & \cdots & \nu_{FMT} \\ \nu_{FM1} & 0 & \cdots & 0 & \nu_{M1}^2 & 0 & \cdots & 0 \\ 0 & \nu_{FM2} & \cdots & 0 & 0 & \nu_{M2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \nu_{FMT} & 0 & 0 & \cdots & \nu_{MT}^2 \end{pmatrix},$$

with $\nu_{F1}^2 =: \sigma_F^2$, $\nu_{M1}^2 =: \sigma_M^2$ and $\nu_{FM1} =: \sigma_{FM}$, and the additional constraints

$$\begin{array}{lll} \nu_{F2}^2 = \sigma_F^2\left(1-\rho^2\right), & \ldots, & \nu_{FT}^2 = \sigma_F^2\left(1-\rho^2\right), \\ \nu_{FM2} = \sigma_{FM}\left(1-\rho^2\right), & \ldots, & \nu_{FMT} = \sigma_{FM}\left(1-\rho^2\right), \\ \nu_{M2}^2 = \sigma_M^2\left(1-\rho^2\right), & \ldots, & \nu_{MT}^2 = \sigma_M^2\left(1-\rho^2\right). \end{array} \tag{3.11}$$

It is clear that this alternative approach requires less constraints and is less computationally demanding than the crude coding of the original covariance structure. Note that the constraints (3.11) can readily be specified in `lavaan`, unlike EQS where these constraints are fixed at approximated values obtained by iterative estimation of the autocorrelation and the latent variances (Bolger & Shrout, 2007).

Using the wide format from the SEM framework has also some consequences for the specification of the mean structure. In the wide format, the mean structure has to be defined for each time point separately, unlike the single-equation approach in the long format. Furthermore, the wide data format also impacts the design matrix **X**. Indeed, the suggested split-up of overtime predictors into a time-specific and time-averaged component introduces some superfluity. More specifically, if one separates an overtime predictor into these two component, the sum of the deviations that defines the time-specific component is zero:

$$\sum_{i=1}^{T}\left(X_{ij} - X_{.j}\right) = 0,$$

and hence, time-specific dyad and member predictors are not independent, making the design matrix non-invertible. This caveat needs to be taken into account in the parametrization as well. This is obtained by replacing the effect of the time-specific component at the last time point by the sum of (opposite) effects of the remaining time points. For instance, the effect $a_{M(XS)}$ of $(X_{MTj} - X_{M.j})$ on Y_{MTj} at the latter time point T is implemented as

$$-\sum_{i=1}^{T-1} a_{M(XS)}(X_{Mij} - X_{M.j}).$$

3.5 Empirical example using an online application

In this section, we will fit the L-APIM on the empirical example data originating from the study of Dewitte et al. (2015). This study investigated 66 heterosexual couples in Flanders on a daily basis over a period of three weeks. Using diary data, men and women reported each morning their sexual and intimate behavior over the past 24 hours. In the evening, they were asked about their individual, relational and partner-related feelings and behavior of the past day. For this paper we will solely focus on the effect of one's own and one's partner reported positive relationship feelings to the next day's perceived intimacy. The intimate acts with their partner were described as the amount of kissing, cuddling and caressing on a 7-point scale from 'not at all' to 'very frequent'. The positive relational feelings were measured as an average of nine items on a 7-point scale: the extent to which they felt happy, satisfied, understood, supported, accepted, loved, in love, connected, and close. We consider the following L-APIM:

$$\begin{cases} Intim_{Fij} = (\mu_F + \eta_{Fj}) + a_{F(PosRelA)}PosRelA_{Fj} \\ \qquad + a_{F(PosRelS)}PosRelS_{Fij} + p_{MF(PosRelA)}PosRelA_{Mj} \\ \qquad + p_{MF(PosRelS)}PosRelS_{Mij} + \varepsilon_{Fij} \\ Intim_{Mij} = (\mu_M + \eta_{Mj}) + a_{M(PosRelA)}PosRelA_{Mj} \\ \qquad + a_{M(PosRelS)}PosRelS_{Mij} + p_{FM(PosRelA)}PosRelA_{Fj} \\ \qquad + p_{FM(PosRelS)}PosRelS_{Fij} + \varepsilon_{Mij} \end{cases}, \quad (3.12)$$

with i referring to the time point $(i = 1, \ldots, T)$, j to the dyad number $(j = 1, \ldots, N)$. The subscripts A and S in the parameters refer to the time-averaged and time-specific effect of the predictor, respectively. As mentioned before, these effects are obtained by constructing corresponding time-specific and time-averaged components of the original variables. We further assume the random effect covariance structure (3.4) and residual covariance structure (3.5).

The research questions that we want to address are:

Q1 how do the average positive relational feelings of a person affect the perceived intimacy of that same person (i.e., $a_{M(PosRelA)}$ and $a_{F(PosRelA)}$ from model (3.12)),

Q2 how do the average positive relational feelings of one's partner affect one's own perception of intimacy (i.e., $p_{FM(PosRelA)}$ and $p_{MF(PosRelA)}$ from model (3.12)),

Q3 how does an increase in positive relational feelings of a person on a particular day (as compared to his/her average positive relational feelings) affect the perceived intimacy of that person (i.e., $a_{M(PosRelS)}$ and $a_{F(PosRelS)}$ from model (3.12)),

Q4 how does an increase in positive relational feelings of one's partner on a particular day (as compared to his/her partner's average positive relational feelings) affect one's own perceived intimacy (i.e., $p_{FM(PosRelS)}$ and $p_{MF(PosRelS)}$ from model (3.12)),

Q5 to what extent does the time-averaged perceived intimacy covary between partners within a dyad (i.e., τ_{MF} from expression (3.4)),

Q6 to what extent is the perceived intimacy of one dyad member related to the perceived intimacy of the partner on a particular day (i.e., σ_{MF} from expression (3.5)),

Q7 how strong is the association between the perceived intimacy on one day with the amount of perceived intimacy on the next day within a given person (i.e., ρ from expression (3.5)).

Several user-friendly tools are already developed for dyadic researcher to fit the cross-sectional APIM (Kenny, 2017; Stas, Kenny, Mayer, & Loeys, 2018). Considering that only SAS can straightforwardly fit the L-APIM within the MLM framework and that the implementation within the SEM framework is possible, but requires tedious coding, we opted to construct an online application using the R-package `shiny` (Chang et al., 2017; RStudio Inc., 2017), that automatically implements the L-APIM within `lavaan`. The application is a user-friendly and free tool that can be found at "`http://fgisteli.shinyapps.io/Shiny_LDD2/`". In this section, we will illustrate the app by fitting an L-APIM on the data of the empirical example, hereby answering the questions Q1-Q7. At the "info" page of the application, one can find a tutorial with extra information and screen shots of each step of the model specification that is required to reproduce the results described below.

Once the application is started, one can upload the data file in the application. Although, the app uses a wide data format for analysis, both long and wide data format are allowed as input. In case of a long data format, the app will transform the data set before fitting the L-APIM. The time range needs to be specified, as well as the indices in the variable names that correspond to the role of the dyad member. In case of indistinguishable dyads, these indices are arbitrary and only necessary for computational reasons. Note that the application assumes equally spaced time points as this restriction is assumed by the autoregressive process. The researcher is allowed to reduce the time range into a range of interest, yet a minimum of three time points is required for the model to be identified.

Next, the mean structure of the model needs to be specified. The application accepts four types of predictors: time-constant-dyad, overtime-dyad, time-constant-member and overtime-member predictors. Overtime predictors are not supposed to be made time-specific or time-averaged in advance. The user of the app is also offered the possibility to relabel the variable names, as well as to grand-mean center the predictors. In our example, the only predictor is positive relational feelings "*PosRel*" and corresponds to an overtime-member variable. We opt for grand-mean centering this variable as well as for the split-up into a time-specific and time-averaged component. The corresponding components are automatically labeled as the variable name with an "S", respectively an "A", at the end of the name (i.e., *PosRelS* and *PosRelA*, respectively). To complete the specification of the mean structure of the model, one needs to report which parameters depend on the role of the dyad member. In case of complete indistinguishable dyads, all model parameters are assumed to be the same for both dyad members. However, in our illustration, we assume all mean structure parameters to be different for men and women. The labels of the model parameters are therefore extended with the dyad member index at the end (e.g., *PosRelSM* refers to the time-specific effect of *PosRel* of the males on the perceived intimacy). The resulting model can be described as in equation (3.12).

Lastly, the covariance structures need to be specified in the app. First, the random effects covariance structure must be defined. As suggested by expression (3.12), the application assumes the intercept to be a mixed effect consisting of a random and fixed part. The user is asked to define the covariance structure of the random intercepts. Since we expect the variation of the overall mean score of perceived intimacy to be different for males and females, and to be correlated, we will assume an unstructured covariance structure. The application also allows a compound symmetry (i.e., same variances for males and females), a heterogeneous variance component (i.e., different variances for males and females, but no correlation) or a variance component (i.e., same variances for males and females and no correlation) structure. Second, the residual covariance structure needs to be defined. As mentioned before, the structure consists of a within-dyad and time-specific within-member part given by the Kronecker product of two matrices in equation (3.5). We will assume an unstructured 2-by-2 covariance structure for the within-dyad part. For the within-dyad part of the covariance structure, the user can again opt for an unstructured, a compound symmetry, a heterogeneous variance component and a variance component covariance structure. In case of indistinguishable dyads, the compound symmetry and variance component covariance structure are plausible candidates. For the time-varying part of the covariance structure, we will assume a first-order autoregressive process. The application also allows a more simple structure in which the residual temporal correlation is assumed to be zero (i.e., an identity matrix).

The (truncated) R-syntax that is generated by the online application can be found in Appendix 3.A. For the interested reader, the corresponding SAS-syntax has been provided in Appendix 3.B. The results of the model fit using the application can be found in Table 3.5.1 and 3.5.2. The results obtained by SAS can be found in Appendix 3.B. Given that the predictor was grand-mean centered and separated in a time-specific and time-averaged component, the intercepts 3.98 and 4.22 reflect the average perceived intimacy over all days for males and females, respectively, at average positive relational feelings. Men and women who report more positive feelings on average (i.e., the effect of $PosRelAM$ on the male outcome and the effect of $PosRelAF$ on the female outcome), also report more perceived intimacy (cfr. Q1). Furthermore, men whose wives report more positive relationship feelings (i.e., the effect of $PosRelAF$ on the male outcome), report more intimacy (cfr. Q2). Both within men and women, an increase in the positive relationship feelings on a specific day (i.e., the effect of $PosRelSM$ on the male outcome and the effect of $PosRelSF$ on the female outcome) is associated with higher perceived intimacy on the next day (cfr. Q3). Moreover, an increase in the positive relationship feelings of the men on a specific day (i.e., the effect of $PosRelSM$ on the female outcome) is associated with higher perceived intimacy of the women the next day (cfr. Q4). A similar time-specific partner effect of the women on their male partner was not found to be significant.

Table 3.5.2 reports the parameter estimates of the covariance structure of the L-APIM. The variation for the average perceived intimacy is 0.46 and 0.35 for males and females, respectively. This implies that 95% of the average perceived intimacy (at average levels of the positive relational feelings) lies between 2.62 and 5.33 for males and between 3.04 and 5.40 for females. The correlation of the average perceived intimacy between males and females is $\frac{0.22}{\sqrt{0.46*0.35}} = 0.54$ (cfr. Q5). The residual variance within a person of the perceived intimacy on a particular day is similar for males and females: 2.62 for males as opposed to 2.47 for females. The association of the daily fluctuations between males and females is stronger than the time-stable correlation, with a magnitude of $\frac{1.72}{\sqrt{2.62*2.47}} = 0.68$. Hence, when the man perceives a lot of intimacy on a specific day, the woman also tends to perceive a lot of intimacy that day, and vice versa (cfr. Q6). It is also interesting to note that same-day correlation between males and females is highly attributable to daily influences. About 89% of the same-day correlation with a magnitude of $\frac{1.72+0.22}{\sqrt{(0.46+2.62)(0.35+2.47)}} = 0.66$ is due to shared variance at day-level, in contrast to the remaining 11% due to shared variance at dyad-level. Lastly, the autocorrelation parameter shows that the amount of intimacy on one day is positively correlated with the amount of intimacy on the next day (cfr. Q7). The magnitude of this correlation is 0.07, which is rather small, but still significant.

Note that in this illustration, we only considered member-predictors. In the online documentation of the app, we also provide a tutorial example that additionally includes a dyad predictor.

Table 3.5.1 Parameter estimates of the fixed effects for the L-APIM for the effect of positive relationship feelings on intimacy.

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Effect on male outcome						
Intercept	3.975	0.100	39.786	<.001	3.779	4.171
PosRelAM	0.547	0.160	3.410	.001	0.233	0.861
PosRelAF	0.339	0.137	2.468	.014	0.070	0.607
PosRelSM	0.295	0.073	4.033	<.001	0.152	0.438
PosRelSF	0.091	0.073	1.249	.212	-0.052	0.233
Effect on female outcome						
Intercept	4.215	0.089	47.511	<.001	4.041	4.388
PosRelAF	0.619	0.122	5.066	<.001	0.379	0.858
PosRelAM	0.118	0.143	0.823	.411	-0.163	0.399
PosRelSF	0.178	0.069	2.582	.010	0.043	0.312
PosRelSM	0.266	0.071	3.747	<.001	0.127	0.405

Table 3.5.2 Estimates of the covariance parameters for the L-APIM for the effect of positive relationship feelings on intimacy.

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Between-dyad						
τ_M^2	0.457	0.111	4.120	<.001	0.240	0.674
τ_F^2	0.346	0.090	3.831	<.001	0.169	0.522
τ_{MF}	0.215	0.081	2.638	.008	0.055	0.374
Within-dyad						
σ_M^2	2.620	0.117	22.364	<.001	2.390	2.849
σ_F^2	2.467	0.107	23.076	<.001	2.257	2.676
σ_{MF}	1.723	0.096	17.965	<.001	1.535	1.911
ρ	0.068	0.025	2.677	.007	0.018	0.117

3.6 Discussion

In this paper, we presented an extension of the cross-sectional APIM: the L-APIM. This model incorporates the four main issues for over-time dyadic data: (a) it includes the effect of partner characteristics on the outcome of a person, (b) it allows for non-independence between the outcomes of two members of a dyad, (c) it takes the non-independence between the different time points within and between the two members of a dyad into account, and (d) it allows to separate the time-varying predictors into a time-specific and time-averaged effect. Its implementation in MLM software such as SAS was already available. Here, we have shown how the model can be fitted within the SEM framework as well.

Fitting the L-APIM within the SEM framework offers several advantages compared to the MLM framework. First, one can incorporate measurement error by considering latent outcome variables in the L-APIM rather than the manifest ones. This is already possible within the cross-sectional setting (Gareau, Fitzpatrick, Gaudreau, & Lafontaine, 2016; Orth, 2013), and an extension toward the longitudinal setting could now be similarly implemented. Second, SEM also allows a latent decomposition of the overtime predictors rather than the manifest approach introduced in this paper. Instead of using the observed means as a proxy for the true underlying mean, one could use a latent mean to represent the time-averaged component of the predictor (Preacher, Zyphur, & Zhang, 2010). In this case, a latent variable should be created for the overtime predictor with all factor loadings equal to one. The effect of the latent variable on the outcome variable would then correspond to the time-averaged effect of the overtime-predictor, while the effect of the residual overtime-predictor would correspond to the time-specific effect of the overtime-predictor. Preacher and Kelley (2011) showed the latter approach is preferable in case of measurement error at the member level, especially in terms of bias for the between-effects. Third, differences in the data structure may imply different treatment of missing data. Within the MLM framework, one typically applies list-wise deletion. This results for example in the exclusion of all records of a person at a particular time as soon as a missing predictor value for either the actor or partner is reported. In the wide data structure, such list-wise deletion would imply the elimination of a complete dyad. However, relying on the full information maximum likelihood (FIML), the SEM framework is able to use all available data (Ledermann & Kenny, 2017). Fourth, tests of indistinguishability are more easily performed within SEM than within MLM. Using these tests, one can investigate to what degree the two dyad members differ by role or not (Gistelinck et al., 2018). Lastly, due to the flexibility of SEM, one can easily relax some of the assumptions imposed by the L-APIM presented in this paper. For instance, variances within dyads can be allowed to vary over time, as well as the autocorrelation parameter and the mean structure parameters.

The application introduced in this paper does not yet incorporate latent outcomes, latent decomposition (Asparouhov & Muthén, 2018) or indistinguishability tests. These extensions are part of future upgrades of the app. Moreover, even though the application only assumes main effects for all predictors, interaction parameters can in principle be included into the L-APIM by including the product of two variables as a new variable. Furthermore, random slopes are not yet allowed into the application.

Lastly, the application does not allow for any mediation analysis yet. One way to incorporate mediation hypotheses is to replace the current APIM's at each time point by its extended APIMeM versions (Ledermann, Macho, & Kenny, 2011). Note, in order to maintain the separation between intra- and inter-individual variation, one needs to split these mediators into time-specific and time-averaged components as well.

Further note that in this paper, we opted for an autoregressive relationship between the residuals in order to incorporate the nonindependence in the data. Yet, one may argue to include an autoregressive relationship between the dependent variables instead. Indeed, a model with first-order autoregressive disturbances and a model with lagged dependent variables are similar, but only under severe restrictions. Hamaker (2005) showed both models to be equivalent in the absence of any explanatory variables and in case the autoregressive parameters are invariant over time. Using lagged dependent variables (of the actor and the partner) to eliminate any residual correlation in the presence of predictors might be a dangerous strategy as it may induce coefficient estimates of the predictor effects to be biased (Achen, 2001; Keele & Kelly, 2005). Moreover, the interpretation of the mean structure parameters changes drastically as well (Jongerling & Hamaker, 2011). Hamaker, Kuiper, and Grasman (2015) also showed that random effects within these models are necessarily correlated with the first lagged outcome variable in order to obtain unbiased parameter estimates. This correlation is easy to specify within the SEM framework, but not in the MLM framework which assumes all predictors to be independent of the random effects.

Throughout the paper, we only considered continuous outcome variables, yet longitudinal adaptations of the APIM with categorical outcomes could be considered as well (Loeys & Molenberghs, 2013; Spain, Jackson, & Edmonds, 2012). Loeys, Cook, De Smet, Wietzker, and Buysse (2014) suggested using the Generalized Estimation Equations (GEE) framework to fit categorical outcomes, but it could be interesting to compare its performance with a SEM approach when fitting longitudinal dyadic data. It has been shown that SEM relying on diagonally weighted leased squares outperforms likelihood based approximations used for Generalized Linear Mixed Models (GLMM) in terms of bias in dyadic data (Josephy, Loeys, & Rosseel, 2016), but no extensions to the longitudinal setting have been investigated yet.

In sum, the L-APIM in the SEM framework provides a rich class of models for analyzing longitudinal dyadic data that allows to explore a wide variety of research questions. We hope that the online application will help applied researchers to fit the model in practice.

References

Achen, C. (2001). *Why lagged dependent variables can suppress the explanatory power of other independent variables.* Retrieved from `https://www.princeton.edu/csdp/events/Achen121201/achen.pdf` (Presented at the Annual Meeting of the Political Methodology, Los Angeles)

Ackerman, R., Donnellan, M., Kashy, D., & Conger, R. (2012). Dyadic data analyses in a developmental context. In B. Laursen, T. Little, & N. Card (Eds.), *Handbook of developmental research methods* (p. 537-556). New York, NY: The Guilford Press.

Asparouhov, T., & Muthén, B. (2018). *Latent variable centering of predictors and mediators in multilevel and time-series models* (Tech. Rep.). UK Education and Research communities.

Atkins, D. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology*, *19*, 98-110. doi: 10.1037/0893-3200.19.1.98

Baird, R., & Maxwell, S. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *12*(2), 175-188. doi: 10.1037/met0000070

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beck, N. (1992). Comparing dynamic specifications: The case of presidential approval. *Political Analysis*, *3*, 27-50. doi: 10.1093/pan/3.1.51

Bentler, P. (2004). *EQS 6 structural equations program manual* (6th ed.). Encino, CA: Multivariate Software, Inc.

Bock, R. (1997). The nominal categories model. In W. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (p. 33-49). New York, NY: Springer.

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

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: The Guildford Press.

Bolger, N., & Shrout, P. (2007). Accounting for statistical dependencies in longitudinal data on dyads. In T. Little, J. Bovaird, & N. Card

(Eds.), *Modeling contextual effects in longitudinal studies* (p. 285-298). Lawrence Erlbaum Associates.

Bovaird, J. (2012). Multilevel structural equation models for contextual factors. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling contextual effects in longitudinal studies* (p. 149-182). Lawrence Erlbaum Associates.

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). Shiny: Web application framework for R [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=shiny`

Cook, W., & Kenny, D. (2004). *A critique of contemporary self-report assessments of family functioning.* (Unpublished manuscript, Maine Medical Center, Portland, ME.)

Cook, W., & Kenny, D. (2005). The actor-partner interdependence model: A model of bidirectional effects in developmental studies. *International Jounal of Behavioral Development*, *29*(2), 101-109. doi: 10.1080/01650250444000405

Cranford, J., Shrout, P., Iida, M., Rafaeli, E., Ypi, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, *32*(7), 917-929. doi: 10.1177/0146167206287721

Curran, P. (2003). Have multilevel models been structural equations models all along? *Multivariate Behavioral Research*, *38*(4), 529-569. doi: 10.1207/s15327906mbr3804_5

Dewitte, M., Van Lankveld, J., Vandenberghe, S., & Loeys, T. (2015). Sex in its daily relational context. *Journal of sexual medicine*, *12*(12), 2436-2450. doi: 10.1111/jsm.13050

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2002). *Analysis of longitudinal data.* Oxford, UK: Oxford University Press.

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121

Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Gareau, A., Fitzpatrick, J., Gaudreau, P., & Lafontaine, M.-F. (2016). Analysing, interpreting, and testing the invariance of the actor-partner interdependence model. *The Quantitative Methods for Psychology*, *12*(2), 101-113. doi: 10.20982/tqmp.12.2.p101

Gistelinck, F., Loeys, T., Decuyper, M., & Dewitte, M. (2018). Indistinguishability tests in the actor-partner interdependence model. *British Journal of Mathematical and Statistical Psychology*, *71*(3), 472-498. doi: 10.1111/bmsp.12129

Gonzalez, R., & Griffin, D. (1999). The correlation analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*(4), 449-469. doi: 10.1037/a0025651

Hamaker, E. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*(3), 404-416. doi: 10.1177/0049124104270220

Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. doi: 10.1037/a0038889

Hamilton, J. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). London, United Kingdom: Routledge.

Jongerling, J., & Hamaker, E. (2011). On the trajectories of the predetermined ALT model: What are we really modeling? *Structural Equation Modeling*, *18*(3), 370-382. doi: 0.1080/10705511.2011.582004

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide* (2nd ed.). Lincolnwood, IL: Scientific Software International, Inc.

Josephy, H., Loeys, T., & Rosseel, Y. (2016). A review of R-packages for random-intercept probit regression in small clusters. *Frontiers in Applied Mathematics and Statistics*, *2*(18), 1-13. doi: 10.3389/fams.2016.00018

Kashy, D., & Kenny, D. (2000). The analysis of data from dyads and groups. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personlaity psychology* (p. 154-477). Cambridge University Press.

Keele, L., & Kelly, N. (2005). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political analysis*, *14*(2), 168-205. doi: 10.1093/pan/mpj006

Kenny, D. (2017, January 5). *DyadR: Web programs for dyadic data analysis.* Retrieved from `http://davidakenny.net/DyadR/DyadRweb.htm`

Kenny, D., & Kashy, D. (2011). Dyadic data analysis using multilevel modeling. In J. Hox & J. Roberts (Eds.), *Handbook of advanced multilevel analysis* (p. 335-370). Routledge.

Kenny, D., Kashy, D., & Cook, W. (2006). *Dyadic data analysis.* New York, NY: The Guilford Press.

Kenny, D., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (p. 243-263). Washington, DC: American Psychological Association.

Laurenceau, J.-P., & Bolger, N. (2005). Using diary methods to study marital and family processes. *Journal of Family Psychology*, *19*(1), 86-97.

Laurenceau, J.-P., & Bolger, N. (2011). Analyzing diary and intensive longitudinal data from dyads. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 407-422). New York, NY: The Guilford Press.

Ledermann, T., & Kenny, D. (2017). Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. *Journal of Family Psychology*, *31*(4), 442-452. doi: 10.1037/fam0000290

Ledermann, T., Macho, S., & Kenny, D. (2011). Assessing mediation in dyadic data using the actor-partner interdependence model. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 595-612. doi: 10.1080/10705511.2011.607099

Loeys, T., Cook, W., De Smet, O., Wietzker, A., & Buysse, A. (2014). The actor-partner interdependence model for categorical dyadic data: a user-friendly guide to GEE. *Personal relationships*, *21*(2), 225-241. doi: 10.111/pere.12028

Loeys, T., & Molenberghs, G. (2013). Modeling actor and partner effects in dyadic data when outcomes are categorical. *Psychological Methods*, *18*(2), 220-236. doi: 10.1037/a0030640

Lyons, K., & Sayer, A. (2005). Longitudinal dyad models in family research. *Journal of Marriage and Family*, *67*(4), 1048-1060. doi: 10.1111/j.1741-3737.2005.00193.x

Mehta, P., & Neale, M. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259-284. doi: 10.1037/1082-989X.10.3.259

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. Roberts (Eds.), *Handbook of advanced multilevel analysis* (p. 335-370). Routledge.

Muthén, L., & Muthén, B. (2012). *Mplus user's guide: Statistical analysis* *with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nestler, S., Grimm, K., & Schönbrodt, F. (2015). The social consequences and mechanisms of personality: How to analyse longitudinal data from individual, dyadic, round-robin and network designs. *European Journal of Personality*, *29*(2), 272–295. doi: 10.1002/per.1997

Newsom, J. (2002). A multilevel structural equation model for dyadic data. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(3), 431-447. doi: 10.1207/S15328007SEM0903_7

Olsen, J., & Kenny, D. (2005). Structural equation modeling with interchangeable dyads. *Psychological methods*, *11*(2), 127-141. doi: 10.1037/1082-989X.11.2.127

Orth, U. (2013). How large are actor and partner effects of personality on relationship satisfaction? the importance of controlling for shared method variance. *Personality and Social Psychology Bulletin*, *39*(10), 1359-1372. doi: 10.1177/0146167213492429

Peugh, P., DiLillo, D., & Panuzio, J. (2013). Analyzing mixed-dyadic data using structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(2), 314-337. doi: 10.1080/10705511.2013.769395

Pinheiro, J., & Bates, M. (2000). *Mixed-effects models in S and S-PLUS.* New York, NY: Springer.

