METHODS FOR ESTIMATING THE HURST EXPONENT OF STOCK RETURNS: A NOTE

JAMAL MUNSHI

ABSTRACT: This note is a further commentary on a previous paper on the chaos theory of stock returns that derives from the alleged detection of persistence in time series data indicated by values of the Hurst exponent H that differs from the neutral value of H=0.5 implied by the efficient market hypothesis (EMH) (Munshi, 2014). A comparison of four different methods for estimating H is presented. Linear regression of log transformed values (OLS) is compared against a numerical approach using the generalized reduced gradient (GRG) method. These methods are applied to two different empirical models for the estimation of H. We find that the major source of error in the empirical estimation of H is the insertion of the extraneous constant C into the empirical model[1].

1. INTRODUCTION

Some methodological flaws in Rescaled Range Analysis of financial markets were explored in a previous paper[2] which addressed certain issues having to do with the empirical test of the Hurst equation (Hurst, 1951) as it is applied to financial markets (Munshi, 2014). The Hurst equation is

Equation 1 $$R/S = \nu^H$$

Empirical tests of this relationship in stock returns is normally carried out using OLS regression and so a logarithmic transformation is used to render it into linear form as

Equation 2 $$\ln(R/S) = H*\ln(\nu)$$

A test of the linear equation as written in Equation 2 requires that the constant term b_0 in the linear model {y = b_0 + b_1*x} be set to {b_0 = 0} and the linear model written as {y = b_1*x} to have a direct correspondence with Equation 2. The empirical test could then proceed by taking the natural logarithms of the values of R/S and ν and using OLS regression to estimate the value of H in Equation 2. However, this is not what R/S analysts do possibly because setting {b_0 = 0} lowers the precision of the linear fit and that in turn increases uncertainty of the b_1 coefficient, and the likelihood that the empirical test will fail

---

[1] Date: February, 2015

Keywords and phrases: finance, financial analysis, efficient market hypothesis, financial markets, chaos theory, stock markets, rescaled range analysis, Hurst exponent, fractal theory of stock markets, stock price behavior, long term memory of stock returns, persistence in stock returns, OLS regression, least squares, linear regression, transformations, logarithmic transformations, nonlinear regression, numerical methods, computational statistics, generalized reduced gradient.

Author affiliation: Professor Emeritus, Sonoma State University, Rohnert Park, CA, 94928, munshi@sonoma.edu

[2] This paper should be read as an addendum to "THERE IS NO CHAOS IN STOCK MARKETS" available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2448648

to find a significant result. It is probable that it is the desire to maintain a higher statistical power of the test that R/S researchers retain the b_0 term and write their linear model as

Equation 3 $$\ln(R/S) = \ln(C) + H*\ln(v)$$

Recognizing the mismatch between Equation 1 (the theory to be tested) and Equation 3 (the alleged empirical model for testing the theory), researchers then work backwards to force the theory to match their empirical test by introducing the constant C and writing Equation 1 as

Equation 4 $$R/S = C*v^H$$

Statistical power is thought to be achieved but at the great cost of not knowing what C represents and therefore of not knowing what theory is being tested. This issue is the main focus of a previous paper on Rescaled Range Analysis (Munshi, 2014). It is noted in that paper that in addition to the problem with C[3], errors may also be introduced by the logarithmic transformation itself (Greene, 2010) (Zeng, 2011). An alternative to OLS is the use of numerical methods to fit the non-linear equation directly. Thus to investigate the reliability and validity issues in the estimation of H we examine the four different approaches involving two different empirical models and two different estimation methods as indicated in Table 1.

	Method for minimizing sum of squared errors	
Empirical model	OLS	GRG
$R/S = C*n^H$	METHOD 1	METHOD 2
$R/S = n^H$	METHOD 3	METHOD 4

Table 1: Four methods for estimating the Hurst exponent H

OLS regression is essentially a minimization problem that with certain assumptions can be solved algebraically to yield an expression for the value of the coefficients of a *linear* model at which the sum of squares of the prediction errors is at a minimum (Draper, 1998)[4]. When the conditions for these algebraic solutions do not exist, generalized extremum seeking convergence analysis may be used to hunt for the minimum. Numerical methods such as linear programming, quasi-Newton, reduced gradient, and generalized reduced gradient (GRG) use a trial-and-error algorithm to converge to an approximate solution (Lee, 2004) (University of Pittsburg, 2015) (Lasdon, 1973) (Lasdon, Nonlinear optimiation, 1973). The *Solver* tool in Microsoft Excel (Microsoft, 2015) includes a GRG tool that may be used for simple non-linear systems such as Equation 1.

