Forecasting a heterogeneous set of mood data

why persistence is hard to beat

P.J. Moore[1]

[1]Oxford Centre for Industrial and Applied Mathematics
University of Oxford

January 17, 2012

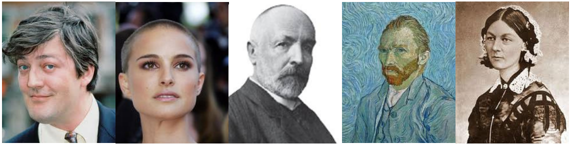
Outline

- This presentation investigates a data set of mood in bipolar disorder. The mood data has been collected every week from patients in Oxfordshire between December 2006 and August 2011.
- We look at data qualities, such as the time series length and the statistics of the intervals between responses.
- We look at the dynamics of the time series by 1) examining stationarity and smoothness and 2) insample forecasting.
- Out of forecasting results using simple exponential smoothing and Gaussian process prediction (kriging) are presented.
- Conclusions are drawn about the heterogeneity of the data and the sampling interval needed for generalising over the set of patients.

Bipolar Disorder

Bipolar disorder

- Bipolar disorder is a condition affecting mood and featuring episodes of mania and depression.
- Mania is a condition in which the sufferer might experience racing thoughts, impulsiveness, grandiose ideas and delusions. Under these circumstances, individuals are liable to indulge in activities which can be damaging both to themselves and to those around them.
- Depression is characterized by low mood, insomnia, problems with eating and weight, poor concentration, feelings of worthlessness, thoughts of death or suicide, a lack of general interest, fatigue and restlessness.



Bipolar Disorder - Subtypes

Diagnostic subtypes of bipolar disorder

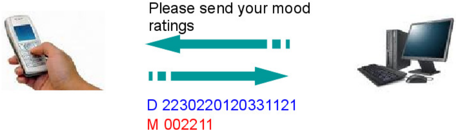
- Bipolar I. At least one manic episode which lasts at least seven days or requires hospital care. Depressive episodes usually occur.
- Bipolar II. At least one major depressive episode with accompanying hypomania - a period of persistently elevated, expansive or irritable mood lasting throughout four days or more.
- Cyclothymia. A history of hypomania and non-major depression.
- Bipolar NOS. Symptoms of mania and depression which do not fit into any categories above.

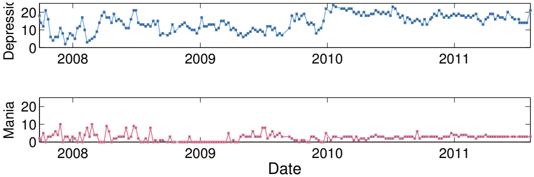
Literature

- The paucity of suitable data has also constrained quantitative studies of mood disorders - many use just mean and standard deviation.
- Two fundamental models for bipolar disorder have been proposed: Daughtery *et al.* [3] and Goldbeter [4] have proposed oscillator models, but don't try to assimilate data.
- Gottschalk et al. [5] analysed daily mood records from 7 rapid cycling patients with bipolar disorder and 28 normal controls and examined the power spectrum and correlation dimension of time series.
- Weekly data has been collected at Oxford since 2006, and two papers have been published using this data. Bopp *et al.* [2] used the data for a longitudenal study and Bonsall *et al.* [1] applied time series models to distinguish two sets of patients.

