Re-scaled Range Analysis of a decade of crime in Baltimore.

Frederick Carlson[1, *]

[1] *University of Vermont*

(Dated: December 14, 2022)

This is a short paper on the use of R/S Hurst Exponent Estimation on Crime Data. The central point of the paper is that the source data for crime has reached a point where we can integrate non-linear analysis to supplement standard linear statistical and time series analysis that has be a fixture of law enforcement management for decades. This work extends work done 25 years ago on a much smaller dataset of the City of Vancouver. The output will be a R/S analysis and Hurst Estimate for each District. The results align very closely to the prior work by Verma, with a range of H being roughly in the range for all calls in the Vancouver study (H=0.622).

I. INTRODUCTION

Crime is a complex sequence of events that occurs when a subject (motivated offender) encounters an object (criminal target) without a functioning protector.

Proposition 1: Crime is not a random event in almost all cases. The subject must interact with the object for a crime to happen. This event is not unexpected as it is one of two unknowns with heavy memory. Please think of the criminal and the target going down two streams that carry them to each other. The initial conditions were not known in advance, but a persistence brings one to the other.

Proposition 2: The Guardian Function (Control) cannot economically sustain being everywhere at once, so a risk-based calculus is established to deploy scarce policing resources best to provide a functioning protective service to as much of the jurisdiction as the police can. The measurement of effectiveness, and the key risk indicators, are overwhelmingly statistical and linear in nature.

Proposition 3: The very nature of modern life ensures that a criminal encounter is feasible and even likely. Geography, internet connectivity, tax systems, social structures, and socioeconomic and legal conditions tied to the locale create a connected environment where even small changes in activity conditions will have an outsized effect.

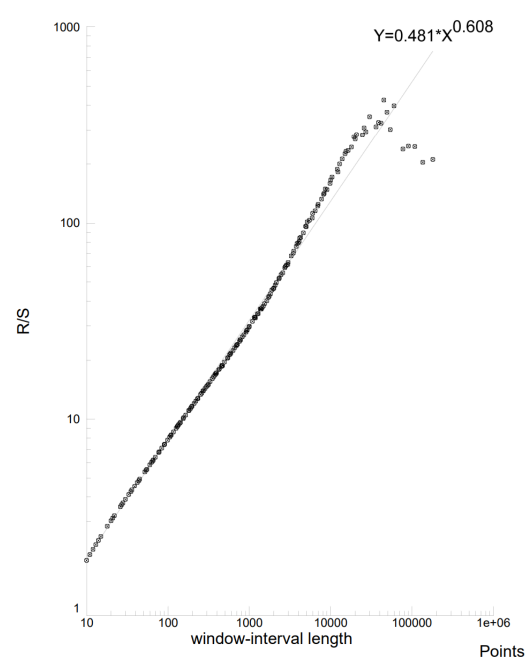
Proposition 4: Crime is a complex system with strong non-linear characteristics that linear and non-linear tools can best be used in conjunction to measure. Deterministic methods such as modern statistical analysis are part of the solution as crime is party deterministic and partly probabilistic with extreme sensitivity to initial conditions. That said, we can assume that all but the most sophisticated law enforcement organizations have linear expectations in a non-linear world.

Proposition 5: In a situation where non-linear and linear determinants coexist is likely to be modeled assuming that we can create time series fractals that are self-similar and scale-invariant. Police Departments can visualize the same crime metrics in a dynamic system as a non-linear one but add a new dimension to their analysis. (Verma)

Example - Hurst Exponents and R/S analysis

The Hurst Exponent is a time series analysis that is related to Brownian motion and is used as a baseline for the persistence of crime in each of the Baltimore Police Districts. What this paper seeks to do is conceptually straightforward. We wish to calculate a re-scaled range estimation of Hurst Exponents for each of the nine police districts in the City of Baltimore. We will also calculate 1/H, which is the fractal dimension. Figure 1 below shows the R/S Plot and H estimation for the full Crime Dataset.

FIG. 1. R/S Plot and H Value (0.608) for Crime in the City of Baltimore 2011-2022 (Benoit – Fractal Analysis Software)



* frederick.carlson@uvm.edu

II. DESCRIPTION OF DATA SETS

The data for this project will come from the Open Baltimore data set, which has ten years of crime data. The last known R/S range analysis on crime data that the author was able to locate was in the City of Vancouver, British Columbia, for several months. This data set goes for ten years. Figure 2 shows the data in the Minitab Statistical Analysis Program.

