Chapter 4

Bayesian Dynamic Modeling to assess differential treatment effects on panic attack frequencies

Abstract

To represent the complex structure of intensive longitudinal data of multiple individuals, we propose a hierarchical Bayesian Dynamic Model (BDM). This BDM is a generalized linear hierarchical model where the individual parameters do not necessarily follow a normal distribution. The model parameters can be estimated on the basis of relatively small sample sizes and in the presence of missing time points. We present the BDM and discuss the model identification, convergence and selection. The use of the BDM model is illustrated using data from a randomized clinical trial to study the differential effects of three treatments for panic disorder. The data involves the number of panic attacks experienced weekly (73 individuals, 10 to 52 time points) during treatment. Presuming that the counts are Poisson distributed, the BDM considered involves a linear trend model with an exponential link function. The final model included a moving average parameter, and an external variable (duration of symptoms pre-treatment). Our results show that cognitive behavioral therapy is less effective on the reduction of panic attacks than serotonin selective re-uptake inhibitors or a combination of both. Post-hoc analyses revealed that males show a slightly higher number of panic attacks at the onset of treatment than females.

This accepted for publiction as: Krone, T., Albers, C. J., & Timmerman, M. E. (2016c). Bayesian dynamic modeling to assess differential treatment effects on panic attack frequencies. *Statistical Modelling (in press)*. doi:10.1177/1471082X16650777

4.1 Intensive longitudinal data to study psychological processes

In recent years the interest in studying the course of psychological processes has increased. One may think of, for example, the effects of psychological interventions across time and the role of external variables on these effects (Elkins & Moore, 2011; Kellett, 2007; Wild et al., 2006). To study these processes, intensive longitudinal data is obtained: data of one or more individuals are gathered repeatedly over time, in such a frequency and over such a time span that it characterizes the trends and dynamics of interest (Hamaker et al., 2015). The measurements can pertain to questionnaires that are administered on a weekly or daily basis, or even multiple times a day. The latter is typically referred to as ecological momentary assessment (EMA) (Larson & Csikszentmihalyi, 1983; Shiffman et al., 2008).

Intensive longitudinal data typically have specific characteristics, which may yield selecting a proper statistical model a challenging task. First, the amount of data available is often rather limited, in terms of the number of observed time points per individual and the number of observed individuals.

Second, missing data may easily occur. Commonly, the amount of incomplete data substantially increases with larger numbers of scheduled time points. The data can be incomplete because of completely missing individuals, or because of incidental missing values, where observations on one or a few time points within a series are lacking, or because of drop-out, where a series lacks observations after a specific time point. The latter may coincide with the drop-out of an intervention (e.g., therapy), but this is not necessarily the case. It is typically wise to use all available data in modeling the intensive longitudinal data, to reduce bias and uncertainty in the model estimates.

Third, intensive longitudinal data is usually collected among multiple individuals. Then the interest is to capture the intra-individual processes as well the interindividual differences in these processes, and possibly to relate these to the external variables. This requires a model that covers the hierarchical structure in the data, where time points are nested within individuals. A popular approach is the multilevel regression model for repeated measures (Bryk & Raudenbush, 1992; Goldstein, 2011; Snijders & Bosker, 1999). However, this model may be overly restrictive in empirical practice, because of its normality assumption of the individual parameters.

Fourth and final, measurements of psychological processes are typically made on discrete scales. Examples include binary scales, such as indicating the absence or presence of a certain behavior, ordinal polytomous scales, such as the well-known Likert-scale, and counts, such as the number of times a certain behavior occurred. The model fit can improve considerably by using a proper distribution

for the scales at hand, rather than the often applied normal distribution. The latter can be an approximation to the discrete scales at best.

These characteristics result in a couple of requirements for a proper intensive longitudinal data analysis method. The method must be able to build a model upon a relatively small amount of data, even though more data will improve the estimation and allow for more complex models to be estimated reliably. The method must be able to deal with missing data. The method should allow for achieving insight into the intra-individual processes and their inter-individual differences. Herewith it is important that the distribution of the individual parameters is not necessarily restricted to normality. Finally, the method must allow for the typically occurring discrete scales (i.e., binary, ordinal polytomous and counts). These requirements lead us to the Bayesian Dynamic Model (BDM) (West & Harrison, 1997), which can handle the combination of requirements mentioned. The remainder of the chapter is organized as follows. In the next section, we will introduce the general BDM-framework. To illustrate the usefulness of the BDM for modeling psychological processes, we present an empirical application to intensive longitudinal count-data from multiple individuals, who participated in a randomized clinical trial. We introduce three variants of the BDM and the three fitted BDMs will be interpreted, and the results will be compared to the previous modeling endeavor which used a frequentist multilevel model. We will conclude with a discussion pertaining to the general use of the BDM for intensive longitudinal data.

