

# Predicting Average Daily Mood using Physical and Sleep Activity with Bayesian Structural Time Series

*Jonathan M. Bryan*

*June 13, 2018*

## Introduction

### Ecological momentary assessment (EMA)

Typical methods for capturing the symptoms and behaviors of psychopathology include clinician assessments (e.g. Clinical Global Impressions (CGI) Scale and Clinically Administered PTSD Scale), self-report and caregiver questionnaires. These assessments are often summarized, global, and retrospective in nature, limiting their accuracy of measurements of real-world phenomena.<sup>1</sup> Clinical and caregiver instruments that measure a subject's psychological state and social functioning are hindered by the bias introduced by an outside observer. Often there are significant discrepancies in how information is both elicited from research subjects and perceived from the observer. Instruments are often validated using a measure of interrater reliability but this only provides quantification of test uncertainty for how subject's answers are perceived. How and when information is elicited from patients can still vary greatly and have significant effects on measurement accuracy.

Self-report questionnaires rely heavily on the ability of subjects to recall previous emotions and experiences, sometimes over significant periods of time. Memories of events constitute an imperfect reconstructive process that inserts systematic bias into information retrieved at recall.<sup>2</sup> Recall is mediated by the uniqueness and emotional intensity of the event to be remembered, as well as the psychological state of the subject at the time of recall. For example, people in negative moods recall more negative information than would be expected at random.<sup>3</sup> The timing of assessment can also influence how subjects respond. For example, a person may remember an event as markedly more negative at night than they would during the afternoon. Ecological momentary assessment or experience sampling is a collection of methods that asks research subjects to repeatedly record mood, behaviors, and environmental interactions in real time.<sup>4</sup> Subjects are asked to complete daily dairies, mood journals, psychometric tests, and other assessments once or multiple times each day.

EMA is motivated by evidence that says reducing the time between daily experiences and measurements of those experiences improves the accuracy of recall. In addition, the integration of EMA within the daily lives of research subjects allows for greater ecological validity or generalizability of results. Daily prompts for patients to record EMA can be driven by some pre-determined or random schedule. Prompts may also be based on specific events experienced by subjects. Several studies have established the validity and reliability of EMA methods to measure changes in affect, mood, and social interactions.<sup>5</sup> Importantly, many EMA methods are temporally and internally consistent, in other words, EMA measures across time and specific items are reliable. Collecting daily records of the psychological state of subjects allows for more sensitive analysis of rates of changes across subjects and psychological states as well as the ordering of changes in reported psychological states and associated events and phenomena.<sup>6</sup>

### Wearable technology

Wearable technology are electronic devices that are worn by the user and whose principal feature is the passive collection of biometric and environmental data. There are many different types of wearable technology, to be worn on various parts of the body, and that carry an array of sensors. Certain models of smartphones constitute wearable technology, as many are able to capture physical activity such as step counts, heart rate,

and sleep activity. Wearable technology is in some sense analogous to medical diagnostic tests as EMA is to discrete retrospective assessments. Wearable technologies give researchers access to continuous measurements of biometric data that would otherwise be collected in discrete clinical settings. Wearables have been used for many biomedical studies, including longitudinal monitoring of symptoms for chronic disease, functional assessments in rehabilitation, and disease prediction.<sup>789</sup>

While EMA may be preferable to discrete retrospective methods, it still places an active burden on research participants and such cognitive load may affect their recorded psychological state.<sup>10</sup> There is active research to understand how wearable sensors may serve as proxy measurements of psychological states in lieu of EMA. Recent technological developments have demonstrated the ability to passively record the intensity and valence (i.e. positive or negative) of emotional states using biometric data rather than through active participation.<sup>1112</sup> Combination measurements including but not limited to skin conductance, heart rate, and blood pressure have been shown to be correlated to emotional arousal.<sup>13</sup>

When not being used as a proxy for psychological states, wearable sensor data has the capability to verify, enrich, and aid in the interpretation of subjective experience data collected using EMA.<sup>14</sup> Environmental triggers from wearable sensors may also increase EMA adherence for subjects enrolled in studies by only prompting subjects when necessary and avoiding redundant questions.<sup>15</sup> Heart rate variability has been shown to characterize some mental illnesses including panic disorders and major depression.<sup>16</sup> Researchers have designed so-called context-aware EMA, that uses a GPS-enabled smart device to prompt research subjects for written, audio, and visual assessment at specific locations.<sup>17</sup>

Critical to the use of wearable technology is the validity and reliability of the sensors to capture the phenomena of interest across the various settings of everyday life. One systematic review of consumer-wearable devices found higher validity and reliability of steps count data and lower validity for energy expenditure and sleep data.<sup>18</sup> Notably the Consumer Technology Association has established a Health and Fitness Technology working group that has developed standards for physical activity, sleep, and consumer EEG monitors. These standards include functional definitions, appropriate sensor measures for different methodologies, and performance criteria for sensor evaluation.