FIG. 2. OpenBaltimore Crime Dataset (Open Baltimore)

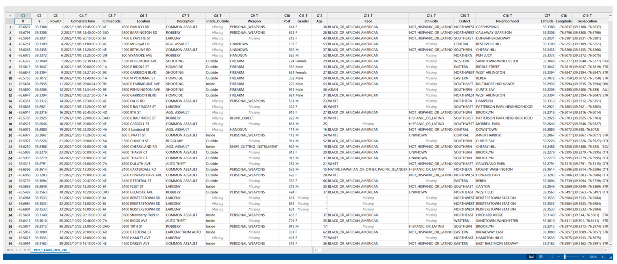
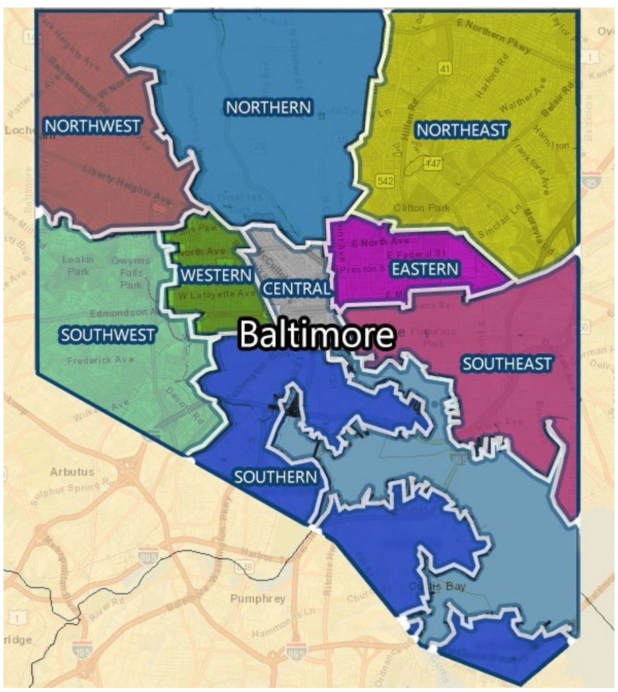


FIG. 3. The Nine Police Districts of Baltimore (BCPD)



III. MODEL

Critical to this paper is the data breakout into the seven police districts. Figure 3 shows the boundaries of the seven districts. These areas are distinctive and have unique crime and complex social characteristics.

The Open Baltimore Dataset has logging data with a sequential log id that serves as a useful timestamp to show the "burstiness" of crime data when the data is broken into different d istricts. This use of log identification timestamps is similar to Verma's in the Vancouver study but with far more data.

The model this paper will be using is a Rescaled Range Analysis Hurst Estimate using a variable window of the Baltimore Time Series Data. The aggregate dataset will be divided into nine districts, and the results will present the type of crime alongside the R/S Analysis. The initial analysis will be done with the Beniot package with a Cross Check with the R Pracma library hurstexp.

Peters provides the overview of the Rescaled Range Hurst Estimation Technique shown below. (Peters p. 61)

$x\,(t, N) = \sum_{z=1}^{t} (i_z - A_n)$

where: $x\,(t, N)$ = deviations over N periods

i_z = number of log entries per day

A_N = average of log entries over N periods

and: Range is = $MAX\,[x\,(t, N)] - MIN\,[x\,(t, N)]$

IV. RESULTS

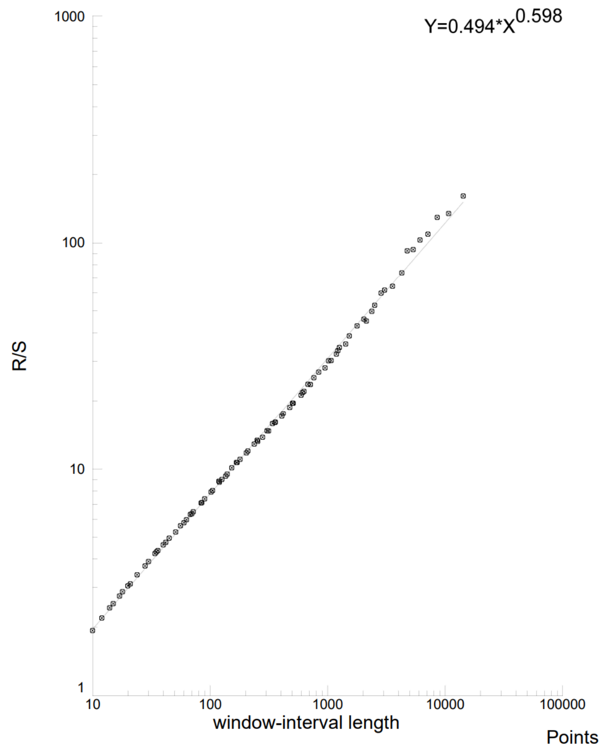
Table I shows the the results of the output of crime data from the Beniot Package. These results show a moderately persistent crime trend for the entire ten years. Interestingly, the H values are extremely close to 25 year old crime data in a completely different city (Vancouver) that was measured in a much shorter time-frame of a month. Figures 4 shows the R/S Output for the Western and Southwestern Districts.

