frontiers in
PSYCHOLOGY

Quantitative Psychology and Measurement

Measurement invariance within and between subjects: A distinct problem in testing the equivalence of intra- and inter-individual model structures.

Janne Adolf, Noémi Schuurman, Peter Borkenau, Denny Borsboom and Conor Vivian Dolan

Journal Name:	Frontiers in Psychology
ISSN:	1664-1078
Article type:	Original Research Article
Received on:	31 Jan 2014
Accepted on:	24 Jul 2014
Provisional PDF published on:	24 Jul 2014
www.frontiersin.org:	www.frontiersin.org
Citation:	Adolf J, Schuurman N, Borkenau P, Borsboom D and Dolan CV(2014) Measurement invariance within and between subjects: A distinct problem in testing the equivalence of intra- and inter-individual model structures.. *Front. Psychol*. 5:883. doi:10.3389/fpsyg.2014.00883

Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures.

Janne Adolf[1]*, Noémi Schuurman[2], Peter Borkenau[3], Denny Borsboom[4], and Conor V. Dolan[5]

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany
[2]Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, The Netherlands
[3]Personality and Diagnostics Group, Department of Psychology, Faculty of Philosophy I, Martin-Luther-University Halle-Wittenberg, Halle, Germany
[4]Psychological Methods Group, Department of Psychology, Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Biological Psychology, Faculty of Psychology and Education, Free University of Amsterdam, Amsterdam, The Netherlands

*** Correspondence**: Janne Adolf, Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.
adolf@mpib-berlin.mpg.de

Keywords: measurement invariance, ergodicity, state-space modeling, latent variables, intra-individual level of analysis

Abstract

We address the question of equivalence between modeling results obtained on intra-individual and inter-individual levels of psychometric analysis. Our focus is on the concept of measurement invariance and the role it may play in this context. We discuss this in general against the background of the latent variable paradigm, complemented by an operational demonstration in terms of a linear state-space model, i.e., a time series model with latent variables. Implemented in a multiple-occasion and multiple-subject setting, the model simultaneously accounts for intra-individual and inter-individual differences. We consider the conditions – in terms of invariance constraints – under which modeling results are generalizable (a) over time within subjects, (b) over subjects within occasions, and (c) over time and subjects simultaneously thus implying an equivalence-relationship between both dimensions. Since we distinguish the measurement model from the structural model governing relations between the latent variables of interest, we decompose the invariance constraints into those that involve structural parameters and those that involve measurement parameters and relate to measurement invariance. Within the resulting taxonomy of models, we show that, under the condition of measurement invariance over time and subjects, there exists a form of structural equivalence between levels of analysis that is distinct from full structural equivalence, i.e., ergodicity. We demonstrate how measurement invariance between and within subjects can be tested in the context of high-frequency repeated measures in personality research. Finally, we discuss problems of measurement variance in relation to problems of non-ergodicity as currently discussed and approached in the literature.

1 Introduction

Population heterogeneity exists when multiple distinct statistical models are required to adequately describe a population (Muthén, 1989). Statistical approaches to investigate and accommodate heterogeneity include, for instance, multi-group modeling (e.g., Jöreskog, 1971; Muthén, 1989), multi-level modeling (e.g., Hox, 2002), and structural equation mixture modeling (e.g., Dolan 2009). In each of these modeling approaches a heterogeneous population is stratified into subpopulations whose members adhere to the same models and differences within are separated from differences between subpopulations (Muthén, 1989). But how small is the smallest subgroup? One could think of a scenario in which breaking up a heterogeneous population into ever smaller subpopulations leads to the smallest subpopulation that is empirically realizable. This is the individual person (Millsap, 2011). Consider, for instance, the five-factor-model (FFM) which states that the dimensions Extraversion, Neuroticism, Agreeableness, Conscientiousness and Openness to Experience are the major sources of inter-individual differences in personality (McCrae & John, 1992). A researcher studying population heterogeneity can now well question, whether the five-factor-model is generally interpretable in the sense that it holds for each individual member of the overall population by addressing "universal" determinants of human behavior (Hamaker, Dolan, & Molenaar, 2005).

Questions of this kind have indeed been posed recently and have been addressed by means of single subject (N = 1) modeling based on the analysis of repeated measurements over occasions (Cattell, 1952; Gregson, 1983; Molenaar, 1985). By contrasting intra-individual with inter-individual difference data, it has been shown that inter-individual modeling results do usually not generalize to the level of the individual. Rather, individual specifics, which remain undetected in standard large sample modeling techniques, seem to be the rule, not the exception (e.g., Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, in press; Brose, Schmiedek, Lövdén, Molenaar, & Lindenberger, 2010; Hamaker et al., 2005; Hamaker, Nesselroade, & Molenaar, 2007; Keldermann & Molenaar, 2007; Molenaar, 2004; Molenaar & Campbell, 2009; Molenaar, Huizenga, Nesselroade, 2003; Nesselroade, 2010; Schmiedek, Lövdén, & Lindenberger, 2009). The increasing interest in individual modeling techniques therefore emphasizes the conceptual continuity between approaches to heterogeneous populations and to the individual. Explicitly stated, single subject modeling accommodates population heterogeneity in its most extreme sense as it does not necessarily involve the generalization of results to other individuals or subpopulations of individuals. Each individual can thus potentially represent a system that is quantitatively or qualitatively unique (Molenaar, 2004).

We have so far conceived of heterogeneity as heterogeneity between individuals, but one may just as well conceive of heterogeneity as heterogeneity within individuals. That is, an individual's system characteristics may display (higher order) stability or variability over time (Molenaar, 2004). To illustrate this, suppose a researcher aims at describing a person with respect to a certain attribute over time. One may now think of an intra-individual distribution of states rather than of a single trait score. Considered over a representative set of situations, this distribution may have relatively stable characteristics over time, e.g., stable mean and variance. These may then be used to differentiate among people and may thus themselves be regarded as personality characteristics (Fleeson, 2001; Hamaker et al., 2007). However, also within individuals, homogeneity cannot be taken for granted but constitutes a (restrictedly) testable assumption. Similarly to questioning to what extent population models generalize to individual population members, one could question to what extent an individual time series model generalizes to (subsets of) single occasions.

The reorientation towards the individual in differential psychology has been motivated by and motivates an integrative consideration of the within- and the between-subject perspective. It therefore provides an optimal setting to address the following guiding questions: Under what conditions are modeling results generalizable (a) over occasions within subjects, (b) over subjects within occasions, and (c) over occasions and subjects simultaneously? Question (c) refers to the conditions that establish a systematic relationship, i.e., equivalence between the structure of intra- and the structure of inter-individual data (given large N and T). Borrowing terminology from statistical mechanics, this situation is termed *ergodicity* in the psychometric literature (e.g., Molenaar, 2004; Molenaar & Campbell, 2009; Molenaar et al. 2003). In the present context, ergodicity is referred to as a situation in which the statistical behavior of a time series observed for a single subject is the same as the statistical behavior of a sample of multiple subjects, obtained at a few occasions (i.e., the definition of an ergodic process according to Molenaar, 2004, p. 208).

Psychological attributes, however, are often represented as latent variables, the study of which requires psychometric measurement. In the context of latent variable modeling the conditions for an ergodic process decompose into invariance constraints on the structural part of the model and invariance constraints on the measurement model. The latter constraints relate to the concept of measurement invariance (MI; Mellenbergh, 1989; Meredith, 1993; Millsap, 2011). In this paper, we discuss how MI ties into the integrated within- and between-subject context. Specifically, we focus on how the concept is to be considered when one is interested in investigating the generalizability of latent variable modeling results along the dimensions time and subject.

The outline of the paper is as follows. Based on the definition as provided by Mellenbergh (1989), we elaborate on MI in the between- and within-subject context, in general terms and operationally in the linear factor model which lends itself well to integrated modeling, i.e., simultaneous modeling of intra-and inter-individual differences. We then proceed to address our guiding questions using a bottom-up approach. That is, in a multiple-subject, multiple-occasion setting, we set up a linear multi-subject latent variable time series model that accounts for intra-individual and inter-individual variability and we implement the model constraints that imply generalizability of results along the dimensions time and subject. We consider these constraints separately at the level of the measurement process and at the level of the latent psychological process. The result is a taxonomy of differently restrictive models ranging from full heterogeneity to full homogeneity between and within individuals. It can be considered a taxonomy of problems[1] a researcher will potentially face when simultaneously modeling intra- and inter-individual variation. We show that MI holding simultaneously over time and subject can be interpreted as constituting a mode of structural equivalence between the intra- and the inter-individual level of analysis that is distinct from full structural equivalence. Using a real data illustration on intra-individual variability in the personality domain (Borkenau & Ostendorf, 1998), we show how researchers can test for MI over subjects and time. In the discussion, we reconsider the assumptions underlying MI testing and review alternative interpretations of and potential approaches to measurement variance within and between subjects.