## Methods

### Data Extraction

All data collection was through a Samsung S8+ smartphone from January 4, 2018 to June 5, 2018. Mood data was recorded using Daylio, a digital diary application that allows users to record their daily mood, activities, and free text diary entries. Mood was recorded on a 5-point scale with qualitative descriptions (e.g. sad, depressed, happy, excited). Some days included multiple mood entries so an average daily mood was calculated for each day. Physical and sleep activity data were recorded using a collection of sensors available on the phone including an accelerometer, gyrometer, and GPS. These data were aggregated into the Samsung Health application that summarizes the raw data for viewing and download. The extracted data included longest idle time, step count, active time, calories, sleep quality, and total sleep time. Predictive mean matching was used for missing data (approximately 7 percent of data was missing).

### Bayesian Structural Times Series

Bayesian structural time series allow for the incorporation of usual time series components including trend and seasonal components, as well as additional regression predictors to model the target response over time.<sup>19</sup> The two basic components of structural time series include the observation and transition equations.

$$y_t = Z_t^T \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t) \quad (\text{Observation equation})$$

$$\alpha_{t+1} = T_t^T \alpha_t + R_t \eta, \quad \eta_t \sim N(0, Q_t) \quad (\text{Transition equation})$$

$y_t$  : observed data at time  $t$   
 $Z_t, T_t$  : structural parameters  
 $H_t$  : observation covariance matrix  
 $\alpha_t$  : latent state parameter  
 $R_t$  : captures linear dependencies and induces full rank of  $Q_t$   
 $Q_t$  : transition covariance matrix

The observation equation relates the response to the latent state variables and structural parameters of the model. The transition equation specifies how the latent states change over time. Structural time series are modular and allow for the flexibility of regression models along with the inclusion of traditional time series components including autoregression, trend, and seasonality.<sup>20</sup>

## Regression Priors and Variable Selection

Sparsity is induced on the regression components by using so-called spike and slab variable selection. The spike specifies each regression parameter under a Bernoulli distribution, where  $\gamma$  is a vector of ones and zeroes denoting whether a variable is included in the model. The slab is a flat prior with large variance (here a normal prior). Prior elicitation uses the expected number of regression parameters, expected  $R^2$ , beta weight, and sigma weight.

$$\gamma \sim \prod_j \pi_j^{\gamma_j} (1 - \pi_j)^{1-\gamma_j}$$

$$\beta_\gamma | \gamma, \sigma^2 \sim N(b_\gamma, \sigma^2 (\Omega_\gamma^{-1})^{-1})$$

