Bayesian Hierarchical Modeling of the Temporal Dynamics of Subjective Well-Being: A 10 Year Longitudinal Analysis

Jeromy Anglim, Melissa K. Weinberg, Robert A. Cummins [1]

Abstract. This study demonstrates, for the first time, how Bayesian hierarchical modeling can be applied to yield novel insights into the long-term temporal dynamics of subjective well-being (SWB). Several models were proposed and examined using Bayesian methods. The models were assessed using a sample of Australian adults (*n* = 1,081) who provided annual SWB scores on between 5 and 10 occasions. The best fitting models involved a probit transformation, allowed error variance to vary across participants, and did not include a lag parameter. Including a random linear and quadratic effect resulted in only a small improvement over the intercept only model. Examination of individual-level fits suggested that most participants were stable with a small subset exhibiting patterns of systematic change.

Keywords. subjective well-being, homeostasis, longitudinal, Bayesian hierarchical models, set-point theory

1. Introduction

Researchers have long been interested in the long-term stability and change of subjective well-being (SWB). Test-retest correlations from longitudinal data (Schimmack & Oishi, 2005) and twin studies (Lykken & Tellegen, 1996), together with the generally small long-term effect of major life events, all attest to the stability of SWB over time. However, test-retest correlations do decline as test-retest intervals increase (Schimmack & Oishi, 2005), and more recent work suggests that some life events lead to long-term changes in SWB for some people. To explain these temporal dynamics, several theoretical models of SWB have been proposed (e.g., Brickman & Campbell, 1971; Cummins, 2015; Easterlin, 2003; Headey & Wearing, 1989). Underpinning the evidence for these theoretical models are various statistical

---

[1] Citation:
Anglim, J., Weinberg, M. K., & Cummins, R. A. (2015). Bayesian Hierarchical Modeling of the Temporal Dynamics of Subjective Well-Being: A 10 Year Longitudinal Analysis. *Journal of Research in Personality,* 59, 1-14.

Jeromy Anglim, Melissa K. Weinberg, and Robert A. Cummins, School of Psychology, Deakin University, Australia. Data and code for presented analyses are available at https://osf.io/qcfnk. Correspondence concerning this article should be addressed to Jeromy Anglim, School of Psychology, Deakin University, 221 Burwood Highway, Burwood, 3125 Victoria, Australia. Email: jeromy.anglim@deakin.edu.au

approaches that have been used to analyze longitudinal datasets (e.g., Charles, Reynolds, & Gatz, 2001; Easterlin, 2003; Ehrhardt, Saris, & Veenhoven, 2000; Headey & Wearing, 1989; Helliwell, 2003; Lucas & Donnellan, 2007; Mroczek & Spiro III, 2005; Orth, Trzesniewski, & Robins, 2010). In particular, various hierarchical modeling and latent variable approaches have provided insights into the nature of SWB dynamics.

While these statistical models have provided useful insights, they also have their limitations. In particular, they have tended to rely on standard distributional assumptions and used a limited set of model comparison tools. More recently, researchers in a wide range of fields, including psychology, have begun to explore the potential of the Bayesian approach to model estimation and comparison (e.g., Anglim & Wynton, 2015; Averell & Heathcote, 2011; Elliott, Gallo, Ten Have, Bogner, & Katz, 2005; Lee, 2008; Nikodijevic, Moulding, Anglim, Aardema, & Nedeljkovic, 2015). Software such as BUGS, Jags, and Stan have made flexible Bayesian model specification more accessible to applied quantitative researchers by reducing the need for the user to specify an algorithm for parameter estimation. Furthermore, the Bayesian approach offers a range of powerful model comparison tools which include model recovery, measures of fit with advanced penalties for model complexity, and checks on whether models recover theoretically important features of the data (Gelman et al., 2013). However, despite their increased accessibility, such models have not yet been applied to longitudinal SWB research.

Thus, the purpose of this paper is to apply the Bayesian approach in order to parsimoniously model the features of long-term change in SWB. We propose several alternative models and show how a Bayesian approach to estimation and model comparison provides novel insights into the temporal dynamics of SWB. We estimate models and apply this approach to 10 years of SWB data from a large representative sample of Australian adults.

1.1. Subjective Well-Being (SWB): An Overview

Subjective well-being (SWB) commonly refers to a broad range of emotional reactions and cognitive evaluations that represent an individual's assessment of their overall life quality (Diener, Suh, Lucas, & Smith, 1999). When measured either by a single global life satisfaction item or by a composite scale based on satisfaction with multiple domains of life (e.g., the Personal Wellbeing Index, International Wellbeing Group, 2013), several robust findings have particular relevance to the current investigation. First, most people report feeling positive about their lives most of the time (Cummins, 1998, 2003, 2013). Second, positive mood provides an explanation for this stability with the combination of happiness, contentment, and alertness accounting for up to 80% of SWB variance (Blore, Stokes, Mellor, Firth, & Cummins, 2011). Third, from the perspective of homeostatic theory, individual differences in this positive affect forms the basis of an affective set-point (Tomyn & Cummins, 2011), and when emotions create a level of SWB different from set-point, a homeostatic system is activated with responsibility for returning SWB to set-point (Cummins, Li, Wooden, & Stokes, 2014).

An essential feature of SWB that can be understood as a consequence of the above is that it tends to be fairly stable over time. Hartmann (1934) provided initial evidence of this, reporting a one-month test–retest correlation of .70 in self-reported general happiness among college students. By the 1970s it was clear that considerable levels of stability in SWB extend over several years (Andrews & Withey, 1976; Palmore & Kivett, 1977). A meta-analysis by Schimmack and Oishi (2005) obtained average test-retest correlations for multi-item scales at 1 year of around $r = .60$, and at 10 years of around $r = .35$, but estimates based on more than 5 years were based on small sample sizes. Supporting a partial genetic basis for this stability, Lykken and Tellegen (1996) found much larger SWB intraclass correlations for monozygotic twins ($r=.44$) than for dizygotic twins ($r=.08$). Finally, many major life events appear to have only a temporary effect on SWB (Headey & Wearing, 1989; Suh, Diener, & Fujita, 1996).

Adding to the understanding of these trends, several strands of evidence suggest that SWB measurement for a given individual is more than just sampling from a stationary distribution. Test-retest correlations do tend to decline somewhat over time and even over one year intervals such correlations are typically less than internal consistency measures of reliability. Furthermore, covariance models that seek to partial out trait and auto-regressive variance have estimated that auto-regressive factors explain almost as much variance as traits (Lucas & Donnellan, 2007). Additional auto-regressive variance may be explained by extreme life events, like approaching death (Gerstorf et al., 2008), marital transition (Lucas, Clark, Georgellis, & Diener, 2003), and acquiring a disability (Lucas, 2007). Finally, studies of overall age effects do suggest that small but meaningful changes in SWB occur over the life course (e.g. Mastekaasa & Moum, 1984).

Despite the demonstration of such small changes, it is the overall stability of SWB over time that has led researchers to propose various stabilizing mechanisms (Cummins, 1995; Cummins, Eckersley, Pallant, Van Vugt, & Misajon, 2003). For example, Brickman and Campbell (1971) proposed that people adjust expectations to changing circumstances while Headey and Wearing (1989; 1992) proposed that stable personality traits systematically influence the experience and perception of life events which, in turn, influences SWB. Finally, Cummins (2015) proposed that HPMood set-points are the key to SWB stability, with systematic change in SWB is caused by homeostatic failure, when an individual's resources are insufficient to effectively counter the level of experienced challenge. Such failure, however, is usually an acute event, with SWB normally recovering to the level of its set point.

1.2. Longitudinal Statistical Models of SWB

Researchers have applied a range of statistical models to study the long-term temporal dynamics of SWB (for a review, see Eid & Kutscher, 2014). Such models almost always included a random intercept and generally adopt either a latent growth curve (e.g., Helson, Jones, & Kwan, 2002; Orth et al., 2010) or a hierarchical modeling approach (e.g., Lucas & Donnellan, 2011; Mroczek & Spiro III, 2005). Stochastic change is typically modeled using a lag parameter, whereas systematic change is commonly modeled using random linear and quadratic effects, although discrete change and growth-mixture models have also been employed (Mancini, Bonanno, & Clark, 2011; Wang, 2007). In particular, trait-state-error models (Kenny & Zautra, 2001) include parameters representing stable and lag components, as well as a state component which includes both occasion specific variance and measurement error (for a review, see Cole, Martin, & Steiger, 2005).

In contrast to latent growth curve models, hierarchical models have the benefit of easily incorporating unequal numbers of observations per participants, as well as placing the emphasis on predicting the criterion variable. A range of other approaches include iterative procedures to explore set points (Cummins et al., 2014), models designed to capture changes in test-retest structure over time (Fraley & Roberts, 2005), and models of momentary measurement error and short to medium term response biases Ehrhardt et al. (2000).

Despite the popularity and insights gained from hierarchical and latent growth curve approaches, they both have several limitations. First, many such models are incorporated into software which makes assumptions that are both difficult to modify and inappropriate for SWB data. For example, individuals differ in within-person variability, but standard models assume that variability is constant over individuals. Second, the data generating process implied by such models is rarely evaluated in terms of whether it captures theoretically relevant features of longitudinal SWB data, as described earlier. Such features also relate to degree of change, distributions of individual scores, and distribution of person-level means. Third, models are only sometimes compared, which in turn raises a number of challenges related to evaluating model complexity. To overcome these limitations, a Bayesian data analytic approach provides a promising framework for refining longitudinal models of SWB.

1.3. Bayesian Hierarchical Modeling

Bayesian hierarchical methods are increasingly applied in psychology to model repeated measures data (e.g., Anglim & Wynton, 2015; Averell & Heathcote, 2011; Lee, 2008; Nikodijevic et al., 2015). Adoption of Bayesian methods has been aided by increased computational power, refinement of algorithms, accessible software (e.g.,WingBugs, JAGS, and Stan) and textbooks relevant to a general applied quantitative audience (e.g., Gelman et al., 2013; Gelman & Hill, 2007; Kruschke, 2010). While Elliott et al. (2005) performed a Bayesian analysis of short term mood data, we are not aware of any attempt to apply Bayesian hierarchical methods to the study of long-term temporal dynamics of SWB.