2 Measurement invariance between and within subjects

2.1 General definition of measurement invariance

The present focus on MI is motivated by the latent variable paradigm which informs conceptual thinking in modern psychology (Bollen, 2002; Borsboom 2008; Borsboom, Mellenbergh, & van Heerden, 2003; Millsap, 2011). Although not directly observable, an attribute such as agreeableness can be conceptualized as manifesting in terms of observable behaviors or reportable attitudes, in this case along the interpersonal dimensions warmth, kindness, ap-

[1] This useful notion was suggested by an anonymous reviewer.

preciation, and consideration (McCrae & John, 1992; Graziano & Tobin, 2009). However, inferences about latent variables on basis of observed indicators are subject to relatively large uncertainty (Borsboom, 2008). MI is one of the psychometric concepts addressing this uncertainty.

A general formal definition of MI in the latent variable paradigm was given by Mellenbergh (1989). Suppose we have a set of indicators $\boldsymbol{Y}$ that together form a psychometric instrument designed to measure a given latent variable Z, and suppose we have a variable X. MI of the indicators with respect to X is defined as independence of the indicators and X conditional on the latent variable, i.e.,

$$f(\boldsymbol{Y}|Z=z)=f(\boldsymbol{Y}|Z=z,X=x) \tag{1}$$

for all values of Z and X, in which $f(\cdot)$ denotes the probability distribution function. Under MI, any effect of X on the indicators is indirect, i.e., mediated through the latent variable (Lubke, Dolan, Kelderman, & Mellenbergh, 2003b). Consequently, significant differences in observed indicator scores are attributable to differences in the targeted latent variable (Z) across units selected on basis of X, e.g., across persons (e.g., Horn & McArdle, 1992; Lubke et al., 2003b; Mellenbergh 1989; Millsap, 2011; van der Sluis et al., 2006, Wicherts & Dolan, 2010).

To illustrate this, imagine we attempted to measure agreeableness (Z) in a given sample using questionnaire $\boldsymbol{Y}$. Let X be the tendency to respond in a socially desirable manner (Holtgraves, 2004; Paulhus & Reid, 2004). If $\boldsymbol{Y}$ was measurement invariant with respect X, any two individuals from the sample having the same level of agreeableness would attain the same score on each item (apart from measurement error effects). Importantly, they would do so independent of their potentially different tendencies to respond in a socially desirable manner. $\boldsymbol{Y}$ would then be considered unbiased with respect to X. On the contrary, if $\boldsymbol{Y}$ was measurement variant or biased with respect to X, for instance due to item contents triggering socially desirable responding, differences in individual's responses would not necessarily be interpretable as differences in agreeableness. They may as well be interpretable as differences in socially desirable responding. Measurement variance or bias thus refers to a replicable difference in item scores which is not due to the targeted latent variable Z (Millsap, 2011). Meaningful comparisons in terms of the targeted latent variable are thus not guaranteed on basis of biased item scores (e.g., Dolan, Roorda, & Wicherts, 2004; Hamaker, 2007; Raykov, Marcoulides, & Li, 2012).

Moreover, biased items can lead to biased estimates of parameters pertaining to the latent variable (Mellenbergh, 1989; Wicherts & Dolan, 2010).The interpretation of the latent variable is then rendered problematic. The converse argument would be that, if MI across persons selected on basis of X holds, the interpretation of the latent variable is the same across these persons (e.g., Borsboom & Dolan, 2007; Dolan et al., 2004; Horn & McArdle, 1992; Lubke, Dolan, Kelderman, & Mellenbergh, 2003a; Mellenbergh, 1989; Nesselroade Gerstorf, Hardy, & Ram, 2007; Raykov et al., 2012; Wicherts & Dolan, 2010). This notion of MI as *theoretical invariance*, as compared to the above notion of *unbiasedness*, can mainly be found for operationalizations of MI in the linear factor model. It is argued that the interpretation of the factor is determined by its relation to the observed indicators (the factor loadings) and that it is unlikely that different factors are related to a fixed set of indicators in exactly the same way (Lubke et al., 2003a).

Regardless of which interpretational notion is employed, in applying the concept of MI, one has to rely on premises which may appear more or less sensible depending on the context. We get back to this in more detail in the discussion.

2.2 Conceptualization of measurement invariance between and within subjects

MI has been investigated extensively in the context of multi-group factor analysis, with groups defined by nominal between-subject variables, such as sex or ethnic background (e.g., van der Sluis, Posthuma, Dolan, de Geus, Colom, & Boomsma, 2006; Wicherts & Dolan, 2010). Mellenbergh's definition, however, is a general one. It is neutral with respect to the nature and format of the potentially biasing variable, the indicator variables, and latent variables, and is thus independent of the psychometric model that relates the indicators to the latent variables (Lubke et al., 2003a; Mellenbergh, 1989; Meredith, 1993; Wicherts & Dolan, 2010). We can therefore draw two conclusions in the present context. First, Mellenbergh's definition should be equally applicable at the between-subject and at the within-subject level (Borsboom & Dolan, 2007). MI can also be considered with respect to time-varying variables relevant within subjects, such as mood or work pressure. For instance, a questionnaire supposed to assess intra-individual fluctuations in the state agreeableness over time may be biased with respect to mood. Then, a person's series of responses over time would reflect not only variations in the state agreeableness but additionally variations in mood. The second conclusion based on Mellenbergh's general definition is, that it is possible to take a more general perspective and consider MI with respect to subject and time (index) itself. This relates back to our introductory questions.[2]

2.3 Operationalization of measurement invariance between and within subjects

Mellenbergh's general MI definition gives rise to testable model constraints when implemented in the context of a concrete latent variable model. The latent variable modeling framework explicitly distinguishes between a (reflective) measurement model, in which the observed indicators are modeled as a function of the latent variables of psychological interest, and a structural model, which concerns the latent variables and their interrelationships. The linear factor model may be viewed as a proper measurement model in which multiple continuous indicators are linearly regressed upon a single continuous latent variable (e.g., Mellenbergh, 1994). In the linear factor model, MI has been associated with the constraints of strict factorial invariance (strict FI; Meredith, 1993) for the standard between-subject context. However, this measurement model features not only in structural equation modeling at the between-subject level (SEM) but also in state-space modeling of time series data at the within-subject level (SSM; Chow, Ho, Hamaker, & Dolan, 2010; Oud, van den Bercken, & Essers, 1990). We argue that strict FI should be equally applicable at the inter-individual and the intra-individual level. That is, strict FI over (subsets of) subjects within occasions, i.e., subject invariant measurement parameters such as factor loadings, intercepts and residual variances should almost certainly imply MI over subjects within occasions. In addition, strict FI over (subsets of) occasions or time within subjects, i.e., time-invariant measurement pa-

[2] The shift in perspective from MI with respect to specific variables to MI over subjects or time has interesting implications (cf. Meredith, 1993, p. 529, theorem 3). MI over subjects implies MI with respect to any variable that varies exhaustively over subjects within the population considered. Equivalently, and under the assumption of an appropriate sampling rate over time, MI over time implies MI with respect to any variable that varies exhaustively within the period of time considered. Hence, by taking this perspective, one automatically accounts for all measured or unmeasured (discrete and finite) background variables that vary along the dimensions time and subject (cf. Lubke et al., 2003b).

rameters, should almost certainly imply MI over time within subjects for the given sampling rate.[3]

3 A bottom-up approach from full heterogeneity to ergodicity

3.1 The baseline model

We now demonstrate the relation between ergodicity and MI in the context of linear stochastic time series models in state-space format (Chow et al., 2010; Durbin & Koopman, 2001; Hamaker & Dolan, 2009; Hamilton, 1994; Harvey, 1989; Oud, et al., 1990). Such models primarily account for intra-individual variation over time. However, by specifying them within many subjects simultaneously we can extend them to multi-subject models. The conditions under which modeling results are generalizable over time, over subjects, and over time and subjects simultaneously may then be expressed in terms of specific invariance constraints. Furthermore, the state-space format incorporates a measurement model and a latent process model which allows distinguishing among constraints that apply to the measurement parameters and constraints that apply to latent parameters. In the following, subscript *i* and *t* refer to subject and discrete time, respectively. We assume equidistant measurement occasions throughout.

