Some Inspired Non-Parametric Portfolio Approaches of James R. Thompson

John A. Dobelman[1]

[1]Rice University, Department of Statistics, 6100 Main Street, Houston, Texas 77005,
dobelman@stat.rice.edu

Abstract
Security returns are an example of phenomena whose distributions defy parametric modeling. Jim Thompson used a variety of non-parametric approaches to develop workable investing solutions in such an environment. We review his ground-breaking exploration of the veracity of the capital asset pricing model (CAPM), and several non-parametric approaches to portfolio formulation including the Simugram^{TM}, variants of his Max-Median rule, and Tukey weightings

Key Words: Non-parametric statistics, simulation, portfolio construction, optimization, MaxMedian, power means

1. Introduction

We are honored and pleased to present this look at James Thompson's non-parametric market portfolio approaches. Professor James R. Thompson (1938-2017) earned his Ph.D. in Mathematics from Princeton in 1965 where John W. Tukey was his thesis advisor. Thompson was well-positioned for the age of Tukey, EDA and numerical solutions. He bridged the robustness/heavy math age of statistics with the bootstrap/computer age. He was one of the first to immediately realize the usefulness of the Bootstrap, back in the time of Julian Simon in 1969, and he became one of its evangelists, from the outset through the optimality phase of its development, to the present. We recall that the hardware processing at the time was accomplished on time-share machines or university computers such as the IBM 360, CDC-6600 or DEC PDP-8.

Like his mentor, Professor Thompson spent half a century thinking “outside the box” and creating or applying de novo methods impacting the modeling of real world problems using statistics and Aristotelian thought processes. Most statisticians familiar with Jim's work think in keywords such as mathematical biology, SIMEST, empirical model-building, cancer modeling, and non-parametric statistics (Dobelman, 2017). The purpose of this paper is to highlight the latter portion of his career in which he took up the problem of modeling and making application in financial markets.

Thompson was well aware that security returns are an example of phenomena whose distributions defy parametric modeling. He was also well-trained in Tukey's Maxim (Tukey, 1962), “Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.” To make progress in computational finance, he proceeded in the manner served him well in biological modeling, i.e. abandoning attempts at closed-form solutions and instead using simulation of fundamental underlying axiomatic probabilistic models.

Figure 1: James R. Thompson

This paper will first present some stylized facts about the security markets, review Thompson's computationally-intensive Simugram™ system, review his work in Capital Asset Pricing Model (CAPM) evaluation and validation, outline his simpler "MaxMedian" system for portfolio management, including extensions, and conclude with his investigation into alternative weighting systems using the Tukey transformation ladder.

2. Almanac of Security Returns

Jim was a staunch believer in exploratory data analysis. In this vein, we present some salient statistics on security market returns and other variables and parameters. More complete treatments may be found in references such as Chakraborti (2011), Cont (2011) and Sewell (2011), but the empirical picture here gives an adequate overview, especially for those readers for whom this subject is not of primary focus.

One of the first things we notice is that the distribution of returns is decidedly non-Gaussian. This will immediately hamper closed-form modeling. In figure 2 we find the empirical density estimate of over 77,000 annual returns from 1985-2008. A similar picture is encountered when looking at all annual (N=380,562) or monthly (N=4,248,428) security returns for the 1926-2017 period. Figure 3 compares the theoretical returns of a standard geometric Brownian motion (GBM) model with the actual returns of General Electric stock; one notes straightaway the persistence in volatility, long-range dependence of autocorrelation, volatility "clumping" and other non-stationarities in the true market returns.

It is tempting when inspecting the empirical distribution of the returns to try and fit a univariate parametric distribution, of which there are many candidate frequency curves. We then have the problem of determining which class of curves the distribution belongs to. This process is highly dependent on the values of the estimated parameters from which the determination is made. In the case of Pearson curves, the first four sample moments are required to determine which regime the curve falls in, and given the sample variability of these estimates, definitive inference eludes us, especially for the small sample sizes encountered in the only 91 years of US security returns generally available to academics.

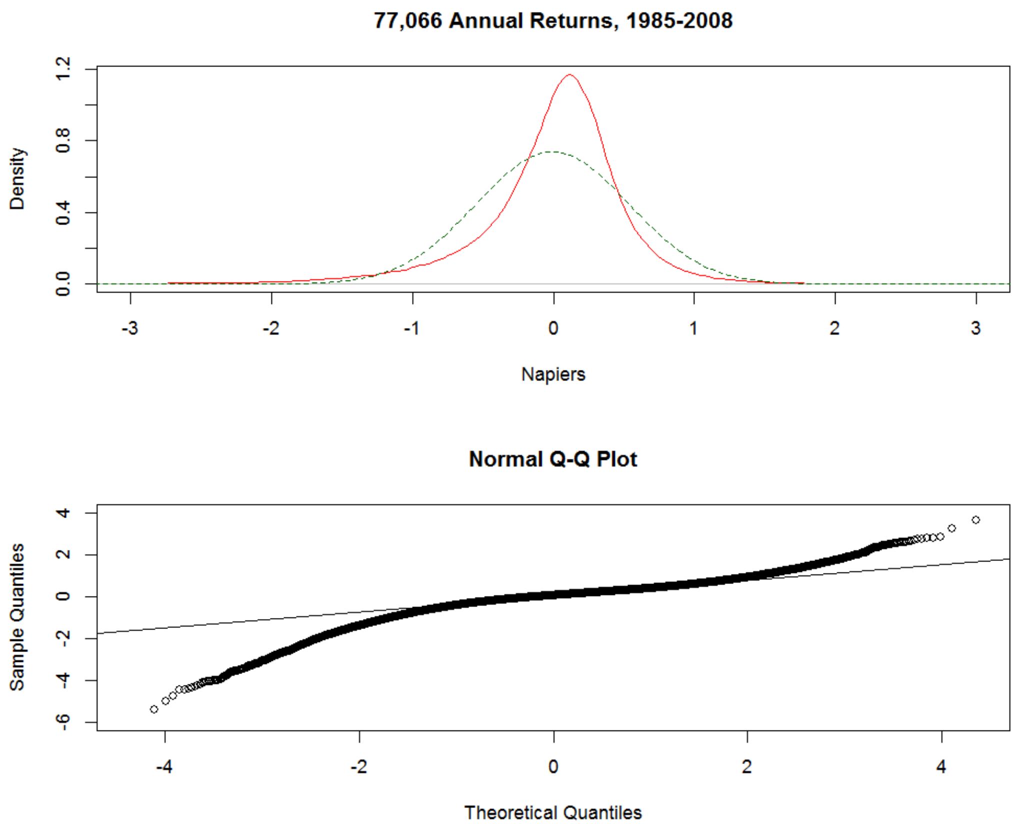

Figure 2: Realized security returns, 1985-2008 vs. normal variates

Summary statistics for financial markets attempt to encapsulate the variability in periodic realizations of these distributional draws. Figure 4 plots the annual returns and cumulative return on $1 from 1970 through 2012 from a particular portfolio construction scheme. The cumulative return plot is inherently difficult for the human eye to process unless it is transformed to linearity. In the case of log(x), an exponential nature is easy to see as a straight line. Scalar summary statistics such as average annual return or compound annual growth rate (equation 1) summarize the 42-year process. Mutual fund literature normally quote arithmetic average returns since they are always higher than geometric average rates of return; only the latter however regenerate realized terminal values.

(1) $$CAGR = \left(\prod_{i=1}^{n} R_i \right)^{1/n} - 1$$, where R_i is the annual gross return.

