Multilevel Dynamic Twin Modeling

N. K. Schuurman, Y. Zheng & C. V. Dolan

Published online: 06 Jul 2021.

Submit your article to this journal ⬀

Article views: 925

View related articles ⬀

View Crossmark data ⬀

Multilevel Dynamic Twin Modeling

N. K. Schuurman[a], Y. Zheng [iD][b], and C. V. Dolan [iD][c]

[a]Utrecht University; [b]University of Alberta; [c]VU Amsterdam

ABSTRACT

Recent developments in the collection and modeling of intensive longitudinal data have enabled us to fit dynamic twin models, in which within-person processes are separated into genetic and environmental components. A well-known dynamic twin model is the genetic simplex model, which is fitted to a few repeated measures for many twins. A more recently developed model is the iFACE model, which is fitted to many repeated measures for a single pair of twins. In this paper, we introduce a missing link between these two models – a multilevel extension that allows for making both population-level and twin-level inferences. We provide a proof-of-principle simulation study for this model, and apply it to an experience sampling data set on 148 monozygotic and 88 dizygotic twins. We use the multilevel model to examine the overlap and differences between the dynamic genetic twin models and the classic twin models, as well as their interpretation.

Twin models conventionally have been used to investigate the relative contributions of genes and environment to psychological and behavioral trait variance based on measurements of stable *inter*-individual differences between twins. Recent developments in the collection and modeling of intensive longitudinal data have offered a novel opportunity to examine the heritability of *intra*-individual differences: the proportion of variance in fluctuations in a phenotype over time that is due to genetic effects.

This type of heritability differs conceptually from the conventional definition of heritability, and may produce very different results. For example, consider a phenotype, such as agreeableness. There could be substantial differences in the agreeableness of different persons, as a result of differences in their genetics, indicating a fairly high inter-individual heritability. When we evaluate whether fluctuations in a person's level of agreeableness from day to day—one day a person is more agreeable than the next—is the result of genetic effects or environmental effects, the results may be quite different. For instance, if the effects of the genetics of a person related to agreeableness do not change much from day to day, the *intra*-individual heritability of agreeableness for that person is low, even if the inter-individual heritability is high.

Intra-individual heritability can be investigated with *dynamic* twin models, in which genetic and environmental components are considered as dynamic processes that vary within persons over time. A well-known dynamic twin model is the genetic simplex model, in which intra-individual (i.e., within-person) differences in genetic and environmental components are modeled for a group of twins. In this model, one common dynamic autoregressive process is assumed for all twins (Boomsma & Molenaar, 1987; Dolan et al., 1991). That is, all the model parameters—including the autoregressive effects- are assumed to be identical for each twin in this model. A more recently developed model is the iFACE model, in which a similar model is fitted for a single twin-pair, based on more intensive longitudinal data (Molenaar, 2011; Molenaar et al., 2012; Nesselroade & Molenaar, 2010). This model allows for the estimation of intra-individual heritability and the contribution of environmental influences specific to each twin in the pair. Hence, the iFACE model allows for different heritabilities of within-person fluctuations in a phenotype for each person.

The interpretation of the intra-individual genetic and environmental components, and that of intra-individual versus inter-individual heritability, is not trivial, yet has received relatively little attention in the field. In this paper, we aim to elucidate these concepts, and the opportunities and pitfalls of dynamic genetic models. To this end, we introduce the "missing link" between the genetic simplex model and the iFACE model—a multilevel extension of these models implemented with Dynamic SEM (DSEM; Asparouhov et al., 2018). This multilevel model allows for making both population-level and twin-level inferences. We discuss the interpretation of its model parameters and intra-individual heritability, as well as the link of the multilevel model with the genetic simplex model and the iFACE model.

In the following sections, we first discuss conventional cross-sectional twin models to highlight the aspects in which these models are distinct from their dynamic counterparts. Next, we briefly discuss the genetic simplex and iFACE model. After that, we introduce the multilevel extension of these models, the interpretation of the model parameters, and the resulting intra-individual heritability estimates. We also discuss how the three dynamic genetic models are related. After that, we discuss the results of a proof-of-principle simulation study on the performance of the multilevel model implemented in Mplus' DSEM (Asparouhov et al., 2018). We

illustrate the interpretation of dynamic twin models via an empirical application on how "zen" twins feel over time, using empirical data from Wichers et al. (2007). We finish with a discussion of the limitations of the current work, and potential directions for future work.

Background: Twin models

Twin models can be used to estimate the proportions of the phenotypic variance that are attributable to genetic effects and to environmental effects (Jinks & Fulker, 1970; Martin & Eaves, 1977; Rijsdijk & Sham, 2002). By comparing the similarity between monozygotic (MZ) and dizygotic (DZ) twin pairs, we can estimate these proportions without actually measuring the environmental or genetic causes of the phenotype, or the genes. Instead of measuring and modeling the genes and environmental characteristics directly, the ensemble of genetic effects and the ensemble of environmental effects are captured in latent variables. By making several assumptions about the MZ and DZ twin data, it is possible to estimate what proportion of phenotypical variance is due to the latent genetic variable, and due to the latent environmental variable (Eaves et al., 1977, we get back to the assumptions in the next subsection). These models were initially developed for studying inter-individual differences based on cross-sectional data. Subsequent extensions of these models to longitudinal data were developed that shift the focus from inter-individual differences, to intra-individual changes in the latent genetic and environmental variables—that is, to changes in these variables within persons over time. In the following we first discuss the ideas behind the classic cross-sectional model, followed by such extensions to longitudinal data.

The classical AE twin model

The classical AE twin model distinguishes two latent variables, denoted A and E. A captures additive genetic effects, and E captures non-shared environmental effects that are unique to each member of each twin pair. More complex models include additional latent variables. For example, the ACE model includes common environmental effects that are shared by each twin member; the ADE model includes a latent variable for dominant genetic effects that captures interactions between alleles (i.e., epistasis; Rijsdijk & Sham, 2002). It is also possible to fit genetic models based on data from other family members, such as parents, other siblings, or cousins, next to or instead of twins (Fulker & Bonné-Tamir, 1982; Heath et al., 1985; Keller et al., 2009). Throughout this paper, however, we will focus on the AE model that includes only the data of MZ and DZ twins, for the sake of simplicity.

In the classic cross-sectional AE model, the phenotype of interest is measured at one occasion in a sample of MZ twin pairs, who share all of their genes, and in a sample of DZ twins, who on average share half of their segregating genes. Hence, if the scores for the MZ twins are more similar, or more strongly correlated than for the DZ twins, this indicates that part of the variance in the phenotype of different persons can be explained by variation in their genes. The classical AE twin model is used to estimate the variance in the phenotype that can be attributed to genetic components, and to unique environmental components. A path-model representation of the cross-sectional AE model is depicted in Figure 1. The measured phenotypes of the twins are denoted 'Phen1′ and 'Phen2′ in Figure 1, for twin 1 and twin 2 of a pair, respectively. Further, the model includes four latent variables, two that capture the additive genetic components for each twin, respectively (A1 and A2), and two that capture the unique environmental components for each twin (E1 and E2). If we can estimate the variances for the latent variables A and E, we can determine how much of the total variance in the phenotype across persons is due to genetic components A, and how much is due to unique (non-shared) environmental components E.

The classic cross-sectional twin models are fitted as a multigroup model, where MZ twins and DZ twins are two different groups, as is depicted in Figure 1. To estimate this model, a number of assumptions need to be made (c.f., Rijsdijk & Sham, 2002, for an overview). Particularly, it is assumed that the genetic components are perfectly correlated with each other in the MZ twins, because they are genetically identical (although there is debate on this, c.f., Charney, 2012; Liu et al., 2018). DZ twins will on average across twins share approximately 50% of their alleles. Based on this assumption, assuming random mating, the expected correlation between the additive genetic variables equals .5. The unique environmental
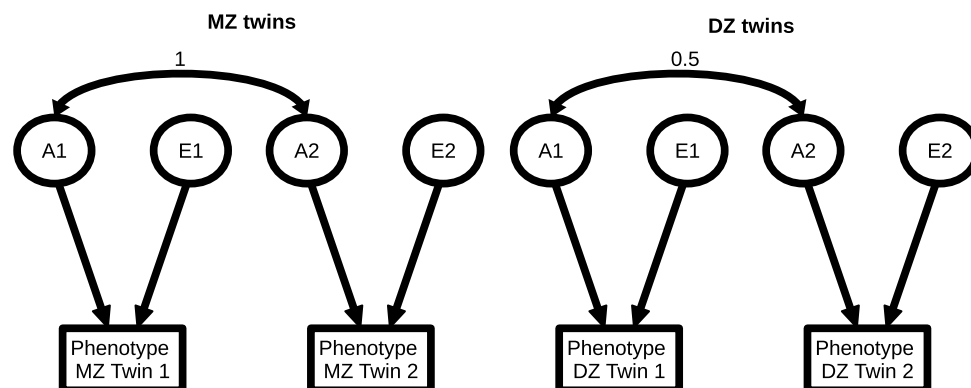
Figure 1. Cross-sectional AE twin model. On the left the model for the mono-zygotic twins is displayed, on the right the model for the di-zygotic twins. For both groups the phenotype of the two twins in each pair load on a factor A that captures additive genetic effect, and a factor E that captures unique environmental effects. For the monozygotic twin pairs, the A factor is perfectly correlated, for the dizygotic twin pairs, the A factor is correlated 0.5. The E factor is uncorrelated among the twins.

components are uncorrelated by definition for MZ and DZ twins.[1] By making these assumptions, it is possible to estimate the variance of the latent additive genetic components A and the latent unique environmental components E. The proportion of the total variance in the phenotype across the twins that is explained by latent variable A, is referred to as the heritability of that phenotype. That is, if the estimated heritability for the cross-sectional model is 0.6, 60% of the variance in the phenotype scores across twins is explained by variance in the genetic components A, and 40% by variance in the unique environmental components E. This procedure is convenient, because the relative contributions of the environment and genetics can be determined without having to measure the relevant specific genes and characteristics of the environment.

Dynamic twin models

The classic cross-sectional model does not take into account that a phenotype, as well as the associated genetic and environmental latent variables, may change over time, or how these changes could influence the estimates of heritability. At first glance, it may seem that the scores of latent variable A should not fluctuate over time given that people's genes do not typically change. It is important to note, however, that what is captured through the latent variables A and E are not the genes and environmental characteristics themselves, but *the effects of* the genes and the environment, respectively. Although genes typically do not change across a person's lifespan, the effects of the genes may change. For example, the expression of genes may change, resulting in their effects becoming gradually stronger or weaker (e.g., genetic amplification; Plomin, 1986). Furthermore, the specific genes relevant to the phenotype may change as well. For instance, the specific genes that affect the phenotype may differ from childhood to adulthood (e.g., genetic innovation and attenuation; Kendler et al., 2008). The environmental effects or relevant environmental characteristics may likewise change over time. To account for such changes in genetic and environmental effects, and to explicitly study changes of these effects over time, dynamic genetic twin models have been developed. Dynamic twin models differ from the classical twin models in that they model change in the phenotype and latent environmental and genetic components within twins over time. That is, genetic and environmental components are considered to be part of a dynamic, within-twin process.

Genetic simplex model

A well-known dynamic twin model is the genetic simplex model (Boomsma & Molenaar, 1987; Dolan et al., 1991), which is applied by means of Structural Equation Modeling (SEM) to panel data that consist of a few repeated measurements (e.g., fewer than 10) for many twin pairs. The genetic simplex model, like classical twin models, makes use of the difference in genetic relatedness between MZ and DZ twins, where for the MZ twin members the additive genetic components are assumed to be perfectly correlated across MZ twins

for each time point, .5 across DZ twins, and the unique environmental components are uncorrelated for all twins at each time point. However, the phenotype scores and the scores on the latent genetic and environmental variables change across measurement occasions. The measurement equation of the genetic simplex model can be expressed as follows:

$$y_{ijt} = \mu_t + \lambda_{At}A_{ijt} + \lambda_{Et}E_{ijt} + \epsilon_{ijt} \quad (1)$$

$$\epsilon_{ijt} \sim N(0, \sigma^2). \quad (2)$$

Here y_{ijt} are the phenotype scores for twin j, member i, at time point t. These scores consist of a mean score μ_t, the genetic contribution to the phenotype $\lambda_{At}A_{ijt}$, the environmental contribution to the phenotype $\lambda_{Et}E_{ijt}$, and occasion-specific residual term ϵ_{ijt} (which includes measurement error). The occasion-specific residuals ϵ_{ijt} are assumed to be normally distributed with a mean of zero and variance σ^2. The genetic and environmental components A and E at one occasion may be related to those at the next occasion. That is, there is dependency between the components over time. In the genetic simplex model, this dependency over time is modeled with an autoregressive process of order 1, where the genetic components A, and the unique environmental components E are regressed on themselves at the nearest previous occasion. This relationship is presented in the following transition equations for components A and E respectively:

$$A_{ijt} = \phi_{At}A_{ijt-1} + v_{Aijt} \quad (3)$$

$$E_{ijt} = \phi_{Et}E_{ijt-1} + v_{Eijt}. \quad (4)$$

Here, the ϕ regression coefficients represent the associations between the scores of the twins across time, and v represent the residuals of the respective processes. To estimate the model, residuals v_{Eijt} and v_{Aijt} are both assumed to be normally distributed with a mean of zero, and variances fixed to 1 for scaling purposes. The correlations between these residuals of A v_{Aijt} are 1 between MZ twin members, .5 between DZ twin members. Correlations among the residuals of E v_{Eijt} are assumed to be equal to zero. As is the case in the classical cross-sectional twin models, the Genetic Simplex model is fitted as a multigroup model, where the two groups are the MZ and DZ twins.

