Extending the Volatility Concept to Point Processes [1]

David R. Brillinger [2]

Statistics Department, University of California, Berkeley, CA, 94720-3860

"So I think that in time series you can't write in a definitive fashion. The field is still going to have a lot of changes in it."

T. W. Anderson quoted in DeGroot (1990).

Abstract

Volatility is a vague concept that can be made precise in a number of ways. This paper investigates the concept for point processes in the time domain. There is review of the time series case, then a volatility measure is developed. Next it is applied to some biomedical data based, RR intervals. of the heart beating. It is seen in the work that in some cases time series volatility might be generated by point process volatility.

MSC: Primary 62M10, 60G55; Secondary 62M99, 91B30.

Keywords: atrial fibrillation, point process, RR intervals, time series, volatility.

[2]Telephone: 510-642-0611, Fax: 510-642-7892, Email: brill@stat.berkeley.edu

1. Introduction

Ted Anderson has been a statistical beacon throughout so many of our academic careers. That light now includes the sparkling talk that he gave recently aged 90. This paper concerns a concept from economic time series analysis, one of Ted's many specialties.

Volatility is an important topic for time series analysis generally and for financial series work particularly. It is a vague concept, capable of being formalized in a variety of ways. One persistent notion is local in time variability, that is things are changing a lot as time advances. Volatility has been measured by the local standard deviation, $\sigma(t)$, or the local variance, $\sigma(t)^2$ for series such as the returns of a security or a market index. Commonly, the higher the volatility the riskier the security, and thus volatility is important information for a security owner or prospective buyer.

The study of volatility can lead to better forecasting of a series, to better understanding of the past, and to better assessment of risk. For example in insurance considerations the safety loaded pure risk premium can take the form

$$\lambda_1 p(t) + \lambda_2 \sigma(t) + \lambda_3 \sigma(t)^2$$

where the λ's are weights, $p(t)$ is the fair premium, and $\sigma(t)$ and $\sigma(t)^2$, are volatility measures at time t. One reference to the insurance case is Daykin et al (1994).

Another measure of financial risk is the Value at Risk (VaR). It has been defined as the maximum expected loss over a specified time period with a given confidence level of occurring. Losses in the time horizon are only to exceed the VaR only with prespecified probability α, see Tsay (2002) or Bouchard and Potters (2003). Estimates of something like $\sigma(t)$ are required for VaR estimation.

The motivating consideration of this present work is the study of the concept of volatility in the point process case. By a point process is meant a non-decreasing sequence of time points, $\{\tau_j\}$, $j = 0, \pm1, \pm2, ...$ It seems worth studying an extension to point processes for reasons including:

a) It appears that generally in the study of time series something more can be learned about that case by studying point process analogs.

b) Point processes are the building blocks of many other processes. This has the implication that volatility in a time series may be generated from volatility in a latent point process. (This will be illustrated later in the paper.)

c) Point processes are an interesting data type in their own right.

d) The subject matters of finance and time series can be brought into the point process case by imagining that there is a common cost associated with each event.

e) One might chose to replace the point process by a 0-1 valued discrete time series.

Things desired in a study of volatility might include: description, detection, formalization, prediction. In this work two measures are proposed, one a function of time and the other a function of a point's index number.

The investigation provided is partly empirical. There is study of a biomedical data set. In fact there are many economic and financial empirical examples of volatility analysis of time series, see Bouchard and Potters (2003), and Tsay (2002) for example.

The layout of the paper includes: discussion of the time series concept, then the point process case. There is material on risk analysis and general discussion.

2. The time series case

2.1 Time series background

The work begins with reference to the development of the volatility concept in the financial world.

On their website Merrill Lynch provides the following definition,

> "*Volatility.* A measure of the fluctuation in the market price of the underlying security. Mathematically, volatility is the annualized standard deviation of returns."

The definition refers to "returns". For a financial entity they are defined as follows: for $t = 0, \pm 1, \pm 2, ...$ if P_t denotes the price of the security of interest at time the return is

$$Y_t = (P_t - P_{t-1})/P_{t-1} \tag{1}$$

This definition fits in with the usage of the word over many years. Return is also sometimes taken to be

$$log \; P_t/P_{t-1}$$

The two definitions are close if the values P_t are not changing quickly.