In figure 4, we have three series; the one with the highest terminal value of $189.6 is labeled MaxMeasures, exhibiting a compound annual growth rate (CAGR) of 13.3%. The Max Measures portfolio strategy is discussed in section 4 of this paper. The second highest terminal value of $49.1 corresponds to the S&P 500 index (with dividends), with CAGR of 9.7%, and the third is that of 30-day US Treasury bills with a terminal value of $8.6 and CAGR of 5.2%. Note the non-intuitive behavior of CAGR's and their associated terminal values. The CAGR difference between T-bills and the S&P 500 index, or the MaxMeasures and S&P 500, is 4.5%, which results in a multiplicative terminal value factor of about 570%. Differences in terminal value due to CAGR can be substantial, for example if the 42-year CAGR on a one-million-dollar investment in the S&P 500 were ½ percent lower (9.2%), the difference in terminal value is almost $10 million dollars.

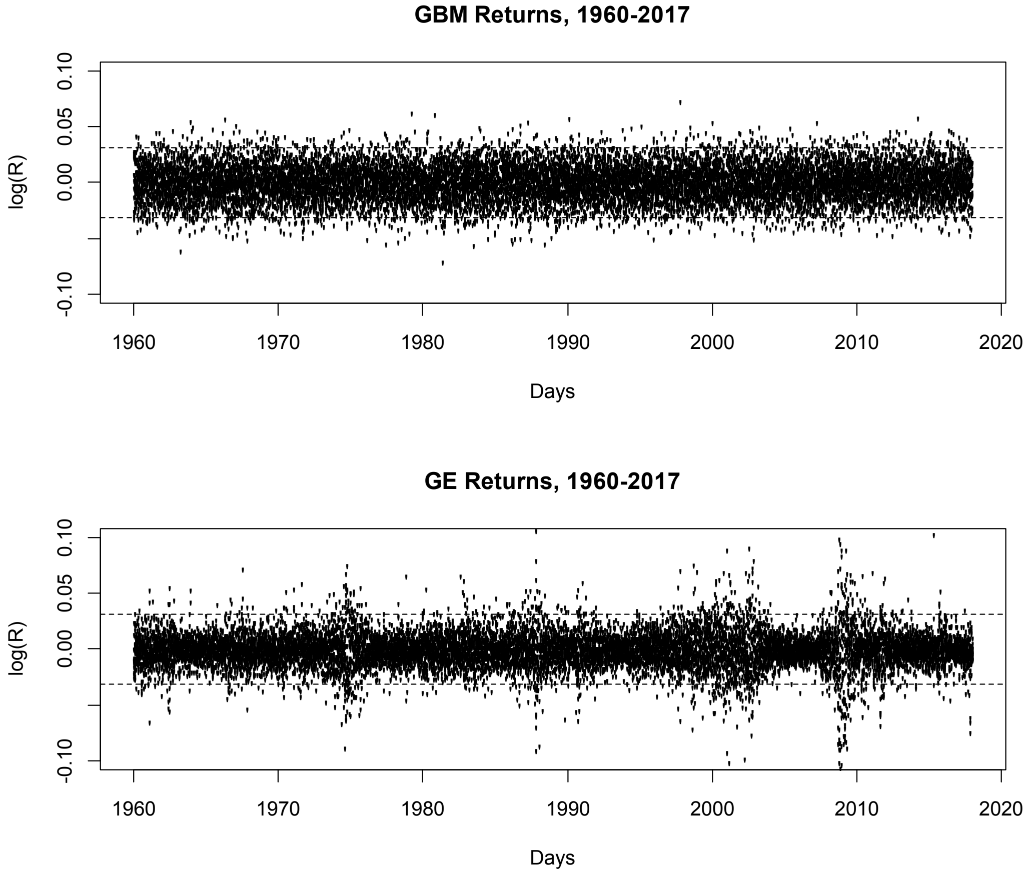

Figure 3: Comparison of geometric Brownian motion (GBM) and actual security returns, 1960-2017

Some comparative market parameters for the study period of 1926-2017 are provided in table 1 below.

Table 1: Market Parameters, 1926-2017

Total number of years study period	92
NYSE average annual return	11.8%
NYSE compound annual growth rate (CAGR)	9.81%
Average excess return (r-r_f)	8.5%
Number of years with negative returns	24 (26%)
Number of years return less than r_f	29 (32%)
GBM parameter estimates (μ,σ)	(.062, .164)
Last Saturday trading	May 24, 1952
Median company survival (years)	6.92

CAGR's for various major US indexes are provided in table 2 below. We quickly note the importance of reinvesting dividends, and that the equal weight indexes substantially outperform the market weighted composition.

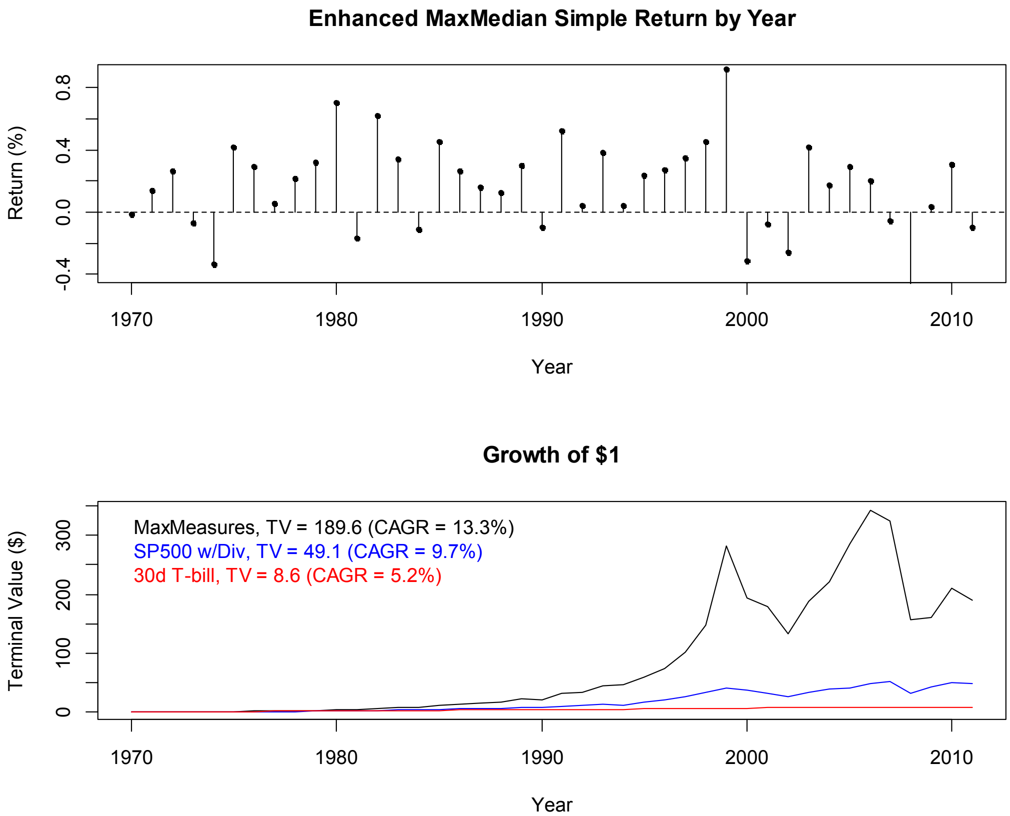

Figure 4: Annual returns and compound growth of $1 for the period 1970-2012

Table 2: Compound annual growth rate (CAGR) of select major U.S. Stock Indexes. VW and EW are correspond to market-value weighted and equal-weighted, respectively. Dividends included or excluded as indicated by D or X. Source: The Center for Research in Security Prices (CRSP).