Each parameter has subscript t in the genetic simplex model, indicating that the model parameters can vary over time. As a result, the contributions of A and E to the phenotype can vary across time points. That is, the heritability of the modeled phenotype can change over time. Hence, the genetic simplex model is typically used to study dynamics over relatively long periods of time across twins on a group level. The heritability of the inter-individual differences is then determined at each measurement occasion, which allows for investigating changes in the heritability over time. For example, Dolan et al. (1991) applied the genetic simplex model to the weight for 83 female twins, which was measured every half year, six times in total. They allowed the heritabilities to

---

[1]Note that the assumption in the AE model is that there are no common environmental effects, which may be unrealistic in some situations. However, the model can be extended to the ACE model to account for this factor by introducing a third latent variable which is correlated at 1 for both MZ and DZ twins.

change from occasion to occasion, and found that the heritabilities of weight were considerable, and fairly stable over time (i.e., heritabilities of .87, .89, .87, .89, .91, and .89).

An important limitation of the genetic simplex model is that it does not allow for inter-individual differences in the dynamic processes across twins; the model parameters are assumed to be the same for each person. Recent studies have discussed several potential sources of substantial inter-individual differences in genetic and enviromental effects, such as stochastic within-person developmental noise (e.g., Molenaar et al., 1993; Molenaar & Raijmakers, 1999) and epigenetic processes that could modify gene expression (e.g., DNA methylation; Bell & Saffery, 2012; Bell & Spector, 2011; Dolan et al., 2015; Kan et al., 2010; Petronis, 2010; Wright et al., 2014). It seems particularly likely that there will be inter-individual differences in the means of twins for most phenotypes, meaning that parameter μ should be allowed to differ from twin to twin. For instance, if we take concentration level as the phenotype of interest, we know that some people generally find it easier to concentrate than others, and hence some people would have a higher average concentration level than others. In the basic genetic simplex model such stable between-person differences are not taken into account. However, it is possible to alter the genetic simplex model to account for such level differences in the phenotypes among different persons by incorporating random means or intercepts in the model (see also Hewitt et al., 1988). In the following, we will specify such a model. To do this, however, we will specify a simplex model where all parameters are the same over time. The resulting stationary Genetic Simplex model with a random mean is strongly related to the other dynamic models discussed in this paper. The equations of the genetic simplex model are altered as follows:

$$y_{ijt} = \mu_{ij} + \lambda_A A_{ijt} + \lambda_E E_{ijt} + \epsilon_{ijt} \qquad (5)$$

$$A_{ijt} = \phi_A A_{ijt-1} + v_{Aijt} \qquad (6)$$

$$E_{ijt} = \phi_E E_{ijt-1} + v_{Eijt}. \qquad (7)$$

The random means are μ_{ij}. These random means may differ between twins, but are assumed to be stable within a twin over time. These means are sometimes referred to as "trait scores," as they reflect traits of particular persons that are stable over time. These traits stand in contrast to the "states" y_{ijt}, which include deviations from the trait at each time point. For example, a person might find it easier to concentrate at the start of a working day, rather than four hours later: This effect would be reflected in a higher concentration score y_{ijt} in the morning than in the evening, while μ_{ij} reflects a general tendency that is stable over time. The fluctuations of the scores y_{ijt} over time are reflected in the deviations at time point t from mean μ_{ij} for each twin ij ($y_{ijt} - \mu_{ij}$). These deviations are captured in the remainder of the measurement equation, $\lambda_A A_{ijt}$, $\lambda_E E_{ijt}$, and ϵ_{ijt}. Hence, by including random means in the model, variance due to stable between-person differences—differences among twins in μ_{ij}—can be distinguished from within-person variance (Hamaker & Grasman, 2015; Hamaker et al., 2015; Hoffman

& Stawski, 2009a). In other words, the stable between-person variance in traits μ_{ij} have been filtered out by including random means, such that any variability in A and E is now the result of within-person fluctuations over time. As a result, for this model we can obtain a different kind of heritability for the phenotype under study: the intra-individual heritability. The intra-individual heritability is the proportion of *within-person* variance in the phenotype ($var(y)_{intra}$) that is the result of within-person variance in the genetic components A_{ijt} over time ($var(A)$). The within-person variance of the phenotype $var(y)_{intra}$ is equal to:

$$var(y)_{intra} = \lambda_A^2 var(A) + \lambda_E^2 var(E) + \sigma^2. \qquad (8)$$

That is, the variance in the phenotype excluding variability that results from stable differences between persons in mu_{ij}. Given that A and E follow an autoregressive process, the (within-person) variance of A ($var(A)$) is equal to

$$var(A) = \frac{var(v_A)}{1 - \phi_A^2}, \qquad (9)$$

and the (within-person) variance of E ($var(E)$) is equal to

$$var(E) = \frac{var(v_E)}{1 - \phi_E^2}. \qquad (10)$$

These functions for the variances of *E* and *A* result because they have an autoregressive process of order 1 (Kim & Nelson, 1999, p. 27). Note that for estimating the model the variances of v_E and v_A ($var(v_A)$, $var(v_E)$) in this equation would be fixed to 1 for scaling purposes. The intra-individual heritability for the genetic simplex model with a random intercept is then equal to:

$$h_{intra}^2 = \frac{\lambda_A^2 var(A)}{\lambda_A^2 var(A) + \lambda_E^2 var(E) + \sigma^2}. \qquad (11)$$

The intra-individual heritability represents the within-person variance in the states y over time that is due to variance in the genetic effects A over time. Hence, this intra-individual heritability is distinct from the inter-individual heritabilities discussed earlier for the non-stationary genetic simplex model without random means and classical twin models. Classic twin models generally aim to determine the between-person heritability of stable traits, that is, the proportion of variance among the traits of different people that can be explained by the genetic variance among those people. In the context of the simplex model with random means, this heritability would be comparable to the between-person heritability of the random means μ_{ij}. That is, to estimate the between-person trait heritability in the genetic simplex model with random means, we would extend it to include an additional between-person AE model on the random means μ_{ij}. This approach to estimating between-person trait heritability for longitudinal data is often taken in the context of genetic latent growth curve models (Long et al., 2017).

Both the the classical between-person heritability of stable traits, and intra-individual heritability, in turn differ from the heritabilities that are estimated per time point in the non-

stationary simplex model that excludes random means: These inter-individual heritabilities show for a particular measurement occasion t the proportion of variance among the phenotype state scores y_{ijt} of different twin pairs that is due to variance in the genetic components A_{ijt}. Given that this basic simplex model does not separate out stable between-person variability from within-person variability in the phenotype, the resulting heritabilities would represent a blend of between-person trait heritability and intra-individual heritability.

In the current work, we will focus mainly on the concept of intra-individual heritability: The proportion of variance in the phenotype y over time that can be explained by variance in the genetic effects A over time. In this context, it is important to note that the intra-individual heritability may actually differ from person to person. This would be the case when the changes over time in the genetic and environmental effects differ from person to person, for example, because one person has different experiences and thus encounters different environmental characteristics than another person. This possibility is not accounted for in the genetic simplex models, because the genetic and environmental components are assumed to have identical dynamic processes. That is, other than the means μ_{ij} in the Genetic Simplex model with random means or intercepts, the model parameters are assumed to be identical for each twin. As a result, the intra-individual heritability specified above reflects an overall intra-individual heritability across all modeled twins. However, a dynamic twin model that does account for person-specific intra-individual processes is the iFACE model developed by Molenaar (2011). We discuss this model in the following subsection.

iFACE model

The idiographic filter ACE (iFACE) model (Molenaar, 2011; Molenaar et al., 2012; Nesselroade & Molenaar, 2010) is a recently introduced dynamic twin model which is applied to time series data for a single twin pair at a time – either MZ or DZ. Hence, unlike the previously described models, the iFACE model is not a multigroup model. Time series data consists of many repeated measures (e.g., >25 measurements), usually measured relatively close in time, such as over the course of a day, a few days or weeks. Typically, psychological time series data are collected with daily diary and ESM measurement procedures (Hamaker & Wichers, 2017). The iFACE model makes use of correlations of 1 between the additive genetic components for a single MZ twin pair, and a to be estimated correlation for a single DZ twin pair which is expected to be around, but not necessarily equal to .5. The correlations between the unique environmental components are equal to zero. The AE rendition of the iFACE model can be expressed as follows for a single twin pair:

$$y_{it} = \mu_i + \lambda_{Ai}A_{it} + \lambda_{Ei}E_{it} + \epsilon_{it} \tag{12}$$

$$\epsilon_{it} \sim N(0, \sigma_i^2) \tag{13}$$

$$A_{it} = \phi_{Ai}A_{it-1} + v_{Ait} \tag{14}$$

$$E_{it} = \phi_{Ei}E_{it-1} + v_{Eit}. \tag{15}$$

Note that the equations are essentially the same as for the genetic simplex model, albeit applied to a single twin pair. The phenotype scores y_{it} for twin i at time point t consist of 1) a mean score μ_i which is stable over time but specific to twin i; 2) the genetic contribution to the phenotype $\lambda_{Ai}A_{it}$; the environmental contribution to the phenotype $\lambda_{Ei}E_{it}$; and occasion-specific residual ϵ_{it}. Furthermore, the changes in the genetic and environmental components are also modeled with an autoregressive process captured in ϕ_{Ai} and ϕ_{Ei}. However, unlike the genetic simplex model, the processes are modeled for each twin pair separately, such that there is no between-twin-pair variance in the model. Furthermore, the modeled processes are *unique to each twin i*: all the parameters may differ between the twins i. The dynamic processes itself are assumed to be stationary in the iFACE model, such that the parameters and heritabilities remain the same over time (hence, subscript t is absent for the model parameters). The (within-person) variance in the phenotype for each twin i is then equal to:

$$var(y_{it}) = \lambda_{Ai}^2 var(A_i) + \lambda_{Ei}^2 var(E_i) + \sigma_i^2, \tag{16}$$

where the person-specific variance of A is equal to

$$var(A_{it}) = \frac{var(v_{Ai})}{1 - \phi_{Ai}^2}, \tag{17}$$

and the person-specific variance of A is equal to

$$var(E_{it}) = \frac{var(v_{Ei})}{1 - \phi_{Ei}^2}. \tag{18}$$

Note that for estimating the model, the variances of v_{Ei} and v_{Ai} ($var(v_{Ai})$, $var(v_{Ei})$) in this equation would be fixed to 1 for scaling purposes. The within-person variances are different for each twin, given that the twins each have unique model parameters. As a result, the intra-individual heritability of the phenotype can also differ for the two twins. This *person-specific* intra-individual heritability reflects the proportion of variance in the twin's phenotype over time that is due to variance in that twin's genetic effects over time. This person-specific heritability can be calculated as follows:

$$h_i^2 = \frac{\lambda_{Ai}^2 var(A_i)}{\lambda_{Ai}^2 var(A_i) + \lambda_{Ei}^2 var(E_i) + \sigma_i^2}. \tag{19}$$

The iFACE model can be used to study the genetic and environmental processes, and estimate the intra-individual heritability, for each twin. For example, Molenaar et al. (2012) applied the iFACE model to multivariate electroencephalogram data from a DZ twin pair, and demonstrated considerable differences in the proportions of genetic and environmental variance between the two twins. Typically, however, time series data is not collected for a single twin pair but for multiple twins pairs. The iFACE model can then be fitted for each twin pair separately—a replicated time series design—resulting in unique parameter estimates for each twin and each twin pair. By inspecting the results across the different twin pairs, the

iFACE model can be used to get an impression of the differences between twin pairs in the processes and heritabilities when the iFACE model. For example, Zheng et al. (2016) applied the iFACE model separately to the daily positive and negative school feelings of multiple 10-year-old MZ twin pairs. They found that the amount of variance in the daily fluctuations in these feelings that was due to unique environmental influences varied greatly across twin pairs. Note that this is done in a two-step procedure, first fitting the iFACE model for each twin, and then in a new analysis summarizing or analyzing the variability among the twins pairs. Hence, it is disregarded that the different twin pairs may to some extent be similar to each other, for example, that all twins may typically have positive autoregressive effects. Furthermore, the two step approach to analyzing the interindividual differences in the within-twin processes does not take into account that there is uncertainty about the parameter estimates of each twin. More ideal may be to actively model and analyze the between-twin differences in the within-twin processes while 1) allowing for differences in the parameters of each twin, 2) taking into account that the twins may be to some extent similar to each other, and 3) taking into account that there is uncertainty about the estimates of each twin by implementing a one-step procedure. We implement this in a multilevel bridge model between the genetic simplex and iFACE model, as we'll discuss in the following sections.

Bridge model between the genetic simplex and iFACE model

Although the genetic simplex model and the iFACE model are applied on different types of data they are clearly closely related. Both model the change in the latent genetic and environmental components, and both do so with an autoregressive process. Both models can be used to shift focus from the heritability of interindividual differences between persons, as is the focus in classic twin models, to the heritability of intra-individual differences within persons. For the simplex model this would require separating the stable between-person variance from the within-person variance in the model, for instance, by including a random intercept or mean. In any case, the estimated heritabilities for genetic simplex models and the iFACE model differ from those in the classical cross-sectional twin model, and the interpretation of the models and resulting heritabilities are not trivial. In the following section, we present what may be considered a missing link between the genetic simplex model and the iFACE model: A multilevel dynamic twin model that is applied to intensive longitudinal data, which consists of time series for multiple twins. We will use this model to further discuss the interpretation of the dynamic genetic models, intra-individual heritability, and the relationship between the different dynamic genetic models.

A multilevel dynamic twin model

In the following we introduce a multilevel dynamic twin model for genetically informative intensive longitudinal data, consisting of time series from multiple twin pairs. Such data allow us to model intra-individual processes of the underlying genetic and environmental effects, as well as explicitly model inter-individual differences in these processes. We will focus on an AE model for MZ and DZ twins for a single phenotypic variable. However, it is possible to fit multilevel dynamic twin models to either MZ twins or DZ twins alone, or to extend the model to include other family members, common environmental effects, or multiple indicators. We return to possible extensions of the model in the discussion. We focus on a 2-level model, where the intra-individual dynamics based on repeated measurements are modeled at level 1, and the inter-individual differences in these dynamics are modeled at level 2. In the following section, we will first discuss the level 1 model, followed by the level 2 model. After that, we discuss the resulting intra-individual heritability estimates based on the model parameters. Finally, we discuss how the multilevel twin model relates to the genetic simplex model and the iFACE model. A graphical representation of the multilevel twin model is included in Figure 2.