The latent process model is formulated as

$$\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t} \tag{2}$$

where $\boldsymbol{\eta}_{i,t}$ is a $q \times 1$ vector of latent variables, the states, which are regressed on themselves at the previous time point, $\mathbf{B}_{i,t}$ is a $q \times q$ matrix of latent regression parameters capturing the auto- and cross-lagged regression relationships among the states over time, and $\boldsymbol{\alpha}_{i,t}$ is a $q \times 1$ vector of latent regression intercepts. The vector $\boldsymbol{\zeta}_{i,t}$ is a $q \times 1$ vector of latent residuals which are assumed to be multivariate normally distributed with mean zero and covariance matrix $\boldsymbol{\Psi}_{i,t}$. The latent residuals are uncorrelated over time and uncorrelated with $\boldsymbol{\eta}_{i,t-1}$. The model-implied mean vector of the latent states, $\boldsymbol{\nu}_{i,t}$, can be expressed as a function of $\boldsymbol{\alpha}_{i,t}$, $\mathbf{B}_{i,t}$, and $\boldsymbol{\nu}_{i,t-1}$. The model-implied covariance-matrix of the latent states, $\mathbf{P}_{i,t}$, can be expressed as a function of $\mathbf{B}_{i,t}$, and $\mathbf{P}_{i,t-1}$ and $\boldsymbol{\Psi}_{i,t}$. Note that although the formal process is driven by a vector autoregressive process of first order, the actual psychological process needs not obey this structure. This so-called single lag structure renders the model fitting process technically convenient. However, any uni- or multivariate autoregressive moving average model can be accommodated (i.e., reformulated in terms of a first order vector autoregressive process) by extending the state vector by the relevant process components (e.g., Hamaker & Dolan, 2009; Harvey, 1989; Shumway & Stoffer, 2011).

The measurement model is formulated as

$$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \tag{3}$$

where $\boldsymbol{y}_{i,t}$ is a $p \times 1$ vector of manifest indicators, $\boldsymbol{\Lambda}_{i,t}$ is a $p \times q$ matrix of factor loadings and $\boldsymbol{\tau}_{i,t}$ is a $p \times 1$ vector of measurement intercepts. The $p \times 1$ vector $\boldsymbol{\varepsilon}_{i,t}$ contains measurement

[3] Under the assumptions that multivariate normality holds, it is unlikely that variation in measurement error variance and variation in specific factor variance cancel each other out across occasions and subjects respectively, and it is unlikely that variation in measurement intercepts and variation in specific factor means cancel each other out across occasions and subjects respectively (cf. Lubke et al. 2003a; Lubke et al., 2003b; Meredith, 1993).

residuals, ideally measurement errors, which are assumed to be multivariate normally distributed with mean zero and covariance matrix $\mathbf{\Theta}_{i,t}$. The measurement residuals are uncorrelated over time and uncorrelated with $\boldsymbol{\eta}_{i,t}$ and $\boldsymbol{\zeta}_{i,t}$. Here, we additionally assume zero correlations among the measurement residuals, i.e., $\mathbf{\Theta}_{i,t}$ is diagonal, satisfying the assumption of local independence. The model-implied mean vector of the indicators, $\boldsymbol{\mu}_{i,t}$ can be expressed as a function of $\boldsymbol{\tau}_{i,t}$, $\mathbf{\Lambda}_{i,t}$, and $\boldsymbol{\nu}_{i,t}$. The model-implied covariance-matrix of the indicators, $\mathbf{\Sigma}_{i,t}$, can be expressed as a function of $\mathbf{\Lambda}_{i,t}$, and $\mathbf{P}_{i,t}$ and $\mathbf{\Theta}_{i,t}$. As noted, this measurement model is equivalent to the linear factor model as it features in standard between-subject SEM (Chow et al., 2010; Oud et al., 1990).

The model in equations (2) and (3) is our baseline model. Note that the model is completely unrestricted with respect to time and subject, meaning that all model parameters can vary in value over time and subjects, but also that the model structure can be subject- and time-dependent. This concerns the dimensionality of the state vector, the pattern of factor loadings, and in the pattern of interrelationships among latent states and latent residuals. As a consequence, the model-implied covariance matrix, and the model-implied mean vector are subject- and time-dependent. Theoretically, the model does thus accommodate full heterogeneity within and between subjects. We now impose increasingly restrictive invariance constraints relating to the dimensions time and subject. We first consider the model constraints that lead from total heterogeneity to MI over time and subjects. We then consider the additional model constraints that eventually result in full invariance over time and subjects, i.e., an ergodic process, as discussed by Molenaar and colleagues (e.g., Molenaar, 2004; Molenaar & Campbell, 2009).The different models are organized in form of a taxonomy. **Figure 1** represents this taxonomy in terms of model equations and verbal terms. As we are interested in the conditions that establish equivalence between the intra- and inter-individual level of analysis, we focus on those models in which we impose constraints simultaneously within and between subjects.

3.2 Modes of equivalence between the intra- and inter-individual level of analysis

We first consider the baseline model as a reference. As presented in equations (2) and (3) neither the measurement model nor the latent process model is restricted over time or over subjects. Note that, technically, the model is not identified until some sort of time-related pattern is imposed. Assuming some pattern would also be indicated from a theoretical perspective. This needs however not involve constraining (measurement) model parameters to be time-invariant. There is thus no equivalence relationship between the intra- and the inter-individual level. A model based on pooled data over occasions and subjects would address a process that is a mixture over time and subjects unconditional and conditional on the latent process (cf. Muthén, 1989). Applying the interpretation of MI as unbiasedness results in the following conclusions. The absence of MI over time within subjects due to time-varying measurement parameters indicates that within any given person there is systematic observed variability over time that is not attributable to the targeted latent variables in $\boldsymbol{\eta}_{i,t}$. Since MI over subjects within time points does also not hold due to person-specific measurement parameters there is systematic observed variability between persons that is not attributable to the targeted latent variables. Different time- and subject-varying variables may cause measurement variance and these associations may be person- and indicator-specific and may change over time. As long as these (unknown) variables and their effects on the indicators are not accounted for, the interpretation of the latent variables as they develop over time and differ over subjects remains complicated. This is in accordance with the notion of MI as theoretical equivalence which holds that the latent variables in $\boldsymbol{\eta}_{i,t}$ are not necessarily interpretable in an invariant sense over time or subjects. That would become directly apparent in an extreme case, in which the measurement model would display different factor loading patterns over time or subjects. In the discussion, we elaborate on recently suggested strategies to handle and explore such a situation.

By constraining all parameters to be invariant over time and subjects we obtain the extreme opposite. The measurement and process model reduce to

$$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \tag{4}$$

and

$$\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t} \tag{5}$$

with

$$\boldsymbol{\varepsilon}_{i,t} \sim N(0, \boldsymbol{\Theta}),$$

$$\boldsymbol{\zeta}_{i,t} \sim N(0, \boldsymbol{\Psi}).$$

An additional requirement ensuring stationarity of the latent process, i.e. time-invariant process characteristics, is that all eigenvalues of matrix **B** are less than one in absolute value (Hamilton, 1994; Molenaar, 2004). Note that the model-implied distributions of observed and latent variables are now independent of subject and time. This model thus represents an operationalization an ergodic process under the assumption of normality (Molenaar, 2004, p. 208). Under these conditions one (intra-individual) process model generalizes across the entire time span and across all subjects in the population considered, i.e., the individual state-space time series models coincide with a standard between-subject longitudinal factor model based on at least two occasions (Molenaar, 2004; Molenaar et al., 2003). Consequently, the between-subject model provides a description of the intra-individual dynamics of each individual in the population and over the entire period of time considered (e.g., Hamaker et al. 2005; Molenaar, 2004; Molenaar & Campbell, 2009).[4] Pooling over persons and time points is feasible as modeling results are fully generalizable between and within subjects.