Current Date:	12/31/17			CAGR		
Index	Begin Date	N	VWRETD	VWRETX	EWRETD	EWRETX
SP500	12/31/1925	92	0.100	0.059	0.118	0.078
NYSE	12/31/1925	92	0.098	0.057	0.123	0.081
AMEX	12/31/1962	55	0.077	0.053	0.126	0.103
NASDAQ	12/31/1972	45	0.106	0.089	0.131	0.116

Figure 5 depicts market volatility and the risk-free interest rate ("r_f"); the latter is usually proxied by U.S. Government Treasury Bills (T-bills) of short duration (such as 30-day bills). The long run average T-bill rate has been about 3.3% over the study period. The mean volatility is 15.3%, or about 1% per day.

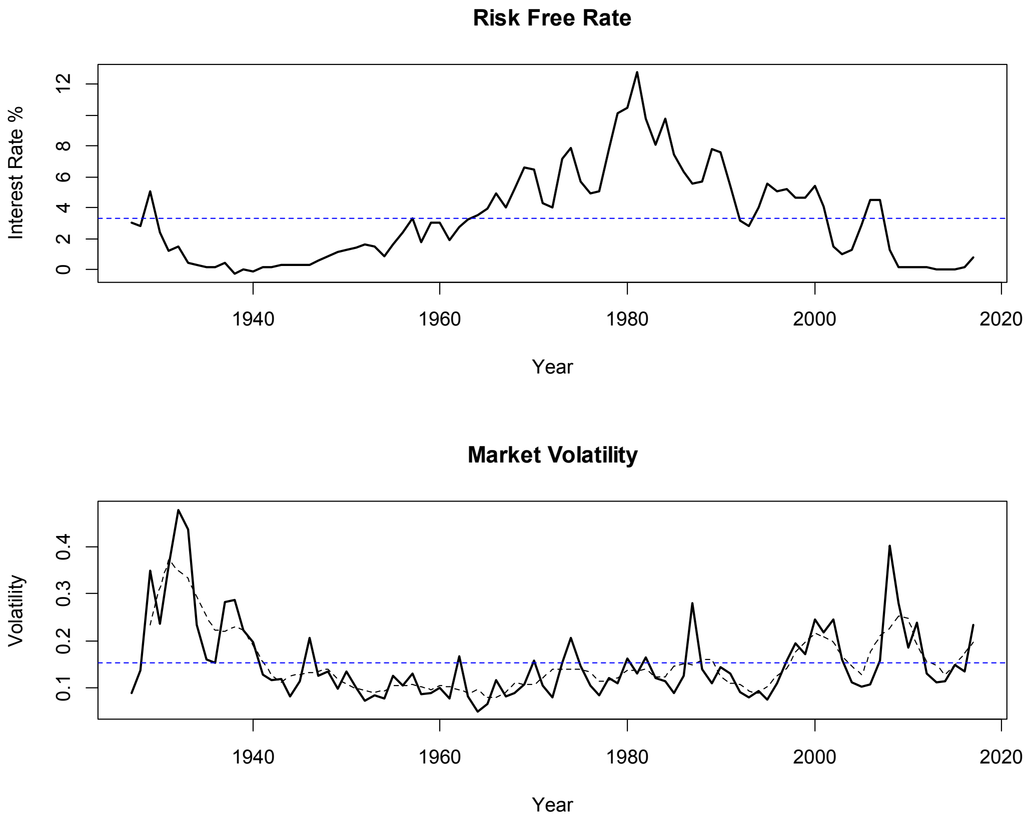

Figure 5:"Risk-free rate" and major stock index annual volatility, 1926-2017. Smoothed volatility curve is 5-year simple moving average

The actual NYSE/AMEX/NASDAQ annual market returns are plotted in figure 6, along with a rolling 2-σ band. We can see the mean return is positive; it equals 11.78%. The cumulative return plot for one dollar is illustrated by a representative index level and is obtained as $y_t = y_0 \prod_{i=1}^{t}(1+r_i)$, with $y_0 = 1$. The generic index level using realized returns is plotted in the lower panel of figure 6, based on a starting index level of 5 in 1926. The actual Dow Jones Industrial Average and S&P 500 (which only had 90 stocks until March 1957) had nominal index levels of 158.54 and 12.74 in 1926, respectively. A diagnostic normal plot of the market returns is provided in figure 7 from which we can again safely reject the normal hypothesis most popularly expounded by Roberts (1959).

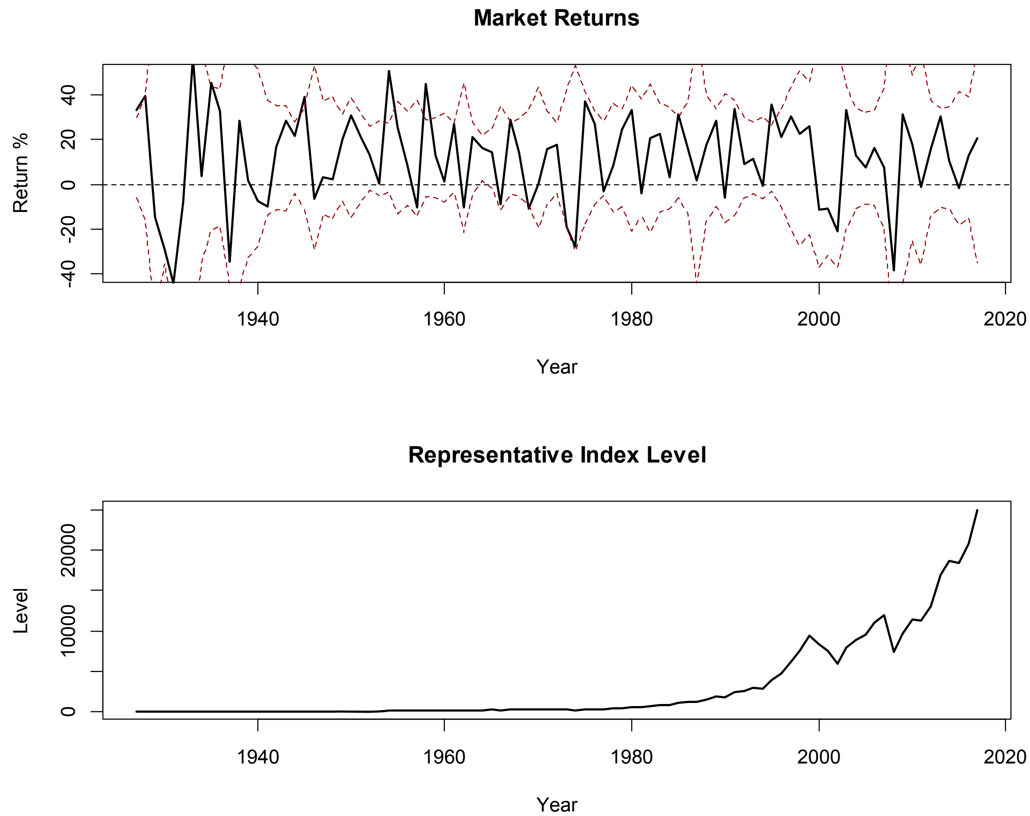

Figure 6: Market returns, 1926-2017 and representative index levels based on the market returns. Ending level is 25,044

Today (June 2018) there are 16,728 stocks traded in the U.S. and Canada on an exchange. Excluding ADR's (foreign) and OTC/BB stocks, the main U.S. exchanges see only 7,100 issues. The number of traded companies grew steadily since the 1920's, peaking at the end of the boom of the late 1990s. Both these totals are down considerably since the dot.com bubble burst in 2000. The total number of stocks is provided on figure 8. The evolution of company counts is a fascinating study, and many methods could be employed to model the birth and death process of a stock. Company counts and trading days per year are given in table 3 and figure 9 below.