Level 1

Level 1 of the multilevel dynamic twin model consists of a measurement equation and a transition equation. The measurement equation relates the observed phenotypic scores to the latent genetic and environmental variables of the repeated measures of each twin. The transition equation describes the dynamic process of the genetic and environmental components of each twin.

Measurement equation

The measurement equation is specified as follows:

$$y_{ijt} = \mu_{ij} + \lambda_{Aij}A_{ijt} + \lambda_{Eij}E_{ijt} + \epsilon_{ijt}. \tag{20}$$

In the measurement equation, the observed phenotype y_{ijt} at each time point t of each twin j and twin member i are divided into four parts. First, a person-specific (random) mean score μ_{ij} that is allowed to be different for each twin member, but remains the same across all time points. That is, μ_{ij} is a random mean, which ensures that we allow for differences in levels among twins. It can be seen as a special case of a random intercept, where the random intercept is equal to the mean. As discussed previously, such random means are sometimes referred to as "trait scores," as they reflect stable traits of particular persons. These traits stand in contrast to "states," which reflect time point specific scores. The fluctuations of the states over time are reflected in the deviations at time point t from mean μ_{ij} of each twin ij. These deviations are captured in the remainder of the measurement equation, $\lambda_{Aij}A_{ijt}$, $\lambda_{Eij}E_{ijt}$, and ϵ_{ijt}.

As discussed previously, here lies the main difference between classic cross-sectional and dynamic twin models: The classic twin models focus on the heritability of traits, using measurements that should reflect overall differences between twins. This between-person heritability is the proportion of the variance in people's trait scores that can be explained by variation in additive genetic effects between people. This heritability would be akin to the between-person heritability of the random means μ_{ij} which are taken in the multilevel model to reflect phenotype trait scores. However, we will focus on the intra-individual heritability of fluctuations in the phenotype within a twin over time, similar to that of the iFACE model. This intra-individual
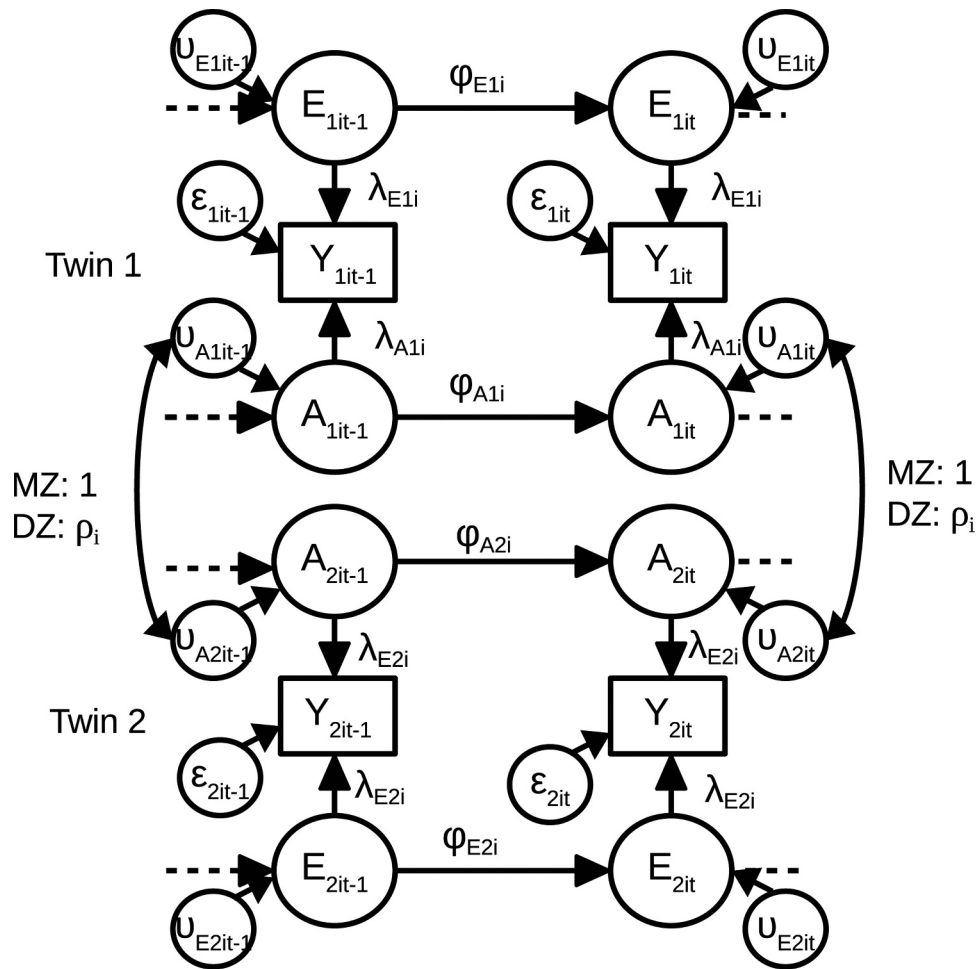
Figure 2. A graphical representation of the multilevel dynamic twin model. It includes random factor loadings λ_{ij}, autoregressive coefficients ϕ_{ij}, random means μ_{ij} of the phenotype y_{ijt}, and random variances σ_{ij}^2 of occasion-specific residuals ϵ_{ijt}, across twins j and twin members i. It also includes random correlations ρ_{DZij} between the genetic components of the DZ twins j. Note that for the MZ twins, we assume that $\phi_{A1i} = \phi_{A2i}$.

heritability is the proportion of variance of the fluctuations of a particular person over time that is due to fluctuations over time in their additive genetic effects.

To this end, the state deviations around the mean are separated into three parts: 1) a score on latent variable A_{ijt} weighted by person-specific factor loading λ_{Aij}. This represents the contribution of the additive genetic effects to the fluctuation at time t, of twin j, member i; 2) a score on latent variable E_{ijt} weighted by factor loading λ_{Eij}. This represents the contribution of the unique environmental effects to the fluctuation at time t, of twin j, member i; and finally 3) a residual score $[i]$ of twin j, member i, at time point t, which captures any deviations from the mean due to other unobserved influences on the system that are *specific to measurement occasion t*, including any measurement error. These residuals are assumed to be normally distributed with a mean of zero and variance σ_{ij}^2, that is,

$$\epsilon_{ijt} \sim N(0, \sigma_{ij}^2). \tag{21}$$

To identify the model, the additive genetic components of the states A_{ijt} of the MZ twin members are assumed to be perfectly correlated across twin members over time; MZ

twin members hence have the same scores A_{ijt}. For DZ twin members, the latent scores A_{ijt} are assumed to be correlated at .5 over time on average across the DZ twins. Note that for individual DZ twin pairs, this correlation may be higher or lower than .5. We allow for the estimation of this genetic correlation at level 2 of the model as will be discussed later. The unique environmental components of the twin members are uncorrelated across both MZ and DZ twins. The fluctuations of the genetic components over time of MZ twin members will be identical in this model, but will differ for the DZ twins. The fluctuations of the unique environmental components are unique to each member of the MZ and DZ twins. In this respect, the model builds on the same ideas as the classic cross-sectional models. The correlation matrix among the latent variables for the MZ and DZ twins implied by the multilevel model is depicted in Table 1.

The factor loadings λ_{ij} determine the relative contributions of the environmental and genetic components to the observed phenotype, and hence in part the heritability. These loadings can vary across twin members, taking into account that for some people the fluctuations in their phenotype over time may depend more on environmental effects, while for others,

Table 1. Correlation matrix of the genetic (A) and environmental (E) latent variables of the two MZ twins and DZ twins over time points t, implied by the multilevel dynamic twin model.

	A_{MZ1j}	A_{MZ2j}	A_{DZ1j}	A_{DZ2j}	E_{MZ1j}	E_{MZ2j}	E_{DZ1j}	E_{DZ2j}
A_{MZ1j}	1	1	0	0	0	0	0	0
A_{MZ2j}	1	1	0	0	0	0	0	0
A_{DZ1j}	0	0	1	0.5*	0	0	0	0
A_{DZ2j}	0	0	0.5*	1	0	0	0	0
E_{MZ1j}	0	0	0	0	1	0	0	0
E_{MZ2j}	0	0	0	0	0	1	0	0
E_{DZ1j}	0	0	0	0	0	0	1	0
E_{DZ2j}	0	0	0	0	0	0	0	1

* This is the average correlation among the genetic latent variables of the DZ twins, the correlation is allowed to vary across the DZ twin pairs. This is discussed in section "A Multilevel Dynamic Twin Model" section "Level 2."

genetic effects play a relatively large role. Note that the interpretation of the latent variables as additive genetic and environmental components is ensured via the assumptions about the correlations among the latent variables of the MZ and DZ twins (1, 0.5, or 0). Hence, the model makes use of the idiographic filter introduced by Nesselroade et al. (2007), and this is also where the "iF" in the iFACE model name stems from (Molenaar & Nesselroade, 2012).

Note again that the latent additive genetic components represent additive effects of genes on intra-individual differences within twin members over time, rather than differences between twins in cross-sectional designs. Hence, for the phenotypes of the MZ or DZ twins to load on the genetic effects A, the twin members need to share variance *in the fluctuations of their phenotypes over time*. In other words, genetic effects on the twins' phenotype are implied when the fluctuations in the phenotypes of MZ twins are more 'similar' from moment to moment than those of DZ twins. As such, to capture such genetic effects, it is especially important to consider the role of the measurement design compared to its role in cross-sectional studies. For example, it is essential that the twin members are measured at approximately the same time to capture shared genetic effects. Furthermore, especially when considering shared environmental effects, it is important to consider whether one measures twins that spend their time mostly together, or apart.

The residual scores ϵ_{ijt} are assumed to be uncorrelated among all twins, and hence can be considered to capture unique environmental effects that are specific to a particular measurement occasion. This term will also capture random measurement error. This contrasts with the unique environmental effects captured in E which are dependent from occasion to occasion, because they are carried over from moment to moment through the autoregressive process (Schuurman & Hamaker, 2019; Schuurman et al., 2015).

Transition equation

The transition equation at level 1 of the model specifies the dependency over time among the genetic and environmental

components. Specifically, the latent variables A_{ijt} and E_{ijt} are modeled with a first order autoregressive process. This is specified as follows, for the genetic and environmental components, respectively:

$$A_{ijt} = \phi_{Aij}A_{ijt-1} + v_{Aijt} \qquad (22)$$

$$E_{ijt} = \phi_{Eij}E_{ijt-1} + v_{Eijt}. \qquad (23)$$

Here, parameters ϕ_{Aij} and ϕ_{Eij} capture the autoregressive effects of the genetic and environmental components, respectively. If the autoregressive coefficient of the genetic components is positive, it indicates that when a twin has relatively high (compared to themselves) additive genetic effects at one occasion, they also tend to have relatively high genetic effects at the next occasion. The larger the autoregressive coefficient, the stronger this carryover of the genetic effects from one moment to the next. If the autoregressive effect is zero, the twins' genetic effects at one occasion cannot be predicted from the previous occasion at all. If it is negative, twins with relatively high genetic effects at the first occasion tend to have relatively low effects the following occasion, although this type of autoregressive effect seems unlikely in this context. The same idea applies to the environmental components. For the genetic components, the autoregressive effects are allowed to differ for each member of the DZ twins. For the MZ twins, the genetic autoregressive effects are allowed to differ across twin pairs j, but are assumed to be the same for the two members i of each MZ twin pair, such that they are assumed to have the same genetic process. For the environmental components, the autoregressive effects may differ for both members of the DZ and MZ twin pairs.

The residuals v_{Aijt} and v_{Eijt} represent fluctuations in the genetic and environmental components over time that are not explained by the autoregressive process. They are assumed to be normally distributed with means of zero. Note that in this model either the factor loadings λ_{ij} or the residual variances of v_{ijt} need to be fixed to a certain value for scaling purposes for each twin. Here, we will fix the residual variances v_{ijt} of the genetic and environmental variables to be equal to one.[2] The correlations between the residuals of A v_{Aijt} are 1 between MZ twin members, .5 between DZ twin members on average across the DZ twins, and correlated at 0 among different twins pairs, over time, and with the unique environmental components. Correlations with the residuals of E v_{Eijt} are generally assumed to be equal to zero.

All parameters in the level 1 model – the mean level of the phenotype μ_{ij}, factor loadings λ_{ij}, residual measurement error variance σ_{ij}^2, and autoregressive effects ϕ_{ij} – are allowed to vary among twin pairs, and among twins of the same pair, aside from ϕ_{Aj} of the MZ twins. Note however that these parameters are assumed to remain the same over time. In particular, we assume that the full model is stationary for each twin (see Hamilton, 1994, p. 259).

---

[2]We choose to scale the variances, rather than the factor loadings, for practical reasons: All level 1 parameters are allowed to vary randomly across twins. By scaling the variance we avoid having to model an additional random variance, but instead model a random loading. Random variances are generally more complicated to model than random loadings. If desired, it is possible to include random variances in the model instead, and scale the loadings, but it would be more involved. An example of how to include random variances is provided later in this section, where we discuss the random variances for the measurement model residuals [ϵ_{ijt}] included in the model.

Level 2

At level 2 of the multilevel dynamic twin model, the between-person differences in the person-specific parameters of level 1 are specified. That is, for the random means μ_{ij}, factor loadings λ_{ij}, residual measurement error variance σ_{ij}^2, and autoregressive effects ϕ_{ij}.

