

Analysis of Affective Instability on Ecological Momentary Assessments

Data: Successive Difference, Variance Decomposition, and Mean

Comparison via Multilevel Modeling

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

SEUNGMIN JAHNG

Dr. Phillip K. Wood, Thesis Supervisor

AUGUST 2007

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

ANALYSIS OF AFFECTIVE INSTABILITY ON ECOLOGICAL MOMENTARY
ASSESSMENTS DATA: SUCCESSIVE DIFFERENCE, VARIANCE
DECOMPOSITION, AND MEAN COMPARISON VIA MULTILEVEL MODELING

presented by Seungmin Jahng,

a candidate for the degree of master of art,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Phillip K. Wood

Professor Timothy J. Trull

Professor Glen Cameron

ACKNOWLEDGEMENTS

I would like to first thank Professor Phillip Wood for his advising me over the whole period of my study in University of Missouri - Columbia. Since the first year I came to the campus, he has supported me both academically and psychologically in every aspect of my life. I also would like to thank Professor Tim Trull for providing me an opportunity to work for the EMA project with him. I have learned a lot about the EMA study and characteristics of psychological disorders from his lab and the project, which encouraged me to write this thesis. Professor Glen Cameron gave me many precious comments and grateful supports to enhance the study and thesis writing. Marika Solhan, one of my lab mates, gave an effort to patient diagnosis and the collection of a special type of complicated data for the study. We had a lot of discussions about how to make the study better. I could have not completed my thesis successfully without any of them.

Returning to my life, it is my family that makes me come to my life from not-easy-going realities. I could not work this study through without the love and support my lovely wife, Miyoung, and my adorable daughters, Jen and Sarah, give to me. I also would like to thank mom and dad for their endless love and encouragement. Thank God for my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	iv
LIST OF TABLES	v
ABSTRACT.....	vi
Chapter	
1. INTRODUCTION	1
2. ISSUES IN ANALYSIS OF INSTABILITY	3
Measures of temporal instability based on successive change: MSSD and PAC	
Variability due to long-term time effects and short-term instability	
Mean comparison of temporal instability: two-stage analysis vs. Multilevel Modeling	
3. AN EXAMPLE ANALYSIS: NEGATIVE AFFECT VARIABILITY AND INSTABILITY IN BORDERLINE PERSONALITY DISORDER AND MAJOR DEPRESSIVE DISORDER PATIENTS.....	13
Method	
Result	
4. DISCUSSION	20
APPENDIX	
1. FIGURES	25
2. TABLES.....	27
REFERENCE:.....	32

LIST OF FIGURES

Figure	Page
1. Three simulated time series plots	25
2. Histograms and lognormal (solid line) and gamma (dashed line) plots of variance and MSSD for original and detrended series.....	26

LIST OF TABLES

Figure	Page
1. Variance decomposition of three simulated time series	27
2. Correlation between temporal instability indices and self-report affective instability measures.....	28
3. Estimates of group difference in temporal instability measure: two-stage approach	29
4. Estimates of fixed and random effects for SSD/SSDD and AC/ACD: generalized MLM.....	30
5. Estimates of group difference in variability due to time factors: two-stage approach.	31

ANALYSIS OF AFFECTIVE INSTABILITY ON ECOLOGICAL MOMENTARY
ASSESSMENTS DATA: SUCCESSIVE DIFFERENCE, VARIANCE
DECOMPOSITION, AND MEAN COMPARISON VIA MULTILEVEL MODELING

Seungmin Jahng

Dr. Phillip K. Wood, Thesis Supervisor

ABSTRACT

Temporal instability of affect is a defining characteristic of some psychological disorders such as Borderline Personality Disorder and mood cycling disorders. Use of Ecological Momentary Assessments (EMA) enables researchers to directly assess such frequent and extreme fluctuations over time. Two specific operationalizations of such temporal instability are proposed: Mean squared successive differences (*MSSD*) and probability of acute change (*PAC*). Additionally, residualizing scores by controlling time effects, such as long-term trend or diurnal effect, at the individual level is useful for identifying artifactual sources of temporal variability due to those time factors. Given that *MSSD* and *PAC* are individual differences measures, it is proposed that these measures be analyzed within generalized multilevel models. An illustrative example using EMA data on negative mood for borderline personality disorder (BPD) and major depressive disorder (MDD) groups is presented which shows that *MSSD* and *PAC* capture affective instability better than within-person variance, and that negative affect reports of the BPD group demonstrate more temporal instability than the MDD group. Versions of *MSSD* and *PAC* which adjust for the differently elapsed time between assessments are also discussed.

1. Introduction

Measurement and analysis of instability has been of great interest in psychological research, especially in the study of mood or affect. For example, mood or affective instability/variability has been characterized as a major symptom or contributing factor in many psychological disorders. Although early work in this area relied on retrospective self-report, direct real-time assessment of mood is possible using Ecological Momentary Assessment (EMA) and Experience Sampling Methods (ESM) (Larson and Csikszentmihalyi, 1983; Scollon, Kim-Prieto, and Diener, 2003; Hufford, Shiffman, Paty, and Stone, 2001). Within these studies, instability is equated with temporal instability and is commonly conceptualized as *frequent and extreme fluctuations over time*. Several statistical indices of temporal instability have been used: The within-person variance/standard deviation, autocorrelation, power spectral density, and mean squared successive difference (*MSSD*). Although each index taps some features of temporal instability, each of these approaches also has limitations. Larsen (1987) pointed out that a good measure of instability should incorporate both the extremity of response as well as the temporal dependency of measurements over time. In other words, a measure of instability should not only index the overall variability of a variable as evidenced by the general dispersion of an individual's scores over time, but also the frequency of fluctuations or successive change over time.