Planalp, E., Du, H., Braungart-Rieker, J., & Wang, L. (2017). Growth curve modeling to studying change: A comparison of approaches using longitudinal dyadic data with distinguishable dyads. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(1), 129-147. doi: 10.1080/10705511.2016.1224088

Preacher, K., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological methods*, *16*(2), 93-115. doi: 10.1037/a0022658

Preacher, K., Zyphur, M., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods*, *15*(3), 209-233. doi: 10.1037/a0020141

Raudenbush, S. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Raudenbush, S., R.T., B., & Barnett, R. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology*, *9*(2), 161-174.

Reis, H., Collins, W., & Bersheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*(6), 844-872. doi: 10.1037/0033-2909.126.6.844

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. doi: 10.18637/jss.v048.i02

RStudio Inc. (2017). *Shiny: Easy web applications in R.* Retrieved from `http://shiny.rstudio.com/`

SAS Institute, I. (2008). *SAS/STAT 9.2 user's guide: The MIXED porcedure.* Cary, NC: SAS Institute Inc.

Spain, S., Jackson, J., & Edmonds, G. (2012). Extending the actor-partner interdependence model for binary outcomes: A multilevel logistic approach. *Personal Relationships*, *19*(3), 431-444. doi: 10.1111/j.1475-6811.2011.01371.x

Stas, L., Kenny, D., Mayer, A., & Loeys, T. (2018). Giving dyadic data analysis away: A user-friendly app for actor-partner interdependence models. *Personal Relationships*, *25*(1), 103-119. doi: 10.1111/pere.12230

Wendorf, C. (2002). Comparisons of structural equation modeling and hierarchical linear modeling approaches to couples' data. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(1), 126-140. doi: 10.1207/S15328007SEM0901_7

Appendix

3.A R code of the illustrative examples

Below, the (truncated) R-syntax of the L-APIM for the empirical example is displayed. This model is automatically generated by the online application.

```
model <- '
  # mean model for Y
IntimF1 ~ alphaF*1 + phi.aF.PosRelA*PosRelAF
 + phi.pMF.PosRelA*PosRelAM
 + tta.aF.PosRelS*PosRelSF1
 + tta.pMF.PosRelS*PosRelSM1
[...]
IntimF20 ~ alphaF*1 + phi.aF.PosRelA*PosRelAF
 + phi.pMF.PosRelA*PosRelAM
 + tta.aF.PosRelS*PosRelSF20
 + tta.pMF.PosRelS*PosRelSM20
IntimF21 ~ alphaF*1 + phi.aF.PosRelA*PosRelAF
 + phi.pMF.PosRelA*PosRelAM
 + tta.aF.PosRelS21*PosRelSF1
 + [...]
 + tta.aF.PosRelS21*PosRelSF20
 + tta.pMF.PosRelS21*PosRelSM1
 + [...]
 + tta.pMF.PosRelS21*PosRelSM20
IntimM1 ~ alphaM*1 + phi.aM.PosRelA*PosRelAM
 + phi.pFM.PosRelA*PosRelAF
 + tta.aM.PosRelS*PosRelSM1
 + tta.pFM.PosRelS*PosRelSF1
[...]
IntimM20 ~ alphaM*1 + phi.aM.PosRelA*PosRelAM
 + phi.pFM.PosRelA*PosRelAF
 + tta.aM.PosRelS*PosRelSM20
 + tta.pFM.PosRelS*PosRelSF20
```

```
IntimM21 ~ alphaM*1
 + phi.aM.PosRelA*PosRelAM
 + phi.pFM.PosRelA*PosRelAF
 + tta.aM.PosRelS21*PosRelSM1
 + [...]
 + tta.aM.PosRelS21*PosRelSM20
 + tta.pFM.PosRelS21*PosRelSF1
 + [...]
 + tta.pFM.PosRelS21*PosRelSF20

  # within-predictor constraints
tta.aF.PosRelS21 == -tta.aF.PosRelS
tta.aM.PosRelS21 == -tta.aM.PosRelS
tta.pFM.PosRelS21 == -tta.pFM.PosRelS
tta.pMF.PosRelS21 == -tta.pMF.PosRelS

  # covariance model for Y
IntimF1 ~~ 0*IntimF1
[...]
IntimF21 ~~ 0*IntimF21
IntimM1 ~~ 0*IntimM1
[...]
IntimM21 ~~ 0*IntimM21

  # random intercepts
iF =~ 1*IntimF1 + [...] + 1*IntimF21
iM =~ 1*IntimM1 + [...] + 1*IntimM21

  # random covariance structure
iF ~~ ivarA*iF + iA*iM + 0*zF1 + [...] + 0*zF21
 + 0*zM1 + [...] + 0*zM21
iM ~~ ivarB*iM + 0*zF1 + [...] + 0*zF21
 + 0*zM1 + [...] + 0*zM21

  # latent variables structure
zF1 =~ 1*IntimF1
[...]
zF21 =~ 1*IntimF21
```

```
zM1 =~ 1*IntimM1
[...]
zM21 =~ 1*IntimM21

  # latent regression structure
zF1 ~ 0*1
zF2 ~ 0*1 + zrho*zF1
[...]
zF21 ~ 0*1 + zrho*zF20
zM1 ~ 0*1
zM2 ~ 0*1 + zrho*zM1
[...]
zM21 ~ 0*1 + zrho*zM20

  # latent covariance structure
zF1 ~~ IntimvarA1*zF1 + IntimA1*zM1
[...]
zF21 ~~ IntimvarA21*zF21 + IntimA21*zM21
zM1 ~~ IntimvarB1*zM1
[...]
zM21 ~~ IntimvarB21*zM21

  # latent model constraints
IntimvarA2 == IntimvarA1*(1 - zrho^2)
[...]
IntimvarA21 == IntimvarA1*(1 - zrho^2)
IntimvarB2 == IntimvarB1*(1 - zrho^2)
[...]
IntimvarB21 == IntimvarB1*(1 - zrho^2)
IntimA2 == IntimA1*(1 - zrho^2)
[...]
IntimA21 == IntimA1*(1 - zrho^2)'
```

3.B SAS code of the illustrative examples

Below, the SAS-syntax of the L-APIM for the empirical example is shown. Note, the data file "dat" corresponds to a long data structure in which the variable *PosRel* has been rearranged into an actor and partner component, and in which the split-up of both components has already been performed.

```
PROC MIXED DATA=dat METHOD=ML COVTEST CL;
 CLASS dyad gender time;
 MODEL intim = gender gender*posrelActorA
  gender*posrelActorS
  gender*posrelPartnerA
  gender*posrelPartnerS / NOINT S CL;
 RANDOM gender / SUBJECT=dyad TYPE=UN G V;
 REPEATED gender time / SUBJECT=dyad
  TYPE=UN@AR(1) R;
RUN;
```

The parameter estimates using this syntax can be found in Table 3.B.1 and 3.B.2. As mentioned in the discussion, missingness has been treated differently in the MLM framework used in SAS compared to the SEM framework used in `lavaan`. Consequently, some deviations between both model fits are to be expected. Yet, similar conclusions are found when fitting the L-APIM on the empirical example data.

Table 3.B.1 Parameter estimates of the fixed effects for the L-APIM for the effect of positive relationship feelings on intimacy obtained by SAS.

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Effect on male outcome						
Intercept	4.084	0.105	39.04	<.001	3.877	4.291
PosRelAM	0.603	0.167	3.62	<.001	0.278	0.929
PosRelAF	0.329	0.142	2.31	.021	0.050	0.608
PosRelSM	0.335	0.078	4.31	<.001	0.183	0.488
PosRelSF	0.075	0.073	1.03	.303	-0.068	0.219
Effect on female outcome						
Intercept	4.313	0.091	47.15	<.001	4.132	4.494
PosRelAF	0.593	0.124	4.78	<.001	0.350	0.837
PosRelAM	0.187	0.145	1.29	.198	-0.098	0.473
PosRelSF	0.174	0.072	2.42	.016	0.033	0.315
PosRelSM	0.278	0.076	3.63	<.001	0.128	0.428

Table 3.B.2 Estimates of the covariance parameters for the L-APIM for the effect of positive relationship feelings on intimacy obtained by SAS.

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Between-dyad						
τ_M^2	0.505	0.122	4.15	<.001	0.331	0.863
τ_F^2	0.351	0.095	3.68	<.001	0.220	0.649
τ_{MF}	0.231	0.088	2.62	.009	0.058	0.404
Within-dyad						
σ_M^2	2.391	0.112	21.30	<.001	2.185	2.627
σ_F^2	2.311	0.107	21.52	<.001	2.114	2.537
σ_{MF}	1.546	0.093	16.63	<.001	1.364	1.728
ρ	0.064	0.027	2.33	.020	0.010	0.118

4

Multilevel autoregressive models when the number of time points is small

Abstract. The multilevel autoregressive model can describe dynamic processes within individuals and allows to differentiate between-person variation from within-person variation. In its basic form, the model essentially includes a random intercept to capture the interindividual variation, and a lagged dependent outcome to account for intraindividual variation. We discuss the performance of the no centering approach, the manifest centering approach and the latent centering approach to fit the multilevel autoregressive model in the setting where the number of time points is small. More specifically, we investigate how standard implementations of these approaches in different software packages deal with the initial conditions problem and endogeneity assumption. We find bias for the parameters of interest, especially when violating the endogeneity assumption. We address this bias within different centering approaches, and end up with two superior approaches.

This chapter is submitted to Structural Equation Modeling: A Multidisciplinary Journal. Gistelinck, F., Loeys, T. & Flamant, N. (n.d.). Multilevel autoregressive models when the number of time points is small. *Structural Equation Modeling: A Multidisciplinary Journal.*

4.1 Introduction

In psychological research, interest not only lies in understanding *interindividual* variation, but also in understanding the change within the individual (i.e., *intraindividual* variation). For this reason, data are ideally collected in a longitudinal fashion, with individuals repeatedly measured over time. From a statistical perspective, a multilevel modeling approach can be used to acknowledge this nested structure. While this modeling framework allows the researcher to separate the interindividual and intraindividual variation, its standard implementation only corrects for time stable sources of non-independence. However, as observations within an individual subject are related to each other over time in a non-stable manner, one should also take into account the temporal correlation, often referred to as *autocorrelation*. Ignoring this type of non-independence may lead to incorrect inference (Fitzmaurice, Laird, & Ware, 2011; Hox, 2010).

Over the last few decades, different extensions within the multilevel framework have been developed in order to control for autocorrelation. On the whole, one could distinguish two major progressions. The first progression incorporates the specification of a residual covariance structure, allowing the residuals of repeated outcomes to covary with one another over time. Assuming equally spaced observations, if one expects the non-independency between the residuals to fade away as the time lag becomes larger, a first-order autoregressive structure for the residual covariance structure may be adopted. That is, the temporal correlation of two adjacent time points is depicted by an autocorrelation parameter ρ ($-1 < \rho < 1$). This progression involved the advancement of existing multilevel software packages (Grimm & Widaman, 2010; Kwok, West, & Green, 2007; Little, 2013) from independent errors to more complex error covariance structures, but it also caused a breakthrough in structural equation modeling (SEM) software (Cole, Cielsa, & Steiger, 2007; Hildreth, 2013). Recently, RDSEM (Residual Dynamic Structural Equation Modeling) in Mplus was presented as a flexible tool to fit such models in SEM (Asparouhov & Muthén, 2018a).

While in the first progression the autocorrelation is typically considered a nuisance, a second progression focuses on the autoregressive relationship between consecutive measurements and allows the prior value to have a direct linear effect on the current value of the outcome. In other words, a lagged dependent variable is included as a predictor in the model. The associated autocorrelation parameter then portrays the time it takes for the individual to recover from a perturbation and restore equilibrium. This second progression is commonly used within the area of econometrics as it allows to model dynamic data-generating processes (Arellano, 2003; Woolridge, 2009), and recently gained popularity in the behavioral sciences as well (Hamaker & Grasman, 2015), leading to the development of DSEM (Dynamic Structural Equation Modeling) in Mplus (Asparouhov, Hamaker, & Muthén, 2018).

One of the most popular models within this second progression is the multilevel autoregressive model or ML-AR(1) model. However, implementing this model with direct autoregressive effect of a lagged outcome on the current outcome harbors some complications. First of all, conditioning on a lagged outcome changes the interpretation of the estimated intercept compared to a model that allows the residuals to be correlated (Jongerling & Hamaker, 2011). As we will show later, an adjusted intercept value is estimated rather than a true underlying common equilibrium. As a result, different centering methods, including manifest mean centering and latent mean centering, have been proposed for directly estimating the equilibrium. Regardless of this centering question, there are two other eminent issues regarding the implementation of the ML-AR(1) model: the initial conditions problem and the endogeneity problem. As the ML-AR(1) model focuses on modeling processes that are ongoing, the initial time point of measurement is affected by an unavailable presample response (i.e., the initial conditions problem). In most approaches, the latter is treated as missing, thereby causing the first time point to be eliminated from the analysis (Zhang, Hamaker, & Nesselroade, 2008). Moreover, the outcome measured at the first time point is thereby considered as an exogenous variable rather than an endogenous one (Bianconcini & Bollen, 2004). Consequently, most software packages will ignore the first outcome score's contribution to the underlying subject-specific equilibrium (i.e., the endogeneity problem). Ignoring this non-independence between the subject-specific equilibrium and the first outcome score (i.e., assuming exogeneity) may cause erroneous inference(Fotouhi, 2005; Skrondal & Rabe-Hesketh, 2014).

Several authors have already compared estimation procedures for the ML-AR(1) models. Jongerling, Laurenceau, and Hamaker (2015) contrasted the no centering approach (based on maximum likelihood estimation), the manifest mean approach (based on maximum likelihood estimation) and the latent mean approach (based on Bayesian estimation) for varying number of time points ($T = 10, 20$ or 50) and subjects ($N = 20, 50$

or 100). While no centering produced very good results in their simulation study, Hamaker and Grasman (2015) noted that manifest person-centering of the lagged predictor revealed considerable bias for the autocorrelation parameter (Hamaker & Grasman, 2015), which is known as Nickell's bias in econometrics literature (Nickell, 1981). This bias is caused by using an observed individual's mean instead of the true underlying one. The observed mean is treated as if its predetermined without taking into account its uncertainty. Allison, Williams, and Moral-Benito (2017), however, showed that the no centering approach yields a biased estimator for the autocorrelation parameter in case the random intercept is assumed to be uncorrelated with the lagged dependent variable, and showed how this bias is removed once allowing for this correlation in the SEM framework. In their seminal paper on DSEM, Asparouhov et al. (2018) claim that also the DSEM framework, which relies on latent centering, can be used to eliminate this dynamic panel bias. Their simulation study considers $T = 10$ to $T = 100$ time points, and $N = 30$ to $N = 100$ subjects. Schultzberg and Muthén (2018) further investigated the question on how many number of subjects and time points are needed for multilevel time series analysis in the DSEM framework, and varied both T and N from 10 to 200. They concluded that DSEM performs better for samples with many subjects and few time points than samples with few subjects and many time points. More recently, McNeish (2019) studied the latter scenario, also referred to as intensive longitudinal data or time series data. In this article, we focus on the first scenario, also referred to as panel data, and compare the performance of different approaches towards fitting a ML-AR(1) when the number of time points is small.

The outline of this paper is as follows. First, we introduce the multilevel first-order autoregressive model and discuss the different modeling approaches. We illustrate how the no centering approach, the manifest mean centering and the latent mean centering approach lead to different conclusions about the autoregressive parameter of our motivating example data. We conduct a simulation study to illuminate the performance of the different approaches. We find the no centering approach assuming endogeneity to perform well, but the latent centering approach in DSEM to perform differently than expected. To gain further insight into the performance of the DSEM-implementation in Mplus within the setting of a small number of time points, we perform a second simulation study exploring the latent centering approach in more detail. We conclude that not treating the first measurement as an endogenous variable but still having it contributing to the latent mean yields the best performance. While we focus on a relatively simple model in this paper, we elaborate on the broader implications of our findings on the analysis of panel data in the discussion.

4.2 The multilevel autoregressive model

As a motivating example throughout this paper, consider the study of Flamant and Soenens (n.d.). These authors investigate pupil's perception of controlling behavior of the teacher and how it affects their motivation, class engagement, psychological need-frustration and general well-being. About 400 pupils from different schools and grades were asked to fill in a weekly questionnaire for four consecutive weeks. Here, we focus on the estimation of the dynamics of the autonomous motivation of the pupil with only those four repeated measures available, and wonder to what extent the motivation of last week affects the current motivation. One of the most promising models in psychological research to model dynamics of such stationary process is the multilevel autoregressive model (Kuppens, Allen, & Sheeber, 2010; Rovine & Walls, 2006). Due to its multilevel nature, it allows researchers to differentiate in interindividual and intraindividual differences of motivation, and due to its autoregressive structure, it also controls for autocorrelation.

Below, the two-level specification of the multilevel first-order autoregressive model or ML-AR(1) model, is presented. A graphical representation of the model can be found in Figure 4.2.1. The level 1 equation describes the outcome scores as a mean score, usually considered as a subject-specific equilibrium, and a temporal deviation from this mean:

$$y_{ti} = \mu_i + z_{ti} \tag{4.1}$$

with i referring to the subject number ($i = 1, \dots, N$) and t to the time point ($t = 1, \dots, T$). The level 2 equation then corresponds to the modeling assumptions of the individual's equilibrium and describes the latter as a grand mean with white noise:

$$\mu_i = \mu + \eta_i \tag{4.2}$$

with $\eta_i \sim N(0, \tau_\eta^2)$.

The individual's deviations at level 1 are modeled as a first-order autoregressive structure with autocorrelation parameter ρ ($-1 < \rho < 1$):

$$z_{ti} = \rho z_{(t-1),i} + \varepsilon_{ti} \tag{4.3}$$

with $\varepsilon_{ti} \sim N(0, \sigma_\varepsilon^2)$. As the residuals ε_{ti} account for the part that is not predicted by the previous $z_{t-1,i}$, it is also referred to as *the innovation*. For instance, in the study of Flamant and Soenens (n.d.), a large residual variance σ_ε^2 would correspond to a high level of perturbation in the autonomous motivation of the participants. While an autocorrelation parameter ρ close to zero would imply little carryover effect of motivation from one measurement occasion to the next, an autocorrelation parameter close to one would imply that there is considerable carryover between consecutive measurement occasions. Hence, the autocorrelation parameter is often referred to as a measure of *inertia*. Note that the equilibrium is assumed to be subject-specific here. If the stability of constructs shows some time-invariant characteristic (e.g., due to the intrinsic motivation of the student), it is shown that the autocorrelation parameter alone is not able to account for this trait (Hamaker, Kuiper, & Grasman, 2015). Even in case the autocorrelation parameter is very close to one, indicating a very high carryover effect, the parameter can only account for temporal stability as its effect nullifies when enough time passes.

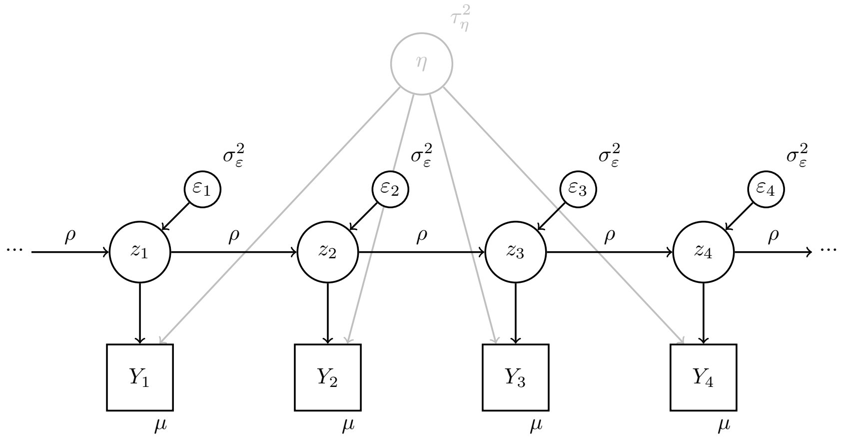


Figure 4.2.1 A schematic overview of the multilevel first-order autoregressive model for four time points.

Models (4.2) and (4.3) further assume the autocorrelation parameter ρ and the level 1 residual variance σ_ε^2 to be fixed instead of random. As mentioned by Jongerling et al. (2015), a person-specific residual variance might be important as it capture differences in sensitivity to innovation. Assuming the residual variance as fixed may induce small bias in the autocorrelation parameter, however. This is explained by the fact that the variance in a ML-AR(1) model both depends on the innovation variance and the autocorrelation parameter. As the bias in the autocorrelation parameter is rather small, we will focus on models with a fixed innovation variance. This choice is also made because we investigate the ML-AR(1) model in the context of panel data (i.e., a small amount of time points). Assuming a person-specific variance might unnecessarily complicate the model and often leads to convergence issues in this setting.

For similar reasons, we also assume a fixed rather than a random autocorrelation parameter. As shown by Schultzberg and Muthén (2018), a random autocorrelation parameter (or random innovation variance) requires a lot more time points in order to obtain a model fit with good performance. Furthermore, incorrectly fitting the autocorrelation parameter as fixed rather than random hardly affects the mean squared error. It is also known that no bias is introduced in the mean structure parameter estimates when treating the parameter as fixed instead of random (Baird & Maxwell, 2016). Finally, as we will only consider a stationary process, the assumption of "$-1 < \rho < 1$" must hold. This implies that the variance of a random inertia cannot be large to begin with. Consequently, if the variance should be small, assuming a random autocorrelation parameter might become futile, resulting in a large mean squared error due to over-specification.

4.3 Implementations of the ML-AR(1) model

Although the ML-AR(1) model specified by equations (4.1) - (4.3) seems straightforward, its implementation is quite challenging. In this section, different approaches towards fitting the ML-AR(1) model will be introduced.

4.3.1 The no centering approaches

The first approach is quite simple. It involves the substitution of the expression (4.2) and (4.3) in expression (4.1). Using the fact that $z_{(t-1),i} = y_{(t-1),i} - \mu_i$, one finds:

$$y_{ti} = \alpha_i + \rho y_{t-1,i} + \varepsilon_{ti}, \tag{4.4}$$

with $\alpha_i \sim N(\alpha, \tau_\alpha^2)$ and

$$\begin{aligned} \alpha &= (1-\rho)\mu, \\ \tau_\alpha^2 &= (1-\rho)^2\tau_\eta^2. \end{aligned} \tag{4.5}$$

Hence, when using model (4.4), one does not estimate the true individual's equilibrium μ or variance τ_η^2, but rather a transformation of them (namely, α and τ_α^2). We will refer to this model as the *NC model*, as it uses the uncentered lagged outcome variable as a predictor in the model.

As expression (4.4) illustrates, one needs a start-up for the autoregressive process (i.e., the initial conditions problem). The simplest solution to this problem is assuming the first outcome variable as exogenous (i.e., predetermined) with its own mean and variance. Given the lack of knowledge about the unobserved previous observations, this seems logical and it is also the most common practice in the multilevel modeling framework. We will refer to this model as the *NC-EXO model* since the first outcome variable is considered exogenous. A graphical representation can be found in Figure 4.3.1. The *NC-EXO model* can be fitted in either the multilevel modeling (MLM) framework, the structural equation modeling (SEM) framework or the Bayesian framework (BAY). It is well known that MLM models have equivalent models within the SEM framework (Curran, 2003). The latter framework uses latent variable to represent theoretical constructs or unmeasured factors (Kline, 2015; Schumacker & Lomax, 2004). As MLM uses the restricted maximum likelihood (REML) estimator by default, asymptotically unbiased estimates are obtained for the (co)variance parameters, in contrast to SEM which typically uses the (unrestricted) maximum likelihood (ML) estimator by default. It is further known that (RE)ML estimation might not be the best choice due to small sample bias. Instead of restricting ourselves to a frequentist approach, one might, therefore, want to appeal to the Bayesian approach (Bolstad & Curran, 2016; Lynch, 2007).

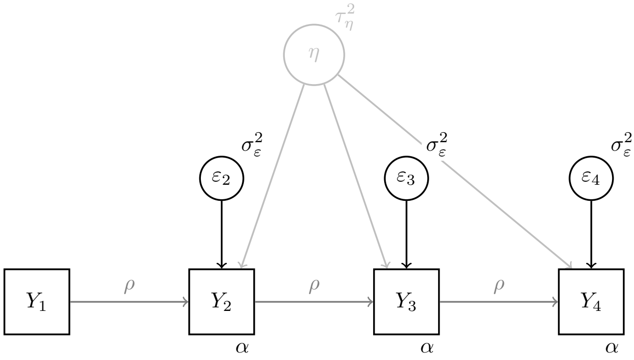


Figure 4.3.1 A schematic representation of the *NC-EXO* parametrization of the uncentered ML-AR(1) model for four time points.

Regardless of the framework used, assuming the first outcome variable to be exogenous causes the *NC-EXO model* to suffer from the endogeneity problem. The lagged outcome variable at the first time point is assumed to be uncorrelated with the random intercept. As suggested by Bollen and Curran (2004), one can solve this endogeneity problem by specifying a relation between the first outcome variable and the random intercept, for instance, by allowing the first outcome variable and the random intercept to be correlated (Teachman, Duncan, & Yeung, 2001). A graphical representation is presented in Figure 4.3.2 and will be referred to as the *NC-ENDO model.*

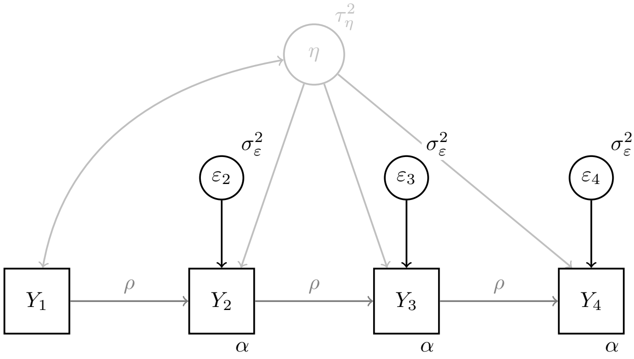


Figure 4.3.2 A schematic representation of the *NC-ENDO* parametrization of the uncentered ML-AR(1) model for four time points.

Using standard MLM software packages, such as HLM (Raudenbush, 2004), Mplus' two-level option (Muthén & Muthén, 2012), SAS's PROC MIXED (SAS Institute, 2008), R packages like `nlme` (Pinheiro & Bates, 2000) or `lme4` (Bates, Mächler, Bolker, & Walker, 2015), all predictor variables, including the lagged dependent variable at the first time point, are considered independent from the random effects by default. Consequently, the *NC-ENDO model* cannot be fitted within standard MLM. Contrary, in the SEM and the BAY framework, one can mitigate the assumption of independence between the exogenous variables and the random effects. If one uses standard SEM packages, such as EQS (Bentler, 2004), LISREL (Jöreskog & Sörbom, 1996), OpenMx (Boker et al., 2011), Mplus (Muthén & Muthén, 2012), Stata's gllamm (Rabe-Hesketh, Skrondal, & Pickles, 2004), SAS's proc CALIS (SAS Institute, 2013) or R's `lavaan` (Rosseel, 2012), one can easily introduce a correlation between the first outcome variable and the latent intercept. Within the BAY framework, adding such correlation is a bit more complex as one needs to assume a multivariate normal distribution for the first outcome variable and the random intercept, but it may be implemented within standard BAY software packages like WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), JAGS (Plummer, 2003) or Stan (Carpenter et al., 2017), using their corresponding R packages `R2WinBugs` (Sturtz, Ligges, & Gelman, 2005), `RJags` (Plummer, 2016) or `RStan` (Stan Development Team, 2018).