The Bayesian hierarchical approach incorporates all the advantages of standard hierarchical modeling, but also offers several additional benefits. First, it allows substantial flexibility in defining the probability model proposed to underlie the data generating process. For example, the distribution of residuals is not required to be constant or normal. Similarly, the distribution of person-level parameters is not required to be normal. Second, the Bayesian approach provides a useful set of model evaluation tools. In particular, posterior predictive checks provide a powerful way of defining theoretical properties of interest and assessing whether a candidate model adequately recovers this feature. This moves beyond simple measures of fit, and seeks to assess the model on a range of theoretically important features using posterior predictive checks (for an overview, see Gelman, Meng, & Stern, 1996; Kruschke, 2013). The checks involve (a) defining a set of features of the sample data that a model should capture; (b) simulating data from the model; and (c) evaluating the model based on whether it produces simulated data that capture the features of the sample data.

While general measures of fit are typically based on the global likelihood of the model, often with a penalty, posterior predictive checks allow more weight to be assigned to particular features when evaluating model performance. Such checks also can provide greater guidance regarding how the model could be improved. They also highlight the relative strengths of different models. For example, posterior predictive checks have been applied to models of SAT scores (Sinharay & Stern, 2003) and learning (Anglim & Wynton, 2015). In the light of this information, we now consider what features of longitudinal SWB data that models should aim to capture.

1.4. Features of the Temporal Dynamics of SWB

There are several important temporal dynamics of SWB that a comprehensive model should capture. First, there is the distribution of person-level means. While people normally have positive levels of SWB, there is a general negative skew. Second, there is the distribution of person-level standard deviations. That is, people differ in the extent to which they fluctuate around their set-point over time (Cummins et al., 2014) and thus individual SWB levels differ from occasion to occasion. This is a neglected feature of emotional dynamics in many statistical models.

Additional consideration needs to be given to the relationship between mean levels of SWB and the standard deviation. This particularly applies to the degree to which lower standard deviations are caused by scale range restriction, whereby people with means closer to scale end points show less variability. Third, there is within-person change in terms of what we label stochastic and systematic change. Stochastic change refers to random change from a previous state, and is typically represented by some form of lag parameter. Systematic change refers to long-term trends that are commonly represented using linear or quadratic models, although many functional forms are possible. Finally, there is the distribution of observation-level residuals. Such residuals are likely to be negatively skewed, showing that while people are generally positive, they may also experience chronic lower levels of SWB caused by homeostatic failure.

1.5. The Present Study

The primary aim of the present study is to demonstrate how the Bayesian approach can be used as a framework to refine and evaluate models of the long-term temporal dynamics of SWB. Illustrating the flexibility of the Bayesian approach, we propose several innovations to existing models of SWB. The primary innovations are to (a) allow within-person error variance to vary between people; (b) include both lag and polynomial trend effects in one model; and (c) include a probit transformation of SWB. The focus of analyses is on comparing the16 models that result by crossing four model features related to within-person error variance (fixed or random), lag effect (present or absent), polynomial effects (random linear, and quadratic excluded or included), and transformation (probit transformed or untransformed).

These models were applied to data from 10 waves of annually collected longitudinal SWB data. The sample size, number of waves, use of a multi-item scale to measure SWB, and use of a representative sample with a broad cross-section of ages, made the longitudinal dataset well-suited to evaluate the competing models. Traditional models have often incorporated polynomial trends or lag effects but almost never incorporated random error or formally evaluated implications of transformations. Thus, we used comparisons of model fits and posterior predictive checks to assess the degree to which random within-person error and probit transformations resulted in superior model fit and whether these modifications altered the need for lag or polynomial effects. We predicted that using a probit transformation and allowing within-person error variance to vary between people would substantially improve model fit. The need for lag or polynomial effects was more of an open question. Whether such parameters are required to parsimoniously represent the temporal dynamics of SWB will be a test-case for the idea of SWB set points.

2. Method

2.1. Participants and Procedure

Participants were drawn from the Australian Unity Longitudinal Wellbeing Study. This national survey has been used to assess the subjective well-being (SWB) of the Australian population since 2001 and by the end of 2013 a total of 30 surveys had been conducted, each of 2,000 people. Participants were initially recruited through telephone numbers, randomly selected from proportional sampling of the Australian population, based on 50 geographic areas. At the end of their telephone interview, participants were asked if they would be willing to participate in a longitudinal written survey. Those who agreed joined the panel, along with others who remained in the longitudinal study from the previous waves. Participants in the longitudinal study completed each wave of measurement at approximately annual intervals. For the present study, participants were retained if they provided at least five waves of data and their data passed several validity checks including consistent age and gender reporting across waves. The final sample involved 1,081 participants (57% female) who completed between 5 and 10 waves (M = 6.8, SD = 1.6) for a total of 7,317 observations. Age at first wave of testing ranged from 18.3 to 87.6 years (M= 58.1, SD = 12.9). Based on the Australian Bureau of Statistics in 2004 (midpoint of first waves), mean age of the adults over 18 in Australian population was 45.7 (SD = 17.9) (Australian Bureau of Statistics, 2014). Thus the sample was somewhat older than the Australian adult population and, in particular, had few participants under 30.

In general, retained participants with five or more observations reported slightly higher SWB with a mean wave SWB score of 7.50 out of 10 compared to 7.40 out of 10 for participants with four or fewer observations (d = 0.07). With regards to demographics, participants providing fewer observations than five observations were on average 3 years younger and were three percentage points more likely to be male.

2.2. Subjective Well-Being (SWB)

The Personal Wellbeing Index (PWI) (Cummins et al., 2003; International Wellbeing Group, 2013) is a well-validated 7-item measure of SWB. Each item is rated on an end-defined, 11-choice scale, anchored by 0 = "no satisfaction at all" to 10 = "completely satisfied". The seven items ask the participant to rate how satisfied they are with a particular domain of life: (1) standard of living, (2) personal health, (3) achieving in life, (4) personal relationships, (5) personal safety, (6) community-connectedness, and (7) future security. An overall measure of SWB is taken as the mean of these seven items. The PWI scale correlated highly with a single-item 0-10 measure of general life satisfaction (i.e., "How satisfied are you with your life as a whole?"); for example, the cross-sectional correlation for the first wave was $r = .78$. In addition, in one wave of the study, the commonly used 5-item Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) was administered; it also correlated highly with the PWI scale ($r = .75$, $n = 696$). Also, the correlation of single-item general life satisfaction with Diener et al's scale ($r = .771$) was similar to the correlation with the PWI scale ($r = .785$). Thus, overall the PWI scale correlates highly with commonly used life satisfaction measures, yet it differs in that it is scored as a unit weighted composite of domain satisfaction scores.

2.3. Data Analytic Approach

We estimated and compared several hierarchical models of the longitudinal SWB data using a Bayesian approach. Common hierarchical linear models used in the literature to date have involved random intercepts, random linear and quadratic effects, and lag effects. Our models include these along with several important extensions, reflecting the flexibility of the Bayesian approach. The person-level standard deviation was allowed to vary between people, lag effects were incorporated using a Bayesian approach, and the non-normal intercepts and residuals were examined by modeling a probit transformation of the SWB data. Using posterior predictive checks, we were able to assess whether the probit transformation replicated distributional features of the raw data after back transformation. To assist others in applying Bayesian methods and to encourage further model development, we make the raw data and code available on https://osf.io/qcfnk.

Before specifying the model, we briefly describe conventions for Bayesian model specification. For those unfamiliar with Bayesian model notation, Kruschke (2013) has an accessible introduction. A Bayesian model requires that the likelihood and the prior be specified. The likelihood is generally a potential data generating process with parameters, and the prior is a probability model of belief about what values the parameters will take prior to seeing the data. Thus, when defining a likelihood we say that the data is drawn from a distribution. For example, $y_{ij} \sim N(\mu_{ij}, \sigma_i^2)$ expresses the idea that the data y_{ij} is distributed (~) normally (N) with a parameters for the mean and variance. It is then common to model the mean function as a function of parameters and covariates (e.g., $\mu_{ij} = \beta_{1i} + \beta_{2i} x_{ij}$) as we would see in regression models. This notation is equivalent to the general regression notation $y_{ij} = \mu_{ij} + e_{ij}$ (or with covariates $y_{ij} = \beta_{1i} + \beta_{2i} x_{ij} + e_{ij}$), where $e_{ij} \sim N(0, \sigma_i^2)$, but is more flexible in that it covers the situation where the model is not neatly divided into a mean and an error component. Often, as is the case with hierarchical models, these parameters themselves are drawn from distributions. For example, in the model below, the person-level intercepts are also drawn from a normal distribution. This is analogous to traditional hierarchical models. However, the Bayesian approach also gives greater flexibility, so for example, distributions other than the normal distribution can be readily used to describe variation in person-level parameters, and distributions can be placed on other terms such as the person-level standard deviation. In this way, the Bayesian approach encourages the analyst to describe a theoretically plausible data generating process and modify it in flexible ways.

2.3.1. Model description

The likelihood part of the model of SWB was

$$
\begin{aligned}
&y_{ij} \sim N(\mu_{ij}, \sigma_i^2) \\
&\mu'_{ij} = \beta_{1i} + \beta_{2i} x_{ij} + \beta_{3i} x_{ij}^2 + \gamma_1 z_{ij} + \gamma_2 z_{ij}^2 + \gamma_3 z_{ij}^3 \\
&\mu_{ij} = \mu'_{ij} + \beta_5 (y_{i(j-1)} - \mu'_{i(j-1)}) \\
&\beta_i \sim \mathcal{N}_4(\mu_\beta, \Sigma) \\
&\sigma_i \sim \exp(\beta_{4i})
\end{aligned}
$$

where y_{ij} represents SWB (0 to 10 on the PWI) for participant $i = 1, ..., N$ at observation $j = 1, ..., n_i$. Observation number was person-mean centered, where $x_{ij} = \bar{x}_{i.} - x'_{ij}$, $x'_{ij} = 1, 2, ..., n_i$, and where $\bar{x}_{i.} = 1/n_i \sum_i^{n_i} i$ is the person-level mean of observation number. Thus, when combined with group-level age effects, intercept β_{1i} represents the predicted SWB for an individual at the middle of their supplied observations. Intercept β_{1i}, linear β_{2i}, quadratic β_{3i}, and variability β_4 parameters were all modeled as varying between people and drawn from a multivariate normal distribution with covariance Σ and a vector of means μ_β, where the linear and quadratic effects where constrained to be zero to avoid multicollinearity with the group-level age effects. Age z'_{ij} was centered at the rounded mean of all age observations (i.e., 62 years, $z_{ij} = 62 - z'_{ij}$). Group-level age effects were modeled with fixed linear γ_1, quadratic γ_2, and cubic γ_3 parameters. Individuals also had their own within-person standard deviation σ_i which was obtained by taking the exponential of β_{3i}. Given β_{3i} is draw from a normally distribution, taking the exponential of β_{3i} results in a log normal distribution which both respects the requirement of σ_i to be greater than zero and reflects expectations that the distribution would be positively skewed. Drawing the vector β_i from a multivariate normal distribution also allows the correlation between parameters to be formally modeled. In particular, we expected intercept β_{1i} and variability either in the form of β_{4i} or σ_i to be negatively correlated. The lag effect β_5 was held constant over individuals and represents the expected increase at the current time point as a multiple of the deviation of the previous time point $y_{i(j-1)}$ from that expected by the model $\mu'_{i(j-1)}$.