Between these two extreme variants is the model in which the invariance constraints only concern the measurement model. Strict FI imposed simultaneously with respect to time and subject implies MI with respect to time and subject and results in the model

$$\mathbf{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \tag{6}$$

and

$$\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t} \tag{7}$$

with

$$\boldsymbol{\varepsilon}_{i,t} \sim N(0, \boldsymbol{\Theta}),$$

[4] If we compared a strictly cross-sectional between-subject structural equation model to an intra-individual state-space time series model, the structural model parameters would necessarily be non-equivalent as soon as lagged relationships are present within the subject (cf. Voelkle, Brose, Schmiedek, & Lindenberger, 2014). Clearly, these can only be estimated based on longitudinal data. However, the model-implied overall covariance and mean-structures of the indicators can still be the same for a cross-sectional between-subject model and a within-subject model.

$$\boldsymbol{\zeta}_{i,t} \sim N(0, \boldsymbol{\Psi}_{i,t}).$$

Note that the conditions for MI over time and subjects concern only the measurement process, that is, invariance of the model parameters over time and subjects conditional on the latent process. Simultaneous MI over time and subjects thus represents a form of structural equivalence between levels of analysis that still allows for substantial heterogeneity with respect to the latent variables and their interrelations over time and over subjects. Consequently, we propose to distinguish between two *modes* of structural equivalence. That is, a mode of measurement equivalence, which involves MI over time and subjects but does not include equivalence of the interrelations among the latent variables and latent residuals, and a distinct mode of full equivalence, which is ergodicity. A model based on data pooled over occasions or subjects would imply a latent process that is a mixture over time and subjects whereas modeling results regarding the measurement process would be generalizable over time and subjects.

Interpreting MI as biasedness of the indicators, this model implies that systematic observed intra-individual as well as inter-individual variability is attributable to the targeted latent variables in $\boldsymbol{\eta}_{i,t}$. The interpretation as theoretical invariance holds that the same latent variables are measured within and between subjects. Systematic within- and between-subject variation can be viewed as variation on the same set of latent variables (cf. Lubke et al., 2003a). The model would thus capture intra-individual dynamics and inter-individual differences therein with respect to the targeted latent variables (cf. Hamaker et al., 2007). In this sense, measurement equivalence could be considered a necessary condition for studying intra- and inter-individual differences pertaining to the latent variables of interest.

4 Illustration

4.1 Purpose of illustration, data description and selection

We show how measurement invariance can be investigated (a) over subjects and (b) over time within a given subject. As we use a modeling approach for stationary time series data we shall limit our illustration to time series models which we assume to be invariant with respect to time. We demonstrate below, that these models allow us to incorporate measurement variance over time to a limited extent.

We use data from Borkenau and Ostendorf (1998) that consist of individual time series of self-ratings on personality items. On 90 successive days, 22 students indicated the degree to which 30 adjectives applied to their daily state. Standard between-subject factor analysis showed that the items measure the inter-individual difference traits Neuroticism, Extraversion, Agreeableness, Conscientiousness and Openness to Experience (e.g., Borkenau & Ostendorf, 1990; Borkenau & Ostendorf, 1998; McCrae & John, 1992). The response format was a 7 point scale with high scores indicating high correspondence between described and perceived state.

For our present illustration, we consider a subset of items and subjects with approximately continuously and normally distributed responses, and the absence of obvious mean-level-trends or variability-changes in the series over time[5]. We focus on three individuals (subjects

[5] We selected subjects based on visual inspection of the frequency distributions and time series plots of their responses. Although the five factor marker items may be considered discrete, they are often treated as continuous in the literature (e.g., Borkenau & Ostendorf, 1998; Hamaker, Dolan, & Molenaar, 2005; Hamaker, Nesselroade, & Molenaar, 2007; Rammstedt & John, 2005). Indeed, Dolan (1994) demonstrated, that treating indicators with at least 7 ordered response

7, 13, and 22), and their responses to the extraversion (“dynamic”, “sociable”, “shy”, “silent”, “lively”, “reserved”) and agreeableness marker items (“selfish”, “good-natured”, “domineering”, “helpful”, “obstinate”, “considerate”). The individual data and descriptive figures are available as supplementary materials.

4.2 Determining the individual state-space time series models

To set up the individual models, we imposed a two-factor measurement model on each individual’s data, such that the extraversion marker items load on one, the agreeableness marker items on a second factor. Note that there is no guarantee that the two-factor model, which would be expected to fit the data in standard inter-individual factor analysis, will fit the individual time series data (e.g., Hamaker et al., 2005; Molenaar, 2004; Molenaar & Campbell, 2009). By means of exploratory factor analysis, one could identify individual factor solutions that would potentially be person-specific (regarding sets of factors and factor loading patterns) and then conduct within-person fit comparisons between the individual models and the two-factor model (e.g., Hamaker et al., 2005; Hamaker et al., 2007). Here, we assume configural invariance over individuals, that is, an invariant number of factors and an invariant factor loading pattern (Meredith, 1993).

We determined the individual process models by modeling the auto- and cross-lagged relationships among the factors using the Fortran program MKF (Dolan, 2010)[6]. This program can fit linear stochastic time series models in state-space format to stationary time series data via the linear, time-invariant Kalman filter algorithm. For correctly specified state-space models the Kalman filter provides optimal estimates of the latent variable states over time and gives rise to ML estimates of the model parameters. Detailed explanations of the estimation procedure can for instance be found in the econometric (e.g., Durbin & Koopman, 2001; Hamilton, 1994; Harvey, 1989) and psychometric literature (e.g., Chow et al., 2010; Oud et al. 1990). Within each individual we contrasted vector auto-regressive processes of first order (VAR(1)), second order (VAR(2)) and of order zero (VAR(0)). In the last case, the factors do not display lagged relationships. We pruned models by fixing to zero non-significant relationships in $\mathbf{B}_{i,t}$ and $\mathbf{\Psi}_{i,t}$ (overall-α=.05). We imposed scaling by fixing the latent intercepts to zero and the latent residual variances to one. The information criteria BIC (Schwarz, 1978) and AIC (Akaike, 1974) served as main indicators for relative model fit but we also conducted Log-Likelihood difference tests where models were nested (α=.05). **Table 1** provides an overview of the results and **Figure 2** shows path diagrammatic representations of the individual models.

According to AIC and BIC, subjects 7 and 22 both display a latent process that involves lagged relationships among the factors. For subject 7 there is only one auto-regressive effect of first order for the agreeableness factor, for subject 22 there is the full set of first- and second-order auto- and cross-lagged regression effects. In case of subject 13 the latent process does not contain any lagged effects among the factors. Within occasions, both factors are correlated within each of the three subjects.

---

categories as continuous, does not affect standard errors and overall test statistics of normal theory maximum likelihood estimation - if the distribution of each indicator is not too skewed. Lubke and Muthén (2004) investigated problematic effects of skewed indicator distributions of pseudo-continuous items in standard confirmatory factor analysis.

[6] The program (including documentation) is available by request from c.v.dolan@vu.nl. All MKF in- and output files for the models fitted are available as supplementary materials. These also include R-code to set up data and input files for MKF, execute MKF, and read MKF output files.

With respect to the individual measurement models, the loadings relating the extraversion indicators to the corresponding factor seem to be relatively homogeneous and reasonably large within each individual (although the measurement residual variances are consistently large). This is different for the agreeableness indicators which are associated not only with more heterogeneous loadings but also with loadings close to zero as in case of the item “helpful”. Especially for subject 7 it is questionable whether one coherent dimension underlies his or her responses to the agreeableness indicators. However, to test this we would have to employ a more explorative approach as outlined above. Note that the loading signs suggest that the factors are inverted in some cases.

Table 1. Comparison of different process models within individuals

Process model	**npars**	**–2LogL**	**AIC**	**BIC**	**χ^2-increase (relative to)**	**df**	**p**
				Subject 7			
VAR(0)	37	1089	1163	1255	10.377(VAR(1))	4	.035
					6.993(VAR(1)*)	1	.008
VAR(1)	41	1079	1161	1263			
VAR(1)*	38	1082	1158	1253	3.384(VAR(1))	3	.336
VAR(2)	45	1095	1185	1297			
				Subject 13			
VAR(0)	37	1522	1596	1689	5.221(VAR(1))	4	0.265
VAR(1)	41	1517	1599	1702			
VAR(2)	45	1515	1605	1718			
				Subject 22			
VAR(0)	37	1212	1286	1378	23.655(VAR(1))	4	.000
VAR(0)*	36	1214	1286	1376	1.815(VAR(0))	1	.178
VAR(1)	41	1188	1270	1373			
VAR(1)*	37	1202	1276	1368	13.366 (VAR(1))	4	.010
VAR(2)	45	1161	1251	1363.7			
VAR(2)*	39	1189	1267	1364.1	27.390 (VAR(2))	6	.000

Note. Model variants denoted with an asterisk are pruned with respect to simultaneous and lagged relationships. The relatively best fitting model according to AIC and BIC is set in italics. χ^2-differences are reported for nested models.

