createspace an Amazon.com company

Digital Proofer

Nonlinear Nonparamet...

Authored by Mr. Fred Viole

8.0" x 10.0" (20.32 x 25.40 cm)
Color on White paper
200 pages

ISBN-13: 9781490523996
ISBN-10: 1490523995

Please carefully review your Digital Proof download for formatting, grammar, and design issues that may need to be corrected.

We recommend that you review your book three times, with each time focusing on a different aspect.

1 Check the format, including headers, footers, page numbers, spacing, table of contents, and index.

2 Review any images or graphics and captions if applicable.

3 Read the book for grammatical errors and typos.

Once you are satisfied with your review, you can approve your proof and move forward to the next step in the publishing process.

To print this proof we recommend that you scale the PDF to fit the size of your printer paper.

NONLINEAR NONPARAMETRIC STATISTICS: Using Partial Moments

Fred Viole

David Nawrocki

© 2013 Viole & Nawrocki. All Rights Reserved

Table of Contents

Foreword

This book introduces a toolbox of statistical tools using partial moments that are both old and new. Partial moment analysis is over a century old but most applications of partial moments have not progressed beyond a substitution for simple variance analysis. Lower partial moments have been in use in finance in portfolio investment theory for over 60 years. However, just as the normal distribution and the variance leads the statistician into linear correlation and regression analysis, partial moments leads us towards nonlinear correlation and regression analysis. Using partial moments as a variance measure is only the tip of the iceberg as the purpose of this book is to explore the entire iceberg.

This partial moment toolbox is the "new" presented in this book. However, "new" always should have some advantage over "old". The advantage of using partial moments is that it is nonparametric and does not require the knowledge of the underlying probability function nor does it require a "goodness of fit" analysis. Partial moments provide us with cumulative density functions, probability density functions, linear correlation and regression analysis, nonlinear correlation and regression analysis, ANOVA, and ARMA/ARCH models. This new toolbox is completely nonparametric and provides a full set of probability hypothesis testing tools without knowing the underlying probability distribution.

In this new advanced approach to nonparametric statistics, we merge the ideas of discrete and continuous processes and present them in a unified framework predicated on partial moments. Through the asymptotic property of partial moments, we show the two schools of mathematical thought do not converge as commonly envisioned. The increased observations approximate the continuous area of a function; versus stabilizing on a discrete counting metric. However, it remains a strictly binary analysis: discrete or continuous. The known properties generated from this continuous vs. discrete analysis affords an assumption free analysis of variance (ANOVA) on multiple distributions.

In our correlation and regression analysis, linear segments are aggregated to describe a nonlinear system. The computational issue is the weighting of the segments. However, since partial moments weigh all observations this consideration is alleviated, ultimately yielding a more robust result with no "butterfly effect" due to our lack of parameters. By building off basic relationships between variables, we are able to perform multivariate analysis with ease and transform "complexity" into "tedious." One major advantage with our work is that the partial moment methodology fully replicates linear conditions or known functions. This trust of methodology is important for transition to chaotic unknowns and forecasting with autoregressive models.

Normalization of data has the unintended consequence of transforming continuous variables to discrete variables while eliminating prior relationships. We present a normalization method that enables a truly apples to apples comparison that retains the finite moment properties of the underlying distribution. In the ensuing analysis of the question variables, we illustrate the distinction between correlation and causation. Using this distinction we offer a definition of causation that integrates historical correlation with conditional probabilities.

Finally, linearity should be a pleasant surprise to encounter in data, not a prerequisite. By eliminating all preconceptions and assumptions, we offer a powerful framework for statistical analysis. The simple nonparametric architecture based on partial moments yields important information to easily conduct multivariate analysis; generating descriptive and inferential statistics for a nonlinear world.

*** All of the functions in this book are available in the R-package 'NNS' available on CRAN: https://cran.r-project.org/web/packages/NNS/

ASYMPTOTICS

$f(\boldsymbol{Newton})$

Abstract

We define the relationship between integration and partial moments through the integral mean value theorem. The area of the function derived through both methods share an asymptote, allowing for an empirical definition of the area. This is important in that we are no longer limited to known functions and do not have to resign ourselves to goodness of fit tests to define f(x). Our empirical method avoids the pitfalls associated with a truly heterogeneous population such as nonstationarity and estimation error of the parameters. Our ensuing definition of the asymptotic properties of partial moments to the area of a given function enables a wide array of equivalent comparative analysis to linear and nonlinear correlation analysis and calculating cumulative distribution functions for both discrete and continuous variables.

"*Imagine how much harder physics would be if electrons had feelings*." - Richard Feynman

INTRODUCTION

Modern finance has an entrenched relationship with calculus, namely in the fields of risk and portfolio management. Calculus by definition is the study of limits and infinitesimal series. However, given the seemingly infinite amount of financial data available we ask the question whether calculus is too restrictive.

In order to utilize the powerful tools of calculus, a function of a continuous variable must be defined. Least squares methods and families of distributions have been identified over the years to assist in this definition prerequisite. Once classified, variables can be analyzed over specific intervals. Comparison of these intervals between variables is also possible by normalizing the area of that interval.

Unfortunately, there are major issues with each of the identified steps of the preceding paragraph. When defining a continuous variable, you are stating that its shape (via parameters) is fixed in stone (stationary). Least squares methods of data fitting make no distinction whether a residual is above or below the fitted value, disregarding any implications thereof. And finally, normalization of continuous variables has been shown to generate discrete variable solutions [1].

Given these formidable detractions, we contend that a proper asymptotic approximation of a function's area "is a better fit" to its intended applications. Parsing variances into positive or negative from a specified point is quite useful for nonlinear correlation coefficients and multiple nonlinear regressions as demonstrated in [2]; and calculating cumulative distribution functions for both discrete and continuous variables [1].

Furthermore, the multiple levels of heterogeneity present in the market structure negate the relevance of true population parameters estimated by the classical parametric method. Estimation error and nonstationarity of the first moment, μ are testaments to the underlying heterogeneity issue; leaving the nonparametric approach as the only viable solution for truly heterogeneous populations. Our ensuing definition of the asymptotic properties of partial moments to the area of a given function enables a wide array of equivalent comparative analysis to the classical parametric approach.

OUR PROPOSED METHOD

Integration and differentiation have been important tools in defining the area under a function $(f(x))$ since their identification in the 17th century by Isaac Newton and Gottfried Leibniz. Approximation of this area is possible empirically with the lower and upper partial moments of the distribution presented in equations 1 and 2.

$$LPM(n,h,x) = \frac{1}{T}\sum_{t=1}^{T}\{\max(h - x_t), 0\}^n \tag{1}$$

$$UPM(q,l,x) = \frac{1}{T}\sum_{t=1}^{T}\{\max(x_t - l), 0\}^q \tag{2}$$

Where x_t is the observation of variable x at time t, h and l are the targets from which to compute the lower and upper deviations respectively, and n and q are the weights to the lower and upper deviations respectively. We set $n, q = 1$ and $h = l$ to calculate the continuous area of the function as demonstrated in [1].

Partial moments resemble the Lebesgue integral, given by

$$f^-(x) = \max(\{-f(x), 0\}) = \begin{cases} -f(x), & if\ f(x) < 0, \\ 0, & otherwise, \end{cases} \tag{3}$$

$$f^+(x) = \max(\{f(x), 0\}) = \begin{cases} f(x), & if\ f(x) > 0, \\ 0, & otherwise. \end{cases} \tag{4}$$

In order to transform the partial moments from a time series to a cross-sectional dataset where x is a real variable, we need to alter equations 1 and 2 to reflect this distinction and introduce the interval $[a,b]$ for which the area is to be computed.

$$LPM(1,0,f(x)) = \frac{1}{n}\sum_{i=1}^{n}\{\max(-f(x_i)), 0\} \qquad if\ x \in [a,b], \tag{5}$$

$$UPM(1,0,f(x)) = \frac{1}{n}\sum_{i=1}^{n}\{\max(f(x_i)), 0\} \qquad if\ x \in [a,b]. \tag{6}$$

We further constrained equations 5 and 6 by setting the target equal to zero for both functions and consider the total number of observations n, rather than the time qualification *T*. The target for the transformed partial moment equations will be a horizontal line, in this instance zero (x-axis); whereby all $f(x) > 0$ are positive and all $f(x) < 0$ are negative area considerations, per the Lebesgue integral in equations 3 and 4.

Lebesgue integration also offers flexibility versus its Riemann counterpart; just as partial moments offer flexibility versus the standard moments of a distribution. Equation 7 illustrates the asymptotic nature of the partial moments as the number of observations tends towards infinity over the interval [*a*,*b*].[1] *This is analogous to the number of irregular rectangle partitions in other numerical integration methods*.

$$\lim_{n\to\infty}[UPM\big(1,0,f(x)\big) - LPM(1,0,f(x))] = \frac{\int_a^b f(x)dx}{(b-a)} \tag{7}$$

Using the proof of the second fundamental theorem of calculus we know

$$F(b) - F(a) = \int_a^b f(x)dx.$$

Yielding,

$$\lim_{n\to\infty}[UPM\big(1,0,f(x)\big) - LPM(1,0,f(x))] = \frac{F(b)-F(a)}{(b-a)} \tag{8}$$

[1] Detailed examples are offered in Appendix A.

Invoking the mean value theorem, where

$$F'(c) = \frac{F(b)-F(a)}{(b-a)} \tag{9}$$

We have

$$F'(c) = \lim_{n\to\infty}[UPM\big(1,0,f(x)\big) - LPM\big(1,0,f(x)\big)] \tag{10}$$

$F'(c)$ using Δx of partition i per the integral mean value theorem shows that

$$F'(c) = \lim_{||\Delta x_i||\to 0}\sum_{i=1}^{n}[f(c_i)(\Delta x_i)] \tag{11}$$

Thus demonstrating the inverse relationship involving:

(i) the distance between irregular rectangle partitions (Δx_i)
(ii) the number of observations (n)

$$\lim_{||\Delta x_i||\to 0}\sum_{i=1}^{n}[f(c_i)(\Delta x_i)] = \lim_{n\to\infty}[UPM\big(1,0,f(x)\big) - LPM\big(1,0,f(x)\big)] \tag{12}$$

Just as integrated area sums converge to the integral of the function with increased rectangle areas partitioned over the interval of $f(x)$,[2] equation 7 shares this asymptote

[2] Provided F is differentiable everywhere on [*a*,*b*] and F' is integrable on [*a*,*b*]. The partial moment term of the equality in equation 12 makes no such suppositions. The total area, not just the definite integral is simply $\left|\int_a^b f(x)dx\right| = \lim_{n\to\infty}[UPM\big(1,0,f(x)\big) + LPM\big(1,0,f(x)\big)]$

equal to the integral of the function. This is demostrated above with equation 12. ***If one can define the function of the asymptotic areas $F'(c)$ (UPM+LPM), then one can find the asymptote or integral of the function directly from observations.***

FINDING THE HORIZONTAL ASYMPTOTE

The horizontal asymptote is the horizontal line that the graph of $F'(c)$ as $n \rightarrow \infty$. This asymptote is equal to $[F(b) - F(a)]/(b - a)$ for the interval $[a,b]$ where $a < b$.

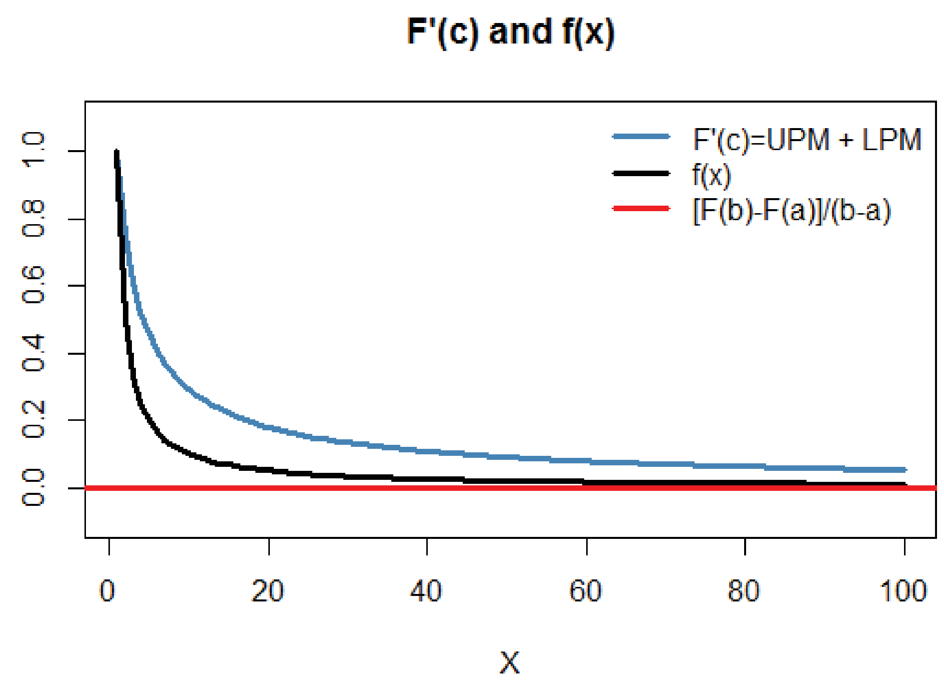


Figure 1. Asymptote of $= \frac{1}{x}$. As the range of the interval increases, we can fit $F'(c)$ or $f(x)$ to determine the asymptote.

Once $F'(c)$ is defined, we can use the method of leading coefficients to determine the horizontal asymptote. Figure 1 above has a horizontal asymptote of zero. However, once $F'(c)$ is defined the dominant assumption is that of stationarity of function parameters at time *t*. Integral calculus is not immune from this stationarity assumption as $f(x)$ needs to be defined in order to integrate and differentiate. Since we are not defining $f(x)$, we have the luxury of recalibrating with each data point to capture the nonstationarity; consequently updating $F'(c)$.

Goodness of fit tests also assume a stationarity on the parameters; detracting from its appeal as a reason to define a function.

DISCUSSION

To define, or not to define: that is the question. If we define $F'(c)$ we can find the exact asymptote, thus area of $f(x)$. If we appreciate the fact that nothing in finance seems to be guided by an exactly defined function, the measured area of $f(x)$ over the interval $[a,b]$ will likely change over time due to the multiple levels of heterogeneity present in the market structure.

Furthermore, if we are going to expand the extra effort to define a function (within tolerances mind you, not an exact fit), does it really matter which function is defined $F'(c) or\ f(x)$? The next observation may very well lead to a redefinition.

Our proposed method of closely approximating the area of a function over an interval with partial moments is an important first step in enjoining flexibility into finance versus integral calculus. We shed the dependence on stationarity, and alleviate the need for goodness of fit tests for underlying function definitions. *Moreover, if the underlying process is stationary then simply increasing the number of observations will ensure a convergence of methods*.

We are hopeful over time this method will be refined and expanded in order to bring a more robust and precise method of analysis then currently enjoyed; while avoiding the pitfalls associated with the parametric approach on a truly heterogeneous population.

APPENDIX A: EXAMPLES OF KNOWN FUNCTIONS USING EQUATION 7

$f(x) = x^2$

To find the area of the function over the interval [0,10] for $f(x) = x^2$, we differentiate according to x yielding $F(x) = \frac{x^3}{3}$. $F(10) - F(0) = \frac{1000}{3} - 0 = 333.33$

Using equation 7 in the 'NNS' package in R, we know $F'(c)$ should converge to $\frac{333.33}{10}$ *or* 33.33.

```
> x=seq(0,10,1);y=x^2;UPM(1,0,y)-LPM(1,0,y)
[1] 35
> x=seq(0,10,.1);y=x^2;UPM(1,0,y)-LPM(1,0,y)
[1] 33.5
> x=seq(0,10,.02);y=x^2;UPM(1,0,y)-LPM(1,0,y)
[1] 33.36667
> x=seq(0,10,.01);y=x^2;UPM(1,0,y)-LPM(1,0,y)
[1] 33.35
```

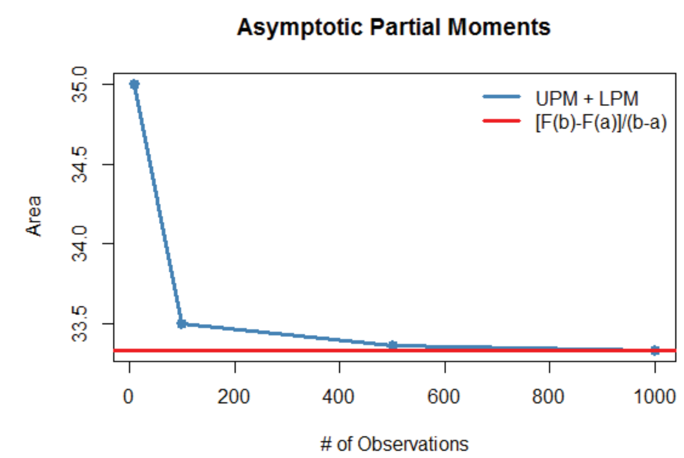


Figure 2. Asymptotic partial moment areas for $\int_0^{10} x^2 dx$.

<u>$f(x) = \sqrt{x}$</u>

To find the area of the function over the interval [0,10] for $f(x) = \sqrt{x}$, we differentiate according to x yielding $F(x) = \frac{2x^{\frac{3}{2}}}{3}$. $F(10) - F(0) = \frac{63.245}{3} - 0 = 21.08$

Using equation 7 in the 'NNS' package in R, we know $F'(c)$ should converge to $\frac{21.08}{10}$ or 2.108.

```
> x=seq(0,10,1);y=sqrt(x);UPM(1,0,y)-LPM(1,0,y)
[1] 2.042571
> x=seq(0,10,.1);y=sqrt(x);UPM(1,0,y)-LPM(1,0,y)
[1] 2.102329
> x=seq(0,10,.02);y=sqrt(x);UPM(1,0,y)-LPM(1,0,y)
[1] 2.107075
> x=seq(0,10,.01);y=sqrt(x);UPM(1,0,y)-LPM(1,0,y)
[1] 2.107638
```

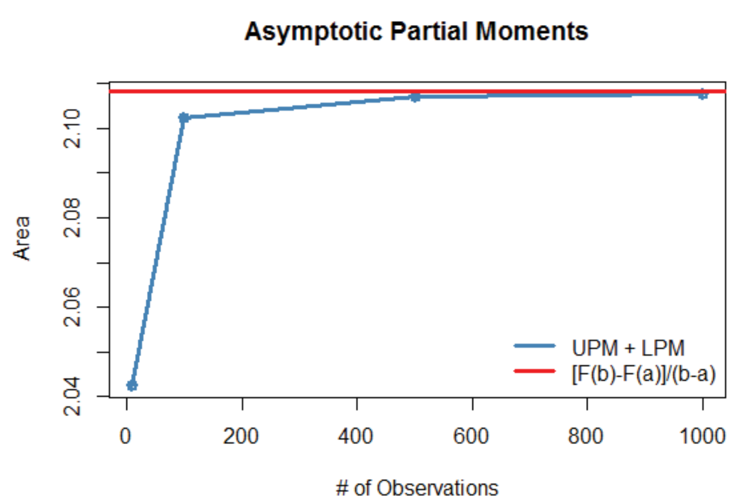


Figure 3. Asymptotic partial moment areas for $\int_0^{10} \sqrt{x}\, dx$.

APPENDIX B: <u>PERFECT UNIFORM SAMPLE ASSUMPTION</u> $\left(\lim\limits_{||\Delta x_i|| \to 0} = \lim\limits_{n \to \infty}\right)$

We can see from an analysis of samples over the interval [0,100] as the number of observations tends towards ∞, the observations approach a perfect uniform sample in Figure 1b. However, when using a sample representing irregular partitions, (more realistic of observations than completely uniform) the length of observations required to achieve perfect uniformity is greater than by assuming it initially. This condition speaks volumes to misinterpretations of real world data when limit conditions are used as an artifact of fitting distributions.

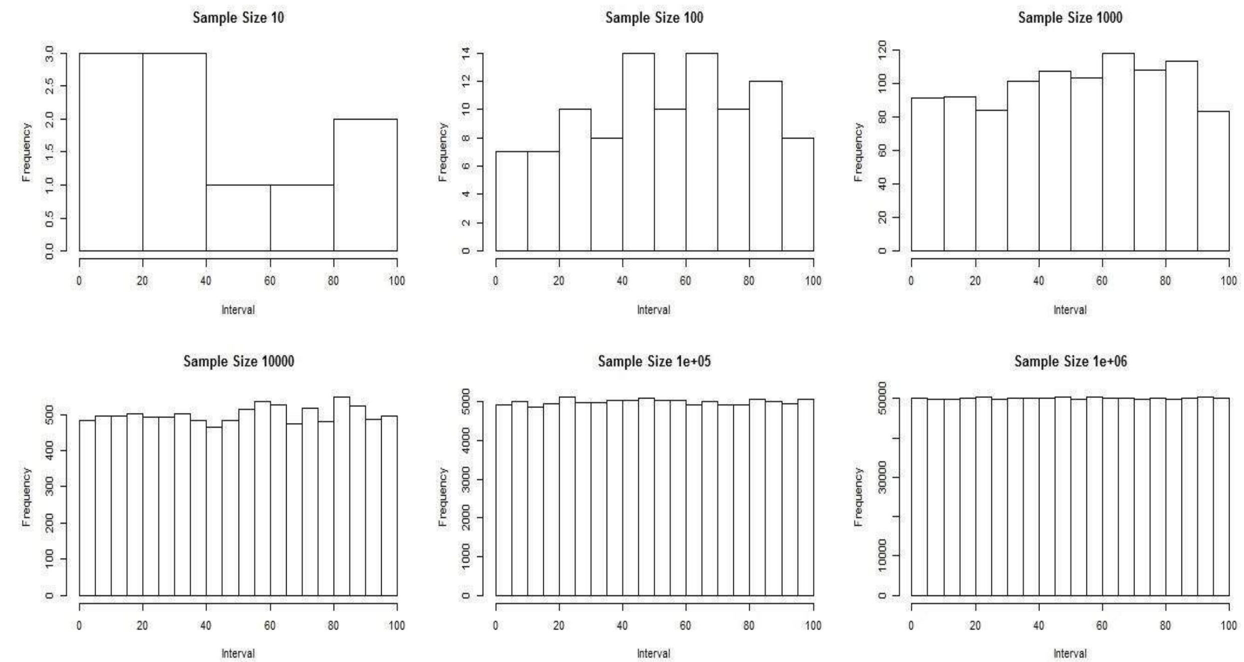


Figure 1b. Randomly generated uniform sample over the interval approaches perfect uniform as number of observations goes to infinity.

DISCRETE VS. CONTINUOUS DISTRIBUTIONS

Cumulative Distribution Functions and UPM/LPM Analysis

Abstract

We show that the Cumulative Distribution Function (CDF) is represented by the ratio of the lower partial moment (LPM) ratio to the distribution for the interval in question. The addition of the upper partial moment (UPM) ratio enables us to create probability density functions (PDF) for any function without prior knowledge of its characteristics. We are able to replicate discrete distribution CDFs and PDFs for normal, uniform, poisson, and chi-square distributions, as well as true continuous distributions. This framework provides a new formulation for UPM/LPM portfolio analysis using co-partial moment matrices which are positive symmetrical semi-definite, aggregated to yield a positive symmetrical definite matrix.

I. Introduction:

The Empirical Cumulative Distribution Function (EDF) should, most of the time, be a good approximation of the true cumulative distribution function (CDF) as the sample set increases. This generalization is at the heart of statistics. Means and variances are used to assign and fit a distribution, but partial moments stabilize with a smaller sample size ensuring a more accurate analysis of the EDF.

The empirical CDF is a simple construct. It is simply the number of observations less than or equal to a target, divided by the total number of observations in a given data set. The problem with extrapolating these results to an assumed true CDF is that the discrete empirical CDF is extremely sensitive to sample size,[3] and any parameter nonstationarity will deteriorate the fit to the true distribution. The paper is organized as follows:

First, we propose a method to derive the CDF and PDF of the EDF, utilizing the upper and lower partial moments (*UPM* and *LPM* respectively) of the EDF. The benefits are obvious, such as compensating for any observed skewness and kurtosis that would force a more esoteric distribution family onto the data. These measurements require zero knowledge of the underlying function and no goodness-of-fit tests to approximate a likely true distribution. Partial moments also happen to exhibit less sample size sensitivity than means and variances as we will discuss later.

Next, this foundation is then used to develop conditional probabilities and joint distribution co-partial moments. Finally, this toolbox allows us to propose a new

[3] Estimated mean average deviations are provided in Appendix A.

formulation for UPM/LPM analysis and we note that each of the co-partial moment matrices are positive symmetrical semi-definite, ensuring a positive symmetrical definite aggregate matrix. This represents a major improvement in the use of partial moment matrices in portfolio theory and avoids the problems with co-semivariance matrices as noted by Grootveld and Hallerbach (1999) and Estrada(2008).

II. Deriving Cumulative Distribution and Partial Density Functions Using Partial Moments

A distribution may be dissected into two partial moment segments using an arbitrary target as shown in Figure 1.

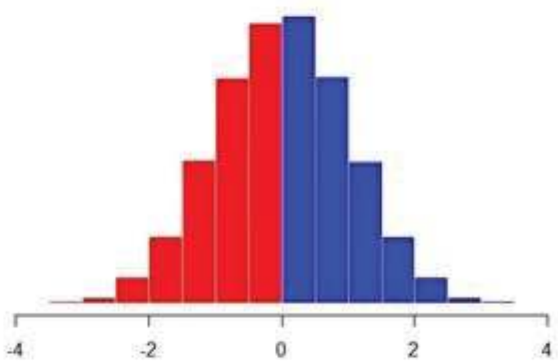


Figure 1. A distribution dissected into its two partial moment segments, red LPM and blue UPM, from a shared target.

The Upper and Lower partial moment formulas are below in Equations 1 and 2:

$$LPM(n,h,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, h - x_t\}^n\right] \tag{1}$$

$$UPM(q,l,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, x_t - l\}^q\right] \tag{2}$$

where x_t represents the observation *x* at time *t*, *n* is the degree of the LPM, *q* is the degree of the UPM, *h* is the target for computing below target returns, and *l* is the target for computing above target returns.[4]

One can visualize how the entire distribution is quantified with the upper and lower partial moment from the same target, (*h* = *l* = *0*) in Figure 1. The area under the function derived from degree one partial moments will approximate the area derived from the integral of the function over an interval [*a*,*b*] asymptotically. This asymptotic numerical integration is shown in Viole and Nawrocki (2012c) and represented with equation (3).

$$\lim_{t\to\infty}[UPM(1,0,f(x)) - LPM(1,0,f(x))] = \frac{\int_a^b f(x)dx}{(b-a)} \tag{3}$$

We use a degree zero (*n*=*q*=0) to generate a discrete analysis, replicating results from the conventional CDF and PDF methodology. Degree one (*n*=*q*=1) is used to generate the continuous results. This is an important distinction, as the discrete analysis is a

[4] Equations 1 and 2 will generate a 0 for degree 0 instances of 0 results.

relative frequency and probability investigation; while the continuous analysis integrates a variance consideration to capture the rectangles of infinitesimal width in deriving an area under a function. Standard deviation remains stable as sample size range increases, thus it is not an accurate barometer of the area of the function to estimate a continuous variable. Figure 2 illustrates the range increase as the number of observations increase for a normal distribution with μ=10 and σ=20 for 5 million random draws from a normal distribution.

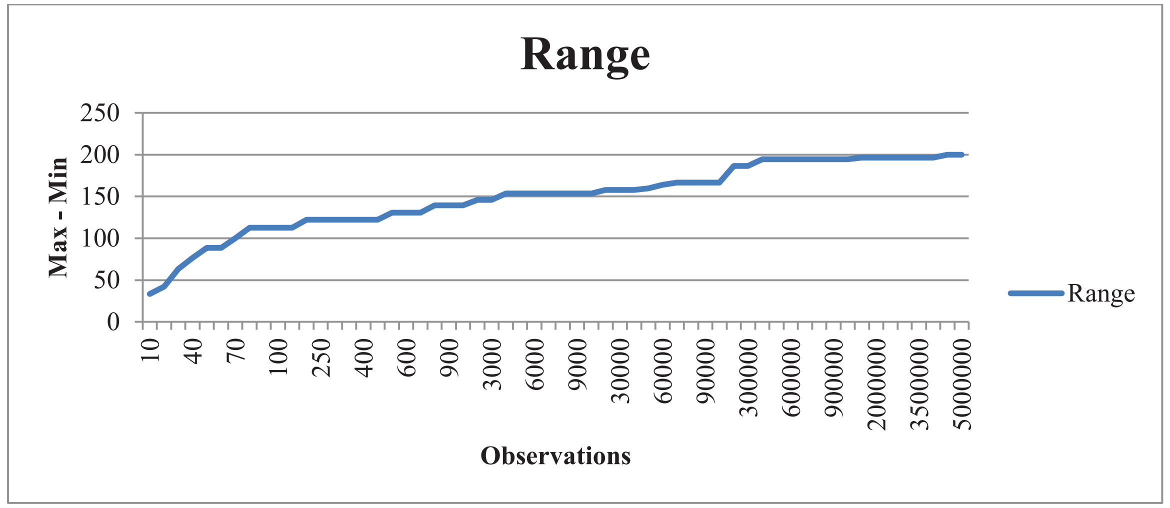


Figure 2. Range for a randomly generated normal distribution μ=10 and σ=20 for 5 million random draws.

Just as the probability of two mutually exclusive events equal one, the sum of the ratios - LPM to the entire distribution; and UPM to the entire distribution (LPM_{ratio} and UPM_{ratio} respectively) plus the point probability, equal one as in equations 8 and 8a.

The point probability is often included in the CDF calculation but it is not uniformly treated as less than or equal to the target.[5]

Theorem 1,

$$P\{X < x\} + P\{X > x\} + P\{X = x\} = 1 \tag{4}$$

If,

$$P\{X \leq x\} = LPM_{ratio}(0, x, X) = \frac{LPM(0, x, X)}{[LPM(0, x, X) + UPM(0, x, X)]} - \frac{\varepsilon}{2} \qquad (5)^{6}$$

$$LPM_{ratio}(0, x, X) = LPM(0, x, X) \tag{5a}$$

$$LPM_{ratio}(1, x, X) \neq LPM(1, x, X) \tag{5b}$$

And,

$$P\{X > x\} = UPM_{ratio}(0, x, X) = \frac{UPM(0, x, X)}{[LPM(0, x, X) + UPM(0, x, X)]} - \frac{\varepsilon}{2} \tag{6}$$

---

[5] There is no consensus language for CDF definitions. Some instances are "$< x$" while others reference "$\leq x$" depending on the distribution, discrete or continuous. We are uniform in our treatment of distributions with
"$\leq x$" for both discrete and continuous distributions. See
http://www.mathworks.com/help/toolbox/stats/unifcdf.html and
http://www.mathworks.com/help/toolbox/stats/unidcdf.html for treatment of the target, x.
[6] It is important to note that $LPM(0, x, X)$ is a probability measure and will yield a result from 0 to 1. Thus, the ratio of $LPM(0, x, X)$ to the entire distribution ($LPM_{ratio}(0, x, X)$) is equal to the probability measure itself, $LPM(0, x, X)$.

$$UPM_{ratio}(0,x,X) = UPM(0,x,X) \tag{6a}$$

$$UPM_{ratio}(1,x,X) \neq UPM(1,x,X) \tag{6b}$$

Since the entire normalized distribution is represented by,

$$\left[\frac{LPM(0,x,X)}{[LPM(0,x,X)+UPM(0,x,X)]}-\frac{\varepsilon}{2}\right]+\left[\frac{UPM(0,x,X)}{[LPM(0,x,X)+UPM(0,x,X)]}-\frac{\varepsilon}{2}\right]+\varepsilon = 1 \tag{7}$$

Where ε is the point probability $P\{X = x\}$. The use of an empty set for ε yields,

$$LPM(0,x,X) + UPM(0,x,X) = 1 \tag{8}$$

$$LPM_{ratio}(1,x,X) + UPM_{ratio}(1,x,X) = 1 \tag{8a}$$

For a discrete distribution, an empty set for target observations lowers both $LPM(0,x,X)$ and $UPM(0,x,X)$ simultaneously so that Equation 8 still equals one with $LPM(0,x,X) = P\{X \leq x\}$ and $UPM(0,x,X) = P\{X > x\}$. The point probability ε for a discrete distribution can easily be computed by the frequency of the specific point divided by the total number of observations. The point probability would be more relevant in a discrete distribution of integers, and has an inverse relationship to the degree of specification of the underlying variable. As the specification approaches infinity, ε approaches zero.

We know from calculus that the $\int_a^b (x)dx = F(b) - F(a)$ and if $F(b) = F(a)$, the integral of a point equals zero. Thus for a continuous distribution, there is no difference between $P\{X < x\}$and $P\{X \leq x\}$ since $\varepsilon = 0$. If one wishes to subscribe to the notion that the sum of an infinite amount of points each equal to zero must sum to one per the integral definition, then equation 7 is simply reduced to equation 8a for continuous variables. However, equation 7 with degree 1 can also be used for the continuous variable to compensate for $\varepsilon > 0$ and generate a normalized continuous probability.

A. <u>Review of the Literature</u>

Guthoff et al (1997) illustrate how the value at risk of an investment is equivalent to the degree zero LPM. We confirm this derivation as the degree zero LPM does indeed provide a normalized solution. However, critical errors were made by Guthoff and in subsequent works by Shadwick and Keating (2002), and Kaplan and Knowles (2004).

The omega ratio is defined as,

$$\Omega(\tau) = \frac{\int_\tau^\infty [1 - F(R)]dR}{\int_{-\infty}^\tau F(R)dR} \tag{9}$$

Where F(.) is the CDF for total returns on an investment and τ is the threshold return. Guthoff and Shadwick and Keating's error was the use of degree one LPM (area) on a degree 0 LPM, the probability CDF of the distribution. Degree one LPM does not need to be performed on the probability CDF as they present.

The Kappa measure is defined as,

$$K_{\mathrm{n}}(\tau) = \frac{\mu - \tau}{\sqrt[n]{LPM_n(\tau)}} \tag{10}$$

Kaplan and Knowles' error was the dismissal of the degree zero LPM (0-th root of something does not exist) which we show equals historical CDF measurements for various distributions. Also, $\sqrt[n]{LPM_n(\tau)}$ forces concavity upon increased *n*, which do not presume such a condition.

The omega ratio (Shadwick and Keating, 2002) and kappa measure (Kaplan and Knowles, 2004) both demonstrate the need for a full derivation of partial moments and their CDF equivalence with full degree explanation and relevance.

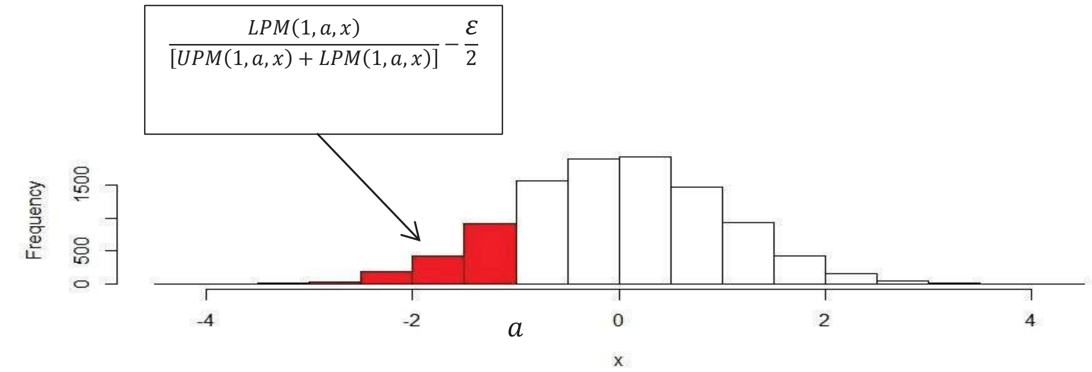


Figure 3. Area of a Probability Density Function represented by the Cumulative Distribution Function of an arbitrary point $\boldsymbol{a}$ for the interval $[-\infty, \boldsymbol{a}]$.