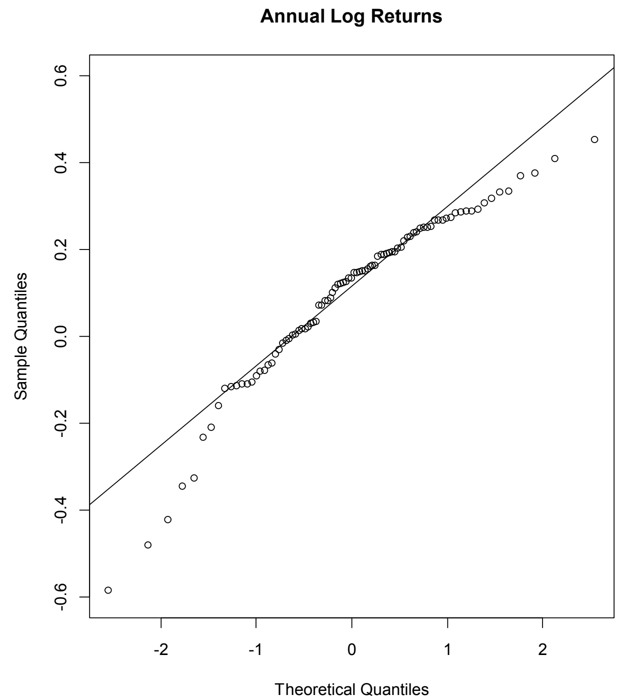

Figure 7: Normal diagnostic plot for market returns shown in figure 6

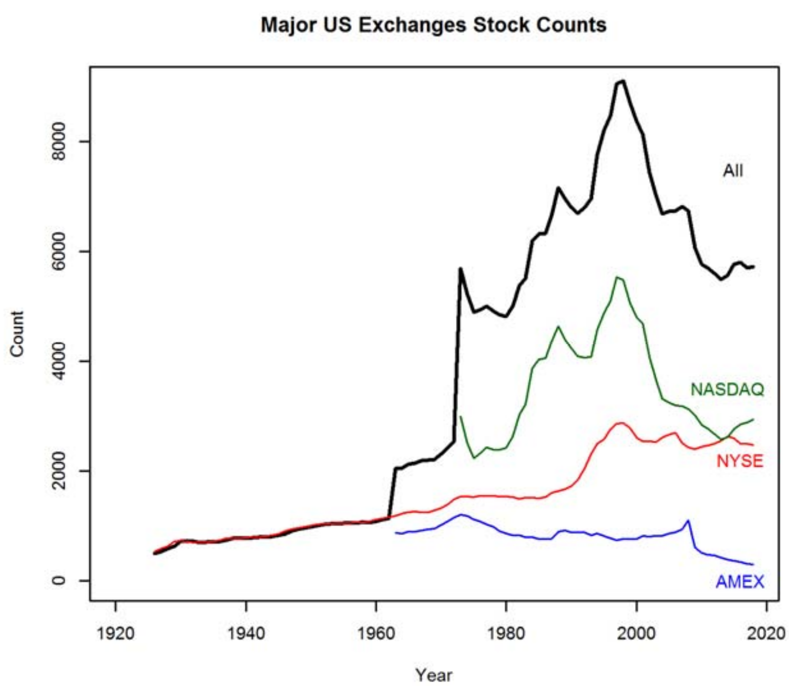

Figure 8: Major U.S. exchanges stock counts, 1926-2017

Table 3: U.S. Exchange-traded stocks as of 6/30/18

Exchange	Count	Percent
Non-traded security	26	0.16
Toronto TSX	1,615	9.65
Canadian Venture	1,405	8.40
NYSE	2,551	15.25
ASE	294	1.76
OTC/BB	10	0.06
Nasdaq	3,044	18.20
PSE	1,380	8.25
Other-OTC	6,132	36.66
Unlisted, evaluated	1	0.01
Not otherwise classified	270	1.61
	16,728	100.00

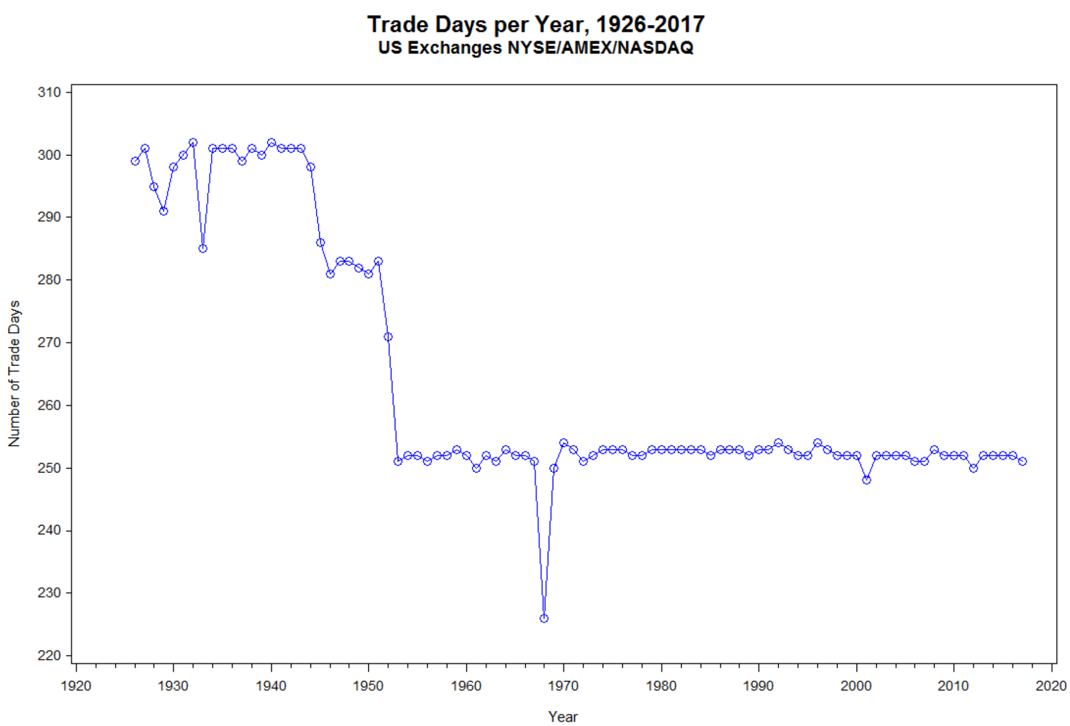

Figure 9: Active trade days on major US exchanges. Anomaly in the modern era was due to the NYSE 6-month closure on Wednesdays to deal with the infamous "paperwork crisis" of 1968. (Traflet 2007)

All company survival times on the major exchanges are given in figure 10. Although the shape of the curve suggests an exponential lifetime, upon closer inspection we see that the sample moments do not match very well, as they should with $X \sim \text{expo}(11.06)$ and

$\hat{\mu}_X = \hat{\beta} = 11.06 = \hat{\sigma}_X$, and we see a discrepancy with very short survival times. A simple gamma fit of X~gamma(.832, 12.28) using method of moment estimators (reasonable given the sample size) provides the appropriate moment values $\hat{\mu}_X = \hat{\alpha}\hat{\beta} = 11.057$ and $\hat{\sigma}_X = \sqrt{\hat{\alpha}\hat{\beta}^2} = 12.12$.