4.3 Addressing MI over subjects

To address MI over subjects we made use of the multi-group modus in MKF treating each individual as a group. FI was then tested via pairwise comparisons between all three subjects. Since we scaled in the latent space by standardizing the conditional latent states, all factor loadings and measurement intercepts are freely estimated and can thus all be subjected to a test of invariance across groups (Raykov et al., 2012). In order to not confound FI constraints with invariance constraints pertaining to the latent level, we freely estimated the latent residual variances in one of the subjects whenever the factor loadings were constrained to equality. Equivalently, we freed the latent intercepts in one of the models, whenever the measurement intercepts were constrained to equality (Raykov et al., 2012; Wicherts & Dolan, 2010). **Table 2** provides an overview of the results.

For all pairwise comparisons between subjects, the AIC and the BIC favored the weakly factorial invariant model. Note that a χ^2-difference-test for instance between the configurally invariant and the strictly factorial invariant model cannot be conducted as the models are not nested. This is due to the freely estimated latent parameters in the strictly factorial invariant model (Raykov et al.; 2012). The finding of subject-invariant factor loadings suggests that the same dimensions underlie the variation within each of the three individuals (Hamaker et al., 2007). These are however not necessarily the dimensions underlying the differences between individuals (Lubke et al., 2003a; Hamaker, 2007) as, according to the fit indices used, uniform bias is likely to be present for at least some of the items. Meaningful comparisons between subjects can be considered feasible as long as they refer to differences in the structure of latent intra-individual variation only. The extent and nature of potential uniform bias between individuals could be the subject of subsequent analyses.

Table 2. Multi-group models with measurement parameters constrained over groups

Measurement models	npars	–2LogL	AIC	BIC	χ^2-increase (relative to)	df	p
Comparison between subjects 7 and 13							
Configural invariance	75	2604	2754	2942			
*Weak FI (**Λ** invariant)*	65	2621	2751	2913			
Strong FI (**Λ, τ** invariant)	55	2797	2907	3044			
Strict FI (**Λ, τ, Θ** invariant)	43	2863	2949	3056	66.087(Strong FI)	12	.000
Comparison between subjects 7 and 22							
Configural invariance	83	2242	2408	2616			
*Weak FI (**Λ** invariant)*	73	2255	2401	2583			
Strong FI (**Λ, τ** invariant)	63	2474	2600	2757			
Strict FI (**Λ, τ, Θ** invariant)	51	2516	2618	2745	42.156(Strong FI)	12	.000
Comparison between subjects 13 and 22							
Configural invariance	82	2684	2848	3053			
*Weak FI (**Λ** invariant)*	72	2701	2845	3025			
Strong FI	62	2787	2911	3066			

($\mathbf{\Lambda}$, $\boldsymbol{\tau}$ invariant)							
Strict FI ($\mathbf{\Lambda}$, $\boldsymbol{\tau}$, $\mathbf{\Theta}$ invariant)	50	6162	6262	6387	3374.630(Strong FI)	12	.000

Note. The relatively best fitting model according to AIC and BIC is set in italics. χ^2-differences are reported for nested models.

4.4 Addressing MI over time

Strict FI over occasions cannot be tested directly, as we confined this illustration to time-invariant models. However, we can investigate whether strict FI over time is violated in a specific sense. We do this by testing for uniform bias of the indicators with respect to a selected time-varying variable *X*. This can be cast in terms of a main-effect of *X* on the indicators additionally to the latent variables (Lubke et al., 2003b).

We extend the time-invariant model for a given individual $i = i^*$ to

$$\boldsymbol{y}_{i*,t} = \boldsymbol{\tau}_{i*} + \mathbf{\Lambda}_{i*}\boldsymbol{\eta}_{i*,t} + \mathbf{\Gamma}_{i*}\mathbf{x}_{i*,t} + \boldsymbol{\varepsilon}_{i*,t} \tag{8}$$

and

$$\boldsymbol{\eta}_{i*,t} = \boldsymbol{\alpha}_{i*,t} + \mathbf{B}_{i*,t}\boldsymbol{\eta}_{i*,t-1} + \mathbf{\Phi}_{i*}\mathbf{x}_{i*,t} + \boldsymbol{\zeta}_{i*,t} \tag{9}$$

where $\boldsymbol{x}_{i*,t}$ is a $r \times 1$ vector of (manifest) variables and $\mathbf{\Gamma}_{i*}$ and $\mathbf{\Phi}_{i*}$ are $p \times r$ and $q \times r$ matrices of regression coefficients. If there is a significant effect of at least one variable in $\boldsymbol{x}_{i*,t}$ on at least one of the indicators, measurement invariance over time would be violated, as - returning to Mellenbergh's definition - the distribution of the indicators is dependent on $\boldsymbol{x}_{i*,t}$ conditional on the latent variables (Lubke et al., 2003b). However, the absence of uniform bias with respect to $\mathbf{x}_{i*,t}$ implies neither MI with respect to this variable (which may still introduce non-uniform bias or be associated with varying measurement residual variances), nor MI with respect to other time-varying variables, let alone MI with respect to time.

We focused on the neuroticism marker item “bad tempered” as a mood indicator and potentially biasing variable in subject 7. The results are shown in **table 3** and the path diagrammatic representation of the corresponding model is displayed in **figure 3**.

The BIC which is more responsive to parsimony than the AIC (Hamaker et al., 2005) favors the model without direct effect of the mood indicator on all indicators and the agreeableness indicators respectively. Both AIC and χ^2-difference test suggest that uniform bias is present for at least one of the indicators. In a given modeling application one could investigate whether uniform bias can be accounted or controlled for also with respect to other potentially biasing covariates. Ultimately however, one needs to decide whether one is willing to discard other forms of bias over time as unlikely or whether actually a modeling approach that incorporates time-varying parameters is the more valid and more interesting alternative. Fitting the “wrong” model to intra-individual data which could be a measurement-invariant or more generally a time-invariant model, will also affect the quality of between-person comparisons. We briefly outline modeling approaches to time-varying dynamics in the discussion.

Table 3. Comparison of models incorporating a potentially biasing variable $\boldsymbol{x}$ for subject 7

Model	**npars**	**–2LogL**	**AIC**	**BIC**	**χ^2-increase (relative to)**	**df**	**p**
$\boldsymbol{y}$, $\boldsymbol{\eta}$ on $\boldsymbol{x}$	52	1010	1114	1244			
$\boldsymbol{\eta}$ on $\boldsymbol{x}$	40	1044	1124	1224	34.250 ($\boldsymbol{y}$, $\boldsymbol{\eta}$ on $\boldsymbol{x}$)	12	.001

$\boldsymbol{y}(a), \boldsymbol{\eta}(a)$ on $\boldsymbol{x}$	45	1034	1124	1237			
$\boldsymbol{\eta}(a)$ on $\boldsymbol{x}$	39	1049	1127	1225	15.061 ($\boldsymbol{y}(a), \boldsymbol{\eta}(a)$ on $\boldsymbol{x}$)	6	.020

Note. $\boldsymbol{y}(a)$ denotes the agreeableness marker items, and $\boldsymbol{\eta}(a)$ denotes the agreeableness factor. We allowed for direct effects of $\boldsymbol{x}$ on the latent variables but did not establish whether these were significant. χ^2-differences are reported for nested models.