4.2 Bayesian Dynamic Model

The BDM, including its generalization the Bayesian Dynamic Generalized Linear Model (West & Harrison, 1997), is a Bayesian interpretation of the state space model. The BDM includes a latent score that is connected to the observed score using the so-called latent state vector. The BDM comprises three equations, namely the link function, the observation equation and the system equation. We will successively present these three equations, in view of jointly modeling intensive longitudinal data of multiple individuals.

Link function We model the distribution p of observed score $y_{t,n}$ of individual n $(n = 1, ..., N)$ at time t $(t = 1, 2, ..., T_n)$ using a latent score $y^*_{t,n}$. The link function we use is equal to the one used in generalized linear models for observed data (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989) and allows for modeling any distribution from the exponential family:

$$\text{Link function}: \qquad p(y_{t,n}) = g^{-1}(y^*_{t,n}).$$

Observation and system equations The observation equation connects the latent score $y^*_{t,n}$ of individual n at time t to the latent state vector $\boldsymbol{\theta}_{t,n}$:

$$\text{Observation equation :} \quad y^*_{t,n} = \boldsymbol{f}_{t,n}\boldsymbol{\theta}_{t,n} + \varepsilon_{t,n}, \tag{4.1}$$

where $\boldsymbol{f}_{t,n}$ $(1 \times r)$ is the design vector at time t for individual n, $\boldsymbol{\theta}_{t,n}$ $(r \times 1)$ is the latent state vector at time t for individual n and $\varepsilon_{t,n}$ is the white noise. The system equation models the evolution of the latent state vector over time:

$$\text{System equation :} \quad \boldsymbol{\theta}_{t,n} = \boldsymbol{G}_{t,n}\boldsymbol{\theta}_{t-1,n} + \boldsymbol{\eta}_{t,n},$$

where $\boldsymbol{G}_{t,n}$ $(r \times r)$ is the innovation matrix and $\boldsymbol{\eta}_{t,n}$ $(r \times 1)$ is the innovation noise vector, at time t for individual n.

Each instance of the vectors $\boldsymbol{f}_{t,n}$ and matrices $\boldsymbol{G}_{t,n}$ must be defined by the researcher. When there is no reason to assume that $\boldsymbol{f}_{t,n}$ and $\boldsymbol{G}_{t,n}$ differ over time and/or over individuals, it is advised to take $\boldsymbol{f}_{t,n}$ and $\boldsymbol{G}_{t,n}$ invariant over time and/or over individuals, to simplify the model and its interpretation. In what follows, we presume that both the design vector and the innovation matrix are invariant over time and individuals. This results in the use of a single vector $\boldsymbol{f}$ and a single matrix $\boldsymbol{G}$ per model.

Noise parameters Typically, it is assumed that the noise, as expressed through $\varepsilon_{t,n}$ and $\boldsymbol{\eta}_{t,n}$, follows a normal distribution with mean zero and standard deviation $\sigma_{\varepsilon_{t,n}}$ and covariance matrix $\boldsymbol{\Sigma}_{\eta_{t,n}}$, respectively. Varying (co)variances of the noise over time and/or individuals may be necessary to achieve a proper fitting model (Jongerling et al., 2015). However, it also complicates the model and may give rise to estimation difficulties due to identification issues and/or lack of data. To avoid these difficulties, the distribution is often taken to be invariant over time and individuals, giving σ_ε and $\boldsymbol{\Sigma}_\eta$ instead of $\sigma_{\varepsilon_{t,n}}$ and $\boldsymbol{\Sigma}_{\eta_{t,n}}$, respectively. As an alternative, advanced techniques such as variance discounting (West & Harrison, 1997, p. 194-195), may be employed. Though typically assumed, the noise distribution is not restricted to normality. Note that when the link function already includes the variance of $y^*_{t,n}$ (as is the case for Poisson data where the variance equals the mean, which is given by the link function), $\varepsilon_{t,n}$ can be omitted from Equation 4.1.

Priors As any Bayesian model, the BDM requires priors chosen by the researcher. The choice of priors generally depends on the expected posterior distribution. If there is little information on what is to be expected, a weak informative prior may be used. An example is a symmetrized reference prior for autocorrelations (Berger & Yang, 1994), which can essentially be used for all instances when it is only known that the parameter to be estimated is an autocorrelation.

When more information is present, an informative prior can be used. For example, when a parameter appeared to be between -0.2 and 0.2 in earlier, well-conducted and trusted studies, a normal prior with mean zero and standard deviation 0.1 may be used. Finally, a non-informative prior may be used, generally an uniform distribution encompassing all possible values for the indicated parameter. However, for parameters with theoretically an infinite range of possibilities, these slow down calculations. Further, a weak informative prior is often available or can be derived from what is known about the expected range of the posterior estimation of the parameters.