Monitoring mood in bipolar disorder

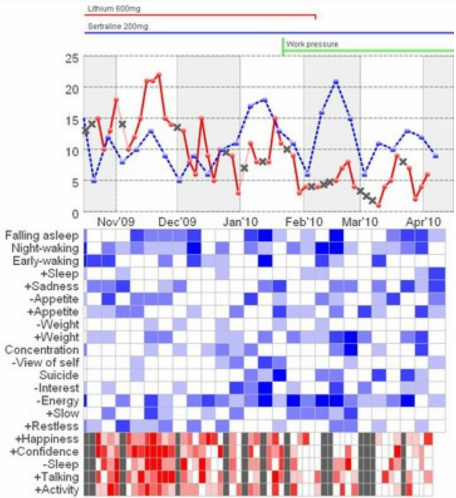
A mood monitoring system





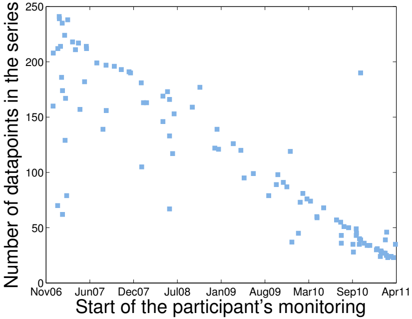
Since 2006, the Department of Psychiatry in Oxford has been monitoring mood in patients with bipolar disorder by using SMS text messages to return questionnaire results. Each week, participants return the questionnaire results as a sequence of integers.

Visualisation of mood



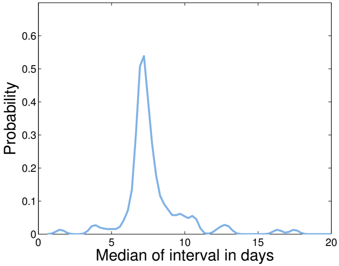
Visualisation used in clinic - depression is in blue and mania in red.

Time series lengths

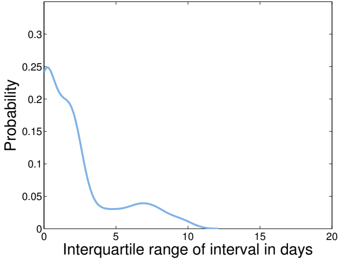


A scatter plot of the number of responses in the time series from each patient where the $x-$axis is the time when the patient enrolled in the monitoring system. Time series length is generally determined by how long the patient has been monitored: the outliers are those patients who return ratings more often than every week.

Response intervals



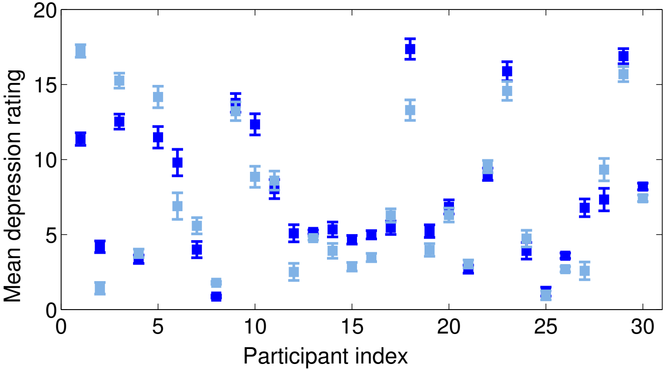
(a) Median interval



(b) Interquartile range of interval

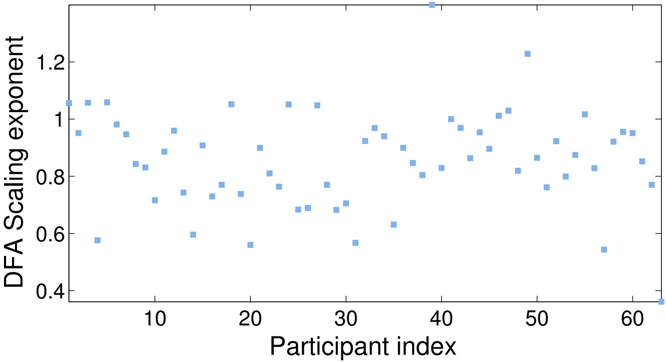
Statistics are for the time interval between responses from the patient. Interval statistics are shown over the set of 100 patient having 23 or more data points in their time series. Density estimates are computed using a Gaussian kernel smoother at 100 points.

Non Stationarity of time series



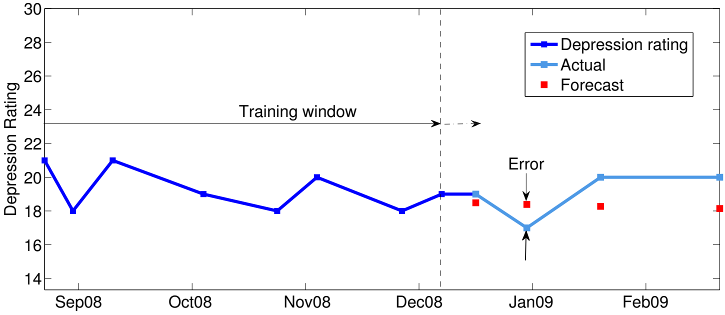
Dark blue and light blue markers represent the sample mean of the first and second halves of the depression time series for 20 patients. The error bars represent the standard error assuming that the numbers are Gaussian distributed and uncorrelated.