$$\frac{1}{\sigma^2} \sim \Gamma\left(\frac{df}{2}, \frac{SS}{2}\right)$$

$$\pi_j : \frac{\text{expected model size}}{\text{number of predictors}}$$

$$b = \mathbf{0}$$

$$\Omega^{-1} = \kappa \frac{a(X^T X + (1 - \alpha) \text{diag}(X^T X))}{n}$$

$$\frac{ss}{df} = (1 - R_{\text{expected}}^2) s_y^2$$

$$df = 1$$

## Model Fitting

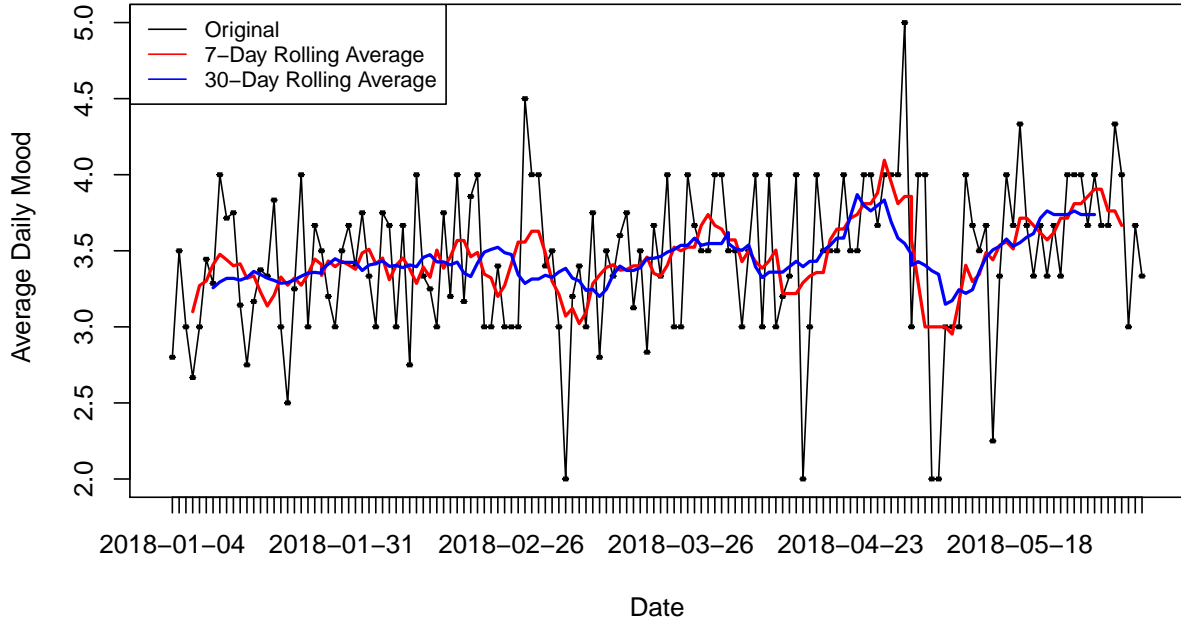
After a model has been specified the principal parts for model fitting include a Kalman filter, Bayesian data augmentation and Markov chain Monte Carlo sampling. For the times series decomposition, the latent state  $\alpha_{t+1}$  is sampled using a Kalman filter which recursively computes the target latent state using the previously observed observations  $(t-1, t-2, \dots, 1)$  based on the conditional independence of the observations given the latent state at times  $t-1, t-2, \dots, 1$ . A Kalman smoother then updates  $\alpha_t$  using all previous observations up to to time  $t$ . The simulated latent states can then be used in conjunction with the observed data to sample the time series model parameters. The regression component is captured within  $Z_t$  of the observed equation and is a single dimension within the  $T_t$  matrix, reducing the number of matrix-matrix multiplication steps needed for the Kalman filter to derive the latent states. We sample the regression parameters once we have

sampled the latent states and time series components. A summary of the model fitting process is depicted below, where  $\alpha$  is the latent states,  $\theta$  are the non-regression time series components,  $\beta$  are the regression coefficients, and  $\sigma_\epsilon$  is the covariance of the regression parameters.

1. Sample  $\alpha_t^s \sim p(\alpha|\theta, \beta, \sigma_\epsilon, y)$  using Kalman filter and smoother
2. Set  $y_t^* = y_t - Z_t^T \alpha_t^s$
3. Sample  $\theta^s \sim p(\theta|\alpha_t^s, y_t^*)$  inverse Wishart full conditional
4. Sample  $\beta^s, \sigma_\epsilon^s \sim p(\beta^s, \sigma_\epsilon^s|\alpha_t^s, y_t^*, \theta^s)$  Gaussian and inverse Wishart full conditional