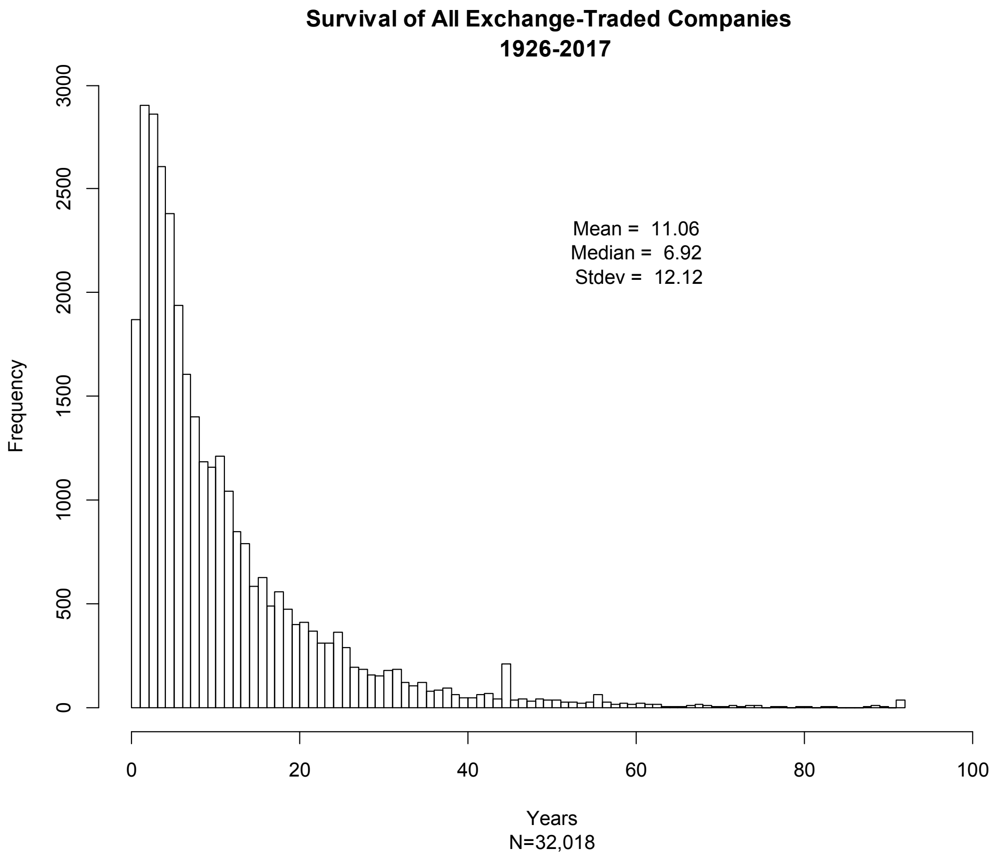

Figure 10: Survival time (in years) of all U.S. exchange traded companies, 1926-2017

2. The Simugram™

Around the turn of the century, Thompson directed his efforts to a serious non-linear optimization computational effort he dubbed the "Simugram" system, which he filed for patent protection in 2003 and which was subsequently granted in 2004. It is an attempt to find holding weights for a given universe of stocks which will provide a portfolio exhibiting the desired profit and risk characteristics using non-parametric resampling of historical returns and non-linear optimization constraints.

Some background in portfolio notation is necessary. A *sized* portfolio consisting of 2 stocks, say {IBM, CAT} is denoted $\Pi = \{N_1 \cdot \text{IBM}, N_2 \cdot \text{CAT}\}$. Assuming we have an amount *I* to invest, position sizes (in numeraire) come from the weights *w* and the trading element prices as $P = I \cdot w = (P_1, P_2)$. With *I*=\$50,000 and *w*=(.5,.5) we have *P*=(\$25,000, \$25,000). If (IBM, CAT) = (\$144, \$154) we find *N*=(173.61, 162.34). Our actual portfolio

$\Pi = (174\,\text{IBM}, 162\,\text{CAT})$ with value extension $\Pi_\$ = (\$25000, \$25000)$. The *value* of the portfolio is thus the sum of the individual positions $P_0 = 1^T \Pi_\$ = \$50{,}000$. Letting the return of each trading element at the end of a period be $r_t = \frac{X_t}{X_{t-1}} - 1$, then the return of the portfolio is found to be $r_p = w^T r$, with standard deviation $\sigma_p = \left(w^T \Sigma w\right)^{1/2}$.

The essence of portfolio construction is finding weights on the universe which result in an acceptable portfolio return (or standard deviation or other risk measures). For example, if one knew stock X_j would perform the best over the coming forward return period, weights of (0,0,0,0,…1,0,0,…0) would guarantee the best return. This of course is cheating. Other weighting schemes might include:

a. Equal weights. $w = \frac{1}{n} \cdot 1_n$ giving $r_p = w^T r = \frac{1}{n}\sum_{i=1}^{n} r_i = \overline{r}$ and

$$\sigma_p^2 = w^T \Sigma w = \frac{1}{n^2} \sum_{i,j} \sigma_{ij}$$

b. Market capitalization weights. $w_i = \frac{MC_i}{\Sigma MC}$

c. Random weights. w=(.3245, .212, …)

The seminal Markowitz (1952) optimization solution optimizes for a target return, a target "risk" (standard deviation), or uses a risk-tolerance utility function. Neither of these approaches are suited to real-world investors. The expected value or other scalar measures of a portfolio are not very useful either (Thompson, 2003, p.33). Instead, based on the simulated risk profile, the empirical cumulative distribution function (ECDF, or "Simugram") of future terminal values of portfolios formed at the end of the investment horizon, we might maximize the lower 20th percentile of return subject to a loss constraint. This is illustrated in figure 11 corresponding to an investment of \$100,000 with a 20-year terminal value of \$417,000.

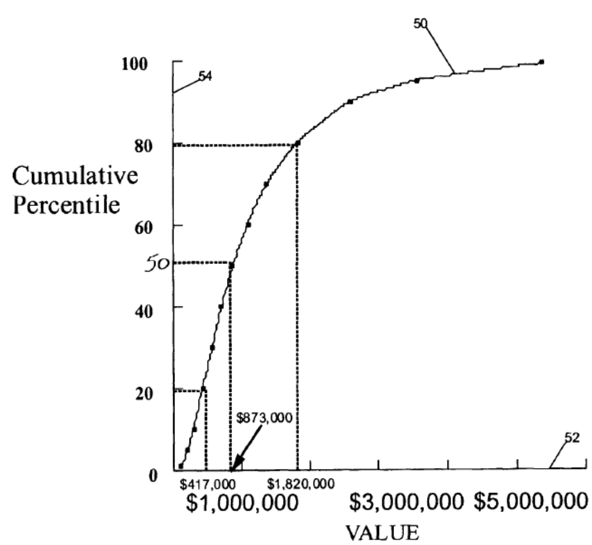

Figure 11: Simugram percentile optimization, taken from U.S. Patent No. 7,720,738

In the Simugram approach the decision variable is the portfolio weights w and the problem is stated as:

$$
\begin{aligned}
&\text{Max } R_{T_C}(w, X, \psi) \\
&\text{s.t.} \\
&\psi_1 : w_T 1_k = \Sigma w = c \\
&\psi_2 : w_i \geq L_w = 0 \\
&\psi_3 : w_i \leq U_w \\
&\psi_{nlc} : \; R_{T_C} \geq R^*
\end{aligned}
\qquad (2)
$$

Where R_{T_C} is the cut-off return (figure 12) with specified minimum, *w* are the weights, *X* are the securities' returns, and ψ_i are members of a constraint set ψ.