5 Discussion

In this paper, we showed how MI (e.g., Mellenbergh, 1989), if present, may facilitate or, if absent, may complicate the generalizability of modeling results within and between subjects. Tying into the ergodicity debate (e.g., Molenaar, 2004), we clarified the relationship between the concepts of MI and ergodicity in the context of general latent variable modeling as well as in a linear multi-subject state-space time series model. We concluded that MI holding simultaneously over time and subjects implies a mode of structural equivalence between the intra- and the inter-individual level of analysis that is distinct from full structural equivalence, i.e., ergodicity. That is, measurement equivalence is a mode of structural equivalence conditional on the latent process. Following common interpretations of measurement invariance, the mode of measurement equivalence could be considered an important condition for integrative latent variable modeling of intra- and inter-individual differences (cf. Ellis & van den Wollenberg, 1993, who stress the importance of local homogeneity in IRT-modeling which is tantamount to measurement equivalence; cf. Millsap, 2011). Using intra-individual time series data from three individuals on daily personality states, we investigated the tenability of MI constraints over subjects and over time. Although strict FI over subjects was absent, the presence of weak FI suggested that between-subject comparisons were feasible with respect to the structure of latent intra-individual variation. We were limited in investigating MI over time due to the time-invariant models we employed. Consequently, we could test for specific MI violations but we did not address unbiasedness with respect to time.

The results of our illustration are in line with a growing body of empirical work investigating potential relationships between the structures of intra- and inter-individual variation and means. So, although we presented measurement equivalence as a less restrictive mode of equivalence between levels of analysis than full structural equivalence, we acknowledge that even this weaker form of structural equivalence may be overly restrictive. We can therefore only stress that the problem of non-ergodicity must in part be viewed as a measurement problem since the violation of measurement invariance with respect to time and subject is a source of heterogeneity within and between individuals (cf. Borsboom, Kievit, Cervone, & Hood, 2009; Nesselroade et al., 2007; Nesselroade et al., 2009). It was the aim of this paper to show that the investigation of measurement related heterogeneity within and between individuals in latent variable modeling qualifies as a problem which is related to but also distinct from the problem of ergodicity.

Regarding a closer examination of measurement related heterogeneity, the presented taxonomy is clearly an abstraction. In practice, the finding of untenable MI constraints is not necessarily the end of an investigation. Modeling application situations falling in the baseline model category and associated problems of measurement variance can be of very different nature. For instance, it may be possible to interpret measurement variance substantively against a given theoretical background (Kelderman & Molenaar, 2007; Millsap & Hartog, 1988). As an example, consider developmental or interventional effects over time, which may manifest as quantitative changes in given parameters, and, more importantly, in changes in the nature or meaning of the psychological entities of interest (Kelderman & Molenaar, 2007, Millsap & Hartog, 1988; Molenaar, 2004; Schmiedek at al., 2009). Also, even if measurement variance is considered a nuisance factor, only a few indicators may display measurement variance. Subsequent analyses may then locate the MI violation in the model and establish whether the number of unbiased indicators is sufficient to proceed with meaningful latent variable modeling, as we have indicated in the illustration (Byrne, Shavelson & Muthén, 1989; Wicherts & Dolan, 2010). Likewise, not all subjects within a sample and not all occasions within a period of time may be affected by measurement variance. It may then be possible to identify intra- or inter-individual variables that explain measurement variance (Mellenbergh, 1989). In the present context, this relates to the concept of conditional equivalence introduced by Voelkle and colleagues (Voelkle et al., 2014). In a simulation study these authors show that full equivalence between inter- and intra-individual model structures can easily be obscured by incorporating single factors that introduce subject- and time-related heterogeneity, e.g., linear mean trends over time, differences between groups of individuals. Conversely, it might be possible to identify such factors for certain constructs and control for them in order to establish conditional equivalence, that is, equivalence for subgroups of individuals and occasions. In case equivalence is well hidden or absent, one can still explore the various types of less restrictive (unconditional) relationships that may arise between intra-individual and inter-individual model characteristics (cf. Brose et al., in press; Kuppens, Allen, & Sheeber, 2010; Montpetit, Bergeman, Deboeck, Tiberio, & Boker, 2010).

These approaches to the links between levels of analysis have yet to be utilized to specifically address measurement variance within and between individuals. To further emphasize why these approaches could be both interesting and necessary given measurement related heterogeneity within and between individuals, let us return to the assumptions, upon which MI is predicated. These concern the existence of the latent variables of interest and the appropriateness of the observed variables as indicators. The first premise holds, that the indicators are – although possibly imperfect, i.e., biased - valid in principle (cf. Meredith, 1964; Meredith, 1993). That is, the indicators are to some extent measuring the variable they were designed to measure (Millsap, 2011) and these psychometric qualities should hold absolutely true or at least hold true for the units of analysis we wish to compare, say, a sample of individuals (Nesselroade, Ram, Gerstorf, & Hardy, 2009). This in turn requires the assumption that the targeted latent variable is indeed given (Mellenbergh, 1989) or a theoretically sensible construct across the selected individuals. As noted by Byrne and Campbell (1999) these premises may be questionable, for instance in applying a measurement instrument in a setting, other than the setting in which it was developed. The setting may be determined by the cultural background of the examinees or the dimension of analysis, e.g., the intra-individual dimension. Hence, a violation of MI with respect to differing setting conditions can be indicative in the following regard. First, it may be that the given test is not valid under some conditions although the latent variable is - on an abstract level - existent or theoretically sensible. The latent variable simply manifests differently under different conditions (e.g., Byrne & Campbell, 1999). Nesselroade and colleagues (2007; 2009) pointed out that a targeted construct (e.g., athletic performance) may be a sensible choice for comparing different individuals – but may require the use of individual-specific indicators ("How well do you play tennis vs. golf?"). Second, a given test may be invalid under certain conditions because the construct is not conceptually sensible across conditions. To label these two scenarios, Byrne and Campbell (1999) refer to the term *construct bias* as opposed to item bias which indicates that the problem has shifted from an "operational" to a "theoretical" problem (Kelderman & Molenaar, 2007, p. 451). The concept of construct bias seems to be highly interesting when contrasting intra- and inter-individual variation. In the light of increasing empirical evidence in favor of substantive individual specifics (e.g., Hamaker et al., 2005; Brose et al., 2010) it raises the following question: To what extent are traditional psychological constructs (and according measurement instruments) that were derived in a between-subject context applicable to intra-individual differences? This is arguably a philosophical question, which has been addressed intensively by Borsboom and colleagues (Borsboom et al., 2003; Borsboom et al., 2009) and by Cervone (2004, 2005). These authors argue that between-subject constructs like extraversion and agreeableness do well in describing inter-individual differences, but are problematic at the level of the individual, where they lack “causal force” (e.g., Cervone, 2004; p. 184). That is, per se, they do not map onto specific psychological mechanisms or processes within the individual, and are thus not suitable to feature as explaining factors in a within-subject model of psychological functioning (Borsboom et al., 2009; van der Maas, Dolan, Grasman, Wicherts, Huizenga, & Raijmakers, 2006). Borsboom et al. (2009) conjecture that there are “infinitely many ways” (p. 88) to achieve a certain outcome on a standard between-subject dimension. The associated constructs thus may lack coherence from an individual-driven perspective, in that they emerge as abstract aggregates only at the level of the population. However, this pessimistic prospect regarding the meaningful application of inter-individual level constructs to the individual can be probed empirically. Millsap employs the term *differential item functioning* rather than the term bias to indicate that “the researcher is unable or unwilling to clearly define the targeted attribute” (Millsap, 2011; p.9). This can be turned into a positive message, namely to explore measurement variance – be it within or between individuals – as a potentially meaningful phenomenon.

An explorative empirical approach to person- and time-related heterogeneity at the level of measurement using the above described strategies and principles can enlighten how measurement instruments that were constructed in the between-subject context function at the within-subject level. This in turn can inform (and be informed by) the elaboration of individual-level concepts and theories (e.g., Cervone, 2005) as well as their implementation in empirical research in terms of operationalizations, measurement devices, and modeling techniques (e.g., Schmiedek et al., 2009). In this sense, it could contribute to building up the theoretical and conceptual foundation that is needed for a true reorientation towards the individual in differential psychology (Molenaar, 2004).