## Analysis

**Fig. 1. Average Daily Mood from 2018-01-04 to 2018-06-05**

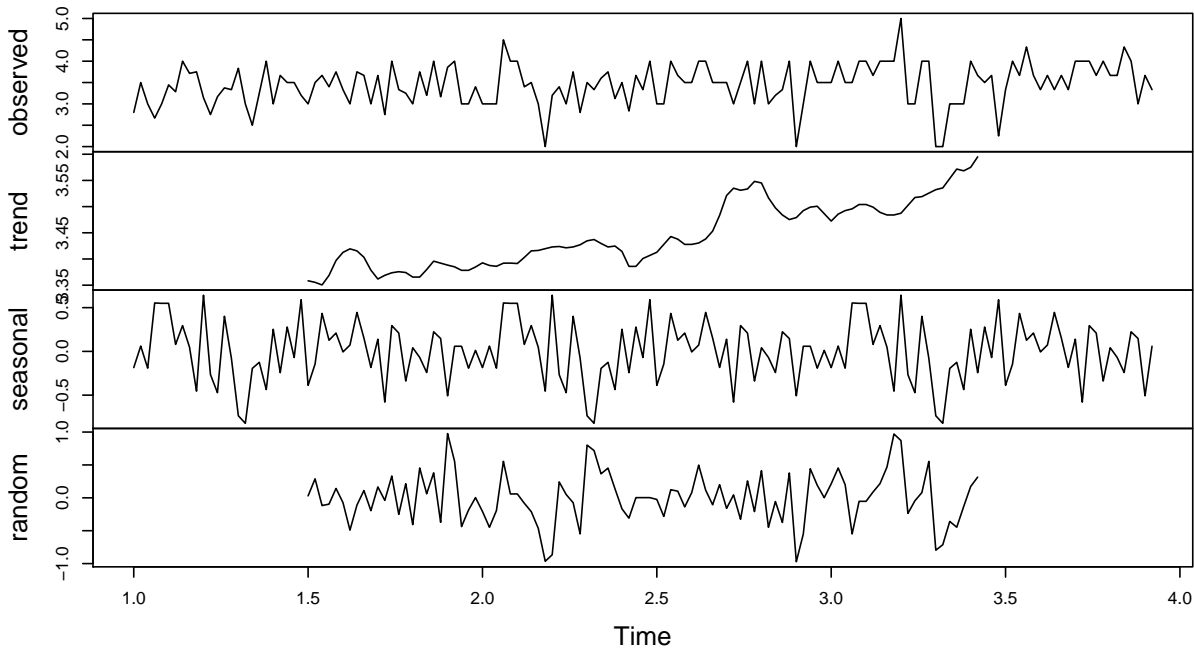


## Exploratory data analysis

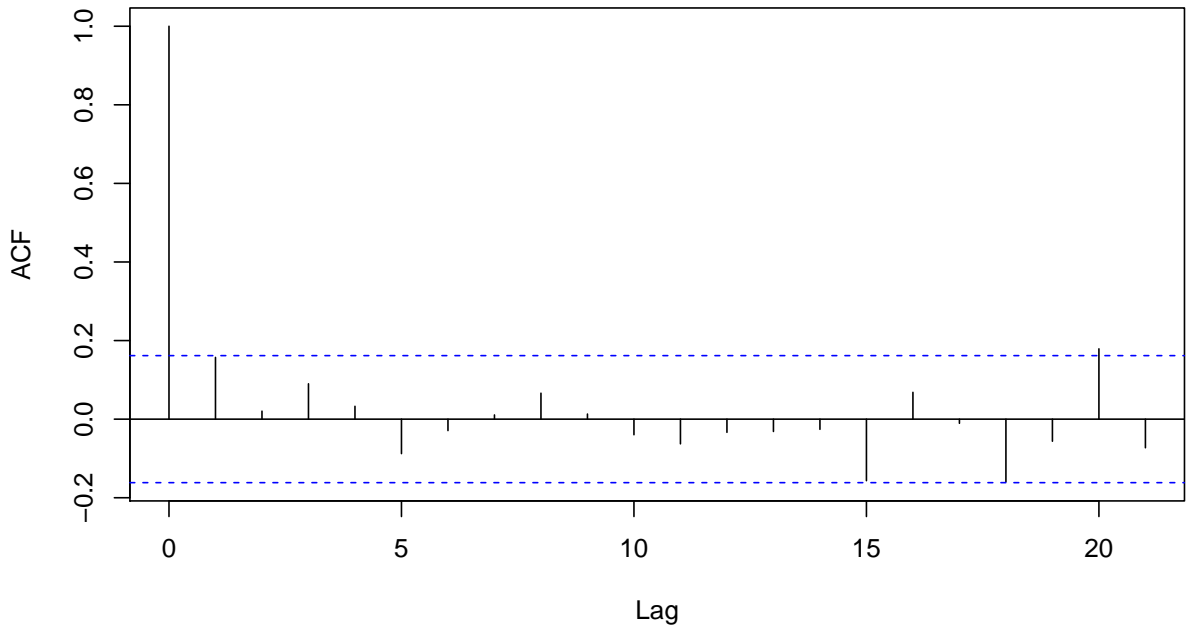
Decomposition of the five months of mood data suggested a non-seasonal, autoregressive pattern with some underlying positive trend (Figure 2). The mean of the mood time series increased from 3.30 to 3.60 from January to the beginning of June suggesting some local linear trend where the mean and slope of the trend follow a random walk. Evaluation of the sample autocorrelation and partial autocorrelation functions suggested the original series is close to stationary but with significant lags at 15 and 20 (Figures 3 and 4). An augmented Dickey-Fuller test (p-value = 0.01) suggested the detrended mood data was stationary with autoregressive lag order 5. Mood was shown to be weakly and positively correlated ( $\rho < 0.14$ ) with all of the regression covariates with the exception of longest idle time. Step count, active time, and calories expended were all strongly positively correlated ( $\rho > .98$ ). This is expected given that the indirect calculation of calories expended is likely some function of active time and step count with both of these variables highly correlated with general physical activity. Sleep quality was positively correlated with total sleep time and longest idle time. Such results suggested

that only a few of the available covariates were likely predictive of average daily fluctuations in mood.