An empirical formula for volatility at time t is provided by

$$se\{Y_s \mid s \; near \; t\}, \tag{2}$$

or its square, with *se* denoting standard error.

In the time series case there are model-based formulas as well. Consider the GARCH(P,Q) series Y_t given by

$$Y_t = \mu_t + \sigma_t \epsilon_t, \tag{3}$$

and

$$\sigma_t^2 = \alpha_0 + \sum_{p=1}^{P} \alpha_p [Y_{t-p} - \mu_{t-p}]^2 + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2 \tag{4}$$

3

for $t = 0, \pm 1, \pm 2, ...$. Here μ_t is the mean level, ϵ_t has zero mean, unit variance, and the $\alpha's, \beta's$ are ≥ 0. The volatility has been defined as σ_t^2, see Tsay (2002). Once the α's and β's have been estimated an estimate of σ_t can be computed and fed into risk computations. A further advantage of employing a GARCH model is that predictions of future Y_t values may be constructed.

In another approach, when time is continuous, one has

$$E[Y(t) - Y(t-h)]^2 = 2[c(0) - c(h)] \approx -2c''(0)h^2 \qquad (5)$$

for small h, provided the derivative exists.

The righthand side of (5) can be recognized as

$$2c(0)\pi^2[E(\#\{crossings\ of\ mean\})]^2 h^2$$

for the stationary normal, see Cramer and Leadbetter (1967), page 194. This leads one to consider the number of crossings of the mean level in the neighborhood of t as another measure of volatility. The corresponding empirical measure is

$$\#\{crossings\ of\ mean | near\ t\} \qquad (6)$$

which may be considered as a local measure of volatility.

If one now views the series $\{Y(t)\}$ as locally stationary, then with this definition volatility is getting at whether the local second derivative of the autocovariance function at 0 changes much as one slides along the series.

Ghysels, Harvey and Renault (1996) presents a broad review of volatility analysis for time series.

2.2 A time series example

A biomedical example is now considered. The MIT-BIH Atrial Fibrilation Database, Goldberger et al (2000), contains a variety of cardiology data sets and analysis programs. The data to be studied here are for case 04043. They are intervals between heartbeats. The last 6000 RR intervals are studied in particular.

Thus one has a time series of interval lengths, $\{X_j\}$ to work with, and the series index, j, is the dimensionless beat number. The units of the intervals X_j are seconds. Successive X_j are graphed in Figure 1. One sees scattered bursts of much higher variability interspersed amongst a background of approximately constant lower variability. One can say that the series is more volatile during the bursts. The bursts are seen to begin and end abruptly, i.e. there is no slow uprise or decay.

In the analyses presented the sequence of RR intervals, $\{X_j\}$, is converted to relative changes in interval length by setting

$$Y_j = (X_j - X_{j-1})/X_{j-1} \qquad (7)$$

4

$j = 2, ..., 6000$. A different stretch of the 04043 data is studied in Tateno and Glass (2000). These researchers work with the successive interval length differences, $X_j - X_{j-1}$ with the goal of detection atrial fibrillation automatically.

The values Y_j are graphed in the upper left panel of Figure 2. The series is similar to Figure 1, but the local trends of Figure 1 have been reduced and the background to the bursts is more orderly.

In Figure 2 the upper right panel shows the volatility measure (2), that is standard errors of batches of 36 interval lengths are plotted. The bottom left panel provides the zero-crossing value (6). This plot is very similar to the previous, but has more background variability at its base.

Each of the the volatility panels does appear consistent with an intuitive notion of volatility as extra variability and fits with the time series plot itself.

A number of GARCH models were fit, but none did well.

3. The point process case

3.1 Point process background

The material in the preceding section has been presented by way of intoducing results for the point process case. Point processes are a common data type and are building blocks of many time series processes. There are both parametric and nonparametric forms of analysis available.

Suppose time is continuous, $-\infty < t < \infty$. One can denote a realization of the point process by a non-decreasing sequence of time points, $\{\tau_j\}$, $j = 0, \pm 1, \pm 2, ...$, or by a non-decreasing function, $N(t)$, counting the number of points up to and including time t, starting at time 0, $N(t) = \#\{\tau_j \leq t\}$. In the present case the τ_j are obtained by summing the lengths of the beat intervals, as in

$$\tau_j = \sum_{k=1}^{j} X_k \tag{8}$$

The approach here is to view the point process as locally stationary, and thereby imagine the parameters to be changing (slowly) in time.