Cumulative Distribution Function (CDF) using partial moments:

$$F_X(x) = P(X \leq x) \tag{11}$$

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{12}$$

Discrete,

$$F(x) = LPM(0, a, x) \tag{13}$$

Continuous,

$$F(x) = LPM_{ratio}(1, a, x) \tag{14}$$

For any distribution the continuous estimate yields,[7]

$$0.5 = LPM_{ratio}(1, \mu, x) \tag{15}$$

[7] Figure 7 offers a visual representation of the difference between continuous and discrete CDFs of the mean.

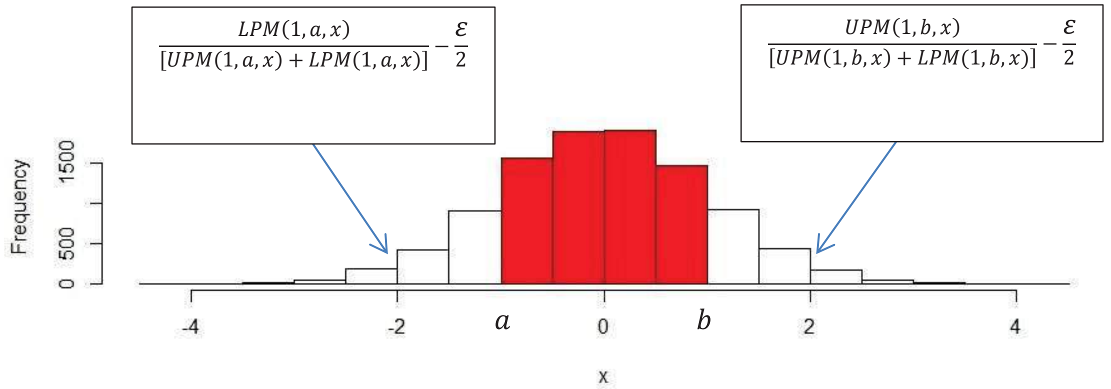


Figure 4. Probability Density Function for the interval $[\boldsymbol{a}, \boldsymbol{b}]$.

Probability Density Function (PDF) using partial moments:

$$P[a \leq x \leq b] = \int_a^b f(x)dx \tag{16}$$

Discrete,

$$P[a \leq x \leq b] = LPM(0,b,x) - LPM(0,a,x) \tag{17a}$$

Continuous,

$$P[a \leq x \leq b] = LPM_{ratio}(1,b,x) - LPM_{ratio}(1,a,x) \tag{17b}$$

B. Methodology Notes:

We generated random distributions for 5 million observations. We then took 300 iterations with different seeds and averaged them. For stability estimates, we generated mean average deviations (MAD) for each statistic over the 300 iterations for observations 30 through 5 million.

The statistics used in the following discussion are as follows: *CHIDF(target)* - Cumulative distribution function for the Chi-square distribution and specified target; *Kurtosis* - Relative Kurtosis measure of the entire sample; *Mean* - μ of the entire sample; *Norm Prob(target)* - Cumulative distribution function for the Normal distribution and specified target; *POIDF(target)* - Cumulative distribution function for the Poisson distribution and specified target; *Range* - Max observation – min observation for the entire sample; *SemiDev* - Semi-deviation of the sample using mean as the target; *Skew* - Skewness measure of the entire sample;

StdDev - Standard deviation of the sample; *UNDF(target)* - Cumulative distribution function for the Uniform distribution.

All of the above mentioned distributions and targets can be easily verified by the reader with statistical software such as the ISML subroutine library. Furthermore, the direct computation of the partial moments can also be easily implemented into such software. The sample parameters generated were as follows:

Normal Distribution: $\mu = 10.00018$ $\sigma = 19.99976$

Poisson Distribution: $\theta = 9.999914$

Uniform Distribution $\mu = 10.00045$

Chi-Square Distribution: $v = 1$ $\mu = 0.999947$

C. Normal Distribution

We compare our metric to the traditional CDF, Φ, of a standard normal random variable.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

The probability generated from the normal distribution converges to $LPM(0,0,X)$ in approximately 90 observations as shown in Figure 5. $LPM(0,0,X)$ stabilizes with less observations than the normal probability (exhibiting a lower MAD) as shown in Appendix A, Table 1a. This is proof that $LPM(0,0,X)$ is indeed the discrete CDF of the distribution for the area less than the target. While the normal probability is less than or equal to the target compared to less than for $LPM(0,0,X)$; the probability of the specific target outcome does not affect the probability to the specification of four decimal places.

The relationship between $LPM_{ratio}(1,0,X)$, $LPM(0,0,X)$ and the normal probability or *Norm Prob(0)* is shown in Figure 5. The further from the mean, the greater the discrepancy between the continuous and discrete CDF as seen in Figure 6. As the area of the distribution increases for the UPM if the target is less than the mean, the continuous CDF will be consistently lower than the discrete CDF. Conversely, as the area of the

LPM increases if the target is greater than the mean, the continuous CDF will be consistently higher than the discrete CDF. This holds for all distribution families. The continuous and discrete probabilities are obviously equal at the endpoints of the distribution, 0 and 1 for minimum and maximum respectively.

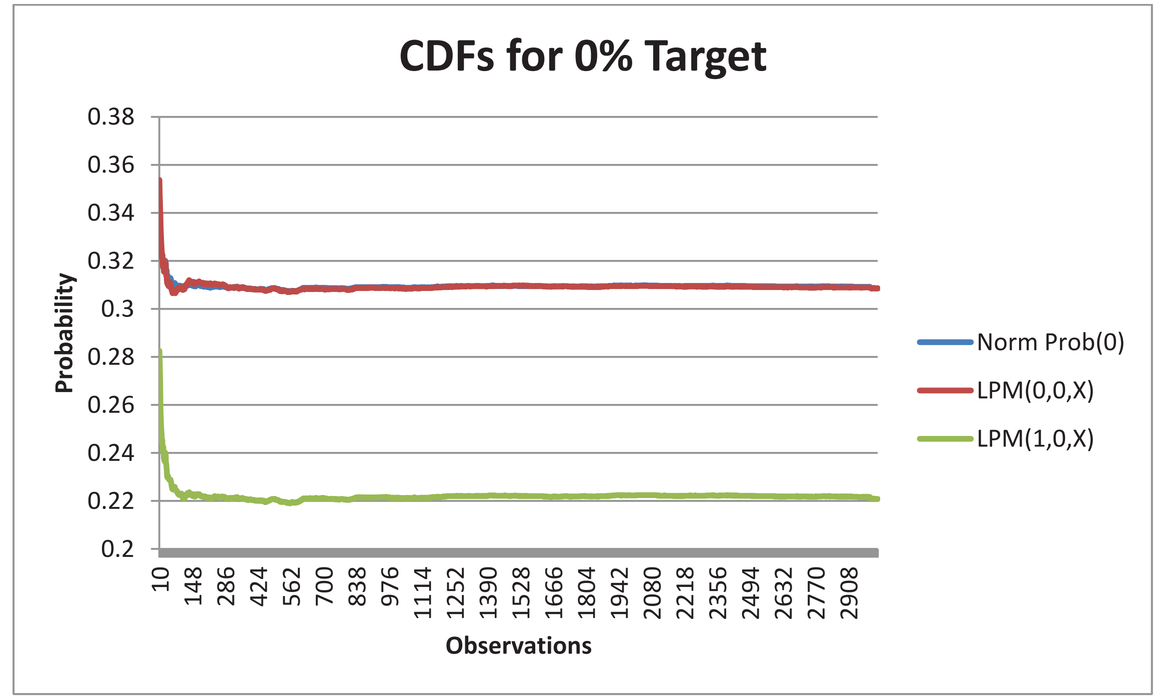


Figure 5. CDF of 0% target for Normal distribution with μ=10 and σ=20 parameter constraints.

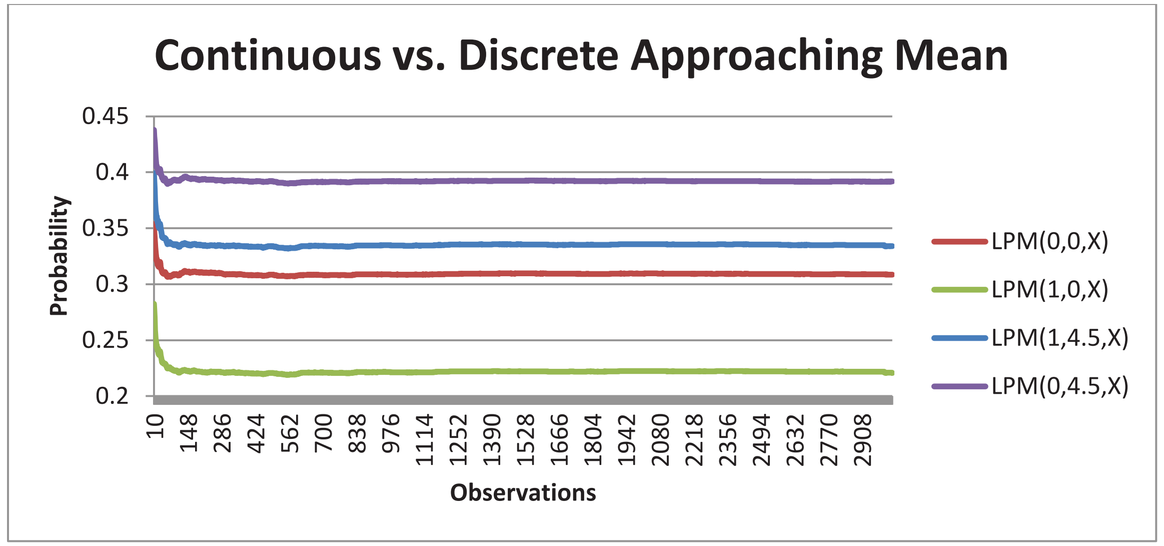


Figure 6. Continuous estimate converges towards discrete estimate as the target approaches sample mean (as *h* is increased from 0 to 4.5). The LPM *n*=0, *h=0* is denoted as LPM(0,0,X), LPM *n*=1, *h*=0) is denoted by LPM(1,0,X), LPM *n*=1, *h*=4.5 is denoted as LPM(1,4.5,X) and the LPM *n*=0, *h*=4.5 is LPM(0,4.5,X).

In Figure 7, the plot shows the convergence of the discrete LPM degree 0 from the mean to the continuous LPM degree 1 using the mean as the target return. The discrete isn't stable until around 1000 observations.

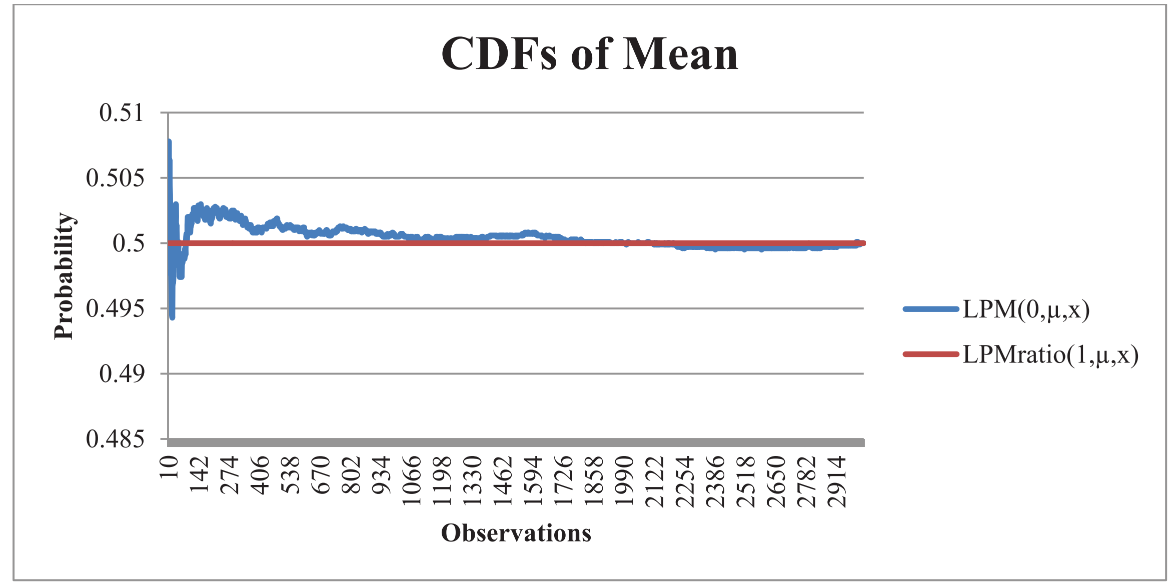


Figure 7. Differences in discrete LPM(0,μ,X) and continuous LPMratio(1,μ,x) CDFs converge when using the mean target for the Normal distribution. $LPM(0,\mu,X) \neq LPM_{ratio}(1,\mu,X)$.

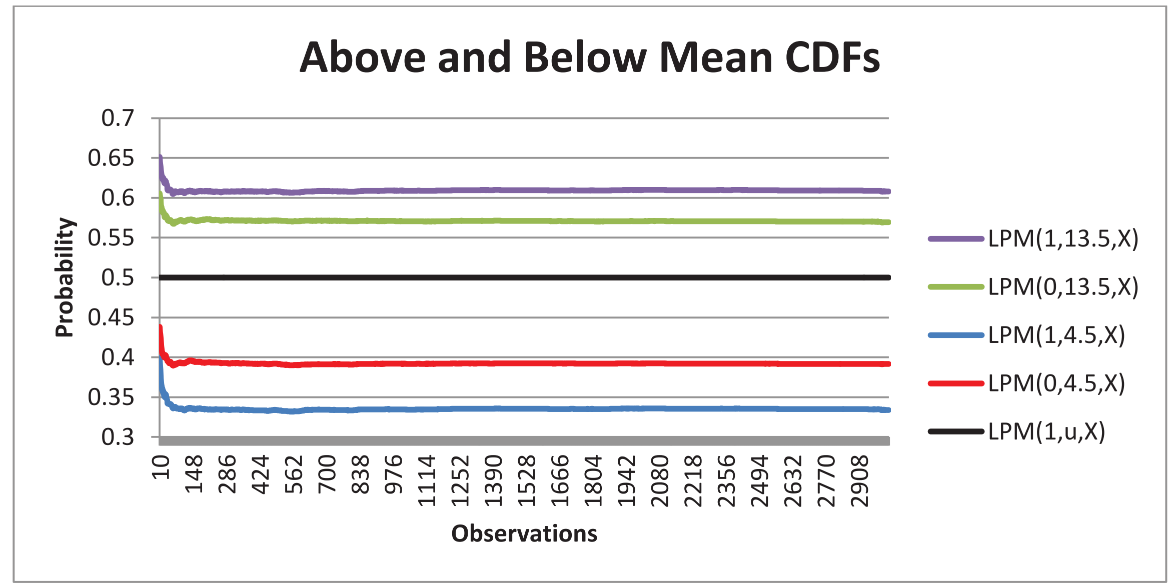


Figure 8. Different locations of the target versus the mean and relationships between discrete and continuous CDFs.

In Figure 8, we used different targets of 4.5%, 9% (mean), and 13.5% and we see that the continuous is outside of the range of the discrete measures. Note that with the mean as the target, the continuous measure is rock solid on the 50% probability.

Normal Distribution Probabilities - 5 Million Draws 300 Iteration Seeds		
Norm Prob($X \le 0.00$) = .3085	LPM(0, 0, X) = .3085	LPM(1, 0, X) = .2208
Norm Prob($X \le 4.50$) = .3917	LPM(0, 4.5, X) = .3917	LPM(1, 4.5, X) = .3339
Norm Prob($X \le$ Mean) = .5	LPM(0, μ, X) = .5	LPM(1, μ, X) = .5
Norm Prob($X \le 13.5$) = .5694	LPM(0, 13.5, X) = .5694	LPM(1, 13.5, X) = .608

Table 1. Final probability estimates with 5 million observations and 300 iteration seeds averaged for the Normal distribution.

In Table 1, we see that the LPM degree 0 provides equivalent probabilities as the Normal Probability function from the IMSL library. The continuous probability using the LPM degree 1 is at 0.5 for the mean as a target and has a lower probability below the mean and a higher probability above the mean as we have noted previously.

D. <u>Uniform Distribution</u>

We compare our metric to the traditional uniform CDF for values less than or equal to *x*.

$$\mathrm{F}(x|A,B) = \begin{cases} 0, & if\ x < A \\ \dfrac{x-A}{B-A}, & if\ A \le x \le B \\ 1, & if\ x > B \end{cases}$$

Table 2 below shows the convergence of our metric to the traditional method for the uniform CDF (UNDF) with a mean of 10. The results are the same as we noted for the normal distribution in Table 1.

Uniform Distribution Probabilities - 5 Million Draws 300 Iteration Seeds		
UNDF($X \le 0.00$) = .4	LPM(0, 0, X) = .4	LPM(1, 0, X) = .3077
UNDF($X \le 4.50$) = .445	LPM(0, 4.5, X) = .445	LPM(1, 4.5, X) = .3913
UNDF($X \le$ Mean) = .5	LPM(0, μ, X) = .5	LPM(1, μ, X) = .5
UNDF($X \le 13.5$) = .535	LPM(0, 13.5, X) = .535	LPM(1, 13.5, X) = .5697

Table 2. Uniform distribution results illustrate convergence of $LPM(0,x,X)$ to UNDF and consistent relationship between $LPM(0,x,X)$ and $LPM_{ratio}(1,x,X)$ above and below the mean target.

E. <u>Poisson Distribution</u>

We compare our metric to the traditional Poisson CDF (POIDF) for values less than or equal to X.

$$f(x) = e^{-\theta}\frac{\theta^x}{x!}$$

Poisson Distribution Probabilities - 5 Million Draws 300 Iteration Seeds		
POIDF($X \le 0.00$) = .00005	LPM(0, 0, X) = 0	LPM(1, 0, X) = 0
POIDF($X \le 4.50$) = .0293	LPM(0, 4.5, X) = .0293	LPM(1, 4.5, X) = .0051
POIDF($X \le$ Mean) = .5151	LPM(0, μ, X) = .5151	LPM(1, μ, X) = .5
POIDF($X \le 13.5$) = .8645	LPM(0, 13.5, X) = .8645	LPM(1, 13.5, X) = .9365

Table 3. Poisson distribution results illustrate convergence of $LPM(0,x,X)$ to POIDF and consistent relationship between $LPM(0,x,X)$ and $LPM_{ratio}(1,x,X)$ above and below the mean target.

F. Chi-Square Distribution

We compare our metric to traditional chi-square CDF (CHIDF) for values less than or equal to X.

$$F(x) = \frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}\int_0^x e^{\frac{-t}{2}} t^{\frac{v}{2}-1}\,dt$$

We set the degrees of freedom for the chi-square equal to one. The reason for this arbitrary selection is the distinct curve generated by this parameter value, and its likeness to the power law distribution. There is no a priori argument that the degrees of freedom will affect our methodology given its non-parametric derivation.

Chi-Squared Distribution Probabilities - 5 Million Draws 300 Iteration Seeds

CHIDF(X ≤ 0) = 0	LPM(0, 0, X) = 0	LPM(1, 0, X) = 0
CHIDF(X ≤ 0.5) = .5205	LPM(0, 0.5, X) = .5205	LPM(1, 0.5, X) = .2087
CHIDF(X ≤ 1) = .6827	LPM(0, 1, X) = .6827	LPM(1, 1, X) = .5
CHIDF(X ≤ 5) = .9747	LPM(0, 5, X) = .9747	LPM(1, 5, X) = .989

Table 4. Chi-Square distribution results illustrate convergence of $LPM(0,x,X)$ to UNDF and consistent relationship between $LPM(0,x,X)$ and $LPM_{ratio}(1,x,X)$ above and below the mean target.

G. Continuous Distributions:

In a discrete measurement with a zero target, there is no difference between a 40% observation and a 70% observation as both will yield a single positive count in the frequency (both were observed in our normal distribution generation with μ=10 and σ=20 parameter constraints). However, there is considerable area between these two observations that merely gets binned in a probability analysis. This undesirable construct also has the ubiquitous quality of scale invariance. Equation (14) measures this neglected area with its inherent variance consideration simultaneously factored with the discrete frequency analysis.

> "*All actual sample spaces are discrete, and all observable random variables have discrete distributions. The continuous distribution is a mathematical construction, suitable for mathematical treatment, but not practically observable*." E.J.G. Pitman (1979).

LPM_{ratio} degree of 1 (*n*=*q*=1) permits us to calculate the area "between the bins." For example, in a roll of a die, the area of the function between 3.1 and 3.9 will be static for the discrete method (based on integer bins 1-6). If the distribution were actually continuous, the variance influence in LPM_{ratio} degree 1 generates an accurate measurement of the area 3.1 through 3.9 for this area between the bins - for uniform and all other distributions. Furthermore, the mean for a die roll is approximately 3.5. LPM_{ratio} degree 1 generates a 0.5 result for the CDF with the 3.5 mean as the target in a uniform distribution ranging from 1 to 6. Unfortunately, per Pitman's observation, we are not able to generate a continuous distribution to observe and verify this notion for target values other than the mean (which we prove always equal 0.5) or endpoints (0 or 1 for sample minimum and maximum). The consistent observed relationship we demonstrated between $LPM_{ratio}(1, x, X)$ and $LPM(0, x, X)$ for targets above and below the mean, offers considerable support of the continuous estimates.

A better example to distinguish between discrete and continuous analysis is the chi-square distribution with degrees of freedom set to one. The range of the observations extended to X=35.1 and resembles the power law function. Considering μ=1.0 and σ=1.414, the discrete probability of a mean return was 0.6827 as shown in Table 4. However, if one envisions the decreasing thin slice of area under the function all the way down the x-axis to the observation X=35.1, this extended result only generates a reading of one in its probability calculation of $x > \mu$. No different than an observation of X=11 which is also a positive count in this example. The frequency of X=11 is the distinguishing characteristic. This difference in area between 11 and 35.1 is considerable and is completely disregarded under discrete frequency analysis. When the variance of that deviation is considered to account for the infinite possible outcomes for the continuous variable, the probability of a mean return drops significantly to 0.5 from 0.6827.

The reason for this is straightforward, $LPM(0, x, X)$ converges to the frequency / counting data set while $LPM_{ratio}(1, x, X)$ retains its area property.

III. Joint Distribution Co-Partial Moments and UPM/LPM Analysis

In this section, we introduce the framework for the joint distribution using partial moments. For more background, Appendix B and Appendix C provide more information on joint probabilities and conditional CDFs. We also replicate the covariance matrix of a two variable normal distribution and its cosemivariance matrix with the variables' aggregated partial moment components. This information provides a toolbox that yields a positive definite symmetrical co-partial moment matrix capable of handling any target and resulting asymmetry, providing a distinct advantage over its cosemivariance counterpart.

The issue in this area traces back to the Markowitz (1959) chapter on semivariance analysis. The cosemivariance matrix in Markowitz is an endogenous matrix that is computed after the portfolio returns have been computed. Because we have to know the portfolio allocations before we can compute the portfolio returns, the cosemivariance matrix is not known until after we have solved the problem. Attempts to solve the mean-semivariance problem with an exogenous matrix, a matrix computed from the security return data, have had problems because the cosemivariance matrix is asymmetric, and therefore, not positive semi-definite. Grootveld and Hallerbach (1999) noted that the endogenous and exogenous matrices are not equivalent. Estrada (2008), however, demonstrates that a symmetric exogenous matrix is a very good approximation for the endogenous matrix. Our purpose is to demonstrate a method that provides a positive semi-definite matrix system that preserves any asymmetry in the underlying process.

First, the LPM and the CLPM are defined as follows:

$$LPM(n,h,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, h - x_t\}^n\right] \tag{18}$$

$$CLPM(n,h,x|y) = \frac{1}{T}\left[\sum_{t=1}^{T} (max\{0, h - x_t\}^n \cdot max\{0, h - y_t\}^n)\right] \tag{19}$$

The Degree 1 Co-LPM (CLPM) matrix is:

$$\begin{bmatrix} LPM(2,h,x) & CLPM(1,h,x|y) \\ CLPM(1,h,y|x) & LPM(2,h,y) \end{bmatrix}$$

$$LPM(2,h,x) = CLPM(1,h,x|x) \tag{20}$$

Since variance is the squared deviation

$$\sigma_x^2 = \mathrm{E}[(x_t - \mu_{xt})^2] = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} (x_t - \mu_{xt})^2 \tag{21}$$

It is also the deviation times itself…the covariance of itself.

$$\sigma_{xx} = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} (x_t - \mu_{xt})(x_t - \mu_{xt}) \tag{22}$$

And the covariance between 2 variables is simply

$$\sigma_{xy} = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} (x_t - \mu_{xt})(y_t - \mu_{yt}) \tag{23}$$

Since semivariance from benchmark B is

$$\Sigma_{xB}^2 = \mathrm{E}\{[min(x - B, 0)^2] = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [\min(x_t - B, 0))^2] \tag{24}$$

Then it is also the cosemivariance of itself

$$\Sigma_{xxB} = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [min(x_t - B, 0)][min(x_t - B, 0)] \tag{25}$$

And the cosemivariance between 2 variables is

$$\Sigma_{xyB} = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [min(x_t - B, 0)][min(y_t - B, 0)] \tag{26}$$

Since LPM degree 2 is equal to semivariance, $LPM(2,B,x) = \Sigma_{xB}^2$

$$LPM(2,B,x) = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [max(B - x_t, 0)]^2 \tag{27}$$

Also equals the Co-LPM degree 1 of the same variable

$$CLPM(1,B,x|x) = \Sigma_{xB}^2 = \Sigma_{xxB} = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [max(B - x_t, 0)][max(B - x_t, 0)] \tag{28}$$

And the Co-LPM degree 1 between 2 variables is

$$CLPM(1,B,x|y) = \left(\frac{1}{T}\right) \cdot \sum_{t=1}^{T} [max(B - x_t, 0)][max(B - y_t, 0)] \quad (29)$$

For two symmetrical distributions x, y with $h = \mu$

Co-LPM Matrix = Co-UPM Matrix

$$\begin{bmatrix} LPM(2,\mu,x) & CLPM(1,\mu,x|y) \\ CLPM(1,\mu,y|x) & LPM(2,\mu,y) \end{bmatrix} = \begin{bmatrix} UPM(2,\mu,x) & CUPM(1,\mu,x|y) \\ CUPM(1,\mu,y|x) & UPM(2,\mu,y) \end{bmatrix}$$

Furthermore, the addition of the Co-LPM matrix, the Co-UPM matrix and the complement set $(CLPM + CUPM)^C$ is equivalent to the covariance matrix on the main diagonal.

$$CLPM(1,h,x|y) + CUPM(1,h,x|y) =$$

$$\begin{bmatrix} LPM(2,h,x) & CLPM(1,h,x|y) \\ CLPM(1,h,y|x) & LPM(2,h,y) \end{bmatrix} + (CLPM + CUPM)^C + \begin{bmatrix} UPM(2,h,x) & CUPM(1,h,x|y) \\ CUPM(1,h,y|x) & UPM(2,h,y) \end{bmatrix}$$

$$= \begin{bmatrix} LPM(2,h,x) + UPM(2,h,x) & CLPM(1,h,x|y) + CUPM(1,h,x|y) \\ CLPM(1,h,y|x) + CUPM(1,h,y|x) & LPM(2,h,y) + UPM(2,h,y) \end{bmatrix}$$

The main diagonal of the aggregated matrix will retain the covariance equivalence under any asymmetry with the following relationship for all targets,

$$\sigma_x^2 = LPM(2,\mu,x) + UPM(2,\mu,x) \quad (30)$$

$$CLPM(1,\mu,x|y) + CUPM(1,\mu,x|y) =$$

$$\frac{1}{T}\left[\sum_{t=1}^{T} (max\{0,\mu - x_t\} \cdot max\{0,\mu - y_t\}) + (max\{0,x_t - \mu\} \cdot max\{0,y_t - \mu\})\right] \quad (31)$$

Equation (31) will generate a zero instead of a negative covariance result, ensuring a positive matrix. This zero (instead of the negative) result does not affect the preservation of information for the instances whereby one variable is above the target and one below. The addition of this observation to the complement set lowers both the CLPM and CUPM. In essence, nothing is something.

We note that each of the co-partial moment matrices are positive symmetrical semi-definite, ensuring a positive symmetrical definite aggregate matrix.

A. Complement Set Matrix

To further analyze the information in the $(CLPM + CUPM)^C$ complement set from diverging target returns between variables, we introduce two new metrics - the diverging lower partial moment ($DLPM$) and diverging upper partial moment ($DUPM$).

$$DLPM(q|n,h,x|y) = \frac{1}{T}\left[\sum_{t=1}^{T}(max\{x_t - h, 0\}^q \cdot max\{0, h - y_t\}^n)\right] \tag{32}$$

$$DUPM(n|q,h,x|y) = \frac{1}{T}\left[\sum_{t=1}^{T}(max\{0, h - x_t\}^n \cdot max\{y_t - h, 0\}^q)\right] \tag{33}$$

Equation (32) provides the divergent LPM for variable Y given a positive target deviation for variable X from shared target h, with the LPM and UPM degrees (n and q respectively) explained earlier in equations 1 and 2. For example, given a 20% observation for variable X and a shared target of 0%, a -10% observation for variable Y will generate a larger $DLPM$ than a -5% observation for variable Y.

Conversely, equation (33) provides the divergent UPM for variable Y given a negative target deviation for variable X.

The matrix of each divergent partial moment will be aggregated to represent the divergent partial moment matrix (DPM). One key feature of this matrix is the main

diagonal consists of all zeros since the divergent partial moment of the same variable does not exist. The degree 1 DPM is presented below.

$$\begin{bmatrix} 0 & DPM(1|1,h,x|y) \\ DPM(1|1,h,y|x) & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & DLPM(1|1,h,x|y) \\ DLPM(1|1,h,y|x) & 0 \end{bmatrix} + \begin{bmatrix} 0 & DUPM(1|1,h,x|y) \\ DUPM(1|1,h,y|x) & 0 \end{bmatrix} \tag{34}$$

Since there only exists four possible interactions between two variables,

$X \leq$ target, $Y \leq$ target	$CLPM(n,h,x|y)$
$X \leq$ target, $Y >$ target	$DUPM(n|q,h,x|y)$
$X >$ target, $Y \leq$ target	$DLPM(q|n,h,x|y)$
$X >$ target, $Y >$ target	$CUPM(q,h,x|y)$

we can clearly see that the sum of the degree 0 probability matrices of all four interactions must equal one, explaining the entire multivariate distribution.

The distinct advantage for the partial moments over semivariance as the preferred below target analysis method is the ability for the partial moments to compensate for any asymmetry.

Under symmetry,

Cosemivariance Matrix = ½ Covariance Matrix

$$\begin{bmatrix} \Sigma_{xx\mu} & \Sigma_{xy\mu} \\ \Sigma_{yx\mu} & \Sigma_{yy\mu} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$$

$$\begin{bmatrix} \Sigma_{xx\mu} & \Sigma_{xy\mu} \\ \Sigma_{yx\mu} & \Sigma_{yy\mu} \end{bmatrix} + \begin{bmatrix} \Sigma_{xx\mu} & \Sigma_{xy\mu} \\ \Sigma_{yx\mu} & \Sigma_{yy\mu} \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \tag{35}$$

Minimizing $\begin{bmatrix} \Sigma_{xx\mu} & \Sigma_{xy\mu} \\ \Sigma_{yx\mu} & \Sigma_{yy\mu} \end{bmatrix}$ creates an imbalance that has no offsetting components to equal the covariance matrix when added to itself. The minimizing of the LPM matrix and the DLPM matrix has a simultaneous inverse effect of increasing the UPM matrix and DUPM matrix, ergo compensating for any asymmetry. This balancing effect holds for any target, not just μ.

$$\begin{aligned} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \sim & \begin{bmatrix} LPM(2,\mu,x) & CLPM(1,\mu,x|y) \\ CLPM(1,\mu,y|x) & LPM(2,\mu,y) \end{bmatrix} \\ & - \begin{bmatrix} 0 & DPM(1|1,\mu,x|y) \\ DPM(1|1,\mu,y|x) & 0 \end{bmatrix} \\ & + \begin{bmatrix} UPM(2,\mu,x) & CUPM(1,\mu,x|y) \\ CUPM(1,\mu,y|x) & UPM(2,\mu,y) \end{bmatrix} \end{aligned} \tag{36}$$

Each of the co-partial moment matrices is positive symmetrical semi-definite, ensuring a positive symmetrical definite aggregate matrix, thus avoiding the endogenous/exogenous matrix problem described by Grootveld and Hallerbach (1999) and Estrada (2008).

In R, using the ‘NNS’ package, we can verify the variance/covariance equivalence.

```
> set.seed(123); x=rnorm(100); y=rnorm(100)

> var(x)
[1] 0.8332328
#Sample:
> UPM(2,mean(x),x)+LPM(2,mean(x),x)
[1] 0.8249005
#Population:
> (UPM(2,mean(x),x)+LPM(2,mean(x),x))*(length(x)/(length(x)-1))
[1] 0.8332328
#Variance is also the co-variance of itself:
> (Co.LPM(1,1,x,x)+Co.UPM(1,1,x,x)-D.LPM(1,1,x,x)-
D.UPM(1,1,x,x))*(length(x)/(length(x)-1))
[1] 0.8332328

> cov(x,y)
[1] -0.04372107
> (Co.LPM(1,1,x,y)+Co.UPM(1,1,x,y)-D.LPM(1,1,x,y)-
D.UPM(1,1,x,y))*(length(x)/(length(x)-1))
[1] -0.04372107
```

IV. Conclusions

We have demonstrated how the LPM degree 0 is equal to the traditionally derived CDF of any assumed distribution. $LPM(0,x,X)$ converges to:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}\,dt,$$

$$F(x|A,B) = \begin{cases} 0, & if\ x < A \\ \frac{x-A}{B-A}, & if\ A \le x \le B, \\ 1, & if\ x > B \end{cases}$$

$$f(x) = e^{-\theta}\frac{\theta^x}{x!},$$

$$F(x) = \frac{1}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})}\int_0^x e^{\frac{-t}{2}} t^{\frac{v}{2}-1}\,dt.$$

The obvious benefit is the distribution agnostic manner of this direct computation, which consumes far less time and cpu effort than bootstrapping a discrete estimate. Furthermore, the stability of the partial moments versus each of the distribution estimates is yet another benefit of our method. Finally, the ability to derive results for a truly continuous variable emphasizes the flexibility of this method.

Any computer generated sample and analysis thereof, is that of a discrete variable. A histogram and bins as commonly performed in Excel by practitioners and academics alike ignores a large area under the function due to this discrete classification. The addition of bins with increased observations does not fill in the area and converge to the continuous

area estimate; it merely creates larger quantities of smaller areas thus keeping the total area constant. Equation (14) makes no such concessions and generates the theoretical continuous area, while maintaining the relationship identified in Equation (15). We note how the continuous CDF is much more pronounced the further from the mean the integral is - compensating for the asymmetry of the additional area "between the bins" that is placed in the proceeding bin during discrete analysis.

Benoit Mandelbrot notes the shorter the measuring instrument, the larger the coastline of Britain; ultimately yielding a result of infinity. This line of reasoning is commensurate with the continuous CDF versus its discrete counterpart; and the infinitesimal subintervals of a continuous distribution. We hope that further research on this method and its applications eventually finds its way to various fields of study.

We show that the Cumulative Distribution Function (CDF) is represented by the ratio of the lower partial moment ratio (LPM_{ratio}) to the distribution for the interval in question. The addition of the upper partial moment ratio (UPM_{ratio}) enables us to create probability density functions (PDF) for any function or distribution without prior knowledge of its characteristics. The ability to derive the CDF and PDF without any distributional assumptions yields a more accurate calculation devoid of any error terms present from a less than perfect goodness of fit, as well as critical information about the tails of the distribution. This foundation is then used to develop conditional probabilities and joint distribution co-partial moments. The resulting toolbox allows us to propose a new formulation for UPM/LPM analysis and we note that each of the co-partial moment matrices are positive symmetrical semi-definite, ensuring a positive symmetrical definite aggregate matrix.