With respect to the person-specific means—the trait scores—it is probable that they may also be more similar for the MZ twins than for the DZ twins, if they have a between-person genetic component. Hence, it is important to take into account that the trait scores may be correlated for the twin members, and that this correlation may be different for MZ twins ($\rho_{\mu MZ}$) and DZ twins ($\rho_{\mu DZ}$).[3] Hence, we assume that the trait scores μ_{ij} are multivariate normally distributed, where twin member 1 is jointly distributed with twin member 2, as follows:

$$\mu_{ijMZ} \sim MvN(\gamma_\mu, \Psi_{\mu MZ}) \tag{24}$$

$$\mu_{ijDZ} \sim MvN(\gamma_\mu, \Psi_{\mu DZ}). \tag{25}$$

The average trait scores γ_μ are assumed to be the same for both members of the MZ twins and the DZ twins: We implemented this in the Mplus code using latent variables, which is available in the supplemental materials. The 2×2 covariance matrices for the traits may differ for the MZ twins and DZ twins, with $\Psi_{\mu MZ}$ and $\Psi_{\mu DZ}$ for the MZ twins and DZ twins, respectively. The variances of the traits of the MZ twins and the DZ twins, $\psi_{\mu MZ}^2$ and $\psi_{\mu DZ}^2$, capture the amount of variation in the person-specific means across twin members. These may differ for MZ twins and DZ twins, but are assumed to be the same for the members of each twin pair (i.e., for DZ1 and DZ2, and for MZ1 and MZ2). The covariances $\psi_{\mu MZ1MZ2}$ and $\psi_{\mu DZ1DZ2}$ capture the associations between the traits for the MZ twins and DZ twins, respectively. For the person-specific autoregressive coefficients and the factor loadings, we assume normal distributions with means γ and variances ψ that are the same for all twins. That is:

$$\phi_{Aij} \sim N(\gamma_{\phi A}, \psi_{\phi A}^2) \tag{26}$$

$$\phi_{Eij} \sim N(\gamma_{\phi E}, \psi_{\phi E}^2) \tag{27}$$

$$\lambda_{Eij} \sim N(\gamma_{\lambda E}, \psi_{\lambda E}^2) \tag{28}$$

$$\lambda_{Aij} \sim N(\gamma_{\lambda A}, \psi_{\lambda A}^2). \tag{29}$$

Here, the fixed effects γ represent the average coefficients across twins, and the variances ψ represent the variance among twin members in these coefficients. For the sake of simplicity, we do not model these random coefficients multivariately, which would allow the random means, loadings, and regression coefficients to correlate among twin members. The model may be extended to account for this; we return to this idea in the discussion.

Next to the means, factor loadings, and autoregression coefficients, the residual variances in the measurement model may also differ from person to person. For example, some people may be more sensitive to external events, or experience more impactful external events, and hence will have more variable scores. To take this into account, we allow the residual variances to vary from person to person (see also Jongerling et al., 2015). We assume a normal distribution for the log transformed variances, as follows:

$$log(\sigma_{ij}^2) \sim N(\gamma_{\sigma^2}, \psi_{\sigma^2}^2). \tag{30}$$

Note that for all random parameters we have assumed that the fixed effects and variances for the random parameters are equal for both twins in a pair, and for MZ and DZ twins except for the variances of the random means. Hence, we assume implicitly that these twins essentially represent the same population with respect to these parameters, as is common in twin modeling. It is straightforward to relax this assumption by specifying different fixed effects and variances for different twins. This would allow for testing whether there are certain structural differences between the dynamic processes of MZ and DZ twins, such as testing whether MZ and DZ twins on average have different means, factor loadings, autoregressive effects, or residual variances. Of course, one would need sufficient twins of both types to estimate these parameters well. If these parameters differ substantively for MZ and DZ twins, this would imply they represent different populations with different characteristics, which implies that MZ/DZ twin model results cannot be generalized to different types of twins, which in turn indicates that they cannot be generalized to non-twin populations.

Finally, we have assumed that the correlation of the genetic components of the DZ twins is equal to 0.5. While this would be expected to be the case on average across DZ twins, the exact overlap between genes can vary from twin pair to twin pair, and hence this correlation may differ for each twin pair to some extent. We can allow this correlation to vary across the DZ twin pairs. We assume a normal distribution for this correlation with the mean correlation fixed to 0.5, and variance $\psi_{\rho_{ADZ}}$, as follows:

$$\rho_{ADZj} \sim N(.5, \psi_{\rho_{ADZ}}^2). \tag{31}$$

Note that by specifying a normal distribution with a mean of 0.5 and a to be estimated variance, it is possible that correlations for twin pairs take on values larger than 1 or smaller than −1. We rely on the data and empirical identification of the model to keep the estimated correlations within a sensible range. The distribution specified can be seen as a prior distribution, and will be updated by the data. However, for very complex models, models with little data, or miss-specified models it will be harder to empirically identify the model. In that situation it can be helpful to specify a distribution that takes into account that the correlations should stay within a range of −1 to 1, for example, with a truncated normal distribution. The same

---

[3]It may also be possible to specify an additional between-person genetic twin model at level 2 for these means, but this is beyond the scope of this paper. We will return to this idea in the discussion.

applies to the normal distribution for the autoregressive effects: for a stationary model, autoregressive parameters are restricted in range. For a simple AR(1) model, the autoregressive coefficient should range from −1 to 1 for a stationary model. In this case, we rely on empirical identification of the model to keep the parameters in a stationary range. If the process is truly non-stationary, however, this approach may result in autoregressive parameters larger than 1 or smaller than −1. This approach is commonly taken for dynamic multilevel models (e.g., Bringmann et al., 2013; Epskamp et al., 2018; Jongerling et al., 2015; Schuurman, Ferrer et al., 2016; Schuurman, Grasman et al., 2016).

Intra-individual heritability

Based on the person-specific parameters in the model, we can calculate a heritability estimate of the intra-individual differences in the phenotype under study. This estimate h^2 reflects the heritability of intra-individual differences, that is, the proportion of variance in fluctuation in the phenotype over time that is due to genetic effects. To obtain this estimate, we first separate the between-person level differences – the estimated trait scores μ_{ij} – from the within-person fluctuations $(y_{ijt} - \mu_{ij})$. For these fluctuations, we determine the proportion of the variance due to genetic and unique environmental effects.

Note that because the factor loadings, autoregressive effects, and residual variances are allowed to differ from person to person, the total variances of the intra-individual differences may differ from person to person, and thus the intra-individual heritabilities may differ from person to person as well. Hence, we will obtain person-specific intra-individual heritability estimates. The within-person variance of the phenotype of a specific person $(var(y_{ij})_{intra})$ can be decomposed as follows, where $var(A_{ij})$ represents the variance of the genetic component of that person across the repeated measurements and $var(E_{ij})$ represents the variance of the environmental component of that person across the repeated measurements:

$$var(y_{ij})_{intra} = \lambda_{Aij}^2 var(A_{ij}) + \lambda_{Eij}^2 var(E_{ij}) + \sigma_{ij}^2, \quad (32)$$

where

$$var(A_{ij}) = \frac{var(v_{Aij})}{1 - \phi_{Aij}^2}, \quad (33)$$

and where

$$var(E_{ij}) = \frac{var(v_{Eij})}{1 - \phi_{Eij}^2}. \quad (34)$$

The functions for the person-specific within-person variances of E and A across the repeated measurements result because A and E have an autoregressive process (Kim & Nelson, 1999, p. 27). Based on these variances, we can calculate person-specific estimates of the intra-individual heritability h_{ij}^2:

$$h_{ij}^2 = \frac{\lambda_{Aij}^2 var(A_{ij})}{\lambda_{Aij}^2 var(A_{ij}) + \lambda_{Eij}^2 var(E_{ij}) + \sigma_{ij}^2}. \quad (35)$$

We can calculate this intra-individual heritability for each twin member separately. Similarly, we can estimate the contribution de_{ij}^2 of the unique environmental effects that are part of the dynamic process, or the contribution me_{ij}^2 of the unique environmental effects that are specific to each measurement occasion (including measurement error), as follows:

$$de_{ij}^2 = \frac{\lambda_{Eij}^2 var(E_{ij})}{\lambda_{Aij}^2 var(A_{ij}) + \lambda_{Eij}^2 var(E_{ij}) + \sigma_{ij}^2} \quad (36)$$

$$me_{ij}^2 = \frac{\sigma_{ij}^2}{\lambda_{Aij}^2 var(A_{ij}) + \lambda_{Eij}^2 var(E_{ij}) + \sigma_{ij}^2}. \quad (37)$$

Based on these person-specific estimates we can also determine some group descriptives, such as the average intra-individual heritability and the variance among the estimated intra-individual heritabilities across the sample of twins.

It is important to note again that the heritabilities estimated for cross-sectional models and dynamic models qualitatively differ, because these models investigate genetic and environmental components of different variables. The cross-sectional models focus on between-person differences in traits that are considered stable over time (at least for the period of the study). The corresponding heritability estimate reflects the proportion of variance of the differences across persons in those traits that is due to genetic effects. The dynamic models, on the other hand, focus on fluctuations within persons over time, that is, on the deviations from the stable traits over time. The intra-individual heritability of a particular person is the proportion of variance in their fluctuation over time that is due to genetic effects. Given the differences between the variables under consideration, there is little reason to expect heritability estimates of classic twin models and dynamic twin models to be similar or comparable, or that one can be easily generalized to another (as is the case generally for cross-sectional, between-person and within-person results, c.f., Adolf et al., 2014; Borsboom et al., 2003; Hamaker, 2012; Hamaker et al., 2015; Hoffman & Stawski, 2009a; Kievit et al., 2013; Molenaar, 2004).

Furthermore, note that the estimated intra-individual heritability may change depending on the time intervals between measurements, because this can affect the estimated dynamic processes for A and E. Revisiting the example from the introduction, consider the intra-individual heritability of agreeableness. The intra-individual heritability reflects to what extent fluctuations in a person's level of agreeableness from year to year, or day to day, or hour to hour, is the result of genetic effects or environmental effects. This is estimated by evaluating what proportion of the variability in the phenotype can be attributed to variation in the genetic effects. However, it may matter whether we evaluate this from year to year, or day to day, or hour to hour. For instance, it could be the case that the effects of genes on agreeableness vary more year to year, than they do hour to hour, and that such a difference is more pronounced for genetic effects than environmental effects. This would indicate that we would expect relatively little variance in A for

measurements taken closely together in time, resulting in a lower intraindividual heritability.

Relation with the genetic simplex and iFACE model

The multilevel dynamic twin model can be considered a bridge model between the iFACE model and the genetic simplex model. An overview of the differences in assumptions of the models is presented in Table 2. The main difference between the multilevel model and the iFACE model, is that in the multilevel model, the simplifying assumption is added that the twin-specific parameters come from common normal population distributions, while the iFACE model makes no such assumption. This assumption implies that the data of one person are to some extent informative about other persons, because the data of each person informs the estimates of the level 2 parameters, and those level 2 parameters in turn inform the estimates of the person-specific parameters at level 1. As a practical result, the number of persons in the sample can compensate for the number of time points, and vice versa, when estimating the multilevel model. A downside nonetheless is that this assumption might be wrong, which could negatively affect our estimates. Note however, that the normality assumption is essentially a prior in the Bayesian multilevel model. Hence, if enough data are available, this prior can become less influential as it becomes dominated by the data, which could limit the negative effects of making an erroneous assumption. Empirical evidence on the impact of the assumption is currently limited, but a recent study by Bulteel et al. (2018) showed that multilevel VAR models had better predictive performance than n = 1 VAR models, because the additional distributional assumptions regularize the model.

Our implementation of the multilevel model may also differ from iFACE implementations in that we specify the autoregressive effects of MZ twins to be identical, and hence essentially identical genetic components A. It may be possible to allow for distinct genetic components A_{ijt} and genetic autoregressive effects ϕ_{Aij} of the MZ twins in the multilevel model: This would entail specifying different latent variables A for both MZ twins, and assuming that the residuals of these latent variables are perfectly correlated. We expect that this model would be considerably harder to estimate.

The introduced multilevel genetic model corresponds to a stationary genetic simplex model, if all parameters in the multilevel model could be assumed to be the same across twins (i.e., no random parameters). This may be considered a stringent assumption, particularly assuming no differences among twins with respect to their mean levels (assuming no trait-like differences among twins). If one were to make this assumption while in reality there are differences in the twins' means, this will affect the resulting heritability estimates. For instance, it is known that if differences among means are disregarded in an autoregressive model, the resulting autoregressive effects would represent a blend of between-person associations and within-person associations; the exact mix depends on how much between-person and within-person variance is present in the data (Hoffman & Stawski, 2009a). Hence, heritability estimates based on a genetic simplex model without random means may also represent a blend of the intraindividual heritability, and the heritability of the traits, as noted earlier. This may be avoided, however, by appropriately person-mean centering the data for the genetic simplex model (c. f., Hamaker & Grasman, 2015; Hamaker et al., 2015), or by specifying a genetic simplex model with a random mean or intercept as discussed in the previous section. An interesting direction for future research is to further investigate how disregarding random effects of the various parameters (means, autoregressive effects, loadings and residual variances) each could affect the parameter estimates and the estimated heritabilities (see also related work by Dolan et al., 2015, on the effect of ignoring random loadings in classical ACE models).

Another difference with the genetic simplex model, is that the genetic simplex model does not necessarily assume stationarity, while we do assume stationarity in the multilevel model. This may not always be a realistic assumption, depending on the duration of the study and the psychological process in question. It may be possible to relax this assumption by extending the model to allow, for instance, for sudden changes (e.g., Chow & Zhang, 2013; De Haan-Rietdijk et al., 2016; Hamaker & Grasman, 2012; Kim & Nelson, 1999) or gradually changing parameters (e.g., Bringmann et al., 2017; Haslbeck et al., 2020; Lo, 2016; Molenaar et al., 1992), although these latter extensions of AR models and the iFACE model have so far been limited to the n = 1 setting.