Roughness of time series



Detrended fluctuation analysis for those time series with length of at least 64 data points. The scaling exponent provides a measure of the roughness of the time series. It can be seen that the data set is heterogeneous.

Forecasting method

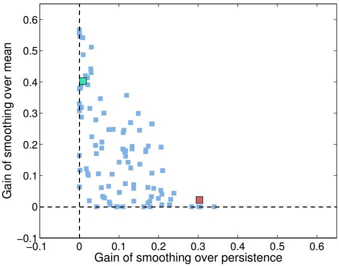


For each time series, a window is extended stepwise from a minimum offset to a maximum length. At each point a next step forecast is made. This process is repeated for all patients, and the root mean square error taken over all forecasts.

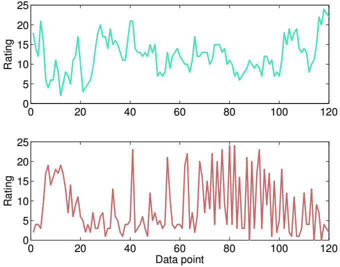
Exponential smoothing

- Exponential smoothing takes a previous forecast and adjusts it with a forecast error to give a next step forecast of $\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$, where α is a constant between zero and one and y_t is the time series response at time t.
- A value of zero for the smoothing parameter then uses the unconditional mean predictor while a value of one gives a persistence forecast.
- The smoothing parameter is found for each patient by searching for the minimum forecast error over the parameter range of zero to one. In this way we can gain some insight into the relative accuracy gains over these two baselines.

Results - insample



(c) Gain over baselines

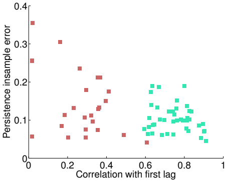


(d) Examples

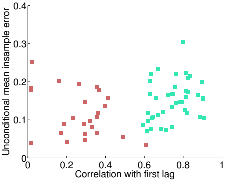
The left panel shows the gain of simple exponential smoothing compared with persistence (X-axis) and insample mean (Y-axis). Points in the top left region have a smoothing parameter close to 1, corresponding to a persistence forecast. Points in the bottom right region have a smoothing parameter close to 0, corresponding to a mean forecast.

First order correlations

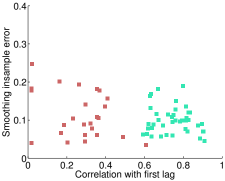
Plots of error against first order correlation for two groups of patients.



(e) Persistence



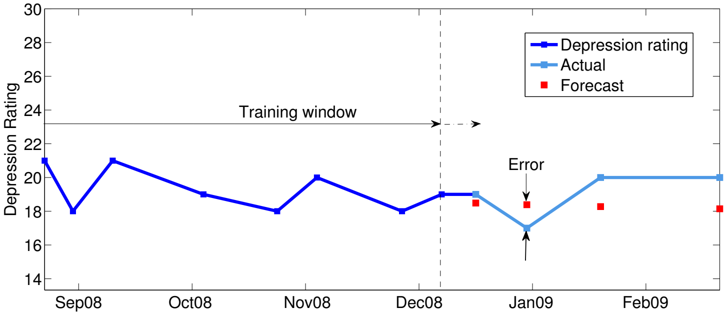
(f) Mean



(g) Smoothing

The red cluster are those time series where mean improves by 10% or more over persistence (25 patients) and the green cluster are those patients where persistence improves by 10% or more over mean (41 patients). Exponential smoothing adapts to the correlation structure of the individual time series.