Of the indices listed above, the within-person variance or standard deviation has been most widely used (Eid & Diener, 1999; Penner, Shiffman, Paty, & Fritzsche, 1994; Larson, Csikzentmihalyi, & Graef, 1980; Gorman & Wessman, 1974; Hoffman & Meyer, 2006; Zeigler-Hill & Abraham, 2006; Stein 1996; Farmer, Nash, & Dance, 2004), because it captures the variability component of instability and is a familiar statistic to many researchers. Use of the within-person standard deviation or variance, however, does not take into account the frequency of fluctuations over time. For example, consider two people who are asked to rate their level of sadness on a 1-5 scale across 100 time points and appear to have an equal amount of instability as measured by the within person standard deviation. Data from the first individual could consist of a report of '4' for the first 50 occasions and '2' for the subsequent 50 occasions (e.g., 4, 4, ..., 4, 4, 2, 2, ..., 2, 2), showing only one shift in rating over time. By contrast, the second individual could produce the same standard deviation by alternating their '2' and '4' ratings from one occasion to the next (e.g., 2, 4, 2, 4, ..., 2, 4). Instability, as operationalized by fluctuation, is quite different for these two individuals: The first respondent has only one mood shift (from occasion 50 to occasion 51), but the other has 99 mood shifts!

2. Issues in analysis of temporal instability

Measures of temporal instability based on successive change: MSSD and PAC

Unlike within-person variance, measures based on successive change quantify temporal instability in terms of both variability and frequency of fluctuations over time. Mean Squared Successive Difference (*MSSD*), originally proposed by von Neumann, Kent, Bellinson, and Hart (1941), is one such measure. *MSSD* is, as its name implies, the average of squared difference between successive observations at occasions $i+1$ and i . The *MSSD* for a time series of n measurement occasions is given by

$$MSSD = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{n-1}$$

The *MSSD* was originally proposed as a measure of the dispersion of scores in nonstationary time series (i.e., time series in which systematic increase or decrease in overall level of response occur over time). This measure was proposed in recognition of the fact that the variance of a series which systematically increases or decreases over time will overestimate the actual dispersion of scores around the general trend line. The *MSSD* has the desirable property that one half the *MSSD* is an unbiased estimator of true variance (von Neumann et al., 1941; Harper, 1967) and, as such, it has a conceptual similarity to the individual-level variance measure just discussed.

Because of its use of adjacent temporal fluctuations, however, the *MSSD* incorporates notions of both variability and successive change.

MSSD is influenced by any change between two consecutive observations, whether big or small, and increasing or decreasing. However, researchers might be interested in a certain amount of change (i.e., a ‘big’ or ‘acute’ change), instead of overall amount of change, and/or change in a particular direction (e.g., a higher level versus lower level). For example, a researcher might be interested in big or acute change in negative mood for Borderline Personality Disorder (BPD) patients, where acute change is operationally defined as an increasing successive change greater than 3 on a 1 to 7 negative mood scale. In this case, the focus of negative mood instability of BPD patients is on the abrupt increase in negative mood from previous moment to next moment. The probability of acute change (*PAC*) is defined by the number of acute changes divided by the total number of changes, i.e.,

$$PAC = \frac{\sum_{i=1}^{n-1} AC_i}{n-1},$$

where $AC_i = 1$, if $x_{i+1} - x_i > k$, $k =$ predetermined cut point, and $AC_i = 0$, otherwise.

The cut-point, k , can be determined theoretically or statistically by researchers. *PAC* can be computed for each individual and used as a measure of instability due to short-term fluctuations. *PAC* complements the *MSSD* index. Although *MSSD* represents instability as an average of *all* (squared) successive change, *PAC* represents it as likelihood of *acute* successive change.

Variability due to long-term time effects and short-term instability

As other researchers pointed out, the *MSSD* (and also *PAC*) is sensitive to short-term, high-frequency fluctuations but not long-term, low-frequency fluctuations (Camm et al., 1996; Berntson, Lozano, & Chen, 2005). For example, consider three time series over 100 occasions: a randomly fluctuating time series without trend, the same random fluctuations with overall linear trend, and the same fluctuations with both linear and quadratic trends (Figure 1).

Table 1 shows that variances of the three series are markedly different; 3.64, 38.67, and 65.90, respectively. Additionally, the overall variance of the series can be partitioned into variance due to time factors (overall linear and quadratic trend) and variance due to random fluctuations (i.e., error variance for detrended series). Note that sum of the type I variances due to each time factor, calculated by dividing type I sum of squares by $n-1$, and residual variance, that is the variance of detrended series, is identical to the variance of original series. Note also that the variance of detrended series is identical across all three series because they have the same short-term fluctuations over and above time effects. Such variance decomposition at the individual level is easily accomplished using readily available software and allows the researcher to decompose overall individual-level variability into variance due to time effects and short-term random fluctuations. For the simulated series, it is clear that the three series have the same variability (variance) due to short-term fluctuations but different variability due to long-term linear and quadratic time effects, which results in the difference in overall variability.

As expected, the *MSSD* and *PAC* of either the original or detrended series produce general measures of short-term variability. Although *MSSDs* and *PACs* of the three series are not so much different to each other as the variances, they are also influenced by the time effects. Having the series detrended, the *MSSDs* and *PACs* of the residuals become identical. This indicates the equality, across the three series, of temporal instability due to short-term fluctuations. Therefore, the *MSSDD* (mean squared successive difference of detrended series) and *PACD* (probability of acute change of detrended series) can be better used as measures of temporal instability due to random short-term fluctuations, over and above long-term linear and quadratic time effects on the series.

In summary, the overall temporal variability/instability can be decomposed into instability due to long-term time effects and instability due to short-term fluctuations by trend analysis, and then the *MSSDD* and *PACD* can be used as measures of temporal instability due to short-term fluctuations.

Mean comparison of temporal instability: two-stage analysis vs. Multilevel Modeling

Although the discussion so far has focused on the operationalization of individual-level measures of instability, the focus of many studies is on whether group differences in instability exist. Researchers may, for example, be interested in knowing whether differences in mood instability exist between individuals with and without a certain diagnosis. Such questions could be addressed via a two-stage

approach: Indices of instability (e.g., within-person variance, *MSSD* or *PAC*) are calculated first for each individual, and then their group mean differences are tested using general linear models. When conducting group mean difference tests in variance, *MSSD*, and *PAC*, one should be aware that the variance, *MSSD*, and *PAC* are not normally distributed in general. Because variance and *MSSD* are average of squared values, they are always non-negative and usually positively skewed in distribution. On the other hand, given that the number of acute changes follows binomial distribution with parameter p , the probability of acute change, and $n-1$, the total number of changes, *PAC* has mean of p and variance of $p(1-p)/(n-1)$. Because *PAC* is dependent on p , the homogeneity of variance assumption in ANOVA and t-test is violated. Additionally, the distribution of a probability, ranging from 0 to 1, is usually not normally distributed, especially when $p < .3$ or $p > .7$ with small n .