The basic population parameters of point processes to be employed here are the rate and the covariance density. To this end suppose that the points are isolated so that there is at most one point in a sufficiently small interval. An algebra of increments and delta functions may be employed in to develop expressions for population parameters, Brillinger (1972). For example one sets

$$E\{dN(t)\} = p_N(t)dt$$

with $p_N(t)$ the point process rate at time t. Supposing one seeks a measure of local dependence one can set

$$cov\{dN(r), dN(s)\} = [p_N(r)\delta(r-s) + q_{NN}(r,s)]drds \tag{9}$$

with $q_{NN}(r,s)$ the covariance density, and $\delta(.)$ the Dirac delta.

For h small one might consider $var\{N(t+h) - N(t)\}$ as a measure of the variability of a point process. Concerning it one has

$$var\{N(t) - N(t-h)\} = \int_{t-h}^{t} \int_{t-h}^{t} [p_N(r)\delta(r-s) + q_{NN}(r,s)]drds$$

$$\approx p_N(t)h + q_{NN}(t,t)h^2 \tag{10}$$

This will become $p_N(t)h$ for very small h.

Poisson variability shows itself in the presence of the term $p_N(t)h$ in (10). The Poisson process is typically the standard against which other point processes are measured. The $q_{NN}(t,t)$ term represents dispersion at time t. Depending on the sign of $q_{NN}(t,t)$ there can be over or under dispersion. It is direct to set down examples, theoretical and empirical, of each type of dispersion.

In the Poisson case the variance is the mean and so one could consider the rate $p_N(t)$, as a measure of variability or volatility. In the case that $p_N(.)$ is a smooth function of time $p_N(t)$ may be estimated by,

$$\int k(t-u)dN(u) / \int k(t-u)du \tag{11}$$

with $k(.)$ a non-negative kernel vanishing outside a short interval including 0. The variance of this estimate in the case of (9) is,

$$p_N(t) \int k(t-u)^2 du / [\int k(t-u)du]^2 + q_{NN}(t,t) \approx p_N(t)/h + q_{NN}(t,t)$$

and one has been led to (10) divided by h, supposing $k(.)$ integrates to 1.

As another method to be led to expression (10), suppose a time series has the form

$$\int a(t-u)dN(u)$$

Then its variance at time t is

$$\int a(t-u)^2 p_N(u)du + \int \int a(t-u)a(t-v)q_{NN}(u,v)dudv$$

6

For the local rate extending over the interval $(-h/2, h/2)$ this is approximately

$$a(0)^2 p_N(t)h + a(0)^2 q_{NN}(t,t)h^2$$

and expression (10) shows itself again.

To tie in with the work of the previous section one can consider the discrete time series, $\{X_j\}$ of successive values $X_j = \tau_j - \tau_{j-1}$, $j = 0, \pm 1, \pm 2, \ldots$ i.e. the interpoint times. One can look for volatility in this interval series as in the previous section.

In summary, the discussion has led to $p_N(t)h + q_{NN}(t,t)h^2$ as a measure of the volatility near t. An estimate of $p_N(t)$ was given above at (11). There are several ways to estimate $q_{NN}(.)$ in the stationary case, see Brillinger (1976). It seemed most convenient here to proceed via a 0-1 time series approximation to the point process. A 0-1 valued series may be obtained as follows. Set

$$Z_k = 1 \ if \ k = \ < \tau_j/h > \tag{12}$$

for some j and for small h with $< . >$ denoting the integral part of its argument for $k = 0, \pm 1, \pm 2, \ldots$, see Brillinger (1987). To have the same time domain one rescales by writing $t = hk$. The parameters of this 0-1 process relate to those of $\{\tau_j\}$ by noting that

$$Z_{hk} = \int_{hk}^{h(k+1)} dN(u)$$

and so its variance is

$$\int \int [p_N(r)\delta(r - s) + q_{NN}(r, s)]drds \approx p_N(t)h + q_{NN}(t,t)h^2$$

and one is again led back to the desired expression (10).

3.2 A point process example

Consideration now turns to the R times or beat times, $\{\tau_j\}$, of the cardiology series. Figure 3 displays various statistics. The upper right panel shows the running rate computed as in (11), with $k(.)$ uniform,. It is computed by splitting the τ_j values into segments of 3.6 seconds in length. Surprisingly the plot is not that unlike the one of Figure 1.