[3] and of using average values of R/S instead of the raw data
[4] Normal Draper and Harry Smith

2. THE ESTIMATED H-VALUES FOR EIGHT STOCK RETURNS TIME SERIES

In the previous paper that this note follows, Rescaled Range Analysis was carried out for eight stock returns series (Munshi, There is no chaos in stock markets, 2014). They included daily returns for IXIC, DJI, BAC, and Cl, and weekly returns for SPX, CAT, BA, and IBM. In this note we compute the H-values of these time series using the four different methods listed in Table 1. The results are summarized in Table 2. The computational details may be found in the relevant Excel file included in the data archive for this paper (Munshi, HurstExponentPaperDataArchive, 2015). Table 2 shows that for some of the stocks there is a great deal of variation among the four methods in the estimated value of H as well as in the precision[5] of the estimate. These differences are easier to appreciate in their graphical depiction in Figure 1 and Figure 2.

Model	Method		IXIC	DJI	BAC	CL	SPX	CAT	BA	IBM
R/S=ln(C) + H*ln(ν)	OLS	C	0.9550	1.0012	0.7339	1.3406	0.6999	1.1850	0.7721	0.9346
		H	0.5282	0.5065	0.5810	0.4352	0.6030	0.4741	0.5872	0.5548
		R2	0.8627	0.8639	0.8594	0.8382	0.8472	0.8211	0.8478	0.8395
		sigma	0.0270	0.0257	0.0301	0.0245	0.0328	0.0283	0.0319	0.0311
		H-0.5	0.0282	0.0065	0.0810	-0.0648	0.1030	-0.0259	0.0872	0.0548
		t-statistic	1.0444	0.2529	2.6910	2.6449	3.1402	0.9152	2.7335	1.7621
		p-value	0.3004	0.8012	0.0092	0.0104	0.0026	0.3637	0.0082	0.0831
R/S=C*ν^H	GRG	C	0.8898	0.6631	0.6356	1.3220	0.9080	1.5396	1.0776	1.2200
		H	0.5433	0.5827	0.6102	0.4399	0.5597	0.4286	0.5309	0.5103
		R2	0.8961	0.8867	0.9368	0.9104	0.9184	0.8375	0.8591	0.8780
		sigma	0.0265	0.0254	0.0288	0.0235	0.0315	0.0280	0.0317	0.0304
		H-0.5	0.0433	0.0827	0.1102	-0.0601	0.0597	-0.0714	0.0309	0.0103
		t-statistic	1.6345	3.2601	3.8224	2.5565	1.8950	2.5480	0.9751	0.3387
		p-value	0.1073	0.0018	0.0003	0.0131	0.0628	0.0134	0.3334	0.7360
R/S=H*ln(ν)	OLS	H	0.5192	0.5067	0.5205	0.4925	0.5332	0.5073	0.5366	0.5416
		R2	0.8624	0.8639	0.8498	0.8232	0.8356	0.8170	0.8413	0.8390
		sigma	0.0270	0.0257	0.0303	0.0247	0.0330	0.0284	0.0320	0.0311
		H-0.5	0.0192	0.0067	0.0205	-0.0075	0.0332	0.0073	0.0366	0.0416
		t-statistic	0.7110	0.2607	0.6772	0.3034	1.0052	0.2573	1.1429	1.3372
		p-value	0.4798	0.7952	0.5008	0.7626	0.3187	0.7978	0.2575	0.1860
R/S=ν^H	GRG	H	0.5250	0.5184	0.5401	0.4846	0.5448	0.4974	0.5425	0.5413
		R2	0.8946	0.8710	0.9186	0.8963	0.9174	0.8055	0.8585	0.8732
		sigma	0.0265	0.0256	0.0291	0.0237	0.0315	0.0286	0.0317	0.0305
		H-0.5	0.0250	0.0184	0.0401	-0.0154	0.0448	-0.0026	0.0425	0.0413
		t-statistic	0.9429	0.7189	1.3773	0.6500	1.4213	0.0910	1.3407	1.3544
		p-value	0.3494	0.4749	0.1734	0.5181	0.1602	0.9278	0.1849	0.1805