Sixteen variants of the above model were formed from crossing four binary model features. First, fixed and random error models were examined. In fixed error models, person-level error variance σ_i was held constant over individuals as is the case in almost all models of SWB in the literature, whereas in random error models, σ_i was allowed to vary over individuals as specified above. Second, models were examined with and without the random polynomial coefficients representing the linear β_{2i} and quadratic β_{3i} effects. Models with only a random intercept were labeled *intercept* and models with linear and quadratic effects of observation were labeled *polynomial*. The model with both linear and quadratic terms is able to capture a broad range of systematic trends. In order to keep number of models to a manageable level, we did not report a model that included a random linear term but not a random quadratic term. Third, models either did or did not include the lag effect β_5. Finally, raw and probit transformed data were modeled. Specifically, the raw data ranged from 0 to 10, and the probit transformation was $\phi((y_{ij}+0.5)/11)$, where ϕ is the standard normal quantile function. This function increases relative differences between scores at the extremes (i.e., towards 0 or 10) and reduces relative differences around the mid-point (i.e., 5 on a 0 to 10 scale). Thus, on the transformed scale, a value of 0 represents the arithmetic midpoint of the raw scale (i.e., 5). The probit transformation is a good choice for transforming psychological data where observed scores are constrained by a forced choice response scale to lie on an interval, but where the latent factor is assumed to be continuous, unconstrained, and approximately normally distributed.

2.3.2. Priors

In general, relatively uninformative priors were used. This allows the data to inform parameter estimates rather than prior expectations. Uniform priors were used for mean intercept μ_{β_1} (0 to 10 for untransformed; -2, to 2 for transformed), the mean of the variation coefficient μ_{β_1} (-4 to 1) and the lag effect β_5 (-0.6 to 0.6). The prior for the covariance matrix for the vector of coefficients β_i was an inverse-Wishart distribution with large variances, zero-valued covariances and 4 degrees of freedom. Priors for group level age effects were normal distributions with mean of zero and variances of 2/30, $2/30^2$, and $2/30^3$ for linear, quadratic, and cubic effects respectively. In order to include the first observation for each participant in the lag models, it was necessary to model a latent observation zero y_{i0} (for a discussion see Congdon, 2007). This was modeled as $y_{i0} \sim N(\mu'_{i0}, \sigma_i^2)$ where μ'_{i0} was calculated as implied above with $x_{i0} = x_{i1} - 1$ and $z_{i0} = z_{i1} - 1$.

2.3.3. Parameter estimation

Posterior parameter estimates were obtained using Markov Chain Monte Carlo (MCMC) methods using JAGS (Plummer, 2003). Trace-plots of MCMC samples were used to examine convergence and determine an appropriate burn-in period. Four chains with 5,000 burn-in and 25,000 sample iterations per chain were used for estimating posterior density estimates. Parameter estimation in a Bayesian MCMC context involves sampling parameter estimates from the posterior density estimate of the sample. The mean of that distribution provides a point estimate and quantiles provide a credible interval (e.g., .025 and .0975 quantiles provide a 95% credible interval). Bayesian credible intervals are analogous to confidence intervals used in frequentist statistics, except that a credible interval expresses belief in the parameter rather than appealing to long run frequency coverage of confidence intervals.

2.3.4. Measures of fit.

We also obtained deviance statistics to measure model fit using the above MCMC methods including mean deviance and Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). Deviance is defined as the -2 log likelihood of the data given the model and parameter values. Because of the uncertainty in the value of the parameters, there is also uncertainty in the deviance. Thus, in a Bayesian context, researchers typically look at the expected deviance integrating over the uncertainty of parameter values. Computationally, using MCMC, a deviance is calculated for every iteration based on the sampled parameters. The mean of this sample is the expected deviance. This Bayesian measure of deviance is analogous to that provided by traditional estimation methods such as maximum likelihood, but differs particularly in the way that it is an estimate that is averaged over the uncertainty of parameter values.

When comparing measures of fit for alternative models, it is important to penalize models that have greater flexibility. In the traditional approaches, researchers often use number of parameters and degrees of freedom for this purpose. However, several researchers have noted that this often fails to capture the actual differences in flexibility between competing models (Pitt & Myung, 2002). DIC (Spiegelhalter et al., 2002) provides a modified measure of deviance that incorporates a data driven penalty for model complexity (i.e., DIC = Mean Deviance + Penalty). When comparing DIC between models, rules of thumb for comparing AIC and related penalized measures of deviance are typically applied. For example, Spiegelhalter et al. (2002) suggests after ruling out computational error due to the finite MCMC sample, differences between models of around 3 to 7 or more indicate that the model with the larger DIC has considerably less support. Thus, given the large sample and potential for MCMC sampling error, we suggest in this paper to interpret differences in DICs between models of 20 or more as indicating a substantive difference in fit.

2.3.5. Posterior predictive checks

To assess the 16 models of interest, posterior predictive checks were performed (Gelman et al., 2013). These checks provide a flexible means for assessing the ability of a given model to capture theoretically relevant features of the data. It involves several steps. First, for a given model, parameters are sampled from the posterior density estimates. Second, data are simulated for each of these sets of parameters. Third, a set of relevant statistics are calculated for each sample. Fourth, the mean and spread of the statistics from the simulated data are compared to the statistics calculated on the actual data. A good model should create simulated data with statistics similar to statistics on the actual data. This approach can also be used to identify shortcomings with a given model and lead to proposals for model improvement. In particular, there is a choice between simulating both observation-level data and participant-level parameters, or, simply simulating observation-level data assuming posterior estimates of participant-level parameters. We chose the more stringent test of simulating participant-level parameters. This requires that the model of the distribution of participant-level parameters is reasonable.

We proposed 11 statistics that we felt a good model of the temporal dynamics of SWB should recover (see Table 4 and Table 5 of the results). They all involved summarizing person-level statistics of SWB scores (e.g., person-level means, SDs, skewness) with summary functions (e.g., mean, SD, skew). We calculated the mean, SD, and skewness of person-level means. Because person-level means are a rough approximation to set points (Cummins et al., 2014), a good model should capture this distribution. We calculated the mean and SD of person-level standard deviations to capture the typical degree to which SWB scores vary within-person and the degree to which people differ in this quantity.

The mean of the person-level lag 1 auto-correlation (i.e., the correlation between the previous time point and the subsequent time point) was calculated to assess the degree to which more proximal time points were more related. Auto-correlation is a feature of longitudinal data and can reflect a range of processes including incremental stochastic change as well as longer term systematic trends.

The mean of person-level skew was calculated. A value of zero indicates a symmetric distribution, negative skew indicates that the tail of the distribution trails off towards lower values, and positive skew indicates that the tail of the distribution trails off towards higher values.

To assess model recovery of outliers, an index of within-person outliers was created as the maximum absolute deviation of a person's SWB scores from their person-level mean. The mean and SD of this outlier index were calculated. Longer term change was evaluated using the SD of the linear regression coefficient obtained for a person when predicting SWB from age. Thus, the unit indicates the expected change in SWB for a one year increase in age. Finally, the correlation between person-level means and person-level standard deviations was obtained.

3. Results

3.1. Descriptive Statistics

Cronbach's alpha reliability was consistently high for the SWB scale. Calculated for each measurement occasion (i.e., 1 to 10) the mean alpha was .87 (range: .86 - .89). A factor analysis of data from the first measurement occasion showed clear support for a one factor solution with item loadings ranging from .60 to .79. The intraclass correlation (ICC1) was .72 indicating that 72% of variance in SWB was due to differences between people . When each participant's scores were averaged over all available times points, the mean was 7.50 (SD = 1.25, skew = -0.95) with 95% of participants having means on the positive side of the scale (i.e., > 5.0). Within-person standard deviations were moderate and highly skewed with a mean of 0.66 (SD = 0.39, skew = 1.80). The correlation between person-level means and person-level standard deviations was r = -.48 indicating that lower levels of SWB were associated with greater volatility in SWB. Further descriptive statistics are provided in Table 4, but will be discussed further under posterior predictive checks.

Figure 1 shows the distributions of all observations, person-level means, deviations from person-level means (unstandardized and standardized, i.e., division by person-level SD) for both raw and probit transformed SWB. It highlights how the probit transformation reduces the skew in the person-level means. It also highlights the kurtosis in unstandardized deviations. However, after division by person-level standard deviations, kurtosis is less prominent. This adds support for allowing person-level standard deviations to vary as kurtosis of unstandardized observations may result from a mixture of normal distributions with different SDs.

Insert Figure 1 about here

Estimates of stability using test-retest correlations were examined for different subsets of the data based on number of waves provided. This involved comparing the correlation between baseline and one year with baseline and final wave. Test-retest correlations were as follows: (a) Five or more waves (baseline with 1 year: r = .77; baseline with 4 years r = .68, n = 1,081; $\Delta r = .77 - .68 = .09$). (b) seven or more waves (baseline with 1 year: r = .76; baseline with 6 years r = .67, n = 559, $\Delta r = .09$). (c) nine or more waves (baseline with 1 year r = .76; baseline with 8 years r = .70, n = 189, $\Delta r = .07$) . In summary, while using different subsets of participants and time-points yields slightly different estimates, in all cases the reduction in test-retest correlations over many years was small. This suggests that there is some, but not a lot, of either systematic or stochastic change.

3.2. General Temporal Trends

Figure 2 shows the relationship between age and SWB using all observations and a model fit with 95% credible interval for group level data based on the probit transformed, polynomial, no lag, random error model. The plot highlights several features of well-being data. First, the model is negatively skewed with most observations in the positive range. Group-level model fits show that SWB generally fluctuated between about 0.2 or 0.3 either side of 7.5. In terms of age trends, SWB is lower for people aged in the 40s and early 50s, as has been commonly reported (e.g. Mastekaasa & Moum, 1984), followed by increases in the 60s and 70s and a decline in the early 80s. There are not enough observations to determine whether SWB is systematically higher in the 20s or lower in the late 80s.