Missing data As stated, the BDM can handle both incidental missing data and drop-outs. Incidental missing data can be handled by not linking the observed score to the latent state when the observed score is missing. In this case, the observation equation and the system equation are estimated, but the link between the observed score and the latent score is not made. For drop-outs, the analysis will stop at T_n, the final observed time point for individual n.

External variables The BDM allows for inclusion of external variables in two ways. First, the external variable can be included as an active covariate. This can be done as a direct effect, for example by considering an external variable as an element of $\boldsymbol{F}_{t,n}$ in Equation 4.1, and as a moderator effect, for example letting an element of $\boldsymbol{\theta}_{t,n}$ be dependent on the level of a covariate. Second, inactive covariates can be implemented post-hoc, by examining the relation between any model parameter and an external variable after the model estimation. This can be done, for instance, by using partial correlations or linear regression, thereby accounting for confounding variables. In our empirical example, we will demonstrate both approaches.

Model estimation The BDM is estimated using Bayesian Markov Chain Monte Carlo (MCMC) estimation. For the MCMC estimation, we use Hamiltonian Monte Carlo (HMC), a generalization of the Metropolis-Hastings algorithm that allows for an efficient estimation of the parameters (Gelman, Carlin, et al., 2013). This is incorporated in the software RStan (R Core Team, 2015; Stan Development Team, 2015), which we used in our modeling.

Model convergence The BDM can be a fairly complex model, especially when parameters are allowed to differ over time and individuals. This may lead to estimation problems due to identification issues, where more than one solution fits the data equally well, or due to too little data in comparison to the complexity of the model. Both of these will result in non-convergence, which implies that the estimation procedure has failed to find the single, optimal solution.

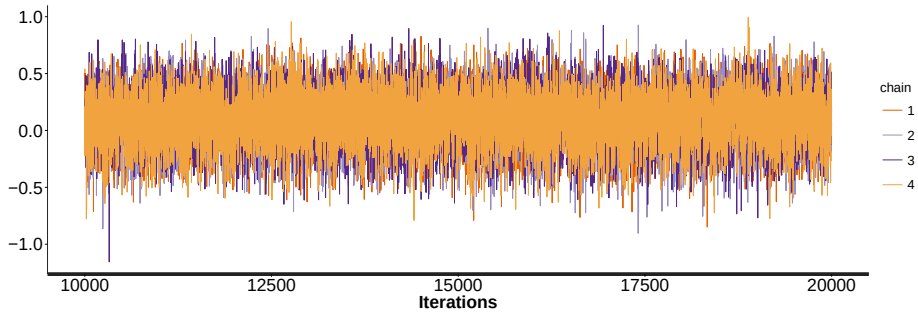

Figure 4.1: Traceplot of a converged parameter with a $\hat{R}$ of 1.00 and posterior mean 0.10

The convergence can be checked through visual inspection of the trace plots or through assessment of the potential scale reduction factor, $\hat{R}$. A trace plot shows the Bayesian MCMC estimates for each parameter at each iteration. If a parameter reaches convergence, the estimates over iterations are highly similar across chains. As a result, the trace plot looks like a fat caterpillar with all chains completely overlapping, except at the fringe of the caterpillar (as shown in Figure 4.1). The $\hat{R}$ expresses the ratio of how much the estimation may change when the number of iterations is doubled; the (ideal) value of 1 indicates that no change is expected (Gelman & Rubin, 1992; Stan Development Team, 2016).

Model selection To select a BDM from a series of competing variants, one may use two strategies. First, the model variants can be compared using the estimated parameters to see which parameters show the most preferable properties. For example, the noise variance is preferred to be small, indicating a proper model fit. Further, model parameters which are close to zero may be superfluous, thereby unnecessary complicating the model.

Second, competing models can be compared considering their fit (i.e., log-likelihood) and number of parameters. To this end, several information criteria can be used, such as the Deviance Information Criterion (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002), the Bayesian Information Criterion (Schwarz, 1978) and the Watanabe-Aikake Information Criterion (WAIC) (Watanabe, 2010). Compared to other information criteria, the WAIC most closely follows the Bayesian methods, as it takes into account the whole posterior distribution, as opposed to just the point estimates (Gelman, Hwang, & Vehtari, 2013). As is usual with information criteria, a lower WAIC indicates a better predictive model accuracy.