4.3.2 The centering approaches

As mentioned above, the no centering approach does not estimate the true individual's equilibrium (variance), but a transformation of it (cf. expression 4.5). Although this does not sound that harmful, it does complicate the interpretation of the model parameter estimates. Looking for a solution, the most popular solution to this rescaling issue is centering (Enders & Tofighi, 2007). If one centers the lagged outcome variable around its subject-specific equilibrium, one only conditions on the within-level of the previous time point and, hence, no longer on the time-invariant effects of the previous time point. One can thus rewrite equation (4.4) as follows:

$$y_{ti} = \mu_i + \rho(y_{t-1,i} - \mu_i) + \varepsilon_{ti} \tag{4.6}$$

with $\mu_i \sim N(\mu, \tau_\mu^2)$ and $\varepsilon_{ti} \sim N(0, \sigma_\varepsilon^2)$. This way, the underlying equilibrium μ_i may be estimated by the model in a straightforward fashion at first sight. However, there is one big catch-22: in order to obtain an estimate of the subject-specific equilibrium, we need the subject-specific equilibrium to center the lagged outcome variable.

Jongerling et al. (2015) suggested to use the observed individual's sample mean $\bar{y}_i$ to center the lagged outcome variable:

$$y_{ti} = \mu_i + \rho(y_{t-1,i} - \bar{y}_i) + \varepsilon_{ti} \tag{4.7}$$

This centering approach is also known as the cluster-mean centering, manifest centering or observed centering. Hence, we will refer to model (4.7) as the *CMC model.* A graphical representation of the model can be found in Figure 4.3.3. An alternative is based on a two-step procedure, in which an estimate for the random intercept is first obtained by fitting an empty model. Here, we will no further discuss this alternative as Jongerling and Hamaker (2011) showed that the differences in performance between both approaches are negligible. Moreover, the first approach is more common practice.

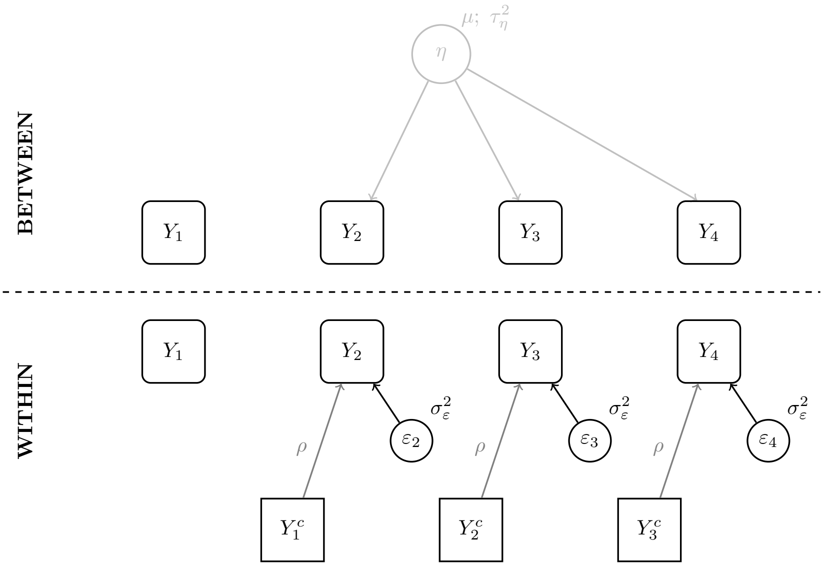


Figure 4.3.3 A schematic representation of the *CMC* parametrization of the cluster-mean centered ML-AR(1) model for four time points.

In this CMC approach, on still has to deal with the initial conditions problem and endogeneity problem. Again, the first outcome variable is assumed as exogenous in order to deal with the initial conditions problem. The *CMC models* are further known to suffer from Nickell's bias (Nickell, 1981): using the observed individual's mean instead of the true underlying one introduces bias in the autocorrelation parameter ρ. An approximation of this bias is given by the following formula:

$$-\frac{1+\rho}{T-1} \tag{4.8}$$

The bias will be especially prominent when the number of time points T is small. Furthermore, it is obvious that Nickell's bias is persistent in case the number of subjects increases. Moreover, due to the negative sign in (4.8), it is even possible to obtain a negative estimate for a true underlying positive autocorrelation.

Interestingly, the simulation studies of Hamaker and Grasman (2015) also showed that, even in case the true underlying mean is used to center, bias is still observed due to the unsolved endogeneity problem. Therefore, a solution should not only avoid Nickell's bias, but should also account for the endogeneity problem. As claimed by Asparouhov and Muthén (2018b), latent centering is able to deal with Nickell's bias. Instead of using the observed equilibrium, one can specify the unobserved subject-specific equilibrium as a latent variable (Lüdtke et al., 2008; Preacher, Zyphur, & Zhang, 2010). The idea is to split the outcome variable into two parts: a within and a between part, which are both considered latent, that is:

$$\begin{aligned} y_{ti} &= \tilde{y}_i + \tilde{y}_{ti} \\ \tilde{y}_i &= \mu + \eta_i \\ \tilde{y}_{ti} &= \rho\tilde{y}_{(t-1),i} + \varepsilon_{ti} \end{aligned} \tag{4.9}$$

with $\eta_i \sim N(0, \tau_\eta^2)$ and $\varepsilon_{ti} \sim N(0, \sigma_\varepsilon^2)$. We will refer to these models as the *LC models*. This latent centering can be performed using DSEM from Mplus in the BAY framework (Muthén & Muthén, 2012). Due to the iterative nature of the Markov Chain Monte Carlo technique, one can use the random intercept's prior in order to perform latent centering of the lagged outcome variable in the posterior distribution, while all model parameter estimates (including the random intercept itself) are updated simultaneously. Asparouhov et al. (2018) further explain that DSEM deals with the initial condition problem by introducing auxiliary variables for the unobserved presample responses which are used in the prediction of the first outcome variable. This variable is unknown and has its own prior distribution based on the sample mean and variance of the outcome variables (Zhang & Nesselroade, 2008).

4.3.3 Empirical Example

As an illustration of the four above described approaches to the ML-AR(1) model, we applied them to our motivating example data on autonomous motivation of primary school pupils. For ease of comparison, we eliminated the incomplete cases, so that 310 pupils with complete data remained in the data set. We used the following statistical frameworks and software to implement the four approaches:

- the *NC-EXO model* within the MLM framework using the `nlme` package in R,
- the *NC-ENDO model* within the SEM framework using the `lavaan` package in R,
- the *CMC model* within the MLM framework using the `nlme` package in R,
- the *LC model* within the BAY framework using DSEM in Mplus.

The model within Mplus was fitted using the R package `MplusAutomation` (Hallquist & Wiley, 2018). In the Bayesian approach, prior distributions, number of burn-in iterations, etc. need to be specified. All specifications were kept at default. The code for the models can be found in the supplementary materials.

The results in Table 4.3.1 reveal remarkable differences between the four approaches. The *NC-EXO* and DSEM approach both find a very strong autocorrelation, larger than 0.70, while the autocorrelation is close to 0.30 for the *NC-ENDO model.* The negative autocorrelation by the *CMC model* is not realistic and may be indicative for the earlier mentioned Nickell's bias. Also the residual variances are comparable between the *NC-EXO* and DSEM, smaller for the *NC-ENDO*, and even more so for the *CMC model.* The intercepts and random intercept variances are only directly comparable within the no centering and centering approaches, respectively. However, after transforming, all approaches estimate the equilibrium for autonomous motivation around 3.35. The *NC-EXO model* finds no evidence for a positive random intercept variance (i.e., estimate at the boundary of the parameter space), while also in the DSEM approach, it is very small. This may explain the large difference between both the *NC-EXO* and *NC-ENDO model.* As the latter has a very small random intercept variance, most variation is incorporated by the autocorrelation parameters, leading to a small random intercept. The estimate for the random intercept variance under the *NC-ENDO model* equals $\frac{0.34}{(1-0.29)^2} = 0.67$, implying that 95% of the participants specific equilibrium lie between 2.09 and 4.66.

Table 4.3.1 Estimates (and standard errors) of the parameters from the ML-AR(1) model on the empirical example with four time points using the different parametrizations.

	No centering approach			Centering approach	
Est.	NC-EXO	NC-ENDO	Est.	CMC	DSEM
α	0.800(0.080)	2.376(0.243)	μ	3.360(0.052)	3.381(0.053)
ρ	0.760(0.023)	0.292(0.071)	ρ	-0.166(0.041)	0.747(0.026)
τ_α^2	<0.001(<0.001)	0.335(0.092)	τ_η^2	0.732(0.067)	0.120(0.074)
σ_ε^2	0.466(0.022)	0.343(0.027)	σ_ε^2	0.289(0.016)	0.459(0.023)

Notes. $\alpha = (1-\rho)\mu$ and $\tau_\alpha^2 = (1-\rho)^2\tau_\eta^2$.

4.3.4 Simulation study 1

Given the discrepancies between the four approaches in our empirical example, we further explore their performance in a simulation study where the underlying truth is known. The data-generating process of the simulation study is as follows:

$$\begin{aligned} y_{ti} &= \mu_i + z_{ti} \\ \mu_i &= 2 + \eta_i \\ z_{ti} &= 0.4 z_{(t-1),i} + \varepsilon_{ti} \end{aligned} \tag{4.10}$$

with i referring to the subject number ($i = 1, \ldots, 50$) and t to the time point ($t = 1, \ldots, T$) in which we varied T from 4 till 20 by steps of 2. The white noises η_i and ε_{ti} follow a normal distribution with mean zero and variance equal to 3 and 3.36, respectively, such that the variance structure of the original residuals z_{ti} has a variance equal to 4. We compare the 4 approaches using 200 replications. Note, the *NC models* estimate a rescaled random intercept with mean equal to $2 * (1 - 0.4) = 1.2$ and variance $3 * (1 - 0.4)^2 = 1.08$.

As mentioned before, one needs to specify the priors in the BAY framework and we again choose to use the default priors of DSEM in Mplus:

$$\begin{aligned} \mu &\sim N(0, \infty) \\ \rho &\sim N(0, \infty) \\ \tau_\eta^2 &\sim IG(-1, 0) \\ \sigma_\varepsilon^2 &\sim IG(-1, 0) \end{aligned} \tag{4.11}$$

with *IG* referring to the inverse gamma distribution. Mplus thus allows for improper distribution as they are advantageous in small samples with respect to bias and mean squared error (Asparouhov & Muthén, 2010). The $IG(-1, 0)$ is approximately equivalent to the uniform distribution with minimum 0 and maximum ∞.

When considering four time points, Mplus' DSEM reported convergence issues for 29 of the 200 repetitions. In these cases, unusual high parameter estimates were found, if a value was reported at all. As a result, we opted to limit the sensitivity of these outliers by using robust summary statistics. The median parameter estimates can be found in Figure 4.4.1, and a similar plot for the median absolute errors (MedAE) can be found in the supplementary materials. As one can see, all approaches, except for the *CMC model*, tend to yield unbiased estimates for the model parameters when at least 10 time points are available. This is consistent with the findings of the simulation study of Hamaker and Grasman (2015). The *CMC model* shows negative bias for the autocorrelation parameter, as predicted by Nickell's bias, even when the amount of time points is 20 (the prediction based on formula (4.8) is indicated by the blue triangle). Consequently, the model also yields biased estimates for the random intercept variance.

Considering the no centering approaches, it is clear that we must deal with the endogeneity problem. As the *NC-ENDO model* shows, the bias from the exogeneity assumption in the *NC-EXO model* can be eliminated by adding a correlation between the first lagged outcome and the random intercept. We also see that as soon as the number of time points is smaller than 6, the *NC-EXO model* estimates the random intercept as almost zero, causing maximal bias in all other variables (cf. the hitch at the start of the curves). The *NC-ENDO model* also illustrates that assuming the first outcome variable as a predetermined variable does not harm the parameter estimation since no bias is observed within this approach. Using Mplus' DSEM to fit the *LC model*, we notice that DSEM does not perform as expected when the number of time points is smaller than 10. This is interesting because no bias was found in a similar simulation study in Asparouhov et al. (2018), investigating latent centering. However, the latter was performed in case of intensive longitudinal data. The question arises what causes the bias. Is it possible that latent centering is not suitable when the number of time points is rather small? Does adding the auxiliary variable in order to resolve the initial condition problem make the estimation unreliable? Could it be that the bias is attributed to specific choices made in the default settings of DSEM, for instance, what is the impact of the choice of the priors? Or, perhaps, the performance of the Bayesian approach for the model in this particular setting is overall deficient?

4.4 Alternative models with latent centering

In order to elucidate the estimation behavior of the *LC model*, we implemented the *LC model* using an open source Bayesian software package, more specifically, the R package `Rjags`. The code for the Bayesian models that we discuss below, can be found in the supplementary materials.

As mentioned above, different causes can be identified that imperil the parameter estimation. Firstly, the bias might be due to the endogeneity problem, which is not resolved by latent centering. In a first implementation, the outcome variable is split up into a latent between and within part at all time points. Figure 4.4.2a represents the *LC-EXO model.* In this implementation, the between part of the first outcome variable does not contribute to the estimation of the random intercept. However, it has its own prior distribution based on the observed mean and variance (cf. auxiliary variable).

We also introduce a similar Bayesian model that deals with the endogeneity problem, retaining the auxiliary variable for the unobserved presample responses, see Figure 4.4.2b. In this case, an association is invoked between the random intercept and the between part of the first outcome variable. The auxiliary variable is introduced at the within-level of the model as a direct start-up of the the autoregressive process. We will refer to this model as the *LC-ENDO-0 model.* The number '0' at the end of this acronym represents the time point used as the start-up of the autoregressive process.

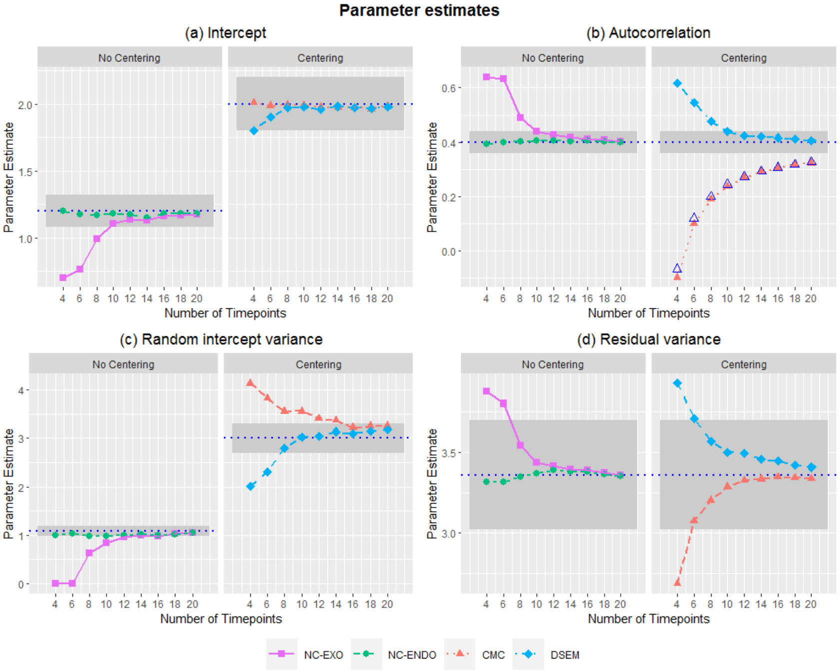
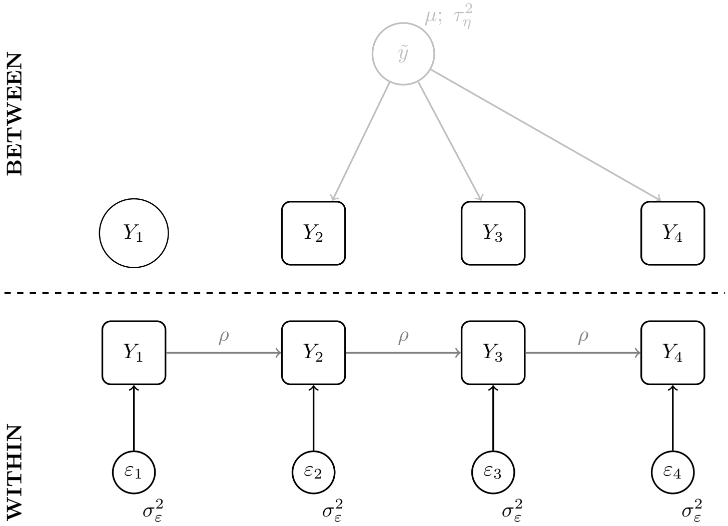


Figure 4.4.1 The parameter estimates of the ML-AR(1) model over 200 repetitions for $t = 4, 6, \ldots, 18, 20$ time points based on the different model parametrizations. The blue dotted line represents the true (transformed) model parameter. The gray band represents the absence of relative bias at a 10% cut-off. The blue triangles in subfigure (b) represents Nickell's bias.

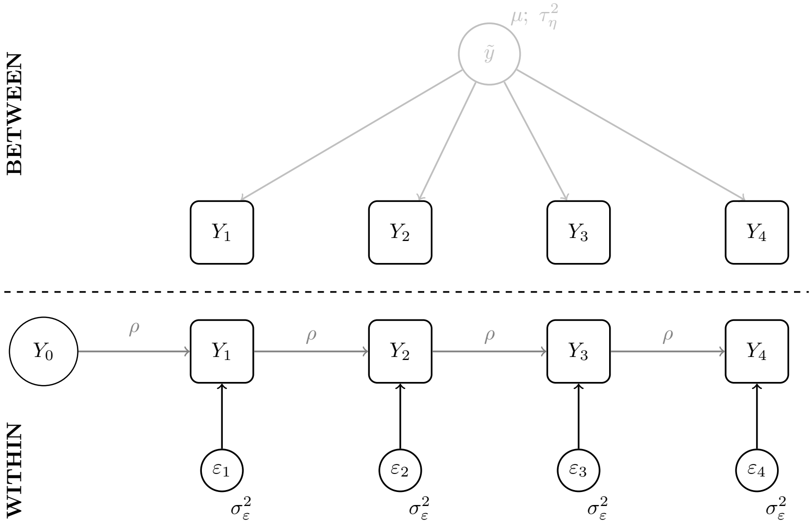
Secondly, the introduction of this auxiliary variable may make the model unreliable in case of panel data. For this reason, a third model implementation is introduced, which removes the auxiliary variable from the *LC-ENDO-0 model*, while preserving the correlation between the first outcome variable and the random intercept. A graphical representation of this model, which we will refer to as the *LC-ENDO-1 model*, can be found in Figure 4.4.2c. In this case, the residual variance of the within-part of the first outcome variable is considered free. As a result, the endogeneity problem has been dealt with and the first time point represents the start-up of the autoregressive process, hence the '1' at the end of the acronym.

The issues above are all related to the specification and assumptions of the model. However, it is possible that the genesis of the bias lies within the statistical framework used. On the one hand, the BAY framework is often preferred in case of panel data as it is supposed to be more robust to small sample bias. However, the selected prior distributions may have an impact on the estimation of the model parameters (Gelman et al., 2013). Therefore, non-informative prior distributions are often suggested. The most frequently used non-informative prior distributions for mean structure parameters, like μ and ρ in our case, is the normal distribution with mean zero and a high variance. For the variance structure parameters, such as τ_η^2 and σ_ε^2, several non-informative prior distributions have been proposed, such as an inverse gamma distribution with a small scale parameter, a uniform distribution with a large scale parameter, a half-Cauchy distribution or a log-normal distribution. Alternatively, one may specify a prior distribution for the standard deviations instead, for example, using a uniform distribution (Zitzmann, Lüdke, & Robitzsch, 2015). As a result, different priors should be used in order to check the prior sensitivity of the model.

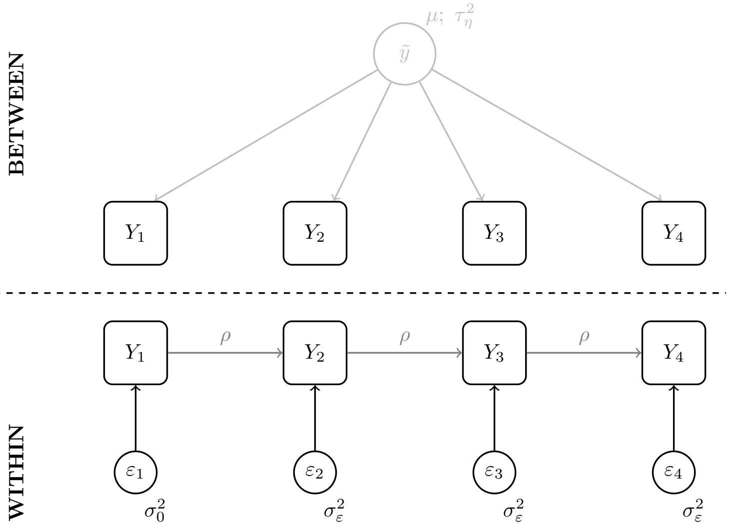
While the BAY approach is somehow robust to small sample bias, it can provide less accurate estimates in case of challenging data constellations, such as small groups, a situation inherent to panel data (Zitzmann, Lüdtke, Robitzsch, & Marsh, 2016). Recently, McNeish (2019) showed that the likelihood which updates the prior distribution carries less weight within the posterior distribution in case of small samples. As a result, the non-informative prior distributions may become unintentionally informative (McNeish & Stapleton, 2016). The reasoning can be similar for small cluster sizes. Therefore, we also considered a frequentist approach which does not use prior distributions, but relies on maximizing the likelihood.



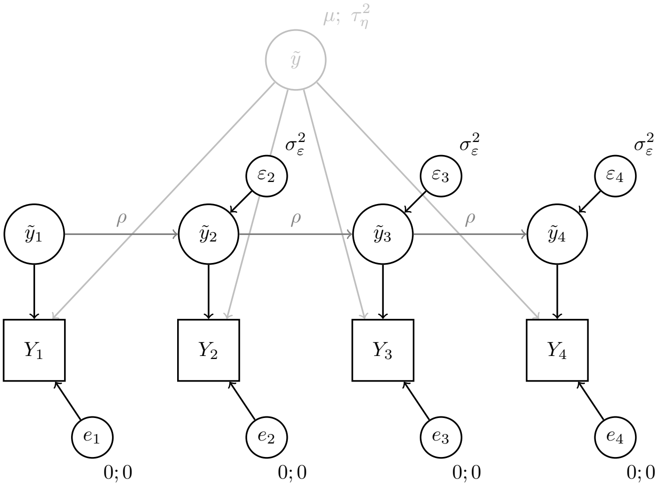
(a) The *LC-EXO model*.



(b) The *LC-ENDO-0 model*.



(c) The *LC-ENDO-1 model*.



(d) The *LC-ENDO-1 model* within the SEM framework.

Figure 4.4.2 A schematic representation of the different *LC models* of the ML-AR(1) model for four time points. Note that the phantom residuals e_t for $t = 1, \ldots, 4$ in Figure 4.4.2d are generated automatically when defining $\tilde{y}$ and $\tilde{y}_t$ and hence, cannot be eliminated from the model.

Within the SEM framework, one can still perform latent centering. This is achieved by defining a latent variable $\tilde{y}_i$ over all time points and a residual latent variable $\tilde{y}_{ti}$ for each time point (Beaujean, 2014). As this split-up explains the observed outcome values completely, the total variance is subdivided into between-level variation and within-level variation. Consequently, one needs to eliminate the original residuals of the outcome variables by (manually) setting their variance to zero. In the SEM framework, it is then possible to deal with the endogeneity problem as well, by adding a correlation between the latent intercept and the latent within-part of the first outcome variable. Moreover, the variance of its within-part can be considered free, allowing the first time point to represent the start-up of the autoregressive process (cf. the initial conditions problem). Hence, we end up with a fourth model, namely an equivalent *LC-ENDO-1 model* within the SEM framework. A graphical representation of this implementation can be found in Figure 4.4.2d.

4.4.1 Simulation study 2

Consider the data-generating process of simulation study 1 as described by expression (4.10). Again, we assume 50 subjects and 4 till 20 time points, varied by steps of 2. In this second simulation study, the four models depicted in Figure 4.4.2 were fitted with `Rjags` and `lavaan`, and were compared to Mplus' DSEM. As `Rjags` neither allows for infinite bounds for prior distributions nor improper priors, we decided to fix the prior distributions for the mean structure parameters to a normal distributions with variance 100,000 in both `Rjags` and Mplus. We also considered three different prior distributions for the variance structure parameters: the uniform distribution with scale parameter 1,000, the inverse gamma distribution with scale parameter 0.001 and the half-Cauchy distribution with scale parameter 10 (not possible in Mplus). Based on the model diagnostics, such as traceplots, the Brooks-Gelman-Rubin plots and the autocorrelation function (Albert, 2009), we decided to keep the thinning fixed at one, to increase the number of chains from two to three and to compute 50,000 MCMC iterations. Note that `Rjags` allows the user to define the burn-in phase separately from the update phase. In the latter phase, `Rjags` also makes a difference between adapting the samplers used in the Markov chain and the initial burn-in period (Plummer, 2015). This is in contrast to Mplus, which uses by default only half of the number of iterations as a burn-in phase and the other half to compute the posterior distribution. Depending on the distance of the Proportional Scale Reduction (PSR) from one (i.e. the convergence criterion), the posterior distribution might be based on a smaller amount of iterations. In order to maintain as much equivalence between `Rjags` and Mplus, the same amount of iterations for the update phase in both software packages were imposed, namely 25,000 (half of 50,000), using “FBITER” within Mplus (Muthén, 2010).

From the simulation results (not shown here), it was clear that the uniform and half-Cauchy distribution provided similar results for all `Rjags` models. The diagnostics of the models based on the inverse gamma prior were bad, which might explain the difference in performance compared to the other priors for both the `Rjags` models as Mplus’ DSEM. Hence, we will further only focus on the result assuming a uniform prior distribution. The simulation results for the median of the parameter estimates are depicted in Figure 4.5.1. A similar plot for the MedAE can be found in the supplementary materials. As Mplus reported convergence issues for 84 of the 200 repetitions at four time points, the impact of these runs on the results are limited with these robust measurements. This non-convergence issue may again explain the jump of the different implementations when going from four to six time points.