Model implementation in Mplus' DSEM and simulation study

We investigated the performance of the multilevel genetic model presented in the previous section with a proof-of-principle simulation study. We implemented the model using

Table 2. Mapped out assumptions that differ for the four dynamic twin models discussed in this paper.

Assumption	Genetic Simplex	Random Intercept Genetic Simplex	iFACE	Multilevel
Parameters fixed/random across twins				
Means/intercepts	fixed	random	random	random
Autoregressive coefficients of A of MZ twins.	fixed	fixed	random	fixed per pair, random across pairs.
Other autoregressive coeffs.	fixed	fixed	random	random
Factor loadings	fixed	fixed	random	random
Measurement error variance	fixed	fixed	random	random
Correlation A DZ twins	fixed	fixed	random	random
Distributional assumptions random parameters across twins	na	Normal distribution[+]	None	Normal distribution[+]
Parameters fixed across time	No	Yes, for random interceps*	Yes*	Yes*

[+]Typically, but other distributions also possible. *This assumption may be relaxed.

Mplus v8's Dynamic SEM (DSEM; Asparouhov et al., 2018). The Mplus model specification is included in the supplemental materials (applied to the empirical example, but the same model specification as is used in the simulation study).

DSEM in Mplus v8 employs a Bayesian Gibbs Sampling procedure to estimate the model parameters. The details of the estimation procedure, including the conditionals for the DSEM framework in Mplus are described in Asparouhov et al. (2018). For an in depth derivation of the conditionals and posteriors for the estimation of a multilevel state space model with latent variables and autoregressive effects, as we have here, we refer the interested reader to Lodewyckx et al. (2011).

For each parameter to be estimated a prior distribution must be specified for the Bayesian analysis. For this implementation of the model, we employed the default priors of Mplus' DSEM, which are conjugate prior distributions, intended to be uninformative. For the measurement equation 20, we need to specify a prior distribution for the mean and variance of 1) the random means; 2) the random factor loadings for A and E; 3) the log-transformed random residual variances. For the transition equations 22 and 23 we need to specify a prior distribution for the mean and variance of the autoregression coefficients, and the variance of the random DZ correlation. By default Mplus applies non-informative normal distributions with a mean of 0 and a variance of 10^{10} for the means of all the parameters. For covariance matrices, Mplus applies Inverse-Wishart distributions, while for single variances it applies Inverse-Gamma distributions. In our current implementation we only model single variances, with the default improper Inverse-Gamma priors with a df of −1 and a scale of 0. This specification is used to get a prior distribution that is uninformative, the distribution nearing a uniform distribution with a wide range. Note that due to the uninformative nature of the prior distributions, the priors allow for a large range of values for each parameter. This is useful for specifying an uninformative prior distribution, but also means allowing for values that the parameter should not, or most likely will not take on. For example, for the variance of the random DZ correlation, a large variance would indicate that the correlations may take on values smaller than −1 or larger than 1. This is typically not problematic, if the model is sufficiently empirically identified by the data. However, particularly for complex models or models with relatively few observations, it can be beneficial to specify more informative prior distributions that restrict the values each parameter may take on. For the variance of the DZ correlation, we may for instance, restrict it to be very small (e.g., 0.01) such that the random correlations will not take on values outside of a −1 to 1 range. Particularly interesting would be to restrict the range based on theory or empirical knowledge about what variances should be expected. However, for this first implementation, we applied the default prior distributions of Mplus, given that these are most likely the prior distributions that will be applied by researchers that use DSEM in practice.

Note that although the multilevel dynamic twin model is in essence a multi-group model, with a group for the DZ twins and a group for the MZ twins, as is the cross-sectional model and Genetic Simplex model, multi-group modeling is not yet available for the DSEM module of Mplus. Instead, we implemented the model as a multivariate model with four observed variables, one for the DZ twins 1, one for the DZ twins 2, and one for the MZ twins 1 and one for the MZ twins 2. The implication of this strategy is that for designs where we have unequal amounts of twins in the four variables, or unequal amounts of repeated numbers for the twins, this results in missing observations for those variables with less observations. These missing observations are estimated alongside the model parameters during the Bayesian estimation procedure. This is done according to the model specified, based on a missing at random (MAR) assumption. The consequence of this is that the estimation procedure takes more time than would be the case if these missing observations were not included.

Simulation conditions and convergence

For the simulation study, we evaluated the performance of the model for various sample sizes. We chose these conditions both to show the consistency of the estimators, and to evaluate the performance of the model for sample sizes we may expect in practice. For models that account for measurement error with a single indicator, as is the case here, a fairly large amount of repeated measurements are necessary for the estimators to perform well (e.g., around 100 or more; Schuurman & Hamaker, 2019; Schuurman et al., 2015). The number of repeated measurements can be compensated to some extent by including many participants. In practice, the number of repeated measurements is typically around 50, although collecting more is becoming more common. The number of twins included is in practice typically similar to or larger than the number of repeated measurements. Hence, the conditions we chose vary from very small samples sizes, to more medium sample sizes that are more common in practice, to relatively large sample sizes, which may be necessary for the model to perform optimally. Particularly, we included the following conditions: 1) 25 MZ and 25 DZ twin pairs and 50 time points, 2) 50 MZ and 50 DZ twin pairs and 50 time points, 3) 50 MZ and 50 DZ twin pairs and 100 time points, 4) 100 MZ and 100 DZ twin pairs and 100 time points, 5) 100 MZ and 100 DZ twin pairs and 200 time points, 6) 200 MZ and 200 DZ twin pairs and 100 time points, and 7) 200 MZ and 200 DZ twin pairs and 30 time points. The last condition is intended to be representative of a situation we may encounter in empirical practice, where we have a large amount of twins, but few repeated measurements.

From preliminary results, it became clear that the residual (measurement error) variance σ^2_{ij} and the factor loading λ_{Eij} for the unique environmental components are the hardest to estimate. There is a trade-off where when the structural part of the environmental effects (loadings and autoregressive effects) is underestimated, the residual variance tends to be overestimated. This is a pattern that may be expected for autoregressive models where measurement error-like variance (the residuals ϵ_{ijt}) is separated from dynamic error variance (the residuals v_{ijt}) with a single indicator (Schuurman & Hamaker, 2019; Schuurman et al., 2015). Separating these sources of variance for a single indicator is relatively hard as it depends on how

strong the dynamic effects are. These types of variance can be distinguished only because the residuals v_{ijt} are carried over across time points via the autoregressive effect, while the residuals ϵ_{ijt} are not. This also implies that when there is no autoregressive effect, the model is no longer identified. Hence, distinguishing the two different types of unique environmental components will likely require relatively large sample sizes, particularly when dynamics are relatively weak. We know from Schuurman et al. (2015) that estimation becomes more difficult when the autoregressive coefficient is close to zero, and when the proportion of measurement error variance to the dynamic variance is large. Hence, we have picked two scenarios for the parameter values of our data-generating model: The first (worst case) scenario has low average autoregressive effect of 0.3 for the environmental components and a relatively large average residual variance (a log variance of −.6), meaning that on average about 56% of the unique environmental variance in y_{ijt} is measurement error-like variance. The second more optimistic scenario has a higher average autoregressive effect equal to 0.6 and smaller average residual variance (log variance of −1.2), meaning that on average about 33% of the unique environmental variance in y_{ijt} is measurement error-like variance. This latter scenario seems more realistic, given that these numbers are closer to what we have seen for measurement error AR models in the empirical literature (Hu et al., 2016; Schuurman & Hamaker, 2019; Schuurman et al., 2015). The true values chosen for the two scenarios are presented in Table 3 and Table 5 respectively.

For each of the conditions, we ran simulations with 100 replications, to keep the computational time feasible.[4] For each replication, we used two chains with 60,000 iterations of which half was discarded as burn-in. We evaluated the convergence of the model by test-running a number of models with multiple chains for each of the different conditions, visually inspecting the trace plots and density plots for the Monte Carlo samples, and by checking the Gelman Rubin statistics (i.e., Potential Scale Reduction Factors; Gelman & Rubin, 1992).

Based on these statistics, the means, and the model parameters related to the additive genetic effects converged well for all conditions, even for the smallest sample sizes. For these parameters the chains mixed well, seemed stable across iterations after the burn-in period, and the Gelman Rubin statistics approached 1. The (log) residual variance σ_{ij}^2 and the environment factor loading λ_{Eij} are harder to estimate: In the condition with low autoregressive coefficients and high measurement error variance, we see that the convergence for these parameters seems poor for the smallest sample sizes (25 or 50 MZ and DZ twins, 50 time points). The estimates for these parameters in these conditions are also not ideal, as we discuss below. For larger sample sizes in the low autoregression/high measurement variance condition, convergence for these parameters is considerably better, but not ideal, with strongly autocorrelated samples, which leads to slow mixing and slow exploration of the posterior. Hence, for estimating these parameters well, taking a large number of samples is desirable; we would consider 60,000 iterations to be a minimum.

For the condition with a higher autoregressive coefficient and smaller error variance, the model converges well for all but the smallest sample size (25 twins, 50 time points). For the smallest sample size the (log) residual variance σ_{ij}^2 and the environment factor loading λ_{Eij} converge sub-optimally, albeit better than in the worst-case condition. For the remaining sample sizes the procedure seems to converge for all parameters. For smaller sample sizes there is stronger autocorrelation among the samples for the residual variance and environmental factor loadings, such that plenty of iterations are advised to estimate these parameters

Table 3. Results of the simulations for the small ϕ_{Eij}, large σ_{ij}^2 condition. Bias, mean absolute error (MAE), and coverage rates (CR) for the 95% credible intervals, for the multilevel dynamic twin model model for different numbers of repeated measures (t) and MZ and DZ twins (n).

Parameter = true	bias			MAE			CR		
	n = 25 t = 50	n = 50 t = 50	n = 50 t = 100	n = 25 t = 50	n = 50 t = 50	n = 50 t = 100	n = 25 t = 50	n = 50 t = 50	n = 50 t = 100
$\gamma_\mu = .55$	.02	.01	.00	.07	.05	.06	.96	.97	.95
$\gamma_{\phi A} = .7$	−.01	.00	.00	.02	.02	.01	.98	.93	.96
$\gamma_{\phi E} = .3$	−.05	−.04	−.05	.10	.08	.07	.86	.84	.83
$\gamma_{\lambda A} = .55$	.00	.00	.00	.02	.01	.01	.95	.94	.95
$\gamma_{\lambda E} = .63$	−.08	.07	.09	.36	.16	.12	.76	.84	.82
$log\gamma_{\sigma^2} = -.6$	−.73	−.45	−.39	.85	.52	.43	.84	.83	.79
$\psi_{\mu MZ}^2 = .3$	.01	.00	.02	.07	.05	.05	.97	.98	.97
$\psi_{\mu DZ}^2 = .3$	.01	.00	.00	.08	.06	.05	.98	.97	.98
$\psi_{\mu MZ1MZ2}^2 = .25$	.05	.02	−.01	.12	.08	.08	.95	.95	.92
$\psi_{\mu DZ1DZ2}^2 = .125$	.06	.03	.03	.08	.07	.06	.98	.94	.95
$\psi_{\phi A}^2 = .005$	.00	.00	.00	.00	.00	.00	.98	.98	.98
$\psi_{\phi E}^2 = .01$	.01	.01	.00	.01	.01	.00	1	1	1
$\psi_{\lambda A}^2 = .008$	.00	.00	.00	.00	.00	.00	.99	.95	.98
$\psi_{\lambda E}^2 = .01$	.01	.01	.00	.01	.01	.01	.96	.97	.99
$log\psi_{\sigma^2}^2 = .1$	.10	.06	.04	.14	.09	.06	.97	.98	.96
$\psi_{\rho DZA}^2 = .01$	.10	.06	.03	.10	.06	.03	.82	.86	.94

[4]Completion of 100 iterations for one replication took approximately 17 to 105 seconds depending on the sample size. In the supplemental materials we have included the results of additional runs of the smallest sample size conditions, to illustrate the stability of the simulation results based on 100 replications.

Table 4. Results of the simulations for the small ϕ_{Eij}, large σ^2_{ij} condition. Bias, mean absolute error (MAE), and coverage rates (CR) for the 95% credible intervals, for the multilevel dynamic twin model model for different numbers of repeated measures (t) and MZ and DZ twins (n).

Parameter = true	bias				MAE				CR			
	n = 100 t = 100	n = 100 t = 200	n = 200 t = 100	n = 200 t = 30	n = 100 t = 200	n = 100 t = 200	n = 200 t = 100	n = 200 t = 30	n = 100 t = 100	n = 100 t = 200	n = 200 t = 100	n = 200 n = 30
$\gamma_\mu = .55$	.00	.01	−.01	.00	.04	.04	.03	.03	.93	.96	.93	.95
$\gamma_{\phi A} = .7$	.00	.00	.00	.00	.01	.01	.00	.01	.99	.97	1	.93
$\gamma_{\phi E} = .3$	−.02	−.01	−.01	−.03	.05	.04	.03	.06	.88	.85	.94	.86
$\gamma_{\lambda A} = .55$	.00	.00	.00	.00	.01	.01	.00	.01	.95	.97	.96	.94
$\gamma_{\lambda E} = .63$	.04	.02	.02	.05	.07	.05	.04	.09	.84	.86	.93	.83
$\log\gamma_{\sigma^2} = -.6$	−.13	−.07	−.06	−.19	.18	.13	.10	.25	.87	.89	.94	.82
$\psi^2_{\mu MZ} = .3$	.01	.01	.00	.01	.04	.04	.03	.03	.98	.94	.99	.93
$\psi^2_{\mu DZ} = .3$	.00	.00	.00	−.01	.04	.04	.03	.04	.98	.94	.93	.93
$\psi_{\mu MZ1MZ2} = .25$	.00	.00	−.01	.00	.06	.05	.04	.05	.96	.97	.98	.96
$\psi_{\mu DZ1DZ2} = .125$	.01	.01	.00	.03	.05	.04	.03	.05	.94	.97	.95	.95
$\psi^2_{\phi A} = .005$	.00	.00	.00	.00	.00	.00	.00	.00	.91	.97	.96	.98
$\psi^2_{\phi E} = .01$	.00	.00	.00	.00	.00	.00	.00	.00	.98	.98	.97	.98
$\psi^2_{\lambda A} = .008$	.00	.00	.00	.00	.00	.00	.00	.00	.96	.97	.94	.97
$\psi^2_{\lambda E} = .01$	.00	.00	.00	.01	.01	.00	.00	.01	.92	.93	.94	.95
$\log\psi^2_{\sigma^2} = .1$	.00	.00	.00	.00	.03	.02	.02	.04	.99	.93	.95	.94
$\psi^2_{\rho DZA} = .01$	.02	.01	.01	.04	.02	.01	.01	.04	.95	.96	.91	.87

optimally; but the chains mix, seem stable, and the Gelman Rubin Statistics approach 1.