Appendix A:

In this section we address any sample size concerns the reader may logically infer. Since these concerns are not specific to our methodology but rather to statistics in general, we offer the results of a separate study comparing the deviations from the large sample sizes reported in the main body of this paper.

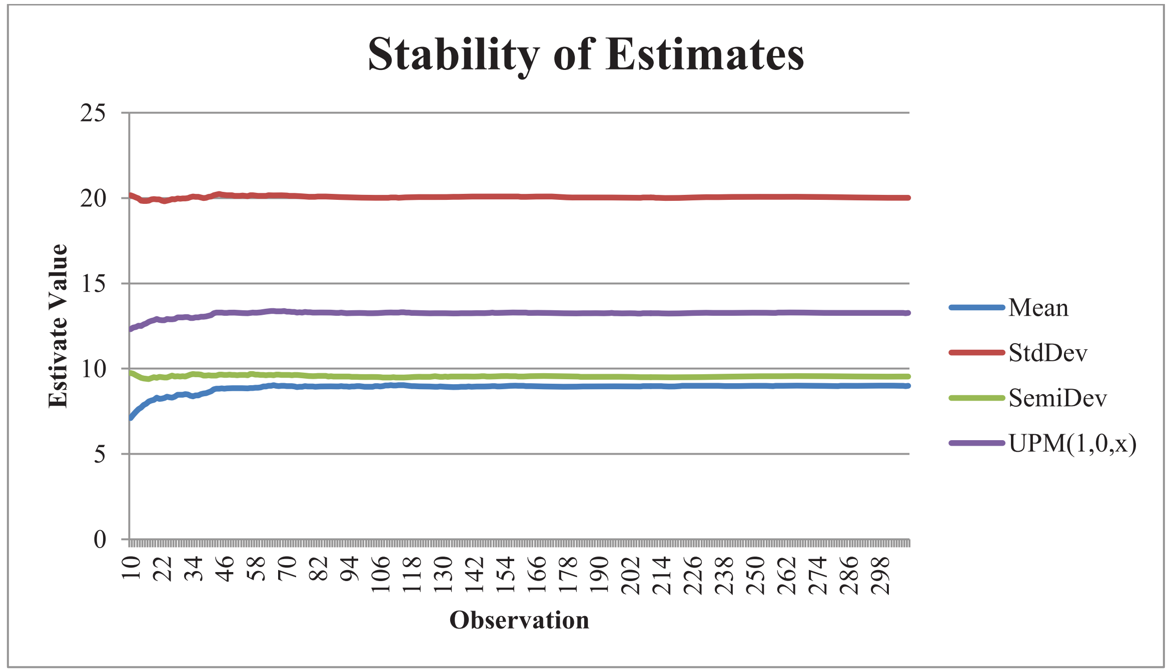


Figure 1a. Visual representation of the stabilization of statistics as sample size increases.

Test	Mean	StdDev	SemiDev	Variance	LPM (0,0,x)	LPM (1,0,x)	UPM (0,0,x)	UPM (1,0,x)	LPM (0,μ,x)	LPM (1,4.5,x)	LPM (1,13.5,x)	LPM (0,4.5,x)	LPM (0,13.5,x)	Norm Prob (4.5)	Norm Prob (13.5)	Norm Prob (0)
MAD 30-100	0.1258	0.0479	0.0409	2.2496	0.0024	0.0582	0.0024	0.0774	0.0012	0.0043	0.0031	0.0023	0.0019	0.0025	0.0024	0.0024
MAD 30-300	0.0678	0.0378	0.0395	2.2107	0.0014	0.0506	0.0014	0.037	0.0013	0.0026	0.0015	0.0014	0.0012	0.0014	0.0012	0.0014
MAD 30-3000	0.021	0.0132	0.0266	0.7485	0.0005	0.0125	0.0005	0.0121	0.0006	0.0007	0.0007	0.0005	0.0004	0.0004	0.0004	0.0004
MAD 100-5 Million	0.0194	0.0106	0.026	0.4914	0.0005	0.011	0.0005	0.0107	0.0006	0.0006	0.0006	0.0004	0.0004	0.0004	0.0004	0.0004
MAD 3000-5 Million	0.0039	0.0043	0.0031	0.1833	0.0001	0.0021	0.0001	0.0027	0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Table 1a. Final Stability of Estimates Test for the Normal distribution.

Test	Mean	StdDev	SemiDev	Variance	LPM (0,0,x)	LPM (1,0,x)	UPM (0,0,x)	UPM (1,0,x)	LPM (0,μ,x)	LPM (1,4.5,x)	LPM (1,13.5,x)	LPM (0,4.5,x)	LPM (0,13.5,x)
MAD 30-100	0.031	0.0208	0.0280	1.9174	0.0005	0.0163	0.0005	0.0206	0.0005	0.0008	0.0006	0.0006	0.0005
MAD 30-300	0.056	0.0184	0.0516	1.9271	0.0008	0.0228	0.0008	0.0342	0.0009	0.001	0.0012	0.0009	0.0007
MAD 30-3000	0.0342	0.0140	0.0330	0.7431	0.0003	0.0169	0.0003	0.0193	0.0002	0.0006	0.0007	0.0003	0.0004
MAD 100-5 Million	0.0283	0.0133	0.0245	0.7108	0.0003	0.0134	0.0003	0.0173	0.0002	0.0005	0.0006	0.0002	0.0004
MAD 3000-5 Million	0.0067	0.0029	0.0023	0.1725	0.0001	0.0019	0.0001	0.0049	0	0.0001	0.0001	0.0001	0.0001

Table 2a. Final Stability of Estimates Test for the Uniform distribution.

Test	Mean	StdDev	SemiDev	Variance	LPM (0,4.5,x)	LPM (1,4.5,x)	UPM (0,4.5,x)	UPM (1,4.5,x)	LPM (0,μ,x)	LPM (1,0,x)	LPM (1,13.5,x)	LPM (0,0,x)	LPM (0,13.5,x)
MAD 30-100	0.0439	0.0322	0.0327	0.2358	0.0053	0.0111	0.0053	0.0327	0.0027	0	0.0012	0	0.0016
MAD 30-300	0.0486	0.0353	0.0346	0.2536	0.0055	0.0109	0.0055	0.0376	0.0029	0	0.0012	0	0.0014
MAD 30-3000	0.0091	0.0115	0.0091	0.079	0.0012	0.0027	0.0012	0.0069	0.0027	0	0.0005	0	0.0004
MAD 100-5 Million	0.0056	0.0082	0.0049	0.0542	0.0006	0.0014	0.0006	0.0046	0.0027	0	0.0005	0	0.0004
MAD 3000-5 Million	0.0009	0.0038	0.0022	0.0242	0.0002	0.0004	0.0002	0.0006	0.0031	0	0.0001	0	0.0001

Table 3a. Final Stability of Estimates Test for the Poisson distribution.

Test	Mean	StdDev	SemiDev	Variance	LPM (0,0.5,x)	LPM (1,0.5,x)	UPM (0,0.5,x)	UPM (1,0.5,x)	LPM (0,μ,x)	LPM (1,0,x)	LPM (1,5,x)	LPM (0,0,x)	LPM (0,5,x)
MAD 30-100	0.036	0.074	0.0009	0.2792	0.0033	0.0009	0.0033	0.0352	0.0032	0	0.004	0	0.0055
MAD 30-300	0.0371	0.0747	0.0012	0.2745	0.0031	0.0011	0.0031	0.036	0.0029	0	0.0039	0	0.0053
MAD 30-3000	0.0102	0.0231	0.0006	0.075	0.0014	0.0006	0.0014	0.0096	0.0008	0	0.001	0	0.0015
MAD 100-5 Million	0.0057	0.0186	0.0004	0.0554	0.0009	0.0004	0.0009	0.0053	0.0006	0	0.0007	0	0.0008
MAD 3000-5 Million	0.0027	0.0095	0.0001	0.0277	0.0003	0.0001	0.0003	0.0026	0.0002	0	0.0003	0	0.0002

Table 4a. Final Stability of Estimates Test for the Chi-Squared distribution.

Appendix B:

Conditional Probabilities:

We illustrate how the partial moment ratios can also emulate conditional probability calculations. We re-visualize the Venn diagram areas in Figure 1b as distribution areas from which the *LPM* and *UPM* can be observed.

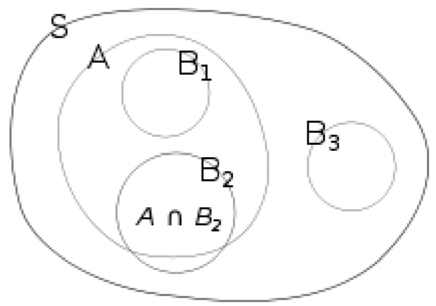


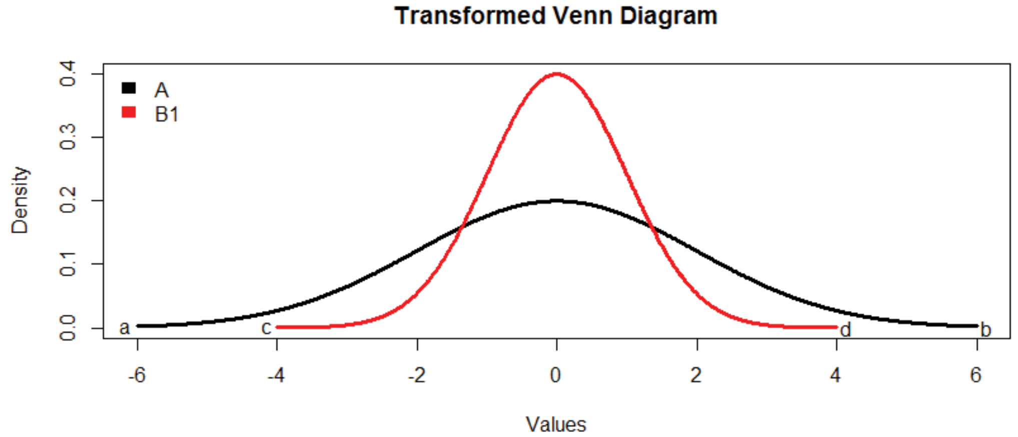
Figure 1b. Venn diagram illustrating conditional probabilities of different areas in the sample space, S.

$P(A|B_1) = 1$

$P(A|B_2) \approx 0.85$

$P(A|B_3) = 0.$

The conditional probability $P(A|B_1) = 1$.

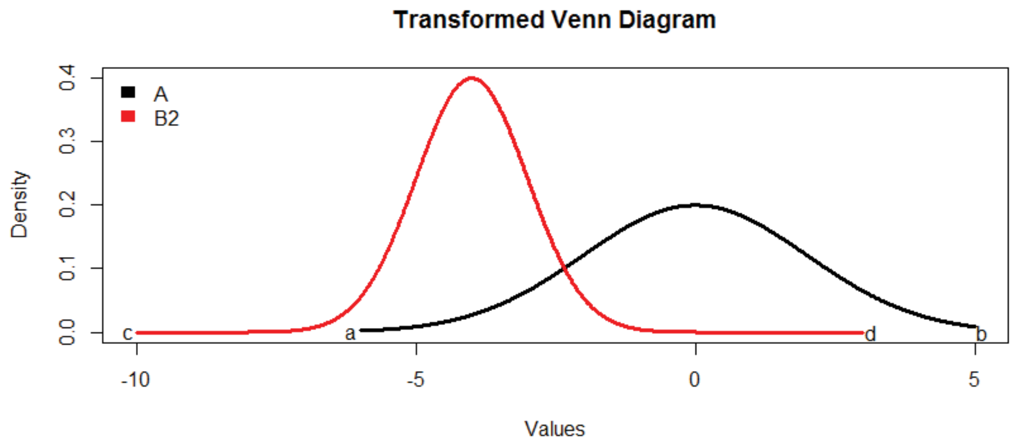


$$1 = 1 - LPM(0, a, B_1) - UPM(0, b, B_1) \tag{B.1}$$

$$1 = UPM(0, a, B_1) - UPM(0, b, B_1) \tag{B.2}$$

$$1 = (1) - (0)$$

The conditional probability $P(A|B_2) \approx 0.85$.

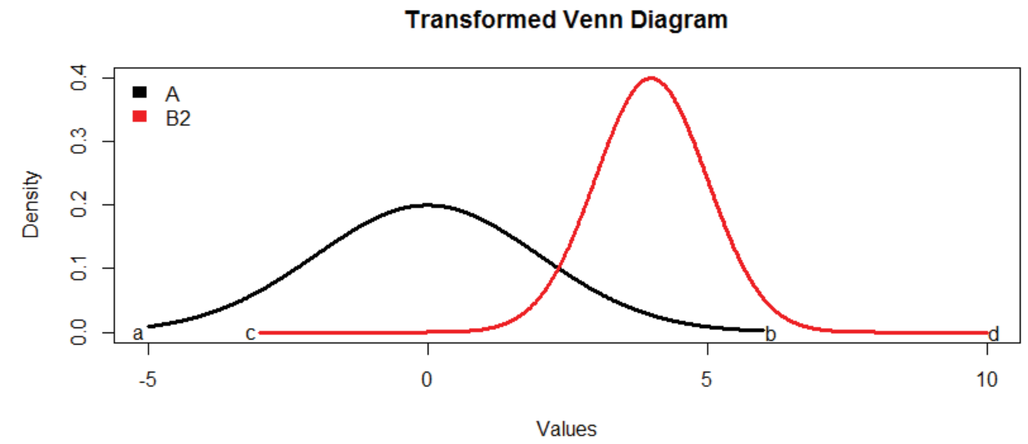


$$0.85 = 1 - LPM(0, a, B_2) - UPM(0, b, B_2) \tag{B.4}$$

$$0.85 = UPM(0, a, B_2) - UPM(0, b, B_2) \tag{B.5}$$

$$0.85 = (.85) - (0)$$

The conditional probability $P(A|B_2) \approx 0.85$.

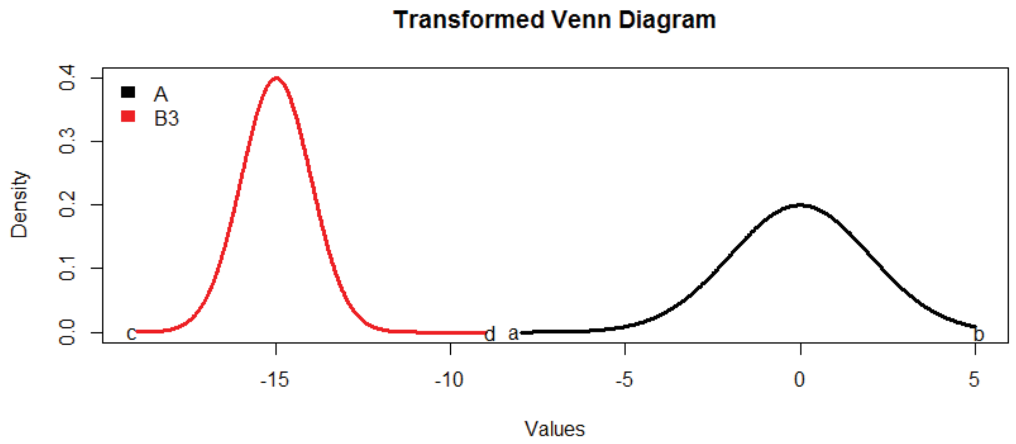


$$0.85 = 1 - LPM(0, a, B_2) - UPM(0, b, B_2) \tag{B.7}$$

$$0.85 = UPM(0, a, B_2) - UPM(0, b, B_2) \tag{B.8}$$

$$0.85 = (1) - (.15)$$

The conditional probability P(A|B_3) = 0.

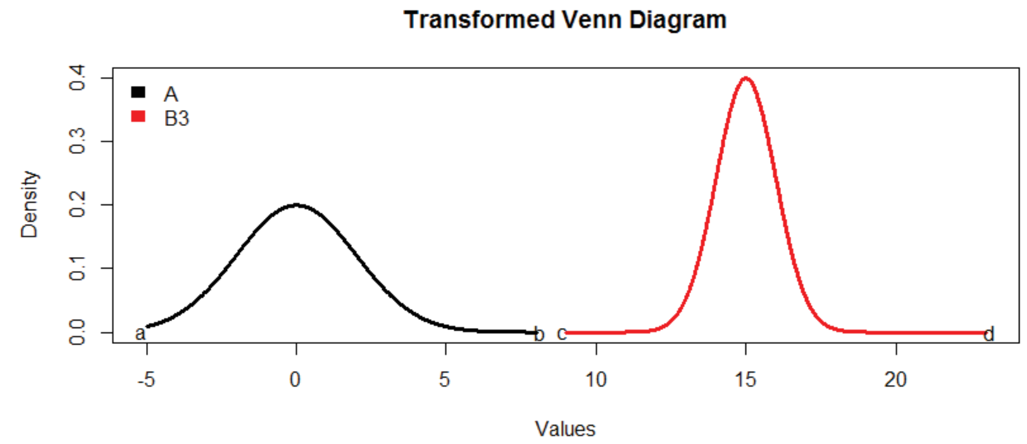


$$0 = 1 - LPM(0, a, B_3) - UPM(0, b, B_3) \quad (B.10)$$

$$0 = UPM(0, a, B_3) - UPM(0, b, B_3) \quad (B.11)$$

$$0 = (0) - (0)$$

The conditional probability P(A|B_3) = 0.



$$0 = 1 - LPM(0, a, B_3) - UPM(0, b, B_3) \quad (B.13)$$

$$0 = UPM(0, a, B_3) - UPM(0, b, B_3) \quad (B.14)$$

$$0 = (1) - (1)$$

Bayes' Theorem:

Bayes' theorem will also generate the conditional probability of A given B, $P(A|B)$ with the formula

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

Where the probability of A is represented by,

$$P(A) = \frac{Area\ of\ A}{Area\ of\ total\ sample\ S} = UPM(0, a, A)$$

And the probability of B is represented by,

$$P(B) = \frac{Area\ of\ B}{Area\ of\ total\ sample\ S} = UPM(0, c, B)$$

Where e is the minimum value target of area (distribution) S; just as a and c are for areas (distributions) A and B respectively (d and b are maximum respective value targets). Thus, if the conditional probability of B given A is (per equation B.2),

$$P(B|A) = \frac{CUPM(0|0, c|a, B|A)}{UPM(0, a, A)}$$

Then,

$$P(A|B) = \frac{\frac{CUPM(0|0, c|a, B|A)}{UPM(0, a, A)} UPM(0, a, A)}{UPM(0, c, B)}$$

Cancelling out $P(A)$ leaves us with Bayes' theorem represented by partial moments, and our conditional probability on the right side of the equality.

$$P(A|B) = \frac{CUPM(0|0, c|a, B|A)}{UPM(0, c, B)}$$

The following table of the canonical breast cancer test example will help place the partial moments with their respective outcomes (R commands in red):

- 1% of women have breast cancer (and therefore 99% do not).
- 80% of mammograms detect breast cancer when it is there (and therefore 20% miss it).
- 10% of mammograms detect breast cancer when it's not there (and therefore 90% correctly return a negative result).
- Using -1 for C & TN instances, and 1 for NC & TP instances[8]

	Cancer (1%)	No Cancer (99%)	Y variable
Test Positive	Co.UPM(0,0,T,C,0,0)=.008	D.LPM(0,0,T,C,0,0)=.099	UPM(0,0,T) = .107
Test Negative	D.UPM(0,0,T,C,0,0)=.002	Co.LPM(0,0,T,C,0,0)=.891	LPM(0,0,T) =.893
X variable	UPM(0,0,C) = .01	LPM(0,0,C) = .99	UPM+LPM=1

Appendix C: Joint CDFs and UPM/LPM Correlation Analysis

Joint CDFs:

The discrete probability that both X is less than some target h_x and Y is less than some target h_y simultaneously is simply the degree 0 co-LPM provided earlier in equation (29).

[8] In R representing 1000 individuals:
> C=c(rep(1,8),rep(-1,990),rep(1,2)); T=c(rep(1,107),rep(-1,893))

$$\Pr[x \leq h_x, y \leq h_y] = CLPM(0|0, h_x|h_y, x|y) \qquad (C.1)$$

This is the discrete CDF of the joint distribution, just how we prove $LPM(0, h, X)$ is the discrete CDF of the univariate distribution.

Where,

$$0 \leq CLPM(0|0, h_x|h_y, x|y) \leq 1 \qquad (C.2)$$

$CLPM(0|0, h_x|h_y, x|y)$ has the following properties for various correlations between the two variables ρ_{xy}, when $h_x = h_y$.[9]

- If $\rho_{xy} = 1$; $CLPM(0|0, h_x|h_y, x|y) = \min\{LPM(0, h_x, x), LPM(0, h_y, y)\}$.
- If $\rho_{xy} = 0$; $CLPM(0|0, h_x|h_y, x|y) = h_x \cdot h_y$
- If $\rho_{xy} = -1$; $CLPM(0|0, h_x|h_y, x|y) = 0$.

An example may help illustrate the relationship. Let's assume the same target $h_x = h_y$ which we arbitrarily select to the 5% CDF level for two normal distributions with μ= 9 and σ= 20. We then ask, what's the probability that both variables will be in the lower 5% of their distribution simultaneously under different correlations?

[9] We leave further asymmetric target analysis for future research.

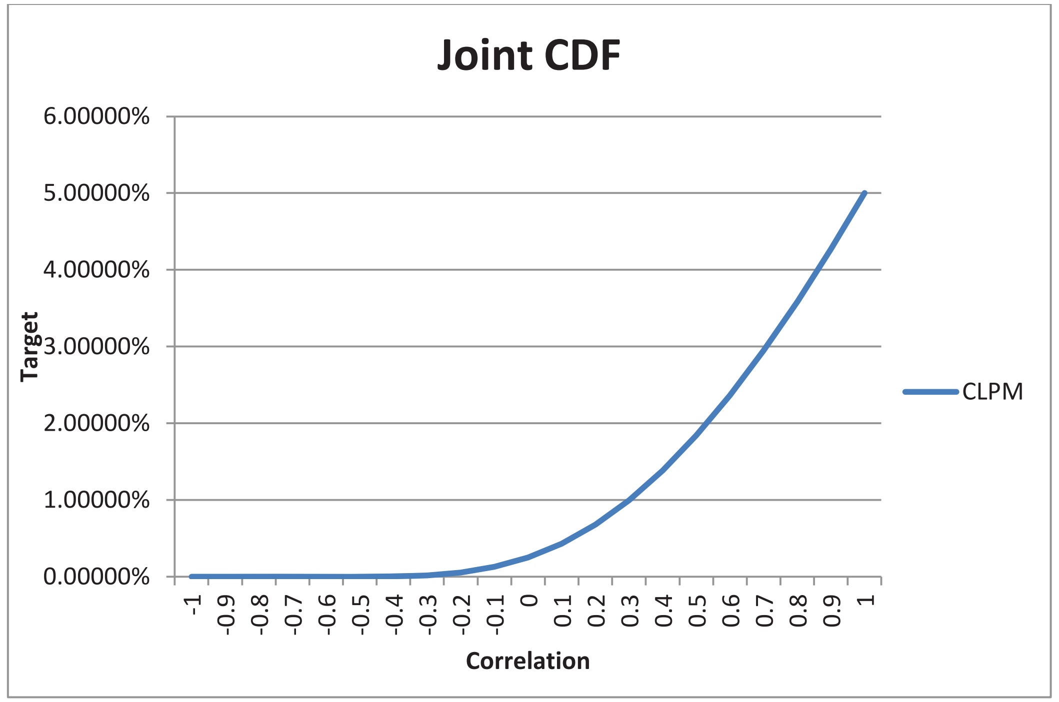


Figure 1C. Hypothetical 5% shared target on two variables (x, y) and the joint CDF for various correlations.

We can deduce the correlation between the assets only with knowledge of the $CLPM$ and $h_x|h_y$. For example, with both our variables and their 5% targets, if the $CLPM(0|0, h_x|h_y, x|y) = 0.25\%$ we know that $\rho_{xy} = 0$.

Equation C.3 will provide the implied correlation for an observed discrete joint CDF, $CLPM(0|0, h_x|h_y, x|y)$. Lucas (1995) provides a framework for estimating the correlation between two events with the following equation which substitutes a binomial event into the standard Pearson correlation coefficient:

$$Corr(A,B) = \frac{P(A \; and \; B) - P(A) \times P(B)}{[P(A)(1-P[A])]^{\frac{1}{2}} \times [P(B)(1-P[B])]^{\frac{1}{2}}} \tag{C.3}$$

From which we can substitute the partial moments for our events

$(x \le h_x, y \le h_y)$, yielding

$$\rho_{xy} = \frac{CLPM(0|0,h_x|h_y,x|y) - LPM(0,h_x,x) \cdot LPM(0,h_y,y)}{\sqrt{[LPM(0,h_x,x) \cdot UPM(0,h_x,x)] \cdot [LPM(0,h_y,y) \cdot UPM(0,h_y,y)]}} \tag{C.4}$$

From our $h_x = h_y = 5\%$ example,

$$\rho_{xy} = \frac{0.25\% - (5\%)(5\%)}{\sqrt{[5\% \cdot 95\%] \cdot [5\% \cdot 95\%]}}$$

$$\rho_{xy} = 0.$$

If the first term in the numerator $(CLPM(0|0,h_x|h_y,x|y))$ equals 0.25%, the implied correlation for that joint CDF is zero. This example also illustrates the independence criterion $(h_x \cdot h_y)$ from a zero correlation.

Partial Moment (Nonlinear) Correlations:

Avoiding the linear dependence of the Pearson coefficient from which Lucas' coefficient is derived, we can use the following relationship in Equation C.5 to determine the nonlinear correlation between two variables $(0|0 \to 0)$.

$$\rho_{xy} = \frac{[CLPM(0,h_x|h_y,x|y) - DLPM(0,h_x|h_y,x|y) - DUPM(0,h_x|h_y,x|y) + CUPM(0,h_x|h_y,x|y)]}{[CLPM(0,h_x|h_y,x|y) + DLPM(0,h_x|h_y,x|y) + DUPM(0,h_x|h_y,x|y) + CUPM(0,h_x|h_y,x|y)]} \tag{C.5}$$

If there is a -1 correlation, then the returns between the variables will always be divergent, thus

$$\rho_{xy} = \frac{[0 - DLPM(0,h_x|h_y,x|y) - DUPM(0,h_x|h_y,x|y) + 0]}{[0 + DLPM(0,h_x|h_y,x|y) + DUPM(0,h_x|h_y,x|y) + 0]} = -1 \tag{C.6}$$

If there is a perfect correlation between two variables, then there will be no divergent returns, thus

$$\rho_{xy} = \frac{[CLPM(0,h_x|h_y,x|y) - 0 - 0 + CUPM(0,h_x|h_y,x|y)]}{[CLPM(0,h_x|h_y,x|y) + 0 + 0 + CUPM(0,h_x|h_y,x|y)]} = 1 \tag{C.7}$$

If there is zero correlation between two variables, then the co- and divergent returns will be of equal frequency or magnitude (degree zero and degree one respectively),

$$CLPM(0,h_x|h_y,x|y) = DLPM(0,h_x|h_y,x|y) = DUPM(0,h_x|h_y,x|y) = CUPM(0,h_x|h_y,x|y)$$

Thus,

$$\rho_{xy} = \frac{[CLPM(0,h_x|h_y,x|y) - DLPM(0,h_x|h_y,x|y) - DUPM(0,h_x|h_y,x|y) + CUPM(0,h_x|h_y,x|y)]}{[CLPM(0,h_x|h_y,x|y) + DLPM(0,h_x|h_y,x|y) + DUPM(0,h_x|h_y,x|y) + CUPM(0,h_x|h_y,x|y)]} = 0 \tag{C.8}$$

Degree one can be substituted to generate correlations whereby the magnitude of the target deviations are compared, generating a dependence coefficient.

Continuous Joint CDF:

The continuous joint CDF can be obtained with the following equation; whereby the ratio of $CLPM(1|1,h_x|h_y,x|y)$ to the entire degree 1 joint distribution will generate the probability percentage. Thus,

$$CLPM_{ratio}(1|1,h_x|h_y,x|y) = \frac{CLPM(1|1,h_x|h_y,x|y)}{[LPM(1,h_x,x)\cdot LPM(1,h_y,y)] + [UPM(1,h_x,x)\cdot UPM(1,h_y,y)]} \tag{C.9}$$

$$\Pr[\mathrm{x} \le h_x, \mathrm{y} \le h_y] = CLPM_{ratio}(1|1,h_x|h_y,x|y) \tag{C.10}$$

NONLINEARITY IS TEDIOUS, NOT COMPLEX

Deriving Nonlinear Correlation Coefficients from Partial Moments

Abstract

We introduce a nonlinear correlation coefficient metric derived from partial moments that can be substituted for the Pearson correlation coefficient in linear instances as well. The flexibility offered by partial moments enables ordered partitions of the data whereby linear segments are aggregated for an overall correlation coefficient. Our coefficient works without the need to perform a linear transformation on the underlying data, and can also provide a general measure of nonlinearity between two variables. We also extend the analysis to a multiple nonlinear regression without the adverse effects of multicollinearity.

1. INTRODUCTION

Chen et al. (2010) explore the problem of estimating a nonlinear correlation (See Figure 1). They note that a generic use statistic such as the Pearson correlation coefficient does not exist for nonlinear correlations. We introduce a generic nonlinear correlation coefficient metric derived from partial moments that can be substituted for the Pearson correlation coefficient in linear instances as well. The flexibility offered by partial moments enables ordered partitions of the data whereby linear segments are aggregated for an overall correlation coefficient.

Partial moments have three main advantages: (1) no distributional assumption is required, (2) partial moments are integrated into economics through expected utility theory (Holthausen, 1981 and Guthoff et al., 1997), and are integrated into statistics as Viole and Nawrocki (2012a) find that partial moments can be used to derive the CDF and PDF of any distribution.

The paper is organized as follows: The next section will cover the development of the measure followed by a section with empirical results. Next, we extend the analysis to a multidimensional nonlinear analysis with an application to nonlinear regression analysis. A final discussion and summary completes the paper.

2. DEVELOPMENT OF NONLINEAR CORRELATION MEASURE

The Pearson correlation coefficient is represented by

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

and is standardized in the range [-1,1]. The covariance and standard deviation cannot isolate and differentiate the information present in each of the four possible relationships between two variables where the target is some reference point:

$X \leq$ target, $Y \leq$ target

$X \leq$ target, $Y >$ target

$X >$ target, $Y \leq$ target

$X >$ target, $Y >$ target

We propose a method of partitioning the distribution with partial moments to capture the information from each linear relationship embedded within a bi- or multivariate relationship (linear or nonlinear). Based on the above four relationships between two variables, a co- or divergent partial moment is constructed to quantify it.[i]

2.1 Co-Partial Moments

$$CLPM(n, h_x|h_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0, h_x - X_t\}^n \cdot max\{0, h_y - Y_t\}^n\right)\right] \quad (1)$$

$$CUPM(q, l_x|l_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{X_t - l_x, 0\}^q \cdot max\{0, Y_t - l_y\}^q\right)\right] \quad (2)$$

where X_t represents the observation X at time t, Y_t represents the observation Y at time t, n is the degree of the LPM, q is the degree of the UPM, h_x is the target for computing below target observations for X, and l_x is the target for computing above target observations for X. For simplicity we assume that $h_x = l_x$.

2.2 Divergent Partial Moments

$$DLPM(q|n, h_x|h_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{X_t - h_x, 0\}^q \cdot max\{0, h_y - Y_t\}^n\right)\right] \quad (3)$$

$$DUPM(n|q, h_x|h_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0, h_x - X_t\}^n \cdot max\{Y_t - h_y, 0\}^q\right)\right] \quad (4)$$

2.3 Definition of Variable Relationships:

$X \leq$ target, $Y \leq$ target $\rightarrow$ $CLPM(n, h_x|h_y, X|Y)$

$X \leq$ target, $Y >$ target $\rightarrow$ $DUPM(n|q, h_x|h_y, X|Y)$

$X >$ target, $Y \leq$ target $\rightarrow$ $DLPM(q|n, h_x|h_y, X|Y)$

$X >$ target, $Y >$ target $\rightarrow$ $CUPM(q, h_x|h_y, X|Y)$

To avoid the blunt covariance and standard deviation dependence of the Pearson coefficient, we can use the following nonparametric formula in equation 5 to determine the correlation (linear or nonlinear) between two variables.

$$\rho_{xy} = \frac{[CLPM(0, h_x|h_y, x|y) - DLPM(0, h_x|h_y, x|y) - DUPM(0, h_x|h_y, x|y) + CUPM(0, h_x|h_y, x|y)]}{[CLPM(0, h_x|h_y, x|y) + DLPM(0, h_x|h_y, x|y) + DUPM(0, h_x|h_y, x|y) + CUPM(0, h_x|h_y, x|y)]} \quad (5)$$

The axiomatic relationship between correlation and co- or divergent returns follows. If there is a -1 correlation, then the returns between the variables will always be divergent, thus,

$$\rho_{xy} = \frac{[0 - DLPM(0, h_x|h_y, x|y) - DUPM(0, h_x|h_y, x|y) + 0]}{[0 + DLPM(0, h_x|h_y, x|y) + DUPM(0, h_x|h_y, x|y) + 0]} = -1 \quad (6)$$

If there is a perfect correlation between two variables, then there will be no divergent returns, thus,

$$\rho_{xy} = \frac{[CLPM(0, h_x|h_y, x|y) - 0 - 0 + CUPM(0, h_x|h_y, x|y)]}{[CLPM(0, h_x|h_y, x|y) + 0 + 0 + CUPM(0, h_x|h_y, x|y)]} = 1 \quad (7)$$

If there is zero correlation between two variables, then the co- and divergent returns will be of equal frequency or magnitude (degree zero and degree one respectively),

$$CLPM(0, h_x|h_y, x|y) = DLPM(0, h_x|h_y, x|y) = DUPM(0, h_x|h_y, x|y) = CUPM(0, h_x|h_y, x|y)$$

Thus,

$$\rho_{xy} = \frac{[CLPM(0, h_x|h_y, x|y) - DLPM(0, h_x|h_y, x|y) - DUPM(0, h_x|h_y, x|y) + CUPM(0, h_x|h_y, x|y)]}{[CLPM(0, h_x|h_y, x|y) + DLPM(0, h_x|h_y, x|y) + DUPM(0, h_x|h_y, x|y) + CUPM(0, h_x|h_y, x|y)]} = 0 \quad (8)$$

Degree one can be substituted for parameters *n* and *q*, to generate correlations whereby the magnitude of the target deviations are compared; thus generating a dependence coefficient.