Figure 4.5.1 reveals that the *LC-EXO model*, which ignores the endogeneity problem, performs rather similar to DSEM. To keep the readability of the figure, the 95%-confidence intervals based on the median absolute differences (MADs) were plotted for only these two models. There are some differences in the autocorrelation parameter and when $T = 4$ in the other model parameters, although the trend is very akin to one another. Figure 4.5.1 further reveals that the *LC-ENDO-0 model* attenuates the bias for the estimate of the autocorrelation parameter compared to the DSEM and the *LC-EXO model.* The estimator of the intercept from the *LC-ENDO-0 model*, however, shows more bias when the amount of time points is rather small. When the *LC-ENDO-1 model* is used, it performs very well, even in case only a small amount of time points are available. This confirms the fact that latent centering can be used in order to resolve the centering problem. Moreover, it can be seen that the corresponding SEM implementation of the *LC-ENDO-1 model* provides unbiased estimates as well and might be preferred over its BAY implementation given that is has less convergence issues when the number of time points is very small.

To summarize, we have found two approaches for fitting the ML-AR(1) that are elusive to bias when applied to panel data. On one hand, if the researcher is not bothered by estimating a rescaled intercept, an uncentered model in which a correlation between the first lagged outcome variable and the random intercept is included, can be used (i.e., the *NC-ENDO model*). On the other hand, if the researcher is interested in directly estimating the equilibrium, latent centering could be used. In this case, the *LC-ENDO-1 model* based on maximum likelihood in the SEM framework performs best. Lastly, the researcher should be aware of the default settings of either type of software packages as those might not always be adequate for valid inference (Muthén & Asparouhov, 2012).

4.5 Discussion

In this paper, we discussed different ways to fit the ML-AR(1) model in light of the endogeneity problem and the initial conditions problem. We found that ignoring the endogeneity problem can lead to severe bias in the autocorrelation parameter. While manifest cluster-mean centering is known to introduce Nickell's bias, we also showed how some implementations of the latent mean centering approach in the Bayesian framework may show bias, especially when the number of the time points is small. This is mostly due to problematic management of the unobserved presample outcome. The no centering approach that properly deals with the endogeneity problem (i.e., the *NC-ENDO model*) and the latent centering approach that does not utilize an auxiliary variable for the unobserved presample outcome (i.e., the *LC-ENDO-1 model*) performed best.

There are several limitations to this conclusion. First, we only considered a simple version of the ML-AR(1) model, with a fixed autocorrelation parameter and a constant residual variance. Although it is likely that our findings still hold in the ML-AR(1) model which relaxes those restrictions, the implementation of the *NC-ENDO* or *LC-ENDO-1 model* for this complex setting in the traditional SEM framework requires some further investigation. Regardless, fitting such models may be too demanding for the limited amount of information available when the number of time points is small. Second, we assumed that all subjects were measured simultaneously at equidistant time points. However, this may not always be true in practice: some time points may be scheduled further apart from each other or time points may differ between subjects. This may complicate the model a lot if the autocorrelation parameter is treated time-specific.

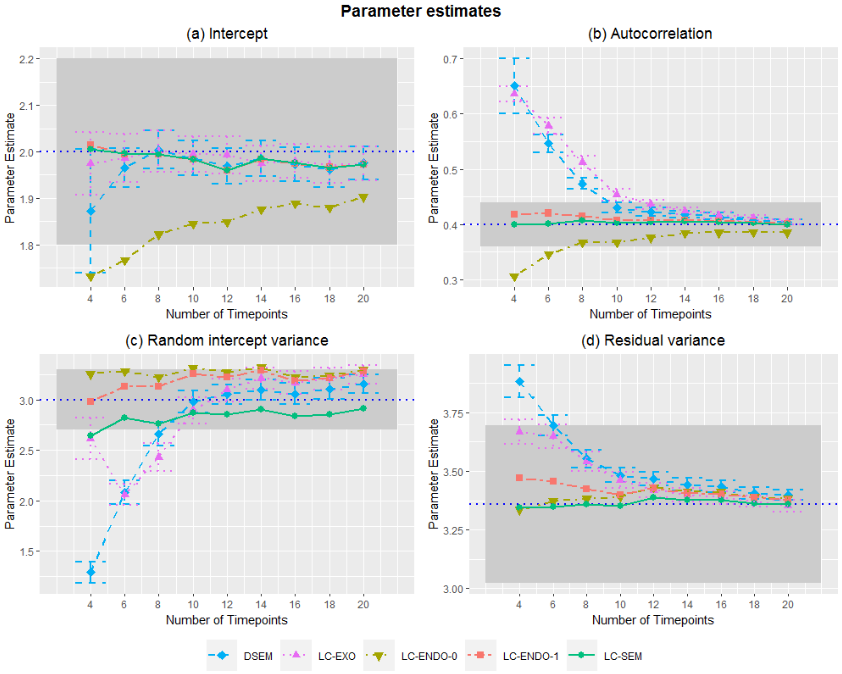


Figure 4.5.1 The parameter estimates of the ML-AR(1) model over 200 repetitions for $t = 4, 6, \ldots, 18, 20$ time points based on the different model implementations for the *LC model*. The blue dotted line represents the true model parameter. The gray band represents the absence of relative bias at a 10% cut-off. The 95%-confidence intervals for the DSEM and *LC-EXO model* has been provided as well.

Third, the ML-AR(1) model discussed in this paper did not include any other predictors in the model, either time-independent or time-varying. Clearly, our findings have implications on more complex models such as the random intercept cross-lagged panel model (RI-CLPM) proposed by Hamaker et al. (2015), or dynamic network models (Bringmann et al., 2013). In case the number of time points is small, the bias in the autoregressive parameter may be substantial when the unobserved presample response and the endogeneity are not appropriately dealt with. Further research is required to explore the impact on the parameter estimates of other predictors. For example, it is possible that the cross-lagged effects of X on Y also show bias when using the wrong approach (Allison et al., 2017).

Finally, it is worth noting that the ML-AR(1) resembles latent state-trait (LST) models (Steyer, Ferring, & Schmitt, 1992). The LST framework can be used to study longitudinal dynamics of psychological attributes too and can, for example, determine the degree to which such attributes reflect stable effects, effects of person-situation interactions, or random measurement error. The single trait-multistate model (STMS model) is the simplest type of LST model which allows the latent trait component to be separated from the state residual components. Similar to the traditional random intercept model, only covariances between the latent states are allowed so they can be captured by the common trait of the STMS model. This assumption might be unreasonable in some situations, hence, the latent state-trait autoregressive model or LST-AR model (Steyer & Schmitt, 1994) relaxes it by including autoregressive effects between the latent state variables at each time point (cf. DSEM). Cole, Martin, and Steiger (2005) proposed the trait-state occasion model (TSO model) as an alternative for the LST-AR model, adding the autoregressive effect between the state residual variables instead (cf. RDSEM). More knowledge may be gained on how reliable and stable estimation of the ML-AR(1) model is best achieved by comparing both frameworks (Lüdtke, Robitzsch, & Wagner, 2018).

References

Albert, T. (2009). *Bayesian computation with R* (2nd ed.). New York, NY: Springer Science & Business Media.

Allison, P., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius: Sociological Research for a Dynamic World*, *3*, 1-17. doi: 10.1177/2378023117710578

Arellano, M. (2003). *Panel data econometrics* (1st ed.). New York, NY: Oxford University Press.

Asparouhov, T., Hamaker, E., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359-388. doi: 10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus* (Tech. Rep.). Mplus Web Notes.

Asparouhov, T., & Muthén, B. (2018a). *Comparison of models for the analysis of intensive longitudinal data.* Retrieved from `http://www.statmodel.com/download/RDSEM.pdf` (Mplus Web Notes)

Asparouhov, T., & Muthén, B. (2018b). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(1), 119-142. doi: 10.1080/10705511.2018.1511375

Baird, R., & Maxwell, S. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *12*(2), 175-188. doi: 10.1037/met0000070

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beaujean, A. (2014). *Latent variable modeling using R* (1st ed.). New York, NY: Routledge.

Bentler, P. (2004). *EQS 6 structural equations program manual* (6th ed.). Encino, CA: Multivariate Software, Inc.

Bianconcini, S., & Bollen, K. (2004). The latent variable-autoregressive latent trajectory model: A general framework for longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *32*(3), 336-383. doi: 10.1177/0049124103260222

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

s11336-010-9200-6

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological methods and research*, *32*(3), 336-383. doi: 10.1177/0049124103260222

Bolstad, W., & Curran, J. (2016). *Introduction to bayesian statistics* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Bringmann, L., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e60188. doi: 10.1371/journal.pone.0060188

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1-32. doi: 10.18637/jss.v076.i01

Cole, D., Cielsa, J., & Steiger, J. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381-398. doi: 10.1037/1082-989X.12.4.381

Cole, D., Martin, N., & Steiger, J. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state occasion model. *Psychological Methods*, *10*(1), 3-20. doi: 10.1037/1082-989X.10.1.3

Curran, P. (2003). Have multilevel models been structural equations models all along? *Multivariate Behavioral Research*, *38*(4), 529-569. doi: 10.1207/s15327906mbr3804_5

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121

Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Flamant, N., & Soenens, B. (n.d.). *Consequences of coping with controlling teachers.*

Fotouhi, A. (2005). The initial conditions problem in longitudinal binary process: A simulation study. *Simulation Modelling Practice and Theory*, *13*(7), 566-583. doi: 10.1016/j.simpat.2005.02.004

Gelman, A., Stern, H., Carlin, J., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis* (3rd ed.). New York, NY: Chapman and Hall/CRC.

Grimm, K., & Widaman, K. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 424-442. doi: 10.1080/10705511.2010.489006

Hallquist, M., & Wiley, J. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analysis in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 621-638. doi: 10.1080/10705511.2017.1402334

Hamaker, E., & Grasman, R. (2015). To center or not to center? investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. doi: 10.3389/fpsyg.2014.01492

Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. doi: 10.1037/a0038889

Hildreth, L. (2013). *Residual analysis for structural equation modeling* (Unpublished doctoral dissertation). Iowa Stat University.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). London, UK: Routledge.

Jongerling, J., & Hamaker, E. (2011). On the trajectories of the predetermined ALT model: What are we really modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 370-382. doi: 10.1080/10705511.2011.582004

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334-349. doi: 10.1080/00273171.2014.1003772

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide* (2nd ed.). Lincolnwood, IL: Scientific Software International, Inc.

Kline, R. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press.

Kuppens, P., Allen, N., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984-991. doi: 10.1177/0956797610372634

Kwok, O., West, S., & Green, S. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research*, *42*(3), 557-592. doi: 10.1080/00273170701540537

Little, T. (2013). *The oxford handbook of quantitative methods in psychology* (1st ed.). New York, NY: Oxford University Press.

Lüdtke, O., Marsh, H., Robitzsch, A., Trautwein, U., Asparouhov, R., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229. doi: 10.1037/a0012869

Lüdtke, O., Robitzsch, A., & Wagner, J. (2018). More stable estimation of the STARTS model: A bayesian approach using markov chain monte carlo techniques. *Psychological Methods*, *23*, 570-593. doi: 10.1037/met0000155

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325-337. doi: 10.1023/A:1008929526011

Lynch, S. (2007). *Introduction to applied bayesian statistics and estimation for social scientists* (1st ed.). Princeton, NJ: Springer Science & Business Media.

McNeish, D. (2019). Two-level dynamic structural equation models with small samples. *Structural Equation Modeling: A Multidisciplinary Journal.* doi: 10.1080/10705511.2019.1578657

McNeish, D., & Stapleton, L. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*(4), 495-518. doi: 10.1080/00273171.2016.1167008

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Tech. Rep.). Mplus Web Notes.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods*, *17*, 313-335. doi: 10.1037/a0026802

Muthén, L., & Muthén, B. (2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417-1426. doi: 10.2307/1911408

Pinheiro, J., & Bates, M. (2000). *Mixed-effects models in S and S-PLUS.* New York, NY: Springer.

Plummer, M. (2003). *JAGS: A program for analysis of bayesian graphical models using gibbs sampling.* (3rd international workshop on distributed statistical computing)

Plummer, M. (2015). JAGS version 4.0.0 user manual [Computer software manual].

Plummer, M. (2016). rjags: Bayesian graphical models using mcmc [Computer software manual]. (R package version 4-6)

Preacher, K., Zyphur, M., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological*

Methods, *15*(3), 209-223. doi: 10.1037/a0020141

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, *69*(2), 167-190. doi: 10.1007/BF02295939

Raudenbush, S. (2004). *HLM6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. Retrieved from `http://www.jstatsoft.org/v48/i02/` doi: 10.18637/jss.v048.i02

Rovine, M., & Walls, T. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. In T. Walls & J. Schafer (Eds.), *Models for intensive longitudinal data* (p. 124-147). Oxford University Press.

SAS Institute, I. (2008). *SAS/STAT 9.2 user's guide: The MIXED procedure.* Cary,NC: SAS Institute Inc.

SAS Institute, I. (2013). *SAS/STAT 13.1 user's guide: The CALIS procedure.* Cary, NC: SAS Institute Inc.

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 495-515. doi: 10.1080/10705511.2017.1392862

Schumacker, R., & Lomax, R. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, New Jersey London: Lawrence Erlbaum Associates.

Skrondal, A., & Rabe-Hesketh, S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society, Series C*, *63*(2), 211-237. doi: 10.1111/rssc.12023

Stan Development Team. (2018). RStan: the r interface to stan [Computer software manual]. (R package version 2.17.3)

Steyer, R., Ferring, D., & Schmitt, M. (1992). States and treits in psychological assessment. *European Journal of Psychological Assessment*, *8*(2), 79-98.

Steyer, R., & Schmitt, T. (1994). The theory of confounding and its application in causal modeling with latent variables. In A. von Eye & C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (p. 36-67). Thousand Oaks, CA: Sage Publications, Inc.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running winbugs from r. *Journal of Statistical Software*, *12*(3), 1-16. doi: 10.7916/D80C55HH

Teachman, J., Duncan, G., & Yeung, D., W.J. an dLevy. (2001). Covariance structure models for fixed and random effects. *Sociological Methods and Researh*, *30*(2), 271-288. doi: 10.1177/0049124101030002005

Woolridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed.). Mason, OH: South-Western Cengage Learning.

Zhang, Z., Hamaker, E., & Nesselroade, J. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(3), 377-402. doi: 10.1080/10705510802154281

Zhang, Z., & Nesselroade, J. (2008). Bayesian estimation of categorical dynamic factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *42*(4), 729-756. doi: 10.1080/00273170701715998

Zitzmann, S., Lüdke, O., & Robitzsch, A. (2015). A bayesian approach to more stable estimates of group-level effects in contextual studies. *Multivariate Behavioral Research*, *50*(6), 688-705. doi: 10.1080/00273171.2015.1090899

Zitzmann, S., Lüdtke, O., Robitzsch, A., & Marsh, H. (2016). A bayesian approach for estimating multilevel latent contextual models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 661-679. doi: 10.1080/10705511.2016.1207179

Appendix

4.A R code of the different ML-AR(1) implementations

4.A.1 The *NC-EXO model* within the MLM framework

```
fit <- lme(yy ~ 1 + yprev,
 random =~ 1|id,
 data = dat_long,
 na.action = na.omit)
summary(fit)
```

4.A.2 The *NC-ENDO model* within the SEM framework

```
model <- '
  # random effects
ri =~ 1*yy.2 + 1*yy.3 + 1*yy.4
ri ~~ vari*ri + delta*yy.1
  # mean structure
yy.1 ~ alpha0*1
yy.2 ~ alpha*1 + rho*yy.1
yy.3 ~ alpha*1 + rho*yy.2
yy.4 ~ alpha*1 + rho*yy.3
  # residual covariance model
yy.1 ~~ resvar0*yy.1
yy.2 ~~ resvar*yy.2
yy.3 ~~ resvar*yy.3
yy.4 ~~ resvar*yy.4
  # model constriants
vari > 0
resvar > 0
resvar0 > 0'

fit <- sem(model, data = dat_wide)
summary(fit)
```

4.A.3 The *CMC model* within the MLM framework

```
fit <- lme(yy ~ 1 + yprev_c,
 random =~ 1|id,
 data = dat_long,
 na.action = na.omit)
summary(fit)
```

4.A.4 The *LC-EXO model* within the BAY framework using Mplus

```
model <- mplusObject(TITLE = "LC_model",
 VARIABLE = "cluster = id; \n lagged = yy(1);",
 ANALYSIS = "TYPE IS TWOLEVEL;
  \n estimator = Bayes;",
 MODEL = " %WITHIN% \n yy (a); \n yy on yy&1 (b);
  \n \n%BETWEEN% \n [yy] (c); \n yy (d);",
 OUTPUT = "TECH8;",
 PLOT = "TYPE = PLOT3;",
 usevariables = colnames(dat_long),
 rdata = dat_long)

fit <- mplusModeler(model,
 modelout = "model.inp",
 run = 1)
coef(fit)
```

4.A.5 The *LC-EXO model* within the BAY framework using Rjags

```
varY <- var(dat_long$yy)
muY <- mean(dat_long$yy)

## Model specifications
modelstring <- 'model{
 # N... total number of subject
 # J... number of timepoints
 # Y... outcome
```

```
  # Likelihood
for(i in 1:N){
 for(t in ((i-1)*J+2):(i*J)){
  Y[t] ~ dnorm(expr[t], invSigmaSqrd.Y)
  expr[t] <- delta[i] + beta.w*z[t-1]
  z[t] <- Y[t] - delta[i]
 }
}

for(i in 1:N){
 Y[(i-1)*J+1] ~ dnorm(expr0[i], invSigmaSqrd.Y)
 expr0[i] ~ dnorm(muY,1/varY)
 z[(i-1)*J+1] <- Y[(i-1)*J+1] - delta[i]
}

for(j in 1:N){
 delta[j] ~ dnorm(alpha, invTauSqrd.Y)
}

  # Prior distributions
alpha ~ dnorm(0, 0.00001)
beta.w ~ dnorm(0, 0.00001)

sigmaSqrd.Y ~ dunif(0,1000)
invSigmaSqrd.Y <- 1 / sigmaSqrd.Y

tauSqrd.Y ~ dunif(0,1000)
invTauSqrd.Y <- 1 / tauSqrd.Y
}'

## Initiate the model
model <- jags.model(textConnection(modelstring),
 data=list('Y'= dat_long$yy,
  'N'= N,
  'J'= tps,
  'muY'= muY,
  'varY' = varY),
 n.chains = 3,
 n.adapt = 5000)
```

```
## Update initial model
update(model, 20000)

## Run MCMC chain
fit <- coda.samples(model,
 variable.names=c('alpha','beta.w',
  'sigmaSqrd.Y','tauSqrd.Y'),
 n.iter=25000)
summary(fit)
```

4.A.6 The *LC-ENDO-0 model* within the BAY framework using Rjags

```
varY <- var(dat_long$yy, na.rm = TRUE)
muY <- mean(dat_long$yy, na.rm = TRUE)

## Model specifications
modelstring <- 'model{
 # N... total number of subject
 # J... number of timepoints
 # Y... outcome

  # Likelihood
for(i in 1:N){
 for(t in ((i-1)*J+2):(i*J)){
  Y[t] ~ dnorm(expr[t], invSigmaSqrd.Y)
  expr[t] <- delta[i] + beta.w*z[t-1]
  z[t] <- Y[t] - delta[i]
 }
}


for(i in 1:N){
 Y[(i-1)*J+1] ~ dnorm(expr0[(i-1)*J+1],
  invSigmaSqrd.Y)
 expr0[(i-1)*J+1] <- delta[i] + beta.w*z0[i]
 z[(i-1)*J+1] <- Y[(i-1)*J+1] - delta[i]
}
```

```
for(j in 1:N){
 delta[j] ~ dnorm(alpha, invTauSqrd.Y)
 z0[j] ~ dnorm(muY,1/varY)
}

  # Prior distributions
alpha ~ dnorm(0, 0.00001)
beta.w ~ dnorm(0, 0.00001)

sigmaSqrd.Y ~ dunif(0,1000)
invSigmaSqrd.Y <- 1 / sigmaSqrd.Y

tauSqrd.Y ~ dunif(0,1000)
invTauSqrd.Y <- 1 / tauSqrd.Y
}'

## Initiate the model
model <- jags.model(textConnection(modelstring),
 data=list('Y'= dat_long$yy,
  'N'= N,
  'J'= tps,
  'muY'= muY,
  'varY' = varY),
 n.chains = 3,
 n.adapt = 5000)

## Update initial model
update(model, 20000)

## Run MCMC chain
fit <- coda.samples(model,
 variable.names=c('alpha','beta.w',
  'sigmaSqrd.Y','tauSqrd.Y'),
 n.iter=25000)
summary(fit)
```

4.A.7 The *LC-ENDO-1 model* within the BAY framework using Rjags

```
## Model specifications
modelstring <- 'model{
 # N... total number of subject
 # J... number of timepoints
 # Y... outcome

  # Likelihood
for(i in 1:N){
 for(t in ((i-1)*J+2):(i*J)){
  Y[t] ~ dnorm(expr[t], invSigmaSqrd.Y)
  expr[t] <- delta[i] + beta.w*z[t-1]
  z[t] <- Y[t] - delta[i]
 }
}

for(i in 1:N){
 Y[(i-1)*J+1] ~ dnorm(delta[i], invSigmaSqrd.Y0)
 z[(i-1)*J+1] <- Y[(i-1)*J+1] - delta[i]
}

for(j in 1:N){
 delta[j] ~ dnorm(alpha, invTauSqrd.Y)
}

  # Prior distributions
alpha ~ dnorm(0, 0.00001)
beta.w ~ dnorm(0, 0.00001)

sigmaSqrd.Y ~ dunif(0,1000)
invSigmaSqrd.Y <- 1 / sigmaSqrd.Y

sigmaSqrd.Y0 ~ dunif(0,1000)
invSigmaSqrd.Y0 <- 1 / sigmaSqrd.Y0

tauSqrd.Y ~ dunif(0,1000)
invTauSqrd.Y <- 1 / tauSqrd.Y }'
```

```
## Initiate the model
model <- jags.model(textConnection(modelstring),
 data=list('Y'= dat_long$yy,
  'N'= N,
  'J'= tps),
 n.chains = 3,
 n.adapt = 5000)

## Update initial model
update(model, 20000)

## Run MCMC chain
fit <- coda.samples(model,
 variable.names=c('alpha','beta.w',
  'sigmaSqrd.Y','sigmaSqrd.Y0','tauSqrd.Y'),
 n.iter=25000)
summary(fit)
```

4.A.8 The *LC-ENDO-1 model* within the SEM framework

```
model <- '
  # latent between-parts
ri =~ 1*yy.1 + 1*yy.2 + 1*yy.3 + 1*yy.4
ri ~ alpha*1
ri ~~ vari*ri + 0*z1 + 0*z2 + 0*z3 + 0*z4

  # latent within-parts
z1 =~ 1*yy.1
z2 =~ 1*yy.2
z3 =~ 1*yy.3
z4 =~ 1*yy.4

z1 ~ 0*1
z2 ~ 0*1 + rho*z1
z3 ~ 0*1 + rho*z2
z4 ~ 0*1 + rho*z3
```

```
z1 ~~ resvar0*z1 + 0*z2 + 0*z3 + 0*z4
z2 ~~ resvar*z2 + 0*z3 + 0*z4
z3 ~~ resvar*z3 + 0*z4
z4 ~~ resvar*z4

  # mean structure
yy.1 ~ 0*1
yy.2 ~ 0*1
yy.3 ~ 0*1
yy.4 ~ 0*1

  # residual covariance model
yy.1 ~~ 0*yy.1
yy.2 ~~ 0*yy.2
yy.3 ~~ 0*yy.3
yy.4 ~~ 0*yy.4

  # model constriants
vari > 0
resvar > 0
resvar0 > 0'

fit <- sem(model, data = dat_wide)
summary(fit)
```

4.B Simulation results: Median Absolute Error plots

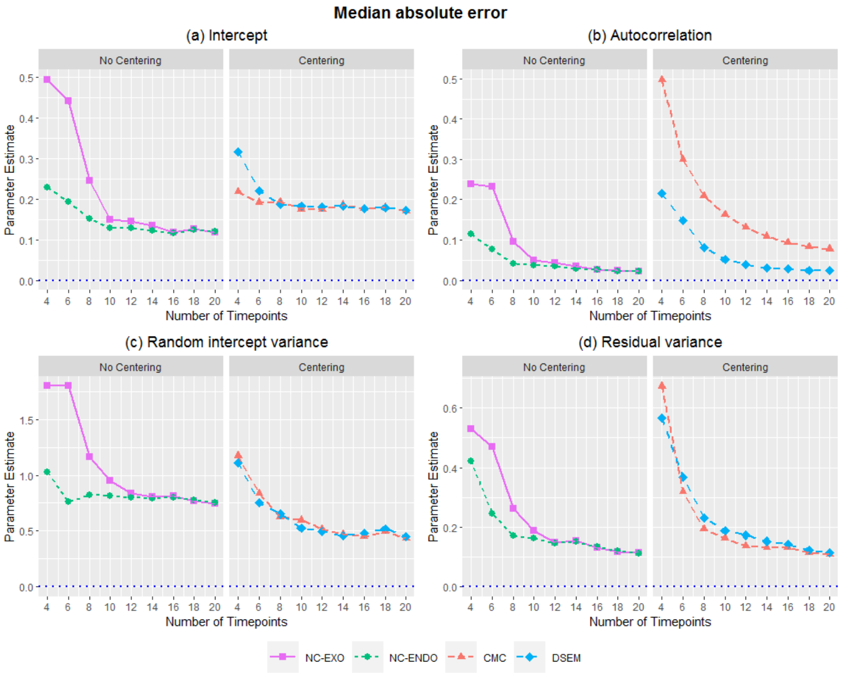


Figure 4.B.1 The median absolute error of the ML-AR(1) model over 200 repetitions for $t = 4, 6, \ldots, 18, 20$ time points based on the different model parametrizations. Similar to the parameter estimates, one could conclude the *NC-ENDO model* to perform better than the *NC-EXO model*, illustrating the necessity to deal with the endogeneity problem. Also, DSEM shows no great improvement (except for the autocorrelation parameter) compared to the *CMC model*.