Simulation results

We evaluate the estimation performance for each condition by determining the bias, absolute mean errors, and coverage rates for the 95% credible intervals for the parameters of the model. The results are presented in Table 3 to 6. We first discuss the results for the worst-case scenario with low autoregressive coefficients and high error variance, followed by the scenario with high autoregressive coefficients and lower error variance.

Relatively small φ_{Eij}, large σ^2_{ij} condition. For the smaller samples sizes of 25 and 50 twin pairs with 50 time points, the estimation procedure for 1 of the 100 replications could not be completed. For the remaining conditions, all replications were successfully completed. We will base our estimation performance results on the completed replications.

As can be seen from Tables 3 and 4, overall, bias and average absolute error decrease, and coverage rates improve, as sample size increase, as we would expect. Most parameters are estimated fairly well, even for low samples sizes. However, there is bias in the model for the dynamics pertaining to the fixed effects for the unique environmental components and residuals, particularly for small sample sizes: There is some trade-off where the structural part of the environmental effects (loadings and autoregressive effects) is underestimated, while the residual variances tend to be overestimated. As discussed previously, effectively distinguishing the two different types of unique environmental effects will require relatively large sample sizes, particularly when dynamics are relatively weak, which is the case in this condition. Due to the bias for the environmental loadings and variances, we see that on a few occasions their coverage rates do not improve as the sample size increases, even though the absolute errors and bias have decreased. In these cases the increased sample sizes increased precision of the estimates, but bias did not decrease

Table 5. Results of the simulations for the larger ϕ_{Eij}, smaller σ^2_{ij} condition. Bias, mean absolute error (MAE), and coverage rates (CR) for the 95% credible intervals, for the multilevel dynamic twin model model for different numbers of repeated measures (t) and MZ and DZ twins (n).

Parameter = true	bias			MAE			CR		
	n = 25 t = 50	n = 50 t = 50	n = 50 t = 100	n = 25 t = 50	n = 50 t = 50	n = 50 t = 100	n = 25 t = 50	n = 50 t = 50	n = 50 n = 100
$\gamma_\mu = .55$	.01	.01	.00	.07	.05	.06	.96	.97	.95
$\gamma_{\phi A} = .7$	.00	.00	.00	.02	.02	.01	.97	.95	.96
$\gamma_{\phi E} = .6$	−.04	−.02	−.01	.06	.04	.02	.90	.97	.97
$\gamma_{\lambda A} = .55$	.00	.00	.00	.02	.01	.01	.96	.95	.95
$\gamma_{\lambda E} = .63$	.04	.02	.01	.06	.04	.02	.88	.91	.89
$\log\gamma_{\sigma^2} = -1.2$	−.24	−.09	−.04	.28	.13	.08	.90	.90	.94
$\psi^2_{\mu MZ} = .3$	.01	.00	.01	.08	.06	.05	.95	.97	.97
$\psi^2_{\mu DZ} = .3$	.00	.00	.00	.09	.07	.05	.98	.96	.97
$\psi_{\mu MZ1MZ2} = .25$	.05	.02	−.01	.12	.09	.08	.96	.97	.96
$\psi_{\mu DZ1DZ2} = .125$	.07	.04	.03	.09	.07	.06	.98	.94	.97
$\psi^2_{\phi A} = .005$	.00	.00	.00	.00	.00	.00	.99	.98	.97
$\psi^2_{\phi E} = .01$	.01	.00	.00	.01	.00	.00	.98	.97	.97
$\psi^2_{\lambda A} = .008$	.00	.00	.00	.00	.00	.00	.99	.96	.93
$\psi^2_{\lambda E} = .01$	.00	.00	.00	.01	.00	.00	.98	.99	.95
$\log\psi^2_{\sigma^2} = .1$	.05	.01	.00	.07	.05	.03	1	.97	.98
$\psi^2_{\rho DZA} = .01$	.09	.05	.03	.09	.05	.03	.85	.86	.94

equivalently with it, resulting in slightly lower or similar coverage rates as in lower sample size conditions.

Note that distinguishing between the variances for ϵ and v_E is relatively hard, but this does not apply to v_A: The genetic residuals are distinguished from the rest due to the different correlation structures for MZ and DZ twins. As a result, the parameters pertaining to the additive genetic effects are generally estimated well. The additive genetic parameters are estimated with little bias and mean absolute errors, and good coverage – even for the smallest sample sizes. As such, we can get a good impression of the intra-individual heritability of the modeled process, even for small sample sizes.

Finally, the DZ correlation variance is generally overestimated, especially for conditions with smaller numbers of twins. The parameter is estimated with very little certainty/precision, and on occasions large values for this variance are sampled, which inflates the point estimate for this parameter. The estimates for this variance improve the most by increasing the number of repeated measures for the DZ twins. An option for smaller sample sizes may be to keep the DZ correlation fixed to 0.5 rather than allowing it to vary across DZ twins. Another option is to make use of more informative priors to restrict the range of the variance of the random DZ correlation to the most realistic values, rather than using Mplus' default uninformative priors.

Relatively large φ_{Eij}, small σ_{ij}^2 condition. For the condition with 100 MZ and 100 DZ twins and 100 time points, the estimation procedure for 1 of the 100 replications could not be completed. For the remaining conditions, all replications were successfully completed. We base our estimation performance results on the completed replications.

For the condition with large ϕ_{Eij}, and small σ_{ij}^2, the overall pattern is very similar to the previously discussed condition, as can be seen from Tables 5 and 6. The environmental parameters are the hardest to estimate, while the genetic parameters are estimated quite well, even for smaller sample sizes. As sample size increases, the bias and mean absolute errors decrease, and coverage becomes closer to 0.95. Overall, the estimates for this condition are considerably better than those for the small ϕ_{Eij}, and large σ_{ij}^2 condition. The bias and average absolute errors are fairly small even for the smallest sample sizes, and from samples sizes of 50 MZ and DZ twins and above the coverage rates for most parameters are around .95 and reach .90-.96 for the environmental fixed effects.

Empirical example: Moment-to-moment feelings of zen

To illustrate the multilevel dynamic twin model, we applied it to an empirical data set collected by Wichers et al. (2007). Wichers et al. (2007) collected momentary assessments of female Belgian MZ and DZ twins with ages ranging between 18 and 46 years using experience sampling method, collecting various measurements of mood, positive affect, negative affect and stress. The twins received random alerts to fill in self-assessments forms 10 times a day for five consecutive days. More details on the data collection can be found in Wichers et al. (2007).

For the present purposes, we focus on 148 MZ and 88 DZ twins from this data set from whom more than 25 repeated measurements were collected for the items "I feel relaxed," "I feel satisfied," and "I feel calm."[5] Each of these items was rated on a seven-point Likert scale, with higher scores indicating higher levels of the respective feeling (1 = "Not at all," 7 = "Very"). For this empirical example, we took the sum score of these items as the indicator of how 'Zen' women felt from moment-to-moment (our phenotype variable). On average, 38 repeated measures were collected per twin member, with a minimum of 26 and a maximum of 50 observations per twin member. Next to these observations, we added missing observations to the data to make the repeated measures more evenly

Table 6. Results of the simulations for the larger ϕ_{Eij}, smaller σ_{ij}^2 condition. Bias, mean absolute error (MAE), and coverage rates (CR) for the 95% credible intervals, for the multilevel dynamic twin model model for different numbers of repeated measures (t) and MZ and DZ twins (n).

Parameter = true	bias				MAE				CR			
	n = 100 t = 100	n = 100 t = 200	n = 200 t = 100	n = 200 t = 30	n = 100 t = 100	n = 100 t = 200	n = 200 t = 100	n = 200 t = 30	n = 100 t = 100	n = 100 t = 200	n = 200 t = 100	n = 200 t = 30
$\gamma_\mu = .55$	.00	.01	−.01	.00	.04	.04	.03	.03	.94	.97	.91	.95
$\gamma_{\phi A} = .7$	.00	.00	.00	.00	.01	.01	.00	.01	.99	.95	.99	.94
$\gamma_{\phi E} = .6$	.00	.00	.00	−.01	.02	.01	.01	.02	.95	.94	.96	.95
$\gamma_{\lambda A} = .55$	.00	.00	.00	.00	.01	.01	.00	.01	.94	.94	.94	.94
$\gamma_{\lambda E} = .63$	.00	.00	.00	.01	.02	.01	.01	.02	.91	.94	.91	.94
$log\gamma_{\sigma^2} = -1.2$	−.01	−.01	−.01	−.03	.05	.04	.04	.07	.92	.95	.88	.93
$\psi_{\mu MZ}^2 = .3$	.00	.00	.00	.01	.04	.04	.03	.04	.98	.97	.96	.93
$\psi_{\mu DZ}^2 = .3$	.00	.00	.00	−.01	.04	.04	.03	.04	.96	.93	.95	.93
$\psi_{\mu MZ1MZ2}^2 = .25$	.00	.00	.00	.00	.06	.05	.04	.05	.96	.98	.98	.95
$\psi_{\mu DZ1DZ2}^2 = .125$	.01	.01	.00	.03	.05	.04	.03	.05	.96	.97	.96	.93
$\psi_{\phi A}^2 = .005$	.00	.00	−.01	.00	.00	.00	.00	.00	.95	.96	.97	.98
$\psi_{\phi E}^2 = .01$	.00	.00	.00	.00	.00	.00	.00	.00	.95	.93	.94	1
$\psi_{\lambda A}^2 = .008$	.00	.00	.00	.00	.00	.00	.00	.00	.93	.93	.92	.95
$\psi_{\lambda E}^2 = .01$	.00	.00	.00	.00	.00	.00	.00	.00	.94	.89	.95	.97
$log\psi_{\sigma^2}^2 = .1$	.00	.00	.00	−.01	.02	.02	.01	.03	.98	.95	.97	.97
$\psi_{\rho DZA}^2 = .01$	.02	.01	.01	.04	.02	.01	.01	.04	.93	.92	.93	.85

[5]Translated here from the Dutch items "Ik voel me ontspannen", "Ik voel me tevreden", and "Ik voel me kalm" respectively. The answer options had a seven point likert scale, with labels for answer categories 1 "niet", 4 "matig", and 7 "zeer".

spaced within each twin, such that there are similar amounts of time in between all of the repeated measurements, and that the measurements of the twin members synchronize over time. We did this by calculating the time intervals between the observations of each twin, and inserting extra missing observations where there were relatively large time intervals. These missing observations are estimated during the analysis based on the model, following a MAR assumption.[6] We followed up with a manual inspection to ensure the timings of the measurements synced up for the repeated measurements of the two twins in each pair.

We investigated the genetic and unique environmental components of the Zen feelings experienced by the twins, particularly if there is carry-over of these components over time; to what extent each contributes to the Zen feelings of the twins; and the individual differences in these effects among twins. To this end, we fit the model as specified in Equation (20) to (31) for the Zen sum score, making use of the DSEM module of Mplus version 8 (Muthen & Muthen, 2017) (see the supplemental materials for the Mplus code).

Judging from the Gelman–Rubin statistics and a visual inspection of the trace plots the model appears to have converged. The samples were however strongly autocorrelated, as we found for smaller sample sizes in the simulation study. Hence we took a large amount of samples (200,000). However, the convergence of the variance of the random DZ correlation was poor: Its chains sometimes shoot up to high values (e.g., 1) and return slowly back to mixing with the other chains near zero. The effects of this are visible in the estimates with unrealistically high point estimate for this variance, as well as a unlikely large credible interval. This is most likely due to that the data contain relatively few DZ twins and repeated measures. As can be seen from our simulation study, many DZ twins and repeated measures are necessary to obtain good estimates for this variance. We fixed the DZ correlation (no variability across twins), to see if

this would impact the estimates/convergence of other parameters in the model: The parameter estimates with fixed and random DZ correlations were extremely similar. In the following paragraphs, we focus on the results of the fully random model to illustrate the interpretation of the model's results and the heritability estimates.

Results for dynamic multilevel twin model

The point estimates and 95% credible intervals for the fixed effects and the variances of the random effects are presented in Table 7 for the model with all random effects. We discuss the results for this fully random model below. Note that we had a relatively low number of time points for each twin, although our sample of twins is relatively large. The results of our simulation study indicate that in these circumstances the estimates of particularly the fixed effects for the unique environmental parameters should be interpreted with caution.

Autoregressive effects & factor loadings

For the additive genetic components, the average autoregressive effect is estimated to be very high and positive (.95), and the estimated variance around this average autoregressive effect across women is very close to zero. As such, for all women we would expect strong autoregressive effects for the genetic components, which indicates that there is very strong carry-over in the genetic effects from moment to moment. This indicates that the genetic effects of Zen feelings do not change a lot over this time period. The average factor loading for the genetic components is estimated to be small (.06), but there is some variability (0.005) in this loading among twins. The estimated autoregressive effects for the unique environmental components are much lower, but positive, with an average of .52 and a variance of .01 across women, indicating autoregressive effects ranging between approximately .3 and .7. That is, for all women we expect some carry-over for the unique environmental components, but for some women this carry-over is stronger than others. Such autoregressive effects and variability therefore are in line with what we typically see in the literature for the fluctuation of affect variables over time (especially when measurement error has been taken into account, as is the case here, c.f., Schuurman & Hamaker, 2019; Schuurman et al., 2015). The factor loadings of the unique environmental components are estimated to be considerably larger than those for the genetic components, with an average loading of .51. However, the variability around this average effect is quite large (.07), such that we would expect loadings ranging from approximately 0 to 1 across women.