Table 2: Comparison of H values of stock returns estimated using four different methods

[5] Measured as R^2

The four different estimation methods listed in Figure 1 are: OLSC, corresponding to Method 1, GRGC, corresponding to Method 2, OLS, corresponding to Method 3, and GRG, corresponding to Method 4 in Table 1. Figure 1 shows that although the H values are fairly stable across estimation methods for some stocks, IXIC for example, for other stocks the value of H varies sufficiently to lead to different conclusions with respect to long term persistence in the series. Much of the divergence is introduced by the inclusion of C in the model. It is apparent in Figure 1 that the H values derived from empirical models that do not include C (OLS and GRG) are more stable than those derived from empirical models that do. A possible motive for the introduction of C is apparent in Figure 2 which shows that R^2 is somewhat higher in models that include the constant C.

Figure 1: Comparison of H values for eight stocks using four different methods

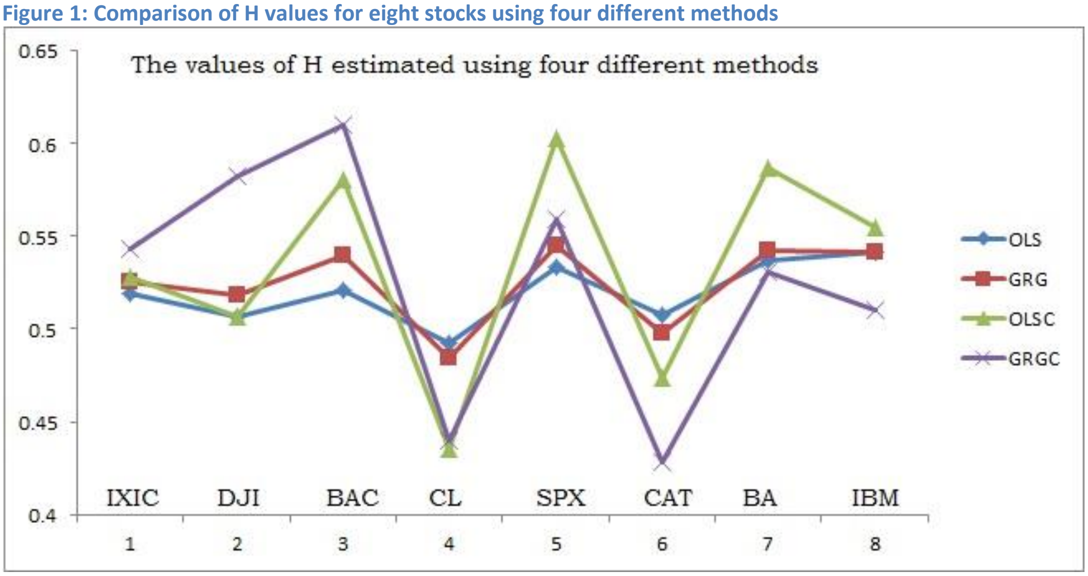
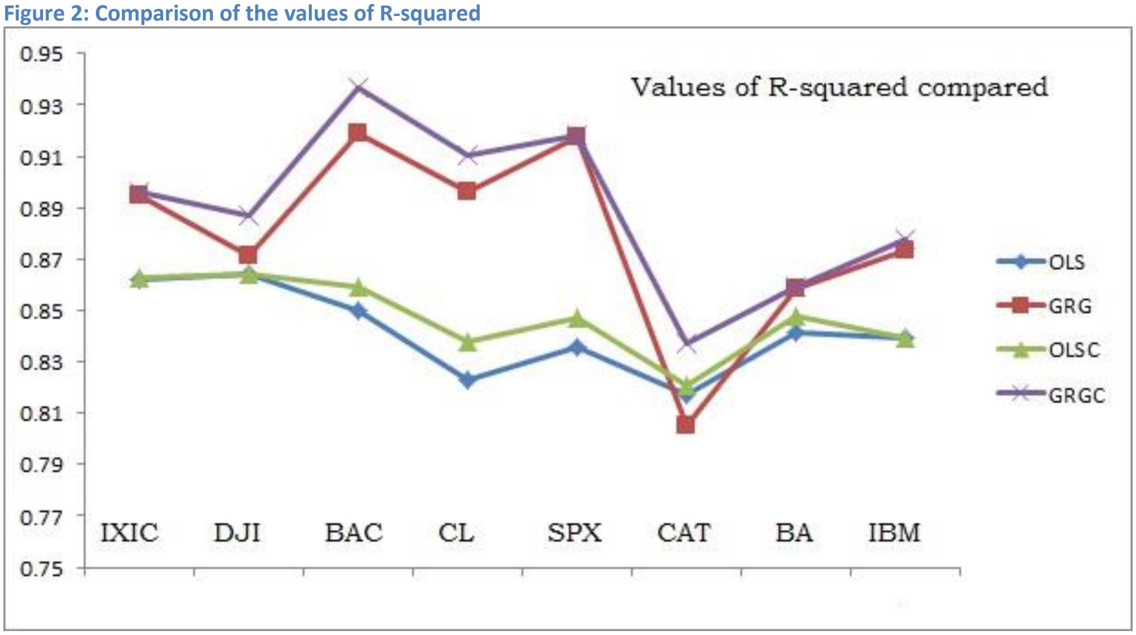