Therefore, traditional t-test or ANOVA approach in which normal error distribution and homogeneity of variance are assumed may not be used to test group mean differences in these instability indices. A possible solution is the use of appropriate transformation. Log transformation of variance and *MSSD* may produce better normal-shaped distribution of these variables. For *PAC*, arcsine, or inverse sine, transformation can be used to address the assumption violation of both homogeneity of variance and normal distribution (there is, however, a better approach to test mean difference in *PAC*. See below).

If an instability measure follows an exponential family of distribution (e.g., normal, gamma, or binomial distribution), a generalized linear model is available to

test group mean difference. For a normally distributed variable, the sample variance is well known to have a scaled chi-square distribution:

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi_{n-1}^2$$

The distribution of *MSSD* is also known to have a good chi-square approximation (Harper, 1967). Because the chi-square distribution is a special case of gamma distribution, generalized linear models with gamma distribution can be used to test a group mean difference of variance and *MSSD*. On the other hand, group mean difference in *PAC* can be tested using logistic regression, because the number of acute change always follows a binomial distribution. In this case, the *PAC* value, ranging continuously from 0 to 1, can not be directly used in the model. Instead, the number of acute change and the total number of changes for each individual are used to estimate parameters of the distribution of *PAC*. All the procedure of two-stage analysis discussed above are also applicable to the variance, *MSSD* and *PAC* for detrended series (i.e., variance of residuals from detrending, *MSSDD* and *PACD*).

Like other two-stage approaches, the above two-stage approach, however, is not without problems: First, the error of estimation of parameters for each individual (e.g., *MSSD* or *PAC*) is not taken into account in the two-stage approach. In the first stage, parameters for each individual are estimated based on observations at individual level. Keep in mind that those estimates have their own error of estimation. In the second stage, we use those estimates as observations in the upper-level modeling and test group mean difference only against the standard error of estimation of the group mean difference (of the estimates), ignoring the fact that we are using estimates rather than the “real” data. If the second stage estimation is conducted

without considering error of estimation at the first stage, the standard error of estimation of group mean difference will be downwardly biased, the test statistic will be too big, and thus probability of making a type I error will be inflated.

Second, a two-stage approach does not take into account the different number of observations across individuals, a situation which often occurs in EMA data. If individuals vary in terms of the number of measurement occasions, the individual variance, *MSSD* or *PAC* should differentially contribute to the group mean.

Moreover, an index of instability is not identically distributed across people, because the number of observations, which is not uniform across individuals, is a parameter of

the distribution n : χ_{n-1}^2 or $Gamma(\frac{n-1}{2}, \beta)$ for variance and *MSSD* or

$Bin(n-1, p)$ for *PAC*. Individuals who have different number of observations are assumed to be sampled from different distributions. The assumption of identical distribution is a fundamental basis in general linear models and most of other statistical tests as well.

Multilevel modeling can address the problems described above. Time series or EMA data that we have discussed so far has a two-level structure; observations are nested within individuals. Each observation is a sample from a random distribution of observations for each individual, and each individual mean is also a sample from a

Second, a two-stage approach does not take into account the different number of observations across individuals, a situation which often occurs in EMA data. If individuals vary in terms of the number of measurement occasions, the individual variance, *MSSD* or *PAC* should differentially contribute to the group mean.

Moreover, an index of instability is not identically distributed across people, because

the number of observations, which is not uniform across individuals, is a parameter of the distribution n : χ_{n-1}^2 or $Gamma(\frac{n-1}{2}, \beta)$ for variance and $MSSD$ or $Bin(n-1, p)$ for PAC . Individuals who have different number of observations are assumed to be sampled from different distributions. The assumption of identical distribution is a fundamental basis in general linear models and most of other statistical tests as well.

Multilevel modeling can address the problems described above. Time series or EMA data that we have discussed so far has a two-level structure; observations are nested within individuals. Each observation is a sample from a random distribution of observations for each individual, and each individual mean is also a sample from a random distribution of individual means for each group. In case of $MSSD$, each squared successive difference is the lower level random variable and individual $MSSD$ is the upper level random realization. In case of PAC , an event of acute change is the lower level random observation and individual PAC is the upper level random variable. In the MLM approach, both lower-level error variance and upper-level error variance are modeled simultaneously, and thus more precise test statistic and correct type I error rate can be produced. In addition, MLM takes the different number of observations at the lower-level into account in estimation and test for the upper-level model. Individuals who have more observations contribute more in estimation of upper-level parameters (e.g., group mean) than individuals who have less observations.

Common Multilevel Modeling assumes normal error distribution at the first level. However, squared successive difference and (event of) acute change are not

normally distributed. Just as the general linear model can be extended to the generalized linear model, MLM can also be extended to the generalized MLM for a non-normal, exponential family of error distribution. For squared successive difference, MLM with gamma error distribution and log link is appropriate because squared successive difference is often distributed as gamma¹. For acute change, logistic MLM can be used because acute change is distributed as binomial. For squared successive difference, $SSD_{(i+1)j} = (x_{(i+1)j} - x_{ij})^2$, where x_{ij} is the i th observation for j th individual, a generalized MLM is given by

$$SSD_{ij} \sim \text{Gamma}(\alpha_j, \beta_j), E(SSD_{ij}) = \alpha_j \beta_j = \mu_j, \text{Var}(SSD_{ij}) = \alpha_j \beta_j^2$$

$$\text{Level-1 Link function: } \eta_j = \log(\mu_j)$$

$$\text{Level-1 structural model: } \eta_j = b_{0j}$$

$$\text{Level-2 model: } b_{0j} = \gamma_{00} + \gamma_{01} \text{Group}_j + u_{0j}, \quad u_{0j} \sim N(0, \tau^2),$$

where γ_{01} is the (log transformed) group difference in overall mean of SSD , i.e., the weighted mean of (log transformed) individual $MSSD$, μ_j . For acute change, $AC_{(i+1)j} = 1$, if $x_{(i+1)j} - x_{ij} > k$, $k =$ predetermined cut point, and $AC_{(i+1)j} = 0$, otherwise, where x_{ij} is the i th observation for j th individual, a logistic MLM is given by

$$AC_{ij} \sim \text{Binomial}(1, p_j), E(AC_{ij}) = p_j = \mu_j, \text{Var}(AC_{ij}) = p_j(1 - p_j)$$

$$\text{Level-1 Link function: } \eta_j = \log\left(\frac{\mu_j}{1 - \mu_j}\right)$$