Table 7. Parameter estimates for the bivariate multilevel MEVAR(1) and VAR(1) model for women in a relationship, modeling the relationship between daily relationship and general positive affect.

parameter	estimate [95% CI] FEs	Variances REs
ϕA	.95 [.94, .97]	.00 [.00, .00]
ϕE	.52 [.43, .60]	.01 [.00, .03]
λA	.06 [.04, .08]	.01 [.00, .01]
λE	.51 [.42, .59]	.07 [.05, .10]
σ^2	.43 [.34, .51]	.16 [.10, .27]
$\log\sigma^2$	−1.17 [−1.41, −.98]	.64 [.51, .82]
ρ_{DZ}	0.5	.24 [.01, .93]
μ_{MZ}	4.81 [4.73, 4.89]	.25 [.19, .34]
μ_{DZ}	4.81 [4.73, 4.89]	.36 [.25, .52]
$\rho_{\mu MZ}$	.25 [.15, .38]	–
$\rho_{\mu DZ}$	.13 [.02, .29]	–
h^2	.12 [.09, .15]	.02 [.01, .03]
de^2	.45 [.39, .59]	.06 [.06, .07]

Note that the variance reported for μ_{MZ} and μ_{DZ} is a *residual* variance, not the total variance. Note that the estimates for h^2 and de^2 concern sample means and variances.
FEs: Fixed effects. REs: Random effects.

Residual (measurement error) variances

The variances of the effects on the phenotype that are specific to each time point are estimated to be .43 on average; but there is also quite some variability in these

---

[6]Note that the DSEM framework in Mplus includes a function 'tinterval' that can be used to add missing values between observations to make the observations more equally spaced as well. We did it manually, because our variables are the phenotypes of the two MZ and two DZ twins, and hence our different variables were not necessarily measured at the same moments. Hence each variable needed unique adjustments, such that we couldn't easily use the tinterval function.
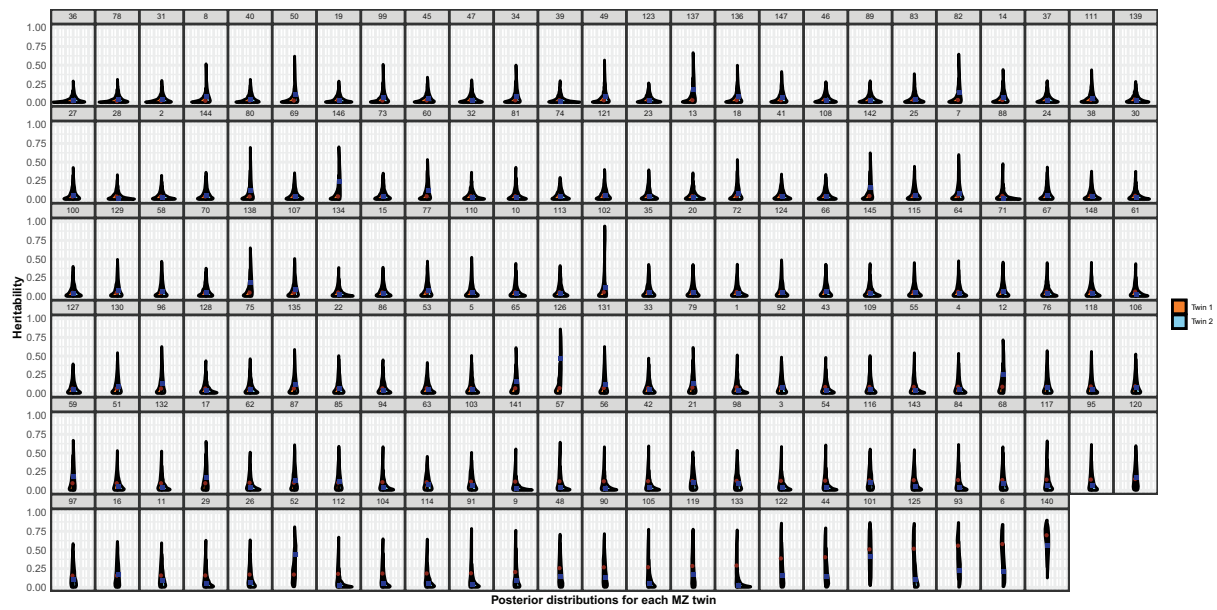
Figure 3. Plots of the posterior distributions for the heritabilities of Zen feelings for MZ twin 1 (blue/right), and MZ twin 2 (orange/left). The dot in each distribution represents the median for that distribution. The tails of the posteriors are trimmed at their 95% credible intervals. The posteriors are ordered across MZ twins 1 based on their median heritability, from low to high (top left to bottom right).
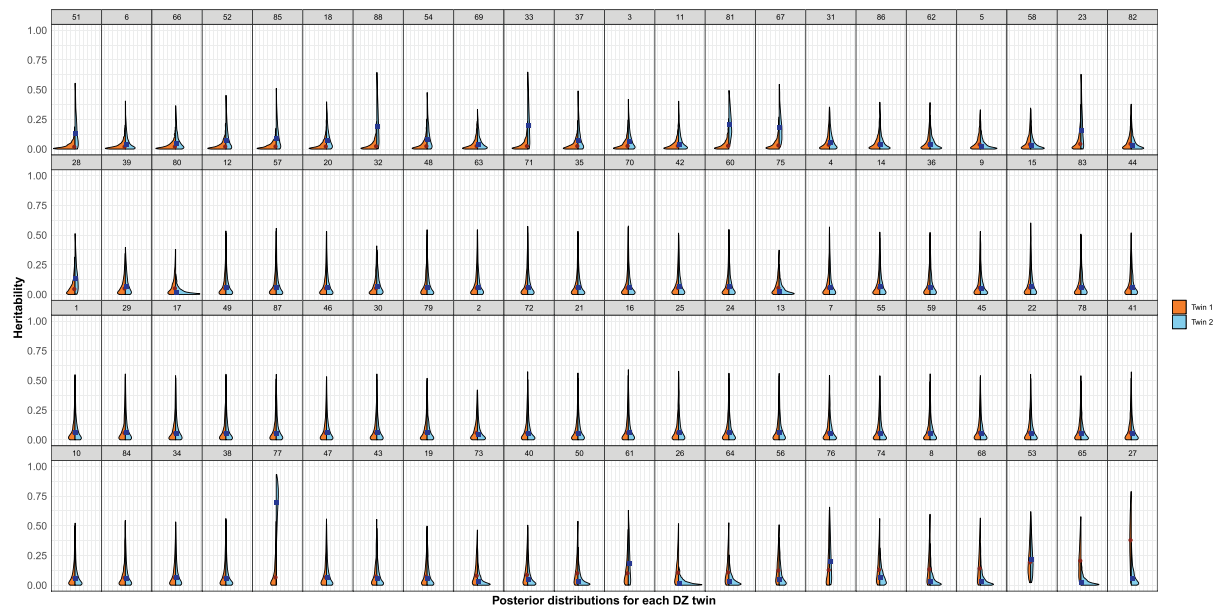

Figure 4. Plots of the posterior distributions for the heritabilities of Zen feelings for DZ twin 1 (blue/right), and DZ twin 2 (orange/left). The dot in each distribution represents the median for that distribution. The tails of the posteriors are trimmed at their 95% credible intervals. The posteriors are ordered across DZ twins 1 based on their median heritability, from low to high (top left to bottom right).

variances across twins. This indicates that some people are either more sensitive to, or experience more, unique environmental effects on their Zen feelings that are unique to each particular measurement occasion, than others.

Intra-individual person-specific heritabilities
Given that there is variability in the autoregressive effects and factor loadings across persons, the estimated intra-individual heritabilities will also differ from person to person. Given that there is considerably more variability in the

loadings and autoregressive effects for the unique environmental components than for the genetic components, differences in heritability can be attributed primarily to the differences in unique environmental effects. On average, across twins the intra-individual heritability of Zen feelings is estimated at 0.13, which indicates that on average about 13% of the moment-to-moment fluctuations in Zen feelings in women are due to additive genetic effects, while the majority of fluctuations of Zen feelings would be due to unique environmental effects. In other words, on average

across women, the moment-to-moment fluctuations in their Zen feelings seems to be predominantly driven by the environment.

There is considerable variation in the estimated heritabilities from woman to woman: for some the contribution is less than 13%, and for others more than 13%. Figure 3 and Figure 4 show the posterior distributions of the person-specific heritabilities for the MZ and DZ twins. From these distributions it can be seen that there is a fair amount of uncertainty around the person-specific heritability estimates, as one would expect given that we had relatively few repeated measures per person. It is clear, however, that for most women the heritabilities are estimated to be roughly between 0 and 0.25.

Trait scores or means

The average trait score (average of the means) for Zen feelings is estimated at 4.81. This indicates that on average, women show moderate to high levels of average experienced Zen feelings. There is considerable variation in the mean levels among the twins (residual variances of .25 for the MZ and .36 for the DZ twins). We see that the correlation among the means of the MZ-twins (.25) is higher than that of the DZ twins (.13), which indicates that trait Zen feelings may be heritable. That is, variation in trait Zen feelings can be explained in part by variation in genetic effects across persons, that is, evaluated on a between-person level, akin to cross-sectional twin models.

Discussion

In this paper, we introduced the multilevel dynamic twin model, and discussed the relation of this novel model to two other dynamic twin models: the genetic simplex model and the iFACE model. All three models can be used to investigate the intra-individual heritability of phenoytypes that fluctuate over time by modeling latent genetic and environmental components with an autoregressive process. However, these three models have different underlying assumptions. The multilevel model connects the genetic simplex model and iFACE model: while the genetic simplex model assumes the same dynamic structure for all twins, the iFACE model is fitted to separate twin pairs and consequently assumes that the dynamics of different people are not informative for others. The multilevel model forms a middle ground between these two models by accounting for differences in the parameters of the dynamic processes of different people, and assuming that the parameters of each person's dynamic model come from a common population distribution. As a result, the model can be used to model person-specific contributions of genetic and environmental effects, as well as individual differences therein.

We emphasize that dynamic twin models estimate intra-individual heritability, which is substantively and conceptually distinct from the classic interpretations or estimates of heritability. While classic cross-sectional studies consider how much variance in the stable differences between persons can be explained by variation in their genes, intra-individual heritability considers how much variance in people's phenotypic fluctuations over time can be attributed to fluctuations in their genetic effects.

Interestingly, an extension of the multilevel model introduced here may be able to estimate both kinds of heritability simultaneously. For example, we find in our empirical example that the random means μ_{ij}, the traits scores are correlated more strongly for MZ twins than DZ twins, which implies that these scores may be heritable. It could be possible – either in one step within the multilevel model, or with a two-step procedure – to specify a full A(C)E classic twin model for the random means, to determine what stable between-person differences are due to additive genetic and environmental effects. The AE model for the random means would then become $\mu_{ij} = A_i^b + E_i^b$. By determining the proportion of variance of between-person additive genetic effects (A_i^b) relative to the total between-person variance (ψ_μ^2), one would obtain the heritability for the stable, trait-like differences between persons in the phenotype (trait-heritability), next to the person-specific intra-individual heritabilities discussed earlier in the paper. Similarly, one may consider investigating the heritability of other model parameters, such as the autoregressive coefficients (e.g., see, Zheng & Asbury, 2019, who implemented an ACE model for autoregressive parameters in an AR(1) model).

The distinction between these two types of heritabilities, trait heritability and intra-individual heritability, is also relevant for cross-sectional twin studies. Although cross-sectional twin studies typically aim to estimate the heritability of stable traits, cross-sectional measurements generally pick up both stable trait variance, as well as some variance due to within-person fluctuations. If these within-person fluctuations are regulated by a dynamic process, and hence are not pure noise, the cross-sectional heritability estimates may also represent a blend of intra-individual and between-person heritability. An interesting endeavor for future research would be to investigate the effects of not separating stable between-person differences and within-person fluctuations on the genetic twin models. Relatedly, it would also be worthwhile to further study the effects of disregarding different kinds of between-person differences in the parameters of dynamic twin models, for example, in their autoregressive effects, factor loadings, variances, or even structural differences in their models.

In this paper, we have provided a proof-of-principle simulation study on the performance of the multilevel dynamic twin model. We found that to estimate the environmental effects well, the multilevel dynamic twin model requires a large number of both twins and repeated measures. The large number of measurements is necessary for distinguishing the dynamic unique environmental effects from time point-specific unique environmental effects and measurement errors in this model. The parameters pertaining to the genetic effects nonetheless can be estimated quite accurately even with smaller sample sizes, which provides promising prospects for estimating intra-individual heritability in less ideal settings.

An option to ease the estimation of the dynamic unique environmental effects and the occasion-specific unique environmental effects could be to use a multiple indicator factor structure as a measurement model for the phenotype (i.e., a dynamic factor twin model). This model may prove easier to empirically identify and estimate; however, it would require establishing indicators that function essentially as parallel tests within persons (c.f., Hu et al., 2016). It arguably would also require establishing measurement invariance across the twins, which may be unrealistic.