Figure 2: Comparison of the values of R-squared



3. A TEST FOR VALIDITY

In Figure 1 we can see that the H values estimated using the OLS and GRG methods are more stable than those derived with the OLSC and the GRGC methods but what is not clear is how these methods compare with respect to validity. Accordingly, a test for validity is devised using the same values of √ used in the stock returns series in Table 2 but this time we compute *synthetic* values of R/S set exactly equal to $√^{0.5}$ so that the Hurst exponent is set to H=0.5. We then introduce "noise" into the R/S values as normally distributed residuals and incrementally increase the range of noise levels until the R-squared values of the estimation models are approximately equal to those in Table 2. Now that we know the true value of H, we can compare the validity of the four methods in estimating H. The results of these computations are shown in Table 3. The computational details may be found in the relevant Excel file included in the data archive for this paper (Munshi, HurstExponentPaperDataArchive, 2015).

Noise	GRG	C	Predicted-H	Actual-H	Error	R-squared
0.12	0	1	0.4901	0.5	0.0198	0.8963
0.12	0	0	0.5002	0.5	0.0004	0.8959
0.12	1	1	0.4968	0.5	0.0064	0.9501
0.12	1	0	0.5005	0.5	0.001	0.95
0.13	0	1	0.5214	0.5	0.0428	0.8938
0.13	0	0	0.5009	0.5	0.0018	0.8924
0.13	1	1	0.5467	0.5	0.0934	0.9311
0.13	1	0	0.5108	0.5	0.0216	0.9251
0.14	0	1	0.5183	0.5	0.0366	0.8769
0.14	0	0	0.5007	0.5	0.0014	0.8759
0.14	1	1	0.5449	0.5	0.0898	0.8967
0.14	1	0	0.5095	0.5	0.019	0.8911
0.15	0	1	0.4979	0.5	0.0042	0.8499
0.15	0	0	0.4999	0.5	0.0002	0.8499
0.15	1	1	0.4808	0.5	0.0384	0.922
0.15	1	0	0.4989	0.5	0.0022	0.92
0.16	0	1	0.4942	0.5	0.0116	0.8303
0.16	0	0	0.4996	0.5	0.0008	0.8302
0.16	1	1	0.5104	0.5	0.0208	0.912
0.16	1	0	0.5027	0.5	0.0054	0.9117
0.17	0	1	0.4839	0.5	0.0322	0.8061
0.17	0	0	0.4991	0.5	0.0018	0.8053
0.17	1	1	0.4664	0.5	0.0672	0.8763
0.17	1	0	0.497	0.5	0.006	0.8707
0.18	0	1	0.4873	0.5	0.0254	0.7889
0.18	0	0	0.4989	0.5	0.0022	0.7884
0.18	1	1	0.5739	0.5	0.1478	0.8567
0.18	1	0	0.5111	0.5	0.0222	0.8418
0.19	0	1	0.4476	0.5	0.1048	0.745
0.19	0	0	0.4975	0.5	0.005	0.7355
0.19	1	1	0.4181	0.5	0.1638	0.7867
0.19	1	0	0.4872	0.5	0.0256	0.7687
0.2	0	1	0.4709	0.5	0.0582	0.7383
0.2	0	0	0.4978	0.5	0.0044	0.7358
0.2	1	1	0.4212	0.5	0.1576	0.7117
0.2	1	0	0.4916	0.5	0.0168	0.6817