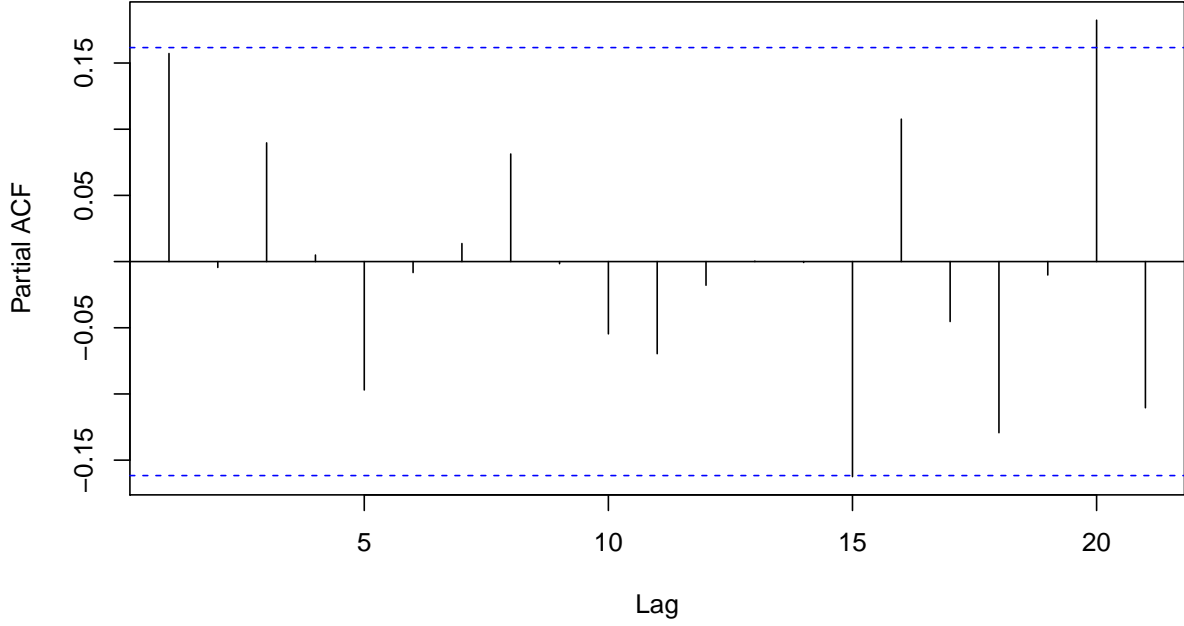
**Fig. 2. Decomposition of Mood Time Series**



**Fig. 3. ACF of Average Daily Mood**



**Fig. 4. PACF of Average Daily Mood**



## Model prediction error and residual analysis

Five models were specified using combinations of local linear trends, seasonality (21 weeks), retrogressive (lag 5), and covariate components (Table 1). The AR(5) model had the largest  $R^2$  and lowest Harvey's goodness of fit statistic, while the local linear trend models, including composite model with AR(5) and regression components, had the largest Harvey's goodness of fit statistic. The difference between the two tests is that Harvey's goodness of fit statistic is analogous to  $R^2$  but relative to a random walk with drift, which may be more relevant given the temporal aspects of the model.

$$1 - \frac{\sum_{i=1}^n \nu_i^2}{\sum_{i=2}^n n(\Delta y_i - \Delta \bar{y})^2} \quad (\text{Harvey's goodness of fit})$$

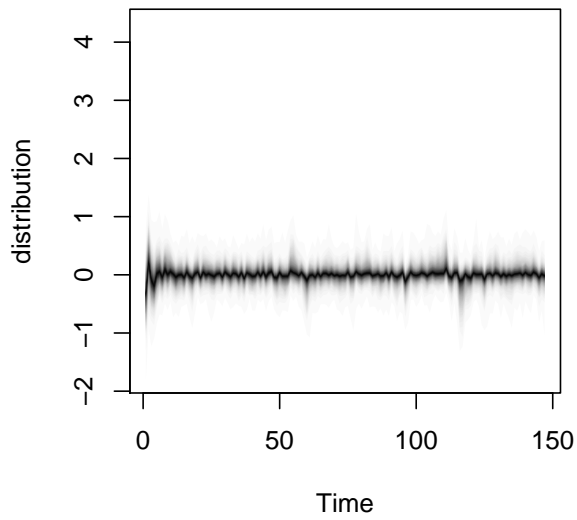
$\nu_i$  : one-step ahead prediction errors

Comparison of cumulative absolute error between the models suggested mood contained no weekly period and that the local linear trend component was a better model for daily changes in mood compared to an autoregressive model at lag 5. The regression components did not increase the predictive power of the model. Residuals of the autoregressive, local linear trend model appeared to be non-normal, with significant right skew (Figure 7) suggesting the model overpredicts more often than it should. The autocorrelation function of the residuals revealed no significant lag correlation (Figure 8). Rerunning the model with a local linear trend using t-distribution errors yielded no significant increase in predictive performance but showed greater congruence with the t-distribution.