Out of sample forecasting

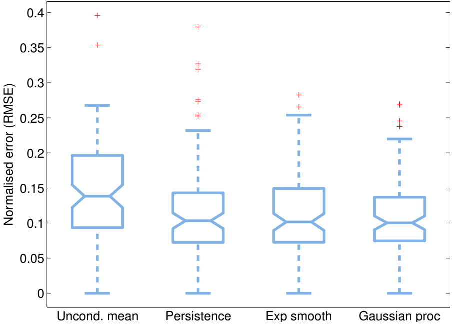


As for insample forecasting except that the model parameters are trained on the first 20 data points and retrained at intervals of 10 points.

Gaussian Process Regression (Kriging)

- For a Gaussian Process the joint distribution between any finite number of these variables is Gaussian. The process is specified by a mean function and a covariance function, which we choose as $k(r) = \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}$ where r is the difference between time indices, l is a length scale and α determines how the covariance changes with r.
- We maximize $p(\mathbf{y}|\mathbf{t}, l, \alpha)$ wrt l, α, σ_n
- The expected value and variance of the function can then be found for any input using the predictive distribution $\mathbb{E}[f_*] = \mathbf{k}_*^{\mathbf{T}}(K + \sigma_n^2 I)^{-1}\mathbf{y}$, where $\mathbf{k}_*$ is the vector of covariances of the test point with the training points, and K is the covariance matrix of the training set.

Results - out of sample



Box plots for out of sample results. The distributions shown are those of the errors of 100 patients, where the error for each patient is the root mean square of the next step forecasts for that patient. Points are drawn as outliers (marked as crosses) if they are larger than $q3 + 1.5 * (q3 - q1)$ or smaller than $q1 - 1.5(q3 - q1)$, where $q1$ and $q3$ are the 25^{th} and 75^{th} percentiles respectively.

Numeric out of sample results

Loss fn.	*Uncond.*	*Persist.*	*Exp. smooth.*	*Gauss. proc.*
RMSE	0.14	0.10	0.10	0.10
MAE	0.10	0.07	0.07	0.08
MdAE	0.07	0.06	0.06	0.07

Table: Out of sample next step forecasting results.

Median of the distribution of errors over the patient set. Each patient's error is estimated by applying one of three loss functions to the next step forecasts: RMSE, MAE and MdAE, and errors are normalised by the maximum of the rating scale (0-27).

Conclusions

- An important feature of this data set is its heterogeneity. Time series returned by the patients differ markedly in length, response interval, self-affinity and stationarity.
- Insample forecasting reveals that 40 of the 100 time series are forecast better by a constant predictor than by last value. The remaining 60 series that are better forecast by persistence are represented mostly by AR(1) processes, when assumed to be autoregressive models.
- Some time series cannot be forecast over the period of a week because there is too much noise or the sampling frequency is too low for the rate of mood change. Other patients exhibit little serial correlation, and so do not benefit from linear forecasting methods.
- In conclusion, we suggest that, if generalizations are to be made across the patient set, then sampling should be more frequent than on a weekly basis.

References

M.B Bonsall, S.M.A Wallace-Hadrill, J.R. Geddes, G.M. Goodwin, and E.A. Holmes.
Nonlinear time series approaches in characterizing mood stability and mood instability in bipolar disorder.
Proceedings of the Royal Society B: Biological Sciences, 2011.

M.J. Bopp, D.J. Miklowitz, G.M. Goodwin, W. Stevens, J.M Rendell, and J.R. Geddes.
The longitudinal course of bipolar disorder as revealed through weekly text messaging: a feasibility study.
Bipolar Disorders, **12**:327 to 334, 2010.

D. Daughtery, T. Roque-Urrea, J. Urrea-Roque, J. Snyder, S Wirkus, and M.A. Porter.
Mathematical models of bipolar disorder.
Commun Nonlinear Sci Numer Simulat, **14**:2897–2908, 2009.

Albert Goldbeter.
A model for the dynamics of bipolar disorders.
Progress in Biophysics and Molecular Biology, 2010.

A. Gottschalk, M.S. Bauer, and P.C. Whybrow.
Evidence of chaotic mood variation in bipolar disorder.
Arch Gen Psychiat, **52**:947 to 959, 1995.