To more fully evaluate the performance of an implementation of the dynamic multilevel model, it would be relevant to include more conditions, including additional sample sizes and parameter values. For example, the study by schuurer on n = 1 AR(1) models with measurement error indicates that the quality of the estimated parameters can differ depending on the value of the autoregressive effects and amounts of measurement error variance. We saw an example of this in our simulation study. It is also worthwhile to investigate the effects of unbalanced designs, where we have more MZ twins than DZ twins and vice versa, to see if and how this affects the estimation, and how this interacts with including a common environmental effects latent variable. In the current multivariate setup, rather than using a typical multi-group model, an additional effect of an unbalanced design is that potentially many missing observations need to be estimated alongside the model parameters. Finally, it would be useful to evaluate the effects of different kinds of prior distributions, rather than Mplus' default options. For example, a more informative prior distribution that restricts the values that the parameters may take on may improve the estimation for the DZ correlations between the genetic latent variable, or for the random variances in the model MBRWish, Gelman (2006).

Despite its complexity, the multilevel model we introduced represents a basic multilevel dynamic twin model, and is hence limited in various ways. For instance, we worked with an AE model for the sake of simplicity, which disregards that there may also be shared environmental effects between both the MZ and DZ twins. This could be taken into account by specifying an ACE model, including a latent variable that models common effects among the twin members. For such a model, it also seems particularly pertinent in the context of longitudinal data, to account for whether the twins are together or apart at each measurement occasion.

One may also consider extending the model by specifying a multivariate model at level 2, allowing all level 2 parameters to be correlated with each other. Estimating such a large covariance matrix for the random parameters with any precision will, however, require many twins. Next to this, extra caution is needed when specifying the prior distribution for this covariance matrix: The typical conjugate Inverse Wishart prior distribution is known to easily induce bias for this type of model when the variances are expected to be close to zero, which would be the case for these random coefficients given their scale (c.f., Schuurman, Grasman et al., 2016; Schuurman & Hamaker, 2019; Song & Ferrer, 2012).

The multilevel twin model, as well as the genetic simplex and iFACE model, account for both unique environmental effects that carry over from occasion to occasion, as well as unique environmental effects that are specific to one occasion which also accounts for measurement error. The models do not account for genetic effects that are specific to each measurement occasion. It is uncertain whether one would expect such effects to contribute considerably to the within person variance. In our empirical example, we found that the autoregressive effects for the genetic effects are quite high, so in this scenario it seems unlikely. However, this may also depend in part on how close in time the repeated measurements are taken. For future work it may be interesting to consider such genetic effects, and how to account for them in the dynamic genetic models.

Furthermore, a limitation of the introduced multilevel dynamic twin model is that it disregards potential gene-environment interactions. Given that genetic effects change over time, it seems likely that environmental effects play some part in this change. The opposite applies as well: changes in genetic effects may play a role in what effects the environment may have. An interesting future development is to measure specific genetic or environmental characteristics and to model their effects on the dynamic processes of genetic and environmental effects. Another interesting option may be to specify a multivariate vector autoregressive structure among the genetic and environmental latent variables, allowing the genetic and environmental components to affect each other over time (e.g., similar to the extended simplex model by Dolan et al., 2014).

Finally, a limitation of both the genetic simplex, iFACE, and the dynamic multilevel model presented here, is that they only consider dynamics in terms of autoregressive processes of order 1. A myriad of other dynamic models could be considered, including higher order autoregressive processes, more complex time series models such as ARIMA models, (stochastic) differential equation models, or models for categorical variables such as Markov models. The autoregressive model of order 1 nonetheless represents a reasonable starting point for dynamic twin models, given that these autoregressive models can serve as a basis for many other dynamic models, such as models from the state space modeling framework (Harvey, 1989; Kim & Nelson, 1999), and the DSEM framework we used in this paper (Asparouhov et al., 2018).

For applications using dynamic twin models, it will generally be of essence to consider the time scale of changes in the phenotype, environmental effects, and genetic effects. For example, in our empirical example, experience sampling measurements were taken multiple times a day, and our results showed that at this time scale there was little change in the genetic effects over time, as one would expect. As a result, the genetic effects can explain only a limited amount of variance over time in the phenotype, which may have been even less if we also took common environmental effects into account in the model. However, most essentially, these results imply that this type of dynamic twin model could be used to investigate the timescale at which one may expect relevant changes in the genetic effects for a particular phenotype, and whether this may differ from person to person.

In sum, although dynamic twin models have received relatively little attention compared to their cross-sectional counterparts, they can provide relevant and novel perspectives on the contributions of genes and environment to psychological and behavioral processes, and potentially on their interplay. The dynamic perspective on the contribution of genes and environment in these dynamic twin models fits in well with the renewed interest in epigenetics (Bell & Saffery, 2012; Bell & Spector, 2011; Kan et al., 2010). We hope that the multilevel dynamic twin model introduced here, as well as the discussion of intra-individual heritability, can help provide a springboard for further investigations of the heritability of intra-individual differences, to complement those of inter-individual

differences by leveraging genetically informative intensive longitudinal data.

ORCID

Y. Zheng 🔟 http://orcid.org/0000-0002-0302-0598
C. V. Dolan 🔟 http://orcid.org/0000-0002-2496-8492

References

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between subjects: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, 5, 883. https://doi.org/10.3389/fpsyg.2014.00883

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

Bell, J. T., & Saffery, R. (2012). The value of twins in epigenetic epidemiology. *International Journal of Epidemiology*, 41, 140–150. https://doi.org/10.1093/ije/dyr179

Bell, J. T., & Spector, T. D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics*, 27, 116–125. https://doi.org/10.1016/j.tig.2010.12.005

Boomsma, D. I., & Molenaar, P. (1987). The genetic analysis of repeated measures. i. simplex models. *Behavior Genetics*, 17, 111–123. https://doi.org/10.1007/BF01065991

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

Bringmann, L., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22, 409–425. https://doi.org/10.1037/met0000085

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, e60188. https://doi.org/10.1371/journal.pone.0060188

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). Var (1) based models do not always outpredict ar (1) models in typical psychological applications. *Psychological Methods*, 23, 740–756. https://doi.org/10.1037/met0000178

Charney, E. (2012). Behavior genetics and postgenomics. *Behavioral and Brain Sciences*, 35, 331–358. https://doi.org/10.1017/S0140525X11002226

Chow, S.-M., & Zhang, G. (2013, October). Nonlinear regime-switching state-space (rsss) models. *Psychometrika*, 78, 740–768. https://doi.org/10.1007/s11336-013-9330-8

De Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, 81, 217–241. https://doi.org/10.1007/s11336-014-9417-x

Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C., Bartels, M., & Boomsma, D. I. (2014). Ge covariance through phenotype to

environment transmission: An assessment in longitudinal twin data and application to childhood anxiety. *Behavior Genetics*, 44, 240–253. https://doi.org/10.1007/s10519-014-9659-5

Dolan, C. V., Molenaar, P. C., & Boomsma, D. I. (1991). Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data. *Behavior Genetics*, 21, 49–65. https://doi.org/10.1007/BF01067666

Dolan, C. V., Nivard, M. G., van Dongen, J., van der Sluis, S., & Boomsma, D. I. (2015). Methylation as an epigenetic source of random genetic effects in the classical twin design. *Advances in Genomics and Genetics*, 5, 305–315. https://doi.org/10.2147/AGG.S46909

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, 30, 1–42. https://doi.org/10.1111/j.2044-8317.1977.tb00722.x

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53, 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fulker, D., & Bonné-Tamir, B. (1982). *Human genetics, part a: The unfolding genome*. Alan R Liss.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1, 514–534. https://doi.org/10.1214/06-BA117A

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511. https://doi.org/10.1214/ss/1177011136

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). Guilford Publications.

Hamaker, E. L., & Grasman, R. P. P. P. (2012). Regime switching state-space models applied to psychological processes: Handling missing data and making inferences. *Psychometrika*, 77, 400–422. https://doi.org/10.1007/s11336-012-9254-8

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

Hamaker, E. L., Kuijper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116. http://dx.doi.org/10.1037/a0038889

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26, 10–15. https://doi.org/10.1177/0963721416666518

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

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

Haslbeck, J. M., Bringmann, L. F., & Waldorp, L. J. (2020). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, 1–30. https://doi.org/10.1080/00273171.2020.1743630

Heath, A. C., Kendler, K. S., Eaves, L. J., & Markell, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics*, 15, 439–465. https://doi.org/10.1007/BF01066238

Hewitt, J., Eaves, L. J., Neale, M., & Meyer, J. (1988). Resolving causes of developmental continuity or "tracking." i. longitudinal twin studies during growth. *Behavior Genetics*, 18, 133–151. https://doi.org/10.1007/BF01067836

Hoffman, L., & Stawski, R. S. (2009a). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, 6, 97–120. https://doi.org/10.1080/15427600902911189

Hu, Y., Nesselroade, J. R., Erbacher, M. K., Boker, S. M., Burt, A., Keel, P. K., Neale, M. C., Sisk, C. L., & Klump, K. (2016). Test reliability at the individual level. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 532–543. https://doi.org/10.1080/10705511.2016.1148605

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, mava, and classical approaches to the analysis of the human

behavior. *Psychological Bulletin*, 73, 311–349. https://doi.org/10.1037/h0029135

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

Kan, K.-J., Ploeger, A., Raijmakers, M. E. J., Dolan, C. V., & Van Der Maas, H. L. J. (2010). Nonlinear epigenetic variance: Review and simulations. *Developmental Science*, 13, 11–27. https://doi.org/10.1111/j.1467-7687.2009.00858.x

Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H., & Eaves, L. J. (2009). Modeling extended twin family data i: Description of the cascade model. *Twin Research and Human Genetics*, 12, 8–18. https://doi.org/10.1375/twin.12.1.8

Kendler, K., Gardner, C., & Lichtenstein, P. (2008). A developmental twin study of symptoms of anxiety and depression: Evidence for genetic innovation and attenuation. *Psychological Medicine*, 38, 1567–1575. https://doi.org/10.1017/S003329170800384X

Kievit, R., Frankenhuis, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, 4, 513. https://doi.org/10.3389/fpsyg.2013.00513

Kim, C.-J., & Nelson, C. R. (1999). *State–space models with regime switching*. The MIT Press.

Liu, C., Molenaar, P. C. M., & Neiderhiser, J. M. (2018). The impact of variation in twin relatedness on estimates of heritability and environmental influences. *Behavior Genetics*, 48, 44–54. https://doi.org/10.1007/s10519-017-9875-x

Lo, L. L. (2016). *Time series analysis and person-specific psychological development: State space modeling applications in behavior genetic and neurocognitive designs* [Unpublished doctoral dissertation]. The Pennsylvania State University.

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, 68–83. https://doi.org/10.1016/j.jmp.2010.08.004

Long, E., Verhulst, B., Aggen, S., Kendler, K., & Gillespie, N. (2017). Contributions of genes and environment to developmental change in alcohol use. *Behavior Genetics*, 47, 498–506. https://doi.org/10.1007/s10519-017-9858-y

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, 38, 79–95. https://doi.org/10.1038/hdy.1977.9

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

Molenaar, P. C. M. (2011). On the limits of standard quantitative genetic modeling of inter-individual variation: Extensions, ergodic conditions and a new genetic factor model of intra-individual variation. In K. E. Hood, C. T. Halpern, G. Greenberg, & R. M. Lerner (Eds.), *Handbook of developmental science, behavior, and genetics* (pp. 626–648). John Wiley & Sons.

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

Molenaar, P. C. M., De Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, 57, 333–349. https://doi.org/10.1007/BF02295422

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

Molenaar, P. C. M., & Raijmakers, M. E. (1999). Additional aspects of third source variation for the genetic analysis of human development and behaviour: A commentary on eaves et al. *Twin Research and Human Genetics*, 2, 49–52. https://doi.org/10.1375/twin.2.1.49

Molenaar, P. C. M., Smit, D. J. A., Boomsma, D. I., & Nesselroade, J. R. (2012). Estimation of subject-specific heritabilities from intra-individual variation: Iface. *Twin Research and Human Genetics*, 15, 393–400. https://doi.org/10.1017/thg.2012.9

Muthen, L. K., & Muthen, B. O. (2017). *Mplus user's guide. eighth edition*. [Computer software manual].

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

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

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*, 465, 721–727. https://doi.org/10.1038/nature09230

Plomin, R. (1986). Multivariate analysis and developmental behavioral genetics: Developmental change as well as continuity. *Behavior Genetics*, 16, 25–43. https://doi.org/10.1007/BF01065477

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, 3, 119–133. https://doi.org/10.1093/bib/3.2.119

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, 21, 206–221. http://dx.doi.org/10.1037/met0000062

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, 51, 185–206. https://doi.org/10.1080/00273171.2015.1065398

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, 24, 70–91. https://doi.org/10.1037/met0000188

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n= 1 psychological autoregressive modeling. *Frontiers in Psychology*, 6, 1038. http://dx.doi.org/10.3389/fpsyg.2015.01038

Song, H., & Ferrer, E. (2012). Bayesian estimation of random coefficient dynamic factor models. *Multivariate Behavioral Research*, 47, 26–60. https://doi.org/10.1080/00273171.2012.640593

Wichers, M. C., Myin-Germeys, I., Jacobs, N., Peeters, F., Kenis, G., Derom, C., Vlietinck, R., Delespaul, P., & Van Os, J. (2007). Evidence that moment-to-moment variation in positive emotions buffer genetic risk for depression: A momentary assessment twin study. *Acta Psychiatrica Scandinavica*, 115, 451–457. https://doi.org/10.1111/j.1600-0447.2006.00924.x

Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., & Xia, K., & others. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, 46, 430–437. https://doi.org/10.1038/ng.2951

Zheng, Y., & Asbury, K. (2019). Genetic and environmental influences on adolescent emotional inertia in daily life. *Journal of Youth and Adolescence*, 48, 1849–1860. https://doi.org/10.1007/s10964-019-01075-2

Zheng, Y., Molenaar, P. C., Arden, R., Asbury, K., & Almeida, D. M. (2016). Person-specific non-shared environmental influences in intra-individual variability: A preliminary case of daily school feelings in monozygotic twins. *Behavior Genetics*, 46, 705–717. https://doi.org/10.1007/s10519-016-9789-z