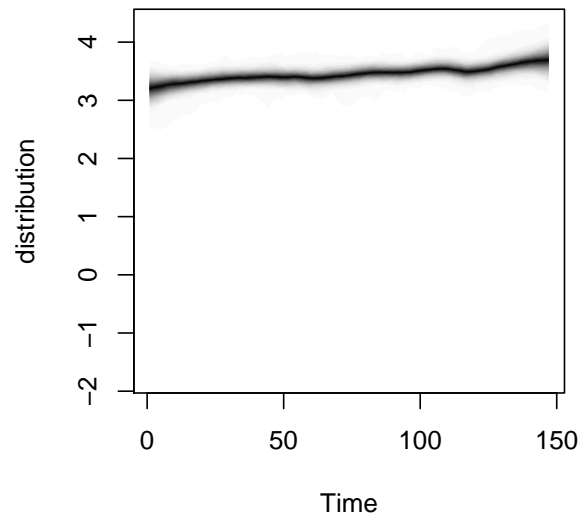
Table 1: Table 1. BSTS Summary Statistics

	Residual SD	1-Step Pred. SD	R-squared	Harvey's GoF
LLT + Seasonal + Reg	0.504	0.554	0.000	0.249
LLT + Reg	0.486	0.515	0.039	0.354
LLT	0.492	0.514	0.013	0.356
AR(5)	0.436	0.572	0.225	0.173
LLT + AR(5)	0.463	0.513	0.125	0.359