¹ The distribution of squared successive difference depends on the distribution of successive difference. If successive differences are normally distributed, the squared successive difference follows a scaled chi-square distribution with one degree of freedom, a special case of gamma distribution. The distribution of successive difference is usually symmetric, especially for the detrended or stationary time series. In most cases, the square of a variable sampled from a symmetric distribution follows a gamma (or an inverse gamma) distribution.

Level-1 structural model: $\eta_j = b_{0j}$

Level-2 model: $b_{0j} = \gamma_{00} + \gamma_{01}Group_j + u_{0j}, \quad u_{0j} \sim N(0, \tau^2),$

where γ_{01} is the (logit transformed) group difference in overall probability of acute change, i.e., the weighted mean of (logit transformed) individual *PAC*, μ_j . The same models are applicable to the detrended data; the generalized MLM with log link and gamma distribution for squared successive difference of detrended series,

$SSDD_{(i+1)j} = (r_{(i+1)j} - r_{ij})^2$, where r_{ij} is the residual of the i th observation for j th

individual, and the logistic MLM for acute change of detrended series, $ACD_{(i+1)j} = 1$, if $r_{(i+1)j} - r_{ij} > k$, $k =$ predetermined cut point, and $ACD_{(i+1)j} = 0$, otherwise, where r_{ij} is the residual of the i th observation for j th individual.

In summary, *MSSD* and *PAC* have better properties than the within-person variance as indices of temporal instability, in that they represent both frequency and extremity of fluctuations over time. Additionally, these indices of detrended time series purely quantify temporal instability due to short term fluctuations, while the variability due to long term trends can be captured in the variance accounted for by time factors from the trend analysis. Although the two-stage analysis can be used to test group difference of the temporal instability indices, the generalized multilevel modeling provides more reliable significance test. The following section will present a sample analysis of negative affective variability and instability on ecological momentary assessments data for people with borderline personality disorder or major depressive disorder.

3. An example analysis: negative affect variability and instability in borderline personality disorder (BPD) and major depressive disorder (MDD) patients

Method

Data presented here are from an ongoing study of affective instability in BPD. Participants were recruited from one of four community mental health outpatient clinics through flyers. Potential participants were screened through chart review and final eligibility for the study was determined through administration of DSM-IV-TR (APA, 2000) Axis I and Axis II structured interviews. Two groups of outpatients were entered into the study, following determination of eligibility into one of two patient groups. The data presented in this article are based on 60 participants: 34 who met DSM-IV-TR (APA, 2000) diagnostic criteria for BPD and who endorsed the diagnostic feature of affective instability; and 26 who met DSM-IV-TR diagnostic criteria for current major depressive disorder or for current dysthymic disorder (MDD/DYS) and did not report affective instability.

All participants were assigned an electronic diary programmed to prompt them to rate a series of mood descriptors from the Positive and Negative Affect Scales-Extended version (PANAS-X; Watson & Clark, 1994). Each mood item was rated on a 1-5 scale, where 1=not at all and 5=extremely. Each participant was asked to rate his or her mood as it had been since the last assessment completed on the electronic diary (i.e., the last prompted response). Here we discuss only the negative

affect composite scores based on 21 items for illustrative purposes. Items come from PANAS 10-item negative affect scale, supplemented by additional items tapping fear, hostility, and sadness. The mean of the 21 items is used as the negative affect score. The study design called for 6 assessments per each day over approximately 4 weeks of consecutive days. However, as expected, the number of days of assessments per each person and the number of assessments per each day differed (16 to 41 days, Median (Mdn) = 29, Interquartile Range (IQR) = 1, per each person; 1 to 9 assessments, Mdn = 6, IQR = 1, per each day). In total, 83 to 193 assessments (Mdn = 156, IQR = 9.5) per each person were conducted. Assessment times within a day varied randomly across people. Additionally, participants were asked to fill out retrospective self-report measures, including *Personality Assessment Inventory-Borderline Features Scale* (PAI-BOR; Morey, 1991) and *Affective Lability Scale* (ALS; Harvey, Greenberg & Serper, 1989) before and after the electronic diary assessments session. ALS and the affective instability subscale in PAI-BOR are measures of affective instability.

Because the calculation of successive difference assumes uniform time interval across all occasions and assessments in the current study were prompted randomly within days, the daily mean values of the diary data were used for the following analysis. Here the main research question is whether BPD patients experienced more instability in negative affect than did MDD/DYS patients across days.

Results

Correlation between temporal instability indices and self-report instability measures

To quantify temporal instability, within-person variance, *MSSD*, and *PAC* were calculated for each individual. Additionally, the same indices were calculated for residuals from trend analysis to produce temporal instability due to short term fluctuations. For detrending, regression analysis with three time factors was conducted for each individual time series. The three time factors are overall linear and quadratic trends across days and weekend effect. To minimize collinearity between linear and quadratic trends across days, ‘days’ was centered at the mean of each individual series. Consequently, the six temporal instability indices were generated; within-person variance, within-person variance of residuals, *MSSD*, *MSSDD*, *PAC*, and *PACD*. For the cut point of acute change (and acute change of detrended series), the 90th percentile of total successive difference (and successive difference of residuals) was used. The cut points were .40 and .41 for acute change and acute change of residuals, respectively.