4.3 Empirical application: differential treatment effects on panic attack frequencies

In this chapter, we re-analyze data described in Van Apeldoorn, Van Hout, Timmerman, Mersch, and Den Boer (2013). In a randomized clinical trial involving panic disorder patients, the differential rate of improvement across three types of therapy was examined. The three treatments involved were Cognitive Behavioural Therapy (CBT), Serotonin Selective Re-uptake Inhibitors (SSRI) and a combination of both (BOTH). As a measure that reflects symptom severity and that is feasible to be measured intensively, the frequency of panic attacks is used. Each patient recorded the number of panic attacks experienced during the previous week. They did so on a weekly basis for the period of one year in which treatment was delivered, including medication taper. Here, we only consider the patients who completed the therapy according to protocol and who provided scores on at least 10 consecutive time points. This results in 73 out of 178 patients, of which $n = 28$ with CBT, $n = 22$ with SSRI and $n = 23$ with BOTH.

4.3.1 Model specification

The core research questions of the study were how the frequencies of panic attacks develop during and after intervention and to what extent this differs across the three treatments, and possibly relates to individual characteristics. The initial state is captured via an individual intercept, and the course across time via an individual slope. Following Van Apeldoorn et al. (2013), we included treatment (CBT, SSRI, BOTH) and level of agoraphobia (no/mild versus moderate/severe) as individual predictors. Earlier research has shown that a panic disorder, when remaining untreated, may become more severe and change in nature (Altamura, Santini, Salvadori, & Mundo, 2005; Federici & Tommasini, 1992). As a result, the intensity of symptoms may increase over time when no treatment is received. To test this, we include the duration of symptoms pre-treatment as a predictor for the panic attack frequency at the start of the treatment.

In empirical data, the noise terms are often not independent for subsequent time points (Goldstein et al., 1994). To assess the presence of autocorrelated noise in our model, we include an moving average mechanism in the noise of the system equation (Box & Jenkins, 1976).

To assess the importance of each of these elements, we will compare three models. Model 1 will include only the predictors for the slope, being treatment and presence of agoraphobia. In Model 2, we will add the moving average mechanism to the system equation. In Model 3, we will add the duration of symptoms pre-treatment as a predictor for the intercept.

4.3.2 Model design

We elicit three different models for the data set.

Link function and observation equation To map our observed count data $y_{t,n}$ to the continuous latent score $y^*_{t,n}$, we consider a Poisson distribution and use an exponential link function:

$$y_{t,n} \sim \text{Poisson}(g^{-1}(y^*_{t,n})), \qquad g(y^*_{t,n}) = \exp(y^*_{t,n}).$$

Earlier studies found an exponential decay in symptoms over time (e.g., Bandelow et al., 2004; Ross, Klein, & Uhlenhuth, 2010; Toni, Perugi, Frare, Mata, & Akiskal, 2004), which is also seen in the observed scores in our data set. Since our link function uses an exponential transformation, we can use a linear function of t on $y^*_{t,n}$ to model the exponential decay in $y_{t,n}$.

As variation is implicitly included in the link function, we can have an observation equation without white noise:

$$y^*_{t,n} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \mu_n \\ \delta_{t,n} \\ \beta_n \end{bmatrix}.$$

where μ_n is the intercept of individual n, roughly corresponds to $y^*_{1,n}$, $\delta_{t,n}$ is difference between the intercept and $y^*_{t,n}$, and β_n is the slope. The link function and observation equation are taken equal for all three models.

System equation For Model 1, the system equation depicts a linear growth model, with slope β_n and intercept μ_n:

$$\begin{bmatrix} \mu_n \\ \delta_{t,n} \\ \beta_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu_n \\ \delta_{t-1,n} \\ \beta_n \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_{t,n} \\ 0 \end{bmatrix}, \qquad \eta_{t,n} \sim N(0, \sigma_\eta), \tag{4.2}$$

where $\eta_{t,n}$ is the innovation error at time t for individual n with standard deviation σ_η.

For Models 2 and 3, we add a moving average mechanism to the system equation, Equation (4.2):

$$\begin{bmatrix} \mu_n \\ \delta_{t,n} \\ \beta_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu_n \\ \delta_{t-1,n} \\ \beta_n \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \eta_{t-1,n} & \eta_{t,n} \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} \psi_{1,n} \\ 1 \end{bmatrix}, \qquad \eta_{t,n} \sim N(0, \sigma_\eta),$$

where $\psi_{1,n}$ is the lag 1 moving average parameter of individual n.