For the next computations it is convenient to replace the point process values by a 0-1 time series constructed as at (12).

The top right panel shows the rate, here called the local average, computed via the 0-1 approach and splitting the series into 201 segments of 176 points. The top two panels are quite similar. The bottom left shows the running variance estimate computed as was

the top right panel. It is surprisingly similar to the top left panel leading one to compute the ratio of local variance to the local average. One sees underdispersion relative to the Poisson. This occurs commonly with approximately periodic processes, as one has here.

4. Risk analysis

Risk is the chance of some unpleasant event occurring. In the present context it might relate to too many points occurring in a brief time interval. With a statistical model that probability can be estimated. An estimate might be simply the historical frequency.

One motivation for the work on volatility for point processes is that risk analyses do arise for point process situations. One can mention earthquakes, wildfires and floods.

Volatility can be a warning that a process is out of control, that a risk has arisen, that predictions have become difficult and that damage can occur if the volatility continues. In a sense volatile periods are outliers and their occurrence needs to be studied and dealt with

5. Extensions

Some problems remain. These include: automatic detection of high volatility, statistical properties of the estimates, extensions to the spatial case, the vector case, the marked point process case, the long-tailed case, and the integer-valued case. One can also consider random effect models.

A basic issue is how to display point process data to bring out the presence or absence of volatility.

6. Summary and discussion.

The work has focused on two measures of local variability or volatility for point processes. One measure is a function of time, based on the τ_j, the other is a function of an index, j, based on the $X_j = \tau_{j+1} - \tau_j$. The measures get at local behavior and appear useful for prediction and process control.

They are computed for a biomedical series that displays bursts of activity and the results appear in accord with one's intuition.

It was also seen that volatility in a latent point process could be the cause of volatility in a time series.

References

Bouchard, J-P. and Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing.* Cambridge, Cambridge.

Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. Pp. 483-513 in Proceedings of the Sixth Berkeley Symposium on Mathematical

Statistics and Probability, (eds. L.M. Le Cam, J. Neyman, E.L. Scott), Berkeley, University of California Press (1972)

Brillinger, D. R. (1976). Estimation of the second-order intensities of a bivariate point process. *J. Royal Statistical Society* 38, 60-66.

Brillinger, D. R. (1987). Analysing interacting nerve cell spike trains to assess causal connections. *Advanced Methods of Physiological System Modelling* I, 29-40

Cox, D. R. (1989). *Analysis of Binary Data*, Second Edition. Chapman and Hall.

Cramer, H. and Leadbetter, R. (1967). *Stationary and Related Stochastic Processes.* J. Wiley, New York.

Daykin, C. D., Pentikainen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries.* Chapman and Hall, London.

DeGroot, M. H, (1989). A conversation with T. W. Anderson. *Statistical Science* 1, 97-105.

Ghysels,E., Harvey, A. C. and Renault, E. (1996). Stochastic volatility. Pp. 119-192 in *Handbook of Statistics*14. Eds. G. S. Maddala and C. R. Rao.

Goldberger A. L., Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215].

Tateno, K. and Glass, L. (2000). A method for detection of atrial fibrillation using RR intervals. *Computers in Cardiology* 27, 391-394.