To investigate how each temporal instability index relates to retrospective self-report affective instability measures, correlations between temporal instability indices and PAI-BOR affective instability subscale (AI) and ALS were analyzed. Before conducting correlation analysis, each temporal instability index was transformed to approximate normal distribution and then Pearson correlation coefficients were calculated with the self-report affective instability measures. Variance, residual variance, *MSSD*, and *MSSDD*, was transformed by log (e.g., $MSSD' = \log(MSSD)$). For *PAC* and *PACD*, arcsine transformation was used (e.g., $PAC' = \sin^{-1}(\sqrt{PAC})$).

Table 2 shows the result of the correlation analysis. As expected, the correlations of *MSSD* and *PAC* with self-report affective instability measures are higher than that of the within-person variance. All six temporal instability indices are significantly correlated with pre- and post-measures of AI. However, the successive-difference-based indices (i.e., *MSSD* and *PAC*) showed higher correlation than the index of dispersion (i.e., within-person variance); *MSSD* ($r = .383, p < .01$) and *PAC* ($r = .342, p < .01$) displayed higher correlation with pre-measured AI than variance ($r = .303, p < .05$), and *MSSDD* ($r = .378, p < .01$) and *PACD* ($r = .371, p < .01$) were more highly correlated with pre-measured AI than variance of detrended residuals ($r = .327, p < .05$). The same pattern was confirmed with post-measured AI (see table 2). More impressively, only *PAC* and *PACD* are significantly correlated with pre-measure of ALS ($r = .277$ and $r = .327$, respectively, $p < .05$) and only *PACD* has a significant correlation with post-measure of ALS ($r = .276, p < .05$). Although the temporal instability indices are not supposed to be successfully validated through the investigation of the correlation with retrospective self-report instability measures, this result suggests that successive-difference-based instability indices quantify instability better than the within-person variance.

The comparison between (correlation of) instability indices of original series and those of detrended series showed that variance and *PAC* for detrended series better detected instability than those for original series (e.g., $r_{(PAC, \text{pre-ALS})} = .277$ and $r_{(PACD, \text{pre-ALS})} = .327$). The correlations of *MSSD* with self-report measures of instability did not differ too much from those of *MSSDD*. Nevertheless, the above

results support that detrending makes the indices of temporal instability more valid and apparent.

Group difference in temporal instability: two-stage approach and multilevel modeling

To compare mean (temporal) affective instability, mean differences in within-variance, *MSSD*, *PAC* and those for detrended series were tested. To decide whether t-test of transformed variables or generalized linear model is more appropriate, the distributions of variance and *MSSD* (for both original and detrended series) were investigated. Figure 2 shows histograms (of variance and *MSSD*) and log-normal and gamma plots fitted to each histogram. Variances of original and detrended series are distributed as more log normal than gamma; Goodness of fit tests of distribution (Kolmogorov-Smirnov test) showed that the distribution of variances of original and detrended series for both the BPD and the MDD/DYS groups have good fit to lognormal distribution (all $ps > .15$). The distribution of *MSSD* and *MSSDD* for both the BPD and the MDD/DYS groups showed better goodness of fit to gamma distribution than lognormal distribution (all $ps > .25$ for gamma and all $ps > .15$ for lognormal). Accordingly, t-test was used for variance of original and detrended series and the generalized linear model with gamma error and log link was used for *MSSD* and *MSSDD*. For *PAC* and *PACD*, logistic regression was used to test mean difference between the two groups.

Table 3 presents two-stage analysis results for the group mean difference of temporal instability measures. Back-transformed parameter estimates are added for interpretation. For logistic regression model, estimates of odds are added instead of

fully back-transformed parameter estimates. Notice the mean difference in (log or logit) transformed or linked variables is expressed as ratio of the two group means (or odds ratio) in back-transformed values.

The BPD group showed significantly greater temporal affective instability than the MDD/DYS group in variance of detrended series ($t_{58} = -2.14, p < .05$), *MSSD* ($\chi^2_{(1)} = 6.64, p < .01$), *MSSDD* ($\chi^2_{(1)} = 6.57, p < .05$), and *PACD* ($\chi^2_{(1)} = 7.70, p < .01$).

Although two-stage approach provided statistical evidence that BPD people has greater affective instability than MDD/DYS people, the accuracy of the tests are open to debate. To provide further evidence, group difference of affective instability was tested using generalized MLM. The GLIMMIX Macro in SAS system was used. For squared successive difference of original and the detrended series (SSD/SSDD), MLM with log link and gamma error distribution was performed. Results showed that the BPD group produced a greater mean in SSD than did the MDD/DYS group ($e^{-2.07} = .126$ and $e^{-2.69} = .068$, respectively; $t_{58} = -2.04, p < .05$). This was also true when the time effects are detrended ($e^{-2.11} = .121$ and $e^{-2.77} = .063$, respectively; $t_{58} = -2.16, p < .05$). For acute change of original and the detrended series (AC/ACD), logistic MLM, (i.e., MLM with logit link and binomial error distribution) was performed. No group differences were found in probability of acute change of either original or detrended series. The results of group differences in successive difference measures are presented in Table 4.

Group difference in variability due to time effects

The variability due to time factors (linear, quadratic and weekend effect) was compared between the BPD and the MDD/DYS groups. Type III variance due to each time factor was calculated for each individual from the regression analysis described in the result of correlation analysis. The choice of type III variance instead of type I was made by the fact that type I sum of squares (and thus type I variance) is affected by the order of factors put in a model (because it is calculated over and above only the factors which are formerly put in the model), while type III sum of squares is not (because it is calculated over and above all the factors in the model).

Table 5 shows group means of variance due to each time factor. Mean difference was tested using generalized linear model with gamma error distribution and log link. There was significant group difference in variance due to weekend effect ($\chi^2_{(1)} = 5.39, p < .05$): BPD group showed larger variance due to weekend effect (mean = .024) than MDD/DYS group (mean = .016). Variability due to overall linear and quadratic trend was not significantly different between the two groups.

In summary, BPD group showed more overall variability of negative affect than MDD/DYS group, because they have more instability due to short-term fluctuations and more variability due to weekend effect than MDD/DYS group.