Predictor variables Including the combination of the treatment group and presence of agoraphobia as predictor for the individual slope yields the following expression for β_n in all three models:

$$\beta_n = \beta_0 + \begin{bmatrix}\beta_{1,1} & \dots & \beta_{1,5}\end{bmatrix} \times \begin{bmatrix} d_{s,m} \\ d_{b,m} \\ d_{c,a} \\ d_{s,a} \\ d_{b,a} \end{bmatrix} + \xi_{\beta,n}, \qquad \xi_{\beta,n} \sim N(0, \sigma_\beta), \tag{4.3}$$

where β_n is the slope for individual n in group $g_{tr,ag}$ with tr (tr: CBT=c, SSRI=s, BOTH=b) being treatment and ag (ag: m = none or mild agoraphobia, a = moderate or severe agoraphobia) being presence of agoraphobia. Furthermore, β_0 is the estimated slope-coefficient for an individual in $g_{c,m}$, $\beta_{1,\cdot}$ is the added effect of the group depicted in the accompanying dummy variable $d_{tr=\cdot,ag=\cdot}$, where $d_{tr=\cdot,ag=\cdot}$ is the 0/1-coded dummy variable indicating whether the patient is in group $g_{tr,ag}$, $\xi_{\beta,n}$ is the individual deviation in β_n and σ_β is the standard deviation of $\xi_{\beta,n}$.

In Models 1 and 2, the intercept μ_n is presumed to follow a normal distribution:

$$\mu_n = \mu_0 + \xi_{\mu,n}, \qquad \xi_{\mu,n} \sim N(0, \sigma_\mu).$$

where μ_0 is the mean estimated intercept, $\xi_{\mu,n}$ is the individual deviation in the intercept and σ_μ is the standard deviation of $\xi_{\mu,n}$.

In Model 3, we include the duration of symptoms pre-treatment x_n as predictor variable for μ_n:

$$\mu_n = \mu_0 + \mu_1 \times x_n + \xi_{\mu,n}, \qquad \xi_{\mu,n} \sim N(0, \sigma_\mu), \tag{4.4}$$

where μ_0 is the estimated intercept when x_n is zero and μ_1 is the effect of x_n on μ_n.

Priors The priors of the model are aimed to be weak-informative. The observed number of panic attacks in this data set ranges from 0 to 12 attacks per week, giving an expected range of $y^*_{t,n}$ between $-\infty$ and 2.5. For the standard deviations (σ_η, σ_β and σ_μ) we set the prior at N(0.5,5) with lower bound zero, creating a prior similar to a half-Cauchy prior (Stan Development Team, 2016). Taking into account the expected range of the latent score, we believe that a wider prior would only delay calculations without improving the model estimates. For β_0, β_1, μ_0 and μ_1 we set the prior at N(0,5), since these parameters are expected to have

relatively small values, not exceeding an absolute value of 2. Finally, for ψ_n we used Berger's symmetrized reference prior (Berger & Yang, 1994), as this prior has shown to be a better prior for autocorrelation parameters than an uniform $[-1, 1]$ prior (Krone et al., 2016a). This prior does not require hyperparameters.

4.4 Results

For each model, we discuss the convergence and the posterior estimates. Further, we interpret the estimated model parameters and compare the models, considering the noise standard deviations and the WAIC. Finally, we show the results for a post-hoc comparison of males and females with regard to the intercept.

4.4.1 Convergence

Each model is estimated using four MCMC chains of 20,000 iterations, of which half the iterations were used for burn-in. Before the results can be interpreted, we must check whether the chains converge. In all three models, all model parameters for all subjects ($\beta_\cdot, \mu_\cdot, \psi_n$ and $\sigma_\cdot$) showed good to very good $\hat{R}$-values (i.e., all below 1.01, except for σ_ε with $\hat{R} = 1.011$ for Models 1 and 2, and 1.012 for Model 3). The traceplots were all proper, as was already expected from the $\hat{R}$-values, that is, fat caterpillars similar to those in Figure 4.1.

4.4.2 Parameters

The slope β_n

The boxplots for the slopes β_n ($n = 1, ..., N = 73$) as estimated for each of the three models, are shown in pane (A) of Figure 4.2. For all three models, the mean β_n in the sample is -0.14, with range -0.23 to -0.05 for Model 1, and range -0.24 to -0.02 for Models 2 and 3. In Table 4.1 the posterior means (and standard deviations) of the slope using Equation (4.3) are presented. In Table 4.2 the estimated mean slopes for each treatment/agoraphobica combination are presented. The differential effects are highly similar over the models, showing only small differences in estimated size of effect. The steepest mean slopes are estimated for $g_{s,m}$ and $g_{b,a}$, followed by $g_{s,a}$ and $g_{b,m}$. The shallowest mean slopes are estimated for $g_{c,m}$ and $g_{c,a}$, for which the mean slope fall outside, or are on the edge of, the 95% credible interval (CrI) of the mean estimated slopes for the other groups.