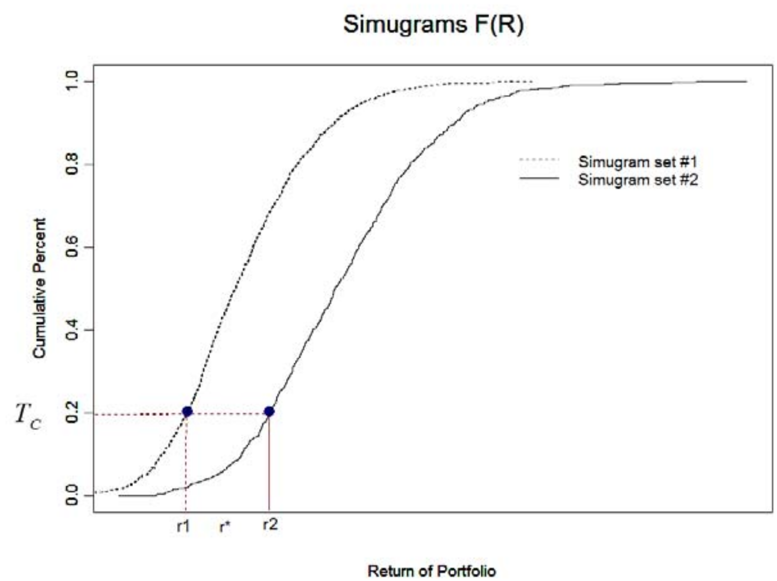

Figure 12: Stochastically dominating simugrams: $F_2 \leq_{st} F_1$

Simugram™ returns are impressive. On the S&P 100 universe, 32-years CAGR's in excess of 15% are typical. Figure 13 shows the outperformance of the Simugram™ system compared with the S&P 500 benchmark for the period 1970-2002. Subsequent work by other researchers have validated these results, and although the 15% annualized returns have been reduced in recent years, the system still provides equivalent returns to those of the equal-weighted total return indexes, with a *manageable* portfolio size, say 20 stocks.

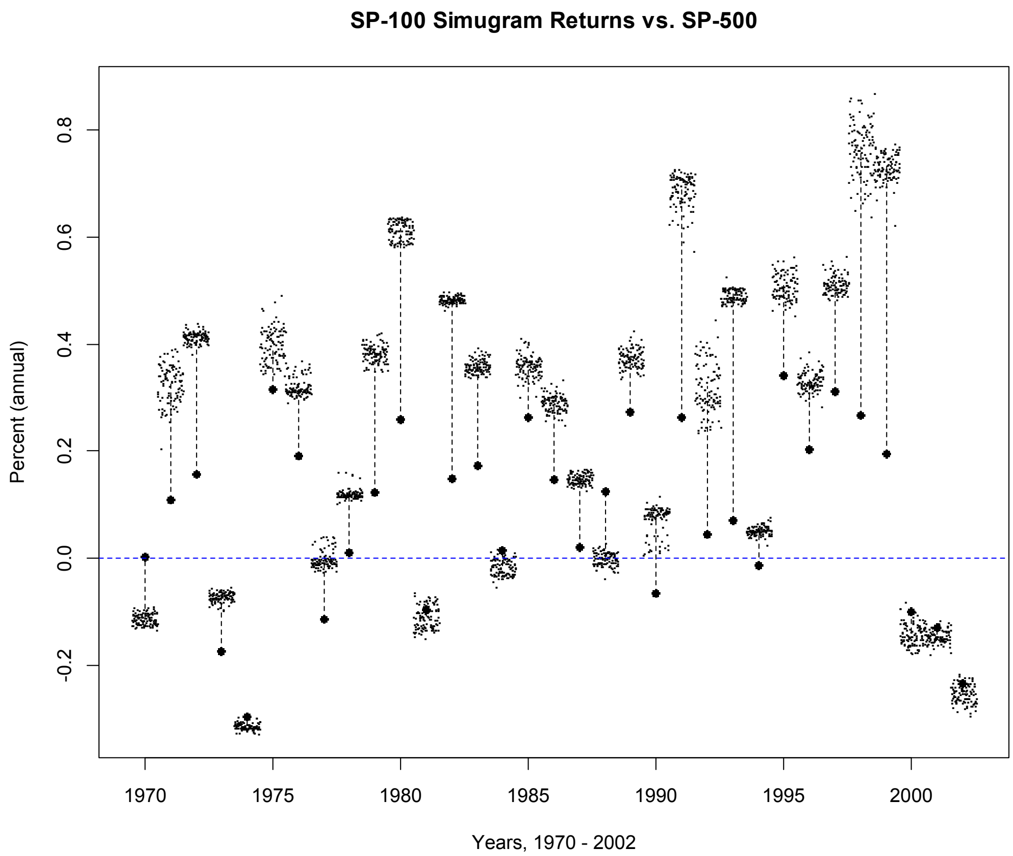

Figure 13: Simugram outperformance over the S&P 500 for the period 1970-2002. Solid dots are S&P return for the year, and light dots are various Simugram trials for that year

3. Capital Asset Pricing Model (CAPM) Validation

In the mid-1990s, Professor Thompson and Jones Graduate School Professor Edward E. Williams began collaborating on entrepreneurship and financial modeling. Around the turn of the 21st century, they focused on critical statistical analyses of the Efficient Market Hypothesis (EMH). Their first joint paper was a critique of the assumptions and conclusions inherent in the famous Black-Scholes-Merton option pricing formula. (Thompson and Williams, 1999). Professor Williams asked Dr. Thompson to take a quantitative look at the Capital Asset Pricing Model (CAPM).

During the 1960s, Treynor (1961), Sharpe (1964), Lintner (1965), and Mossin (1966) proposed economic models that purported to demonstrate how financial markets operate. This early work became known as the CAPM. The models had their foundation with Markowitz (1952) since he had used expected return, variance, and covariance notions to build his theory of portfolio analysis. Since he was dealing with the future, the forecasting task for his theory was formidable. This problem was dispatched by the development of the EMH since past returns and risk calculations could be substituted for future ones. Many assumptions are required for analytic tractability. According to Professors Focardi and Fabozzi (2004, p.512), "The CAPM is an abstraction of the real world capital markets and, as such, is based on some assumptions. These assumptions simplify matters a great deal, and some of them may even seem unrealistic. However, these assumptions make the CAPM more tractable from a mathematical standpoint. The CAPM assumptions are as follows:

Assumption 1: Investors make investment decisions based on the expected return and variance of returns.
Assumption 2: Investors are rational and risk averse.
Assumption 3: Investors subscribe to the Markowitz method of portfolio diversification.
Assumption 4: Investors all invest for the same period of time.
Assumption 5: Investors have the same expectations about the expected return and variance of all assets.
Assumption 6: There is a risk-free asset and investors can borrow and lend any amount at the risk-free rate.
Assumption 7: Capital markets are completely competitive and frictionless."

However "unrealistic" these assumptions may be in real life, their use admits a beautiful theory. One of the key results of the CAPM is the so-called Capital Market Line (CML). According to Wojciechowski and Thompson (2006), "..the Capital Market Line contains all volatility and return pairs that correspond to a portfolio constructed from a linear combination of the market portfolio and a riskless instrument." Specifically, the return from a portfolio on the Capital Market Line is constructed as $r_p = \alpha r_f + (1-\alpha) r_M$, where r_M is the market portfolio return, r_f is the riskless instrument return, and $\alpha > 0$ is a constant representing the allocation between the market portfolio and the riskless instrument. Note, $\alpha > 1$ is interpreted as borrowing, or leveraging, to invest in the market portfolio."