Tsay, R. S. (2002). *Analysis of Financial Time Series.* J. Wiley, New York.
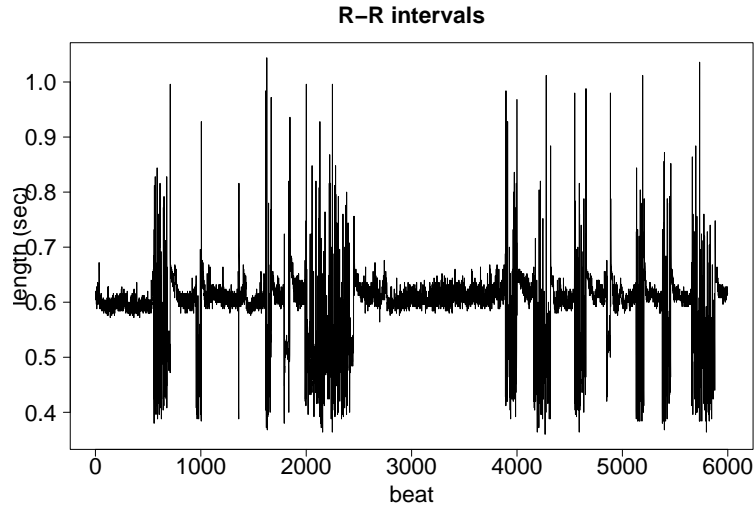
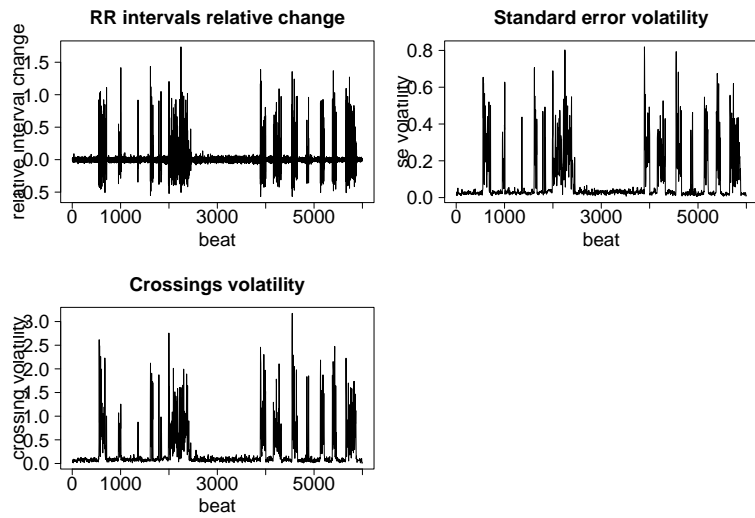Figure 1: 6000 RR interval lengths plotted against beat number.

Figure 2: Upper left, the series Y_t. The other panels provide the measures (2) and (6) of volatility for the RR intervals as a time series.
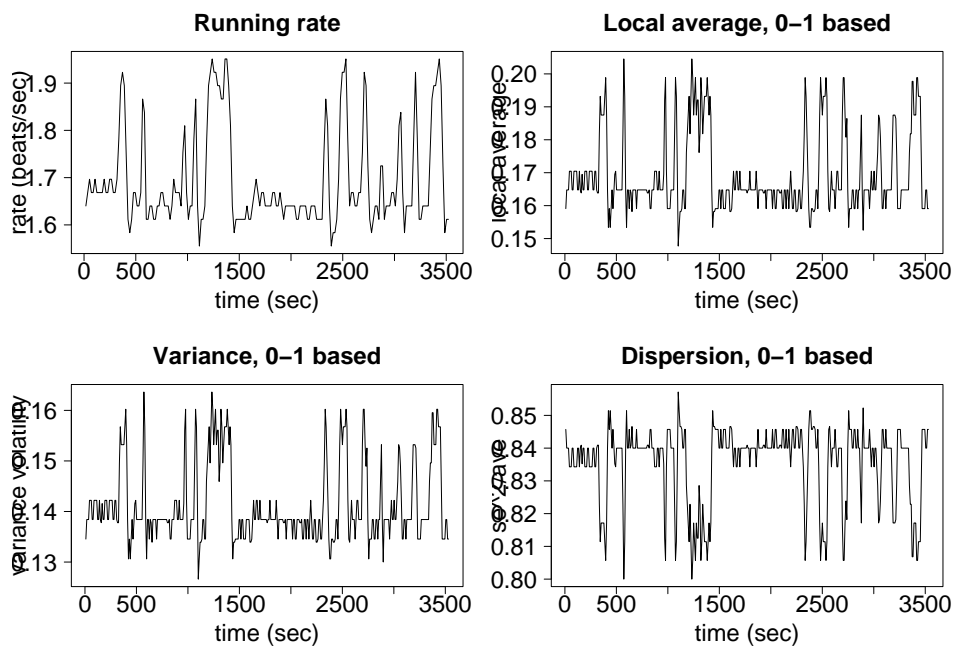
Figure 3: Running rate of the point process of R times, τ_j, computed via expression (11). The top right and bottom left are computed by splitting the 0-1 series into 201 segments of 176 points. The bottom right is the variance divided by the local average, i.e. the index of dispersion. One notes that it is always less than .86 .