Insert Figure 2 about here

Figure 3 shows the relationship between measurement occasion and SWB scores, at the individual-level, for a random sample of 15 participants who provided 10-years of data. Focusing only on the pattern of the data, the overall impression is one of underlying stability, with clear individual differences in the average levels of SWB. There are also a small number of cases that seem to incrementally change relative to their previous observations. This pattern is consistent with some systematic influence. Finally, there are a few instances where SWB dropped abruptly and was typically restored on the subsequent measurement occasion.

Insert Figure 3 about here

3.3. Model Estimates

The set of models were estimated based on (a) whether or not a probit transformation was applied; (b) whether a lag effect was estimated; (c) whether person-level error was allowed to vary over people; and (d) whether random linear and quadratic terms were included. Table 1 presents model fit statistics for all models. While deviance based measures of fit cannot be compared across untransformed and transformed variable models, the general pattern of fit statistics was similar. Allowing the within-person standard deviation to vary over people led to a large reduction in deviance, and despite the penalty associated with estimating additional parameters, the Deviance Information Criteria (DIC) was also much less. Inclusion of a lag effect increased the deviance and DIC. Adding random linear and quadratic terms led to a massive reduction in the deviance, but only a small reduction in

DIC and this reduction in DIC appeared somewhat smaller for the transformed data.

Thus, based on DIC the preferred model had random error, no lag effect, and a random polynomial effect. In general, including both the random linear and quadratic effects gives the model a great deal more flexibility. The improvement in fit from including the random polynomial effect was less for random error models than for fixed error models, and for transformed rather than untransformed data. It may be that this flexibility is needed more for the untransformed and fixed error data in order to recover from other problems related to outliers and skewness. Furthermore, the improvement in DIC from including the polynomial effect was relatively small compared to whether random error or the lag effect was included.

Insert Table 1 about here

Parameter estimates for the random error, no lag, random intercept and random polynomial models are presented in Table 2 for the raw scale and Table 3 for the probit transformed scale. For all models the mean within-person standard deviation (mean of σ_i) is a little over half the standard deviation of person-intercepts $\sigma_{\beta_{\text{intercept}}}$ highlighting that between person variability in SWB was much greater than within-person variability. The ratio of the standard deviation of σ_i to the mean of σ_i captures how much individuals differed in their within-person standard deviations. In particular, it was slightly larger for the untransformed model 0.49 (0.34 / 0.69) than for the transformed model 0.40 (0.08/ 0.20). The negative correlation between intercept and SD between-person coefficients was also smaller for the probit transformed variable.

Insert Table 2 about here

The group-level age effects were centered on the rounded mean age of all observations (62 years). Thus, the mean of the intercept ($\mu_{\beta_{\text{intercept}}}$) can be interpreted as the expected value at that age. Linear and quadratic terms were small and significant in both transformed and untransformed models, while the cubic term was only significant in the untransformed model . The group-level trend line is summarized in Figure 2.

Insert Table 3 about here

As mentioned above, mean deviance and DIC was not reduced by the inclusion of the lag parameter. Given the emphasis of existing trait-state-error models in the literature on an auto-regressive term, we further investigated this result. Across the eight models estimates of the lag effect ranged from .03 to .24. In general random error models had a slightly larger lag effect (.03 to .05 larger than fixed error models), and models that included the random polynomial effect had a much smaller lag effect (.10 to .17 smaller than intercept models). For the random error probit intercept model, when a lag effect was included the estimate of the lag effect was .24, 95% CI [.20, .28]. Given that it was a significant lag effect, it was surprising that mean deviance was greater when a lag effect was added to the model. To further investigate this result, we temporarily fixed the lag effect to be the sample estimate (.24), but the model without the lag effect still had a much lower mean deviance. Examination of model fits showed how the small changes in predictions implied by the lag effect did not visually seem to capture the model predictions. One plausible explanation of this result is that our approach does not treat the time before the first observation as known. While we could have discarded the first observation as outcome data and used it as the basis for estimating the lag effect, such a lag effect would be confounded by the fact that this first time point would contribute substantially to the reliability of estimation of the participant-level mean and thus should be positive purely because of this artifact.

As a comparison, we also compared parameter estimates using the above Bayesian approach with standard frequentist multilevel modeling approaches using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2013) in general and the nlme package in R (Pinheiro, Bates, DebRoy, & Sarkar, 2013) to examine the lag model. For the random intercept models without lag effects the parameter estimates the frequentist estimates were almost identical to the corresponding Bayesian estimates. Lag models and linear and quadratic models generally required some re-specification to be implemented but parameter estimates were similar. The models with random standard deviations could not readily be implemented.

3.4. Individual-Level Model Predictions and Credible Intervals

Figure 3 shows the model fits and 95% credible intervals for the two best performing models, i.e., the probit transformed, no lag, random error models with either (a) a random intercept or (b) random polynomial. The credible intervals show how these two models capture individual differences in within-person variation. They also capture the skewness of that variation, as person-level means deviate from the scale mid-point.

Comparing the model fits for the two models is particularly useful as the random intercept model embodies a strict form of set-point theory whereas the random polynomial model allows for systematic trends over multiple years. It is worth noting that the random intercept model does allow for small changes over time due to group-level age related changes. Overall, there are several cases where the polynomial model appears to provide improved fit to the data. However, the polynomial model also has substantially greater flexibility which may result in curves capturing noise. Table 1 showed how the polynomial model resulted in a massive reduction in the model deviance, but only a small reduction in the parsimony adjusted measure of fit, DIC. The model fits in Figure 3 appear consistent with this interpretation whereby polynomial curves are fitting a mixture of some systematic change and some fitting to noise in the data.

3.5. Posterior Predictive Checks

To further evaluate the proposed models we ran posterior predictive checks. Interpreting posterior predictive checks involves first interpreting the sample statistics of interest when applied to the sample data. We then examined the degree to which the models are able to generate simulated data with statistics consistent with the sample data.

3.5.1. Statistics on sample data.

The sample data (see the "dataset" column of Table 4) shows that the typical within-person standard deviation (0.66) was around half the standard deviation of the means (1.25). It also shows that the SD of within person SDs (0.39) is fairly large relative to the mean of the within-person SDs (0.66), thereby suggesting the importance of including a parameter that captures this variability. The mean of means of 7.50 combined with the SD of within-person means (1.25) is consistent with previous reports that 95% of the data from such samples lie above the scale arithmetic average of 5.0 due to the combined influence of SWB set points (Cummins et al., 2014) and homeostasis (Cummins, 2015).

Insert Table 4 about here

The absence of a positive lag 1 autocorrelation is somewhat surprising (mean lag 1 autocorrelation of -.06). Several theories suggest that SWB stability involves restoring values to a base-line or set point. Additionally, there is much evidence that life events can temporarily alter well-being, and that, if the level of challenge is strong and maintained, the reduced level of SWB can become chronic. Where such effects are operating, a positive lag1 autocorrelation would be expected. This is not evident in these data.

Insert Table 5 about here

Negative skew was present in both between and within-person-levels. However, the skew in between-person means (−0.94) was much larger than the mean skew of within-person data (−0.17). Thus, while people occasionally experience SWB outside their set point-range, they are more likely to experience a chronic reduction rather than a chronic elevation. Interpreting the outlier statistic is a little more challenging, but it does indicate that average maximum deviation from a person's mean is 1.08 units, which is a little under double the mean of SDs. It may also be that the SD of the outliers is a little larger than one might normally expect based on normal distribution assumptions. This reflects the non-normality of the actual data distribution, the fact that few people have substantial outliers, and is in conformity with homeostasis theory.

When a linear regression was fit to individual SWB trajectories, the mean linear change was close to zero, but at the individual-level some values increased and some decreased. However, a certain amount of variability in individual-level linear change is due to error in estimating such a coefficient for each participant on only 5 to 10 data points.

The correlation between the within-person means and within-person SDs was -.48. This shows that that greater variability is associated with lower means. While this is partially due to scale constraints, this is only part of the reason. Specifically, scale constraints dictate that the maximum possible standard deviation for person *i* on a 0 to 10 scale is $\sqrt{\bar{y}_{i.}(10-\bar{y}_{i.})}$ where $\bar{y}_{i.}$ is the person-level mean for person *i*. To inspect the degree to which the correlation between means and SDs was driven by scale constraints, we compared the sample correlation between person-level means and (a) raw person-level SDs ($r = -.48$) or (b) person-level SDs divided by the maximum possible SD given the person's scale mean ($r = -.25$). Thus, about half of the correlation was explained by scale constraints. The remainder is consistent with prediction based on homeostasis theory.

3.5.2. Model recovery of statistics.

Results of the posterior predictive checks for candidate models are presented in Table 4 for untransformed data and Table 5 for transformed data. The values for the models are the mean statistic for the posterior samples, and statistics are bolded where they appear to capture the sample statistic or are notably better than other models at capturing the sample statistic. This bolding is purely a device to facilitate interpretation and a more sophisticated interpretation should emphasize the proximity of the statistic of the model simulated data to the sample dataset.

Several major points can be made. First, models where standard deviations were allowed to vary were better able to capture within-person standard deviations, outliers, and the correlation between the mean and standard deviations. Interestingly the correlation between person-level means and person-level standard deviations was captured with a fixed standard deviation when the probit transformation was applied. Second, the autocorrelation and SD of linear change statistics provide indices of the within-person change. Without a lag or polynomial parameter, the lag 1 autocorrelation and the SD of linear change was slightly underestimated. Introduction of either the lag effect or the polynomial terms seemed sufficient to capture these statistics. Third, transformed data were much better at capturing the skewness in the data. That said, while the untransformed model implies that there is no skew, the probit transformed model captured about two thirds of the skew in the sample data.

Thus, overall, the random error transformed models with lag or polynomial effects performed best in the posterior predictive checks as determined by greater recovery of sample statistics. The transformed intercept-only random error model preferred by DIC also performed well, but appeared to fail to capture small amounts of more systematic within-person change. Based on principles of parsimony, the posterior predictive checks would favor the lag model over the polynomial because the lag model involves a single fixed parameter, whereas the polynomial model is estimating two additional parameters per participant.

4. Discussion

The set of models present a picture of the temporal dynamics of SWB over time frames ranging from 5 to 10 years. The core findings are that (a) the underlying level of SWB is quite stable over time and much of the variability from year to year is not systematic; (b) people vary in how much they vary over time, (c) people with lower person-level means tend to have greater variability, (d) the distribution of person-level means and residuals is negatively skewed. Importantly, the inclusion of random error and a probit transformation reduced the need for lag or polynomial trend components. Overall, without denying the existence of some systematic change, the results in the present study, albeit grounded in its particular sample and measure of SWB, provide somewhat greater support for the notion of set points and SWB homeostasis than would be implied by some of the existing SWB literature.