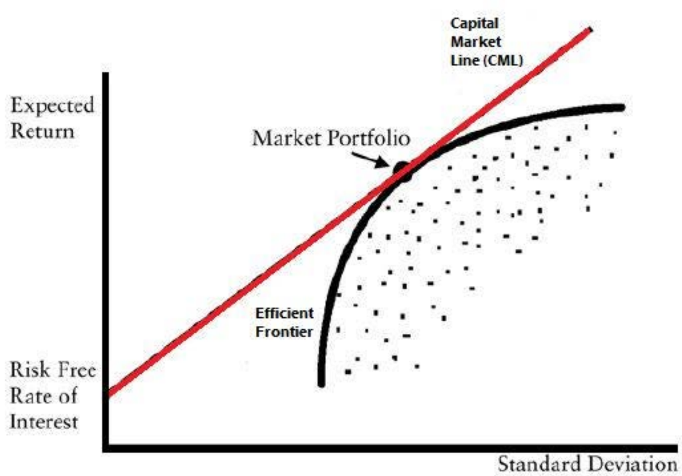

Figure 14: The Capital Market Line (CML)

If the return and the standard deviation of returns for the market portfolio are (σ_M, μ_M), then the CML is the line joining points $(0, r_f)$ and (σ_M, μ_M). The equation for the line is $\mu = r_f + ((\mu_M - r_f)/\sigma_M)/\sigma$. (One recognizes the slope of this line is related to the "market price of risk."). Portfolios with a return of μ and standard deviation of returns σ satisfying this equation lie on the CML. Wojciechowski and Thompson (2006) define performance and outperformance as follows:

> "If a random portfolio falls above the Capital Market Line, we say the random portfolio has outperformed. Otherwise, it has underperformed. We measure the amount of outperformance by the distance a point is above the Capital Market Line."

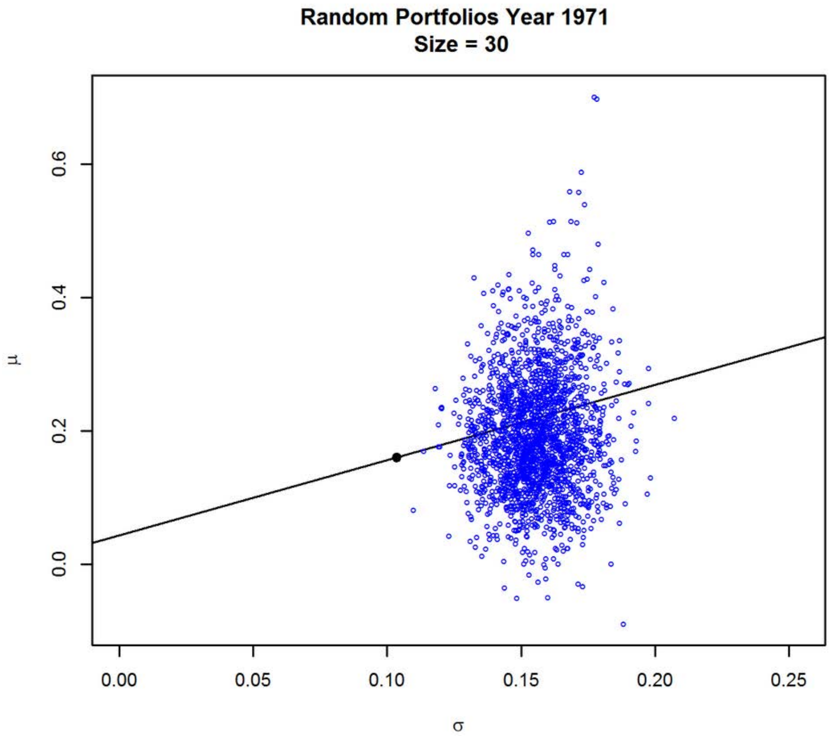

Figure 15: Annual returns and annualized volatility of each of the random portfolios generated for the year 1971. Open circles are the random portfolios, the black dot is the Market portfolio (S&P 500)

We recently performed this exercise and a representative year is presented in figure 15. One can eyeball that approximately 38% of the random portfolios lie above the CML, indicating outperformance compared with the passive market portfolio. A graphic representing the outperformance over the 48-year period 1970 through 2017 is seen in figure 16. As the portfolio size increases, the dispersion of the portfolios is reduced but the center of the clouds remains the same, and the percent of under- and out-performance tends to gather at the 100% and 0% levels.

Thompson and Wojciechowski conclude:

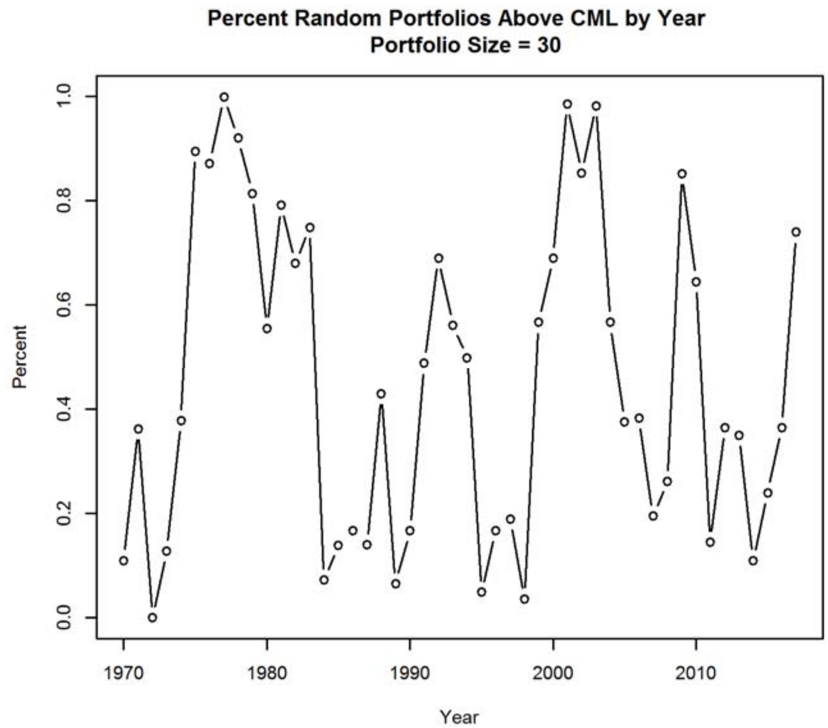

Figure 16: The percentage of random portfolios that outperform the market portfolio by year. It is clear from random selection that there are many opportunities for a competent portfolio manager to do better than the market portfolio.

> "This is hardly what we would expect if passive investing were truly optimal. If random selection can exhibit such behavior, then a competent, professional portfolio manager should be able to do as well as, if not better than the random selection. Again, we would like to make it clear that we are not promoting random selection as a valid investment strategy. The proclamation here is that if random selection can do better than a market cap weighted portfolio, then a competent active portfolio manager should also be able to outperform the market cap portfolio."

4. "Everyman's" MaxMedian Rule for Portfolio Management"

The Simugram^{TM} is computationally intensive. In fact, Thompson claimed it is an NP-hard problem. He obtained workable results using the Nelder-Mead direct search approach, followed later by the Torczon method. For years, he had a close working relationship with the U.S. Army Research Office, and several colleagues there asked him to devise a simple investing program that almost anyone with some basic computational skills could implement. Thus the MaxMedian was born in 2005, and presented at the U.S. Army Conference on Applied Statistics (Thompson, 2007). Since Professor Thompson was a historiographic generalist and aficionado in addition to his other loves, he dubbed it "Everyman's[1] MaxMedian Rule." It is a non-proprietary rule to enhance the return to ordinary investors not unwilling to do a little data crunching, but not requiring parallel processing and proficiency in programming languages such as C. Moreover, MaxMedian

[1] The *Somonyng (Summoning) of Everyman*, John Skot, c.1530

portfolio results meet or exceed the equal-weighted S&P 500 benchmark performance with only 20 stocks. This is very important, since the equal-weighted portfolio was impossible to invest in (at the time of the MaxMedian publication), and obtaining these results with a portfolio of only 20 stocks is most attractive.