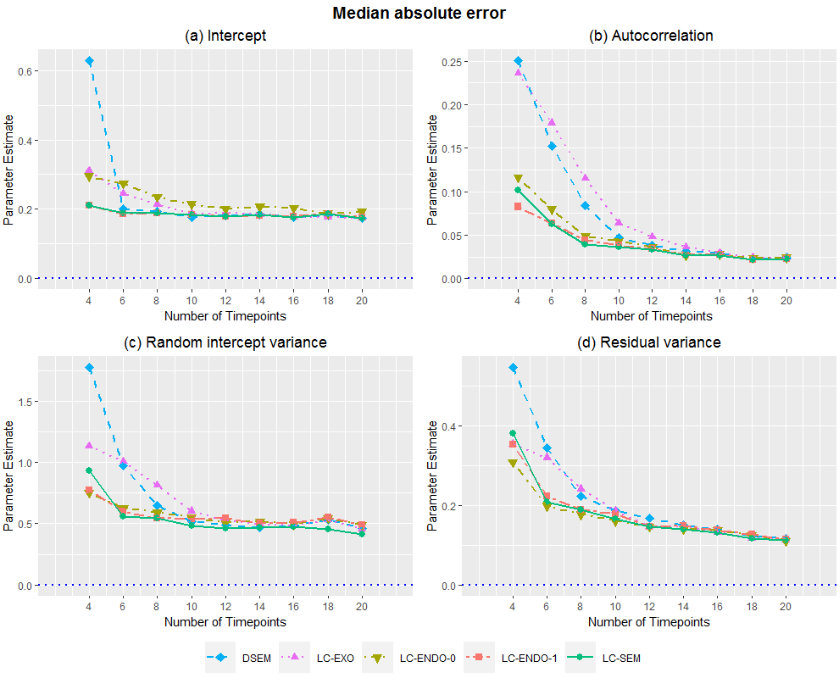


Figure 4.B.2 The median absolute error of the ML-AR(1) model over 200 repetitions for $t = 4, 6, \ldots, 18, 20$ time points based on the different model implementations for the *LC model*. Similar to the parameter estimates, DSEM and *LC-EXO* perform very similar (except for the autorcorrelation parameter). The other implementations perform very similar, with small discrepancies at 4 time points.

5

Multilevel autoregressive models for longitudinal dyadic data

Abstract. In social and behavioral science, dyadic research has become more and more popular. In case of cross-sectional dyadic data, one can apply the actor-partner interdependence model or APIM when both an antecedent and an outcome are measured in each dyad member. In this paper, we introduce an extension of the cross-sectional APIM to the longitudinal setting, called the LD-APIM or lagged-dependent APIM. It allows to investigate the effects of an antecedent on an outcome given the previous outcome, and extends the multilevel autoregressive model to dyadic data. To facilitate the use of this complex model, a user-friendly Shiny-application, called the *LDDinSEM*-application, was developed. The app automatically fits the LD-APIM on the uploaded data set within the structural equation modeling (SEM) framework. We illustrate both the model and the app using an empirical example assessing the actor and partner effects of positive relationship feelings on next day's intimacy in heterosexual couples.

This chapter will be submitted to the special issue 'Multilevel modeling: Different research fields and research designs' of the journal 'Testing, Psychometrics, Methodology in Applied Psychology'. Gistelinck, F. & Loeys, T. (n.d.). Multilevel autoregressive models for longitudinal dyadic data. *Testing, Psychometrics, Methodology in Applied Psychology.*

5.1 Introduction

Psychologists are often interested in the processes underlying social and behavioral phenomena. As these phenomena are often interpersonal by definition (Reis, Collins, & Bersheid, 2000), the popularity of collecting dyadic data rather than individual data has increased exponentially over the last decade. From a statistical point of view, analyzing such data can be quite a challenge: the subjects are related to one another, violating the independence assumption of most standard statistical procedures (Gonzalez & Griffin, 1999). Indeed, members of the same dyad will behave more (dis)similar than two random people. One of the most popular and widely used statistical model to analyze dyads is the actor-partner interdependence model or APIM (Kashy & Kenny, 2000). This model estimates the effect of an antecedent on the behavior or emotional status of a person within the dyadic context. More specifically, the model incorporates the effect of one's own characteristics or feelings on his/her outcome score (i.e., the *actor effect*), as well as the effect of one's partner characteristics or feelings on his/her own outcome score (i.e., the *partner effect*). Furthermore, it accounts for *interdependence* by allowing the outcomes of the dyad members to be correlated with each other.

Research becomes even more challenging and sophisticated when these dyads are measured repeatedly over time. Consider, for example, the study of 66 heterosexual couples in Flanders (Dewitte, Van Lankveld, Vandenberghe, & Loeys, 2015). These couples were asked to fill in a daily diary study on sexual behavior over a period of three weeks. Both men and women were inquired about their sexual and intimate behavior, as well as their individual, relational and partner-related feelings and behavior of the past day. For now, focus on the association between the positive feelings about the relationship and the perception of intimacy within couples. The richness of questions that can be raised from such longitudinal dyadic data, is immense: "How do the average positive relational feelings of one's partner affect one's own perception of intimacy, given the intimacy perceived on the previous day?", "To what extent do the average perceived intimacies correlate between man and woman within a dyad?", "How strong does yesterday's perceived intimacy affects today's perceived intimacy within a given person?", etc.

Such longitudinal dyadic data (LDD) come with an additional amount of statistical challenges (Gistelinck & Loeys, 2019). Since longitudinal data are characterized by a nested structure, in which occasions are clustered within the unit of analysis, such data are often analyzed using multilevel models (Snijders & Bosker, 2011). In the absence of any predictors, the random intercepts in such models can be viewed as the individual's trait score or *equilibrium*. However, if repeated measures are taken close in time, the current measure does not only depend on the trait, but is likely to be predictable from the preceding measure as well. If a person perceives his/her relationship positively today, it will likely be rated positively tomorrow as well (Cranford et al., 2006). Standard multilevel approaches assuming independent residuals fail to account for such *autocorrelation*. Erroneous inference is obtained in case this temporal correlation is ignored (Fitzmaurice, Laird, & Ware, 2011; Singer & Willett, 2013).

For the analysis of longitudinal individual data with autocorrelation, two major progressions within the multilevel framework can be distinguished. In the first progression, one controls for autocorrelation by specifying a more complex covariance structure for the residuals, such as a first-order autoregressive process. In the second progression autocorrelation is explicitly modeled through the inclusion of a lagged dependent variable (i.e., the dependent variable at the previous time point) as a covariate. Such models have been referred to as multilevel first-order autoregressive or ML-AR(1) models (Kuppens, Allen, & Sheeber, 2010; Rovine & Walls, 2006), and are used to investigate the effect of an antecedent on the dependent variable while controlling for the previous level of the outcome variable. It may seem quite straightforward to implement the latter in standard multilevel software, but Gistelinck and Loeys (n.d.) recently showed that there are some challenges with the ML-AR(1) model. The first issue is the initial conditions problem. In the study of Dewitte et al. (2015), we can not regress, for example, the perceived intimacy on the first day of measurement on its previous level, as it is simply not available. Some approaches, such as the DSEM (Asparouhov, Hamaker, & Muthén, 2018), will seemingly 'impute' some presample responses, whilst other will only start regressing from the second observation to avoid the missing lagged predictor. The second issue is the endogeneity assumption. In a ML-AR(1) model, it is unreasonable to assume that the random intercept is independent from the lagged predictor. For instance, in the study of Dewitte et al. (2015), the underlying perceived intimacy within an individual (i.e., the trait or random intercept) will also affect the perceived intimacy of the first day. Allison, Williams, and Moral-Benito (2017) showed that ignoring this correlation can seriously bias the estimator for the effect of the lagged outcome, as well as the coefficients of the other predictors. For the setting of longitudinal individual data, Gistelinck and Loeys (n.d.) investigated how both issues can best be tackled in the structural equation modeling (SEM) framework. Especially when the number of time points is smaller, one should specify a correlation between the latent intercept and the first lagged outcome to avoid bias due to endogeneity, while the first outcome should be treated as exogenous (i.e., not regressed on an 'imputed' presample response).

When considering longitudinal dyadic data, we could expand the APIM for cross-sectional data to longitudinal data along the lines of the two aforementioned progressions in the individual context. The extension of the APIM in the first progression was proposed by Bolger and Laurenceau (2013), and was further developed in the SEM framework by (Gistelinck & Loeys, 2019). Here, we will propose an extension of the APIM in the second progression, taking into account the above-mentioned issues. We will refer to this model as the LD-APIM or lagged dependent actor-partner interdependence model. Implementing the LD-APIM may become quite intricate for a dyadic researcher and form a genuine barrier. Therefore, we will present in this paper a user-friendly application with its implementation. Even in the case of cross-sectional dyadic data, the need for such a comprehensive analytic tool is apparent. One of the most cited APIM papers involves a user-friendly guide for fitting the APIM using SAS or HLM (Campbell & Kashy, 2002). Similar tutorials with implementations in other statistical software packages, such as Mplus, also gained quite some interest over the last years (Fitzpatrick, Gareau, Lafontaine, & Gaudreau, 2016). However, these require a license for the software under consideration. The recently developed *APIM_SEM*-application (Stas, Kenny, Mayer, & Loeys, 2018), which is part of a bigger project called *DyadR* (Kenny, 2017), allows users to fit standard or more complex APIMs for cross-sectional dyadic data. The application uses a point and click interface, but does not require any software. In this paper, we develop a similar application, called the *LDDinSEM*-application, to fit the LD-APIM for longitudinal dyadic data.

The article is organized as follows. First, a motivating example is presented. Next, we will introduce the LD-APIM and discuss how the model tackles the different issues of LDD. After, we illustrate how the LD-APIM is easily fitted using the *LDDinSEM*-application on the empirical example. We end with a discussion.

5.2 Data and Hypotheses

As a motivating example throughout this paper we consider a Flemish daily diary study on sexual behavior in 66 heterosexual couples (Dewitte et al., 2015). Every morning during three weeks both members of the couple were asked about their sexual and intimate behavior since the last time they had filled out their morning diary (i.e., sexual behavior over the past 24 hours). Every evening the participants were asked to report on their individual, relational, and partner-related feelings and behavior experienced during that day. Here we will only focus on the association between the positive feelings about the relationship and the next morning's perception of intimacy. Positive relationship feelings were computed as the average of nine items on a 7-point Likert scale: the extent to which they felt happy, satisfied, understood, supported, accepted, loved, in love, connected, and close. The amount of intimacy was measured by the amount of kissing, cuddling and caressing rating from 'not at all' to 'very frequent' using a 7-point Likert scale.

Several interesting research questions can be raised from such longitudinal dyadic data:

(Q1) Do people who have generally more positive relationship feelings, also report a higher perception of intimacy, given their perception the day before?

(Q2) Do people who have partners with generally more positive relationship feelings, also report a higher perception of intimacy, given their perception the day before?

(Q3) Given yesterday's perception of intimacy, how does an increase or decrease in one's own positive relationship feelings (as compared to his/her average feelings) affect today's perception of intimacy?

(Q4) Given yesterday's perception of intimacy, how does an increase or decrease in one's partner positive relationship feelings (as compared to their average feelings) affect one's own perception of intimacy?

(Q5) To what extent does yesterday's perception of intimacy affect today's perception of intimacy?

From the formulation of the questions, one can clearly see that (Q1) and (Q2) look at differences between individuals (i.e. a *between* or *time-averaged* effect), while questions (Q3) and (Q4) are concerned with effects within an individual (i.e. a *within* or *time-specific* effect). Longitudinal data allow to disentangle those effects. Furthermore, (Q1) and (Q3) relate to the effect of one's own feelings on one's own behavior (i.e. an *actor* effect), while (Q2) and (Q4) relate to the effect of one's partner feelings on one's own behavior (i.e. a *partner* effect). Using dyadic data, we can study such dynamics. As we look at distinguishable dyads (i.e., individuals within the dyad are distinguished by their gender role), we can allow all those effects to be different for men and women. It is further worth noting that the questions (Q1)-(Q4) look at the effects of positive feelings on the perception of intimacy from a person while controlling for the perception of intimacy from that person on the day before. That is, we are interested in investigating whether positive relationship feelings have any *additional* effect given the within-individual dynamics of perception of intimacy. Throughout the paper, we make the assumption that today's perception of intimacy only depends on yesterday's perception of intimacy, but not on any further lags, although in principle this assumption could easily be relaxed.

To answer the above questions, we can model the effect of positive relationship feelings ($PosRel$) on intimacy ($Intim$) in females and males as follows:

$$\begin{cases} Intim_{Fij} = (\mu_F + \eta_{Fj}) + \rho_F Intim_{Fi-1,j} \\ \qquad + a_{F(PosRelA)} PosRelA_{Fj} + p_{MF(PosRelA)} PosRelA_{Mj} \\ \qquad + a_{F(PosRelS)} PosRelS_{Fij} + p_{MF(PosRelS)} PosRelS_{Mij} + \varepsilon_{Fij} \\ \\ Intim_{Mij} = (\mu_M + \eta_{Mj}) + \rho_M Intim_{Mi-1,j} \\ \qquad + a_{M(PosRelA)} PosRelA_{Mj} + p_{FM(PosRelA)} PosRelA_{Fj} \\ \qquad + a_{M(PosRelS)} PosRelS_{Mij} + p_{FM(PosRelS)} PosRelS_{Fij} + \varepsilon_{Mij} \end{cases}, \tag{5.1}$$

with i referring to the time point ($i = 2, \ldots, 21$), j to the dyad number ($j = 1, \ldots, 66$). The parameters a_F and a_M are the actor effects for the females and the males, respectively, while the parameters p_{MF} and p_{FM} are the partner effects for the females and the males, respectively (F refers to female and M refers to male). The variables $PosRelA$ and $PosRelS$ refer to the time-averaged and time-specific component of positive relationship feelings, respectively. The parameters ρ_F and ρ_M capture the autoregressive process of perceived intimacy for females and males, respectively. Clearly, the parameters $a_{F(PosRelA)}$ and $a_{M(PosRelA)}$ address (Q1), while $a_{F(PosRelS)}$ and $a_{M(PosRelS)}$ answer (Q3). Questions (Q2) and (Q4) on the partner effects can be addressed by the parameters $p_{MF(PosRelA)}$ and $p_{MF(PosRelA)}$, and $p_{MF(PosRelS)}$ and $p_{FM(PosRelS)}$, respectively. We assume all those effects to be fixed (i.e., no variation between individuals). The parameters μ_F and μ_M capture the intercept, which we allow to vary between individuals through the random intercepts η_{Fj} and η_{Mj} for females and males, respectively. The residuals ε_{Fij} and ε_{Mij} capture all the variability that is not explained by the predictors in the model.

In addition to questions (Q1)-(Q5), researchers may also be interested in further exploring the dynamics within and between dyads. For example,

(Q6) Do women (or men) that generally have a high average perception of intimacy, typically have a male (or female) partner with high average perception of intimacy?

(Q7) If a woman (or man) has a high perception of intimacy on a particular day, will her male partner have a high perception of intimacy that day as well.

In other words, (Q6) wonders how trait-like couples behave in their perception, while (Q7) addresses how similar couple members are on specific occasions. To illustrate how those questions are answered, we will first formalize the model more generally in the next section.

5.3 The lagged dependent actor-partner interdependence model

Suppose one wants to explore the effect of an antecedent X on current behavior Y given past behavior, and that both X and Y are repeatedly measured for each dyad member. Then, one can fit the LD-APIM in order to examine this association:

$$\begin{cases} Y_{Fij} = (\mu_F + \eta_{Fj}) + \rho_F Y_{F,i-1,j} + a_{F(X)} X_{Fij} + p_{MF(X)} X_{Mij} + \varepsilon_{Fij} \\ Y_{Mij} = (\mu_M + \eta_{Mj}) + \rho_M Y_{M,i-1,j} + a_{M(X)} X_{Mij} + p_{FM(X)} X_{Fij} + \varepsilon_{Mij} \end{cases}, \tag{5.2}$$

with i referring to the time point ($i = 2, \ldots, T$) and j to the dyad number ($j = 1, \ldots, N$). A graphical representation of the model can be found in Figure 5.3.1. For ease of explanation, we will assume the dyads to be distinguishable and refer to the first dyad member as females ('F') and second dyad member as males ('M'). Of course, other types of distinguishable dyads are also possible (e.g., brother and sister, oldest and youngest child, etc.), and one can add constraints to the model in order to allow for indistinguishable dyads too (e.g., twin brothers, same-sex couples, etc.) (Gistelinck, Loeys, Decuyper, & Dewitte, 2018; Olsen & Kenny, 2005). In our motivating example, X represents the positive relationship feelings, while Y corresponds to the perceived intimacy. From Figure 5.3.1, it is clear that the LD-APIM is an extension of the cross-sectional APIM towards the longitudinal setting as the latter is fitted at each point of measurement. Obviously, additional features are integrated into the LD-APIM to deal with the statistical challenges inherent to LDD.

First, as the LD-APIM focuses on modeling the association between an antecedent X and an outcome variable Y, one has to acknowledge the difference between the *time-averaged* and *time-specific* effect of overtime or time-varying predictors. If the researcher opts to ignore this difference, the estimated actor and partner effects would portray a mixture of both effects, leading to deceptive conclusions (Enders & Tofighi, 2007). Moreover, there may be unmeasured common causes of the outcome and the predictor as well. For instance in our example, some unmeasured personality traits may affect both the positive relationship feelings and perception of intimacy. As shown by Talloen, Loeys, and Moerkerke (2019), estimators of the time-specific effect on the outcome are unbiased if such separation in effects is made, even in the presence of such unmeasured upper level confounders or measurement error. As a result, we recommend a split-up of the overtime predictors into a time-averaged and time-specific component. For example, in case of an overtime-member predictor X (i.e., a time-varying predictor with different scores for each dyad member, like $PosRel$), the former is obtained by calculating the cluster-mean, that is, by calculating the mean value $X_{F.j}$ and $X_{M.j}$ over all time points for each female and male, respectively. The time-specific component is then computed by cluster-mean centering the original predictor variables: $X_{Fij} - X_{F.j}$ and $X_{Mij} - X_{M.j}$ for each dyad j at each time point i. The LD-APIM model is able to subsume both components as follows:

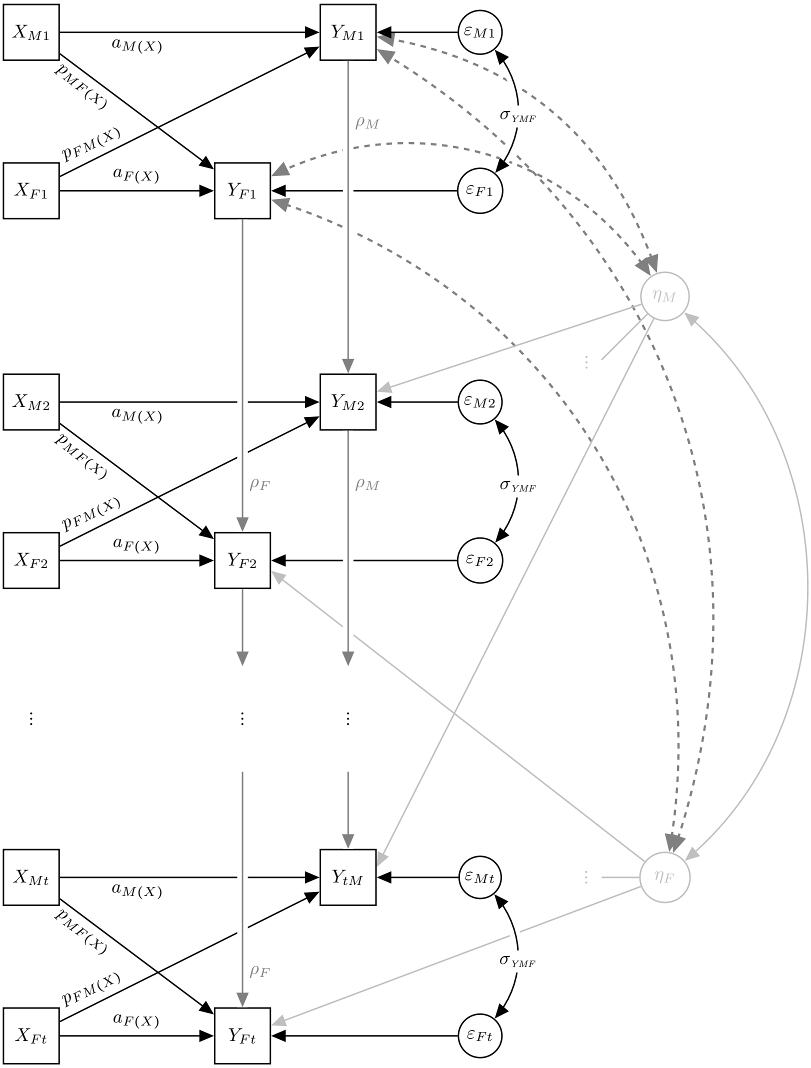


Figure 5.3.1 A graphical representation of the LD-APIM or lagged-dependent APIM.

$$\begin{cases} Y_{Fij} = (\mu_F + \eta_{Fj}) + \rho_F Y_{Fij} + a_{F(XA)} X_{F.j} + p_{MF(XA)} X_{M.j} \\ \qquad\qquad + a_{F(XS)}(X_{Fij} - X_{F.j}) + p_{MF(XS)}(X_{Mij} - X_{M.j}) + \varepsilon_{Fij} \\ Y_{Mij} = (\mu_M + \eta_{Mj}) + \rho_M Y_{Mij} + a_{M(XA)} X_{M.j} + p_{FM(XA)} X_{F.j} \\ \qquad\qquad + a_{M(XS)}(X_{Mij} - X_{M.j}) + p_{FM(XS)}(X_{Fij} - X_{F.j}) + \varepsilon_{Mij} \end{cases}, \tag{5.3}$$

with the indexes XA and XS in the model parameters referring to the time-averaged and time-specific effect of the predictor X, respectively. So, in expression (5.1), the time-averaged components $PosRelA_{Fj}$ and $PosRelA_{Mj}$ correspond to $PosRel_{F.j}$ and $PosRel_{M.j}$, while the time-specific components $PosRelS_{Fj}$ and $PosRelS_{Mj}$ correspond to $(PosRel_{Fij} - PosRel_{F.j})$ and $(PosRel_{Mij} - PosRel_{M.j})$. This split-up procedure is similarly performed in case of overtime-dyad variables (i.e., time-varying predictor with identical scores for both dyad members). Note, as the values of the time-specific component sum up to zero (i.e., $\sum_{i=1}^{T} X_{Mij} - X_{M.j} = \sum_{i=1}^{T} X_{Fij} - X_{F.j} = 0$), the design matrix of the model in the wide format is no longer invertible. In order to avoid this redundancy as implemented in the design matrix, one can replace the effect of the time-specific component at the last time point by the sum of (opposite) effects of the remaining time points.

Second, as mentioned in the introduction, one needs to account for two types of non-independence: the non-independence between the members of the same dyad and the non-independence within (and across) dyad members. Things become even more complex for the latter if one wants to make a distinction between general developmental principles of each dyad member (i.e., *time-invariant* sources of variation or stable traits) and occasion-specific variability (i.e., *time-varying* sources of variation). Within the LD-APIM, interindividual sources of variation are accounted for by including random effects, η_{Fj} and η_{Mj}, for the females and males, respectively. These effects represent random variation around the intercepts μ_F and μ_M, and are allowed to correlate with each other. More specifically, the random intercepts for females and males are assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \eta_{Fj} \\ \eta_{Mj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_F^2 & \tau_{FM} \\ \tau_{FM} & \tau_M^2 \end{pmatrix}\right). \tag{5.4}$$

The parameters τ_F^2 and τ_M^2 are related to the variability in the average outcomes (i.e. the trait) in females and males, respectively. The parameter τ_{FM} encompasses the covariation between the traits of females and their male partners. The latter parameter will thus be informative to answer questions such as (Q6).

Of course, one has to account for intraindividual sources of variation as well. As pointed out by Laurenceau and Bolger (2011), this part of the covariance structure can become quite complex as it must capture sources of variation from the dyad members at a particular time point, including the non-independence between the members of the same dyad, as well as for the autocorrelation within both dyad members. In the LD-APIM, the former source of variation is similarly captured as in the cross-sectional APIM: allow both dyad members to correlate with each other at any time point. In other words, the residuals ε_{Fij} and ε_{Mij} from expression (5.2) for females and males, respectively, are allowed to correlate at each time point i $(i = 1, \ldots, T)$:

$$\begin{pmatrix} \varepsilon_{Fij} \\ \varepsilon_{Mij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_F^2 & \sigma_{FM} \\ \sigma_{FM} & \sigma_M^2 \end{pmatrix}\right), \tag{5.5}$$

with σ_F^2 and σ_M^2 the residual variances for females and males, respectively. The parameter σ_{FM} captures the occasion-specific covariation within dyads, and can be used to address questions like (Q7).

Finally, we discuss the latter source of variation: the autocorrelation. In case one assumes equally spaced observations for both dyad members in which the non-independence between outcome scores fades away as the time lag becomes larger, one can expect adjacent time points to be correlated by a first-order autoregressive process with autocorrelation parameter ρ $(-1 < \rho < 1)$. Considering the first progression discussed in the introduction, this first-order autoregressive process can be integrated into the residual covariance structure Gistelinck and Loeys (2019). However, following the second progression, we want to focus on the autoregressive relationship between concurrent measurements by including lagged dependent variables $Y_{F,i-1,j}$ and $Y_{M,i-1,j}$ as a predictor in the model (Hamaker & Grasman, 2015). Conceptually, this corresponds to conditioning all model parameters on the previous outcome variable, while keeping the residual covariance structure simple. Then, the autocorrelation parameter ρ, which may be different between both dyad members, represents the carryover effect. If ρ is close to zero, then it takes little time for a dyad member to recover from a perturbation and equilibrium (i.e., the return to the average trait) is easily restored. If ρ is close to one, a perturbation will have a substantial effect on the subsequent outcome score. As explained by Gistelinck et al. (2018), conditioning on the lagged dependent outcome also implies that the intercepts μ_F and μ_M from (5.2) no longer reflect the underlying equilibrium, but rather a transformation of it (up to a factor $\frac{1}{1-\rho_F}$ and $\frac{1}{1-\rho_M}$, respectively).

Adding a lagged dependent outcome variables into the model may seem straightforward but encompasses further issues: the initial conditions problem and the endogeneity problem. As expression (5.2) suggests, one needs to define a start-up process for the autoregressive process (Skrondal & Rabe-Hesketh, 2014). If one conditions on the previous outcome score, how can one condition on an unavailable presample response at the first measurement? Assuming the first outcome variable as exogenous (i.e. predetermined) is often considered common practice. Alas, it causes the first measurement to be eliminated from the analysis (Zhang & Nesselroade, 2008). Although this might be of minor concern in case of intensive longitudinal data, it poses an earnest ado in case of a small number of timepoints due to the endogeneity problem. As the underlying trait affects all time points, including the initial time point, erroneous inference is obtained when the first outcome variable is considered to be independent from the random intercept (Achen, 2001). As illustrated by Allison et al. (2017), this bias can be avoided by allowing the first outcome variable to be correlated with the random intercept. In Gistelinck and Loeys (n.d.), several implementations of the multilevel first-order autoregressive model, or ML-AR(1) model, have been compared within different statistical software packages. A model that allows the initial outcome variable to be exogenous while being correlated with the random intercept, yields unbiased estimators for the effect of the lagged dependent outcome (even in case the number of time points is rather small). Such an implementation can easily be achieved within the SEM framework. We adapted these finding towards the dyadic setting: we allowed the random intercept of the males and females to correlate with the first outcome variable of both the males and the females, while the latter are otherwise kept exogenous with their own mean and variance.