4.1. Features of Temporal Dynamics of SWB

4.1.1. Stability

Correlations of SWB over time were generally larger and exhibited smaller reductions than some previous studies (Lucas & Donnellan, 2012; Schimmack & Oishi, 2005). One likely explanation for the larger absolute correlations, but perhaps not the reduced reduction in correlations over time, is that some previous longitudinal studies have relied on single item measures and sometimes used fewer response options than the 0 to 10 scale used in the present study. In contrast, the PWI scale used in the present study had seven items which provide more reliable measurement. The broad range of domains measured in PWI may also result in domain compensation (Best, Cummins, & Lo, 2000). As a follow-up analysis, we compared test-retest correlations between our 7-item PWI and a single item global life satisfaction (GLS) measure. Using data with at least seven waves, correlations between baseline and one year were $r = .63$ for GSL and $r = .76$ for PWI, and between baseline and six years were $r = .52$ for GLS and $r = .67$ for PWI. Thus, consistent with principles of attenuation due to imperfect reliability, the test-retest correlations were attenuated more for GLS, but the pattern of only modest declines in correlations over time was similar for both measures.

Another broad explanation for differences is the study design and sample. The sample comprised a mid- to older-adult population who might be expected to have more established life circumstances. For example, Lucas and Donnellan (2011) found that four year personality test-retest correlations exhibited an inverted-u shape that tended to be lower in the 20s peaking in the middle age (40s to 60s) and reducing in the mid-70s and 80s. Consistent with this pattern follow-up analysis indicated that the intraclass correlation (i.e., an index of stability) was not smaller for younger participants (mostly starting late 20s and early 30s), and there was a slight reduction in stability for participants over the age of 75 (ICC = .65 compared to whole sample ICC = .72). The Australian sample may also be more stable given that it is a developed, literate, democratic, politically-stable country with reasonable levels of social security, potentially providing a relatively stable environment where dispositional tendencies may be more influential. All of these points highlight the need to extend the modeling in this paper to other longitudinal datasets in order to further clarify the relative impact of modeling approach versus sample and study design on understanding stability and change in SWB.

4.1.2. Aggregate age trends.

Consistent with earlier studies (e.g. Mastekaasa & Moum, 1984) the overall age trend suggests that SWB reaches a low point in the 40s gradually increases into the 50s and 60s where it may begin to plateau and possibly declines again towards the late 70s and 80s. Overall, the size of these effects were small to moderate, with group-level changes varying about .2 or .3 from the mean of around 7.5 over the ages of examination (around ages 30 to 85). While the dataset does not allow for an assessment of whether such effects are age or cohort based, several reasonable explanations could be offered for the age related trends. Such challenges include stresses associated with work, parenting, and marital relationships.

4.1.3. Within-person variability

The results show that people meaningfully differ in how much their SWB varies from year to year. Although such change has been modeled in the experience sampling literature (e.g., Kuppens, Oravecz, & Tuerlinckx, 2010; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007), this fundamental characteristic is missing from most hierarchical and covariance models of long term temporal dynamics. While the finding is partially explained by scale constraints, whereby participants who have means closer to scale end-points have less potential variability, even after this is controlled, a substantial correlation remains. Many theories of short term emotional experience include emotional volatility (e.g., Kuppens et al., 2010) as a component with bipolar disorder reflecting a clinical extreme. Tendencies for variability over short time frames may impact the longer-term evaluations implied by SWB measurement. There may also be a longer timeframe analogue to the tendency for short term volatility. Homeostasis theory also offers a possible explanation in that lower values are more likely to represent homeostatic failure and, so, increased volatility (Cummins et al., 2014). Finally, while models of SWB often downplay the role of the environment, it seems likely that people would differ in the degrees to which their environments are volatile over longer time frames. Insecure employment, financial difficulties, intermittent health problems, and unstable relationships are all potential candidates for increases volatility. Future research should seek to disentangle measurement, individual differences, and environmental predictors of within-person variability. The Bayesian framework provides a useful framework for performing such predictions.

4.1.4. Non-normality

The results also reinforced the point that SWB scores have a strong negative skew, rather than being normally distributed. A probit transformation is a principled transformation for mapping a constrained scale onto the real number line. It resulted in SWB data with substantially reduced skew. While we imagine that it is common knowledge amongst SWB researchers that SWB is negatively skewed, this assumption is rarely incorporated into formal models of temporal dynamics. Future work could further explore these transformations or could seek to adapt the Bayesian model to use a skewed or constrained distribution (e.g., a beta distribution).

Importantly, the use of the probit transformed SWB data and the use of a random within-person standard deviation had implications for model differences (see Table 1). Specifically, the improvement in model fit, as indexed by DIC from including the random linear and quadratic effects, varied substantially based on whether random error and probit transformations were applied. On untransformed data with fixed error variance, including the random polynomial terms resulted in a large improvement in DIC. In contrast, when the random polynomial was added to the probit transformed random error model, DIC became worse. It is likely that some of what appears to be meaningful linear and quadratic effects relates to capturing outliers and skew in the data. In particular, the probit transformation better reflects the actual distribution of residuals, and random within-person standard deviation allows participants with different levels of year-to-year volatility to not be constrained to a single variance model. Thus, accounting for these important features of residuals reduces outliers and reduces the potential for linear and quadratic effects to overfit the data.

4.1.5. Systematic and stochastic individual change

The results have implications for understanding the extent to which there is stochastic or systematic change in SWB. If we focus on the probit-transformed random error models, we observe support for both a high degree of stability and a small amount of meaningful change. As already discussed, the group-level data supports the existence of small systematic changes in SWB over the adult life course which, barring cohort effects or differential survival rates, should be the manifestation of average individual-level changes. Also, posterior predictive checks suggest that inclusion of either a fixed lag effect or random polynomial change is able to better capture the auto-correlation in the data. We also have data from a range of other studies that have found that certain events such as unemployment, divorce, and acquiring a disability may have medium to long-term effects on SWB (e.g., Gerstorf et al., 2008; Lucas, 2007; Lucas et al., 2003), as well as state-trait-error models which assign relatively large effects to auto-correlation rather than trait variance (e.g., Lucas & Donnellan, 2007).

In contrast, supporting the claim that much of SWB is random variation around a set point, we see that the probit transformed random error intercept model had a DIC only a little less than the polynomial model, and with the exception of a small inaccuracy in estimating auto-correlation, posterior predictive checks were very good for the intercept model. A synthesis may be that both people who are susceptible to large systematic change in SWB, and the events that give rise to such change, are uncommon. This would explain the small improvements in capturing auto-correlation and variance in the linear change by including a lag effect or random polynomial effects. Furthermore, large year-to-year variability seems to be retained even in the subset of cases that appear to exhibit more systematic change. The functional form and timing of these changes is also likely to be quite variable. Thus, in the absence of external time-varying covariates to explain these changes, participants with difficult-to-model temporal change may be better explained by the model estimating a larger within-person standard deviation. Greater sample sizes for participants for even longer than 10 years may be required for these longer term effects to manifest over and above the background variability.

Given the smaller reductions in test-retest correlations over time observed in the present study (cf. Lucas & Donnellan, 2012; Schimmack & Oishi, 2005), future work should seek to clarify whether the reduced importance of the lag effect in the present study is due to unique features of the sample or whether it generalizes to other longitudinal datasets. If it does generalize, then this would require a more rigorous statistical examination of the meaning of the lag effect in the present Bayesian model and in previously used models of SWB that incorporate auto-correlation (e.g., Eid & Kutscher, 2014; Kenny & Zautra, 2001; Lucas & Donnellan, 2012). For example, the model presented in this paper treats the lag effect as a latent variable, whereas some models treat the lag effect as an observed variable (for a discussion, see Congdon, 2007). Treating a lag effect as latent captures the inherent uncertainty in estimating what is the lag effect and may prevent effects artificially being attributed to lag effect when they really are an aspect of latent person-intercepts. A second issue relates to the general challenge of model comparison, of which there are extensive debates particularly around the suitability of mean deviance and DIC in multilevel models (Spiegelhalter, Best, Carlin, & Linde, 2014). It is partially for this reason that we also examined posterior predictive checks which, in contrast to DIC, suggested that lag effects provided a small but meaningful improvement. However, more generally, applied Bayesian model comparison is a rich area that remains an active topic of research (Gelman et al., 2013).

4.2. Implications for Theories of SWB

Overall, the results provide fairly strong support for the notion of individual set points with a few refinements. First, people differ in their variability around their set point. Second, this variability tends to be greater for people with a lower set point. Third, there are small amounts of systematic individual and group-level change in SWB, but this tends to be dwarfed by person-level means and year-to-year variation.

4.3. Implications for Modeling SWB

A question for researchers modeling longitudinal SWB data is whether they should adopt a Bayesian or the more common frequentist approach. Ardent Bayesians may argue that their approach is the only consistent way to perform inference. However, a more pragmatic perspective involves weighing up the pros and cons of both approaches. Major benefits of the Bayesian approach are (a) greater flexibility in specifying the likelihood (i.e., the implied data generating process), (b) a set of powerful model comparison tools, and (c) more direct parameter estimation. The major challenges of the Bayesian approach are (a) the technical challenge of learning how to specify and estimate models and (b) the time it can take to fit complex models to moderately large datasets.

Generally, when uninformative priors are used and the likelihood specified is the same, results will not differ much between a frequentist and a Bayesian approach. Thus, for example, when researchers want to run a basic multilevel model with a few different predictors, it will often be simpler to rely on standardized tools. However, if researchers are wanting to more precisely model the nuances of complex multilevel data, then the Bayesian approach provides a simpler way to model these features. Thus, the Bayesian approach is particularly useful in enabling a deeper examination of well-established research questions in psychology, particularly when such questions have a multilevel component.

In summary, the greater benefit of the Bayesian approach comes from being able to specify more flexible models. This flexibility is facilitated particularly by software such as BUGS, Jags, and Stan, which provide the researcher with a wide range of options for specifying a likelihood and prior, while at the same time shielding the researcher from the mathematical complexities of specifying an algorithm for sampling from the posterior.