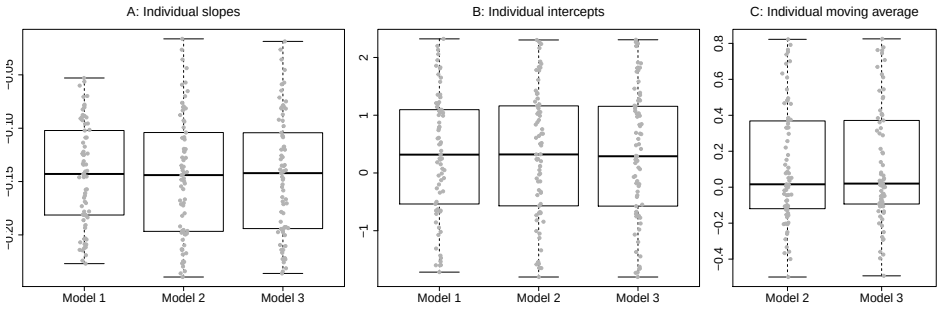

Figure 4.2: Boxplots of individual parameter values (shown by the dots) within the sample per model for (A) the slope β_n, (B) the intercept μ_n and (C) the moving average parameter ψ_n

	Model 1		Model 2		Model 3	
	mean (sd)	95% CrI	mean (sd)	95% CrI	mean (sd)	95% CrI
μ_0	0.32 (0.17)	-0.02; 0.65	0.31 (0.17)	-0.04;0.63	0.10 (0.23)	-0.36; 0.54
μ_1					0.03 (0.02)	-0.01; 0.07
σ_μ	1.24 (0.16)	0.97; 1.57	1.27 (0.15)	1.00;1.60	1.27 (0.15)	1.00; 1.60
β_0	-0.10 (0.02)	-0.15; -0.06	-0.09 (0.03)	-0.15;-0.05	-0.09 (0.02)	-0.14; -0.05
$\beta_{1,1}$	-0.09 (0.04)	-0.17; -0.01	-0.11 (0.04)	-0.20;-0.02	-0.11 (0.04)	-0.20; -0.02
$\beta_{1,2}$	-0.04 (0.03)	-0.11; 0.03	-0.06 (0.04)	-0.13;0.02	-0.06 (0.04)	-0.13; 0.02
$\beta_{1,3}$	0.01 (0.03)	-0.05; 0.08	0.01 (0.03)	-0.05;0.08	0.01 (0.03)	-0.05; 0.08
$\beta_{1,4}$	-0.06 (0.04)	-0.14; 0.01	-0.08 (0.04)	-0.17;-0.00	-0.08 (0.04)	-0.16; -0.00
$\beta_{1,5}$	-0.09 (0.04)	-0.17; -0.02	-0.10 (0.04)	-0.18;-0.03	-0.10 (0.04)	-0.18; -0.03
σ_β	0.05 (0.02)	0.01; 0.08	0.06 (0.02)	0.02;0.09	0.06 (0.02)	0.02; 0.09
σ_η	0.33 (0.03)	0.27; 0.38	0.26 (0.03)	0.21;0.32	0.26 (0.03)	0.21; 0.31

Table 4.1: Posterior mean (and standard deviation) with 95% credible interval (CrI) estimates for the model parameters of Models 1, 2 and 3.

The Intercept

In pane (B) of Figure 4.2 the boxplots of the estimated μ_n per model can be seen. For Models 1 and 2, the mean estimated μ_n is 0.32, with a range of -1.7 to 2.3 and of -1.8 to 2.3, respectively. For Model 3, the mean estimated μ_n is 0.30 with a range of -1.8 to 2.3. The posterior means (and standard deviations) for μ_0 and σ_μ for all three models can be seen in Table 4.1.

In Model 3, μ_n is estimated using Equation (4.4). As can be seen in Table 4.1, the posterior mean of μ_1 is smaller than the accompanying standard deviation and the 95% CrI for the posterior mean of μ_1 includes zero, suggesting that the duration of symptoms pre-treatment is not or only weakly related to the frequency of panic attacks at the start of therapy.

	Model 1	Model 2	Model 3
$g_{c,m}$	-0.10 (0.02)	-0.09 (0.03)	-0.09 (0.02)
$g_{s,m}$	-0.20 (0.05)	-0.21 (0.05)	-0.20 (0.05)
$g_{b,m}$	-0.14 (0.04)	-0.15 (0.05)	-0.15 (0.04)
$g_{c,a}$	-0.09 (0.04)	-0.08 (0.04)	-0.08 (0.04)
$g_{s,a}$	-0.16 (0.05)	-0.18 (0.05)	-0.17 (0.05)
$g_{b,a}$	-0.20 (0.05)	-0.20 (0.05)	-0.20 (0.05)

Table 4.2: Mean slope per group for Models 1, 2 and 3, calculated by adding β_0, which is the mean slope of $g_{c,m}$, to the $\beta_{1,\cdot}$ for the relevant condition.