The presented modeling approach has the following limitations, however, that would restrict such an explorative endeavor. First, we based our modeling on the linear, time-invariant Kalman filter and ML estimation which led to time-invariant time series models. Time-varying model parameters can – to some extent – be accommodated using the extended Kalman filter (e.g., Chow & Zhang, 2013; Chow, Zu, Shifren, & Zhang, 2011) or a Bayesian approach (e.g., Del Negro & Otrok, 2008). Second, we employed a multi-group approach, i.e., a two-step procedure to address inter-individual differences in intra-individual dynamics. Inter-individual differences in intra-individual model parameters can be quantified and modeled directly using a Bayesian multi-level approach (e.g., Lodewyckx, Tuerlinckx, Kuppens, Allen, & Sheeber, 2011). Note, however, that multi-group modeling is in principle less restrictive than hierarchical modeling. In the present context, it did not impose any restrictions across individuals apart from applying the same modeling framework to each individual’s data. That is, within individuals, we assumed continuous, normal variables, at the manifest and latent level, which were linearly related to each other. Our reliance on the linear factor model here is expedient, although we are satisfied linear modeling of 7 point scales is adequate. Generalized linear modeling of intra-individual time series to accommodate discrete indicators is possible (cf. van Rijn, Dolan, & Molenaar, 2010), but at present depends on software development. Non-normally distributed continuous indicators (due to nonlinear effects) can be approximated by mixtures of (un-)conditional normal distributions (e.g., Klein & Moosbrugger, 2010). Note that in our case of single-subject models, mixture models return us to time-varying models (Hunter, 2014), which are increasingly discussed in the psychometric literature.

6 Acknowledgments

Janne Adolf thanks her colleagues from the "intra-person behavioral dynamics" project at the Max Planck Institute for Human Development for their helpful comments and discussion input.

Janne Adolf is a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE, www.imprs-life.mpg.de; participating institutions: MPI for Human Development, Freie Universität Berlin, Humboldt-Universität zu Berlin, University of Michigan, University of Virginia, University of Zurich).

Conor V. Dolan is support by the European Research Council (Genetics of Mental Illness: ERC-230374 awarded to Dorret I Boomsma).

7 References

Akaike, H. (1974). A new look at statistical model identification. *IEEE transactions on Automatic Control, 19*, 716-723. doi 10.1109/TAC.1974.1100705

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634. doi: 10.1146/annurev.psych.53.100901.135239

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and individual differences, 11*, 515–524. doi: 10.1016/0191-8869(90)90065-Y

Borkenau, P., & Ostendorf, F. (1998). The big five as states: How useful is the five factor model to describe inraindividual variations over time? *Journal of Research in Personality, 32*, 202–221. doi: 10.1006/jrpe.1997.2206

Borsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25–53. doi: 10.1080/15366360802035497

Borsboom, D., & Dolan, C. V. (2007). Theoretical equivalence, measurement, invariance, and the idiographic filter. *Measurement, 5*, 236–263. doi: 10.1080/15366360701765020

Borsboom, D., Kievit. R. A., Cervone, D., & Hood, B. S. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In: J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudary (Eds.). *Developmental process methodology in the social and developmental sciences*, pp. 67–97. New York: Springer.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*, 203–219. doi: 10.1037/0033-295x.110.2.203

Brose, A., Schmiedek, S., Lövdén, M., Molenaar, P. C. M., & Lindenberger, U. (2010). Adult age differences in covariation of motivation and working memory performance: Contrasting between-person and within-person findings. *Research in Human Development, 7(1)*, 61–78. doi: 10.1080/15427600903578177

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, U., Schmiedek, F. (in press). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Psychology*. doi: 10.1002/per.1961

Byrne, B. M. & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology, 30*, 555–574. doi: 10.1177/0022022199030005001

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105(3)*, 456–466. doi: 10.1037/0033-2909.105.3.456

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

Cervone, D. (2004). The architecture of personality. *Psychological Review, 111*, 183–204. doi: 10.1037/0033-295X.111.1.183

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology, 56*, 423–452. doi: 10.1146/annurev.psych.56.091103.070133

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

Chow, S.-M. & Zhang, G. (2013). Non-linear regime-switching state-space (rsss) models. *Psychometrika, 78(4)*, 740–768. doi: 10.1007/S11336-013-9330-8

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

Del Negro, M. & Otrok, C. (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. *Federal Reserve Bank of New York Staff Reports*, No. 326.

Dolan, C. V. (2004). Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, 47, 309–326. doi: 10.1111/j.2044-8317.1994.tb01039.x

Dolan, C. V. (2009). Structural equation mixture modeling. In: R. E. Millsap, A. Maydeu-Olivares (Eds.). *The Sage handbook of quantitative methods in psychology*, pp. 568–591. Thousand Oaks, CA Sage Publications Ltd.

Dolan, C. V. (2010). *MKFM6: Multi-group, multi-subject stationary time series modeling based on the Kalman filter*.

Dolan, C. V., Roorda, W., Wicherts, J. M. (2004). Two failures of spearman's hypothesis: The GATB in holland and the JAT in south africa. *Intelligence, 32*, 155–173. doi: 10.1016/j.intell.2003.09.001

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

Ellis, J. L., & van den Wollenberg, A. L. (1993). Local homogeneity in latent trait models: A characterization of the homogeneous monotone IRT model. *Psychometrika*, *58*, 417–429. doi: 10.1007/bf02294649

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 89(6)*, 1011–1027. doi: 10.1037/0022-3514.80.6.1011

Graziano, W. G. & Tobin, R. M. (2009). Agreeableness. In: M. R. Leary and R. H. Hoyle (Eds). *Handbook of individual differences in social behavior*, pp. 46–61. New York, NY, US: Guilford Press.

Gregson, R. A. M. (1983). *Time series in psychology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. Inc.

Hamaker, E. L. (2007). How to inspect fruit. *Measurement, 5 (4)*, 250–253. doi: 10.1080/15366360701775961

Hamaker, E. L., & Dolan, C. V. (2009). Idiographic data analysis: Quantitative methods - from simple to advanced. In J. Valsiner, P. C. M. Molenaar, M. Lyra and N. Chaudhary (Eds). *Dynamic Process Methodology in the Social and Developmental Sciences*, pp. 191–216. New York: Springer-Verlag.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: Raw data maximum likelihood. *Structural Equation Modeling, 10*, 352–379. doi: 10.1207/s15328007sem1003_2

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

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait-state model. *Journal of Research in Personality, 41*, 295–315. doi: 10.1016/j.jrp.2006.04.003

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

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

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30(2)*, 161–172. doi: 10.1177/0146167203259930

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18(3)*, 117–144. doi: 10.1080/03610739208253916

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

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

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426. doi: 10.1007/BF02291366

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

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65(4)*, 457–474. doi: 10.1007/BF02296338

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

Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B., & Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *Journal of Mathematical Psychology, 55(1)*, 68–83. doi: 10.1016/j.jmp.2010.08.004

Lubke, G. H., Dolan, C V., Kelderman, H., & Mellenbergh, G. J. (2003 a). On the relationship between- and within group differences and measurement invariance in the common factor model. *Intelligence, 31*, 543–566. doi: 10.1016/s0160-2896(03)00051-5

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003 b). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology, 56*, 231–248. doi: 10.1348/000711003770480020

Lubke, G. H. & Muthén, B. O. (2004) Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11*, 514–534. doi: 10.1207/s15328007sem1104_2

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127–143. doi: 10.1016/0883-0355(89)90002-5

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29*, 223–236. doi: 10.1207/s15327906mbr2903_2

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29(2)*, 177–185. doi: 10.1007/BF02289699

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. doi: 10.1007/BF02294825

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73(3)*, 574–584. doi: 10.1037/0021-9010.73.3.574

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50,* 181–202. doi: 10.1007/bf02294246

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

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

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology,* pp. 339–360. Dordrecht: Kluwer Academic Publishers.

Montpetit, M. A., Bergeman, C. S., Deboeck, P. R., Tiberio, S. S., & Boker, S. M. (2010). Resilience-as-process: Negative affect, stress, and coupled dynamical systems. *Psychology and Aging, 25*, 631–640. doi: 10.1037/a0019268

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54,* 557–585. doi: 10.1007/BF02296397

Nesselroade, J. R. (2010). On an emerging third discipline of scientific psychology. In P. C. M. Molenaar & K. M. Newell (Eds). *Individual Pathways of Change: Statistical Models for Analyzing Learning and Development*, pp. 209–218. Washington, DC: American Psychological Association.