4.4. Limitations, Future Research, and Conclusion

There are several opportunities to extend on the findings and modeling approach used in the present study. First, we did not seek to model observation-level SWB as a latent variable. Such an approach can be incorporated into the Bayesian approach and is a logical next step for Bayesian hierarchical models of SWB. Second, we focused on lag and polynomial structures of change. Future work could explore a broader range of temporal processes. In particular, different random-walk lag-related mechanisms may be appropriate and latent trajectory mixture models also have merits. Third, it would be interesting to examine latent trajectory models to assess the claim in homeostasis theory about a subset of the low data demonstrating greater volatility. Fourth, future work could explicitly compare state-trait-error models and covariance modeling approaches to the models presented. There may also be scope for further syntheses of the two approaches. For example, assumptions of state-trait-error models could be further incorporated into the Bayesian approach or random error, while probit transformations could be incorporated into state-trait-error models. Fifth, the models presented could readily be extended to incorporate between-person covariates such as personality or family upbringing. Such models could also be extended to not only allow prediction of person-level means, but also differences in within-person standard deviations.

In summary, the present study makes both methodological and substantive contributions to the study of the temporal dynamics of SWB. Methodologically, the Bayesian approach provides a flexible modeling framework. The benefits of this flexibility were illustrated using the proposed models that went beyond traditional techniques, by incorporating important assumptions regarding the non-normality of SWB and the variability of within-person variation. By providing raw data and source code we hope that the present study will motivate other researchers to apply Bayesian hierarchical methods to the study of the longitudinal dynamics of SWB.

5. References

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. New York: Plenium Press.

Anglim, J., & Wynton, S. K. (2015). Hierarchical Bayesian Models of Subtask Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 957-974. doi: 10.1037/xlm0000103

Australian Bureau of Statistics. (2014). 3101.0 Australian Demographic Statistics. Table 59. Estimated Resident Population By Single Year of Age, Australa.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology, 55*(1), 25-35. doi: 10.1016/j.jmp.2010.08.009

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(4)

Best, C. J., Cummins, R. A., & Lo, S. K. (2000). The quality of rural and metropolitan life. *Australian Journal of Psychology, 52*, 69-74.

Blore, J. D., Stokes, M. A., Mellor, D., Firth, L., & Cummins, R. A. (2011). Comparing multiple discrepancies theory to affective models of subjective wellbeing. *Social Indicators Research, 100*(1), 1-16. doi: 10.1007/s11205-010-9599-2

Brickman, P., & Campbell, D. T. (1971). *Hedonic relativism and planning the good society*. New York: Academic Press.

Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology, 80*(1), 136-151. doi: 10.1037/0022-3514.80.1.136

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

Congdon, P. (2007). *Bayesian statistical modelling*. West Sussex, England: John Wiley & Sons.

Cummins, R. A. (1995). On the trail of the gold standard for subjective well-being. *Social Indicators Research, 35*(2), 179-200. doi: 10.1007/BF01079026

Cummins, R. A. (1998). The second approximation to an international standard for life satisfaction. *Social Indicators Research, 43*(3), 307-334. doi: 10.1023/A:1006831107052

Cummins, R. A. (2003). Normative life satisfaction: Measurement issues and a homeostatic model. *Social Indicators Research, 64*(2), 225-256. doi: 10.1023/A:1024712527648

Cummins, R. A. (2013). Positive Psychology and Subjective Wellbeing Homeostasis: A critical examination of congruence. In D. Moraitou & A. Efklides (Eds.), *Quality of Life: A Positive Psychology Perspective* (pp. 67-86). New York: Springer

Cummins, R. A. (2015). The theory of subjective wellbeing homeostasis: A contribution to understanding life quality. In F. Maggino (Ed.), *A Life Devoted to Quality of Life - Festschrift in Honor of Alex C. Michalos*. Dordrecht, Netherlands: Springer.

Cummins, R. A., Eckersley, R., Pallant, J., Van Vugt, J., & Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social Indicators Research, 64*(2), 159-190. doi: 10.1023/A:1024704320683

Cummins, R. A., Li, L., Wooden, M., & Stokes, M. (2014). A demonstration of set-points for subjective wellbeing. *Journal of Happiness Studies, 15*, 183-206. doi: 10.1007/s10902-013-9444-9

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment, 49*(1), 71-75.

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective Weil-Being: Three Decades of Progress. *Psychological Bulletin, 125*(2), 276-302.

Easterlin, R. A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences, 100*(19), 11176-11183. doi: 10.1073/pnas.1633144100

Ehrhardt, J. J., Saris, W. E., & Veenhoven, R. (2000). Stability of life-satisfaction over time. *Journal of Happiness Studies, 1*(2), 177-205. doi: 10.1023/A:1010084410679

Eid, M., & Kutscher, T. (2014). Statistical models for analyzing stability of change in happiness *Stability of happiness: Theories and evidence of whether happiness can change* (pp. 263-304). Amsterdam: Elsevier.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., & Katz, I. R. (2005). Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics, 6*(1), 119-143. doi: 10.1093/biostatistics/kxh022

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112*(1), 60. doi: 10.1037/0033-295X.112.1.60

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. New York: NY: CRC press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: NY: Cambridge University Press.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6*(4), 733-760.

Gerstorf, D., Ram, N., Estabrook, R., Schupp, J., Wagner, G. G., & Lindenberger, U. (2008). Life satisfaction shows terminal decline in old age: longitudinal evidence from the German Socio-Economic Panel Study (SOEP). *Developmental Psychology, 44*(4), 1148. doi: 10.1037/0012-1649.44.4.1148

Hartmann, G. W. (1934). Personality traits associated with variations in happiness. *Journal of Abnormal and Social Psychology, 29*, 202-212. doi: 10.1037/h0073108

Headey, B., & Wearing, A. (1989). Personality, life events, and subjective well-being: toward a dynamic equilibrium model. *Journal of Personality and Social Psychology, 57*(4), 731-739. doi: 10.1037/0022-3514.57.4.731

Headey, B., & Wearing, A. J. (1992). *Understanding happiness: A theory of subjective well-being*. Melbourne, Victoria, Australia: Longman Cheshire.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling, 20*(2), 331-360. doi: 10.1016/S0264-9993(02)00057-3

Helson, R., Jones, C., & Kwan, V. S. (2002). Personality change over 40 years of adulthood: hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology, 83*(3), 752-766. doi: 10.1037/0022-3514.83.3.752

International Wellbeing Group. (2013). *Personal Wellbeing Index Manual*. Melbourne: Deakin University. http://www.deakin.edu.au/research/acqol/iwbg/wellbeing-index/index.php.

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

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial introduction with R*. Burlington, MA: Academic Press.

Kruschke, J. K. (2013). Posterior predictive checks can and should be Bayesian: Comment on Gelman and Shalizi,'Philosophy and the practice of Bayesian statistics'. *British Journal of Mathematical and Statistical Psychology, 66*(1), 45-56. doi: 10.1111/j.2044-8317.2012.02063.x

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*(6), 1042-1060. doi: 10.1037/a0020962

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion, 7*(2), 262-274. doi: 10.1037/1528-3542.7.2.262

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review, 15*(1), 1-15. doi: 10.3758/PBR.15.1.1

Lucas, R. E. (2007). Long-term disability is associated with lasting changes in subjective well-being: evidence from two nationally representative longitudinal studies. *Journal of Personality and Social Psychology, 92*(4), 717-730. doi: 10.1037/0022-3514.92.4.717

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: reactions to changes in marital status. *Journal of Personality and Social Psychology, 84*(3), 527-539. doi: 10.1037/0022-3514.84.3.527

Lucas, R. E., & Donnellan, M. B. (2007). How stable is happiness? Using the STARTS model to estimate the stability of life satisfaction. *Journal of Research in Personality, 41*(5), 1091-1098. doi: 10.1016/j.jrp.2006.11.005

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology, 101*(4), 847-861. doi: 10.1037/a0024298

Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research, 105*(3), 323-331. doi: 10.1007/s11205-011-9783-z

Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7*(3), 186-189. doi: 10.1111/j.1467-9280.1996.tb00355.x

Mancini, A. D., Bonanno, G. A., & Clark, A. E. (2011). Stepping off the hedonic treadmill. *Journal of Individual Differences, 32*(3), 144-152. doi: 10.1027/1614-0001/a000047

Mastekaasa, A., & Moum, T. (1984). The perceived quality of life in Norway: regional variations and contextual effects. *Social Indicators Research, 14*(4), 385-419. doi: 10.1007/BF00300450

Mroczek, D. K., & Spiro III, A. (2005). Change in life satisfaction during adulthood: findings from the veterans affairs normative aging study.

Journal of Personality and Social Psychology, 88(1), 189-202. doi: 10.1037/0022-3514.88.1.189

Nikodijevic, A., Moulding, R., Anglim, J., Aardema, F., & Nedeljkovic, M. (2015). Fear of self, doubt and obsessive compulsive symptoms. *Journal of Behavior therapy and Experimental Psychiatry* doi: 10.1016/j.jbtep.2015.02.005

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: a cohort-sequential longitudinal study. *Journal of Personality and Social Psychology, 98*(4), 645-658. doi: 10.1037/a0018769

Palmore, E., & Kivett, V. (1977). Change in life satisfaction: A longitudinal study of persons aged 46-70. *Journal of Gerontology, 32*(3), 311-316. doi: 10.1093/geronj/32.3.311

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2013). R Development Core Team (2012) nlme: linear and nonlinear mixed effects models. R package version 3.1-103. *R Foundation for Statistical Computing, Vienna*

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*(10), 421-425. doi: 10.1016/s1364-6613(02)01964-2

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.* Paper presented at the Proceedings of the 3rd international workshop on distributed statistical computing.