2.4 Visualization of the Partitions Using Means as Targets:

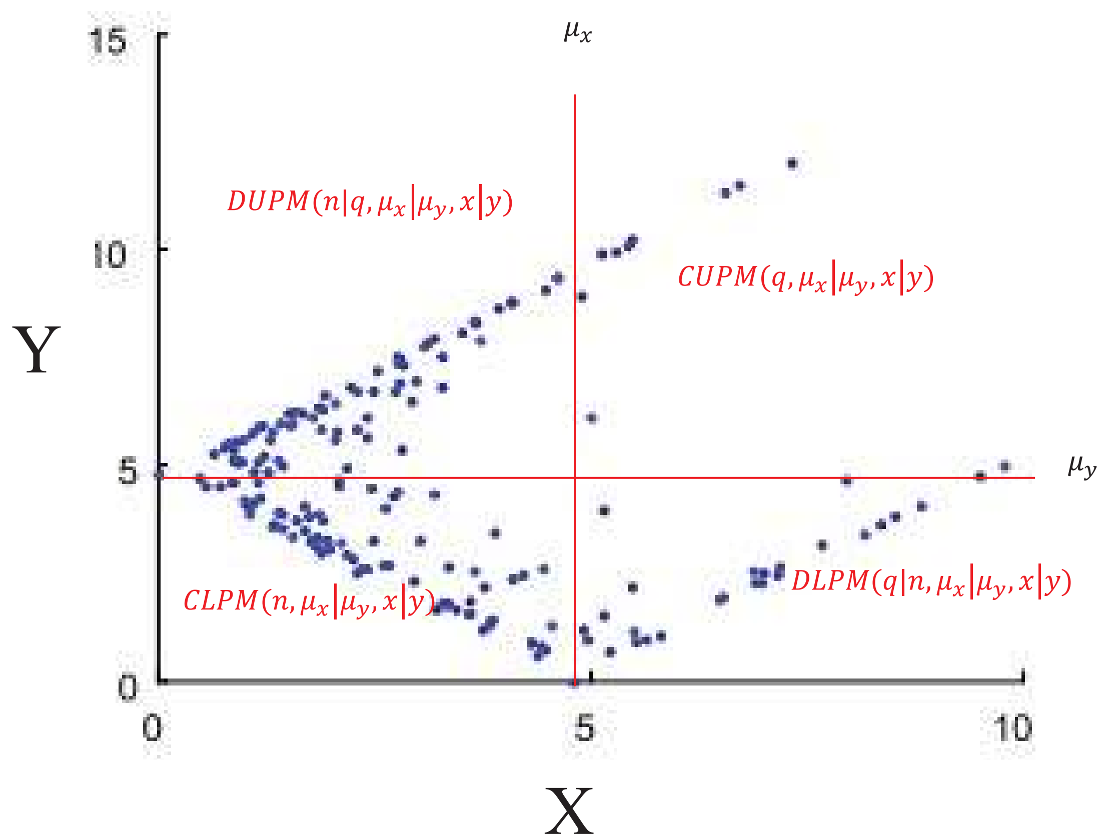


Figure 1. 1st order partitioning of the distribution based on variable relationships with co- and divergent partial moments on an observed nonlinear correlation in a microarray study from Chen et al. (2010).

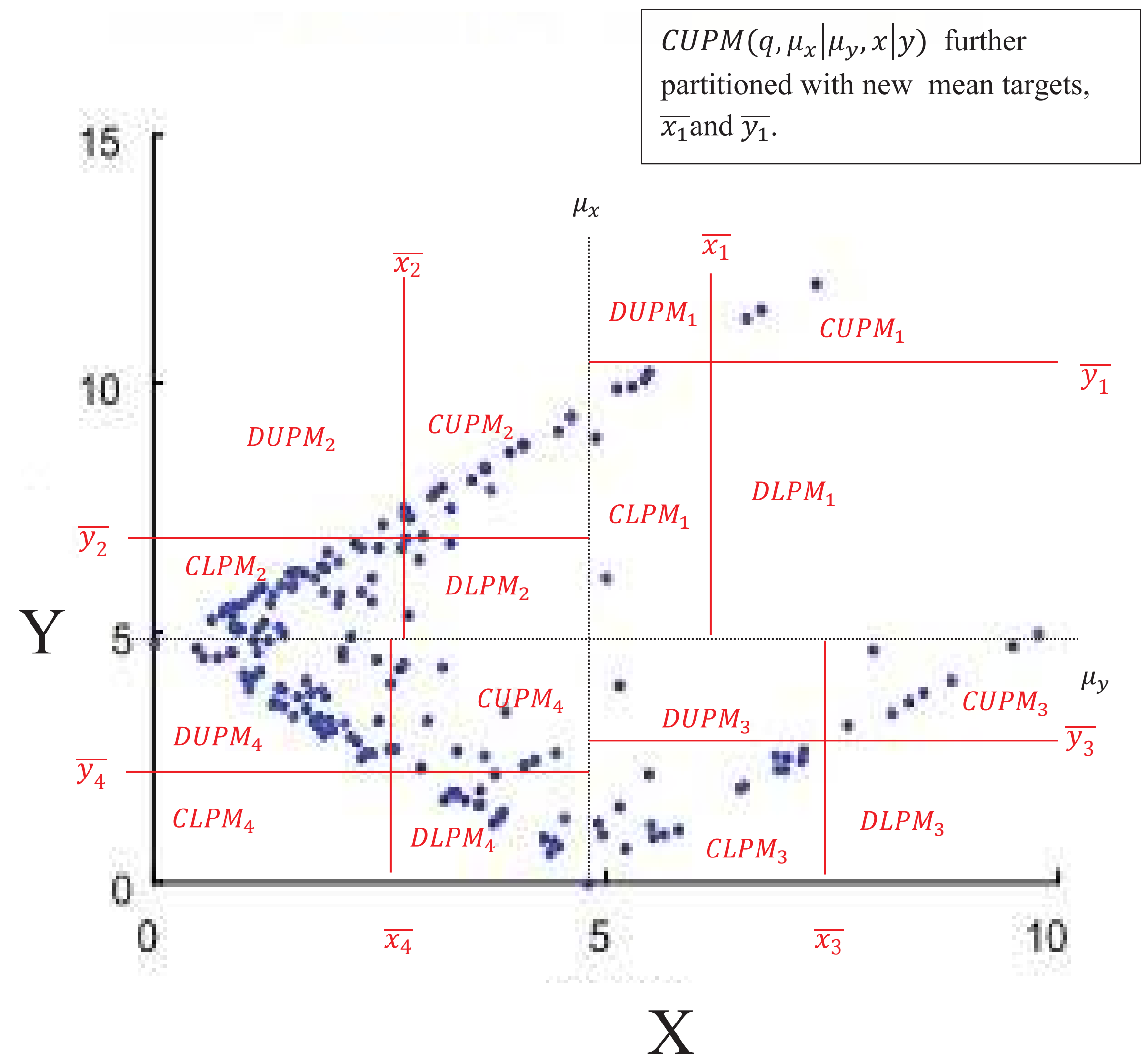


Figure 2. 2nd order partitioning of the microarray study based on means of partial moment subsets as targets.

2.5 Definition of Variable Subsets:

$\{x_1, y_1\} \in CUPM(q, \mu_x|\mu_y, x|y)$

$\{x_2, y_2\} \in DLPM(q|n, \mu_x|\mu_y, x|y)$

$\{x_3, y_3\} \in CLPM(n, \mu_x|\mu_y, x|y)$

$\{x_4, y_4\} \in DUPM(n|q, \mu_x|\mu_y, x|y)$

2.6 Definition of Subset Means:

$$\overline{x_1} = \frac{\sum_{n=1}^{n} x_{1_n}}{n} \qquad \overline{y_1} = \frac{\sum_{n=1}^{n} y_{1_n}}{n}$$

$$\overline{x_2} = \frac{\sum_{n=1}^{n} x_{2_n}}{n} \qquad \overline{y_2} = \frac{\sum_{n=1}^{n} y_{2_n}}{n}$$

$$\overline{x_3} = \frac{\sum_{n=1}^{n} x_{3_n}}{n} \qquad \overline{y_3} = \frac{\sum_{n=1}^{n} y_{3_n}}{n}$$

$$\overline{x_4} = \frac{\sum_{n=1}^{n} x_{4_n}}{n} \qquad \overline{y_4} = \frac{\sum_{n=1}^{n} y_{4_n}}{n}$$

2.7 Defintion of Subset Partial Moments:

$$CUPM_1(q, \overline{x_1}|\overline{y_1}, X_1|Y_1) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0, x_{1_t} - \overline{x_1}\}^q \cdot max\{0, y_{1_t} - \overline{y_1}\}^q\right)\right] \tag{9}$$

$$DLPM_1(q|n, \overline{x_1}|\overline{y_1}, X_1|Y_1) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{x_{1_t} - \overline{x_1}, 0\}^q \cdot max\{0, \overline{y_1} - y_{1_t}\}^n\right)\right] \tag{10}$$

$$CLPM_1(n, \overline{x_1}|\overline{x_1}, X_1|Y_1) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0, \overline{x_1} - x_{1_t}\}^n \cdot max\{0, \overline{y_1} - y_{1_t}\}^n\right)\right] \tag{11}$$

$$DUPM_1(n|q, \overline{x_1}|\overline{y_1}, X_1|Y_1) = \frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0, \overline{x_1} - x_{1_t}\}^n \cdot max\{y_{1_t} - \overline{y_1}, 0\}^q\right)\right] \tag{12}$$

For a 3rd order analysis for example, one needs to then compute the 12 remaining subset partial moments (in addition to the four identified in equations 9-12 above) using the appropriate subset mean targets for each quadrant. The total amount of subset means will be less than or equal to $4^{(N-1)}$ where N is the number of orders specified.[ii]

The eventual correlation metric is accomplished by adding all CUPM's and CLPM's (positive correlations) and subtracting DUPM's and DLPM's (negative correlations) in the numerator, while summing all 16 co- and divergent partial moments representing the entire distribution in the denominator per equation 13 below.

$\rho_{xy} =$

Numerator:

$(CLPM_1 + CLPM_2 + CLPM_3 + CLPM_4 - DLPM_1 - DLPM_2 - DLPM_3 - DLPM_4 - DUPM_1 - DUPM_2 - DUPM_3 - DUPM_4 + CUPM_1 + CUPM_1 + CUPM_3 + CUPM_4)$

Denominator:

$(CLPM_1 + CLPM_2 + CLPM_3 + CLPM_4 + DLPM_1 + DLPM_2 + DLPM_3 + DLPM_4 + DUPM_1 + DUPM_2 + DUPM_3 + DUPM_4 + CUPM_1 + CUPM_1 + CUPM_3 + CUPM_4)$

(13)

2.8 Dependence:

We can also define the dependence present between two variables as the sum of the absolute value of the per quadrant correlations. Stated differently, when all of the per quadrant observations are either the CLPM & CLPM, or DLPM & DUPM, the variables are dependent upon one another.

$$\eta(X,Y) = |\rho_{CLPM}| + |\rho_{CUPM}| + |\rho_{DLPM}| + |\rho_{DUPM}| \tag{14}$$

Where the CLPM quadrant's correlation is given by

$$|\rho_{CLPM}| = \left|\frac{CLPM_4 + CUPM_4 - DLPM_4 - DUPM_4}{CLPM_4 + CUPM_4 + DLPM_4 + DUPM_4}\right|$$

Equation 14 describes the amount of nonlinearity present in each quadrant when the negative correlations are equal in frequency or magnitude (depending on degree 0 or 1 respectively) to the positive correlations.

When $\eta(X,Y)$ equals one, there is maximum *dependence* between the two variables. As $\eta(X,Y)$ approaches 0, the relationship is approaching maximum *independence*.

3. EMPIRICAL EVIDENCE:

Third order partitions are shown and calculated in R. The 1st order partition is the thick red line (per Figure 1), the 2nd order partition is the thin red line (per Figure 2) and the 3rd order partition is the dotted black line.

Linear Equalities:[iii]

$Y = 2X$

```
> x=seq(-3,3,.01);y=2*x
> cor(x,y)
[1] 1
> NNS.dep(x,y,print.map = T)
$Correlation
[1] 1
$Dependence
[1] 1
```

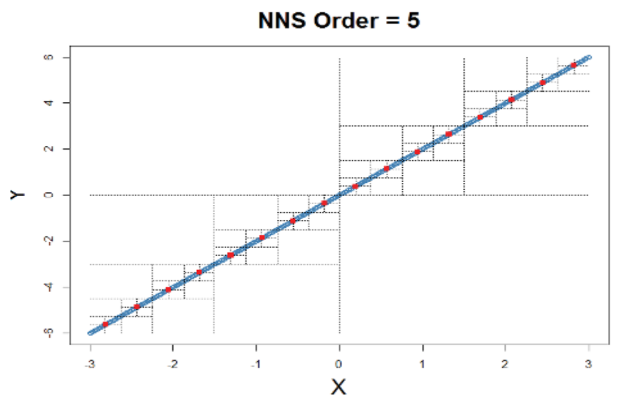


Figure 3. Linear positive relationship between two variables (X, Y).

$Y = -2X$

```
> x=seq(-3,3,.01);y=-2*x
> cor(x,y)
[1] -1
> NNS.dep(x,y,print.map = T)
$Correlation
[1] -1
$Dependence
[1] 1
```

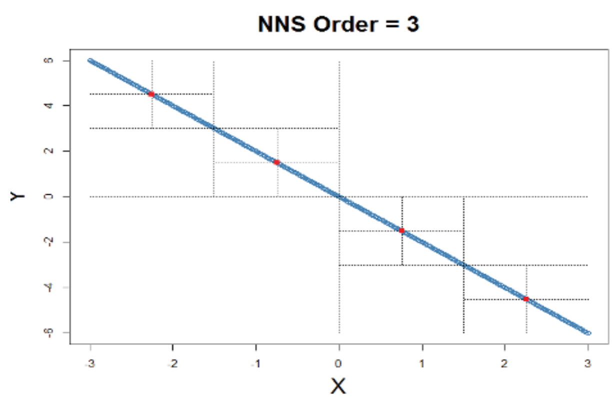


Figure 4. Linear inverse relationship between two variables (X, Y).

Nonlinear Differences:

$Y = X^2$ for positive X

```
> x=seq(0,3,.01);y=x^2
> cor(x,y)
[1] 0.9680452
> NNS.dep(x,y,print.map = T)
$Correlation
[1] 0.9994402
$Dependence
[1] 0.9994402
```

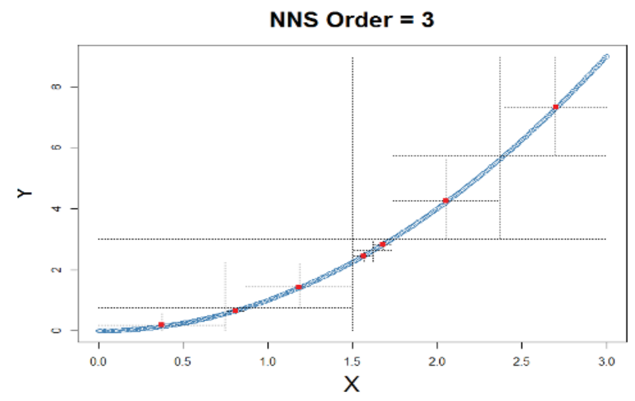


Figure 5. Nonlinear positive relationship between two variables (X, Y).

$Y = X^2$

```
> x=seq(-3,3,.01);y=x^2
> cor(x,y)
[1] 7.665343e-17
> NNS.dep(x,y,print.map = T)
$Correlation
[1] -0.001647721
$Dependence
[1] 0.9993975
```

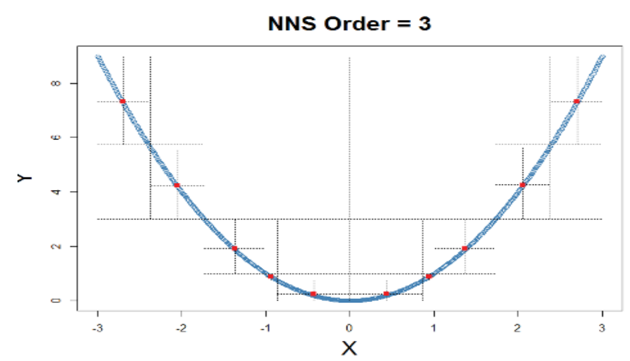


Figure 6. Nonlinear relationship between two variables (X, Y).

As the exponential function increases in magnitude, we actually find it to retain its linear relationship…

$\boldsymbol{Y = X^{10}}$

```
> x=seq(0,3,.01);y=x^10
> cor(x,y)
[1] 0.6610183
> NNS.dep(x,y,print.map = T)
$Correlation
[1] 0.9812511
$Dependence
[1] 0.9812511
```

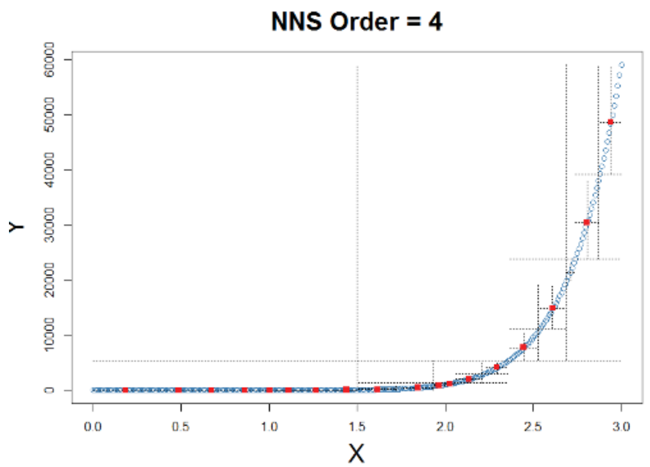


Figure 7. Nonlinear positive relationship between two variables (X, Y).

And a completely nonlinear clustered dataset, where coefficient weighting due to partition occupancy is exemplified.

$\boldsymbol{Y = undetermined\ f(x)}$

```
>
cor(cluster.df[,3],cluster.df[,4])
[1] -0.6275592
> NNS.dep(cluster.df[,3],
cluster.df[,4],print.map = T)
$Correlation
[1] -0.1020994
$Dependence
[1] 0.2637387
```

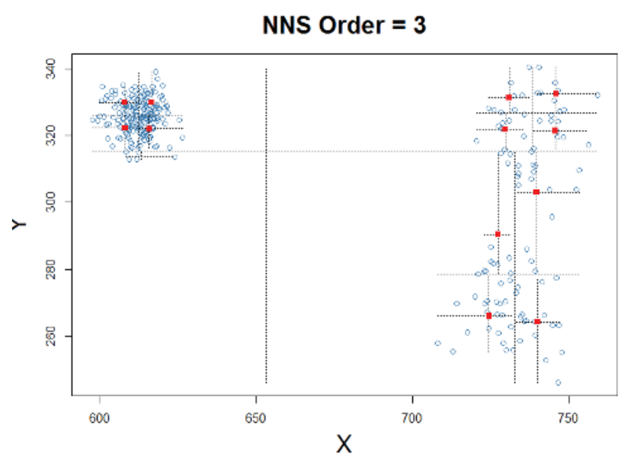


Figure 8. Nonlinear relationship between two variables (X, Y).

4. MULTIDIMENSIONAL NONLINEAR ANALYSIS:

To find the 1st order aggregate correlation for more than two dimensions, the method is similar to what was just presented. Instead of co- and divergent partial moments, we are going to substitute co- and divergent partial moment *matrices* into equation 5. A *n* x *n* matrix for each of the interactions (CLPM, DLPM, DUPM and CUPM) per Viole and Nawrocki (2012a), can be constructed and treated analogously to the direct partial moment computation.

Thus,

$$CLPM_{matrix}(0,h_x \dots h_n, x \dots n) = \begin{pmatrix} CLPM(0,h_x|h_x,x|x) & \cdots & CLPM(0,h_x|h_n,x|n) \\ \vdots & \ddots & \vdots \\ CLPM(0,h_n|h_x,n|x) & \cdots & CLPM(0,h_n|h_n,n|n) \end{pmatrix} \qquad (15)$$

Yielding,

$$\rho_{x \dots n} = \frac{[CLPM_{matrix}(0,h_x \dots h_n,x \dots n) - DLPM_{matrix}(0,h_x \dots h_n,x \dots n) - DUPM_{matrix}(0,h_x \dots h_n,x \dots n) + CUPM_{matrix}(0,h_x \dots h_n,x \dots n)]}{[CLPM_{matrix}(0,h_x \dots h_n,x \dots n) + DLPM_{matrix}(0,h_x \dots h_n,x \dots n) + DUPM_{matrix}(0,h_x \dots h_n,x \dots n) + CUPM_{matrix}(0,h_x \dots h_n,x \dots n)]} \qquad (16)$$

Whereby the final result will be an equal sized *n* x *n* matrix,

$$\rho_{x \dots n} = \begin{pmatrix} \rho_{xx} & \cdots & \rho_{xn} \\ \vdots & \ddots & \vdots \\ \rho_{nx} & \cdots & \rho_{nn} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & \rho_{xn} \\ \vdots & \ddots & \vdots \\ \rho_{nx} & \cdots & 1 \end{pmatrix}$$

To derive the overall correlation, we need to sterilize the main diagonal of 1's (which are self-correlations) with the following formula,

$$\rho_{x\dots n} = \frac{\left[\sum \begin{pmatrix} \rho_{xx} & \cdots & \rho_{xn} \\ \vdots & \ddots & \vdots \\ \rho_{nx} & \cdots & \rho_{nn} \end{pmatrix} - n\right]}{n^2 - n} \tag{17}$$

Again, if the variables are all below or above their respective targets at time t, the CLPM and CUPM matrices respectively will capture that information. If the variables are i.i.d., the likelihood that one variable would diverge at time t increases as n increases, reducing $\rho_{x\dots n}$.

Further order partition analysis can be translated to the multidimensional by creating matrices for each of the identified subsets for all of the variables.

4.1 Nonlinear Regression Analysis:

The target means from which the four partial moment matrices are calculated also serve as the basis for a nonlinear regression. By plotting all of the mean intersections, the linear segments will fit the underlying, nonparametrically. The increased order of portioning will generate more intersecting points (maximum of $4^{(N-1)}$) for a more granular analysis. Below is an example with 3rd order partitioning, generating a fit to the linear data.

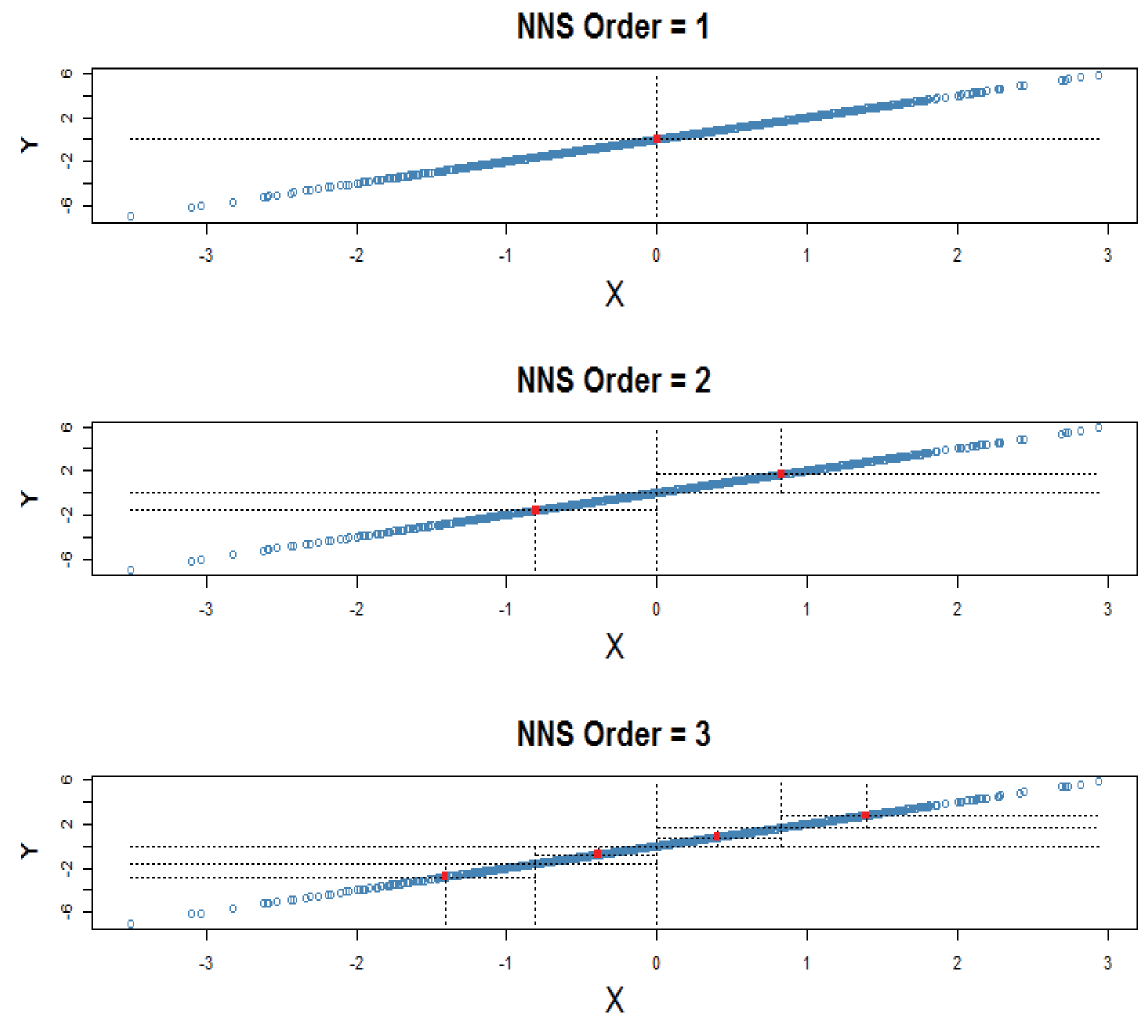


Figure 9. Nonparametric regression points for a linear relationship between (X, Y). Orders progressing restricted to the previous partition boundary.

We can also perform this on nonlinear relationships. Below is an example with 3rd order partitioning, generating a fit to an exponential relationship between the variables.

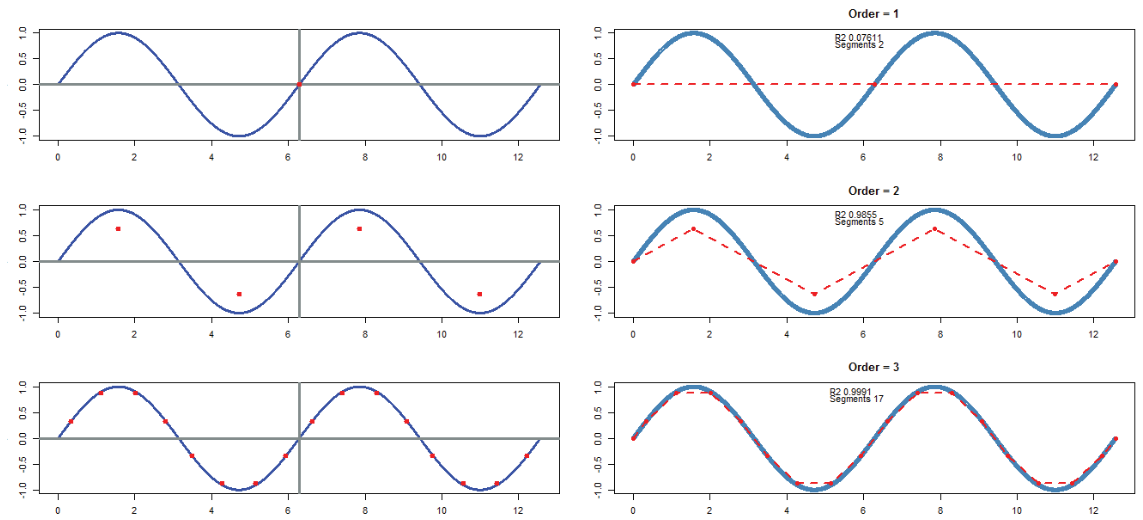


Figure 10. Nonparametric regression points for a nonlinear relationship between (X, Y). As partition orders increase, the curve is better fit.

Generating a multiple variable nonlinear regression analysis requires creating a synthetic variable. This variable, X* is the weighted average of all of the explanatory variables. The weighting is the nonlinear correlation derived from the *n* x *n* matrix where the explanatory variables are on the same row as the dependent variable which will have a 1.0 self-correlation. Thus, an explanatory variable with zero correlation to the dependent variable will be excluded from consideration.

Thus,

$$X^* = \frac{\sum_{i=1}^{n}(\rho_{y,x_i})(x_i)}{n} \tag{18}$$

And the nonlinear multiple regression can be performed in kind to the two variable example above with means of Y, X* as intersection points. This is similar to a nonparametric local means regression only the number of means has to be a factor of 4 due to the four partial moment matrices per each analysis.

Figure 11 below is the nonlinear correlation matrix and the subsequent weightings for the multiple variable nonlinear regression using SPY as the dependent variable with TLT, GLD, FXE, and GSG as explanatory variables.[iv] The data involved 100 daily observations from 5/8/12 through 9/27/12 for all variables. As shown in Viole and Nawrocki (2012c) partial moments asymptotically converge to the area of the function, and stabilize with approximately 100 observations.

```
> NNS.cor(ReturnsDF,order=3)
            GSG         GLD         TLT        FXE         SPY
GSG  1.00000000 -0.10111213 -0.05050505 0.06070809  0.11111111
GLD -0.10111213  1.00000000  0.23232323 0.21212121  0.03030303
TLT -0.05050505  0.23232323  1.00000000 0.15151515 -0.23242629
FXE  0.06070809  0.21212121  0.15151515 1.00000000  0.23232323
SPY  0.11111111  0.03030303 -0.23242629 0.23232323  1.00000000
```

Figure 11. Nonlinear correlation matrix for 5 variables (SPY, TLT, GLD, FXE, GSG). Highlighted row isolates the coefficients for equation 18.

In this example per equation 18 our aggregated explanatory variable is,

$$X^* = \frac{-0.23(TLT) + 0.03(GLD) + 0.23(FXE) + 0.11(GSG)}{4}$$

Again, there are no multicollinearity issues with the explanatory variables, it simply does not matter if they are correlated or not. Below in figure 13 is the graph of this analysis with our 3rd order fit.

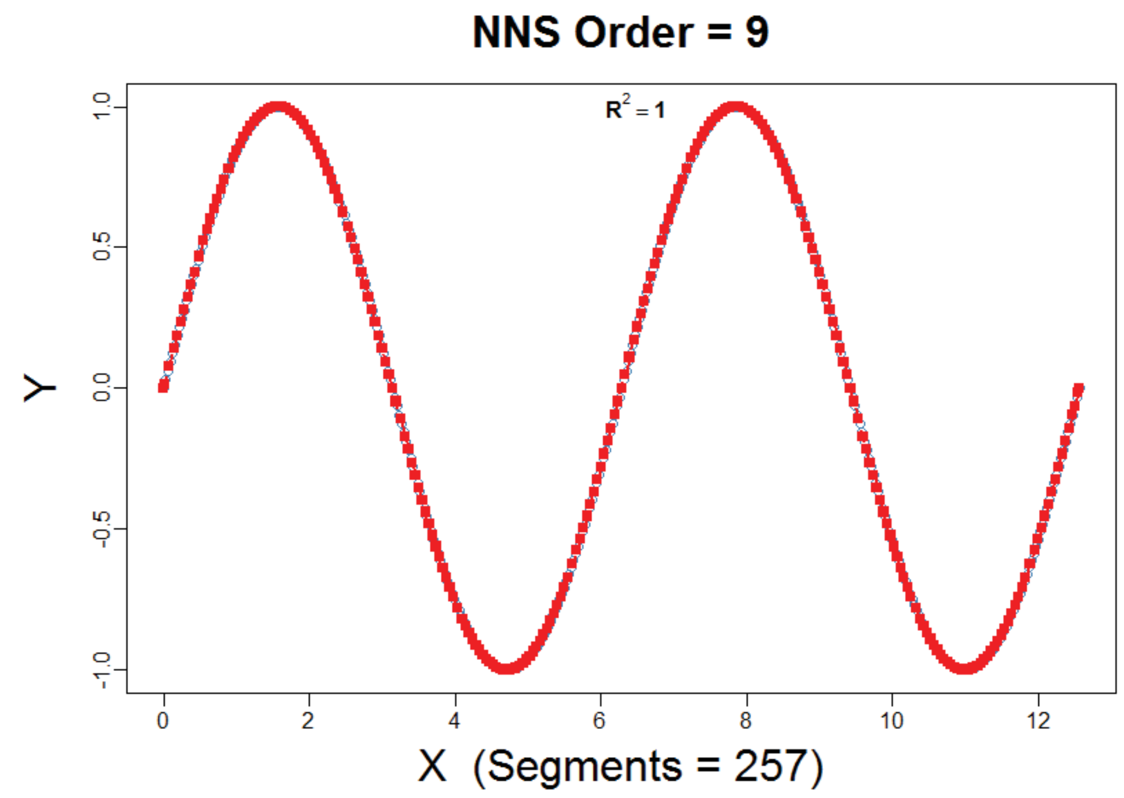


Figure 12. Our 9th order fit for a sine wave function of X.

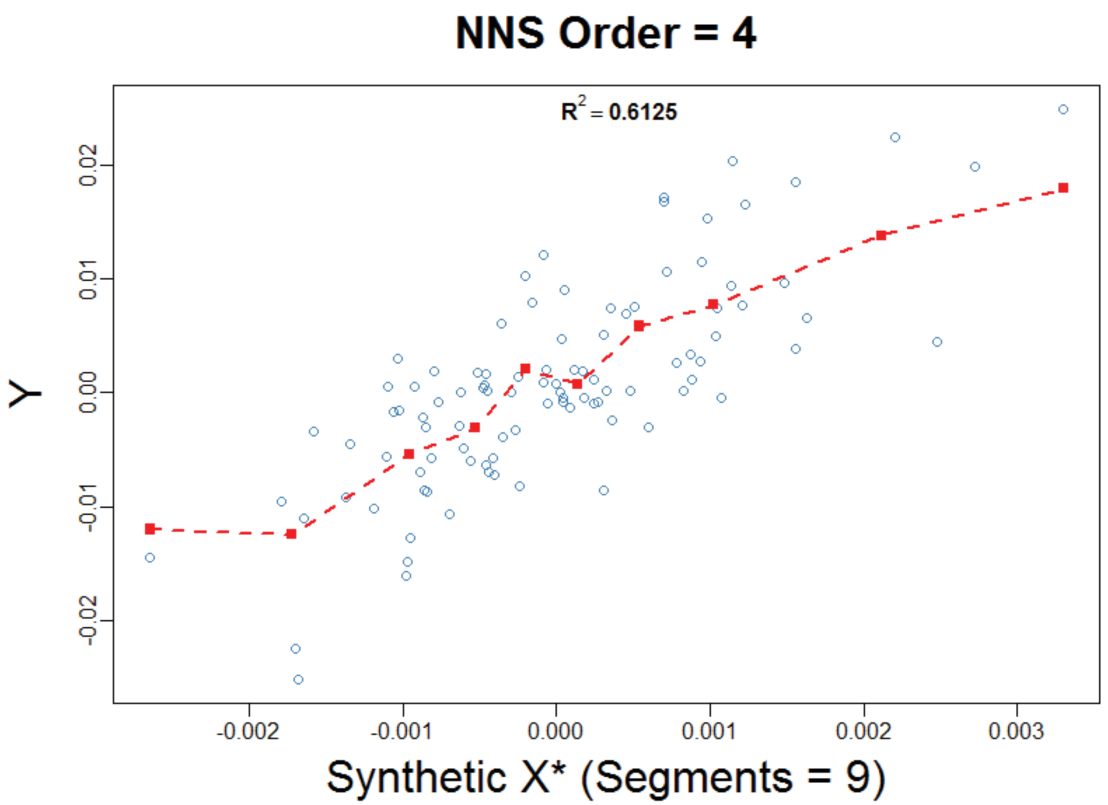


Figure 13. Our 4th order fit for an undetermined function of X*.

5. DISCUSSION AND SUMMARY:

There is no argument as to why the partition cannot be further specified N times, ultimately yielding a 4^N number of segments. The partial moments are direct computations, just as other statistics such as means and variances. The obvious benefit is the ability to parse what was referred to as "noise" into valid information. Due to the fact that individual observations are weighted by $\left(\frac{1}{T}\right)$, the number of observations in each segment will weigh the segment accordingly; thus affirming outlier observation status for such instances where a segment has minimal occupancy.

The purpose of this paper was to put forth a nonparametric, nonlinear correlation metric where Chen et al. (2010) note, "there is no commonly use statistic quantifying nonlinear correlation that can find a similarly generic use as Pearson's correlation coefficient for quantifying linear correlation." Our linear sum of the weighted micro does indeed capture the aggregate correlation. But, unlike Pearson's single correlation coefficient, we also generate the information necessary to reconstruct the relationship from the individual partial moment matrices. As for a direct policy statement resulting from the nonlinear regression analysis; it would have to assume the form of a conditional equation whereby each linear segment is defined for a specific range of the explanatory variable(s).

Autoregressive Modeling

ABSTRACT

Using component series from a given time series, we are able to demonstrate forecasting ability with none of the requirements of the traditional ARMA method, while strictly adhering to the definition of an autoregressive model. We also propose a new test for seasonality using coefficient of variance comparisons for component series, and then extend this proposed method to non-seasonal data. The resulting effect is that of conditional heteroskedasticity on the forecast with more accurate forecasts derived from implementing nonlinear regressions into the component series.