The Moving Average

The moving average parameter is included in Models 2 and 3. Panel (C) of Figure 4.2 shows the distribution of the posterior means of ψ_n within the sample. For both models, the ψ_n has a mean of 0.12 and median of 0.02 in the sample, with a range of -0.50 to 0.82 for Model 2 and a range of -0.49 to 0.83 for Model 3. The estimated standard deviation of the individual ψ_n ranges from 0.23 to 0.71 for both models. Out of the 73 individuals, for both models only five 95% CrIs did not include zero.

4.5 Comparison of models

We compare the three models to see whether the model is improved by including the duration of symptoms pre-treatment as an external variable, and the moving average. When considering σ_μ, σ_β and σ_η, the differences between the models are small. For σ_μ and σ_β, the posterior means are slightly smaller for Model 1 than for Models 2 and 3. The σ_η of Model 1 falls above the 95% CrI of the posterior estimate of σ_η for Model 2 and 3, for which σ_η is similar in size. This implies that Model 1 shows slightly less noise when estimating μ_n and β_n, but more noise for the estimation of $\delta_{t,n}$ and thus for the estimated latent score.

Second, we compare the models using the Watanabe-Aikake Information Criterion (Watanabe, 2010) with the functions as provide by the package 'loo' in R (Vehtari, Gelman, & Gabry, 2015). A lower WAIC indicates a better predictive model accuracy. The WAIC is 30,312 for Model 1, 27,823 for Model 2 and 27,616 for Model 3, thereby favouring Model 3.

Combining the error standard deviations and the likelihood estimates, we can infer that the effect of including the moving average term (ψ_n) on the model fit is stronger than the effect of including the duration of symptoms pre-treatment.

To give a visual reference of the resulting fit to the sample data, Figure 4.3 shows the mean observed score and the mean estimated score per condition across time for Model 3.

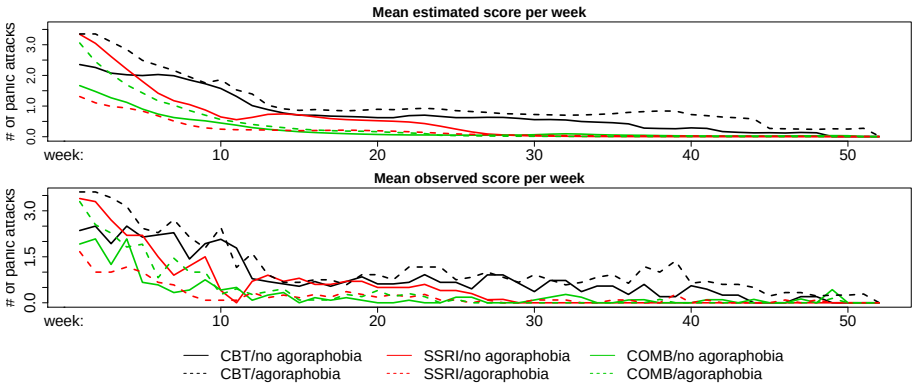

Figure 4.3: Mean observed score and mean estimated score using Model 3 per condition per week

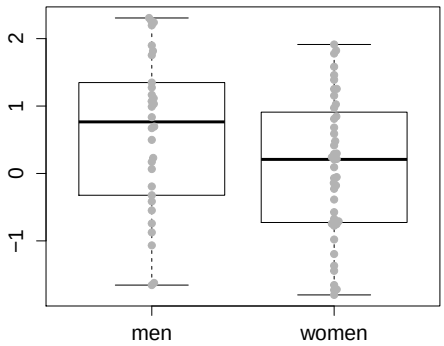

Figure 4.4: Boxplots of individual estimates (shown by the dots) for μ_n with Model 3 per gender

4.6 Post-hoc: intercept and gender

Earlier studies found that compared to women, men wait longer before asking for help and are less able in recognizing symptoms of mental stress (Khlat, Legleye, & Sermet, 2013; Swami, 2012). This may result in a larger number of symptoms experienced before treatment is sought. To see if this is reflected in our sample we compare the intercept, μ_n as found in Model 3, between men and women post-hoc.

As can be seen in Figure 4.4, the difference between the groups is small, and the variation is larger within each group than between groups. To test the difference, we use a Bayesian t-test (Morey & Rouder, 2011; Morey, Rouder, & Jamil, 2015), with H_0 : the true difference in mean equals 0, and H_a : the true difference in mean is unequal to zero. The prior of the effect size of the difference in mean is a Cauchy$(1/\sqrt{2})$ distribution. We found only anecdotal evidence in favor of H_a, with a Bayes Factor of 1.32. We conclude that men may have equal to slightly higher numbers of panic attacks than females at the start of the treatment.