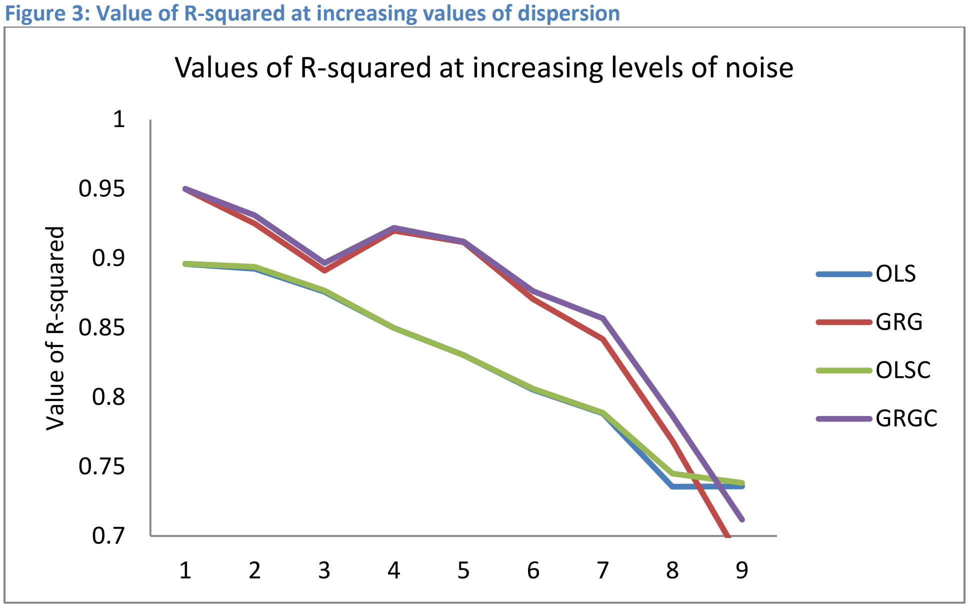
Table 3: Values of H and R-squared for synthetic data

It was found by trial and error that a noise range of 12% to 20% included the entire range of R-squared values observed in the empirical data for stocks. Thus, values of H and R-squared are computed for each of nine levels of random dispersion of the data and for each of four estimation methods. The nine levels of dispersion are shown in the column labeled Noise in Table 3. The error of each estimation is computed as the absolute value of the difference between the estimated value of H and the known value of H=0.5 and is reported in Table 3 as a fraction of the known value of H=0.5. The columns labeled GRG and C in Table 3 contain dummy coding to identify the estimation method. The values for GRG and C in Table 3 may be interpreted as 1=yes and 0=no. These values relate to Table 1 according to:

GRG=0, C=1 => OLSC => Method 1. OLS is used and the constant C is included in the estimation model.
GRG=1, C=1 => GRGC => Method 2. GRG is used and the constant C is included in the estimation model.
GRG=0, C=0 => OLS => Method 3. OLS is used and the constant C is left out of the estimation model.
GRG=1, C=0 => GRG => Method 4. GRG is used and the constant C is left out of the estimation model.

The results in Table 3 may be viewed graphically in Figure 3, Figure 4, and Figure 5[6]. Figure 3 supports the intuition that the effect of increasing levels of dispersion is to lower the value of R-squared. Figure 3 also contains the interesting information that the insertion of the constant C into the model does not appear to have a significant effect on R-squared. Instead what we find is that R-squared is generally higher for GRG models than they are for OLS models.

Figure 3: Value of R-squared at increasing values of dispersion



[6] These graphs are included in the data archive for this paper (Munshi, HurstExponentPaperDataArchive, 2015).