4. Discussion

Although within-person variance is the most-widely used measure of variability/instability over time, *MSSD* is a more conceptually appealing measure of temporal instability. A variance estimate does not differentiate variability due to short-term fluctuations from variability due to long-term change. More importantly, variance only represents dispersion of states, without considering temporal order. On the other hand, because *MSSD* captures only short term fluctuation, longer-period effects in affect may be overlooked using this measure alone. Along with regressions of such overall trend and periodic effects, the *MSSD* can be used to determine whether the overall variability is caused by short-term fluctuations or other long-term trends. Identifying the source of temporal variability is important so that factors influencing instability/variability can be explored. If overall variability is heavily influenced by time factors, the unknown factors of the variation of states are also probably time-dependent. If variability is more influenced by short-term fluctuations, state is possibly influenced by short-term fluctuating factors.

Both within-person variance and *MSSD* are known as time domain measures of temporal variability in contrast with its frequency domain measures, mainly obtained by power spectral density (PSD) analysis (Camm et al., 1996; Berntson, Lozano, & Chen, 2005). PSD analysis, which produces a set of spectral density functions, is an analytic technique to decompose a time series with cyclical components into several underlying sinusoidal (sine and cosine) functions (see

Kendall, 1976 or Bloomfield, 1976 for more details). As a prism decomposes a beam of sun light into different wave lengths or cyclical components that make up white sun light, a successful spectral analysis might uncover a few recurring cycles of different lengths in the time series of interest. It is more mathematically complicated but fascinating than time domain measures. In heart rate variability studies, in which temporal variability is mainly investigated, spectral analysis has been the most popular statistical model for measurement of variability.

Nevertheless, application of spectral analysis to the assessment of affective instability has difficulties in providing an overall index of affective instability that reflects both amplitude and frequency of fluctuations. Researchers, who use spectral analysis to measure mood instability, often investigate total power spectral density, which is essentially identical to a within-person variance that does not consider temporal order of measurements, or the proportion or percentage of arbitrarily-divided power bins which does not consider the degree of amplitude of fluctuations (Larsen, 1987; Woyshville, et al., 1999; Gottschalk, Bauer, Whybrow, 1995; see Ebner-Priemer, et al., 2006). More importantly, an appropriate theoretical base of sinusoidal cyclicity of a time series is, at least, required to apply the spectral analysis. In a study with bipolar disorder patients, for example, apparent cyclicity was not found even though theory of this disorder supports cyclical mood fluctuations (Gottschalk, Bauer, Whybrow, 1995).

The decomposition of overall variability into long- and short-term variability/instability, by trend analysis and *MSSD*, and the decomposition of total power spectral density, by PSD analysis, provide similar information. But the former

decomposition provides simpler and more interpretable information than the latter. Camm et al. (1996) pointed out ‘... , unless special investigations are performed ... , the results of the frequency-domain analysis are equivalent to those of the time domain analysis, which is easier to perform.’ However, studies comparing results from the two different approaches of variance decomposition are needed.

Other measures of temporal instability which are not considered here have been proposed. Autocorrelation, for example, measures correlational dependency of order k between each i_{th} element and $(i-k)_{\text{th}}$ element of time series, i.e., sequences of measurements taken at equally spaced time intervals. A k_{th} order autocorrelation provides an index of how well scores at time t_i correlate with scores at t_{i-k} , i.e., the magnitude of the *stability* of states between two measurement points (Stein, 1996). Although autocorrelation is a good measure of temporal dependency, it does not reflect extremity or degree of amplitude of fluctuations. In other words, if two time series have the same pattern of change trajectory with different amplitude, they will have the same autocorrelation coefficient (Ebner-Priemer, et al., 2006). For this reason, autocorrelation is not a good measure of instability.

Although *MSSD* and *PAC* have good properties as indices of temporal instability, they are limited to assessments with fixed uniform time interval. Measures of temporal instability using successive change assume that time intervals between consecutive observations are uniform (i.e., $t_i - t_{i-1}$ is same for all i 's, where t_i is time at occasion i). If this condition is met, the successive difference has same meaning across all occasions. In some studies, however, observations are irregularly spaced over time (e.g., at randomly selected times within a day or as occurs when gaps in the

series occur when a participant is unable to respond due to sleeping at night or not being able to come in during the weekend). In those cases, successive differences have a different interpretation, depending on the time interval involved. An amount of change over an hour implies greater or more frequent change than the same amount of change over nine hours, for example. One way of adjusting the successive difference score for the amount of elapsed time is to divide the successive difference by

$\sqrt{(t_i - t_{i-1}) / \text{median}(t_i - t_{i-1})}$. Standardized successive difference (SD) with time interval, $(t_i - t_{i-1})$, is, therefore:

$$\text{Standardized } SD_i = \frac{(x_i - x_{i-1})}{\sqrt{(t_i - t_{i-1}) / \text{median}(t_i - t_{i-1})}}.$$

By dividing a time interval by its median in the denominator, the time intervals are rescaled with median of 1. The standardized successive difference has the same value as the unstandardized one when the time interval is equal to the median of the total time interval. When the time interval is longer than the median, the standardized difference becomes smaller than the unstandardized one. In the same fashion, the standardized difference becomes larger than the raw difference when time interval is shorter than the median. The choice of median instead of mean is because median is robust to a skewed distribution, commonly found in randomly spaced time intervals of EMA data. The square root is used to attenuate the effect of the difference in time interval. If an occasion has a successive difference score with an interval and another occasion has the same difference score with the nine times longer interval than the first one, standardized difference of the first is three times greater than the second, by means of square root. If attenuation is not made, the ratio

of the two standardized successive differences is nine to one, which seems to produce over-standardization. Although we propose the square root for attenuation, researchers may use cubic or quartic root for more attenuation of over-standardization. Notice that if the time interval is same across all occasions, standardization has no effect because the denominator of standardized successive difference is 1 across all occasions. When this standardization is applied to the data used in this study where time interval within days were randomly spaced, similar results were obtained for the standardized version of *MSSD* and *PAC*. However, this standardization procedure has not been used previously in EMA studies and needs more investigation to be used.