INTRODUCTION

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series.[10]

In this article we aim to present a method of autoregressive modeling strictly adhering to the above definition. We accomplish this by using a linear regression of like data points excluded from the total time series. For instance, in monthly data, we will examine the "January" data points autonomously to generate the ex ante "January" observation.

Testing for seasonality of each of the monthly classifications will alert us weather to incorporate other months' data in the linear regression. Through simple examples, we will show how the steps of:

- Model Identification
- Model estimation
- Diagnostic Testing
- Forecasting

Will be reduced to that of:

- Separating like classifications
- Testing for seasonality
- Regression / Forecasting

We will also demonstrate how the ARIMA requirement of stationarity of the time series is no longer necessary to forecast while no data will be lost to differencing techniques.

---

[10] http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc444.htm

METHODOLOGY

In his 2008 article, Wang explains how to use Box-Jenkins models for forecasting. He uses an example of the quarterly electric demand in New York City from the first quarter of 1995 through the fourth quarter of 2005.

Figure 1 clearly shows that the demand data are quarterly seasonal trending upward; consequently, the mean of the data will change over time. We can define that a stationary time series has a constant mean and has no trend overtime. A plot of the data is usually enough to see if the data are stationary. In practice, few time series can meet this condition, but as long as the data can be transformed into a stationary series, a Box-Jenkins model can be developed. As defined above, this time series is not stationary.

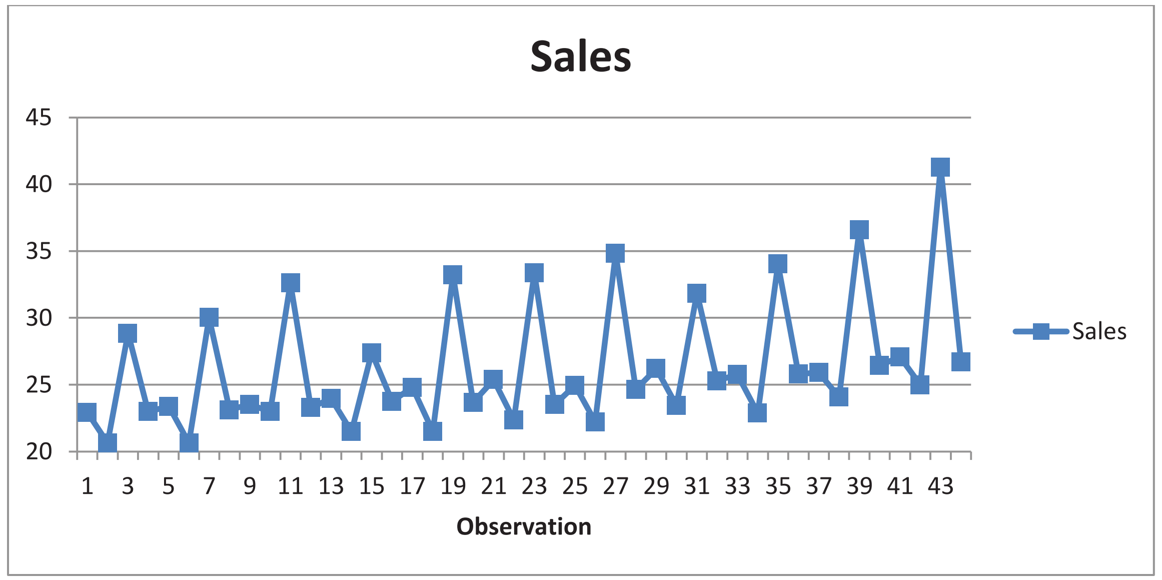


Figure 1. Recreation of data set from Wang [2008] based on quarterly electric demand in New York City from the first quarter of 1995 through the fourth quarter of 2005.

I. COMPONENT SERIES

Our first step is to break the time series down into like classifications. In this example, first quarter data will be aggregated to form a first quarter time series. The vectors of observation number and sales are given below

Observation number = {1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41}

Sales = {22.91, 23.39, 23.51, 23.97, 24.81, 25.37, 24.95, 26.21, 25.76, 25.91, 27.08}

Vectors for Quarters 2 through 4 will be created analogously using every fourth observation starting from the corresponding quarter number and the sales data.

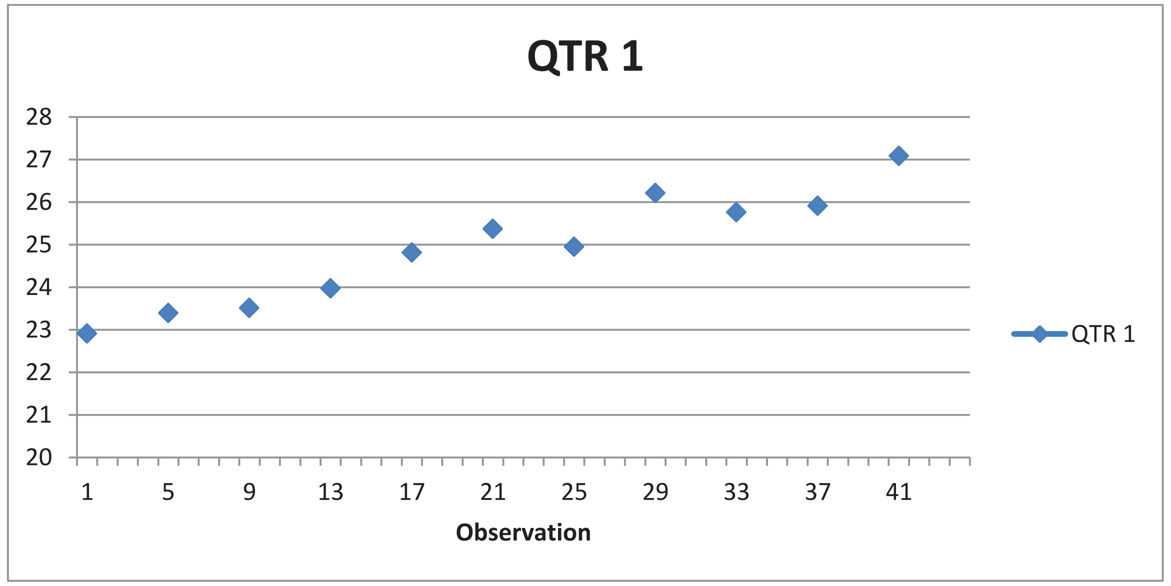


Figure 2. First quarter series isolated from original time series.

II. SEASONALITY

In order to test for seasonality, outside of the recommended "eyeball test" of the plotted data, we propose another method. If each of the quarterly series' coefficient of variance (σ/ μ) is less than the total sample coefficient of variance, seasonality exists. In our working example, the variances and means are presented in table 1 below.

	Full Sample	QTR 1	QTR 2	QTR 3	QTR 4
σ	4.589798	1.261198	1.313679	3.632291	1.306242
μ	26.23295	24.89727	22.47545	33.09091	24.46818
σ/ μ	0.174963	0.050656	0.058449	0.109767	0.053385

Table 1. Variances and means for full sample vs. each quarterly series. The coefficient of variance (σ/ μ) is less than the sample for all component series, indicating seasonality present in the data.

In monthly time series from 1/2000 through 5/2013 for the S&P 500, we find the total coefficient of variation to equal 0.158665526 with the "Janurary" series coefficient of variation equal to 0.16710549, thus negating the seasonality consideration (and enabling the data for a conditional heteroskedasticity treatment we will illustrate later).[11]

[11] Plots of total and monthly series are in the Appendix.

III. LINEAR REGRESSION

In order to adhere to the autoregressive definition provided in the introduction, we need to use a linear regression on the prior values of a variable. We have just created a subset of those values with like classifications to perform the regression.

Figure 3 below is the linear regression of the QTR 1 series. The regression equation is

$$y = 0.0961x + 22.878$$

Thus, our estimate for the next QTR 1 observation (the 45th observation overall)[12] is

$$y = 0.0961*45 + 22.878$$

$$y = 27.203$$

This is fairly close to the Box-Jenkins model result provided in Wang [2008] of 27.40. Again, we have lost no observations due to differencing in order to transform the data into a stationary series. Aside from the nonstationarity of the quarterly series, we note the linear approximation of the data as evidenced by the high $R^2$of 0.9297. This linearity is not necessary as will be discussed later when we introduce the nonlinear regression method to the discussion.

[12] The same series can be regressed on its own index, for this example (1:11).

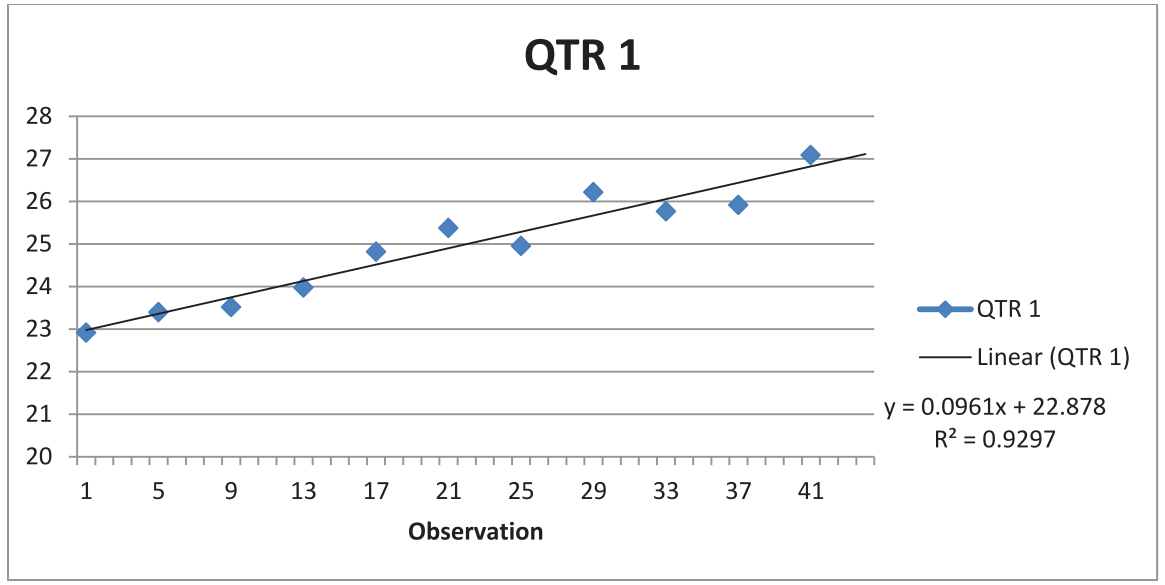


Figure 3. QTR 1 plot with linear regression.

We extend the analysis to all four quarter series and generate the forecasts based on the linear regression of each series in figure 4 below. You will note the overall pattern resemblance of the estimates to the seasonal data set.

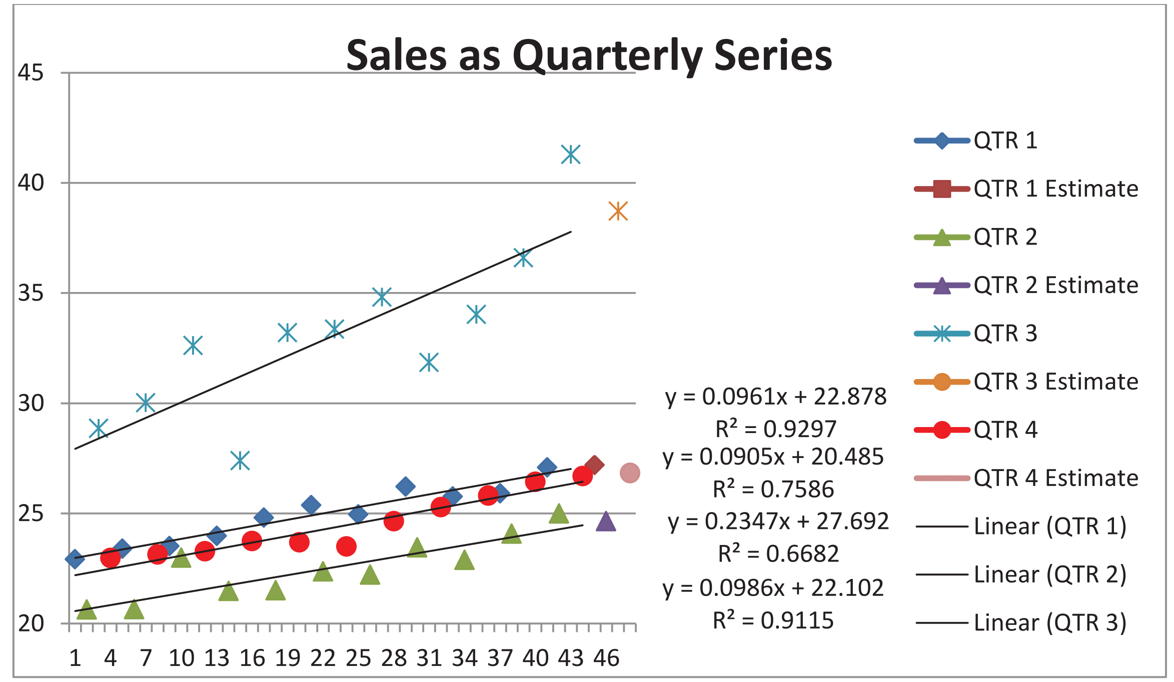


Figure 4. All quarterly plots with associated linear regressions and estimates for each quarterly series.

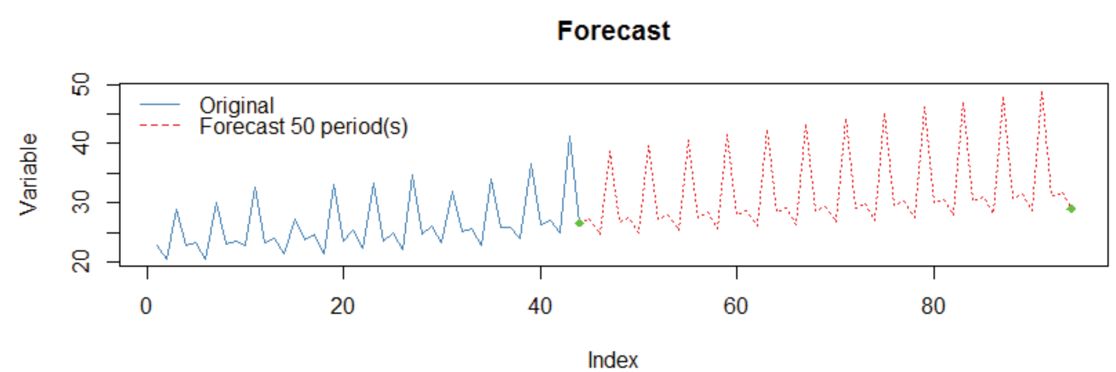


Figure 5. 50 period forecast using static 4 period lag and linear regression.

IV. CONDITIONAL HETEROSKEDASTICITY

We noted earlier that under seasonality of the data, it is a simple regression of the component series to generate a forecast. However, under the absence of perfect seasonality this is not the case. When a single seasonal period is not identified, we use a weighted average of all identified seasonal components.

Figure 6 illustrates the seasonal components to the Wang [2008] quarterly time series (data provided in Appendix). Note the strong seasonal presence in periods 4 and 8.

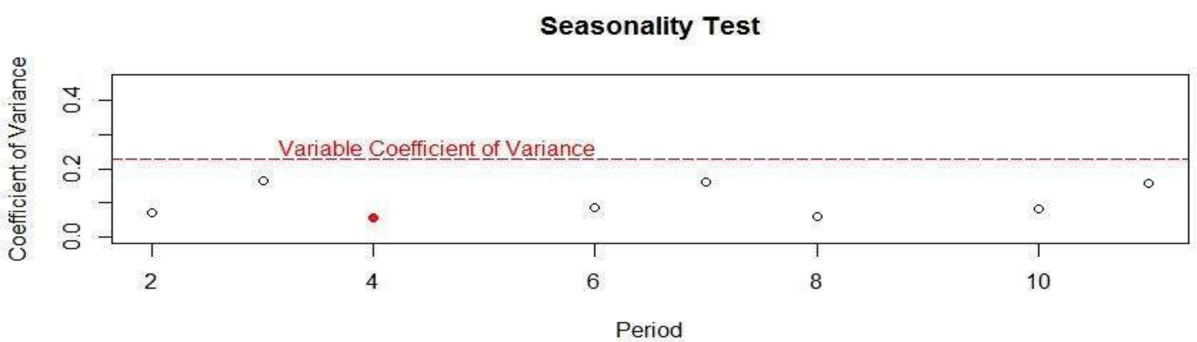


Figure 6. Periods (*i*) where $\frac{\sigma_i}{\mu_i} < \frac{\sigma_x}{\mu_x}$ for variable (*x*).

Period (*i*)	Coefficient of Variance $\frac{\sigma_i}{\mu_i}$	Variable Coefficient of Variance $\frac{\sigma_x}{\mu_x}$
2	0.07176943	0.1769858
3	0.16419383	0.1769858
4	0.05599103	0.1769858
6	0.08503594	0.1769858
7	0.15964245	0.1769858
8	0.06053440	0.1769858
10	0.08217461	0.1769858
11	0.15878767	0.1769858

Table 1. Coefficients of variance for all periods versus the variable coefficient of variance.

In this example, we perform 8 component regressions and the forecast output weights are determined by summing the inverses of each period's coefficient of variance.

Period)	**Intercept**	**+**	**β (t+1)**	**=**	**Forecast**
2)	24.6275325	+	0.3797007 (23)	=	33.36065
3)	23.1120879	+	0.3990549 (15)	=	29.09791
4)	22.5900000	+	0.3845455 (12)	=	27.20455
6)	23.874286	+	1.256071 (8)	=	33.92286
7)	25.87466667	+	0.03914286 (7)	=	26.14867
8)	22.786	+	0.728 (6)	=	27.154
10)	20.075	+	2.945 (5)	=	34.8
11)	23.110	+	0.999 (5)	=	28.105

Period (*i*)		Observations (t+1)	Output Weight
2		23	0.283950617
3		15	0.185185185
4		12	0.148148148
6		8	0.098765432
7		7	0.086419753
8		6	0.074074074
10		5	0.061728395
11		5	0.061728395
	SUM	81	1.0
Period (*i*)		Inverse Coefficient of Variance $\frac{\mu_i}{\sigma_i}$	Output Weight
2		13.93351	0.153293933
3		6.090362835	0.067005065
4		17.86000365	0.196492513
6		11.75973359	0.129378451
7		6.263998078	0.068915368
8		16.5195327	0.181744895
10		12.16920896	0.133883424
11		6.297718204	0.069286351
	SUM	90.89406702	1.0

Table 2. Forecast output weights for all periods demonstrating seasonality.

Forecast * Averaged Output Weight	**=**	**Weighted Forecast**
33.36065* 0.218622275	=	7.293381202
29.09791* 0.126095125	=	3.669104596
27.20455* 0.172320331	=	4.687897049
33.92286* 0.114071942	=	3.869646502
26.14867* 0.077667561	=	2.03090341
27.154* 0.127909485	=	3.473254147
34.8* 0.09780591	=	3.40364566
28.105* 0.065507373	=	1.841084715
Weighted Forecast Sum =		**30.269**

This technique places equal consideration on the number of observations in a component series and its coefficient of variance. Again, it should be reserved for instances of truly unknown seasonal periods and be more effective than a single seasonal factor on a test set from the sample.

NONLINEAR REGRESSION

There is not a strong argument as to why a linear regression is required in the autoregressive model. Perhaps it was due to the time in which the models were derived? Regardless, we can use a nonlinear regression method to derive more accurate forecasts than the stipulated linear regression. This option will handle the nonlinearity of the component series.

So even if the data for the component series resembles the sine wave function as in figure 7 below (we are highlighting the nonlinearity of the data, stationarity is irrelevant) we will be able to generate a more accurate series forecast. We can see that the linear regression would suggest a positive data point (in green), yet the nonlinear regression based on partial moments from Viole and Nawrocki [2012] would suggest a decidedly negative observation for their forecasts.

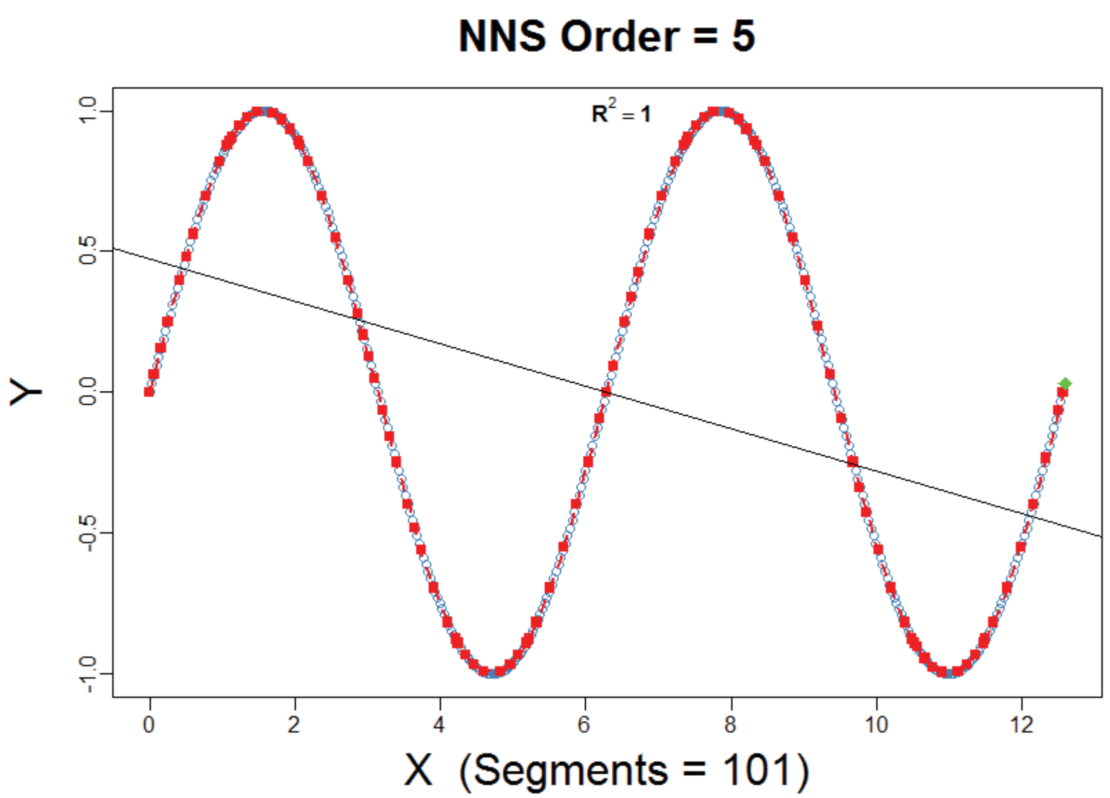


Figure 7. Nonlinear regression on a hypothetical component series used to highlight the inadequacy of a linear regression for forecasting even component series, let alone total series.

DISCUSSION

We have closely approximated the results from a Box-Jenkins method with an autoregressive model with no stationarity requirement, no model identification, capable of handling nonlinearity. The absence of requirements and the retention of all of the original data is a promising starting point to adhere to the definition of the process.

We have also introduced a method of detecting seasonality in time series data. This technique can be used in conjunction with existing methods to confirm the results found in tests with normalized data (typically autocorrelation plots of differenced data). In the absence of seasonality, we offer a simple procedure for giving equal representation of other component variance which typically influences the component series via conditional heteroskedasticity.

APPENDIX: Wang[2008] dataset.

Obs #	Value	Obs #	Value
1	22.9	33	25.76
2	20.63	34	22.88
3	28.85	35	34.02
4	22.97	36	25.8
5	23.39	37	25.91
6	20.65	38	24.07
7	30.02	39	36.6
8	23.13	40	26.43
9	23.51	41	27.08
10	22.99	42	24.99
11	32.61	43	41.29
12	23.28	44	26.69
13	23.97		
14	21.48		
15	27.39		
16	23.75		
17	24.81		
18	21.51		
19	33.2		
20	23.68		
21	25.37		
22	22.36		
23	33.36		
24	23.5		
25	24.95		
26	22.22		
27	34.81		
28	24.64		
29	26.21		
30	23.45		
31	31.85		
32	25.28		

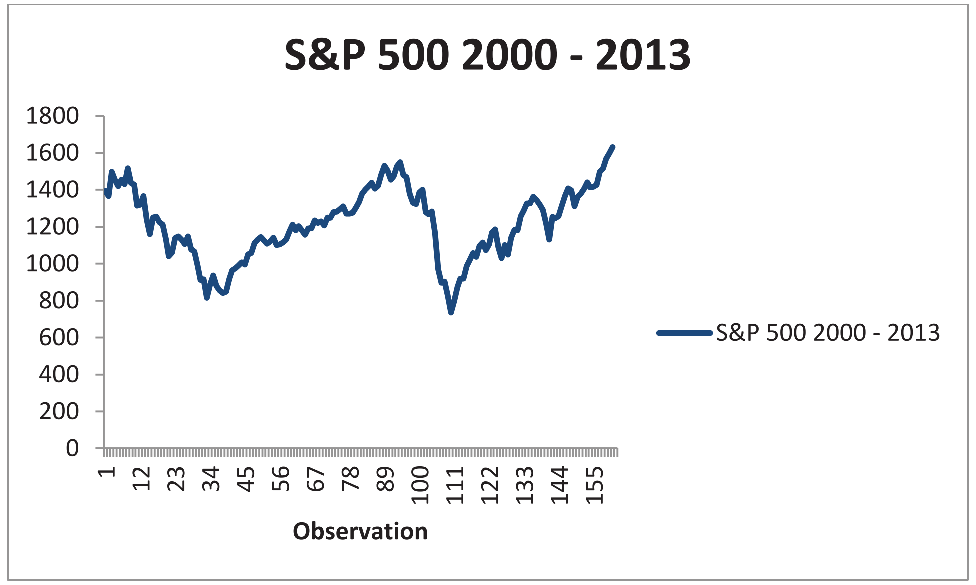


Figure 1A. S&P 500 monthly returns 1/2000 – 5/2013.

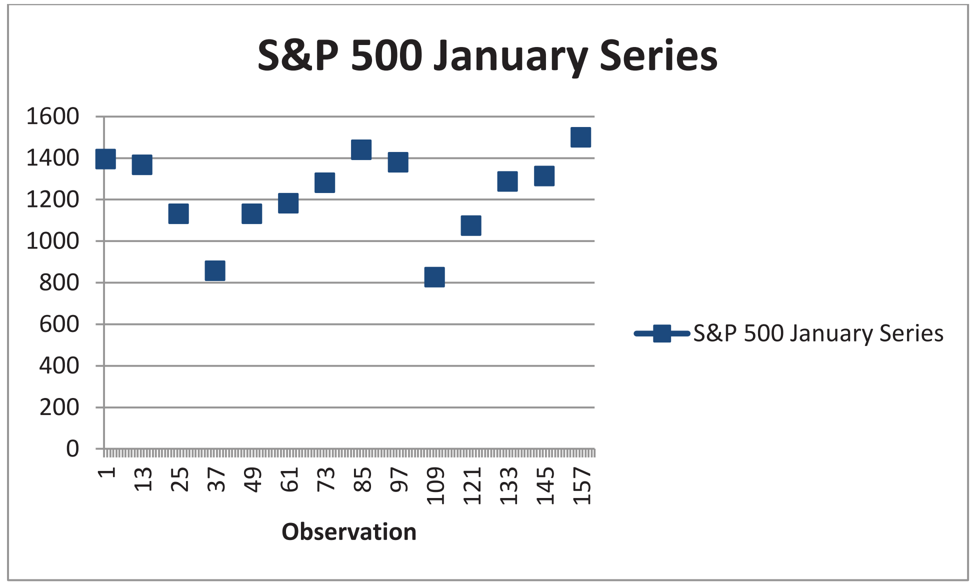


Figure 2A. S&P 500 January only returns 1/2000 – 5/2013.

APPLES TO APPLES COMPARISONS

NonLinear Scaling Normalization with Variance Retention

ABSTRACT

We present a nonlinear method of scaling to achieve normalization of multiple variables. We compare this method to the standard linear scaling and the quantile normalization methods. We find our overall normalized distribution to be more representative of the original data set with regards to standard moments of individual variables. We also find our normalized results to have an overall lower standard deviation versus both the linear scaling and quantile normalization results for variables with similar distributions.

INTRODUCTION

Normalization is the preferred technique for aligning and then comparing various data sets. However, this technique often loses the variance properties associated with the underlying distributions. The results are catastrophic on continuous variables, such that they are effectively transformed into discrete variables. Viole and Nawrocki [2012a] demonstrate this undesirable transformation for normalized variables.

We propose a new method of normalization that improves upon the linear scaling technique by incorporating a nonlinear association metric as proposed in Chen [2010], and Viole and Nawrocki [2012b]. In essence the typical linear scaling method assumes a linear relationship between variables.

We then compare these normalized data sets using our proposed nonlinear scaling technique, the linear scaling method, and quantile normalization.

METHODS

Linear Scaling

Linear scaling uses each set as a reference once, then averaging all of the iterations. This way original series for all is considered in the final normalization. It is an equitable treatment of the data, yet blunt in its approach.

The Genomics and Bioinformatics Group of the NIH describe the linear scaling process as:[13]

In practice, for a series of chips, define normalization constants C_1, C_2,…, by:

$$C_1 = \sum_{genes} f_1^{gene}, C_2 = \sum_{genes} f_2^{gene}, \text{and so on,}$$

where the numbers f_1^{gene} are the fluorescent intensities measured for each probe on chip *i*. Select a common total intensity K (eg. the average of the C_i's). Then to normalize all the chips to the common total intensity K, divide all fluorescent intensity readings from chip *i* by C_i., and multiply by K.

Quantile Normalization

The goal of the Quantile method is to make the distribution of probe intensities for each array in a set of arrays the same. Quantile normalization assumes that the distribution of gene abundances is nearly the same in all samples. For convenience Bolstad et al. [2003] take the pooled distribution of probes on all chips. Then to normalize each chip they compute for each value, the quantile of that value in the distribution of probe intensities; they then transform the original value to that quantile's value on the reference chip. In a formula, the transform is

$$x_{norm} = F_i^{-1}\left(F_{ref}(x)\right), \qquad (1)$$

[13] http://discover.nci.nih.gov/microarrayAnalysis/Affymetrix.Preprocessing.jsp

where F_i is the distribution function of chip *i*, and F_{ref} is the distribution function of the reference chip.

A quick illustration of such normalizing on a very small dataset:[14]

Arrays 1 to 3, genes A to D

A	5	4	3
B	2	1	4
C	3	4	6
D	4	2	8

For each column determine a rank from lowest to highest and assign number i-iv

A	iv	iii	i
B	i	i	ii
C	ii	iii	iii
D	iii	ii	iv

These rank values are set aside to use later. Go back to the first set of data. Rearrange that first set of column values so each column is in order going lowest to highest value. (First column consists of 5,2,3,4. This is rearranged to 2,3,4,5. Second Column 4,1,4,2 is rearranged to 1,2,4,4, and column 3 consisting of 3,4,6,8 stays the same because it is already in order from lowest to highest value.)

The result is:

A	5	4	3	becomes	A	2	1	3
B	2	1	4	becomes	B	3	2	4
C	3	4	6	becomes	C	4	4	6
D	4	2	8	becomes	D	5	4	8

[14] http://en.wikipedia.org/wiki/Quantile_normalization

Now find the mean for each row to determine the ranks

A (2 1 3)/3 = 2.00 = rank i

B (3 2 4)/3 = 3.00 = rank ii

C (4 4 6)/3 = 4.67 = rank iii

D (5 4 8)/3 = 5.67 = rank iv

Now take the ranking order and substitute in new values:

A	iv	iii	i
B	i	i	ii
C	ii	iii	iii
D	iii	ii	iv

becomes:

				Original		
A	5.67	4.67	2.00	**5**	**4**	**3**
B	2.00	2.00	3.00	**2**	**1**	**4**
C	3.00	4.67	4.67	**3**	**4**	**6**
D	4.67	3.00	5.67	**4**	**2**	**8**

This is the new normalized values. The new values have the same distribution and can now be easily compared.

OUR PROPOSED METHOD

The nonlinear association between variables is an important metric. It is also quite new to the literature. Chen et al. [2010] propose a method by using a rank transformation on the underlying data, while Viole and Nawrocki [2012b] propose a method based on the partial moments of the underlying data. VN will be the method employed for this analysis.

We define the amount of nonlinearity association present between two variables as.

$$\eta(X,Y) = |\rho_{CLPM}| + |\rho_{CUPM}| + |\rho_{DLPM}| + |\rho_{DUPM}| \tag{2}$$

Where,

Co-Partial Moments

$$CLPM(n, h_x|h_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}(max\{0, h_x - X_t\}^n \cdot max\{0, h_y - Y_t\}^n)\right] \tag{3}$$

$$CUPM(q, l_x|l_y, X|Y) = \frac{1}{T}\left[\sum_{t=1}^{T}(max\{X_t - l_x, 0\}^q \cdot max\{0, Y_t - l_y\}^q)\right] \tag{4}$$

where X_t represents the observation X at time t, Y_t represents the observation Y at time t, n is the degree of the LPM, q is the degree of the UPM, h_x is the target for computing below target observations for X, and l_x is the target for computing above target observations for X. For notational simplicity we assume that $h_x = l_x$ and $h_y = l_y$.

Divergent Partial Moments

$$DLPM(q|n,h_x|h_y,X|Y)=\frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{X_t-h_x,0\}^q\cdot max\{0,h_y-Y_t\}^n\right)\right] \quad (5)$$

$$DUPM(n|q,h_x|h_y,X|Y)=\frac{1}{T}\left[\sum_{t=1}^{T}\left(max\{0,h_x-X_t\}^n\cdot max\{Y_t-h_y,0\}^q\right)\right] \quad (6)$$

Definition of Variable Relationships:

$X\leq$ target, $Y\leq$ target → $CLPM(n,h_x|h_y,X|Y)$

$X\leq$ target, $Y>$ target → $DUPM(n|q,h_x|h_y,X|Y)$

$X>$ target, $Y\leq$ target → $DLPM(q|n,h_x|h_y,X|Y)$

$X>$ target, $Y>$ target → $CUPM(q,h_x|h_y,X|Y)$

Equation 2 describes the amount of nonlinearity present when the negative correlations (D-PM's) are equal in frequency or magnitude (depending on degree 0 or 1 respectively) to the positive correlations (C-PM's).

The nonlinear correlation between two variables is given by

$$\rho_{xy}=\frac{[CLPM(0,h_x|h_y,x|y)-DLPM(0,h_x|h_y,x|y)-DUPM(0,h_x|h_y,x|y)+CUPM(0,h_x|h_y,x|y)]}{[CLPM(0,h_x|h_y,x|y)+DLPM(0,h_x|h_y,x|y)+DUPM(0,h_x|h_y,x|y)+CUPM(0,h_x|h_y,x|y)]} \quad (7)$$

When $\eta(X,Y)$ equals one, there is maximum *dependence* between the two variables. As $\eta(X,Y)$ approaches 0, it is approaching maximum quadrant linearity. Per Viole and Nawrocki [2012b], the instances of maximum linearity $\eta(X,Y)=0$, are associated with maximum nonlinear correlation readings $\rho_{xy}=1\ or -1$. Thus the use of dependence is more aptly defining the nonlinear association between variables. For a complete treatment on nonlinear correlations and associations please see Viole and Nawrocki [2012b].

Using this nonlinear association metric as a factor in the normalization iterative process produces very different results than the assumed 1 (linearity) from the standard linear scaling method.

Figure 1 below illustrates the process for a 2 gene and a 4 gene example. Each gene has the desired property of serving as the reference gene (RG) in the process once. This consideration is identical to the standard linear scaling technique. From each RG's total intensity, we derive the RG factor for each gene to the RG. Simple enough. However, we then multiply each gene's observations by the RG factor *and* the nonlinear association between the genes $\eta(X,Y)$.