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology, 89*(3), 395-406. doi: 10.1037/0022-3514.89.3.395

Sinharay, S., & Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference, 111*(1), 209-221. doi: 10.1016/S0378-3758(02)00303-8

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(3), 485-493.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639. doi: 10.1111/1467-9868.00353

Suh, E., Diener, E., & Fujita, F. (1996). Events and subjective well-being: only recent events matter. *Journal of Personality and Social Psychology, 70*(5), 1091-1102. doi: 10.1037/0022-3514.70.5.1091

Tomyn, A. J., & Cummins, R. A. (2011). Subjective wellbeing and homeostatically protected mood: Theory validation with adolescents. *Journal of Happiness Studies, 12*(5), 897-914. doi: 10.1007/s10902-010-9235-5

Wang, M. (2007). Profiling retirees in the retirement transition and adjustment process: examining the longitudinal change patterns of retirees' psychological well-being. *Journal of Applied Psychology, 92*(2), 455-474. doi: 10.1037/0021-9010.92.2.455

Table 1
Overall Model Fit Statistics for Candidate Models

	Untransformed			Transformed		
	Deviance	Penalty	DIC	Deviance	Penalty	DIC
No Lag						
Fixed error intercept	16730	1024	17754	-1851	1024	-827
Random error intercept	13873	2000	15873	-3848	2083	-1765
Fixed error polynomial	15626	1708	17334	-3259	2032	-1227
Random error polynomial	12709	2992	**15702**	-5409	3565	**-1845**
Lag						
Fixed error intercept	16981	1030	18010	-1602	1029	-573
Random error intercept	14265	1944	16210	-3517	2067	-1450
Fixed error polynomial	16107	1559	17666	-3102	1998	-1104
Random error polynomial	13348	2713	16061	-5156	3477	-1680

Note. Model fit statistics can not be compared across transformed and untransformed data. Models with smaller deviance are better fitting and models with larger penalties are more complex. DIC incorporates both fit and complexity whereby smaller models are generally preferred. Smallest DIC for transformed and untransformed data are shown in bold.

Table 2

Parameter Estimates for Untransformed, Random Error, No Lag Models

	Intercept			Polynomial		
	M	Lower CI	Upper CI	M	Lower CI	Upper CI
Mean of intercepts $\mu_{\beta_{\text{intercept}}}$	7.56	7.48	7.64	7.54	7.46	7.62
SD of intercepts $\mu_{\beta_{SD}}$	-0.47	-0.50	-0.44	-0.55	-0.58	-0.51
Mean of SDs σ_i	0.69	0.68	0.71	0.65	0.64	0.67
SD of intercepts $\sigma_{\beta_{\text{intercept}}}$	1.19	1.13	1.24	1.19	1.13	1.25
SD of linear $\sigma_{\beta_{\text{linear}}}$				0.09	0.08	0.10
SD of quadratic $\sigma_{\beta_{\text{quadratic}}}$				0.04	0.03	0.04
SD of Beta SDs $\sigma_{\beta_{SD}}$	0.45	0.42	0.48	0.48	0.45	0.51
SD of SDs σ_i	0.34	0.32	0.36	0.34	0.32	0.37
$\text{cor}(\beta_{\text{intercept}}, \beta_{SD})$	-0.68	-0.73	-0.62	-0.67	-0.72	-0.62
Age linear effect θ	0.008810	0.003640	0.013850	0.012490	0.006420	0.018900
Age quadratic θ	-0.000410	-0.000620	-0.000210	-0.000280	-0.000510	-0.000050
Age cubic θ	-0.000020	-0.000030	-0.000010	-0.000020	-0.000030	-0.000010

Note. M, Lower CI, and Upper CI are the mean and lower and upper 95% credible intervals of the posterior density estimate of parameters. Intercept, linear and quadratic effects refer the random effects of time, whereas age effects are fixed effects of age.

Table 3
Parameter Estimates for Probit Transformed, Random Error, No Lag Models

	Intercept			Polynomial		
	M	Lower CI	Upper CI	M	Lower CI	Upper CI
Mean of intercepts $\mu_{\beta_{\text{intercept}}}$	0.67	0.65	0.69	0.66	0.64	0.69
SD of intercepts $\mu_{\beta_{SD}}$	-1.68	-1.71	-1.65	-1.79	-1.82	-1.75
Mean of SDs σ_i	0.20	0.20	0.20	0.18	0.18	0.18
SD of intercepts $\sigma_{\beta_{\text{intercept}}}$	0.34	0.33	0.36	0.34	0.33	0.36
SD of linear $\sigma_{\beta_{\text{linear}}}$				0.04	0.03	0.04
SD of quadratic $\sigma_{\beta_{\text{quadratic}}}$				0.02	0.02	0.02
SD of Beta SDs $\sigma_{\beta_{SD}}$	0.36	0.34	0.39	0.39	0.36	0.42
SD of SDs σ_i	0.08	0.07	0.08	0.07	0.07	0.08
$\text{cor}(\beta_{\text{intercept}}, \beta_{SD})$	-0.21	-0.29	-0.13	-0.21	-0.30	-0.13
Age linear effect θ	0.002490	0.000870	0.004080	0.004250	0.002280	0.006180
Age quadratic θ	-0.000160	-0.000230	-0.000100	-0.000120	-0.000200	-0.000040
Age cubic θ	-0.000010	-0.000010	0.000000	-0.000010	-0.000010	0.000000

Note. M, Lower CI, and Upper CI are the mean and lower and upper 95% credible intervals of the posterior density estimate of parameters. Intercept, linear and quadratic effects refer the random effects of time, whereas age effects are fixed effects of age.

Table 4
Posterior Predictive Checks for Models: Untransformed Data

		No Lag				Lag			
		Intercept		Polynomial		Intercept		Polynomial	
Statistic	Dataset	FE	RE	FE	RE	FE	RE	FE	RE
Mean($mean_i$)	7.50	**7.50**	**7.50**	**7.51**	**7.51**	**7.50**	**7.51**	**7.51**	**7.50**
SD($mean_i$)	1.25	**1.27**	**1.22**	**1.26**	**1.23**	**1.26**	**1.21**	**1.26**	**1.21**
Skew($mean_i$)	-0.94	-0.01	-0.11	0.00	-0.10	0.00	-0.18	0.00	-0.14
Mean(SD_i)	0.66	0.73	**0.66**	0.73	**0.67**	0.72	**0.66**	0.73	**0.67**
SD(SD_i)	0.39	0.23	**0.38**	0.23	**0.38**	0.23	**0.38**	0.23	**0.38**
Mean($AR1_i$)	-0.06	-0.15	-0.15	**-0.09**	**-0.08**	**-0.06**	**-0.04**	**-0.06**	**-0.04**
Mean($skew_i$)	-0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean($outlier_i$)	1.08	**1.16**	**1.05**	**1.17**	**1.07**	**1.15**	**1.06**	**1.16**	**1.06**
SD($outlier_i$)	0.72	0.40	**0.64**	0.43	**0.64**	0.41	**0.64**	0.42	**0.64**
SD($beta\_age_i$)	0.22	**0.18**	**0.18**	**0.20**	**0.19**	**0.20**	**0.21**	**0.20**	**0.20**
Cor($mean_i$, SD_i)	-0.48	0.00	**-0.50**	0.00	**-0.48**	0.00	**-0.51**	0.00	**-0.49**

Note. FE = Fixed error (within person SD); RE = Random error (within person SD). Values indicate mean statistic for samples generated from posterior density estimates. Dataset is the statistic for the sample data. Mean, SD, Skew and Cor are functions of participant-level statistics. Arguments in parentheses with subscript i indicate that it is a property of the individual (e.g., mean SWB score for ith individual). See method for details.

Table 5

Posterior Predictive Checks for Models: Probit Transformed Data

		No Lag				Lag			
		Intercept		Polynomial		Intercept		Polynomial	
Statistic	Dataset	FE	RE	FE	RE	FE	RE	FE	RE
Mean($mean_i$)	7.50	**7.48**	**7.49**	**7.48**	**7.49**	**7.48**	**7.50**	**7.48**	**7.49**
SD($mean_i$)	1.25	**1.22**	**1.22**	**1.23**	**1.23**	**1.22**	**1.21**	**1.23**	**1.22**
Skew($mean_i$)	-0.94	**-0.54**	**-0.56**	**-0.54**	**-0.55**	**-0.54**	**-0.57**	**-0.55**	**-0.56**
Mean(SD_i)	0.66	0.70	**0.66**	0.73	**0.70**	0.69	**0.66**	0.73	**0.69**
SD(SD_i)	0.39	0.26	**0.37**	0.29	**0.37**	0.26	**0.37**	0.29	**0.37**
Mean($AR1_i$)	-0.06	-0.15	-0.15	**-0.05**	**-0.03**	**-0.06**	**-0.05**	**-0.04**	**-0.01**
Mean($skew_i$)	-0.17	**-0.12**	**-0.10**	**-0.12**	**-0.11**	**-0.12**	**-0.10**	**-0.12**	**-0.11**
Mean($outlier_i$)	1.08	**1.11**	**1.05**	**1.18**	**1.12**	**1.10**	**1.04**	**1.17**	**1.11**
SD($outlier_i$)	0.72	0.45	**0.61**	0.54	**0.64**	0.46	**0.61**	0.53	**0.63**
SD($beta\_age_i$)	0.22	**0.17**	**0.17**	**0.21**	**0.20**	**0.20**	**0.20**	**0.21**	**0.20**
Cor($mean_i$, SD_i)	-0.48	**-0.52**	**-0.48**	**-0.49**	**-0.49**	**-0.51**	**-0.48**	**-0.49**	**-0.49**

Note. FE = Fixed error (within person SD); RE = Random error (within person SD). Values indicate mean statistic for samples generated from posterior density estimates. Dataset is the statistic for the sample data. Mean, SD, Skew and Cor are functions of participant-level statistics. Arguments in parentheses with subscript i indcate that it is a property of the individual (e.g., mean SWB score for ith individual). See method for details.