Figure 4: Predicted value of H at increasing levels of dispersion

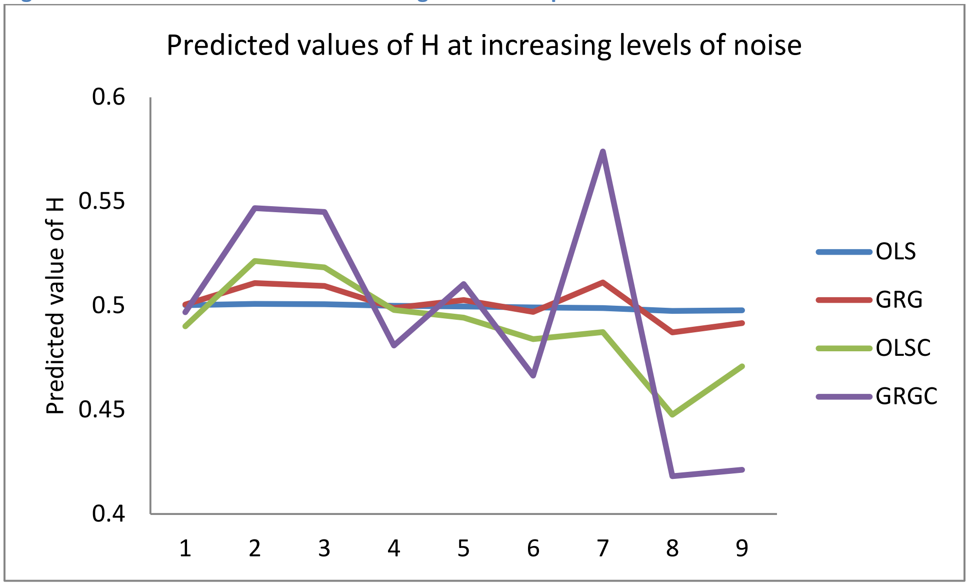
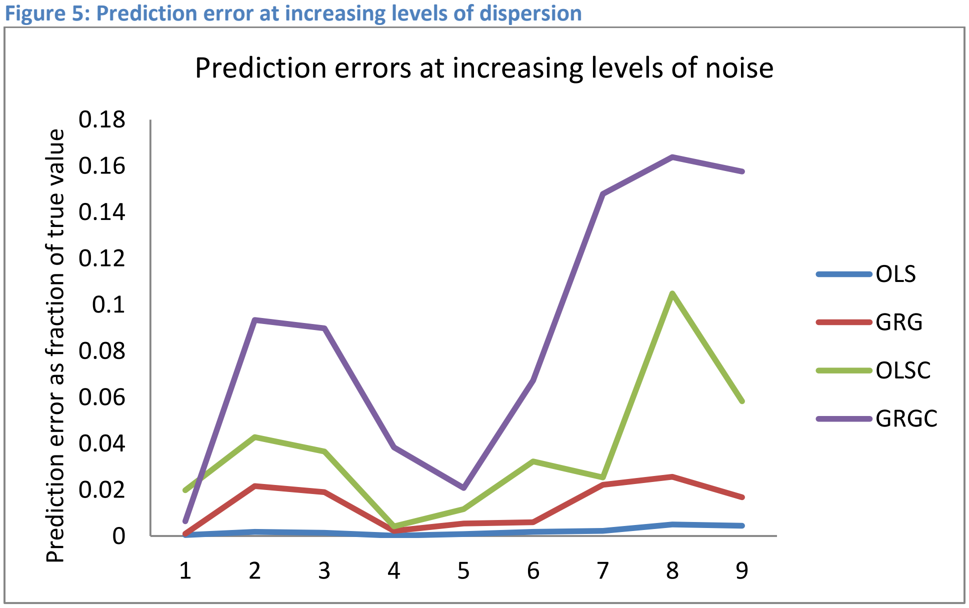


Figure 4 shows that OLSC an GRGC are rather unstable as estimators of H with large swings in value. The relative instability of the C-models increases sharply at higher levels of dispersion. In contrast, the OLS an GRG estimation methods appear to be relatively stable with OLS displaying a more stability than GRG.

Figure 5: Prediction error at increasing levels of dispersion



In terms of prediction error, we find that OLS consistently generates the best estimation and GRGC the worst although GRG performs better than OLSC. The data appears to indicate the insertion of the constant C into Hurst's equation presents a serious problem in the estimation of The Hurst exponent.

A general linear model is used to assess the effect of Noise, GRG, and C on Error. The results are shown in Table 4 and the computational details are available in the synthetic data Excel file included in the data archive for this paper (Munshi, HurstExponentPaperDataArchive, 2015).

	Coeff	*SE*	*t Stat*	*P-value*
Intercept	-0.1023411	0.0334995	-3.0549999	0.0045118
Noise	0.59175	0.2016436	2.9346335	0.0061334
GRG	0.0306333	0.0104128	2.9418837	0.0060219
C	0.0546111	0.0104128	5.2445986	9.754E-06

Table 4: General linear model for Error

As in the previous paper we set our error rate for hypothesis tests at $\alpha=0.001$ to be consistent with normal practice in Bayesian Statistics (Johnson, 2013). In strict conformance with $\alpha=0.001$ we find that the only significant effect is that of C. On average the error in estimating H is higher by 0.0546 when the constant C is inserted into Hurst's equation. On this basis and in view of our analysis of Figures 3, 4, and 5, we can eliminate Method-1 and Method-2 (OLSC, GRGC) from consideration. Of the remaining two methods, the greater stability of OLS when compared with GRG[7] appears to support the use of OLS for the empirical estimation of The Hurst exponent.