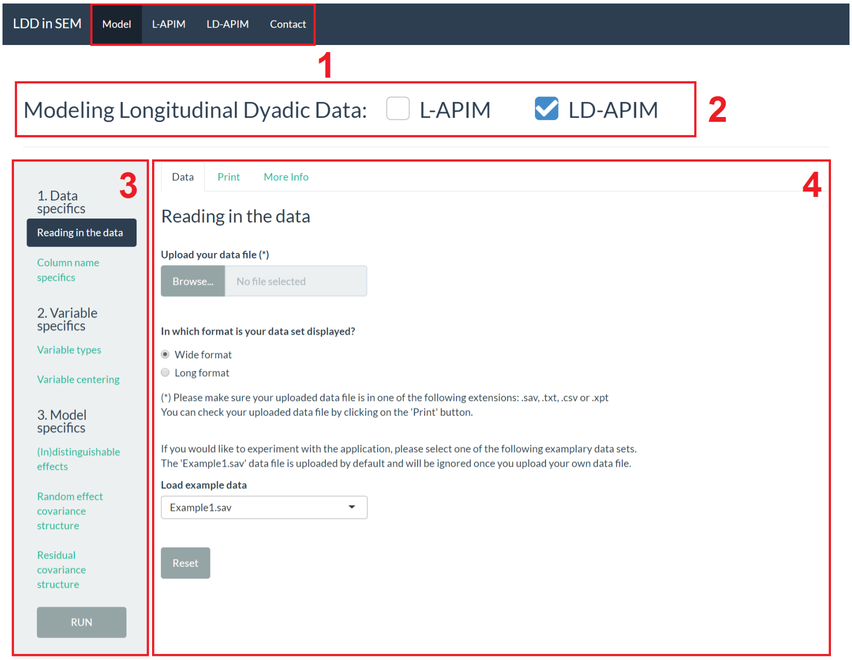
Specifying the above described model with the aforementioned caveats within SEM software is a tedious task. In order to support researchers in their quest for answers, we constructed a Shiny-application within RStudio (RStudio Inc., 2017), called the *LDDinSEM*-application. It is a user-friendly and free web application with a point-and-click interface (Chang, Cheng, Allaire, Xie, & McPherson, 2017). The user does not need any software license nor any specialist knowledge on statistical software. Thanks to the application, researchers can upload their data set and specify the LD-APIM appropriate to their research questions. The app then automatically fits the model on the data set using `lavaan` (Rosseel, 2012) behind the scenes. Afterwards, the app provides the user with summary tables for the model parameters, model-based figures of the effects, as well as the original `lavaan` syntax and the option to download the (transformed) data set. All findings, including tables and figures, are also wrapped into one big summary file, which the user can download. To illustrate the accessibility of the app, we will discuss the different options of the app in much detail on our motivating example in the next section.

5.4 Fitting the LD-APIM using a Shiny-application

5.4.1 General lay-out of the application

We will fit the LD-APIM using the *LDDinSEM*-application, which can be found on "`https://fgisteli.shinyapps.io/Shiny_LDD2/`". In the application, one can distinguish four tabs at the top (see Figure 5.4.1): (a) "Model", where the user can specify the model for his/her LDD, (b) "L-APIM", an information page about the longitudinal APIM without lagged dependent variables, (c) "LD-APIM", an information page about the lagged dependent APIM, and (d) "Contact", a page with contact information of the developer of the application. While the L-APIM is discussed in Gistelinck and Loeys (2019), we focus here on the implementation of the LD-APIM and how one specifies the latter model in the "Model" tab.

As we want to fit the LD-APIM here, one should tick the button of the LD-APIM in the second row in Figure 5.4.1. From the menu on the left, it is clear that the application only needs three steps to implement and fit the LD-APIM: (1) the user uploads the data and adds some information about the data set such that the application gets how the data is structured and what the variable names are, (2) the user specifies the type for all predictor variables in the data and how they should be utilized for the remainder of the analysis, (3) the user specifies the LD-APIM in terms of the mean structure, the random effects covariance structure and the residual covariance structure. Once these three steps are completed, one can click on the ‘RUN’ button and the app will fit the requested LD-APIM. We will now give more information about these three steps below.



Modeling Longitudinal Dyadic Data: ☐ L-APIM ☑ LD-APIM

Reading in the data

Upload your data file (*)

No file selected

In which format is your data set displayed?

● Wide format
○ Long format

(*) Please make sure your uploaded data file is in one of the following extensions: .sav, .txt, .csv or .xpt
You can check your uploaded data file by clicking on the 'Print' button.

If you would like to experiment with the application, please select one of the following examplary data sets.
The 'Example1.sav' data file is uploaded by default and will be ignored once you upload your own data file.

Load example data

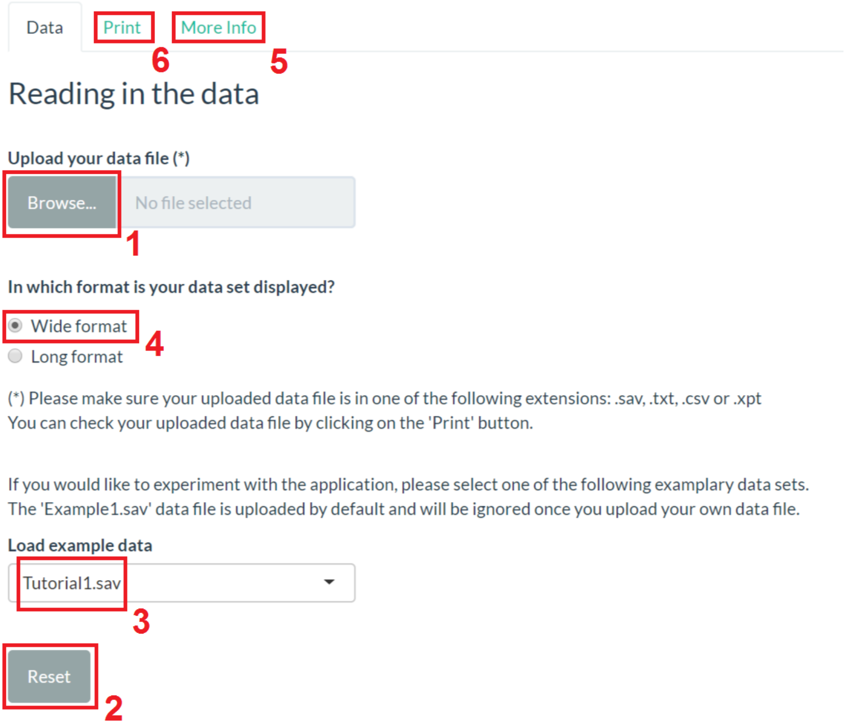
Example1.sav

Figure 5.4.1 The opening page of the *LDDinSEM*-application and its four main areas, (1) refers to the four main pages of which the app exists (page to fit the model, an information page about the L-APIM, an information page about the LD-APIM and a contact page of the developer, (2) allows the researcher to switch between the L-APIM and LD-APIM, (3) corresponds to the three main steps to fit an L(D)-APIM on LDD, (4) displays the current selected step of the application in the third area. The latter area always consists of at least two tabs: the step itself and a tab with more information about the step in consideration.

5.4.2 Step 1: Data specifics

As mentioned above, the first step to fit the LD-APIM is to upload your data set. Researchers can upload their own data set by clicking on the ‘Browse’ button (see label 1 in Figure 5.4.2). In case the user uploaded a wrong data set, it can be removed by clicking on the ‘Reset’ button below (label 2). The application also provides several example data sets (label 3) for the user to experiment or reproduce the finding of the tutorials from the “LD-APIM” tab. To reproduce the analysis on positive feelings on perceived intimacy that we will present below, one has to select the ‘Tutorial1.sav’ data file. Although the underlying program `lavaan` assumes the data set to be structured into the wide format (i.e., one line of information for each dyad), the application allows a long format as well (i.e., one line of information for each dyad member at a particular time point). One just needs to indicate the correct structure of the data set (label 4). One can find more information about the different data formats in the ‘More Info’ tab (label 5). To check whether the data file is correctly uploaded, take a look at the ‘Print’ tab (label 6). Our example data are structured in a wide format. The dependent variable perceived intimacy from day one till day 21 can be found in the columns ‘IntimF1’ to ‘IntimF21’ for females and in the columns ‘IntimM1’ to ‘IntimM21’ for males.

The application assumes the variable names to have a specific format: the dyad member and/or time indices are located at the end of the column name with or without specific separation symbols. In our case, the time range goes from 1 to 21 because the couples were interviewed daily for three consecutive weeks (label 1 at Figure 5.4.3). Note that the application allows to restrict the analysis to a particular time range. For example, if the study included a test period which should not be included in the analysis, these time points can be excluded from the time range. However, the application assumes time points to be consecutive and equally spaced, and the data should at least contain 3 time points for the model to be identified. The dyad member indices represent the label that corresponds to the different roles of the dyad members (label 2). These indices are part of the original variable names and should be reported in the two boxes (random order allowed). In our example, ‘M’ and ‘F’ have to be filled out, referring to males and females, respectively. If one is working with indistinguishable dyads, the labels are necessary for computational matters, but an adapted model can be specified in the third step of the application (see below). At label (3) in Figure 5.4.3, the user has to specify the separation symbols between the variable name and both types of indices. For instance, if column names were constructed as ‘Intim_5.M’, an underscore ‘_’ and a dot ‘ . ’ should be filled out in that order. However in our case, no separation symbols are included in column names, so these boxes are left empty.



Reading in the data

Upload your data file (*)

In which format is your data set displayed?

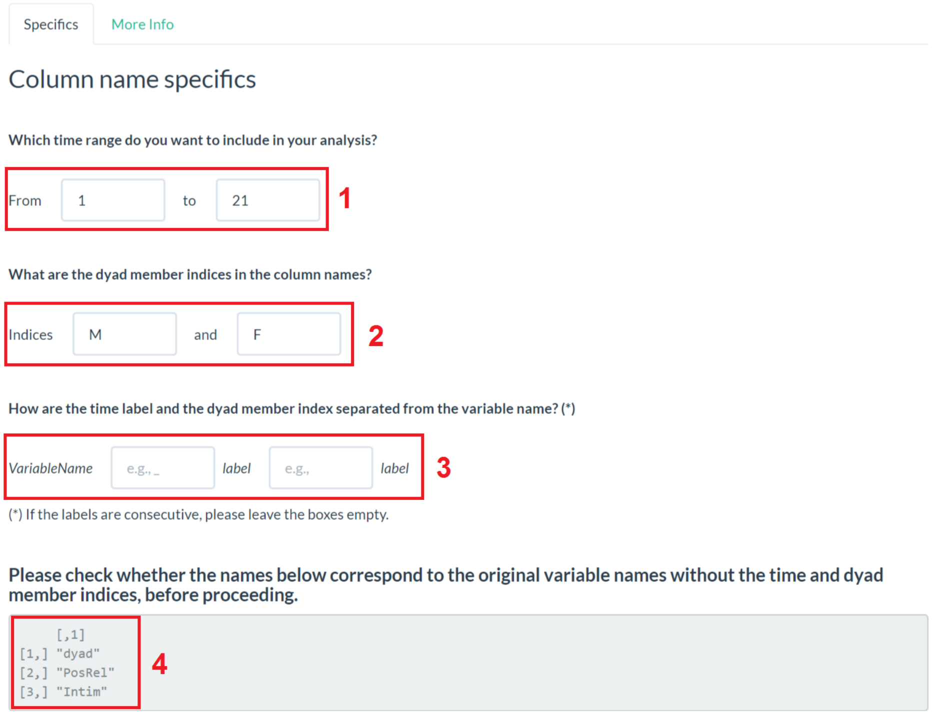
● Wide format
○ Long format

(*) Please make sure your uploaded data file is in one of the following extensions: .sav, .txt, .csv or .xpt
You can check your uploaded data file by clicking on the 'Print' button.

If you would like to experiment with the application, please select one of the following examplary data sets.
The 'Example1.sav' data file is uploaded by default and will be ignored once you upload your own data file.

Load example data: Tutorial1.sav

Figure 5.4.2 The 'Reading in the data' page of the *LDDinSEM*-application with (1) the 'Browse' button to upload the data set, (2) the 'Reset' button to reset the uploaded data, (3) the example data sets, (4) the data format of the uploaded/selected data set, (5) the 'More Info' tab with extra information about the different data formats, and (6) the 'Print' tab to check whether the data was correctly uploaded.



Column name specifics

Which time range do you want to include in your analysis?

From 1 to 21

What are the dyad member indices in the column names?

Indices M and F

How are the time label and the dyad member index separated from the variable name? (*)

VariableName e.g., _ label e.g., label

(*) If the labels are consecutive, please leave the boxes empty.

Please check whether the names below correspond to the original variable names without the time and dyad member indices, before proceeding.

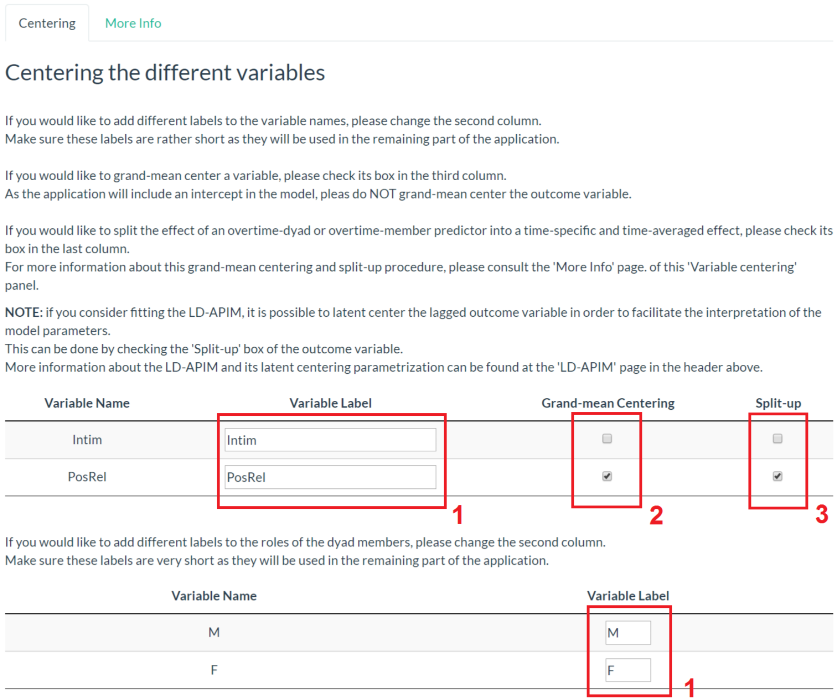
```
     [,1]
[1,] "dyad"
[2,] "PosRel"
[3,] "Intim"
```

Figure 5.4.3 The ‘Column name specifics’ page of the *LDDinSEM*-application with (1) the time range to include in the analysis, (2) the labels of the dyad member roles, (3) the separation symbols in the column names between the variable name and time/dyad member indexes, and (4) the reconstructed variable names by the application.

5.4.3 Step 2: Variable specifics

Once the application read in the data correctly, one has to specify which variables to include in the LD-APIM and what type these variables are. When considering LDD data, one can distinguish four different types of dyadic predictors, depending on whether the variable is measured on the dyad- or member-level and whether the variable is time-invariant or time-varying. More specifically, time-constant-dyad variables correspond to time-invariant variables measured on the dyad-level. The values of these variables are the same for both dyad members and are constant over time (e.g., the season in which a short-period study was performed). Overtime-dyad variables correspond to time-varying variables measured on the dyad-level. The values of these variables are the same for both dyad members and change over time (e.g., the amount of hours the couple spent together at each day). Time-constant-member variables differ between both dyad members, but are constant over time (e.g., the age of each dyad member). Overtime-member variables correspond to time-varying variables measured at the member-level. The values of these variables differ between both dyad members and change over time (e.g., the amount of experienced happiness at each day). It is clear that the predictor *PosRel*, which corresponds to the positive relational feelings, represent an overtime-predictor on member level. Obviously, *Intim* coincides with the dependent variable, and should be added at the corresponding place at the "Variable Types" tab.

Before one can tell the application how to include these variables into the model, the application allows for some preprocessing. In the second column of both tables at the "Variable centering" page depicted in Figure 5.4.4, the user is allowed to abbreviate the variable names and dyad member indexes (label 1). As these new labels will be used in the remainder of the application, this option avoids references that are too long. In the third column (label 2), the user can grand-mean center the variables (except for the outcome variable as the LD-APIM assumes an intercept in the model). In this empirical example, we choose to grand-mean center *PosRel*. In the latter column (label 3), one can opt to split-up the overtime predictors. As it is possible for the time-specific effect of positive relational feelings to differ from its time-averaged effect, it is advised to differentiate both components. The application will refer to these components as *PosRelS* and *PosRelA* for the time-specific and the time-averaged component, respectively.



Centering the different variables

If you would like to add different labels to the variable names, please change the second column.
Make sure these labels are rather short as they will be used in the remaining part of the application.

If you would like to grand-mean center a variable, please check its box in the third column.
As the application will include an intercept in the model, pleas do NOT grand-mean center the outcome variable.

If you would like to split the effect of an overtime-dyad or overtime-member predictor into a time-specific and time-averaged effect, please check its box in the last column.
For more information about this grand-mean centering and split-up procedure, please consult the 'More Info' page. of this 'Variable centering' panel.

NOTE: if you consider fitting the LD-APIM, it is possible to latent center the lagged outcome variable in order to facilitate the interpretation of the model parameters.
This can be done by checking the 'Split-up' box of the outcome variable.
More information about the LD-APIM and its latent centering parametrization can be found at the 'LD-APIM' page in the header above.

Variable Name	Variable Label	Grand-mean Centering	Split-up
Intim	Intim	☐	☐
PosRel	PosRel	☑	☑

If you would like to add different labels to the roles of the dyad members, please change the second column.
Make sure these labels are very short as they will be used in the remaining part of the application.

Variable Name	Variable Label
M	M
F	F

Figure 5.4.4 The 'Variable centering' page of the *LDDinSEM*-application with the possibility to (1) relabel the variables and dyad member indexes, (2) grand-mean center the predictor variables, and (3) split-up the overtime-predictors into a time-averaged and time-specific component.

5.4.4 Step 3: Model specifics

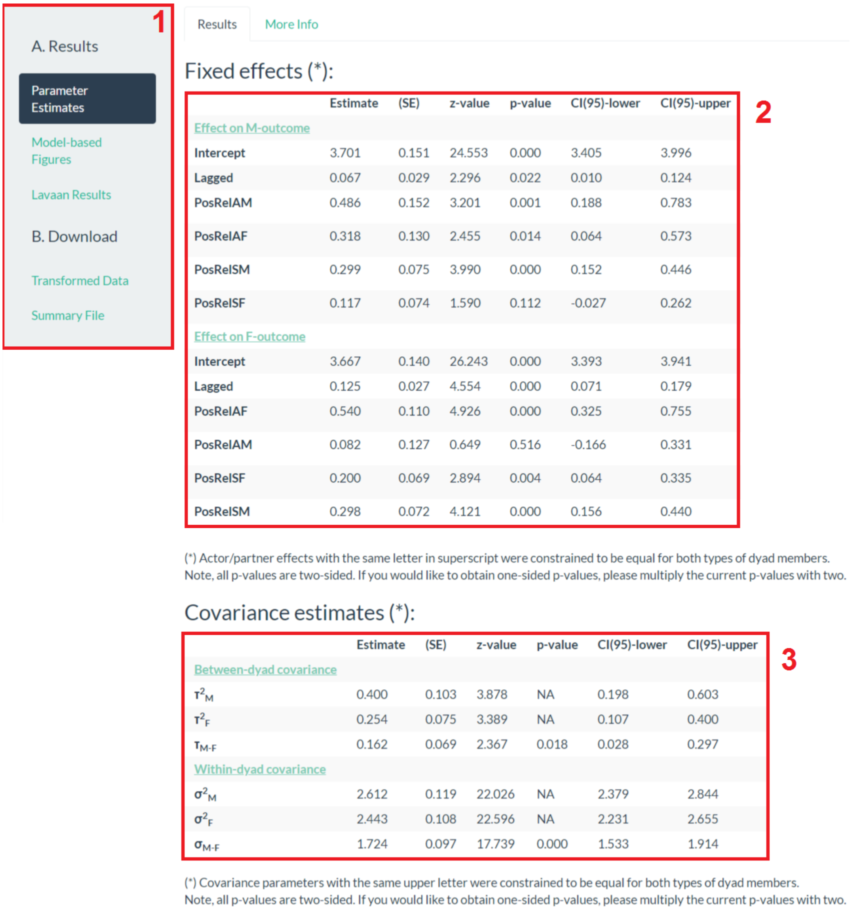
In this final step, the LD-APIM must be specified. First, the application asks to describe the mean structure at the "(In)distinguishable Effects" page. With heterosexual couples, we will allow the intercept of the model to be different for males and females. With indistinguishable dyads, such as same-sex couples, one might opt these intercepts to be the same. Similar decisions have to be made about the actor and/or partner effects of the overtime/time-constant dyad/member predictors. In our empirical example, we will assume all actor and partner effects of both the time-averaged and time-specific component of *PosRel* to be different for males and females, so all these boxes should be checked. The LD-APIM also contains a lagged outcome variable, for which the effect may differ between men and women as well.

Second, the random effects covariance structure has to be defined at the "Random effect covariance structure" page. As we are working with distinguishable dyads, we allow the variation of these random effects to differ between men and women, and to be correlated. As a result, an unstructured covariance structure is most appropriate. As one can see, one could fix this correlation to zero by choosing for the heterogeneous variance component structure. If dyads are indistinguishable, a compound symmetry or a homogeneous variance component structure might be more appropriate. More information about these different covariance structure can be found in the "More Info" tab of this page.

Third, the user is asked to specify the residual covariance structure at a particular time point. Similar to the random effects covariance structure, four options are available depending on whether you want to allow for different variances for both dyad members, and whether you want to allow the residuals to correlate with each other. In our example, we will assume an unstructured covariance structure. This way, we account for the non-independence between the two dyad member at each time point, while we allow the residual variance to be different for men and women.

5.5 Results and Download

With these three steps, the application is able to fit the LD-APIM. Depending on the complexity of the model and the size of the data set, the running-step might take a few minutes. Once the application has fitted the model, the results will be displayed in two parts (see label 1 in Figure 5.5.1). The first part depicts the results of the model estimation, while the second part contains different download options.



Fixed effects (*):

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Effect on M-outcome						
Intercept	3.701	0.151	24.553	0.000	3.405	3.996
Lagged	0.067	0.029	2.296	0.022	0.010	0.124
PosRelAM	0.486	0.152	3.201	0.001	0.188	0.783
PosRelAF	0.318	0.130	2.455	0.014	0.064	0.573
PosRelSM	0.299	0.075	3.990	0.000	0.152	0.446
PosRelSF	0.117	0.074	1.590	0.112	-0.027	0.262
Effect on F-outcome						
Intercept	3.667	0.140	26.243	0.000	3.393	3.941
Lagged	0.125	0.027	4.554	0.000	0.071	0.179
PosRelAF	0.540	0.110	4.926	0.000	0.325	0.755
PosRelAM	0.082	0.127	0.649	0.516	-0.166	0.331
PosRelSF	0.200	0.069	2.894	0.004	0.064	0.335
PosRelSM	0.298	0.072	4.121	0.000	0.156	0.440

(*) Actor/partner effects with the same letter in superscript were constrained to be equal for both types of dyad members.
Note, all p-values are two-sided. If you would like to obtain one-sided p-values, please multiply the current p-values with two.

Covariance estimates (*):

	Estimate	(SE)	z-value	p-value	CI(95)-lower	CI(95)-upper
Between-dyad covariance						
τ^2_M	0.400	0.103	3.878	NA	0.198	0.603
τ^2_F	0.254	0.075	3.389	NA	0.107	0.400
$\tau_{M\text{-}F}$	0.162	0.069	2.367	0.018	0.028	0.297
Within-dyad covariance						
σ^2_M	2.612	0.119	22.026	NA	2.379	2.844
σ^2_F	2.443	0.108	22.596	NA	2.231	2.655
$\sigma_{M\text{-}F}$	1.724	0.097	17.739	0.000	1.533	1.914

(*) Covariance parameters with the same upper letter were constrained to be equal for both types of dyad members.
Note, all p-values are two-sided. If you would like to obtain one-sided p-values, please multiply the current p-values with two.

Figure 5.5.1 The opening page of the *LDDinSEM*-application after hitting the 'RUN' button and its (1) extended menu with the results and download section, (2) the fixed parameter estimates, and (3) the covariance parameter estimates of the LD-APIM.

5.5.1 Parameter estimates

As Figure 5.5.1 shows, the first page in the results section displays the parameter estimates of the LD-APIM for both dyad members. The first table corresponds to the mean structure parameters estimates: the upper part for the first type of dyad member (in our case the males), and the bottom part for the second type of dyad member (in our case the females). In case of indistinguishable dyads, this separation is still made using the arbitrary labels from the "Data specifics" step, but some model parameter estimates might have been fixed to be equal for both dyad member roles. This will be indicated by an identical superscript at the end of the parameter. The table contains, next to the point Estimate, the standard error (SE), the corresponding z-value, *p*-value and the 95%-confidence interval (i.e., CI(95)-lower and CI(95)-upper). Note that all *p*-values mentioned in this table are two-sided.

Based on the first table, we can address the first five research questions. Men and women who report more positive feelings on average (i.e., the effect of $PosRelAM$ on the male outcome and the effect of $PosRelAF$ on the female outcome), also report more perceived intimacy, given the perceived intimacy the day before (Q1). Furthermore, men whose wives report more positive relationship feelings (i.e., the effect of $PosRelAF$ on the male outcome), report more intimacy, given the perceived intimacy the day before (Q2). For both men and women, an increase in the positive relationship feelings on a specific day (i.e., the effect of $PosRelSM$ on the male outcome and the effect of $PosRelSF$ on the female outcome) is associated with higher perceived intimacy on the next day, given the perceived intimacy the day before (Q3). Moreover, an increase in the positive relationship feelings of the men on a specific day (i.e., the effect of $PosRelSM$ on the female outcome) is associated with higher perceived intimacy of the women the next day, given the perceived intimacy the day before (Q4). A similar time-specific partner effect of the women on their male partner was not found to be significant. It is also clear that there is a significant carryover effect for both males and females from the perceived intimacy on one particular day to the next, although the magnitude of this carryover effect is rather moderate (Q5). Note, given that the predictor was grand-mean centered and separated in a time-specific and time-averaged component, the intercepts $\frac{3.70}{1-0.07} = 3.97$ and $\frac{3.67}{1-0.13} = 4.19$ reflect the average perceived intimacy over all days for males and females, respectively, at average positive relational feelings.