We repeat this process with every gene serving as the RG and then average all of the RG factored observations for each gene. The result is a fully normalized distribution for each gene with variance retention of the original data set.

	Gene	ORIGINAL DATA SET			Our Proposed Method							Linear Scaling Method						
					(UPM + LPM) "Total Intensity"	Reference Gene Factor	Nonlinearity Association Factor	Normalized Data Iterations			Mean	(UPM + LPM) "Total Intensity"	Reference Gene Factor	Nonlinearity Association Factor	Normalized Data Iterations			Mean
Reference Gene >>>	A	5	4	3	4.0000	1.0000	**1.0000**	5.0000	4.0000	3.0000	4.0000	4.0000	1.0000	**1**	5.0000	4.0000	3.0000	4.0000
	B	2	1	4	2.3333	1.7143	**0.6547**	2.2445	1.1223	4.4891	2.6186	2.3333	1.7143	**1**	3.4286	1.7143	6.8571	4.0000
	A	5	4	3	4.0000	0.5833	**0.6547**	1.9094	1.5275	1.1456	1.5275	4.0000	0.5833	**1**	2.9167	2.3333	1.7500	2.3333
Reference Gene >>>	B	2	1	4	2.3333	1.0000	**1.0000**	2.0000	1.0000	4.0000	2.3333	2.3333	1.0000	**1**	2.0000	1.0000	4.0000	2.3333
Final Values		Original Data						Non-Linear Scaling Normalized Values							Linear Scaling Normalization Values			
	A	5	4	3				3.4547	2.7638	2.0728	2.7638				3.9583	3.1667	2.3750	3.1667
	B	2	1	4				2.1223	1.0611	4.2445	2.4760				2.7143	1.3571	5.4286	3.1667

	Gene	ORIGINAL DATA SET			Our Proposed Method							Linear Scaling Method						
					(UPM + LPM) "Total Intensity"	Reference Gene Factor	Nonlinearity Association Factor	Normalized Data Iterations			Mean	(UPM + LPM) "Total Intensity"	Reference Gene Factor	Nonlinearity Association Factor	Normalized Data Iterations			Mean
	A	5	4	3	4.0000	1.1667	**0.6547**	3.8188	3.0551	2.2913	3.0551	4.0000	1.1667	**1**	5.83333	4.66667	3.50000	4.66667
	B	2	1	4	2.3333	2.0000	**1.0000**	4.0000	2.0000	8.0000	4.6667	2.3333	2.0000	**1**	4.00000	2.00000	8.00000	4.66667
	C	3	4	6	4.3333	1.0769	**0.7857**	2.5385	3.3846	5.0769	3.6667	4.3333	1.0769	**1**	3.23077	4.30769	6.46154	4.66667
Reference Gene >>>	D	4	2	8	4.6667	1.0000	**1.0000**	4.0000	2.0000	8.0000	4.6667	4.6667	1.0000	**1**	4.00000	2.00000	8.00000	4.66667
	A	5	4	3	4.0000	1.0833	**0.9820**	5.3191	4.2552	3.1914	4.2552	4.0000	1.0833	**1**	5.41667	4.33333	3.25000	4.33333
	B	2	1	4	2.3333	1.8571	**0.7857**	2.9184	1.4592	5.8367	3.4048	2.3333	1.8571	**1**	3.71429	1.85714	7.42857	4.33333
Reference Gene >>>	C	3	4	6	4.3333	1.0000	**1.0000**	3.0000	4.0000	6.0000	4.3333	4.3333	1.0000	**1**	3.00000	4.00000	6.00000	4.33333
	D	4	2	8	4.6667	0.9286	**0.7857**	2.9184	1.4592	5.8367	3.4048	4.6667	0.9286	**1**	3.71429	1.85714	7.42857	4.33333
	A	5	4	3	4.0000	0.5833	**0.6547**	1.9094	1.5275	1.1456	1.5275	4.0000	0.5833	**1**	2.91667	2.33333	1.75000	2.33333
Reference Gene >>>	B	2	1	4	2.3333	1.0000	**1.0000**	2.0000	1.0000	4.0000	2.3333	2.3333	1.0000	**1**	2.00000	1.00000	4.00000	2.33333
	C	3	4	6	4.3333	0.5385	**0.7857**	1.2692	1.6923	2.5385	1.8333	4.3333	0.5385	**1**	1.61538	2.15385	3.23077	2.33333
	D	4	2	8	4.6667	0.5000	**1.0000**	2.0000	1.0000	4.0000	2.3333	4.6667	0.5000	**1**	2.00000	1.00000	4.00000	2.33333
Reference Gene >>>	A	5	4	3	4.0000	1.0000	**1.0000**	5.0000	4.0000	3.0000	4.0000	4.0000	1.0000	**1**	5.00000	4.00000	3.00000	4.00000
	B	2	1	4	2.3333	1.7143	**0.6547**	2.2445	1.7143	6.8571	3.6053	2.3333	1.7143	**1**	3.42857	1.71429	6.85714	4.00000
	C	3	4	6	4.3333	0.9231	**0.9820**	2.7193	3.6923	5.5385	3.9834	4.3333	0.9231	**1**	2.76923	3.69231	5.53846	4.00000
	D	4	2	8	4.6667	0.8571	**0.6547**	2.2445	1.7143	6.8571	3.6053	4.6667	0.8571	**1**	3.42857	1.71429	6.85714	4.00000
Final Values		Original Data						Non-Linear Scaling Normalized Values							Linear Scaling Normalization Values			
	A	5	4	3				4.0118	3.2095	2.4071	3.2095				4.79167	3.83333	2.87500	3.83333
	B	2	1	4				2.7907	1.5434	6.1735	3.5025				3.28571	1.64286	6.57143	3.83333
	C	3	4	6				2.3818	3.1923	4.7885	3.4542				2.65385	3.53846	5.30769	3.83333
	D	4	2	8				2.7907	1.5434	6.1735	3.5025				3.28571	1.64286	6.57143	3.83333

Figure 1. Iterative procedure on 2 gene and 4 gene example.

Note the original identical values between genes A & C (non-linear relationship) BLACK CIRCLES.

Linear Scaling: 7.69% difference in normalized values

Our Method: 1.05% difference in normalized values.

Note the original identical values between genes A & C (non-linear relationship) RED CIRCLES.

Linear Scaling: 7.69% difference in normalized values

Our Method: 0.54% difference in normalized values.

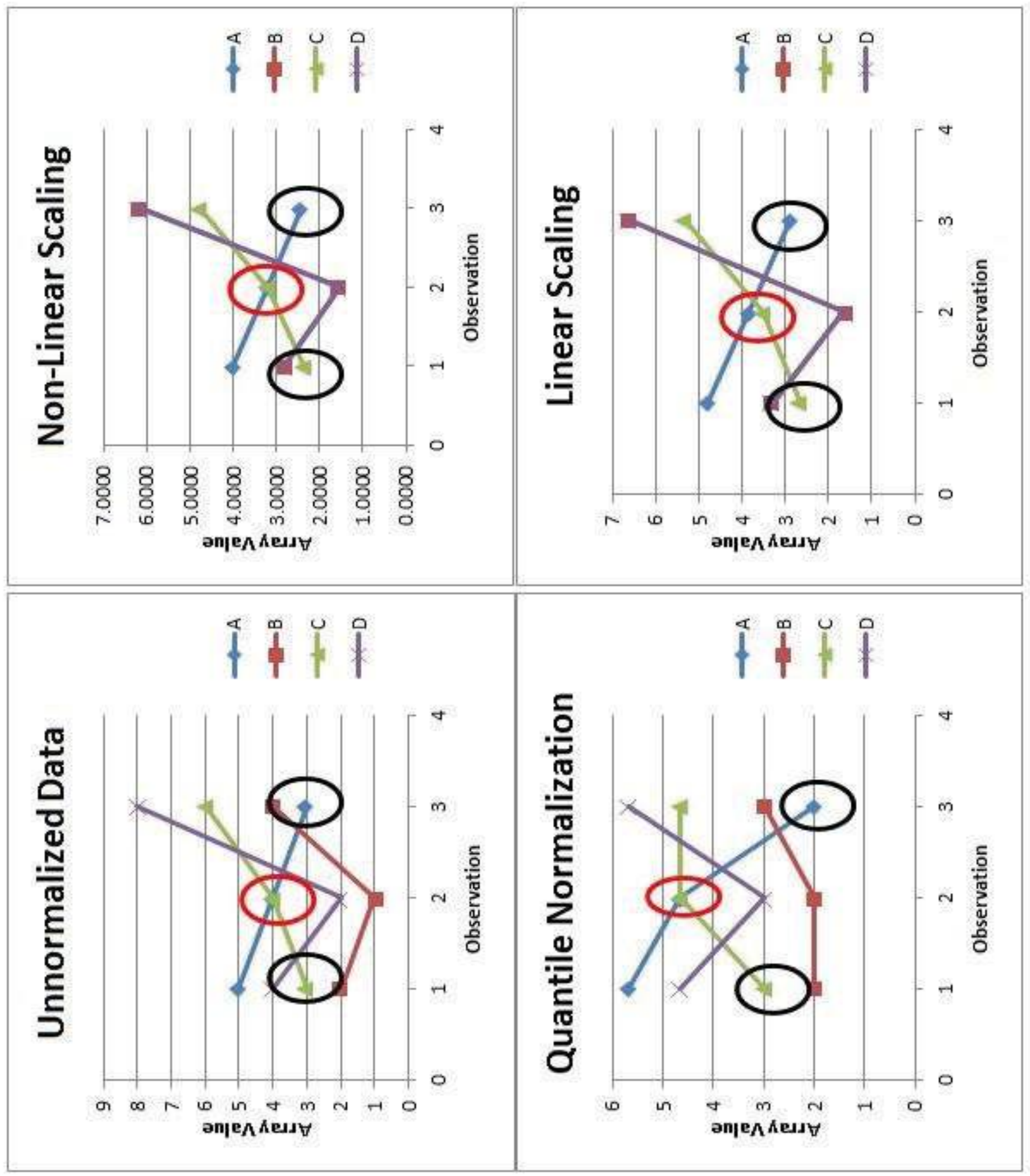


Figure 2. Graphical representation of 2 gene and 4 gene example.

We now present the results of this method on four financial variables SPY, TLT, GLD, and FXE. The nonlinear association between self and cross financial time-series is well noted. This is an important test, since gene distributions are roughly similar, how does this method work on the most stochastic variables?

Figure 3 below illustrates the results. Our method visually represents the original data set more clearly and also retains the finite moment relationships that the linear scaling method enjoys. We note the strong influence the nonlinear association has on the normalized series, as SPY is distinct due to its very low correlation to any of the other time series. ***Thus, the more correlated the series are, the lower the variance of the normalized population.***

The problem with quantile normalization is that if the distributions do not intersect, the quantile ranks remain static and the normalized value is simply the mean. This is exemplified below with the financial variables. Obviously this is not an issue with gene arrays, however, it speaks to the ad hoc nature of the method. We see in Figure 3 below quantile normalization does succeed in creating the same distribution for all, however, they are all uniform distributions.

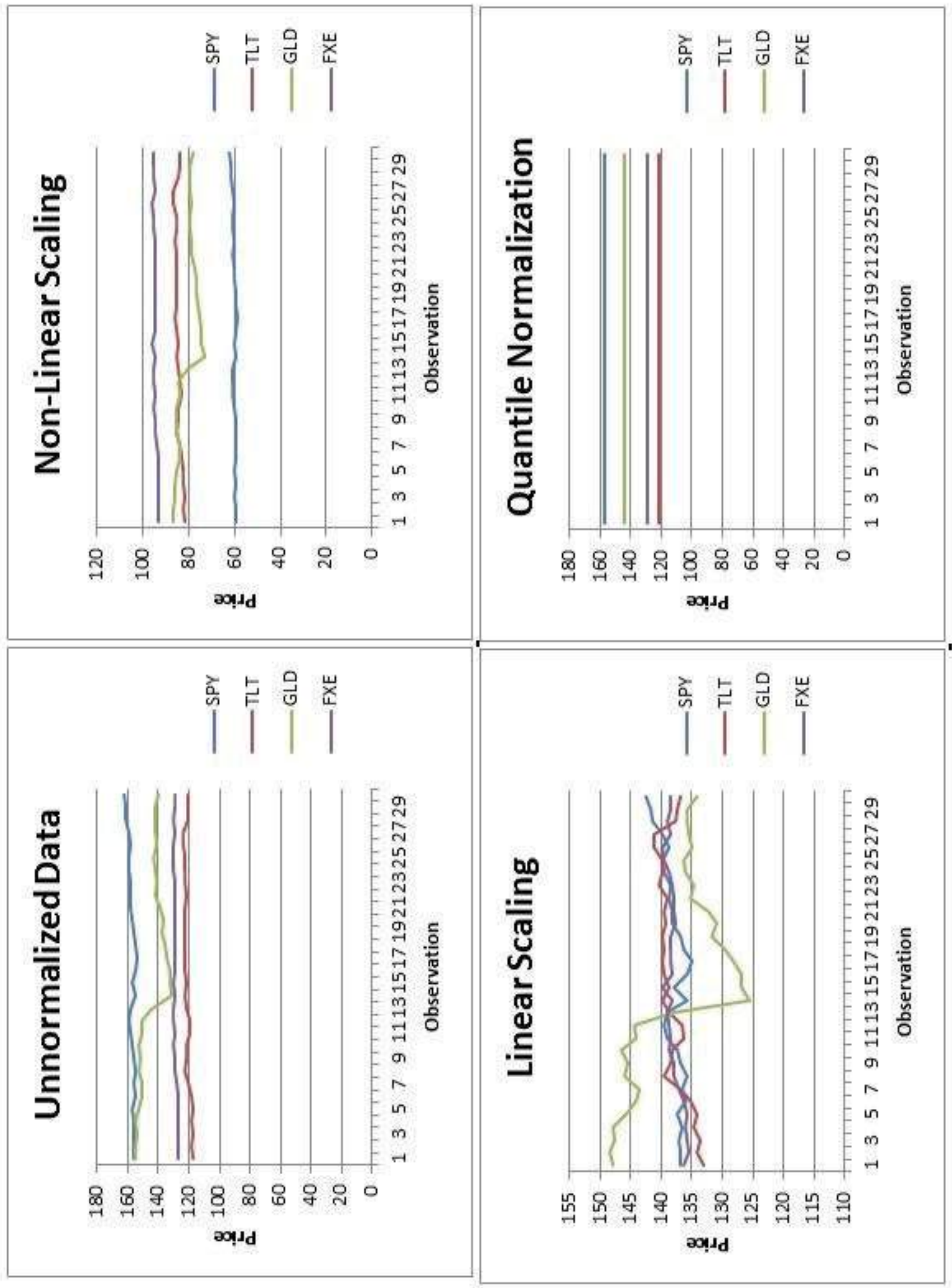


Figure 3. Graphical representation of 4 financial variable example.

ORDERS OF MAGNITUDE DIFFERENCES REMOVED

The method also successfully removes orders of magnitude differences between variables. Below in figure 4 is an example illustrating the results on MZM ($ billions scale), S&P 500 (point scale) and the US 10 Year Yield (% scale).

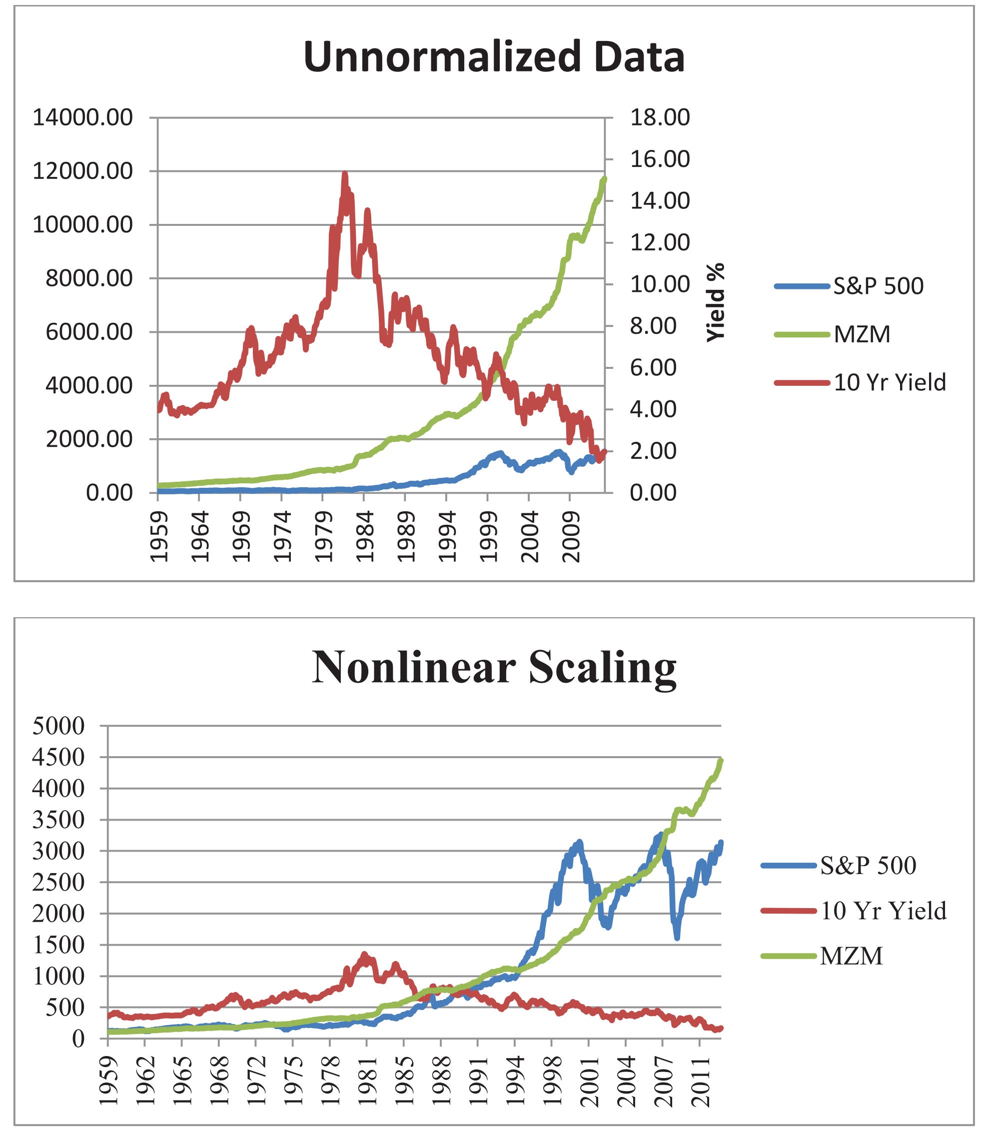


Figure 4. Orders of magnitude differences removed from 3 financial variables.

DISCUSSION

Note the tighter overall distribution from our method versus the linear scaling method. Also note the variance properties of the each of the distributions versus the quantile normalization. We are tighter and more representative of the original data set for similar distributions. When the distributions vary considerably, the nonlinear association will be reflected in the variance of the normalized series.

We also have retained mean differences between the distributions for nonlinear variables. This characteristic is lost via its use *as the normalizing factor* in the linear scaling technique. Factoring the nonlinear association between variables is imperative in noting the nonlinear differences. **Moreover, if the variable relationship is linear, our method retains the relationship between variables!**

Bolstad et al. [2003] note,

> "The four baselines shifted slightly lower in the intensity scale give the most precise estimates. Using this logic, one could argue that choosing the array with the smallest spread and centered at the lowest level would be the best, but this does not seem to treat the data on all arrays fairly."

Our method does treat all of the data on all of the arrays fairly. We use each array as a RG and utilize its nonlinear association (which uses all observations equally) with all other arrays equally.

ANOVA Using Continuous Cumulative Distribution Functions

Abstract

Analysis of Variance (ANOVA) is a statistical method used to determine whether a sample originated from a larger population distribution. We provide an alternate method of determination using the continuous cumulative distribution functions derived from degree one lower partial moment ratios. The resulting analysis is performed with no restrictive assumptions on the underlying distribution or the associated error terms.

INTRODUCTION

Analysis of Variance (ANOVA) is a statistical method used to determine whether a sample originated from a larger population distribution. This is accomplished by using a statistical test for heterogeneity of means by analysis of group variances. By defining the sum of squares for the total, treatment, and errors, we then obtain the P-value corresponding to the computed F-ratio of the mean squared values. If the P-value is small (large F-ratio), we can reject the null hypothesis that all means are the same for the different samples. However, the distributions of the residuals are assumed to be normal and this normality assumption is critical for P-values computed from the F-distribution to be meaningful. Instead of using the ratio of variability between means to the variability within each sample, we suggest an alternative approach.

Using known distributional facts from samples, we can deduce a level of certainty that multiple samples originated from the same population without any of the assumptions listed below.

ANOVA ASSUMPTIONS

When using one-way analysis of variance, the process of looking up the resulting value of F in an F-distribution table, is proven to be reliable under the following assumptions:

- the values in each of the groups (as a whole) follow the normal curve,
- with possibly different population averages (though the null hypothesis is that all of the group averages are equal) and
- equal population standard deviations (SD).

The assumption that the groups follow the normal curve is the usual one made in most significance tests, though here it is somewhat stronger in that it is applied to several groups at once. Of course many distributions do not follow the normal curve, so here is one reason that ANOVA may give incorrect results. It would be wise to consider whether it is reasonable to believe that the groups' distributions follow the normal curve.

Of course the different population averages imposes no restriction on the use of ANOVA; the null hypothesis, as usual, allows us to do the computations that yield F.

The third assumption, that the populations' standard deviations are equal, is important in principle, and it can only be approximately checked by using as bootstrap estimates the sample standard deviations. In practice, statisticians feel safe in using ANOVA if the largest sample SD is not larger than twice the smallest.[15]

[15] http://math.colgate.edu/math102/dlantz/examples/ANOVA/anovahyp.html

KNOWN DISTRIBUTIONAL FACTS FROM SAMPLES

Viole and Nawrocki [2012a] offer a detailed examination of CDFs and PDFs of various families of distributions represented by partial moments. They find that the continuous degree 1 LPM ratio is .5 from the mean of the sample. No deviations, for every distribution type, regardless of number of observations, period. Thus when a sample mean is compared to the population, the further the population continuous degree 1 LPM ratio from the sample mean target is from 0.5, the less confident we are that sample belongs to that population.

$$LPM_{ratio}(1,h,X) = \frac{LPM(1,h,X)}{[LPM(1,h,X) + UPM(1,h,X)]} \tag{1}$$

Where,

$$LPM(n,h,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, h - x_t\}^n\right] \tag{2}$$

$$UPM(q,l,x) = \frac{1}{T}\left[\sum_{t=1}^{T} max\{0, x_t - l\}^q\right] \tag{3}$$

where x_t represents the observation x at time t, n is the degree of the LPM, q is the degree of the UPM, h is the target for computing below target returns, and l is the target for computing above target returns. $h = l = \mu$ throughout this paper.

Tables 1 through 4 illustrate the consistency of the degree 1 LPM ratio across distribution types.

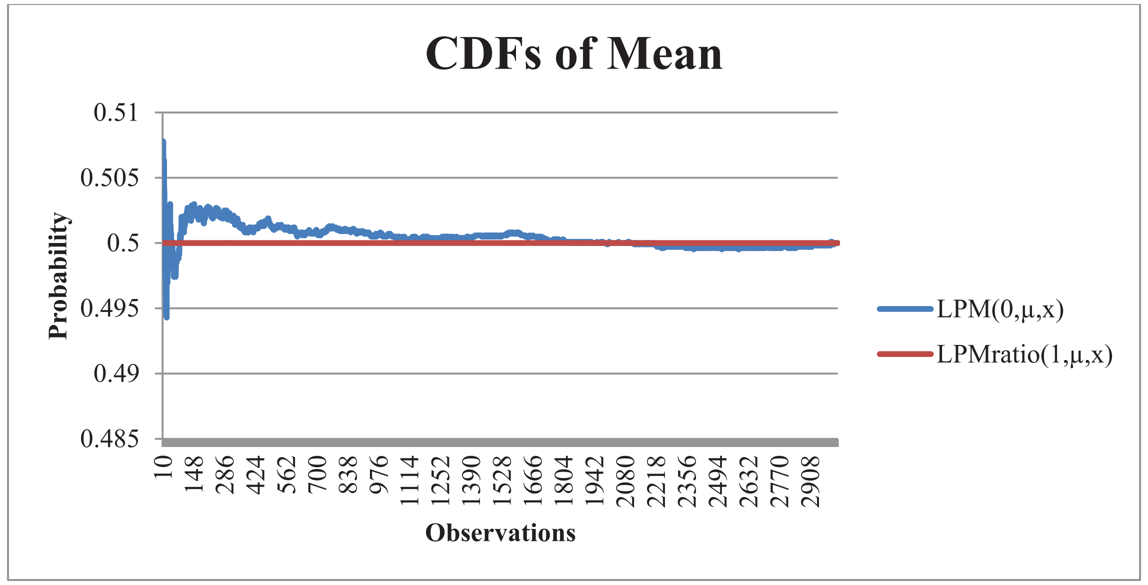


Figure 1. Differences in discrete $LPM(0,\mu,X)$ and continuous $LPM_{ratio}(1,\mu,X)$.CDFs converge when using the mean target for a Normal distribution. $LPM(0,\mu,X) \neq LPM_{ratio}(1,\mu,X)$.

Normal Distribution Probabilities - 5 Million Draws 300 Iteration Seeds

Norm Prob(X ≤ 0.00) = .3085	LPM(0, 0, X) = .3085	LPM(1, 0, X) = .2208
Norm Prob(X ≤ 4.50) = .3917	LPM(0, 4.5, X) = .3917	LPM(1, 4.5, X) = .3339
Norm Prob(X ≤ Mean) = .5	LPM(0, μ, X) = .5	**LPM(1, μ, X) = .5**
Norm Prob(X ≤ 13.5) = .5694	LPM(0, 13.5, X) = .5694	LPM(1, 13.5, X) = .608

Table 1. Final probability estimates with 5 million observations and 300 iteration seeds averaged for the Normal distribution. Bold estimate is the continuous $LPM_{ratio}(1,\mu,X) = 0.5$.

Uniform Distribution Probabilities - 5 Million Draws 300 Iteration Seeds

UNDF(X ≤ 0.00) = .4	LPM(0, 0, X) = .4	LPM(1, 0, X) = .3077
UNDF(X ≤ 4.50) = .445	LPM(0, 4.5, X) = .445	LPM(1, 4.5, X) = .3913
UNDF(X ≤ Mean) = .5	LPM(0, μ, X) = .5	**LPM(1, μ, X) = .5**
UNDF(X ≤ 13.5) = .535	LPM(0, 13.5, X) = .535	LPM(1, 13.5, X) = .5697

Table 2. Final probability estimates with 5 million observations and 300 iteration seeds averaged for the Uniform distribution. Bold estimate is the continuous $LPM_{ratio}(1,\mu,X) = 0.5$.

Poisson Distribution Probabilities - 5 Million Draws 300 Iteration Seeds

POIDF(X ≤ 0.00) = .00005	LPM(0, 0, X) = 0	LPM(1, 0, X) = 0
POIDF(X ≤ 4.50) = .0293	LPM(0, 4.5, X) = .0293	LPM(1, 4.5, X) = .0051
POIDF(X ≤ Mean) = .5151	LPM(0, μ, X) = .5151	**LPM(1, μ, X) = .5**
POIDF(X ≤ 13.5) = .8645	LPM(0, 13.5, X) = .8645	LPM(1, 13.5, X) = .9365

Table 3. Final probability estimates with 5 million observations and 300 iteration seeds averaged for the Poisson distribution. Bold estimate is the continuous $LPM_{ratio}(1,\mu,X) = 0.5$.

Chi-Squared Distribution Probabilities - 5 Million Draws 300 Iteration Seeds

CHIDF(X ≤ 0) = 0	LPM(0, 0, X) = 0	LPM(1, 0, X) = 0
CHIDF(X ≤ 0.5) = .5205	LPM(0, 0.5, X) = .5205	LPM(1, 0.5, X) = .2087
CHIDF(X ≤ 1) = .6827	LPM(0, 1, X) = .6827	**LPM(1, 1, X) = .5**
CHIDF(X ≤ 5) = .9747	LPM(0, 5, X) = .9747	LPM(1, 5, X) = .989

Table 4. Final probability estimates with 5 million observations and 300 iteration seeds averaged for the Chi-Squared distribution. Bold estimate is the continuous $LPM_{ratio}(1,\mu,X) = 0.5$.

METHODOLOGY

We propose using the mean absolute deviation from 0.5 for the samples in question. This result compared to the ideal 0.5 will then answer the ANOVA inquiry whether the samples originated from the same population.

First we need the average of all of the sample means, $\bar{\boldsymbol{\mu}}$. Then we can compute each sample's absolute deviation from the mean of means.

$$D_i = |LPM_{ratio}(1, \bar{\mu}, X) - LPM_{ratio}(1, \mu, X)| \tag{4}$$

Which reduces to,

$$D_i = |LPM_{ratio}(1, \bar{\mu}, X) - 0.5|$$

The mean absolute deviation for *n* samples is then

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|LPM_{ratio}(1, \bar{\mu}, X_i) - 0.5| \tag{5}$$

Yielding our measure of certainty ρ associated with the null hypothesis that the samples in question belong to the same population

$$\rho = \frac{(0.5 - MAD)^2}{0.5} \tag{6}$$

The next section will provide some visual confirmation of this methodology with confirming classic ANOVA analysis.

EXAMPLES OF OUR METHODOLOGY

Figure 1 below illustrates 3 hypothetical sample distributions. The dotted lines are the sample means $\boldsymbol{\mu}$, ***which we know have an associated*** $\boldsymbol{LPM_{ratio}(1, \mu, X) = 0.5}$. The solid black line is the mean of means $\bar{\boldsymbol{\mu}} = \mathbf{19.81}$, and associated LPM ratio deviations from 0.5 can be visually estimated.

Source	Sum of Squares	*df*	Mean Square	*F* Value	*P* Value
Between Groups	0.566	2	0.283	0.224	> .05
Within Groups	34.108	27	1.263		
Total	34.674	29			

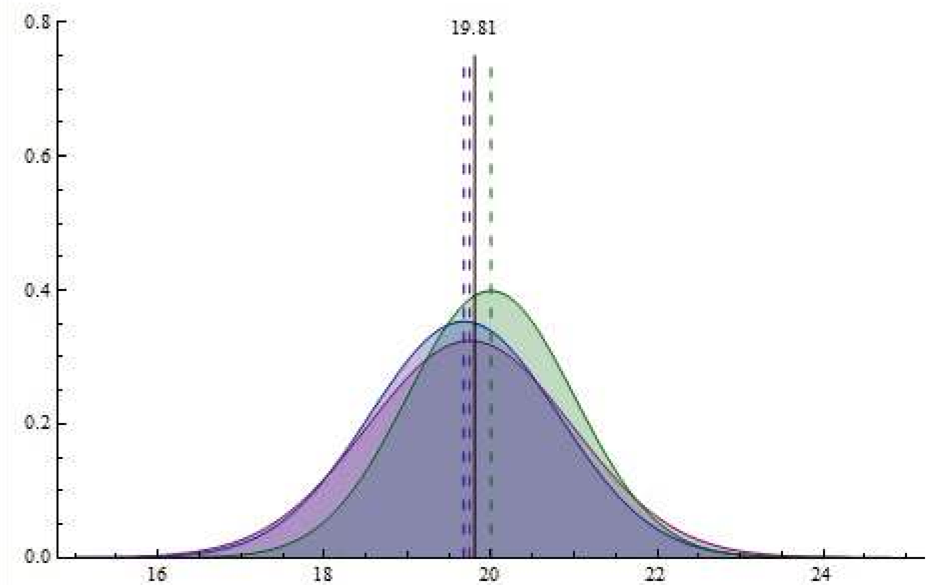


Figure 1. 3 samples from the same population.

We can see visually that the $\boldsymbol{LPM_{ratio}(1, \bar{\mu}, X)}$ for these 3 samples is approximately 0.52, 0.51, and 0.48 for blue, purple and green respectively. The mean absolute deviation from .5 is equal to .0167. Thus we are certain $(\rho = 0.934)$ these 3 samples are from the same population.

According to the F-Values and associated degrees of freedom,

$$\boldsymbol{F_{.05}(2, 27\ df) = 3.3541}$$

$$\boldsymbol{F_{.01}(2, 27\ df) = 5.4881}$$

The classic ANOVA would reach the same conclusion even at $P\ value < .01$.

Figure 2 below illustrates 3 hypothetical sample distributions, only more varied than the previous example. The dotted lines are the sample means $\boldsymbol{\mu}$, ***which we know have an associated*** $\boldsymbol{LPM_{ratio}(1, \mu, X) = 0.5}$. The solid black line is the mean of means $\boldsymbol{\bar{\mu} = 20.48}$, and associated LPM ratio deviations from 0.5 can be visually estimated.

Source	Sum of Squares	*df*	Mean Square	*F* Value	*P* Value
Between Groups	34.833	2	17.416	13.787	< .01
Within Groups	34.108	27	1.263		
Total	68.941	29			

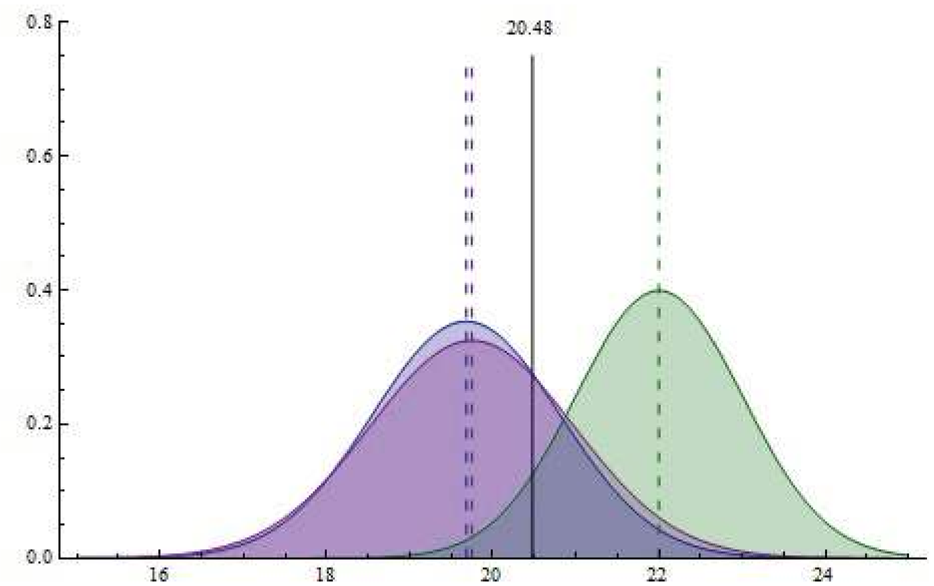


Figure 2. 3 samples not from the same population.

We can see visually that the $\boldsymbol{LPM_{ratio}(1, \bar{\mu}, X)}$ for these 3 samples is approximately 0.65, 0.63, and 0.2 for blue, purple and green respectively. The mean absolute deviation from .5 is equal to .1933. Thus we are not certain $(\rho = 0.376)$ these 3 samples are from the same population. The null hypothesis of a same population was rejected by classic ANOVA at $P\ value < .01$.

Figure 3 below illustrates 3 hypothetical sample distributions, only more varied than the previous example. The dotted lines are the sample means, ***which we know have an associated*** $\boldsymbol{LPM_{ratio}(1, \mu, X) = 0.5}$. The solid black line is the mean of means $\boldsymbol{\bar{\mu} = 20.48}$, and associated LPM ratio deviations from 0.5 can be visually estimated.