Accordingly we reproduce the relevant section of Table 2 in Table 4 as our best estimate of H for the eight stock return series. The high p-values in the last row show clearly that none of these values of H stands as evidence of persistence or of long term memory in stock returns and that these data do not serve as evidence against the efficient market hypothesis (EMH).

Model	Method		IXIC	DJI	BAC	CL	SPX	CAT	BA	IBM
R/S=H*ln(n)	OLS	H	0.5192	0.5067	0.5205	0.4925	0.5332	0.5073	0.5366	0.5416
		R2	0.8624	0.8639	0.8498	0.8232	0.8356	0.8170	0.8413	0.8390
		sigma	0.0270	0.0257	0.0303	0.0247	0.0330	0.0284	0.0320	0.0311
		H-0.5	0.0192	0.0067	0.0205	-0.0075	0.0332	0.0073	0.0366	0.0416
		t-statistic	0.7110	0.2607	0.6772	0.3034	1.0052	0.2573	1.1429	1.3372
		p-value	0.4798	0.7952	0.5008	0.7626	0.3187	0.7978	0.2575	0.1860

Table 5: Our best estimate for values of H for the eight stock series

[7] Figure 1, Figure 4, and Figure 5

4. CONCLUSIONS

The usual procedure for the empirical estimation of The Hurst exponent in stock returns is to insert the constant C into the Hurst equation and then to use logarithmic transformation and OLS regression. Our study shows that the use of the constant C corrupts the estimation and is the probable cause of a large number of spurious findings in this field of research. We also find that although the logarithmic transformation may introduce a bias into the data, the OLS results are still superior to numerical methods such as the generalized reduced gradient method. We propose that The Hurst exponent in time series data should be estimated by a logarithmic transformation of Hurst's equation to ln(R/S) = H*ln(√) where R/S is the rescaled range and √ is the sub-sample size. The value of H can then be estimated using OLS regression with the y-intercept restricted to $b_0 = 0$.

5. REFERENCES

Draper, N. (1998). *Applied regression analysis.* Wiley.

Greene, W. (2010). *Nonlinear regression models.* Retrieved 2015, from stern.nyu.edu: http://people.stern.nyu.edu/wgreene/MathStat/GreeneChapter7.pdf

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 770.

Johnson, V. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, http://www.pnas.org/content/110/48/19313.full.

Lasdon, L. (1973). *Nonlinear optimiation.* PN.

Lasdon, L. (1973). *Nonlinear optimization using the generalized reduced gradient method.* Retrieved 2015, from dtic.mil: http://www.dtic.mil/dtic/tr/fulltext/u2/774723.pdf

Lee, H.-T. (2004). *A study of the generalized reduced gradient method with different search directions.* Retrieved 2015, from http://joqm.ctu.edu: http://joqm.ctu.edu.tw/Download/joqm/930920--QM1-3.pdf

Microsoft. (2015). *Microsoft.com*. Retrieved 2015, from Microsoft.com: http://microsoft.com

Munshi, J. (2014). *There is no chaos in stock markets.* Retrieved 2015, from ssrn.com: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2448648

Munshi, J. (2015). *HurstConstantPaperDataArchive.* Retrieved 2015, from dropbox.com: https://www.dropbox.com/sh/5ba54wlc5zh2h2u/AACNhdIw9pJcUGI99ybXwE_Pa?dl=0

University of Pittsburg. (2015). *Reduced gradient methods.* Retrieved 2015, from pitt.edu: http://www.pitt.edu/~jrclass/opt/notes5.pdf

Zeng, W. S. (2011). *Bias correction in logarithmic regression.* Retrieved 2015, from researchgate.net: https://www.researchgate.net/publication/262877343_BIAS_CORRECTION_IN_LOGARITHMIC_REGRESSION_AND_COMPARISON_WITH_WEIGHTED_REGRESSION_FOR_NON-LINEAR_MODELS