Nesselroade, J. R., Gerstorf, D., Hardy, S. A., & Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement, 5*, 217–235. doi: 10.1080/15366360701741807

Nesselroade, J. R., Ram, N., Gerstorf, D., Hardy, S. A. (2009). Rejoinder to commentaries on Nesselroade, Gerstorf, Hardy, and Ram. *Measurement*, *7*, 17–26. doi: 10.1080/15366360802715361

Oud, J. H. L., van den Bercken, J. H., & Essers, R. J. (1990). Longitudinal factor score estimation using the Kalmanfilter. *Applied Psychological Measurement*, *14*, 395–418. doi: 10.1177/014662169001400406

Paulhus, D. L. & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60(2)*, 307–317. doi: 10.1037/0022-3514.60.2.307

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit. Diagnostica, 51, 195–206. doi: 10.1026/0012-1924.51.4.195

Raykov, T., Marcoulides, G. A., & Li, C. H. (2012). Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Eduational and Psychological Measurement, 72(6)*, 954–974. doi: 10.1177/0013164412441607

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging, 24*, 841–857. doi: 10.1037/a0017799

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. doi: 10.1214/aos/1176344136

Shumway, R. S., & Stoffer, D. S. (2011). *Time series analysis and its applications: With R examples*. New York: Springer.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A Dynamical Model of General Intelligence: The Positive Manifold of Intelligence by Mutualism. *Psychological Review, 113(4)*, 842–861. doi: 10.1037/0033-295X.113.4.842

van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the dutch WAIS-III. *Intelligence*, *34*, 273–289. doi: 10.1016/j.intell.2005.08.002

van Rijn, P., Dolan, C. V., & Molenaar, P. C. M. (2010). State space methods for item response modeling of multisubject time series. In P. C. M. Molenaar & K. M. Newell (Eds). *Individual Pathways of Change: Statistical Models for Analyzing Learning and Development*, pp. 125–135. Washington, DC: American Psychological Association.

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

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. Educational Measurement: *Issues and Practice*, *29(3)*, 39–47. doi: 10.1111/j.1745-3992.2010.00182.x

8 Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

9 Figure legends

Figure 1. **Model taxonomy in terms of model equations and verbalized form.**

Figure 2. **Relatively best fitting models for subjects 7, 13 and 22.** Paths fixed to zero are not drawn. Note that these include the regression parameters of the vector eta on the constant, i.e., vector alpha, which are fixed to zero for scaling purposes. Paths fixed to one are dashed. These include the latent residual variances in order to provide a latent metric. Freely estimated paths are drawn in black and parameter point estimates are provided. Items denoted with e are extraversion marker items, whereas items denoted with a are agreeableness marker items. The numerical ordering of the items employed here corresponds to the ordering of the items as given in the data description section. Index i is dropped as the models describe single individuals.

Figure 3. **Individual model for subject 7 including the neuroticism marker item “bad tempered” as a fixed regressor.** According to this representation, the neuroticism item possibly affects the agreeableness marker items above the potential effect it through the agreeableness factor.

Figure 1.JPEG

Dimension person / between-subject level

Dimension time / within-subject level

	No restrictions	Invariance constraints on the measurement model	Invariance constraints on the measurement and latent model
No restrictions	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\Lambda}_{i,t}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta}_{i,t})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i,t})$ No invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta}_{t})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i,t})$ No invariance over time Measurement invariance over subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{t} + \boldsymbol{\Lambda}_{t}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{t} + \mathbf{B}_{t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta}_{t})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{t})$ No invariance over time Measurement invariance and process invariance over subjects
Invariance constraints on the measurement model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta}_{i})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i,t})$ No invariance over subjects Measurement invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i,t} + \mathbf{B}_{i,t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i,t})$ Measurement invariance over time and subjects	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{t} + \mathbf{B}_{t}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{t})$ Measurement invariance over time Measurement invariance and process invariance over subjects
Invariance constraints on the measurement and latent model	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau}_{i} + \boldsymbol{\Lambda}_{i}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i} + \mathbf{B}_{i}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta}_{i})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i})$ No invariance over subjects Measurement invariance and process invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha}_{i} + \mathbf{B}_{i}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi}_{i})$ Measurement invariance over subjects Measurement invariance and process invariance over time	$\boldsymbol{y}_{i,t} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\ \boldsymbol{\eta}_{i,t} + \boldsymbol{\varepsilon}_{i,t}$ $\boldsymbol{\eta}_{i,t} = \boldsymbol{\alpha} + \mathbf{B}\ \boldsymbol{\eta}_{i,t-1} + \boldsymbol{\zeta}_{i,t}$ $\boldsymbol{\varepsilon}_{i,t} \sim N\ (0, \boldsymbol{\Theta})$ $\boldsymbol{\zeta}_{i,t} \sim N\ (0, \boldsymbol{\Psi})$ Measurement invariance and process invariance over time and subjects

Figure 2.JPEG

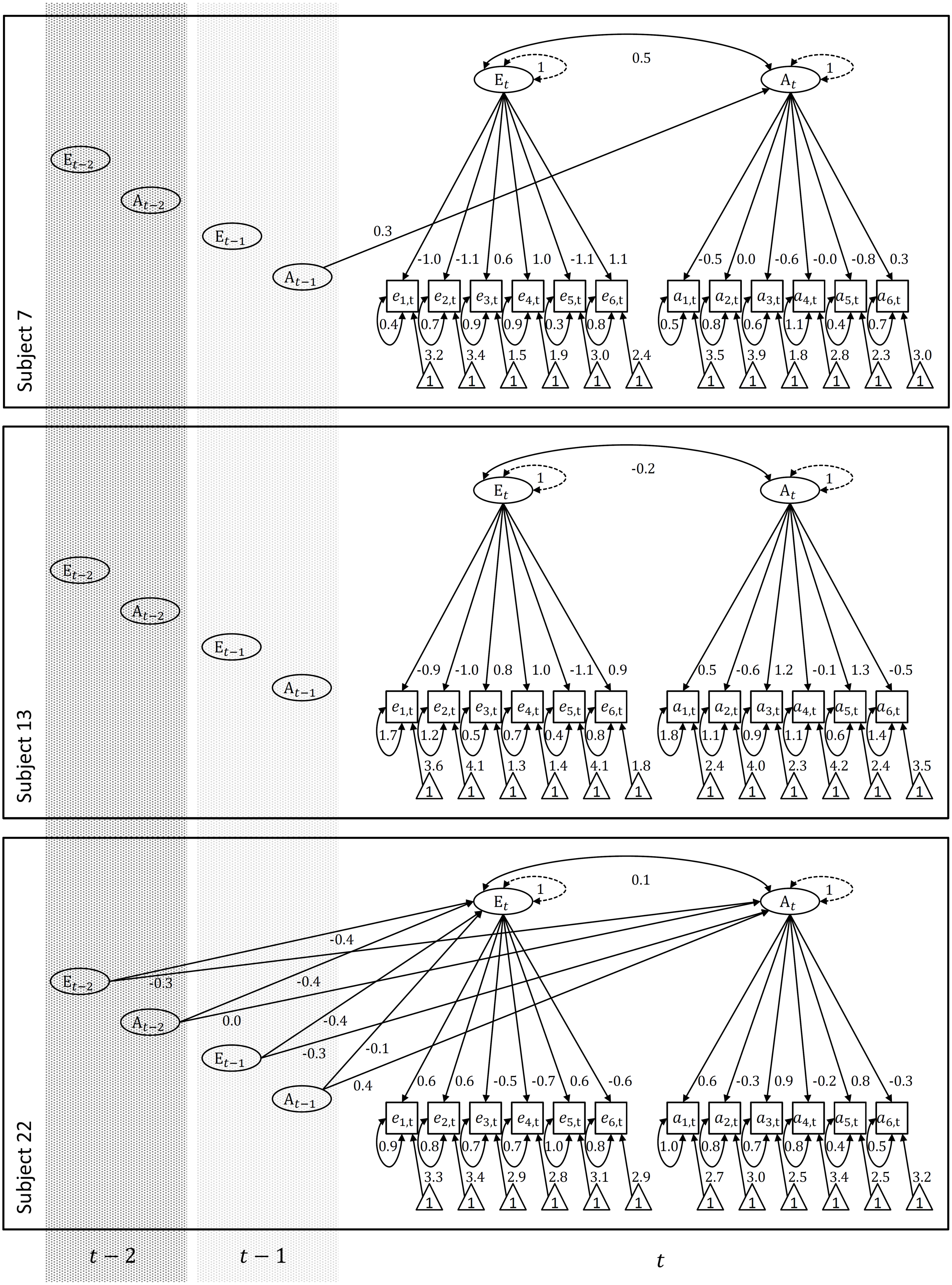

Figure 3.JPEG