Source	Sum of Squares	*df*	Mean Square	*F* Value	*P* Value
Between Groups	34.833	2	17.416	17.188	< .01
Within Groups	27.358	27	1.013		
Total	62.191	29			

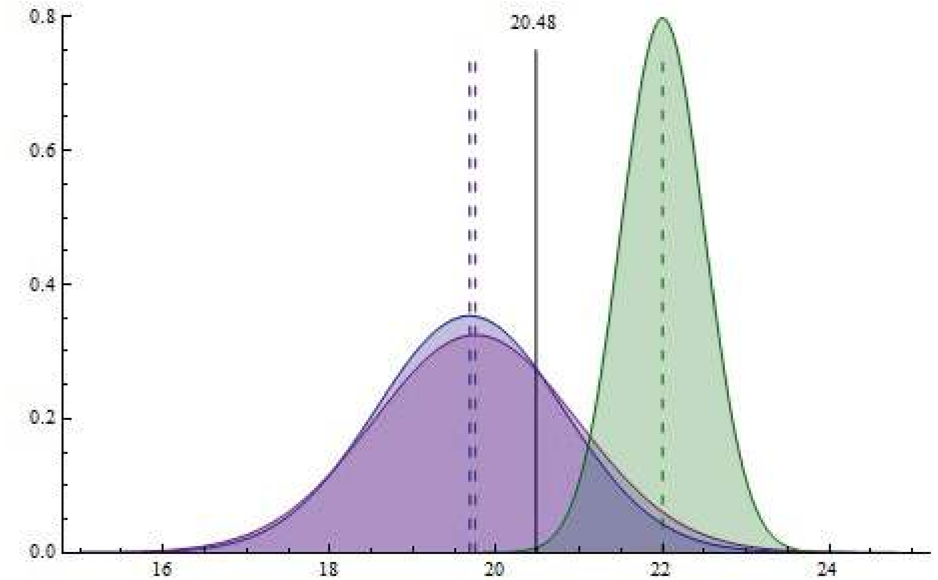


Figure 3. 3 samples not from the same population.

We can see visually that the $\boldsymbol{LPM_{ratio}(1, \bar{\mu}, X)}$for these 3 samples is approximately 0.65, 0.63, and 0.01 for blue, purple and green respectively. The mean absolute deviation from .5 is equal to .2567. Thus we are more certain $(\rho = 0.237)$ than the previous example that these 3 samples are NOT from the same population.

SIZE OF EFFECT

In the previous sections, we identified whether a difference exists and demonstrated how to assign a measure of uncertainty to our data. We focus now on how to ascertain the size of the difference present. The use of confidence intervals is often suggested as a method to evaluate effect sizes. Our methodology assigns the interval to the effect without the standardization or parameterization required for traditional confidence intervals.

The first step is to derive a sample mean for which we would be 95% certain the sample mean belongs to the population. We calculate the lower 2.5% of the distribution with a LPM test at each point to identify the inverse, akin to a value-at-risk derivation. We perform the same on the upper portion of the distribution with a UPM test. This two sided test results in a negative deviation from the population mean $(\boldsymbol{\mu}^{*-})$ and a corresponding positive deviation from the mean $(\boldsymbol{\mu}^{*+})$. It is critical to note that this is not necessarily a symmetrical deviation, since any underlying skew will alter the CDF derivations for these autonomous points.

The effect size then is simply, the difference between the observed mean $(\boldsymbol{\mu})$ and a certain mean associated within a tolerance either side of the population mean $(\boldsymbol{\mu}^{*-}\ \boldsymbol{and}\ \boldsymbol{\mu}^{*+})$.

$$(\boldsymbol{\mu}-\boldsymbol{\mu}^{*-}) \leq \boldsymbol{effect} \leq (\boldsymbol{\mu}-\boldsymbol{\mu}^{*+}).$$

DISCUSSION

Viole and Nawrocki [2012c] define the asymptotic properties of partial moments to the area of any *f*(x). Thus, it makes intuitive sense that increased quantities of samples and observations will provide a better approximation of the population. Given this truism, the degrees of freedom do not properly compensate the number of observations.

We can see below that increasing the number of distributions from two to three and increasing the number of observations from 30 to 100 does not have an order of magnitude effect on the F-Values.

2 distributions and 3 distributions with 30 observations each:

$\boldsymbol{F_{.05}(1,59\ df) = 4.004}$ $\boldsymbol{F_{.05}(2,88\ df) = 3.1001}$

2 distributions and 3 distributions with 100 observations each:

$\boldsymbol{F_{.05}(1,199\ df) = 3.8886}$ $\boldsymbol{F_{.05}(2,298\ df) = 3.0261}$

The t-test concerns are simply nonexistent under this methodology, thus multiple 2 distribution tests can be performed. For example, if 15 samples are all drawn from the same population, then there are 105 possible comparisons to be made leading to an increased type-1 error rate. The mean absolute deviation for 2 distributions' $\boldsymbol{LPM_{ratio}(1,\bar{\mu},X)}$ would have to be > 0.025 to be less than 95% certain (0.475/.5) the distributions came from the same population. This translates to a substantial percentage difference in means. It is not hard to visualize such an extreme scenario such as Figure 4 below.

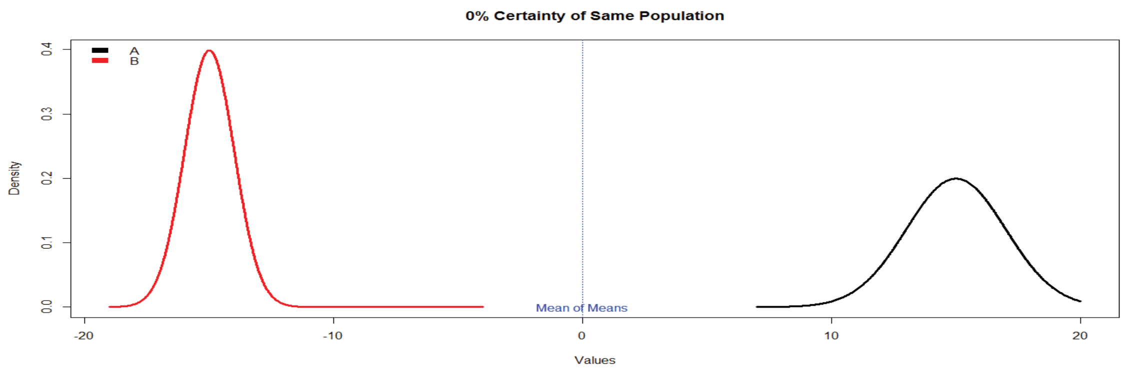


Figure 4. 2 samples not from the same population.

Given this scenario whereby $\boldsymbol{LPM_{ratio}(1, \bar{\mu}, A) = 1.0}$ and $\boldsymbol{LPM_{ratio}(1, \bar{\mu}, B) = 0}$, the mean absolute deviation from $\boldsymbol{\bar{\mu} = 0.5}$ thus $\rho = 0$. Therefore, we are certain these distributions came from different populations.

Again, we have no assumptions on the data to generate this analysis and compensate for any deviation from normality either in the distribution of returns or the distribution of error terms. We substitute our level of certainty ρ for an F-test and associated P-value based ANOVA; the latter has been the subject of increasing debate recently and should probably be avoided.[16]

[16] http://news.sciencemag.org/sciencenow/2009/10/30-01.html?etoc
http://www.sciencenews.org/view/feature/id/57091/description/Odds_Are_Its_Wrong

CORRELATION

≠

CAUSATION

Causation

Abstract

We identify the necessary conditions to define causation between two variables. We compare this to Granger causality and the convergent cross mapping method to illustrate the theoretical differences. Our proposed method avoids the reciprocal Granger and nonlinearity concerns. We loosely share a procedural step with the convergent cross mapping method in so much that our lagged variable time-series are normalized. The resulting normalized variables permit relevant conditional probability and correlation statistics to be generated and used to determine causation.

INTRODUCTION

Correlation does not imply causation. We have known this to be the case for decades, however, the often misapplication of correlation to causation speaks volumes to the suspicion that correlation and causation are entwined…but how? Fischer Black [1984] offers multiple normative cases explaining how causality can only be demonstrated with experimentation. Black's argument is indirectly identifying the conditional probability associated with a causal relationship and is explicit in our proposed measure of causality.

$$\boldsymbol{C}(\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \boldsymbol{P}(\boldsymbol{X}|\boldsymbol{Y}) * \boldsymbol{\rho}_{\boldsymbol{X},\boldsymbol{Y}} \tag{1}$$

CAUSATION(*X* → *Y*) = CONDITIONAL PROBABILITY(*X*|*Y*) * CORRELATION(*X*,*Y*)

Conditional Probability: The probability that an event will occur, given that one or more other events have occurred.

Correlation: A mutual relationship or connection between two or more things.

Correlation is a reciprocal relationship between two things. Conditional probability is not necessarily a reciprocal relationship between two things. This distinction is critical in factoring correlation to define the correlation/causation link.

HISTORICAL CAUSALITY TESTS

GRANGER CAUSALITY

Granger causality (GC) measures whether one event (*X*) happens before another event (*Y*) and helps predict it. According to Granger causality, past values of *X* should contain information that helps predict *Y* better than a prediction based on past values of *Y* alone. The formulation is based on a linear regression modeling of stochastic processes.

This technique immediately raises some well documented concerns, namely, linearity, stationarity and of course the appropriate selection of variables. Any proposed substitute should be able to address these basic data set concerns.

CONVERGENT CROSS MAPPING

Sugihara et al. [2012] examine an approach specifically aimed at identifying causation in ecological time series called convergent cross mapping (CCM). They demonstrate the principles of their approach with simple model examples, showing that the method distinguishes species interactions (*X*, *Y*) from the effects of shared driving

variables (*Z*). Attractor reconstruction is used to determine if two time series variables belong to the same dynamic system and are thus causally related.

Points on manifolds *X* and *Y* will only be nearest neighbors if *X* and *Y* are causally related. CCM uses the historical record of *Y* to estimate the states of *X* and vice versa. With longer time series the reconstructed manifolds are denser, nearest neighbors are closer, and the cross map estimates increase in precision. This convergence is used as a practical criterion for determining causation, further exposed by measuring the extent to which the historical record of *Y* values can reliably estimate states of *X*. CCM hypothesizes that this reliable estimate holds only if *X* is causally influencing *Y*.

> "In dynamical systems theory, time-series variables (say, *X* and *Y*) are causally linked if they are from the same dynamic system (Dixon et al. [1999], Takens [1981], Deyle et al. [2011])—that is, they share a common attractor manifold *M*." Sugihara et al. [2012]

Figure 1 is a reproduction from their paper illustrating the manifold relationship.

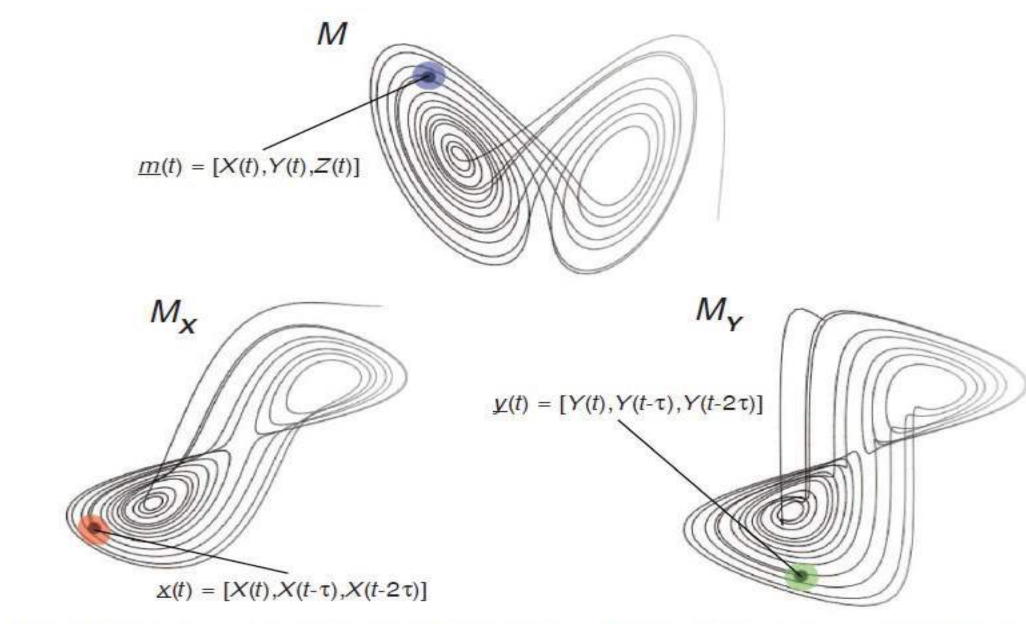


Figure 1. Manifold relationship from Sugihara et al. [2012].

Separability Requirement

Sugihara et al. note the key requirement of GC is *separability*, namely that information about a causative factor is independently unique to that variable. Conditional probability is also independently unique to that variable. Separability is characteristic of purely stochastic and linear systems, and GC can be useful for detecting interactions between strongly coupled (synchronized) variables in nonlinear systems. Conditional

probabilities are not restricted to these specific characteristics. Separability reflects the view that systems can be understood a piece at a time rather than as a whole. By normalizing the variables, we retain the whole system view perspective.

Our proposed measure avoids the GC problems of nonlinearity by normalizing the variables with a nonlinear scaling method. It also avoids the Granger problems of reverse causality since ***the Venn areas (conditional probabilities) would have to be identical in size and shape and location to permit reverse causality.***

OUR PROPOSED METHOD

The first step in our method is to normalize the variables in order to determine the conditional probability between the two variables in question. In an experiment setting, conditional probability is controlled quite easily; in fact, this is the main argument of Black [1984]. To determine the conditional probability, we need a shared histogram for variables X and Y. This is not all dissimilar to the approach in the convergent cross map technique, with the common attractor manifold for the original system M used to describe M_X and M_Y.

1) *Normalize the variables*. Viole and Nawrocki [2013] (VN) present a method for normalizing variables with a nonlinear scaling method that reflects the inherent nonlinear association between the variables within the scaling factor. The

normalized variables retain their variance and other finite moment characteristics. This is important to accurately derive the conditional probability of the new normalized variables. This is also critical in addressing the nonlinearity between variables where GC fails.

The CCM manifolds M_X and M_Y are constructed from lagged coordinates of the time series variables to retain past information. We accomplish the retention of lagged information via the normalization of each variable against lagged values of itself (τ and 2τ), resulting in normalized variables X' and Y'.

We then normalize X' and Y' to each other via the VN process of nonlinear scaling to generate the shared histogram resulting in X'' and Y''.

2) *Derive the correlation between normalized variables.* VN [2012] offer a method of deriving nonlinear correlation coefficients from partial moments that **fully replicate Pearson's correlation coefficient in linear variable relationships**. This is an important advantage at our disposal, and one Granger did not have access to at the time of his work. Given the lack of linear relationships between variables, any linear consideration will prove ineffectual. Furthermore, the normalization procedure in step 1 *significantly* reduces the nonlinearity between variables, allowing for a visual confirmation of the nonlinear correlation coefficients.

3) *Derive the conditional probabilities*. Using the partial moments of each of the resulting distributions will allow us to derive the conditional probabilities of the normalized variables.

$$LPM(n,h,X) = \frac{1}{T}\sum_{t=1}^{T}\{\max(h - X_t)\,,0\}^n \tag{1}$$

$$UPM(q,l,X) = \frac{1}{T}\sum_{t=1}^{T}\{\max(X_t - l)\,,0\}^q \tag{2}$$

Where X_t is the observation of variable X at time t, h and l are the targets from which to compute the lower and upper deviations respectively, and n and q are the weights to the lower and upper deviations respectively.

The next section will discuss deriving conditional probabilities from partial moments of the normalized distributions of X'' and Y''. **Partial moments are asymptotic approximations of the area of an interval** (in this instance as shown later, the entire distribution) **for any $f(x)$**. This nonparametric flexibility captures the nonstationarity associated with variables, which often spoils attempts at estimating true population parameters. Convergence, the first "C" in CCM, is demonstrated as the number of observations increases. Our method also benefits from increased observations as partial moments gain stability as the number of observations increases.

CONDITIONAL PROBABILITIES

We illustrate how the partial moment ratios can also emulate conditional probability calculations. We re-visualize the Venn diagram areas in Figure 2 as distribution areas from which the *LPM* and *UPM* can be observed.

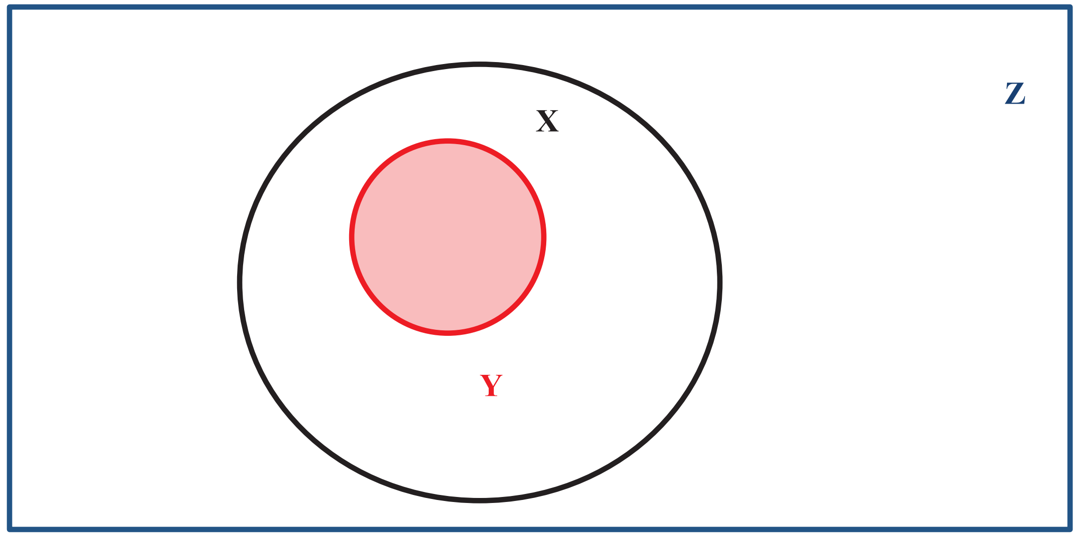
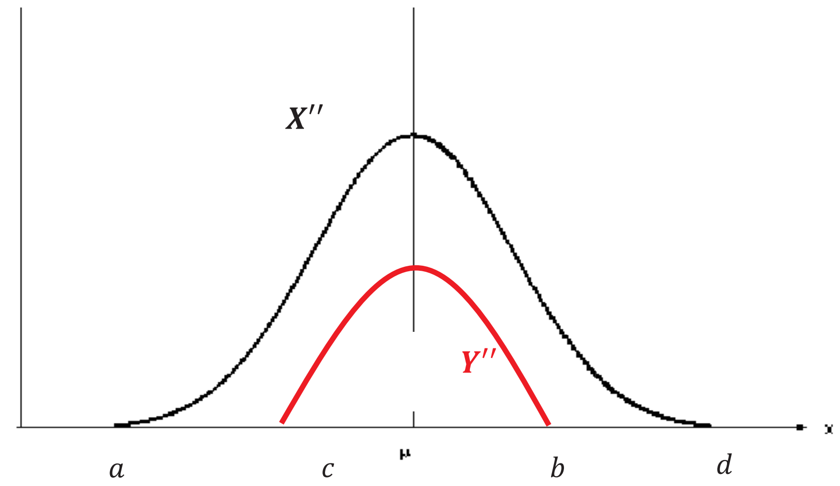


Figure 2. Venn diagram illustrating conditional probabilities X,Y in sample space Z. $P(X|Y) = 1$.

The conditional probability $\boldsymbol{P(X|Y) = 1}$ reconstructed as normalized distributions. The following degree 0 partial moment relationships will yield the conditional probability of $\boldsymbol{X''}$ given $\boldsymbol{Y''}$.

Figure 3. Normalized Data Sets $\boldsymbol{P(X''|Y'') = 1}$



$$\boldsymbol{P(X''|Y'')} = 1 - LPM(0, a, Y'') - UPM(0, b, Y'') \quad (3)$$

$$\boldsymbol{P(X''|Y'')} = UPM(0, a, Y'') - UPM(0, b, Y'')$$

$$\boldsymbol{P(X''|Y'')} = (1) - (0)$$

If $\boldsymbol{X}$ is chewing tobacco and $\boldsymbol{Y}$ is rare tongue cancer, does $\boldsymbol{X}$ cause $\boldsymbol{Y}$? Axiomatically, there exists a conditional probability between the two variables. However, we know nothing about the relationship between them, in fact, if the correlation is negative we could state that $\boldsymbol{X}$ cures $\boldsymbol{Y}$! We assume (know) this to not be the case, but it illustrates the necessity to define the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ further than just their conditional probability.

Per figure 3 above, given the conditional probability $\boldsymbol{P(X|Y)=1}$, and if a positive correlation exists such that measured increases (decreases) in $\boldsymbol{X}$ result in measured increases (decreases) in $\boldsymbol{Y}$ (correlation $\boldsymbol{\rho_{X,Y}=1}$), we can state definitively that $\boldsymbol{X}$ causes $\boldsymbol{Y}$.

$$\boldsymbol{C(X\rightarrow Y)=P(X|Y)*\rho_{X,Y}}$$

$$\boldsymbol{C(X\rightarrow Y)=1*1}$$

$$\boldsymbol{C(X\rightarrow Y)=1}$$

The reciprocal case does not necessarily hold as we can see from the figure above. Since $\boldsymbol{X}$ can occur without the occurrence of $\boldsymbol{Y}$, $\boldsymbol{P(Y|X)<1}$, thus reducing $\boldsymbol{C(Y\rightarrow X)}$ regardless of correlation since $\boldsymbol{\rho_{X,Y}=\rho_{Y,X}}$. **In order for reciprocity of causality to occur, $\boldsymbol{P(X|Y)=P(Y|X)}$.**

ADDITIVITY OF CAUSATION

$\boldsymbol{C(X\rightarrow Y)}$ is also additive such that

$$\sum_{i=1}^{n}\boldsymbol{C(X_{1\ldots n}\rightarrow Y)=1} \tag{4}$$

Below is a figure whereby$\boldsymbol{P(X|Y)<1}$. This is an important realization and primarily the problem with finance and Bayes' application to finance and economics. Identifying the

independent variables to satisfy equation 4 is nearly impossible in the social sciences and is a prominent argument in Black [1984].[17]

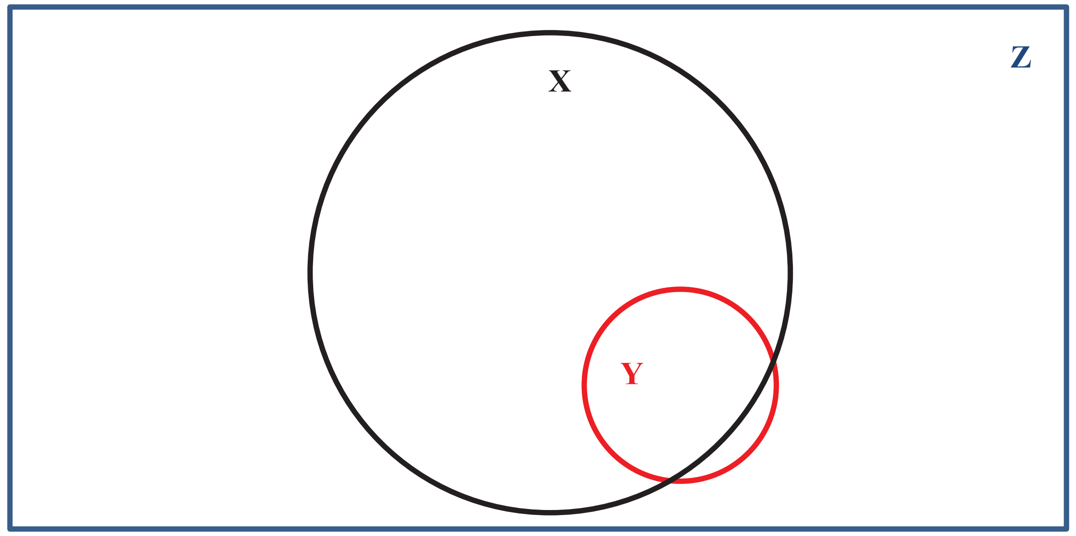
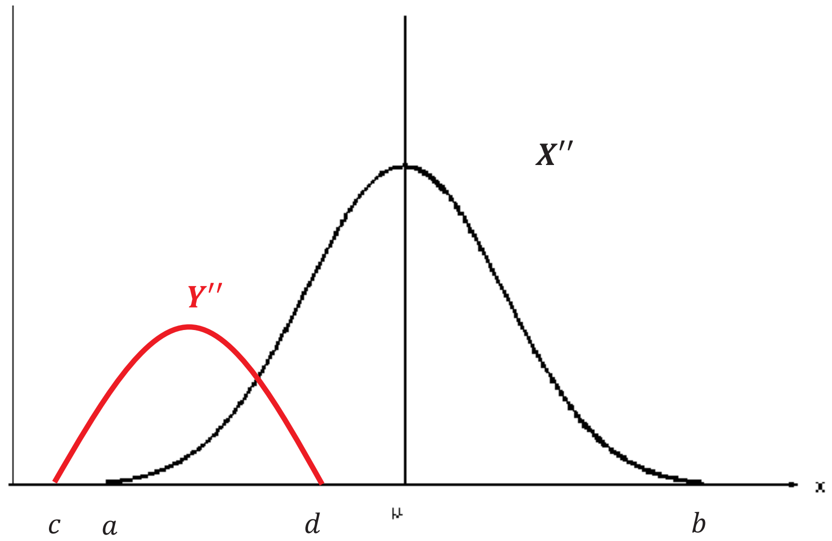


Figure 4. Venn diagram illustrating conditional probabilities $\boldsymbol{X, Y}$ in sample space $\boldsymbol{Z}$. $\boldsymbol{P(X|Y)\sim 0.85}$.

The conditional probability $\boldsymbol{P(X|Y)\sim 0.85}$ reconstructed as normalized distributions.

[17] We do not rule out the possibility of multiple causes. However, multiple highly causative independent variables would then by necessity be exceptionally correlated with conditional probability overlays. This observation satisfies the conditions of the omitted variable bias whereby the omitted variable: 1) must be a determinant of the dependent variable; and 2) must be correlated with one or more of the included independent variables.

Figure 5. Normalized Data Sets $P(X''|Y'') \sim 0.85$



$$P(X''|Y'') = 1 - LPM(0, a, Y'') - UPM(0, b, Y'')$$

$$P(X''|Y'') = UPM(0, a, Y'') - UPM(0, b, Y'')$$

$$P(X''|Y'') = (.85) - (0)$$

If the correlation between variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is the same as our theortetical assumption from the prior example $(\boldsymbol{\rho_{X,Y} = 1})$, then

$$\boldsymbol{C(X \to Y) = 0.85 * 1}$$

$$\boldsymbol{C(X \to Y) = 0.85}$$

Then by the additive assumption, there exist other variable(s) to explain the causation of $\boldsymbol{Y}$ for the remaining 0.15 while factoring their specific correlations as well. It should be noted that it is irrelevant which side of the distribution $\boldsymbol{Y}$ overlaps $\boldsymbol{X}$.

BAYES' THEOREM

Bayes' theorem will also generate the conditional probability of X given Y, $P(X|Y)$ with the formula

$$P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}.$$

Where the probability of X is represented by,

$$P(X) = \frac{Area\ of\ X}{Area\ of\ total\ sample\ Z} = UPM(0, a, X)$$

And the probability of Y is represented by,

$$P(Y) = \frac{Area\ of\ Y}{Area\ of\ total\ sample\ Z} = UPM(0, c, Y)$$

Where e is the minimum value target of area (distribution) Z; just as a and c are for areas (distributions) X and Y respectively (d and b are maximum respective value targets). Thus, if the conditional probability of Y given X is (per equation 3),

$$P(Y|X) = \frac{CUPM(0|0, c|a, Y|\ X)}{UPM(0, a, X)}$$

Then,

$$P(X|Y) = \frac{\frac{CUPM(0|0, c|a, Y|\ X)}{UPM(0, a, X)} UPM(0, a, X)}{UPM(0, c, Y)}$$

Cancelling out $P(X)$ leaves us with Bayes' theorem represented by partial moments, and our conditional probability on the right side of the equality.

$$P(X|Y) = \frac{CUPM(0|0, c|a, Y|X)}{UPM(0, c, Y)}$$

■

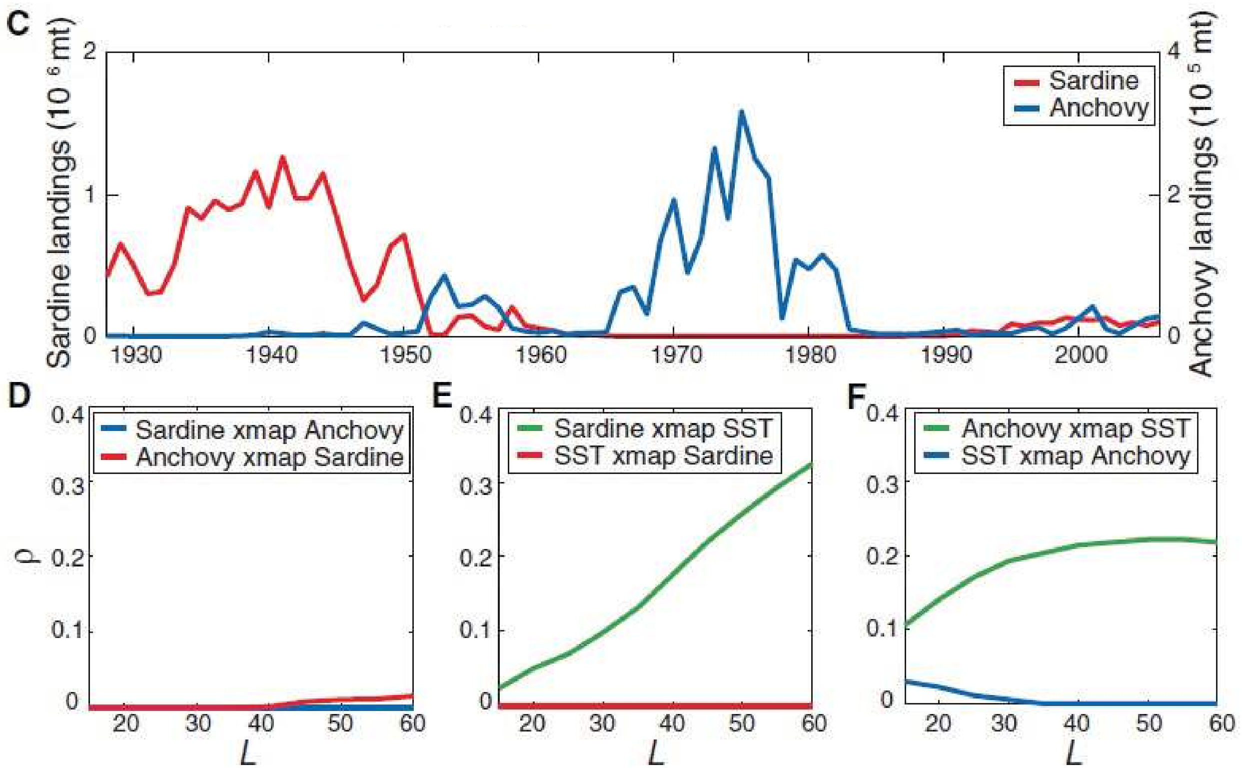
MULTIVARIATE CAUSATION MATRIX

We can construct a multivariate causation matrix summarizing all causative influences per variables in question. We first use our method on the Sardine-Anchovy-Sea Surface Temperature example in Sugihara et al. and compare our results to the CCM method. We then apply our method to the S&P 500 – 10 Year Treasury Yield – Money Supply relationship.

Sugihara et al. Sardine – Anchovy – SST Example Replication

Sugihara et al. examine the relationship among Pacific sardine landings, northern anchovy landings, and sea surface temperature (SST). Figure 7 below, reproduced from Sugihara et al. panel C shows the California landings of Pacific sardine and northern anchovy, while panels D to F show the CCM (or lack thereof) of sardine versus anchovy, sardine versus SST, and anchovy versus SST respectively. Sugihara et al. contend this shows that sardines and anchovies do not interact with each other and that both are weakly forced by temperature.

Figure 7. Reproduced Figures 5C through 5F from Sugihara et al. [2012].



This example raises an important correlation consideration, especially when the differences in variables are in orders of magnitude. The sardine landings (left y-axis) and anchovy landings (right y-axis) in figure 7 are represented in different orders of magnitude for their unnormalized observations. Linear correlation coefficients are ill suited for such analysis. Figure 8 from VN[2012] illustrates the VN correlation coefficient differences under such an extreme scale consideration ($Y = X^{10}$) versus the Pearson correlation coefficient.

$Y = X^{10}$

```
> x=seq(0,3,.01);y=x^10
> cor(x,y)
[1] 0.6610183
> NNS.dep(x,y,print.map = T)
$Correlation
[1] 0.9812511
$Dependence
[1] 0.9812511
```

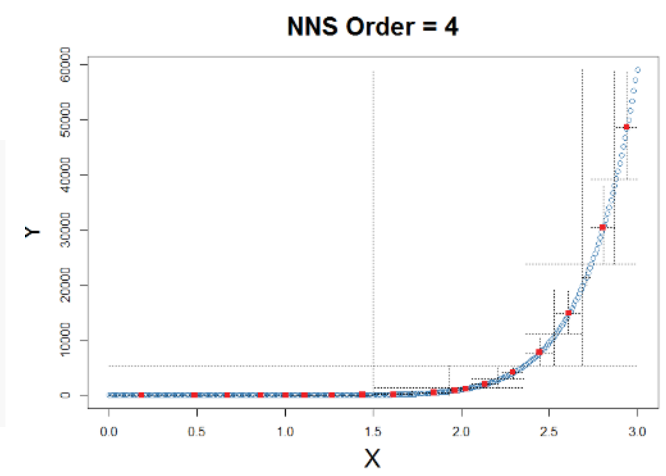


Figure 8. Correlation coefficients for nonlinear relationship on extreme scale. Source Viole and Nawrocki [2012b].

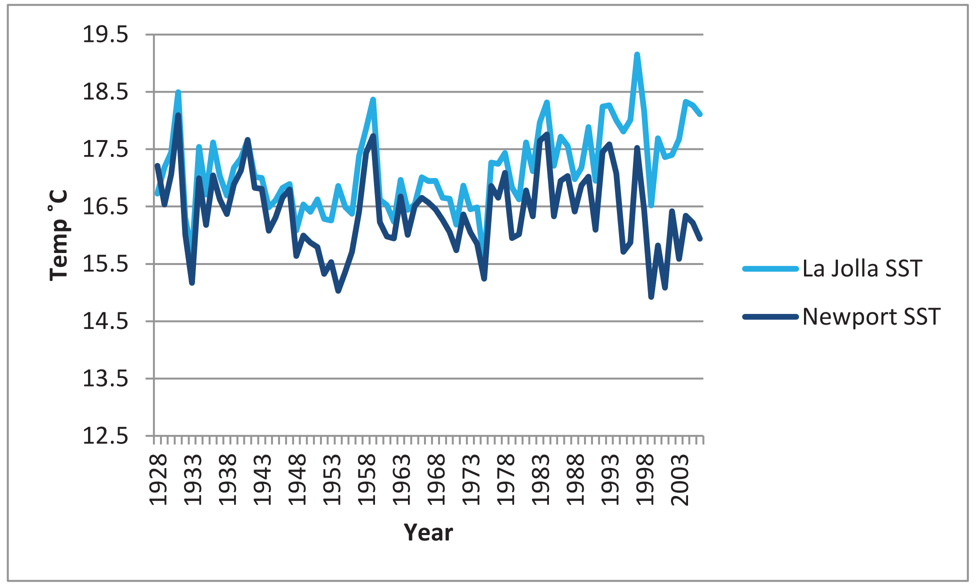
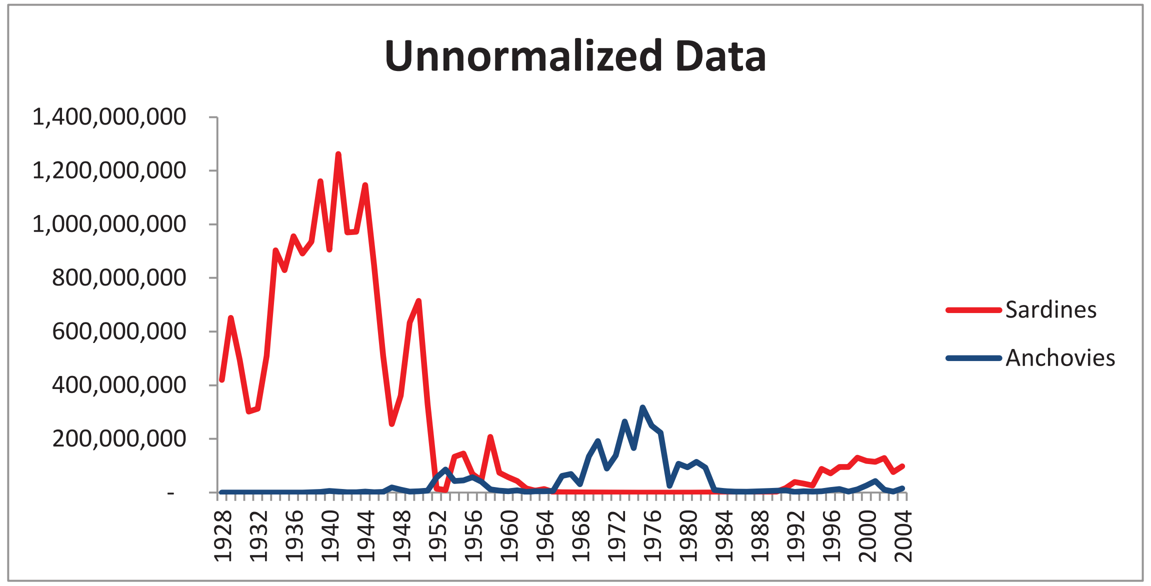


Figure 9. Newport and La Jolla SST relationship visualized. Newport Beach SST data were used for anchovy data set versus La Jolla SST for sardine data set per Sugihara et al. procedure.