In this article, we described why *MSSD* and *PAC* are good measures of temporal instability and how they can be used along with trend analysis to locate sources of temporal variability. Moreover, use of multilevel modeling application for group mean comparison of *MSSD* and *PAC* was introduced in order to take error of estimation in individual *MSSD* and *PAC* and unbalanced design into account.

The importance of measurement and analysis of temporal instability increases as interest in the cause and effect of affective instability and use of EMA data increase. I hope that the suggested procedure of measuring and analyzing temporal instability will inspire more investigation of affective instability using EMA data.

APPENDIX 1: FIGURES

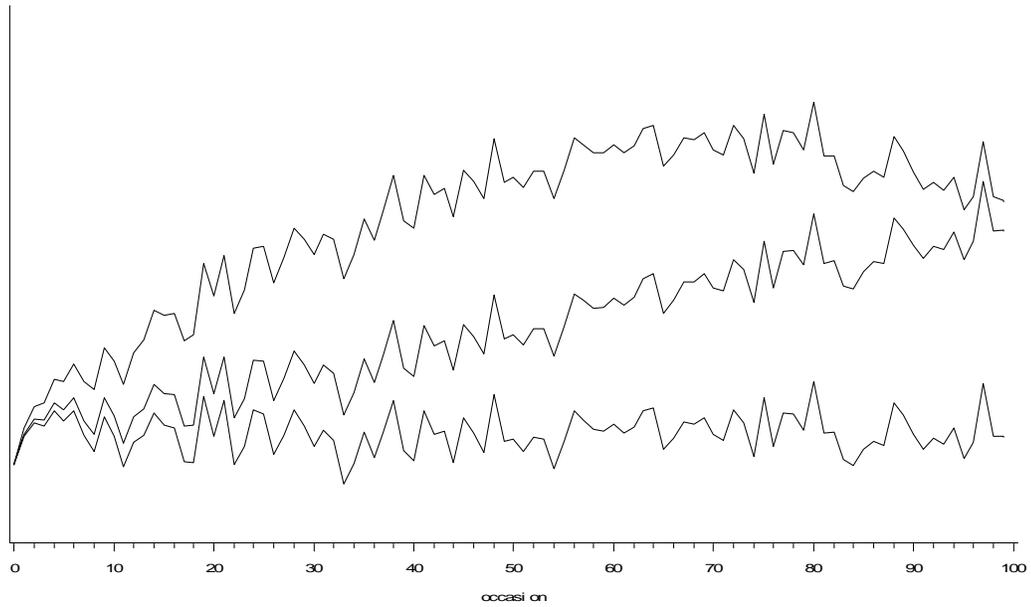


Figure 1. Three simulated time series plots. The lower series are random fluctuations with no trend. The middle series are the same random fluctuation with linear trend. The upper series are the same random fluctuations with linear and quadratic trends. See also Table 1.

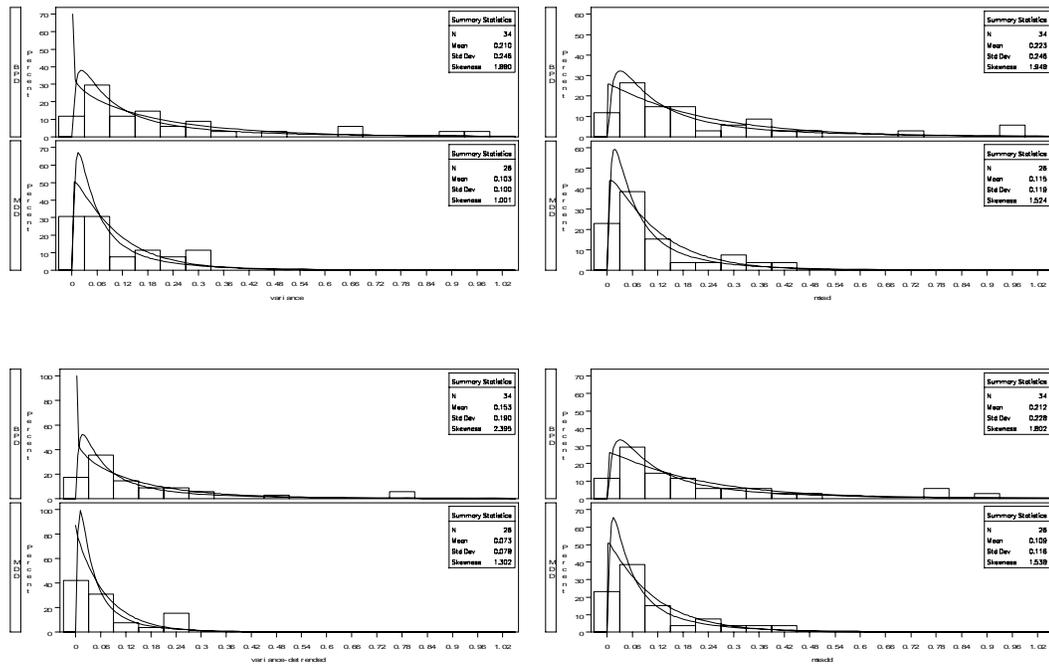


Figure 2. Histograms and lognormal (solid line) and gamma (dashed line) plots of variance and MSSD for original and detrended series.