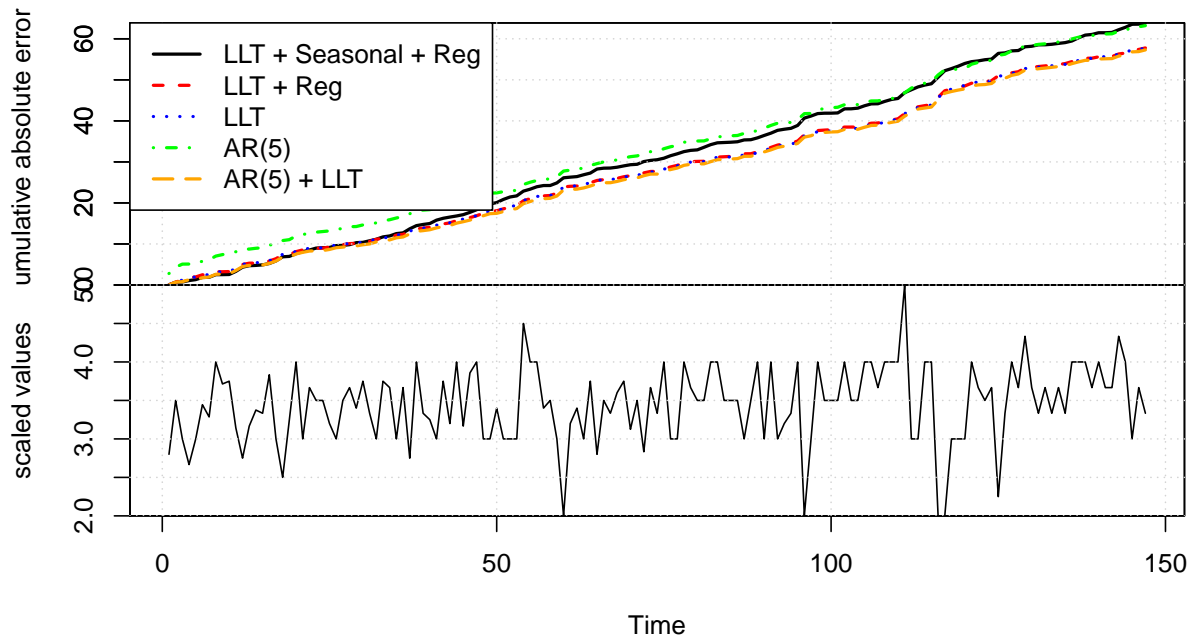
**Fig. 5. Ar5**



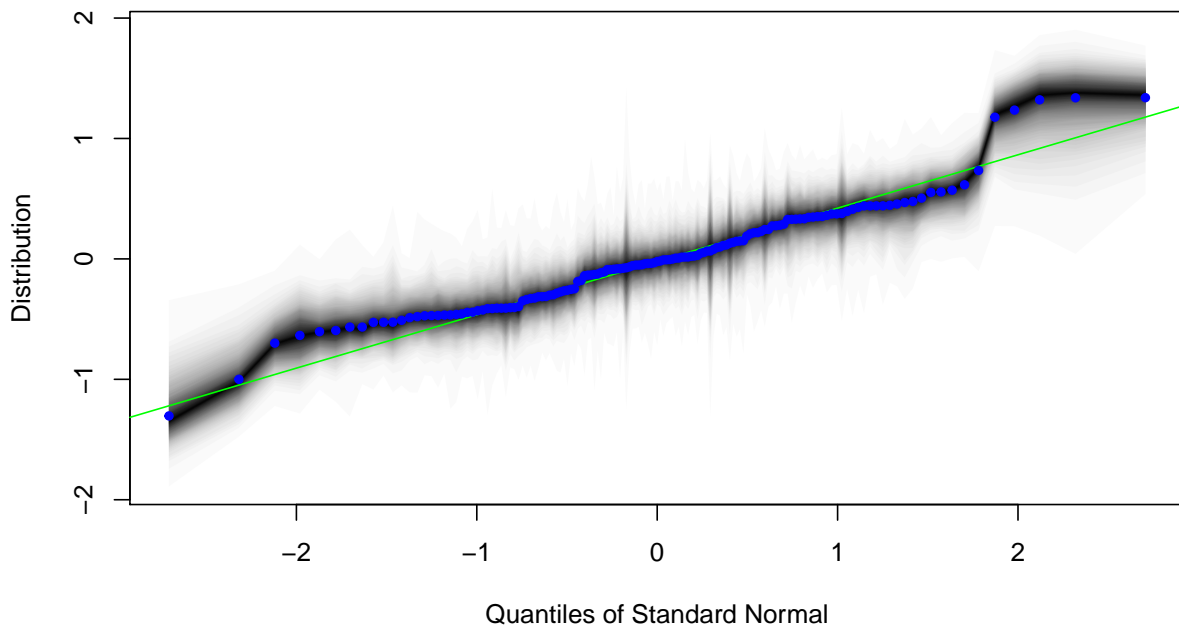
**Fig. 5. trend**



**Fig. 6. Average Daily Mood BSTS Comparison**

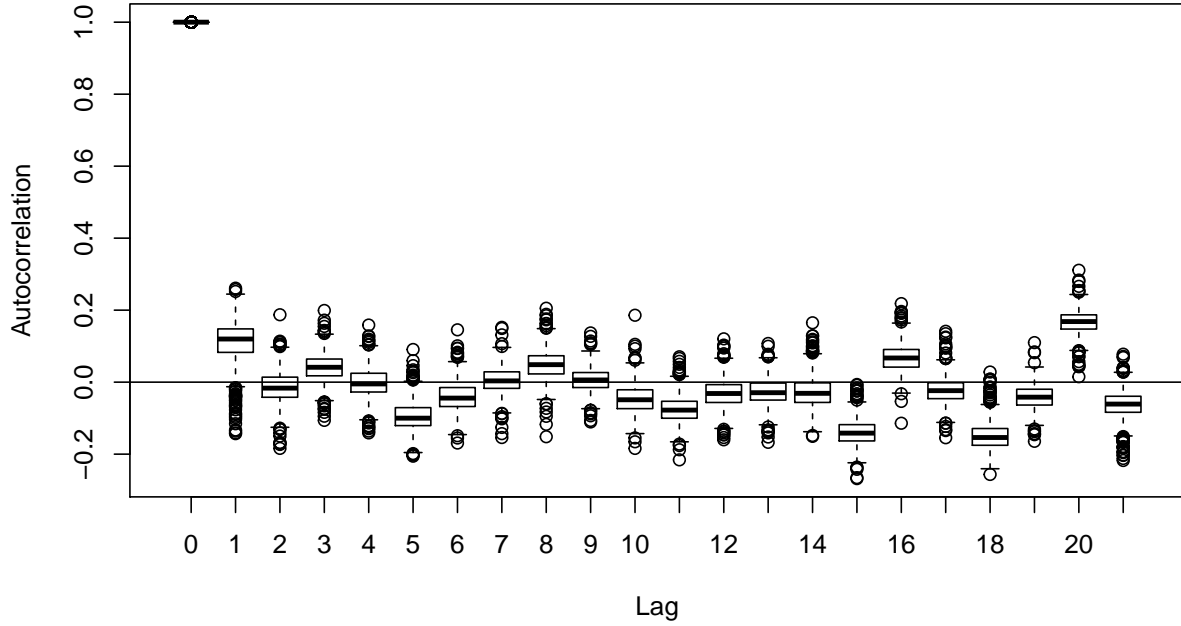


**Fig. 7. AR(5) Local Linear Trend BSTS QQ-Plot**





**Fig. 8. AR(5) Local Linear Trend BSTS ACF**



## Influence of Physical and Sleep Activity

Bayesian structural time series use model averaging, such that an inclusion probability for the regression components can be calculated. None of the regression components had an inclusion probability above 3 percent for the models that included them. This further corroborated the data from the model prediction error that the physical and sleep activity data explain little of the variation in average daily mood.