The second table at the bottom of the page contains the covariance parameter estimates. The upper part of the table contains the random effects covariance parameters, while the lower part of the table contains the residual covariance parameters. Similar to the first table, the point Estimate, the standard deviation (SE), z-value, *p*-value and 95%-confidence interval (i.e., CI(95)-lower and CI(95)-upper) are included in the table for each covariance parameter. Again, the reported *p*-values are two-sided. Based on this table, we can answer the last two research questions. The random intercept variances actually represent temporal stability and indicates that 95% of the average perceived intimacy (at average levels of the positive relational feelings) lies between $\frac{3.70-1.96*\sqrt{0.40}}{1-0.07} = 2.64$ and $\frac{3.70+1.96*\sqrt{0.40}}{1-0.07} = 5.30$ for males, and between 3.06 and 5.32 for females. The correlation of the average perceived intimacy between males and females of a dyad is $\frac{0.16}{\sqrt{0.40*0.25}} = 0.51$ (Q6). The residual variance within a person of the perceived intimacy on a particular day is similar for males and females: 2.61 for males as opposed to 2.44 for females. The correlation of the daily fluctuations between males and females has a magnitude of $\frac{1.72}{\sqrt{2.61*2.44}} = 0.68$ (Q7). Hence, when the man perceives a lot of intimacy on a specific day, the woman also tends to perceive a lot of intimacy that day, and vice versa. It is worth noting that $\frac{1.72}{1.72+\frac{0.16}{(1-0.13)(1-0.07)}} = 90\%$ of the non-independence between two members of a dyad at a particular day is attributed to the day-level covariance, while the remaining 10% is due to the trait-like or time-stable connection between the two members.

5.5.2 Model-based figures and `lavaan` results

On the "Model-based Figures" page (see, Figure 5.5.2), the user can select for which predictor a model-based figure needs to be displayed. A small descriptive table is also presented for both the outcome and the selected predictor. At the bottom of the page, the effect itself is plotted in a figure. It is clear from this plot that both the actor and partner effects of the time-specific positive relational feeling have a positive effect on perceived intimacy for both the females and the males. However, the partner effects have a higher impact on females than on males.

Figures More Info

Model-based Figures

For which predictor would you like to investigate the effect on the outcome?

PosRel, its time-specific component

Descriptive table of the outcome and predictor of interest.

	Mean	(SD)	Min	Max
IntimM	4.018	1.903	1.000	7.000
IntimF	4.221	1.811	1.000	7.000
PosRelSM	0.000	0.745	-3.650	2.472
PosRelSF	0.000	0.755	-4.157	1.902

Model-based effects of the predictor of interest on the outcome.(*)

(*) These effects should be interpreted as if the other predictors are put to zero. Hence, these figures only have a proper interpretation if all continuous predictors are grand-mean centered. Beware, if you included a predictor to represent the interaction of other predictors, these figure no longer make any sense.

Figure 5.5.2 The ‘Model-based Figures’ page of the *LDDinSEM*-application after hitting the ‘RUN’ button.

For those researchers who are familiar with R and/or `lavaan`, we also provided a page with more technical results, see the “Lavaan Results” page. In the first tab, the model specification of the LD-APIM within `lavaan` is included. We encourage the user to copy/paste this syntax within RStudio and experiment with the model. For instance, if the user wants to allow the intercept to be time-varying, one can simply change the label of the intercept to become different for each time point. For the sake of completeness, we also provided the original `lavaan` output and table. The latter was used to compute the summary tables of the “Parameter estimates” page.

5.5.3 Transformed data and summary file

Before the user can truly experiment with the model, he/she also needs the data set used to fit the model. These can be found at the first page within the "Download" section. This *transformed* data set differs from the original uploaded data set with respect to (a) reshaping in case the data set was not structure in the wide format, (b) relabeling as the user might have relabeled all variable names, (c) indexing as the column names have been adapted to correspond to the model syntax of the LD-APIM from the application, (d) centering as predictors could be grand-mean centered, and (e) split-up as we allowed the user to split-up overtime variables into a time-averaged and a time-specific component.

We also provided a summary file with all the results at the "Summary File" page. More specifically, the file, which can be downloaded as a pdf, a Word-doc or a webpage, contains (a) the changes to the original data set as mentioned above, (b) the LD-APIM expressed in equations and a general graphical representation of the LD-APIM (i.e., not adapted to the particular model expressed by the equations), (c) the tables with the parameter estimates, and (d) all model-based figures.

5.6 Discussion

This paper presents an extension of the cross-sectional APIM towards longitudinal dyadic data (LDD), called the lagged-dependent APIM or LD-APIM. The model tackles different statistical challenges inherent to LDD, but also makes several assumptions that in principle can be relaxed. First, the model assumes that the actor and/or partner effects of the predictors are constant over time. Although it is reasonable to make this assumption in our motivating example, it may be relaxed. This can easily be achieved by adjusting the `lavaan`-syntax from the app accordingly. Second, the model only assumes main effects, but interaction effects could technically be included as well by defining a new variable that corresponds to the interaction values. Third, the LD-APIM introduced here allows for a first-order lagged dependent variable. It is possible that this does not completely resolve the autocorrelation, and that higher order lagged dependent variables need to be included (Wilkins, 2018). Fourth, we did not allow for random slopes, neither for the actor or partner effects, nor for the autocorrelation effect. For example, one might want to allow for person-specific autocorrelation as it represents a person's regulatory weakness

(Hamaker, 2012). In addition, one may want to allow for person-specific residual variances in order to capture differences in sensitivity to perturbations (Jongerling, Laurenceau, & Hamaker, 2015). Unfortunately, these modifications are not straightforward to implement in the current application. Fifth, indistinguishability tests could be added to the application. This way, empirical evidence can be obtained about the difference in roles of the member within a dyad (Gistelinck et al., 2018). Next, although the application uses the SEM framework in order to fit the LD-APIM, the application does not yet allow for latent outcome variables or predictors. For the latter, a latent decomposition rather than a manifest decomposition may then be preferred in order to take measurement error into account (Lüdtke et al., 2008). Lastly, we only considered the LD-APIM in the context of continuous outcome variables. It has already been shown how the cross-sectional APIM can be adapted for binary and count outcomes (Loeys, Cook, De Smet, Wietzker, & Buysse, 2014; Spain, Jackson, & Edmonds, 2012) within the multilevel modeling framework. As suggested by Josephy, Loeys, and Rosseel (2016), diagonally weighted leased squares can be used within the SEM framework in order to allow for categorical outcomes as well. However, further investigation is needed to confirm its performance in the context of LDD.

Despite the limitations and assumptions that we make in the defaults of the application, we hope that the *LDDinSEM*-application will inspire dyadic researchers to contemplate the wealth of research questions that can be addressed with longitudinal dyadic data.

References

Achen, C. (2001). *Why lagged dependent variables can suppress the explanatory power of other independent variables.* Retrieved from `https://www.princeton.edu/csdp/events/Achen121201/achen.pdf` (Presented at the Annual Meeting of the Political Methodology, Los Angeles)

Allison, P., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius: Sociological Research for a Dynamic World*, *3*, 1-17. doi: 10.1177/2378023117710578

Asparouhov, T., Hamaker, E., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359-388. doi: 10.1080/10705511.2017.1406803

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: The Guilford Press.

Campbell, L., & Kashy, D. (2002). Estimating actor, partner, and interaction effects for dyadic data using PROC MIXED and HLM: A user-friendly guide. *Personal Relationships*, *9*(3), 327-342. doi: 10.1111/1475-6811.00023

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). Shiny: Web application framework for R [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=shiny`

Cranford, J., Shrout, P., Iida, M., Rafaeli, R., Ypi, R., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, *32*(7), 917-929. doi: 10.1177/0146167206287721

Dewitte, M., Van Lankveld, J., Vandenberghe, S., & Loeys, T. (2015). Sex in its daily relational context. *Journal of sexual medicine*, *12*(12), 2436-2450. doi: 10.1111/jsm.13050

Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121

Fitzmaurice, G., Laird, N., & Ware, J. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Fitzpatrick, J., Gareau, A., Lafontaine, M.-F., & Gaudreau, P. (2016). How to use the actor-partner interdependence model (APIM) to estimate different dyadic patterns in Mplus: a step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(1), 74-86. doi: 10.20982/tqmp.12.1.p074

Gistelinck, F., & Loeys, T. (n.d.). *Multilevel autoregressive models when the number of time points is small.*

Gistelinck, F., & Loeys, T. (2019). The actor-partner interdependence model for longitudinal dyadic data: An implementation in the SEM framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(3), 329-347. doi: 10.1080/10705511.2018.1527223

Gistelinck, F., Loeys, T., Decuyper, M., & Dewitte, M. (2018). Indistinguishability tests in the actor-partner interdependence model. *British Journal of Mathematical and Statistical Psychology*, *71*(3), 472-498. doi: 10.1111/bmsp.12129

Gonzalez, R., & Griffin, D. (1999). The correlation analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*(4), 449-469. doi: 10.1037/a0025651

Hamaker, E. (2012). Why researchers should think within-person: A paradigmatic rationale. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 43-61). New York, NY: The Guilford Press.

Hamaker, E., & Grasman, R. (2015). To center or not to center? investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. doi: 10.3389/fpsyg.2014.01492

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334-349. doi: 10.1080/00273171.2014.1003772

Josephy, H., Loeys, T., & Rosseel, Y. (2016). A review of R-packages for random-intercept probit regression in small clusters. *Frontiers in Applied Mathematics and Statistics*, *2*(18), 1-13. doi: 10.3389/fams.2016.00018

Kashy, D., & Kenny, D. (2000). The analysis of data from dyads and groups. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personlaity psychology* (p. 451-477). Cambridge, UK: Cambridge University Press.

Kenny, D. (2017, March 27). *DyadR: Web programs for dyadic data analysis.* Retrieved from `http://davidakenny.net/DyadR/DyadRweb.htm`

Kuppens, P., Allen, N., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984-991. doi: 10.1177/0956797610372634

Laurenceau, J.-P., & Bolger, N. (2011). Analyzing diary and intensive longitudinal data from dyads. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 407-422). New York, NY: The Guilford Press.

Loeys, T., Cook, W., De Smet, O., Wietzker, A., & Buysse, A. (2014). The actor-partner interdependence model for categorical dyadic data: a user-friendly guide to GEE. *Personal relationships*, *21*(2), 225-241. doi: 10.111/pere.12028

Lüdtke, O., Marsh, H., Robitzsch, A., Trautwein, U., Asparouhov, R., & Muthén, B. (2008). The multilevel latent covariate model; a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229. doi: 10.1037/a0012869

Olsen, J., & Kenny, D. (2005). Structural equation modeling with interchangeable dyads. *Psychological methods*, *11*(2), 127-141. doi: 10.1037/1082-989X.11.2.127

Reis, H., Collins, W., & Bersheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*(6), 844-872. doi: 10.1037/0033-2909.126.6.844

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. Retrieved from `http://www.jstatsoft.org/v48/i02/` doi: 10.18637/jss.v048.i02

Rovine, M., & Walls, T. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. In T. Walls & J. Schafer (Eds.), *Models for intensive longitudinal data* (p. 124-147). Oxford, UK: Oxford University Press.

RStudio Inc. (2017). *Shiny: Easy web applications in R.* Retrieved from `http://shiny.rstudio.com/`

Singer, S., & Willett, J. (2013). *Applied longitudinal data analysis.* Oxford, UK: Oxford University Press, Inc.

Skrondal, A., & Rabe-Hesketh, S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society, Series C*, *63*(2), 211-237. doi: 10.1111/rssc.12023

Snijders, T., & Bosker, R. (2011). *Multilevel analysis. an introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Spain, S., Jackson, J., & Edmonds, G. (2012). Extending the actor-partner interdependence model for binary outcomes: A multilevel logistic approach. *Personal Relationships*, *19*(3), 431-444. doi: 10.1111/j.1475-6811.2011.01371.x

Stas, L., Kenny, D., Mayer, A., & Loeys, T. (2018). Giving dyadic data analysis away: A user-friendly app for actor-partner interdependence models. *Personal Relationships*, *25*(1), 103-119. doi: 10.1111/pere.12230

Talloen, W., Loeys, T., & Moerkerke, B. (2019). Consequences of unreliability of cluster means and unmeasured confounding on causal effects in multilevel mediation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(2), 191-211. doi: 10.1080/10705511.2018.1533406

Wilkins, A. (2018). To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis. *Political Science Research and Methods*, *6*(2), 393-411. doi: 10.1017/psrm.2017.4

Zhang, Z., & Nesselroade, J. (2008). Bayesian estimation of categorical dynamic factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *42*(4), 729-756. doi: 10.1080/00273170701715998

6

General discussion

Human emotions and behavior are often influenced by the person closest to that human. As a result, dyadic research has become more and more popular in social and behavioral science. When analyzing dyads, it is obvious for the researcher to wonder whether the roles of two members of a dyad are to be considered different or not. It profoundly affects the statistical analysis. Although the question can be answered using theoretical arguments, it is also possible to collect empirical evidence for the degree of indistinguishability. In this dissertation, we took a closer look at these indistinguishability tests (ITs) and examined some related issues that have not been investigated before. We found that (a) REML estimation should be preferred above ML estimation with respect to covariance structure parameters, and (b) sequential tests with multiple testing correction were less liberal than global tests when the amount of dyads are rather small.

Interest often lies in the causes of within-subject changes, while accounting for between-subjects variability related to these changes. In this case, the applied researcher needs to be able to analyze longitudinal dyadic data. Instead of extending longitudinal individual models towards the dyadic setting, we extended a particular cross-sectional dyadic model towards the longitudinal setting. In this dissertation, we opted to extend the APIM towards the longitudinal setting due to its dominant use and applicability within social and behavioral science, for example, when considering diary data that are not centered around a specific event. As a result, a new class of longitudinal dyadic models was introduced, consisting of the L-APIM and LD-APIM. Unlike other longitudinal dyadic models such as the dyadic LGCs, the L-APIM or LD-APIM (further abbreviated as L(D)-APIM) focuses on the effect of an antecedent on the outcome scores. In other words, it emphasizes the influence of a predictor on the outcome variable rather than the trajectory of the outcome variable itself.

The two different extensions of the APIM were obtained based on two major approaches to incorporate temporal correlation in the longitudinal individual setting. The first approach allows the residuals of the model to be correlated with one another. However, this residual covariance structure becomes quite complex in the case of dyadic data. Even in the case where we make the assumption of stable (co)variances over time for both dyad members and a first-order autoregressive process between adjacent time points, most MLM software packages do not support the complex residual covariance structure. Except when using SAS, it is next to impossible to fit this advanced model, referred to as the L-APIM. Consequently, an optimal implementation within the SEM framework was considered. The second approach adds a lagged dependent variable to the model in order to condition on the previous time point. This way the residual covariance structure is kept simple, but other issues now present themselves, such as the initial conditions problem and the endogeneity problem. As these issues are quite complex, we first considered solving them in the absence of any covariates within the individual setting. We found that an implementation within the SEM or BAY framework yields good results provided that the initial value was allowed to correlate with the random intercept, and no auxiliary variable was used for the missing presample response, especially when the number of time points is rather small. These findings were used to resolve similar issues within the dyadic setting and we ended up with the LD-APIM within the SEM framework.

In order to promote these new models, an online user-friendly application was developed, called the *LDDinSEM*-application. It makes the models more accessible to the dyadic researcher, whether their area of interest is educational science, staff management, organizational psychology or social psychology. Moreover, the user does not need software licenses to fit these models and is not required to have any technical knowledge of the software that performs the analysis. Thanks to the point-and-click interface, the user is guided through the different steps of fitting the L(D)-APIM: (a) uploading the data, (b) specifying the different variables in the data, and (c) describing the model itself. After clicking the ‘RUN’ button, the model is fitted for the user, who can download a summary file with all the results. For users who are more familiar with the technical and statistical details, the application also includes information about the model specification. For instance, the app allows the user to consult the original coding. This way, the user can copy and paste the model within RStudio. He or she can then experiment with the model and customize it.

6.1 Limitations

Of course, there are always some limitations involved and related issues to be (re)considered. In chapter 2, we discussed indistinguishability tests for longitudinal dyadic data. However, that chapter's longitudinal dyadic model can be considered as quite simple, even incorrect in a sense, as it ignores time-specific correlation between the different time points and the non-independence between the two members of a dyad. It could be possible that if we reproduce the simulation study for the more complex models such as the L(D)-APIM, the advantage of sequential testing might disintegrate and global testing should be preferred instead. Moreover, we only considered the performance of the tests in terms of the Type I error, but one should consider the power of the test as well. Even though the sequential tests showed lower Type-I error, one still has to investigate whether the power of these tests is acceptable. If so, we may update the current *LDDinSEM*-application, allowing researchers to perform these indistinguishability tests with a simple click on the computer.

Most empirical examples in this dissertation make use of the longitudinal dyadic data provided by Dewitte, Van Lankveld, Vandenberghe, and Loeys (2015). In the previous chapters, we focused mainly on the variables "perceived intimacy" and "positive relationship feelings". As one might have noticed, in the first chapter, the effect of perceived intimacy of the previous day on the current positive relationship feelings was considered, while in chapter 3 and 5, the effect of positive relationship feelings of the previous day on the current amount of perceived intimacy was estimated. As often is the case in social and behavioral science, both directions make sense. If a person feels positive about his or her relationship, that person will show more intimacy towards the partner. Contrariwise, if a person experiences a lot of intimacy from his or her partner, that person may experience more positive feelings towards their relationship. The current design of the L(D)-APIM only allows the researcher to estimate causal effects in which the direction is fixed. Consequently, explanatory variables are assumed to be predetermined. However, one may consider an adaptation of the model similar to the first-order vector autoregression or VAR models from the individual setting (Brandt & Williams, 2007). The adapted L(D)-APIM would not only consider actor and partner effects of the antecedent on the current outcome, but it would also allow the current outcome score of both dyad members to influence their own and their partner's predictor scores of the following time point. This would imply that the predictor and outcome score alternate each other as independent and dependent variable throughout time, which might present some statistical complications that need to be investigated.

Social and behavioral researchers are often interested in theoretical constructs, such as emotions, depression, and trust. These constructs are latent and cannot be directly observed. Over the years, different questionnaires have been developed with good psychometric properties, enabling researchers to scale a person with respect to this construct. For instance in this dissertation, the Dyadic Adjustment Scale was used in order to provide a measurement of the relationship satisfaction. Usually this measurement is obtained by taking the average of the person's responses over different items. For this reason, all variables in the L(D)-APIM are considered to be observed. Statistically speaking, it might be better to consider the construct as latent in order to incorporate measurement error: the averaged scores are not perfect measurements of the construct it represents. This idea already provoked a class of models within the individual setting, called dynamic factor analysis, which has been extrapolated to the dyadic case in some studies (Ferrer & Nesselroade, 2003; Ferrer & Widaman, 2008). In Gareau, Fitzpatrick, Gaudreau, and Lafontaine (2016), the authors showed how to include measurement error in the cross-sectional APIM within the SEM framework. One could adapt the L(D)-APIM in a similar way in order to allow the model for measurement error.

The L(D)-APIMs are in fact first-order models due to the use of the first-order autoregressive structure. They assume that the non-independence between the outcome scores separated by two time points, is explained by the intermediate time point. For instance in the individual setting, the correlation between the outcome scores Y_{M1} and Y_{M3} is explained by Y_{M2}, and because we are working with dyadic data, it also means that the correlation between Y_{M1} and Y_{F3} is explained by Y_{M2} and Y_{F2} (in the case of heterogeneous couples). In other words, the temporal correlation is completely explained by considering the previous outcome measurement. However, in practice this might not always be the case. For instance, in econometrics, higher order lagged dependent variables are often included into the model to further explain the autocorrelation between the residuals (Enders, 2010). In case of the LD-APIM, the model can easily be adapted to include these higher order lagged dependent variables. The statistical issues are similarly resolved for higher orders as in the first-order setting, although, the interpretation might change for the overtime predictors included in the model. In case of the L-APIM, higher order non-independence embodies a higher order moving average process. Its implementation within the SEM framework looks straightforward at first sight, yet, one should be aware of statistical complications and technical issues.

Note that the L(D)-APIMs also form a class of models that assume *weak stationarity*. This implies that the statistical characteristics of the process do not change over time: the mean levels of the variables of the process do not change over time, (co)variances between variables of the process are constant over time, and patterns of temporal dependency only change in function of the time lag. In other words, all model parameters are assumed to be time-invariant and shifting the process over time should not affect its dynamics. However, this assumption might be unrealistic as most human aspects are prone to change as they take place in a bigger environment. For instance, the emotional interaction between two individuals from the start of their marriage till the end of their divorce definitely shows ups and downs due to this life changing event. Consequently, the L(D)-APIM cannot be used to fit such data. Fortunately, it is possible to relax the stationarity assumption by allowing the L(D)-APIM's parameters to become time-dependent (Bringmann, Ferrer, Hamaker, Borsboom, & Tuerlinckx, 2018). Along a similar line of thought, one may argue that the L(D)-APIM does not allow any subject-specific model parameter, except for the latent intercept. The current L(D)-APIM does not allow any random slopes for the explanatory variables, random autocorrelation to represent a person's regulatory weakness, or random residual variances to represent a person's perturbation sensitivity (Jongerling, Laurenceau, & Hamaker, 2015). In either case, more information is needed to implement these random effects within the SEM framework and sample size requirements have to be identified in order to avoid convergence issues within these adaptations.

6.2 Further research

Throughout the dissertation, simulation studies were performed with complete data. However, complete data is hardly available in real life as it often contains missing values. There are in fact three types of missingness: missingness completely at random (i.e., the missing data generating mechanism is independent from the (un)observed measurements), missingness at random (i.e., the missing data generating mechanism solely depends on the observed measurements), and missingness not at random (i.e., the missing data generating mechanism depends on unobserved measurements). The first type of missingness is considered as a minor concern as the model based on the complete cases will still obtain asymptotically unbiased estimates, although less efficient, provided that the model is correctly specified. In the second type of missingness, the missing values pattern is under control and knowledge of the researcher. Statistical analyses based on the likelihood of the observed outcomes, such as the ML estimation performed in the SEM framework, still yield valid results if the variables that explain the missingness are included in the model (Jansen, Van den Troost, Molenberghs, Vermulst, & Gerris, 2006). However, in case of the last type of missingness, things are a bit more complex as the missing values pattern is beyond the control of the researcher. In this case, one talks about informative drop-out. The reason why the missing values occur cannot be explained by the observed variables as they depend on unseen observations. Hence, the missingness embodies extra information that needs to be taken into account. Since the L(D)-APIM does not contain the possibility to include this extra information, biased results may be obtained. More research should be performed to extend the L(D)-APIM to fit the model and the missingness mechanism. For instance, one could adopt the shared-parameter modeling framework from the individual setting (Creemers et al., 2010). Unlike the selection modeling framework or pattern mixture modeling framework which uses conditional models for either the missingness process or the measurement process respectively, the shared-parameter modeling framework uses latent variables or random effects in order to simultaneously drive both the missingness and measurement process. Sensitivity analysis should be performed to explore the influence of missingness not at random on the inference of the L(D)-APIM.

Similarly, the longitudinal dyadic data used in the analysis of this dissertation are assumed to be balanced. The time points are assumed to be equally spaced so that the assumption of a first-order autoregressive structure would make sense. Additionally, all dyads are assumed to be measured on the same occasions. In case of violation of the first data assumption, the time grid can be completed with missing values. In this case, the L(D)-APIM can still be fitted, although estimation will be less efficient. Violation of the second data assumption may be more difficult to solve, especially because of the characteristic use of partner effects within the L(D)-APIM. The interpretation of the partner effects do not make any sense if the antecedent is measured, for example, two days in advance for one dyad, and one day in advance for another dyad. A possible solution is to allow the mean structure parameters of the L(D)-APIM to depend on the time lag between the antecedent and the outcome variable. However, implementing such a model might become quite cumbersome. Moreover, the adapted model might experience convergence issues due to its complexity.

The L(D)-APIM introduced in this dissertation focuses on outcomes that are continuously distributed. Some outcome scores do not follow this assumption. For example, "the effect of relationship dissatisfaction on the decision to masturbate or not" considers a dichotomous outcome variable, or "the effect of coping on the number of fights between the couple" considers counts. As a result, a categorical version of the L(D)-APIM might be practical for researchers as well. There already exists a categorical version of the cross-sectional APIM, albeit in the MLM framework (Loeys, Cook, De Smet, Wietzker, & Buysse, 2014). Hence, extending this categorical APIM towards a categorical L(D)-APIM might be inadequate as the MLM framework lacks to deal with the statistical challenges of LDD. Within the SEM framework, it is suggested that diagonally weighted least squares can be used for categorical outcomes (Josephy, Loeys, & Rosseel, 2016). So, perhaps a similar approach could be used to adapt the L(D)-APIM. Of course, chapter 2 about indistinguishability tests should be reconsidered within the context of categorical data as well. For instance, in the case of Poisson distributed outcomes (i.e., counts), no additional hypotheses about the variance components exist as the latter is completely characterized by the mean components.

When analyzing longitudinal data, research questions can in fact be subdivided into three sets: (a) questions related to the course and shape of the process, (b) questions related to the moderators of change, and (c) questions related to mediators of change (Laurenceau, Hayes, & Feldman, 2007). As the L(D)-APIM focuses on the second type of questions (while controlling for the first type of questions), further research should include the possibility to answer questions of the third type using the L(D)-APIM. Real life data is often prone to mediators: a third variable through which the predictor influences the outcome variable. For instance, intrusive behavior might affect the level of stress of a person through its influence on trust in the partner. Most mediation analysis are performed within cross-sectional models, ergo the APIM has already been adapted to the APIMeM in order to allow for mediators (Ledermann, Macho, & Kenny, 2011). However, the causal relationships between the variables often take time to unfold when considering mediation. As a result, mediation analysis in longitudinal data might provide more reliable inference (Selig & Preacher, 2007). Future research could entail the extension of the L(D)-APIM to incorporate mediators. For the L-APIM, one could extend the APIMeM on a similar way as we extended the APIM towards the L-APIM. For the LD-APIM, the model might be adapted in a similar way as MacKinnon (2008), who extended autoregressive cross-lagged models to include mediators in case of longitudinal individual data. To the best of our knowledge, mediation analysis in the context of longitudinal dyadic data in which the predictor, the mediator and the outcome have been measured at all occasions, can be considered as unknown territory and should be included in future work.

As the L-APIM and LD-APIM were introduced as two prototypes of a new class of longitudinal dyadic models, it might be interesting to contrast both models to one another. Notwithstanding their profoundly different origin (marginal vs conditional), it may still make sense to compare them or combine their characteristics into an overarching model. Hamaker (2005) already compared the use of lagged dependent variables to residuals correlated in a first-order autoregressive manner. However, this comparison was made in the absence of any predictors. It may be interesting to extrapolate their findings to models such as the L(D)-APIM which include explanatory variables. Similarly, this new class of longitudinal dyadic models could be extended with other extensions of cross-sectional dyadic models. Possibly, there will be other statistical issues involved when developing a longitudinal version for the common fate model or mutual influence model. This can be considered in future research.