Figure 9 illustrates the (nonlinear) relationship between Newport and La Jolla SST. The VN correlation coefficient under a less extreme scale consideration versus the Pearson correlation coefficient are .43 and .6541 respectively. The extreme scaling differences, present even after normalization, argue for the more accurate nonlinear VN correlation coefficient. Figure 10 represents the results of the VN normalization process. Sugihara et al. use a first difference normalization technique with unintended consequences as will be discussed later.



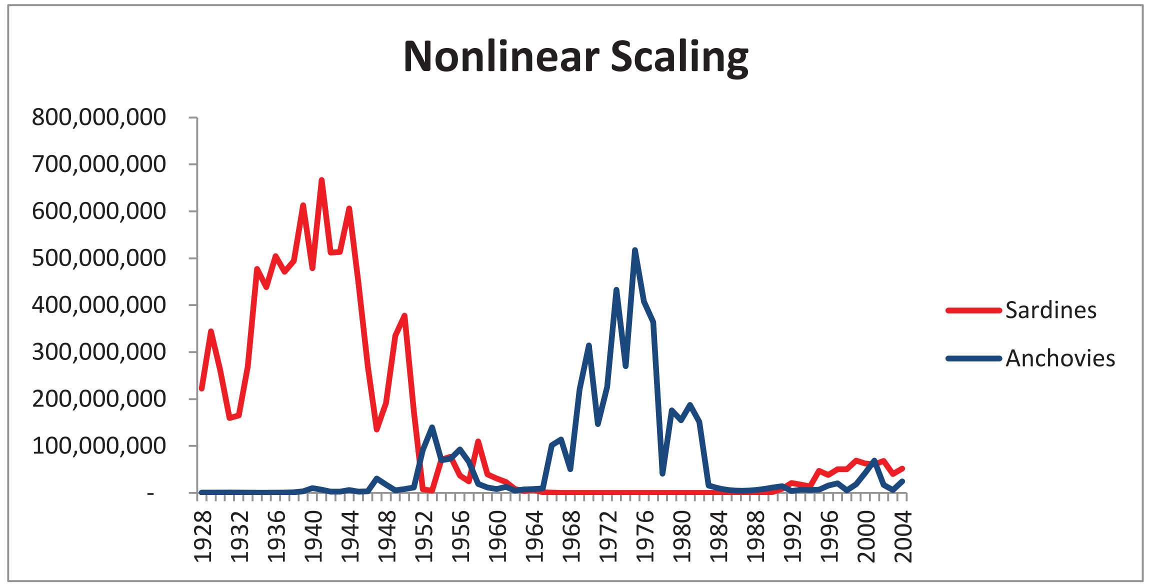


Figure 10. Unnormalized and Normalized Sardine and Anchovy landings per the VN process. Successfully eliminating orders of magnitude differences while maintaining distributional properties.

Table 1. Sardine-Anchovy data set with $\tau = 2$ for normalization. A: The conditional probability matrix; B: The VN ρ on the normalized data; C: The Pearson ρ on the normalized data (for comparison to VN results); D: Causality matrix.

A

		$P(X''\|Y'')$ X	
		Sardines	Anchovies
Y	Sardines	-	.775
	Anchovies	1.0	-

B

		$\rho_{X'',Y''}$ X	
		Sardines	Anchovies
Y	Sardines	-	(.5663)
	Anchovies	(.5663)	-

C

		Pearson $\rho_{X'',Y''}$ X	
		Sardines	Anchovies
Y	Sardines	-	(.358)
	Anchovies	(.358)	-

A*B=D

		$C(X \rightarrow Y)$ X	
		Sardines	Anchovies
Y	Sardines	-	(.4388)
	Anchovies	(.5633)	-

Table 2. Sardine-SST data set with $\tau = 2$ for normalization. A: The conditional probability matrix; B: The VN ρ on the normalized data; C: The Pearson ρ on the normalized data (for comparison to VN results); D: Causality matrix.

A

		$P(X''\|Y'')$ X	
		Sardines	SST
Y	Sardines	-	.008
	SST	1.0	-

B

		$\rho_{X'',Y''}$ X	
		Sardines	SST
Y	Sardines	-	(.157)
	SST	(.157)	-

C

		Pearson $\rho_{X'',Y''}$ X	
		Sardines	SST
Y	Sardines	-	(.18)
	SST	(.18)	-

A*B=D

		$C(X \rightarrow Y)$ X	
		Sardines	SST
Y	Sardines	-	(.0013)
	SST	(.157)	-

Table 3. Anchovy-SST data set with $\tau = 2$ for normalization. A: The conditional probability matrix; B: The VN ρ on the normalized data; C: The Pearson ρ on the normalized data (for comparison to VN results); D: Causality matrix.

A

			$P(X''\|Y'')$ X	
			SST	Anchovies
	Y	SST	-	1.0
		Anchovies	.005	-

B

			$\rho_{X'',Y''}$ X	
			SST	Anchovies
	Y	SST	-	(.0067)
		Anchovies	(.0067)	-

C

			Pearson $\rho_{X'',Y''}$ X	
			SST	Anchovies
	Y	SST	-	.1459
		Anchovies	.1459	-

A*B=D

			$C(X \rightarrow Y)$ X	
			SST	Anchovies
	Y	SST	-	(.0067)
		Anchovies	(.00003)	-

Sugihara et al. Sardine – Anchovy – SST Example Discussion

Sugihara et al. [2012] declare from the implementation of the CCM method on the sardine – anchovy – SST dataset,

> "In addition, as expected, there is no detectable signature from either sardine or anchovy in the temperature manifold; obviously, neither sardines nor anchovies affect SST."

We concur that there is no anchovy signature in the SST data. However, there is a very slight sardine signature. Obviously sardines do not affect SST, but we are measuring their presence through landing data. Given this semantic clarification, perhaps the sardines pick up on another diminishing variable which is more sensitive to other water conditions (salinity?) and also have inverse causal relationships. The sardines leave (diminished presence) due to this omitted variable, and the SST subsequently rises. *The sardines did not cause the water temperature increase, they anticipated the rise and left.*

> "Thus, although sardines and anchovies are not actually interacting, they are weakly forced by a common environmental driver, for which temperature is at least a viable proxy. Note that because of transitivity, temperature may be a proxy for a group of driving variables (i.e., temperature may not be the most proximate environmental driver)." Sugihara et al. [2012].

We measure the presence of sardines and presence of anchovies as inversely related (nonlinearly) due to the substantial difference in correlations between the VN and Pearson correlation coefficients; and in a manner consistent with the bidirectional coupling case from Sugihara et al. The minimal net effect sardine-anchovy of (.1275) also suggests another variable at play. We are not here to prove causation of sardine and anchovy landing data, as the authors' focus of finance and economics precludes them from accurately selecting relevant variables. However, we do offer a contending insight to the Sugihara et al. conclusion using exclusively nonlinear techniques.

This striking linear vs. nonlinear difference occurs in the very first step, the normalization techniques on the raw data. Sugihara et al. use the first difference in data points to normalize the data in CCM. This standard normalization technique results in a Pearson correlation of -.073 and equally paltry .0278 VN correlation coefficient for sardines versus anchovies. However, this is compared with a -.3579 Pearson and -.67 VN correlation coefficient on the raw data. Table 3 below presents the Pearson correlation coefficient for the raw data set, the Sugihara et al. first differences data set, and the VN normalized data set.

Table 3. Normalization effects on Pearson correlations and resulting correlation matrices.

	Raw Data Pearson ρ				1st Differences Normalized data Pearson ρ			
	SST	**Anchovy**	**Sardine**	**SST(NB)**	**SST**	**Anchovy**	**Sardine**	**SST(NB)**
SST	1	(.3043)	(.10)	.6541	1	(.13)	.017	.8694
Anchovy	(.3043)	1	(.358)	(.2431)	(.13)	1	(.073)	(.0632)
Sardine	(.10)	(.358)	1	.1607	.017	(.073)	1	.0403
SST(NB)	.6541	(.2431)	.1607	1	.8694	(.0632)	.0403	1

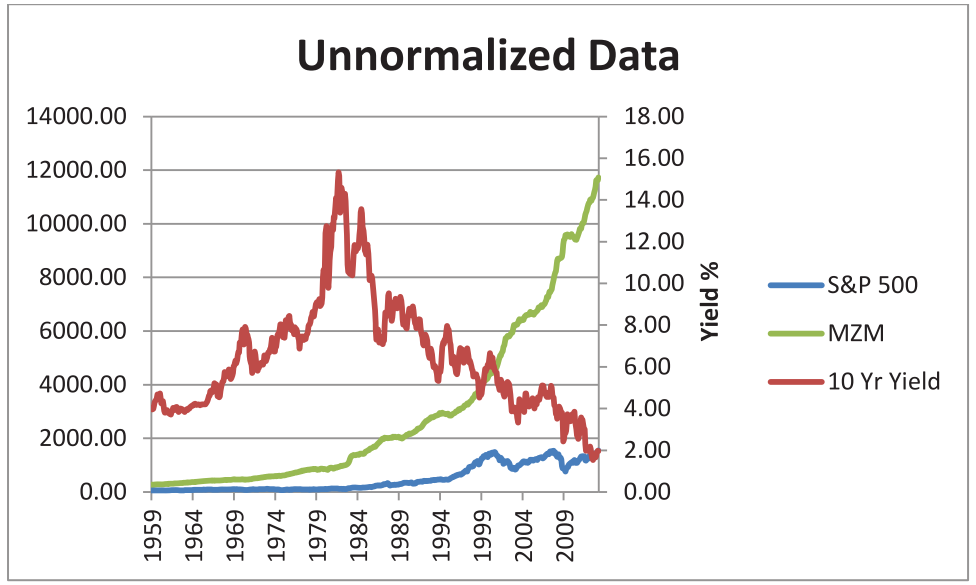
VN Normalized data Pearson ρ

	SST	**Anchovy**	**Sardine**	**SST(NB)**
SST	1	(.3043)	(.10)	.6541
Anchovy	(.3043)	1	(.358)	(.2431)
Sardine	(.10)	(.358)	1	.1607
SST(NB)	.6541	(.2431)	.1607	1

A closer examination of the normalization processes reveals the VN nonlinear scaling method retains the *identical* results for both Pearson and VN correlation coefficients while the first differences method eliminates the underlying sardine-anchovy-SST relationships.

<u>Money Supply – S&P 500 – 10 Year US Treasury Yield Example</u>

We present the findings of on the S&P 500 – 10 Year Treasury Yield – Money Supply relationship through our method using a three variable normalization versus the two variable prior example.[v]



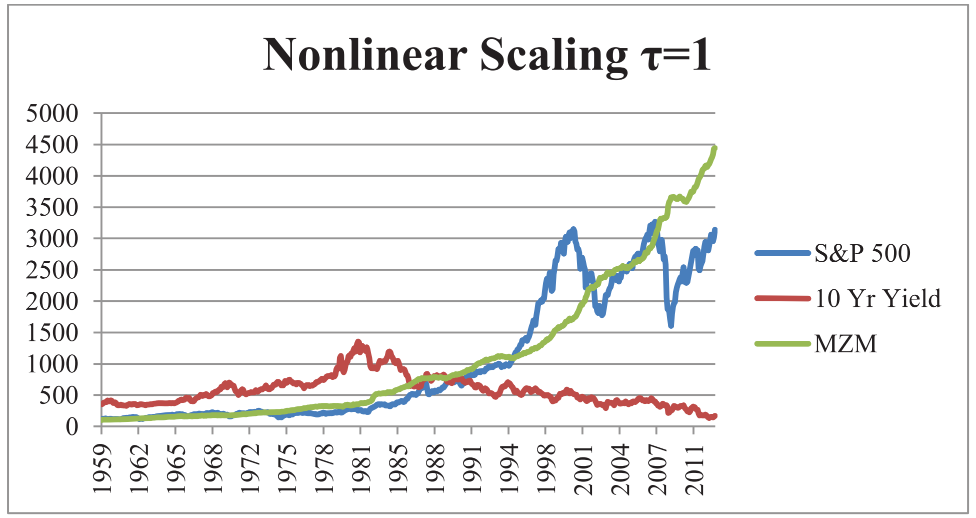


Figure 11. Visual representation of the unnormalized (top) dual y-axis and final normalized variables (τ=1) single y-axis using the method presented in Viole and Nawrocki [2013]. Also illustrates the ability for true multivariable normalization.

Figure 11 illustrates the effects of the nonlinear scaling normalization process on multiple variables. The resulting normalized variables are analogous to the manifolds offered in CCM and present the system as a whole for consideration by placing them on a shared axis.

One important feature is that $\boldsymbol{MZM''}$ has a conditional probability equal to one given the events of both the $\mathbf{10}\,\boldsymbol{Year\ Yield''}$ and the $\boldsymbol{S\&P500''}$. All of the normalized data points fit within the normalized range for $\boldsymbol{MZM''}$ per figure 11 above. These numbers are in red in section A of table 4 below.

The correlation coefficient in section B of table 4 represents the 3rd order nonlinear correlation coefficient as demonstrated in VN [2012]. This offers a distinct insight versus its linear alternative, the Pearson correlation coefficient.

Table 4. Financial variable dataset with $\tau = 1$ for normalization. A: The conditional probability matrix; B: The VN ρ on the normalized data; C: Causality matrix with cumulative causation in the bottom row and cumulative effect in far right column.

A

		$P(X''\|Y'')$ X		
		S&P 500	10 Year Yield	MZM
Y	SPY	-	.6867	1.0
	10 Year Yield	1.0	-	1.0
	MZM	.9074	.6651	-

B

		$\rho_{X'',Y''}$ X		
		S&P 500	10 Year Yield	MZM
Y $\rho_{Y'',X''}$	SPY	1.0	(.2841)	.5031
	10 Year Yield	(.2841)	1.0	(.5287)
	MZM	.5031	(.5287)	1.0

C

		$C(X \rightarrow Y)$ X		
		S&P 500	10 Year Yield	MZM
Y	SPY	-	(.1940)	.5031
	10 Year Yield	(.2841)	-	(.5287)
	MZM	.4565	(.3517)	-
	$\sum C(X \rightarrow Y)$	.1724	(.5457)	(.0256)

<u>**We can state that MZM is a cause to S&P 500 prices and inverse cause to 10 year Treasury yields net of the bidirectional coupling the variables share.**</u> It should be noted that the linear Pearson correlation resulted in extremely high correlations, and consequently causation for these same variable sets($\rho_{X'',Y''} > .90$). These above results are consistent (and stronger) with the asymmetrical bidirectional coupling predator – prey example in Sugihara et al. and with Black's casual argument on the intertwined relationship between money stock and economic activity.

Rogalski and Vinso [1977] through GC firmly reject the hypothesis that causality runs unidirectionally from past values of money to equity returns. Their results are consistent with the hypothesis that stock returns are not purely passive but perhaps influence money supply in some complicated fashion. Our results showing asymmetrical bidirectional coupling directly support Rogalski and Vinso's contention.

DISCUSSION

Fischer Black had a very insightful article on causation in "The Trouble with Econometric Models." Black recommends experiments to isolate the causal variable in question, *conditional probability*. He illustrates several examples identifying the specification error associated with conditional probability as the *cause* of the lack of causality. Black sums it up beautifully,

We just can't use correlations, with or without leads and lags, to determine causation.

That's as true today as it was decades ago. However, we can now say:

We need correlations and conditional probabilities, with and without leads and lags, to determine causation.

Granger causality was predicated on prediction instead of correlation to identify causation between time-series variables. Stochastic variables predicated on nonlinear relationships do not lend themselves to prediction, especially if they are not strongly synchronized.

> "Therefore, information about X(t) that is relevant to predicting Y is redundant in this system and cannot be removed simply by eliminating X as an explicit variable. When Granger's definition is violated, GC calculations are no longer valid, leaving the question of detecting causation in such systems unanswered." Sugihara et al. [2012]

While CCM was not designed to compete with GC, rather is specifically aimed at a class of system not covered by GC (nonseparable, weakly coupled systems effected by shared driving variables), our method is aimed at all systems. We normalize the variables to lagged observations of themselves, nonlinearly. We normalize the normalized variables to the other normalized variables of interest, nonlinearly. We generate nonlinear correlations between the normalized variables. All of the nonlinear methods employed fully replicate linear situations as demonstrated in Viole and Nawrocki [2012b, 2013].

The authors' main focus is economics and finance. This binding condition inhibits them from extending the analysis to other areas such as biology or ecological systems as the convergent cross mapping method exemplifies without collaboration. We could provide many more axiomatic examples of known (and unknown) conditional probabilities as Black does for support (or rejection) of causation, but experimentation and empirical analysis will ultimately serve as proof to this theoretical work. We look forward to extending the discussion to other fields in search of these experiments, thus satisfying the conditional probability requirement in proving causation.

APPENDIX A

EMPIRICAL CONDITIONAL PROBABILITY EXAMPLE

Earlier we illustrated the conditional probability for a given occurrence using partial moments from normalized variables. However, if we wish to further constrain the conditional distribution to positive and negative occurrences we need to use co-partial moments of reduced the reduced observation count. This differs from a joint probability where the number of observations is not reduced to the conditional occurrences.

The following example will generate the conditional probability of a specific occurrence with Bayes' theorem, then with our method. Given 100 observations of 10 Year yield returns and S&P 500 returns (normalized by percentage return), what is the probability that given an interest rate increase, stocks rose?

Using the following data in Table 1A, we are after the bold red numbers:

	S&P 500	10 Year Yield		S&P 500	10 Year Yield
1/1/2005	**2.56%**	0.95%	5/1/2007	**3.95%**	2.81%
2/1/2005	-1.50%	-0.24%	6/1/2007	**3.19%**	1.27%
3/1/2005	1.53%	-1.19%	7/1/2007	**0.22%**	7.11%
4/1/2005	**-0.40%**	7.62%	8/1/2007	0.41%	-1.98%
5/1/2005	-2.58%	-3.62%	9/1/2007	-4.44%	-6.83%
6/1/2005	1.18%	-4.72%	10/1/2007	2.88%	-3.26%
7/1/2005	2.01%	-3.44%	11/1/2007	**2.80%**	0.22%
8/1/2005	**1.65%**	4.40%	12/1/2007	-5.08%	-8.76%
9/1/2005	**0.17%**	1.90%	1/1/2008	1.08%	-1.21%
10/1/2005	0.13%	-1.42%	2/1/2008	-7.03%	-9.19%
11/1/2005	**-2.81%**	6.01%	3/1/2008	-1.75%	0.00%
12/1/2005	**3.74%**	1.78%	4/1/2008	-2.84%	-6.35%
1/1/2006	1.98%	-1.55%	5/1/2008	**3.98%**	4.73%
2/1/2006	1.31%	-1.12%	6/1/2008	**2.36%**	5.29%
3/1/2006	**-0.16%**	3.34%	7/1/2008	**-4.52%**	5.52%
4/1/2006	**1.33%**	3.23%	8/1/2008	-6.46%	-2.22%
5/1/2006	**0.65%**	5.56%	9/1/2008	1.90%	-3.04%
6/1/2006	**-0.94%**	2.38%	10/1/2008	-5.16%	-5.28%
7/1/2006	-2.90%	0.00%	11/1/2008	**-22.81%**	3.20%
8/1/2006	0.57%	-0.39%	12/1/2008	-9.27%	-7.63%
9/1/2006	2.11%	-4.21%	1/1/2009	-0.62%	-37.75%
10/1/2006	2.35%	-3.33%	2/1/2009	**-1.37%**	4.05%
11/1/2006	**3.40%**	0.21%	3/1/2009	**-7.23%**	13.01%
12/1/2006	1.84%	-2.79%	4/1/2009	-6.16%	-1.76%
1/1/2007	1.98%	-0.87%	5/1/2009	**11.35%**	3.83%
2/1/2007	**0.54%**	4.29%	6/1/2009	**6.20%**	11.59%
3/1/2007	1.44%	-0.84%	7/1/2009	**2.59%**	12.28%
4/1/2007	-2.65%	-3.45%	8/1/2009	1.04%	-4.40%

	S&P 500	10 Year Yield		S&P 500	10 Year Yield
9/1/2009	**7.60%**	0.84%	1/1/2012	1.37%	-1.50%
10/1/2009	3.39%	-5.44%	2/1/2012	4.50%	-0.51%
11/1/2009	2.19%	-0.29%	3/1/2012	3.91%	0.00%
12/1/2009	**1.89%**	0.29%	4/1/2012	**2.68%**	9.67%
1/1/2010	**2.03%**	5.44%	5/1/2012	-0.20%	-5.69%
2/1/2010	**1.18%**	3.83%	6/1/2012	-3.31%	-13.01%
3/1/2010	-3.11%	-1.08%	7/1/2012	-1.34%	-10.54%
4/1/2010	**5.61%**	1.08%	8/1/2012	2.71%	-5.72%
5/1/2010	**3.85%**	3.17%	9/1/2012	**3.16%**	9.35%
6/1/2010	-6.22%	-11.84%	10/1/2012	**2.81%**	2.35%
7/1/2010	-3.78%	-6.65%	11/1/2012	-0.39%	1.73%
8/1/2010	-0.33%	-6.12%	12/1/2012	-3.06%	-5.88%
9/1/2010	0.69%	-10.87%	1/1/2013	**1.97%**	4.15%
10/1/2010	3.15%	-1.87%	2/1/2013	**4.00%**	10.48%
11/1/2010	4.32%	-4.24%	3/1/2013	**2.13%**	3.60%
12/1/2010	**2.30%**	8.31%	4/1/2013	2.52%	-1.02%
1/1/2011	**3.49%**	17.57%			
2/1/2011	**3.26%**	2.99%			
3/1/2011	**2.96%**	5.45%			
4/1/2011	-1.27%	-4.87%			
5/1/2011	**2.05%**	1.46%			
6/1/2011	0.51%	-8.75%			
7/1/2011	-3.89%	-5.51%			
8/1/2011	2.90%	0.00%			
9/1/2011	-11.15%	-26.57%			
10/1/2011	-0.97%	-14.98%			
11/1/2011	**2.80%**	8.24%			
12/1/2011	1.58%	-6.73%			

Defining the probabilities as:

P(SI) = probability of the S&P 500 increasing
P(SD) = probability of the S&P 500 decreasing
P(II) = probability of interest rates increasing
P(ID) = probability of interest rates decreasing

	Interest Rate Increase	Interest Rate Decrease	Interest Rate Unchanged	Total
S&P Increase	35 **CUPM**	28 **DLPM**	2	65 UPM
S&P Decrease	9 **DUPM**	24 **CLPM**	2	35 LPM
S&P Unchanged	0	0	0	0
Total	44 **UPM**	52 **LPM**	4	100

Table 2A. Bayes' Theorem probabilities identified and displayed from the data in table 1A. Corresponding partial moments quadrants also represented.

According to Bayes' theorem $P(SI|II) = \frac{P(II|SI)\,P(SI)}{P(II)}$

$$P(SI|II) = \frac{\left(\frac{35}{65}\right)\left(\frac{65}{100}\right)}{\left(\frac{44}{100}\right)} = \left(\frac{35}{44}\right) = 79.55\%$$

This example raises an immediate concern - in the instance where there is a zero return, the observation is neither a gain nor a loss. These observations are highlighted in grey in table 1A. When an observation equals a target in the partial moment derivations, that observation is placed into an empty set; analogous to the unchanged column in the table above. Empty sets reduce both the lower and upper partial moments, thus their effect is symmetrical to the resulting statistics.

Using our method:

Figure 1A below illustrates the normalized distributions from the data in table 1A. Using equation 3, we can see that the S&P 500 degree zero upper partial moment from the minimum 10 Year Yield observation is equal to .7955. The S&P 500 degree zero upper partial moment from the maximum 10 Year Yield observation is equal to zero. Thus, the conditional probability of a positive S&P 500 return given an increase in 10 Year Yields is equal to 79.55%, represented by the lighter shaded blue.

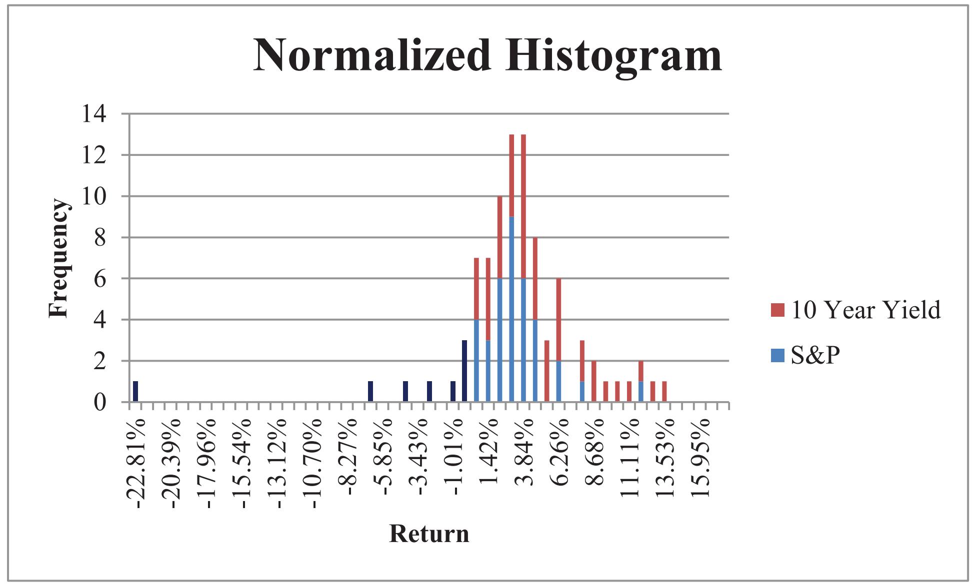


Figure 1A. Graphical representation of conditional probability of positive S&P500 return given an increase in 10 Year Yields.

Alternatively, we can derive the same conclusion with conditional partial moments. The frequency of positive 10 Year Yield returns is represented by the degree zero upper partial moment from a zero target, where *X*= S&P 500 and *Y* = 10 year yield.

$$UPM(0,0,Y) = \frac{1}{T}\sum_{t=1}^{T}\{\max(Y_t - 0), 0\}^0 = 0.44$$

In R where sp = S&P 500 and ten.yr = 10 year yield:

```
> UPM(0,0,ten.yr)
[1] 0.44
```

The number of occurrences is (0.44 * T) which yields 44 in this example. Using T* as our reduced universe of observations, we compute the conditional upper partial moment for a direct computation of the conditional probability from the underlying time series.

$$CUPM(0|0,0|0,X|Y) = \left(\frac{1}{T^*}\right) \cdot \sum_{t^*=1}^{T^*}[max(X_{t^*} - 0,0)]^0\,[max(Y_{t^*} - 0,0)]^0$$

In our example,

$$CUPM(0|0,0|0,X|Y) = .7955$$

And in R:

```
> Co.UPM(0,0,sp,ten.yr,0,0)/UPM(0,0,ten.yr)
[1] 0.7954545
> UPM(0,0,sp[ten.yr>0])
[1] 0.7954545
```

But, this result isn't particularly interesting or innovative since degree zero partial moments are frequency and counting statistics – just as in the Bayes derivation.

However, the method permits an easy conversion to a conditional expected shortfall measure whereby the average S&P increase given an increase in interest rates can be computed by changing the degree of the X term to 1 from 0.

$$CUPM(1|0,0|0,X|Y) = \left(\frac{1}{T^*}\right) \cdot \sum_{t^*=1}^{T^*} [max(X_{t^*} - 0,0)]\,[max(Y_{t^*} - 0,0)]^0$$

In our example the average S&P 500 increase given an increase in interest rates is,

$$CUPM(1|0,0|0,X|Y) = 1.5\%$$

And in R:

```
> (Co.UPM(1,0,sp,ten.yr,0,0)-D.UPM(1,0,sp,ten.yr,0,0))/UPM(0,0,ten.yr)
[1] 0.01495909
> UPM(1,0,sp[ten.yr>0])-LPM(1,0,sp[ten.yr>0])
[1] 0.01495909
```

Both methodologies yield the same conditional probability which is not surprising given the simple frequency requirement of the underlying calculation and same associated targets for the partial moments. However, since partial moments are already used in portfolio analysis their flexibility in constructing other relevant statistics is often overlooked.

REFERENCES

REFERENCES

Billingsley, P. (1968), "*Convergence of Probability measure,"* John Wiley and Sons, New York, third edition.

Black, Fischer [1984]. "The Trouble with Econometric Models." *Financial Analysts Journal*, Vol. 38, No. 2, pp. 29-37.

B. M. Bolstad, R. A. Irizarry, M. Astrand and T. P. Speed [2003] "A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias." *Bioinformatics*, Vol. 19, Number 2, p. 185-193.

Chen, Y. A., Almeida, J., Richards, A., Muller, P., Carroll, R., and Rohrer, B. [2010]. "A Nonparametric Approach to Detect Nonlinear Correlation in Gene Expression." *Journal of Computational and Graphical Studies*, Volume 19, Number 3, p. 552-568.

Dixon, P. A., M. J. Milicich, and G. Sugihara. [1999]. "*Episodic fluctuations in larval supply*." *Science*, Vol. 283, pp. 1528-1530.

Deyle, E.R, and Sugihara G. [2011]. *Generalized Theorems for Nonlinear State Space Reconstruction*. PLoS ONE 6(3): e18295.

Estrada, Javier, (2008). "Mean-Semivariance Optimization: A Heuristic Approach," *Journal of Applied Finance*, v18(1), 57-72.

Granger, C. [1969]. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods". *Econometrica*, Vol. 37, No. 3, pp. 424-438.

Grootveld, Henk and Winfried Hallerbach. (1999), "Variance vs. Downside Risk: Is There Really That Much Difference." *European Journal of Operational Research*, v114, 304-319.

Guthoff, A, Pfingsten, A., Wolf, J (1997), "*On the compatibility of Value-at-risk, other risk concepts and expected utility maximization,*" in: Hipp, C. et.al. (eds).

Guthoff, A., Pfingsten, A. and J. Wolf (1997). "*On the Comapatibility of Value at Risk,Other Risk Concepts, and Expected Utility Maximization"*; in: Hipp, C. et.al. (eds.): *Geld, Finanzwirtschaft, Banken und Versicherungen: 1996; Beiträge zum 7. Symposium Geld, Finanzwirtschaft, Banken und Versicherungen an der Universität Karlsruhe vom 11.-13. Dezember 1996*, Karlsruhe 1997, p. 591-614.

Holthausen, D. M. (1981). "A Risk-Return Model With Risk And Return Measured As Deviations From a Target Return." *American Economic Review*, v71(1), 182-188.

Kaplan, P. and Knowles, J. (2004), “Kappa: A Generalized Downside Risk-Adjusted Performance Measure.” *Journal of Performance Measurement*, 8(3), 42-54.

Lucas, D. (1995). “Default Correlation and Credit Analysis.” *Journal of Fixed Income*, Vol. 11, pp. 76-87.

Markowitz, Harry. 1959, *Portfolio Selection*. (First Edition). New York: John Wiley and Sons.

Pitman, E.J.G. (1979). “Some Basic Theory for Statistical Inference.” London, Chapman and Hall.

Rogalski, R. J., and Vinso, J. D. [1977] "Stock Returns, Money Supply, and the Direction of Causality." *Journal of Finance*, September 1977, pp. 1017-1030.

Shadwick, W. and Keating, C. (2002), “A Universal Performance Measure.” *Journal of Performance Measurement*, Spring 2002, pp. 59-84, 2002.

G.R. Shorack, J.A. Wellner, (1986), “Empirical processes with applications to statistics,” Wiley Series in Probab. and Math. Stat.: Probab. and Math. Stat. John Wiley & Sons, Inc., New York.

Sugihara, G., May, R., Ye, H., Hsiech, C., Deyle, E., Fogarty, M., Much, S. [2012]. “Detecting Causality in Complex Ecosystems.” *Science*, Vol. 338, pp. 496-500.

Takens, F. [1981] in *Dynamical Systems and Turbulence*, D. A. Rand, L. S. Young, Eds. (Springer-Verlag, New York, 1981), pp. 366–381.

Viole, F. and Nawrocki, D. [2012a]. “Deriving Cumulative Distribution Functions & Probability Density Functions Using Partial Moments.” Available at SSRN: http://ssrn.com/abstract=2148482

Viole, F. and Nawrocki, D. [2012b]. “Deriving Nonlinear Correlation Coefficients from Partial Moments.” Available at SSRN: http://ssrn.com/abstract=2148522

Viole, F. and Nawrocki, D. [2012c]. “f(Newton).” Available at SSRN: http://ssrn.com/abstract=2186471 .

Viole, F. and Nawrocki, D. [2013]. “Nonlinear Scaling Normalization with Variance Retention”. Available http://ssrn.com/abstract=2262358

A.W. van der Vaart, J.A. Wellner, Jon A. (1996), “Weak convergence and empirical processes.” With applications to statistics. Springer Series in Stat. Springer-Verlag, New York.

Wang, G.S., [2008]. “A Guide to Box-Jenkins Modeling.” *Journal of Business Forecasting*; Spring 2008, Vol. 27 Issue 1, p19

http://demonstrations.wolfram.com/SingleFactorAnalysisOfVariance/

---

[i] Newton proved the integral of a point in a continuous distribution to be equal to zero.
[ii] If no data exists in a subset, no mean is calculated.
[iii] The horizontal line as in the equation $Y = 1$ (point probability) yields a 0 correlation for both Pearson’s correlation and our metric.
[iv] All variables in the regression are exchange traded funds (ETFs) that trade in US markets: SPY is the S&P 500 ETF, TLT is the Barclays 20+ year Treasury Bond ETF, GLD is the Gold Trust ETF, FXE is the Euro Currency ETF, and GSG is the S&P GSCI Commodity Index ETF.
[v] The data are monthly series from 01/01/1959 through 04/01/2013. They are available from FRED with links to graphs and data for each of the variables listed.

http://research.stlouisfed.org/fred2/graph/?id=SP500

http://research.stlouisfed.org/fred2/graph/?s[1][id]=GS10

http://research.stlouisfed.org/fred2/series/MZMNS?rid=61

Proof

Printed By Createspace

createspace™
an Amazon.com company
Digital Proofer