## Discussion

Analysis suggested two primary insights: 1) average daily mood likely contained no weekly seasonal or autoregressive component; and 2) the physical and sleep activity covariates were not predictive of average daily mood. Rather average daily mood appears to be best modeled by a local linear trend model. This suggests that average daily mood is driven by some underlying and unobserved trend with stochastic fluctuations. Research has demonstrated that mood is strongly affected by interpersonal interactions which was not captured in this study.<sup>21</sup> Limitations of the data included the relatively short time frame of the data, making longer term seasonal components (e.g. monthly and yearly) unable to be measured. Unobserved physical and sleep conditions may have also contributed to average daily mood but were not captured by the data collection methods used. In comparison to other methods, EMA requires a higher time commitment from research subjects because more measurements are needed. Like other self-reported measurements, EMA lacks independent verification at the point of measurement. Research has also shown that bias in subject responses can occur over very short time frames (e.g. 30-minutes) putting some limitations on certain types of EMA assessments, such as daily diaries, that ask subjects to summarize their psychological state over the course of the day.<sup>22</sup>

## References

- <sup>1</sup> Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4, 1-32
- <sup>2</sup> Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, 236(4798), 157-161.
- <sup>3</sup> Clark, D. M., & Teasdale, J. D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of abnormal psychology*, 91(2), 87.
- <sup>4</sup> Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*.
- <sup>5</sup> Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*(pp. 35-54). Springer Netherlands.
- <sup>6</sup> Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, 31(1), 13.
- <sup>7</sup> Patel, S., Chen, B. R., Mancinelli, C., Paganoni, S., Shih, L., Welsh, M., . . . & Bonato, P. (2011, August). Longitudinal monitoring of patients with Parkinson's disease via wearable sensor technology in the home setting. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 1552-1555). IEEE.
- <sup>8</sup> Bonato, P. (2005). Advances in wearable technology and applications in physical medicine and rehabilitation.
- <sup>9</sup> Oresko, J. J., Jin, Z., Cheng, J., Huang, S., Sun, Y., Duschl, H., & Cheng, A. C. (2010). A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. *IEEE Transactions on Information Technology in Biomedicine*, 14(3), 734-740.
- <sup>10</sup> Fletcher, R. R., Tam, S., Omojola, O., Redemske, R., & Kwan, J. (2011, August). Wearable sensor platform and mobile application for use in cognitive behavioral therapy for drug addiction and PTSD. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 1802-1805). IEEE.
- <sup>11</sup> Fletcher, R. R., Dobson, K., Goodwin, M. S., Eydgahi, H., Wilder-Smith, O., Fernholz, D., . . . & Picard, R. W. (2010). iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 215-223.
- <sup>12</sup> Matheson, R., & MIT News Office. (2017, July 12). Wearable device reveals consumer emotions. Retrieved from <http://news.mit.edu/2017/wearable-device-reveals-consumer-emotions-0712>
- <sup>13</sup> Fletcher, R. R., Dobson, K., Goodwin, M. S., Eydgahi, H., Wilder-Smith, O., Fernholz, D., . . . & Picard, R. W. (2010). iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 215-223.
- <sup>14</sup> Gaggioli, A., Pioggia, G., Tartarisco, G., Baldus, G., Corda, D., Cipresso, P., & Riva, G. (2013). A mobile data collection platform for mental health research. *Personal and Ubiquitous Computing*, 17(2), 241-251.
- <sup>15</sup> Hicks, J., Ramanathan, N., Kim, D., Monibi, M., Selsky, J., Hansen, M., & Estrin, D. (2010, October). AndWellness: an open mobile system for activity and experience sampling. In *Wireless Health 2010* (pp. 34-43). ACM.
- <sup>16</sup> Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236. <sup>17</sup> Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., & Bao, L. (2003, April). A context-aware experience sampling tool. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 972-973). ACM.
- <sup>18</sup> Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 159.

- <sup>19</sup> Scott, S. L., & Varian, H. R. (2014). Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2), 4-23.
- <sup>20</sup> Scott, Steven L. Bayesian Structural Time Series Models. 10 Aug. 2015, docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbmxzdGV2ZXRoZWJheWVzaWFufGd4OjI2ZGEwMTk4M2VmOWRkOTE.
- <sup>21</sup> Bolger, N., DeLongis, A., Kessler, R. C., & Schilling, E. A. (1989). Effects of daily stress on negative mood. *Journal of personality and social psychology*, 57(5), 808.
- <sup>22</sup> Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain*, 104(1-2), 187-194.