Note that in this dissertation, time was used as an indicator for the temporal precedence in the causal process. One can also be interested in research questions with a much bigger role for time, for instance, questions related to the within-person acceleration and its association with the state of the process or questions about the degree to which the behavior of the two members fluctuates together. These questions can only be answered in case time is considered as a continuous variable instead of a categorical one, as in differential equation models (Boker & Laurenceau, 2006) or cross-spectral models (Cook, 2003). Future research may entail how to extend these longitudinal dyadic models with explanatory variables.

This dissertation was based on dyadic data, however, the idea that people are correlated with the person closest to him or her can be extrapolated to larger groups of related people. Instead of considering groups consisting of two people, one could consider groups of three people (triads), four people (quadruples), or more. In family research, groups of four people are not uncommon as a family of interest is often restricted to consist of two parents and two siblings. In case everybody rates everyone and everybody is rated by everyone, a social relations model can be used. However, one can also consider a social network analysis if one would like to model groups of related subjects in general (Bringmann et al., 2013). Further research may involve a thorough investigation of these longitudinal network models and their applicability within social and behavioral science.

6.3 Conclusion

In this dissertation, the main content involved the statistical challenges when analyzing longitudinal dyadic data. As distinguishability is one of the major features of dyadic data, we first reconsidered different aspects of indistinguishability tests. We hope that the dyadic researcher becomes more familiar with those tests and the way they are best performed. Whether the dyads are distinguishable or indistinguishable, we also provided two different longitudinal dyadic models based on the APIM: the L-APIM and LD-APIM. The first model uses an advanced residual covariance structure to account for the temporal correlation, while the second model works conditionally on the previous time point by including a lagged dependent variable. Both models have been implemented in a user-friendly free web-based application so users do not need any software license or technical knowledge of the statistical package used to fit the model. Despite the assumptions made by these models, we believe the L-APIM and the LD-APIM to be useful models and we hope that this dissertation will encourage applied dyadic researchers to exploit their longitudinal dyadic data using these models, perhaps via the *LDDinSEM*-application.

References

Boker, S., & Laurenceau, J.-P. (2006). Dynamical systems modeling: An application to the regulation of intimacy and disclosure in marriage. In T. Walls & J. Schafer (Eds.), *Models for intensive longitudinal data* (p. 195-218). New York, NY: Oxford University Press.

Brandt, P., & Williams, J. (2007). *Multiple time series models.* Thousand Oaks, CA: Sage Publications, Inc.

Bringmann, L., Ferrer, E., Hamaker, E., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate behavioral research*, *53*(3), 293-314. doi: 10.1080/00273171.2018.1439722

Bringmann, L., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS ONE*, *8*(4), e96588. doi: 10.1371/journal.pone.0060188

Cook, W. (2003). Quantitative methods for deductive (theory-testing) research on parent-child dynamics. In L. Kuczynski (Ed.), *Handbook of dynamics in parent-child relations* (p. 347-372). Thousand Oaks, CA: Sage Publications, Inc.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., & Kenward, M. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal data. *Biometrical Journal*, *52*(1), 111-125. doi: 10.1002/bimj.200800235

Dewitte, M., Van Lankveld, J., Vandenberghe, S., & Loeys, T. (2015). Sex in its daily relational context. *Journal of sexual medicine*, *12*(12), 2436-2450. doi: 10.1111/jsm.13050

Enders, W. (2010). *Applied econometric time series* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Ferrer, E., & Nesselroade, J. (2003). Modeling affective processes in dyadic relations via dynamic factor analysis. *Emotion*, *3*(4), 344-360. doi: 10.1037/1528-3542.3.4.344

Ferrer, E., & Widaman, K. (2008). Dynamic factor analysis of dyadic affective processes with inter-group differences. In N. Card, J. Selig, & T. Little (Eds.), *Modeling dyadic and interdependent data in the developmental and behavioral sciences* (p. 107-137). New York, NY: Routledge.

Gareau, A., Fitzpatrick, J., Gaudreau, P., & Lafontaine, M.-F. (2016). Analysing, interpreting, and testing the invariance of the actor-partner interdependence model. *The Quantitative Methods for Psychology*, *12*(2), 101-113. doi: 10.20982/tqmp.12.2.p101

Hamaker, E. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*(3), 404-416. doi: 10.1177/0049124104270220

Jansen, I., Van den Troost, A., Molenberghs, G., Vermulst, A., & Gerris, J. (2006). Modeling partially incomplete marital satisfaction data. *Sociological Methods & Research*, *35*(1), 113-136. doi: 10.1177/0049124106289163

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334-349. doi: 10.1080/00273171.2014.1003772

Josephy, H., Loeys, T., & Rosseel, Y. (2016). A review of R-packages for random-intercept probit regression in small clusters. *Frontiers in Applied Mathematics and Statistics*, *2*(18), 1-13. doi: 10.3389/fams.2016.00018

Laurenceau, J.-P., Hayes, A., & Feldman, G. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, *27*(6), 382-695. doi: 10.1016/j.cpr.2007.01.007

Ledermann, T., Macho, S., & Kenny, D. (2011). Assessing mediation in dyadic data using the actor-partner interdependence model. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 595-612. doi: 10.1080/10705511.2011.607099

Loeys, T., Cook, W., De Smet, O., Wietzker, A., & Buysse, A. (2014). The actor-partner interdependence model for categorical dyadic data: a user-friendly guide to GEE. *Personal relationships*, *21*(2), 225-241. doi: 10.111/pere.12028

MacKinnon, D. (2008). *Introduction to statistical mediation analysis.* New York, NY: Erlbaum.

Selig, J., & Preacher, K. (2007). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2-3), 144-164. doi: 10.1080/15427600902911247

7

Summary

In social and behavioral science, human phenomena are often examined within close relationships. How you feel, what you believe, the way you behave: it depends on your own input, but also on the input of the person closest to you. Therefore, observations in social and behavioral science often refer to pairs or dyads rather than single persons. As a result, the popularity of dyadic data is increasing exponentially. Instead of collecting data on one person, the researcher investigates both persons involved in the interpersonal relationship. Standard statistical methods fail due to the natural connection between the two members of a dyad, called non-independence. New statistical models are needed in order to analyze this type of data. Over the last decades, several types of models have been suggested, with the actor-partner interdependence model or APIM as one of the most prominent ones. This model enables the applied researcher to estimate a person's own input (actor effect) and the partner's input (partner effect) on his/her own outcome score, while controlling for the non-independence between both subjects (interdependence). In this dissertation, the goal is to extend the APIM towards the longitudinal setting, that is, in the case dyads are measured repeatedly over time.

Before any statistical analysis is performed on dyadic data, whether cross-sectional or longitudinal, one has to check whether the dyads should be considered as distinguishable (like in heterosexual couples) or indistinguishable (like in same-sex couples). This question can be answered theoretically by identifying a meaningful variable that distinguishes the roles of the members of a dyad (e.g., gender). However, if one wants to answer this question empirically, tests of indistinguishability have to be performed. In this dissertation, we take a closer look at these tests and investigated possible optimization in case the sample size was rather small. We concluded that sequential tests, in which REML estimation was used for the covariance components, might outperform a global test within the SEM framework.

When the interest lies in understanding the intraindividual variation (i.e., understanding the change within a member of a dyad) next to the interindividual variation, the researcher has to measure the dyads repeatedly over time, that is, he or she has to collect longitudinal dyadic data. In this dissertation, the goal was to identify the statistical issues related to longitudinal dyadic data and to try to tackle these complications. There already exist longitudinal dyadic models, albeit with emphasis on the dynamic process of the outcome variable. However, the most interesting research questions in longitudinal dyadic data focus on the influence of a predictor on an outcome variable. For instance, “How does your partner’s intrusive behavior on average affect your relationship satisfaction?”, “Does the way you rate your relationship today depend on the amount of intimacy you perceived yesterday?”, or “Does the way you think your partner understands you influences your happiness more than the way your partner actually understands you?”, these are only a few examples to illustrate the richness of this type of questions.

In order to answer such kind of research questions, a new class of models is needed which focuses on the causal relationship of a predictor and the outcome variable over time. Therefore, we opted to extend the APIM towards the longitudinal setting. This adapted model should deal with the most notorious problems concerning longitudinal dyadic data. First, it has to handle two types of non-independence: the non-independence due to the dyadic nature of the data as well as the non-independence characteristic to the repeated measurements design (autocorrelation). Here, researchers preferably want to make a distinction between time-stable sources of variation (i.e., allowing for random effects) and time-varying sources of variation. Second, it has to include both actor and partner characteristics. Third, it has to be able to disentangle the effect of a time-varying predictor into a time-averaged and time-specific effect.

As one can imagine, the first issue was the most challenging in the APIM’s extension and two different approaches were used. In the first approach, we opted for a marginal extension of the APIM, called the longitudinal actor-partner interdependence model or L-APIM. It is the most natural extension of the cross-sectional APIM as it essentially fits the APIM at each time point. The two types of non-independence mentioned above, were incorporated into the L-APIM by allowing the residuals to be correlated with one another. As this covariance structure became quite complex, an alternative implementation was necessary to fit the model, which was found within the SEM framework. In the second approach, we opted for a conditional extension of the APIM, called the lagged dependent actor-partner interdependence model or LD-APIM. In this approach, the residual covariance structure is kept simple by conditioning on the previous outcome score. This method is quite controversial as it introduces other statistical problems, such as the initial conditions problem, the endogeneity problem and the centering problem. If one conditions on the previous outcome score, one needs a start-up process at the first occasion (initial conditions problem). This start-up process may collide with the estimation of the random effects present in the model and may cause biased inference (endogeneity problem). Moreover, conditioning on the previous outcome score influences the interpretation of the intercept (centering problem). The LD-APIM was developed in order to resolve these issues within the SEM framework.

Both models are implemented in a free web-based Shiny-application with a point-and-click interface. This way, we hope to encourage applied dyadic researchers to use these models in order to exploit their longitudinal dyadic data. The application guides the user throughout the model implementation thanks to the information tab panels at each step. It automatically fits this L(D)-APIM and allows the user to summarize, visualize and download the results. Tutorials are also provided to demonstrate the use of the app. Moreover, the application displays the original code so that more experienced users can relax the defaults of the L(D)-APIM and adapt the L(D)-APIM.

8

Nederlandstalige samenvatting

In sociale wetenschappen en gedragswetenschappen worden menselijke gedragingen vaak bestudeerd binnen het kader van hechte relaties. Hoe een persoon zich voelt, de dingen waarin hij of zij gelooft, de manier hoe hij of zij zich gedraagt, enz. hangen niet alleen af van die persoon zelf, maar ook van de persoon die nauw in relatie staat met hem of haar. Observaties van menselijke gedragingen worden daarom minder beschouwd als eigen aan een individu, maar steeds meer in de context van een paar of dyade. Het is duidelijk dat de populariteit van dyadische data exponentieel toeneemt. In plaats dat de onderzoeker gegevens verzamelt over één persoon, verzamelt deze gegevens over beide personen waaruit de interpersoonlijke relatie bestaat. Standaard statistische methodes slagen er niet in om rekening te houden met de natuurlijk connectie die bestaat tussen de twee leden van een dyade, welke *non-independence* wordt genoemd. Er zijn innovatieve statistische modellen nodig om dergelijke soort gegevens te analyseren. De voorbije jaren werden al verschillende soorten modellen uitgewerkt, met het *actor-partner interdependence model* of APIM als één van de meest vooraanstaande. Het model is in staat om zowel de eigen invloed van een lid van een dyade (*actor effect*) als de invloed van het andere lid van de dyade (*partner effect*) te schatten, rekening houdend met de *non-independence* tussen beide leden (*interdependence*). Het doel van dit proefschrift was om dit model uit te breiden naar de longitudinale context, waarbij dyades meermaals worden geobserveerd over een bepaalde tijdsperiode.

Voordat de analyse van cross-sectionele of longitudinale dyadische data wordt uitgevoerd, dient men eerst na te gaan of de dyades *distinguishable* (zoals bij heteroseksuele koppels) of *indistinguishable* (zoals bij homoseksuele koppels) zijn. Theoretisch kan deze vraag worden beantwoord door een variabele te identificeren die de rollen van beide leden van een dyade op een betekenisvolle manier onderscheiden (zoals geslacht). Empirisch kan deze vraag ook worden beantwoord met behulp van *indistinguishability tests*. In dit proefschrift werden dergelijke testen onderzocht en geoptimaliseerd in het geval dat het aantal tijdspunten relatief klein is. We konden besluiten dat de globale test binnen het SEM framework minder goed presteert dan de stapsgewijze *indistinguishability tests* op basis van REML.

Als de nadruk ligt op zowel de intra-individuele variantie (d.w.z. de verandering binnen een lid van een dyade), alsook op de inter-individuele variantie, dan moeten dyades herhaaldelijk gemeten worden. Dergelijke verzamelde gegevens noemt men longitudinale dyadische data. Het doel van dit proefschrift was om de statistische problemen met dergelijke data te identificeren en de bijhorende complicaties aan te pakken. De bestaande longitudinale dyadische modellen focussen zich vooral op het dynamische proces van de uitkomstvariabele. De meest interessante onderzoeksvragen bij longitudinale dyadische data behandelen echter de invloed van een predictor op de uitkomstvariabele, bijvoorbeeld, "Hoe beïnvloedt het opdringerige gedrag van je partner gemiddeld jouw eigen relatietevredenheid?", "Hangt jouw relatietevredenheid van vandaag af van de mate van intimiteit die je gisteren hebt waargenomen bij je partner?" of "Wordt jouw geluk bepaald door het feit dat jouw partner je begrijpt of door het feit dat je denkt dat jouw partner je begrijpt?". Het zijn slechts enkele voorbeelden die de omvang van dit soort vragen illustreert.

Om dergelijke onderzoeksvragen te beantwoorden, hebben we nood aan een nieuwe klasse modellen die zich focussen op de causale relatie tussen een predictor en een uitkomstvariabele over de tijd heen. We hebben daarom het APIM uitgebreid naar het longitudinale geval. Dit aangepaste model zou in staat moeten zijn om de meest belangrijke problemen in verband met de longitudinale dyadische data aan te pakken. Ten eerste moet het model twee vormen van *non-independence* in rekening brengen: de ene omwille van de dyadische eigenheid van de data en de andere omwille van de herhaalde metingen in de data (ook wel *autocorrelatie* genoemd). Daarenboven willen onderzoekers hierbij een onderscheid maken tussen tijdsonafhankelijke (d.w.z. random effecten toelaten in het model) en tijdsafhankelijke bronnen van variantie. Ten tweede moet het model zowel *actor* als *partner* eigenschappen bevatten. Ten derde moet het model ook in staat zijn om het effect van een tijdsafhankelijke predictor op te splitsen in een effect gemiddeld over de tijd en in een tijd-specifiek effect.

Zoals men zich kan voorstellen, is het eerste probleem het meest uitdagend bij de uitbreiding van het APIM en twee verschillende aanpakken werden gebruikt. In het eerste geval werd er een marginale uitbreiding van het APIM bepaald, ook wel het *longitudinal actor-partner interdependence* *model* of L-APIM genoemd. Het is de meest logische uitbreiding van het cross-sectionele APIM aangezien het model in wezen op elk tijdstip het APIM implementeert. Het L-APIM brengt de twee eerder genoemde vormen van *non-independence* in rekening door de residuen met elkaar te laten correleren. Aangezien de covariantiestructuur hierbij erg complex werd, zochten we naar een alternatieve implementatie om dit model te kunnen fitten binnen het SEM framework. In het tweede geval werd er een conditionele uitbreiding van het APIM bepaald, ook wel het *lagged dependent actor-partner interdependence* of LD-APIM genoemd. Hierbij wordt de covariantiestructuur eenvoudig gehouden door te conditioneren op de vorige uitkomstscore. Deze methode wordt als vrij controversieel gevonden aangezien het andere statistische problemen met zich meebrengt, onder andere het *initial conditions problem*, het *endogeneity problem* en het *centering problem*. Als men op het vorige tijdstip conditioneert, moet men een opstartprocedure definiëren voor het initiële tijdspunt (*initial conditions problem*). Deze opstartprocedure kan botsen met de schatting van de *random effecten* uit het model, waardoor er foute conclusies kunnen gevormd worden (*endogeneity problem*). Meer nog, wanneer er geconditioneerd wordt op het vorige tijdstip, wijzigt de interpretatie van het intercept (*centering problem*). Het LD-APIM werd zodanig geconstrueerd dat deze de bovenstaande problemen aanpakt binnen het SEM framework.

Beide modellen werden gebundeld in een gratis online Shiny-applicatie met een “point-and-click” interface. Op deze manier hopen we toegepaste dyadische onderzoekers aan te moedigen om deze modellen te gebruiken bij het analyseren van hun longitudinale dyadische data. De applicatie begeleidt de gebruiker doorheen de verschillende stappen om het gepaste L(D)-APIM te specifiëren met behulp van informatietabbladen die bij elke stap zijn voorzien. Dit L(D)-APIM wordt automatisch op de data toegepast en de app laat de gebruiker toe om de resultaten samen te vatten, te visualiseren en te downloaden. Er zijn ook handleidingen ter beschikking om het gebruik van de app verder toe te lichten. De applicatie geeft ook de originele code weer zodat meer ervaren gebruikers de modelassumpties kunnen versoepelen en het L(D)-APIM kunnen aanpassen.

9

Data Storage Fact Sheets

Data Storage Fact Sheets Chapter 1

```
1. Contact details
===============================================================

1a. Main researcher
---------------------------------------------------------------
- name: Fien Gistelinck
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: fien.gistelinck@ugent.be


1b. Responsible Staff Member (ZAP)
---------------------------------------------------------------
- name: Prof. dr. Tom Loeys
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: tom.loeys@ugent.be

If a response is not received when using the above contact
details, please send an email to data.pp@ugent.be or contact
Data Management, Faculty of Psychology and Educational
Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.


2. Information about the datasets to which this sheet applies
===============================================================

* Reference of the publication in which the datasets are
  reported:
Gistelinck, F. (2019). Statistical challenges in modeling
longitudinal dyadic data (Doctoral dissertation). Ghent, BE:
```

```
Ghent University.

* Which datasets in that publication does this data sheet
  apply to?
This data storage fact sheet refers to the raw data of the
illustrative examples and corresponding R scripts related to
the first chapter of the reference of the publication.


3. Information about the files that have been stored
=============================================================

3a. Raw data
-------------------------------------------------------------

* Have the raw data been stored by the main researcher?
  [X] YES / [ ] NO
  If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [ ] research group file server
  - [X] other (specify):
The raw data was provided by Céline Hinnekens and Olivia De
Smet, who performed the original experiments, respectively.
As such, they also possess the raw data.

* Who has direct access to the raw data
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [X] other (specify):
The raw data was provided by Céline Hinnekens and Olivia De
Smet, who performed the original experiments, respectively.
As such, they also possess the raw data.
```

```
3b. Other files
-------------------------------------------------------------

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to
    reported results. Specify: ...
  - [ ] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify:
R scripts to analyze the raw data
  - [ ] files(s) containing information about informed
    consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files
    and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [ ] research group file server
  - [ ] other: ...

* Who has direct access to these other files (i.e., without
intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...

4. Reproduction
=============================================================
* Have the results been reproduced independently?:
  [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:
```

Data Storage Fact Sheets Chapter 2

1. Contact details
===

1a. Main researcher

- name: Fien Gistelinck
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: fien.gistelinck@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Prof. dr. Tom Loeys
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: tom.loeys@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
===

* Reference of the publication in which the datasets are reported:

Gistelinck, F., Loeys, T., Decuyper, M. & Dewitte, M. (2018). Indistinguishability tests in the actor-partner interdependence model. British Journal of Mathematical and Statistical Psychology, 71(3), 472-498.

* Which datasets in that publication does this data sheet apply to?

This data storage fact sheet refers to the raw data of the illustrative examples, the generated data of the simulation study and corresponding R scripts related to the reference of the publication.

```
3. Information about the files that have been stored
=============================================================

3a. Raw data
-------------------------------------------------------------
* Have the raw data been stored by the main researcher?
  [X] YES / [ ] NO
  If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [ ] research group file server
  - [X] other (specify):
The raw data was provided by Mieke Decuyper and Marieke
Dewitte, who performed the original experiments of the
illustrative examples. As such, they also possess the
raw data.

* Who has direct access to the raw data
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [X] other (specify):
The raw data was provided by Mieke Decuyper and Marieke
Dewitte, who performed the original experiments of the
illustrative examples. As such, they also possess the
raw data.


3b. Other files
-------------------------------------------------------------

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to
    reported results. Specify: ...
```

```
  - [ ] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify:
R scripts to generate the data for the simulation study,
R scripts to perform the simulation study, and
R scripts to analyze the raw data.
  - [ ] files(s) containing information about informed
    consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files
    and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [ ] research group file server
  - [ ] other: ...

* Who has direct access to these other files
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...


4. Reproduction
=============================================================
* Have the results been reproduced independently?:
  [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:
```

Data Storage Fact Sheets Chapter 3

```
1. Contact details
=============================================================

1a. Main researcher
-------------------------------------------------------------
- name: Fien Gistelinck
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: fien.gistelinck@ugent.be


1b. Responsible Staff Member (ZAP)
-------------------------------------------------------------
- name: Prof. dr. Tom Loeys
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: tom.loeys@ugent.be

If a response is not received when using the above contact
details, please send an email to data.pp@ugent.be or contact
Data Management, Faculty of Psychology and Educational
Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.


2. Information about the datasets to which this sheet applies
=============================================================

* Reference of the publication in which the datasets are
  reported:
Gistelinck, F. & Loeys, T. (2019). The Actor-Partner Inter-
dependence Model for Longitudinal Dyadic Data: An Implemen-
tation in the SEM Framework. Structural Equation Modeling:
A Multidisciplinary Journal, 26(3), 329-347.

* Which datasets in that publication does this data sheet
  apply to?
This data storage fact sheet refers to the raw data of the
empirical example, the generated data of the simulation
study and corresponding R and SAS scripts related to the
reference of the publication.
```

```
3. Information about the files that have been stored
=============================================================

3a. Raw data
-------------------------------------------------------------
* Have the raw data been stored by the main researcher?
  [X] YES / [ ] NO
  If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [ ] research group file server
  - [X] other (specify):
The raw data was provided by Marieke Dewitte, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.

* Who has direct access to the raw data
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [X] other (specify):
The raw data was provided by Marieke Dewitte, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.


3b. Other files
-------------------------------------------------------------

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to
    reported results. Specify: ...
  - [ ] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify:
```

```
R/SAS scripts to generate the data for the simulation study,
R/SAS scripts to perform the simulation study, and
R/SAS scripts to analyze the raw data.
  - [ ] files(s) containing information about informed
    consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files
    and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [ ] research group file server
  - [ ] other: ...

* Who has direct access to these other files
 (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...


4. Reproduction
=============================================================
* Have the results been reproduced independently?:
  [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:
```

Data Storage Fact Sheets Chapter 4

```
1. Contact details
=============================================================

1a. Main researcher
-------------------------------------------------------------
- name: Fien Gistelinck
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: fien.gistelinck@ugent.be


1b. Responsible Staff Member (ZAP)
-------------------------------------------------------------
- name: Prof. dr. Tom Loeys
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: tom.loeys@ugent.be

If a response is not received when using the above contact
details, please send an email to data.pp@ugent.be or contact
Data Management, Faculty of Psychology and Educational
Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.


2. Information about the datasets to which this sheet applies
=============================================================

* Reference of the publication in which the datasets are
  reported:
Gistelinck, F., Loeys, T. & Flamant, N. (n.d.). Multilevel
autoregressive models when the number of time points is small.
Structural Equation Modeling: A Multidisciplinary Journal.

* Which datasets in that publication does this data sheet
  apply to?
This data storage fact sheet refers to the raw data of the
empirical example, the generated data of the simulation
study and corresponding R scripts related to the reference
of the publication.
```

```
3. Information about the files that have been stored
=============================================================

3a. Raw data
-------------------------------------------------------------
* Have the raw data been stored by the main researcher?
  [X] YES / [ ] NO
  If NO, please justify:

* On which platform are the raw data stored?
  - [X] researcher PC
  - [ ] research group file server
  - [X] other (specify):
The raw data was provided by Nele Flamant, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.

* Who has direct access to the raw data
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [X] other (specify):
The raw data was provided by Nele Flamant, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.

3b. Other files
-------------------------------------------------------------

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to
    reported results. Specify: ...
  - [ ] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify:
R scripts to generate the data for the simulation study,
R scripts to perform the simulation study, and
```

```
R scripts to analyze the raw data.
  - [ ] files(s) containing information about informed
    consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files
    and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [ ] research group file server
  - [ ] other: ...

* Who has direct access to these other files
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...


4. Reproduction
=============================================================
* Have the results been reproduced independently?:
  [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:
```

Data Storage Fact Sheets Chapter 5

```
1. Contact details
=============================================================

1a. Main researcher
-------------------------------------------------------------
- name: Fien Gistelinck
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: fien.gistelinck@ugent.be


1b. Responsible Staff Member (ZAP)
-------------------------------------------------------------
- name: Prof. dr. Tom Loeys
- address: H. Dunantlaan 2, 9000 Gent
- e-mail: tom.loeys@ugent.be

If a response is not received when using the above contact
details, please send an email to data.pp@ugent.be or contact
Data Management, Faculty of Psychology and Educational
Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.


2. Information about the datasets to which this sheet applies
=============================================================

* Reference of the publication in which the datasets are
 reported:
Gistelinck, F. \& Loeys, T. (n.d.). Multilevel auto-
regressive models for longitudinal dyadic data. Testing,
Psychometrics, Methodology in Applied Psychology.


* Which datasets in that publication does this data sheet
  apply to?
This data storage fact sheet refers to the raw data of the
empirical example, the generated data of the simulation
study and corresponding R scripts related to the reference
of the publication.
```

```
3. Information about the files that have been stored
=============================================================

3a. Raw data
-------------------------------------------------------------
* Have the raw data been stored by the main researcher?
  [X] YES / [ ] NO
  If NO, please justify: ...

* On which platform are the raw data stored?
  - [X] researcher PC
  - [ ] research group file server
  - [X] other (specify):
The raw data was provided by Marieke Dewitte, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.

* Who has direct access to the raw data
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [X] other (specify):
The raw data was provided by Marieke Dewitte, who performed
the original experiment of the empirical example. As such,
she also possesses the raw data.


3b. Other files
-------------------------------------------------------------

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to
    reported results. Specify: ...
  - [ ] file(s) containing processed data. Specify: ...
  - [X] file(s) containing analyses. Specify:
```

```
R scripts to generate the data for the simulation study,
R scripts to perform the simulation study, and
R scripts to analyze the raw data.
  - [ ] files(s) containing information about informed
    consent
  - [ ] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files
    and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [X] individual PC
  - [ ] research group file server
  - [ ] other: ...

* Who has direct access to these other files
  (i.e., without intervention of another person)?
  - [X] main researcher
  - [ ] responsible ZAP
  - [ ] all members of the research group
  - [ ] all members of UGent
  - [ ] other (specify): ...


4. Reproduction
=============================================================
* Have the results been reproduced independently?:
  [ ] YES / [X] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:
```