APPENDIX 2: TABLES

Table 1. Variance decomposition of three simulated time series

Series	Original series		Detrended series		Variance due to time factor (Type I)			
	Variance	MSSD	PAC	Variance	MSSD	PAC	Linear	Quadratic
							trend	trend
No trend	3.638	6.972	.101	3.623	6.972	.101	.014	.001
Linear	38.671	7.023	.091	3.623	6.972	.101	35.046	.001
Linear + Quadratic	65.897	7.175	.111	3.623	6.972	.101	45.238	17.035

Table 2. Correlation between temporal instability indices and self-report affective instability measures

Instability index		pre-measure		post-measure	
		AI	ALS	AI	ALS
Original	Variance	.303*	.152	.436**	.202
series	MSSD	.383**	.167	.531**	.176
	PAC	.342**	.277*	.523**	.255
Detrended	Variance	.327*	.182	.491**	.242
series	MSSD	.378**	.166	.530**	.165
	PAC	.371**	.327*	.524**	.276*

AI – affective instability subscale of PAI-BOR. ALS – Affective

Lability Scale. * p < .05, ** p < .01

Table 3. Estimates of group difference in temporal instability measure:

two-stage approach

		Estimate			
Variable	transformation/link	BPD	MDD	Difference/Ratio	Standard error
Variance	log	-2.21	-2.81	-0.60	0.31
	original value	0.11	0.06	0.55	
Variance (detrended)	log	-2.53	-3.20	-0.67*	0.32
	original value	0.08	0.04	0.51	
MSSD	log	-1.50	-2.16	-0.66**	0.26
	original value	0.22	0.12	0.52	
MSSDD	log	-1.55	-2.21	-0.66*	0.26
	original value	0.21	0.11	0.52	
PAC	logit	-2.12	-2.42	-0.30	0.17
	odds	0.12	0.09	0.74	
PACD	logit	-2.06	-2.54	-0.48**	0.17
	odds	0.13	0.08	0.62	

* $p < .05$, ** $p < .01$

Table 4. Estimates of fixed and random effects for
SSD/SSDD and AC/ACD: generalized MLM

Parameter	Original series		Detrended series	
	Estimate	SE	Estimate	SE
	SSD		SSDD	
Intercept	-2.07**	0.20	-2.11**	0.20
Group difference (BPD – MDD/DYS)	-0.62*	0.30	-0.66*	0.31
Level 2 variance	1.19		1.24	
Level 1 variance	4.13		3.91	
	AC		ACD	
Intercept	-2.25**	0.18	-2.17**	0.19
Group difference (BPD – MDD/DYS)	-0.32	0.28	-0.60	0.31
Level 2 variance	0.76		0.92	
Level 1 variance	0.81		0.78	

* p < .05; ** p < .0001.

Table 5. Estimates of group difference in variability due to time factors:

two-stage approach

		Estimate		
Variable	transformation/link BPD	MDD	Difference/Ratio	Standard error
Days -	log	-3.747-4.150	-0.403	0.426
Linear	original value	0.024 0.016	0.668	
Days-	log	-4.386-4.780	-0.394	0.320
Quadratic	original value	0.012 0.008	0.674	
Weekend	log	-4.084-5.118	-1.035*	0.446
	original value	0.017 0.006	0.355	

* p < .05

REFERENCE

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Press.
- Berntson, G. G., Lozano, D. L., & Chen, Y. (2005). Filter properties of root mean square successive difference (RMSSD) for heart rate. *Psychophysiology*, *42*, 246–252.
- Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- Camm, A. J., Malik, M., Bigger, J. T., Breithardt, G., Cerutti, S., Cohen, R. J., Coumel, P., Fallen, E. L., Kennedy, H. L., Kleiger, R. E., Lombardi, F., Malliani, A., Moss, A. J., Rottman, J. N., Schmidt, G., Schwartz, P. J., & Singer, D. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, *93*, 1043-1065.
- Ebner-Priemer, U. W., Kuo, J., Kleindienst, N., Welch, S. S., Reisch, T., Reinhard, I., Lieb, K., Linehan, M. M., & Bohus, M. (2006). State affective instability in borderline personality disorder assessed by ambulatory monitoring. *Psychological Medicine*, *37*, 1-10.
- Eid, M. & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*, 662-676.

- Farmer, R. F., Nash, H. M. & Dance, D. (2004). Mood patterns and variations associated with personality disorder pathology. *Comprehensive Psychiatry*, 45, 289–303.
- Gorman, B. S., & Wessman, A. E. (1974). The relationship of cognitive styles and moods. *Journal of Clinical Psychology*, 30, 18-25.
- Gottschalk, A., Bauer, M. S., Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, 52, 947–959.
- Harper, W. M. (1967). The distribution of the mean half-square successive difference. *Biometrika*, 54, 419-434.
- Harvey, P. D., Greenberg, B. R., & Serper, M. R. (1989). The Affective Lability Scales: Development, reliability and validity. *Journal of Clinical Psychology*, 45, 786-793.
- Hoffman, B. U., & Meyer, T. M. (2006). Mood fluctuations in people putatively at risk for bipolar disorders. *British Journal of Clinical Psychology*, 45, 105–110.
- Hufford, M. R., Shiffman, S., Paty, J., & Stone, A. A. (2001). Ecological momentary assessment: real-world, realtime measurement of patient experience. In J. Fahrenberg, & M. Myrtek (Eds.), *Progress in ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 69-92). Kirkland, WA: Hogrefe & Huber.
- Kendall, M. (1976). *Time series*. London, England: Charles Griffin.

- Larsen, R. (1987). The stability of mood variability: a spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology*, *52*, 1195–1204.
- Larson, R. W., Csikzentmihalyi, M., & Graef, R. (1980). Mood variability and the psychosocial adjustment of adolescents. *Journal of Youth and Adolescence*, *9*, 469-489.
- Larson, R., & Csikszentmihalyi, M. (1983). The Experience sampling method. *New Directions for Methodology of Social & Behavioral Science*, *15*, 41-56.
- Morey, L. C. (1991). *Personality assessment inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Penner, L. A., Shiffman, S., Paty, J. A., and Fritzsche, B. A. (1994). Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology*, *66*, 712-721.
- Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*, 5-34.
- Stein, K. F. (1996). Affect instability in adults with a borderline personality disorder. *Archives of Psychiatric Nursing*, *10*, 32–40.
- von Neumann, J., Kent, R. H., Bellinson, H. R., Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, *12*, 153–162.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. Retrieved from University of Iowa, Department of Psychology Web site:
<http://www.psychology.uiowa.edu/Faculty/Watson/Watson.html>.

Woyshville, M. J., Lackamp, J. M., Eisengart, J. A. & Gilliland, J. A. (1999). On the meaning and measurement of affective instability: Clues from chaos theory.

Biological Psychiatry, 45, 261–269.

Zeigler-Hill, B., & Abraham, J. (2006). Borderline personality features: Instability of self-esteem and affect. *Journal of Social and Clinical Psychology*, 25, 668-687.