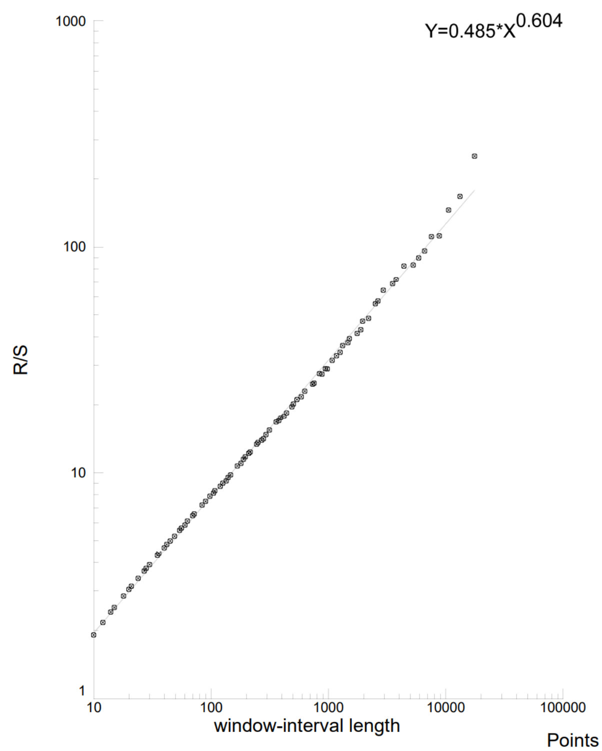
TABLE I. H and Fractal Dimension by District

	H (Beniot)	Fractal Dimension (Beniot)	H (Vancouver Study)
All Districts	0.608	1.644	0.622
Central	0.627	1.595	
Western	0.598	1.672	
Eastern	0.595	1.681	
Southwest	0.604	1.656	
Northern	0.592	1.689	
Northeast	0.622	1.607	
Northwest	0.599	1.669	
Southeast	0.622	1.607	
Southern	0.606	1.650	

Hurst Case	Interpretation
H=0.50	Random Walk
0<H<0.50	Anti-Persistent
0.50<H<1.0	Persistant

FIG. 4. R/S - Western and Southwestern Districts





V. CONCLUDING REMARKS

Figure Five shows the cross check from a completely different package using the hurstexp function from the R Practical Mathematics (Pracma). The one this paper will use to cross check is the "Corrected R/S Hurst Exponent.

The H estimated from the R package is 0.600 for the Southwest District and 0.571 for the Western District. This is a difference of -0.004 for the Southwest District and -0.027 for the Western District.

These results seem solid but should be taken with a grain of salt. This paper, like the issues in the Verma paper has a very loose estimate of the number of log entries per day. While this seems to hold up with prior research, a much more exhaustive examination of the time series assumption of calls is needed. An integration of the work by Melgarejo, et al, would be a good extension of this paper. (MELGAREJO and OBREGON)

FIG. 5. R Code to Cross Check H for Western and Southwestern Districts (PRACMA: Hurst – R Documentation)

```
Southwestern

> xgn <- rnorm(1024)
> xlm <- numeric(1024); xlm[1] <- 0.1
> for (i in 2:53000) xlm[i] <- 4 * xlm[i-1] * (1 - xlm[i-1])
> hurstexp(xgn)

Simple R/S Hurst estimation:         0.5577965
Corrected R over S Hurst exponent:   0.6000586
Empirical Hurst exponent:            0.6125285
Corrected empirical Hurst exponent:  0.5722207
Theoretical Hurst exponent:          0.5404756

Western

> xgn <- rnorm(1024)
> xlm <- numeric(1024); xlm[1] <- 0.1
> for (i in 2:42000) xlm[i] <- 4 * xlm[i-1] * (1 - xlm[i-1])
> hurstexp(xgn)
Simple R/S Hurst estimation:         0.5508644
Corrected R over S Hurst exponent:   0.5711356
Empirical Hurst exponent:            0.5026615
Corrected empirical Hurst exponent:  0.4675594
Theoretical Hurst exponent:          0.5404756
```

VI. METHODS

The methods used in this paper were pulled from preexisting computer software from the Benoit Fractal Analysis Package (Trusoft Corporation) and from the R Packages "Pracma" and "methods." The author of this paper did not manually compute the R/S estimate.

The data from the OpenBaltimore site was wrangled from Excel and Minitab.

ACKNOWLEDGMENTS

We are grateful for the support and friendship from everyone at the Vermont Complex Systems Center.

Works Cited

Baltimore Police Department, https://www.baltimorepolice.org/.

Benoit – Fractal Analysis Software, http://www.trusoft-international.com/.

MELGAREJO, MIGUEL, and NELSON OBREGON. “Multifractal Approach to the Analysis of Crime Dynamics: Results for Burglary in San Francisco.” Fractals, vol. 25, no. 05, 2017, p. 1750043., https://doi.org/10.1142/s0218348x17500438.

Open Baltimore, https://data.baltimorecity.gov/.

Peters, Edgar E. Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility. John Wiley & Sons, 1996.

“PRACMA: Hurst – R Documentation.” Quantargo, https://www.quantargo.com/help/r/latest/packages/pracma/2.3.3/hurst.

Schroeder, Manfred Robert. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. Dover Publications, 2020.

Verma, Arvind. “The Fractal Dimension of Policing.” Journal of Criminal Justice, vol. 26, no. 5, 1998, pp. 425–435., https://doi.org/10.1016/s0047-2352(98)00021-x.