4.7 Discussion

In this chapter, we explained how the BDM can handle relatively small data sets containing missing data points and dropouts, with a multilevel structure and non-normally distributed observed data. The BDM for modeling intensive longitudinal data is showcased by a re-analysis of data from Van Apeldoorn et al. (2013), consisting of the number of panic attacks experienced by 73 patients, measured per week. We used three models to examine the effects of three external variables and a moving average element. Model 1 included the treatment and presence of agoraphobia as predictors for the slope, in Model 2 a moving average parameter was added and in Model 3 the duration of symptoms pre-treatment was added as a predictor for the intercept. For the random effects of our predictor variables and our innovation noise distribution, we choose to use the normal distribution. If necessary, it is possible to use a different distribution, for example a Students-t distribution for smaller datasets.

In all three models the slope depends on the treatment an individual received and on the presence of agoraphobia. The CBT treatment shows a slower decrease in symptoms than the other two treatments, both for the individuals without and with agoraphobia. Furthermore, the SSRI treatment shows a stronger decrease for individuals without than for individuals with agoraphobia. This contrasts to the BOTH treatment, which shows a stronger decrease in symptoms for individuals with agoraphobia. All effects of agoraphobia are only trends. The results pertaining to the effect of the treatment and presence of agoraphobia in individuals on the slope are consistent with those found by Van Apeldoorn et al. (2013).

In Model 2, the moving average parameter was added. As a result, the standard

deviation of the error of the parameter estimation increased, but the standard error of the latent variable decreased. The WAIC decreased strongly from Model 1 to 2, which indicates that the moving average parameter is an important part of the model, even though the parameter is close to zero for a large number of individuals in the sample.

In Model 3, the duration of symptoms pre-treatment was included as a predictor for the intercept. Though the WAIC indicated that Model 3 fitted best, the 95% CrI of the estimated posterior mean for the predictor effect included zero. This indicates that it is uncertain whether the duration of symptoms has an effect on the intercept. A post-hoc test was conducted using Model 3 to compare the intercepts for men and women. The Bayesian t-test suggest anecdotal evidence for a difference in means.

An important question is how the BDM improves estimation compared to the frequentist multilevel model used by Van Apeldoorn et al. (2013). First, in contrast to the multilevel model, the BDM allows for adjustment to the estimation during the time series through the noise of the latent state vector. Opposed to the white noise as used in the multilevel model, the innovation noise is included in the current latent vector, which is used to estimate the latent vector at the next time point. As a result, the divergence from an expected score at a certain time point is used in the BDM to adjust estimations of later time points.This larger flexibility of the BDM results in a better fit of the observed data. Second, the BDM estimates the individual parameters, opposed to only estimating the distribution from which they are drawn. This allows for inspection of these individual parameters, for example for post-hoc analyses. Third, the combination of random coefficients (β_n, μ_n) and fixed coefficients (ψ_n) as used in Model 3, would not have been possible within the frequentist multilevel framework.

As an alternative to the BDM, the data could have been analyzed with a State Space Model (SSM), which is the frequentist counterpart to the BDM. The SSM is a highly versatile model for intensively measured, functionally related data, such as intensive longitudinal data (Durbin & Koopman, 2012; Petris et al., 2009; Pole, West, & Harrison, 1994). The SSM allows for modeling of non-normally distributed data (Durbin & Koopman, 2012) and time-dependent parameters, such as found in threshold models (Haan-Rietdijk, Gottman, Bergeman, & Hamaker, 2014; Hamaker & Grasman, 2012). However, to our knowledge no SSM has been developed yet that incorporates both missing data and non-normally distributed observed data simultaneously. The hierarchical Poisson SSM proposed by Terui et al. (2010) allows for non-normally distributed data. However, this estimation approach can be applied to complete data only, implying that scores should be available for all individuals at all time points.

The number of individuals in the data set used for our empirical study is small compared to the number of conditions; 73 individuals divided over six conditions.

By combining multiple individuals, the estimation of the distribution of a parameter improves (Krone et al., 2016d), but it is not yet clear how exactly this relates to each other in complicated models as the one discussed in this chapter. Another important property of the data is the length of the time series; earlier studies stated that a length of at least fifty time points is preferred for a simple single-subject moving average model (Box & Jenkins, 1976; Krone et al., 2016a). As our data set contained time series with a minimum of ten and a maximum of fifty-two data points, the model may at points be too complex for the data.

Models used in empirical studies often include a number of different parameters. For trend models as in this chapter, important parameters are the intercept, the slope parameter and eventually added autoregressive or moving average parameters, of which several parameters are often influenced by external variables. When the dynamics are the main interest, parameters such as the mean score, autoregressive and moving average coefficients, and for multivariate models the cross-lag coefficients, are of importance. The efficiency with which these parameters are estimated in such complex models, both with regard to individuals and to time points, is not yet thoroughly studied. Thus, the amount of data needed for reliable estimates in such complex models is an interesting topic for further studies.