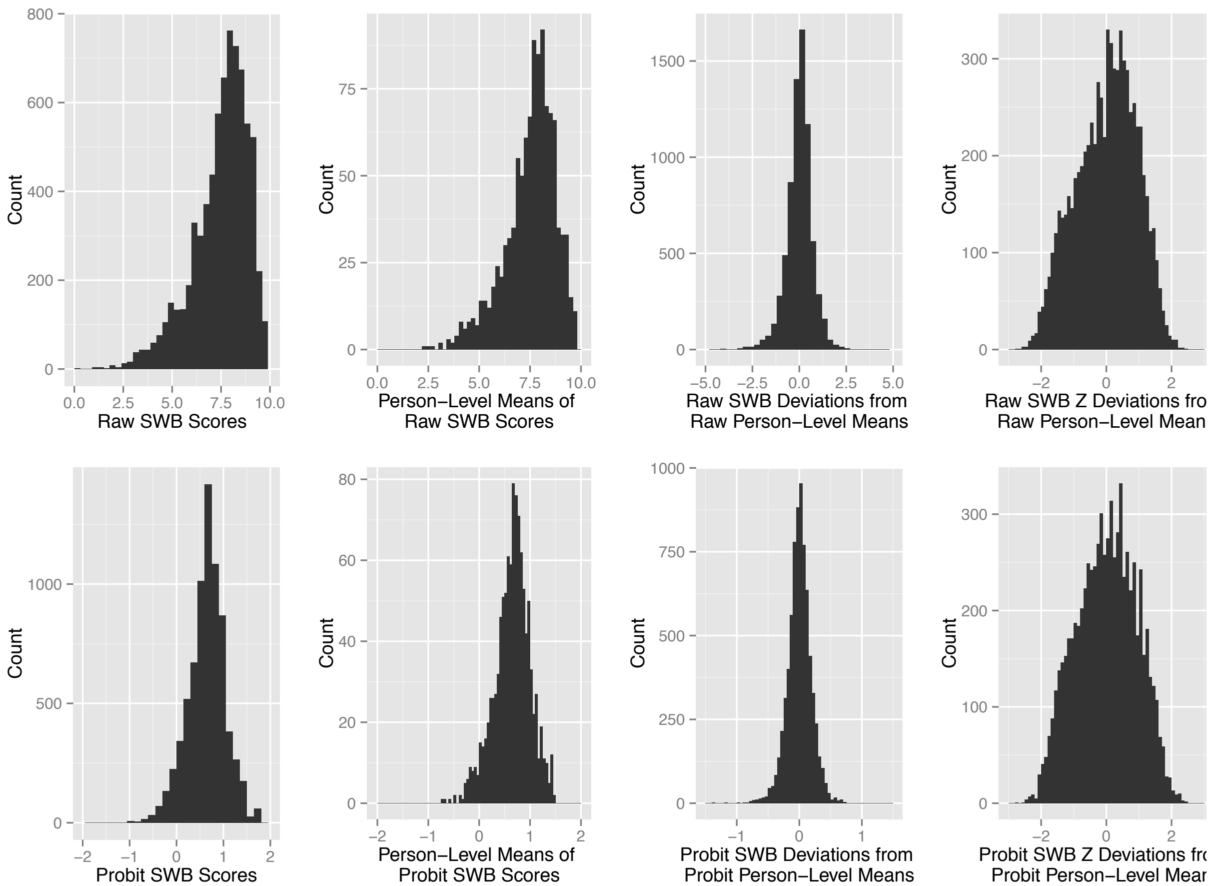

Figure 1. **Histograms showing the distribution for observations, person-level means, and deviation of observations from person-level means, and standarized (Z) deviations from person-level means for both raw and probit transformed SWB.**

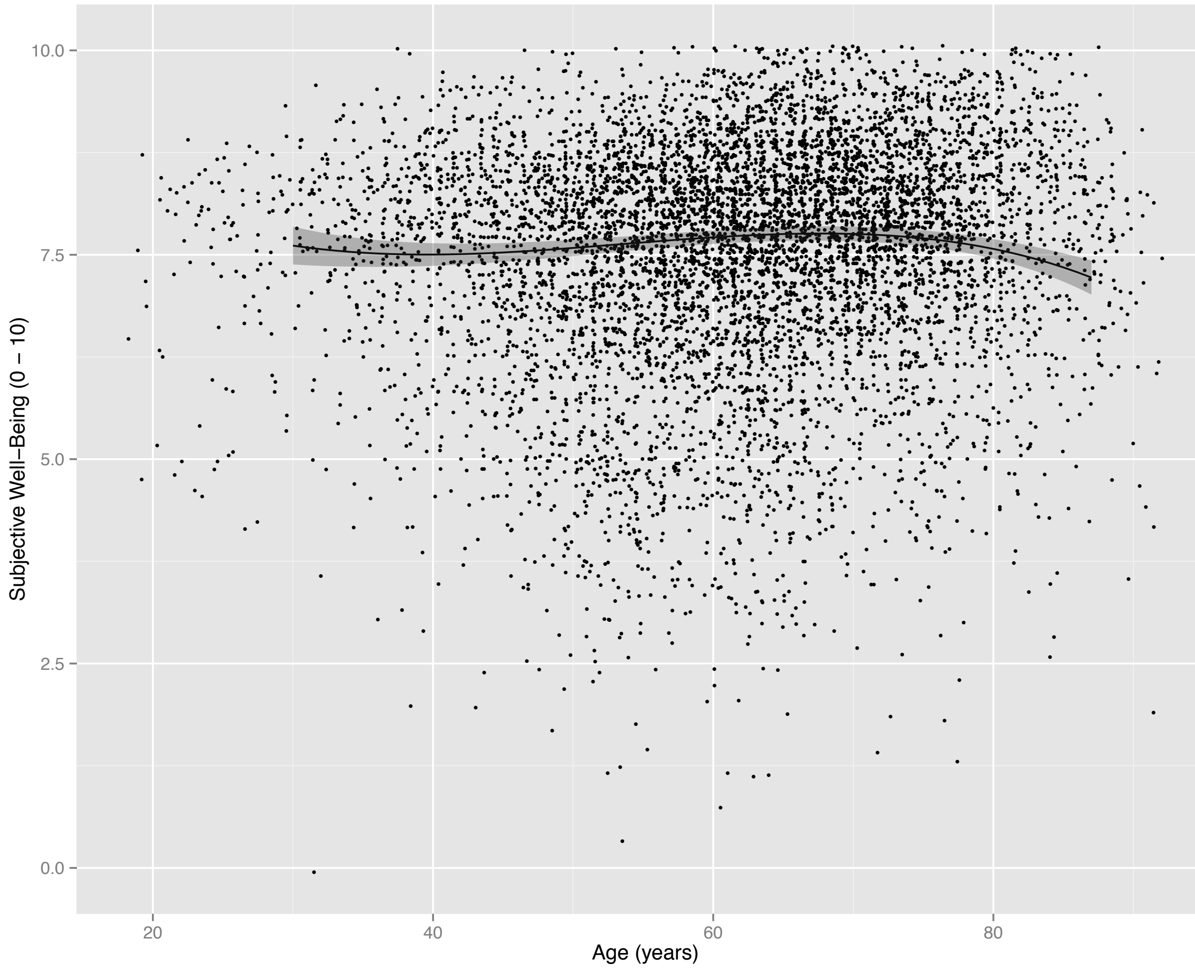

Figure 2. **The relationship between age and Subjective Well-Being. Points are raw data after a small amount of random noise has been applied to reduce the overlap of points and better highlight the density of the data. Line of best fit and error band represents predictions and 95% credible intervals for estimates of group-level age effect based on probit transformed, polynomial, no lag, random error model. Predictions are truncated to only show data with at least 100 data points above or below the given age.**

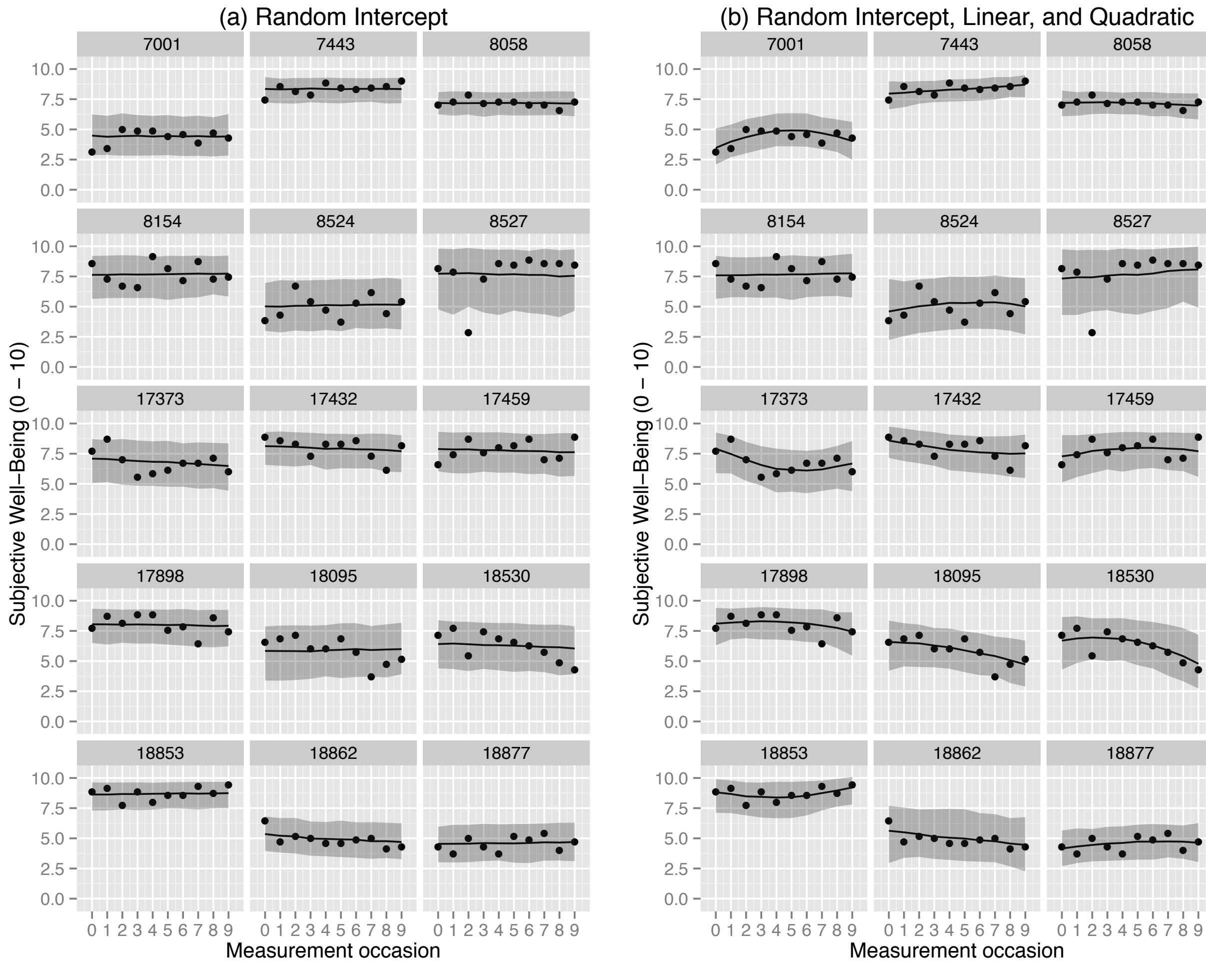

Figure 3. Relationship between measurement occasion and SWB for a random sample of participants who provided data for all 10 measurement occasions. Each cell is one participant where participant ID is shown above. The same participants are used in both Figure 3 (a) and 3 (b). These two models represent the two best fitting models. In each cell the dark line indicates the expected value and the shaded area is the 95% credible interval for the given participant on the given measurement occasion based on the probit transformed, no lag, random error model with either (a) only a random-intercept or (b) random polynomial (i.e., random intercept, linear, and quadratic terms).