The “classical” MaxMedian rule is as follows:

1. Collect the previous year's daily returns r(i,j) for all stocks in the S&P 500 at the time of portfolio formation.
2. Each year, calculate the 500 yearly median values for r(j,t)
3. Invest equally in the 20 stocks with the highest median returns.
4. Hold for one year (and one day), then liquidate.

Unquestionably, one could choose other statistics than the median for ranking purposes, such as mean, various percentiles, etc. Dozens of graduate and undergraduate students in Rice University’s Market Models class, and in independent research, have studied various permutations to the rule, such as different lookback and holding periods, other rebalancing rules, cash investment and constituent dropout programs, different weighting, etc. Academic researchers have also validated and extended the concept, such as Affinito (2009), Ernst and Miao (2017), and Tooth (2012). They have all studied the Rule's efficacy over different study periods and have found the phenomenon persists. Table 4 provides typical results.

Table 4: MaxMedian vs. Benchmark Returns, 1958-2014

Dividends Excluded[1]

Measure	Equal Weight	Market Weight	MaxMedian
Terminal value	276.0	52.1	342.9
57-Year CAGR	10.4%	7.2%	10.8%

1. Source: Ernst, Thompson and Miao (2017)

To date, the researcher who studied this most extensively was Sarah M. Tooth (1988-2014). In her master's thesis (2012) she looked at what she coined the *MaxMeasures* family of ranking strategies. Besides considering ranking based on percentiles other than the 50th (median), she also looked at generalized, or power means (Tooth and Dobelman, 2016). These are defined for $x_i > 0 \, \forall i$ and $p \in \mathbb{R}$, $M_p(x) = \left(\frac{1}{N}\sum_1^N x_i^p\right)^{1/p}$. The arithmetic mean is a special case with *p=1*. Useful limiting cases are the geometric mean and the max and min:

$$\text{(3)} \qquad M_p(x) = \begin{cases} \left(\frac{1}{N}\sum_1^N x_i^p\right)^{1/p} & p \in \mathbb{R} \\ \left(\prod_i^N x_i\right)^{1/N} & p \to 0 \\ \max x & p \to \infty \\ \min x & p \to -\infty \end{cases}$$

Recall the market index benchmarks reproduced in table 5 below. The calculated benchmarks in Tooth (2012) are very similar, as seen in table 6.

Table 5: Compound annual growth rate (CAGR) of select major U.S. Stock Indexes. VW and EW are correspond to market-value weighted and equal-weighted, respectively. Dividends included or excluded as indicated by D or X. Source: The Center for Research in Security Prices (CRSP).

Current Date:	12/31/17			CAGR		
Index	Begin Date	N	VWRETD	VWRETX	EWRETD	EWRETX
SP500	12/31/1925	92	0.100	0.059	0.118	0.078
NYSE	12/31/1925	92	0.098	0.057	0.123	0.081
AMEX	12/31/1962	55	0.077	0.053	0.126	0.103
NASDAQ	12/31/1972	45	0.106	0.089	0.131	0.116

Table 6: Selected U.S. Index benchmarks for 1970-2011 and associated decade CAGR's (Tooth 2012)

	Overall	1970s	1980s	1990s	2000s
With Dividends					
Value Weighted S&P500	9.90	5.83	17.60	18.36	-0.69
Equal Weighted S&P500	12.40	8.91	19.97	15.54	6.02
Mean Return	12.66	10.18	18.70	16.02	6.53
Without Dividends					
Value Weighted S&P500	6.49	1.57	12.54	15.49	-2.51
Equal Weighted S&P500	9.10	4.50	15.41	12.88	4.20

MaxMeasures results are reproduced in figure 7. Note that the 50th percentile represents the "classic" MaxMedian methodology. We observe Professor Thompson's inspired intuition in the use of the median ranking statistic. The 25th percentile has equivalent return and a volatility close to that of the indexes, but the median's 24% volatility is not objectionable. Regarding power means, recall that $p=-1$ corresponds to the harmonic mean, $p=0$ the geometric mean, and $p=1$ the arithmetic mean. The harmonic mean was chosen as a canonical ranking statistic for the MaxMeasures results for its understabability, simplicity, improved CAGR of 14.4% and tolerable volatility. Is it worth it to use MaxMeasures in place of MaxMedian? The CAGR difference of 14.4% vs. the S&P 500's 9.9% over 40 years is a terminal value factor of 591%. One could fund five vs. a single grandchild.

Table 7: MaxMeasures results, 1970-2011, for selected percentile and power means rankings

	CAGR	Mean	Med	σ
S&P500	9.90	11.43	14.90	17.72
Percentile				
0.05	12.33	13.47	14.82	15.82
0.25	12.58	13.80	14.97	16.12
0.50	12.67	15.50	18.21	23.61
0.75	10.17	16.41	18.88	35.40
0.95	11.54	18.42	17.50	40.53
Power				
-10.0	14.63	16.53	19.69	20.54
-1.0	14.44	17.88	18.65	27.83
0.0	13.43	17.29	21.12	29.06
1.0	12.99	16.92	21.32	28.73
10.0	11.94	18.67	18.99	39.66

The average and max percent drawdowns shown in table 8 are for the sample period 1970 thru 2011, including the total number of negative years (TNY) and maximum run of negative years (MNY). Here we see for the S&P 500 index, the maximum drawdown for the study period was 38%, which flushed many investors (short-term, we presume) out of the markets.

Table 8: MaxMeasures drawdown and loss summary

	Average	Max	Total	TNY	MNY
S&P500	-17.38	-38.44	-104.30	6	2
Percentile					
0.05	-12.59	-24.26	-75.56	6	2
0.25	-10.27	-29.06	-82.14	8	2
0.50	-18.38	-54.69	-183.78	10	3
0.75	-26.29	-63.96	-315.53	12	3
0.95	-23.30	-62.36	-302.96	13	3
Power					
-10	-9.94	-37.16	-99.45	10	3
-1	-15.55	-46.91	-186.57	12	3
0	-17.76	-51.89	-213.15	12	3
1	-19.65	-53.03	-235.77	12	3
10	-28.75	-62.62	-287.49	10	3

Although this does not comport with the EMH, exiting the market at the bottom of a drawdown period is usually the worst thing to do, in that one misses the characteristically large rebound which occurs in the recovery year. For example, at the bottom of the stock market in March 2009, the return one year later was 48%. Just holding would result in a 25% increase for the period.

We may notice that MaxMedian has a max drawdown of 54%, which is painful for fund investors. Higher percentile ranking methods have worse drawdowns than the median, as do most of the power means except for -10. The harmonic mean (*p=-1*) has substantially smaller drawdowns that the MaxMedian. The 5th percentile ranking strategy has almost the same CAGR as the median, but with substantially smaller drawdowns. The truly conservative investor might choose this MaxMeasures approach if she were comparing her strategy to the of holding the (now-available) equal weighted index. In any case we recall that MaxMedian generates returns equivalent to the equal weighted indexes but with only 20 stocks. After a little mode searching, Ms. Tooth's ending recommendation was for using a lookback period of 5 months and *p=-.6*, but for simplicity, publicly advocated a one-year lookback period with the harmonic mean.

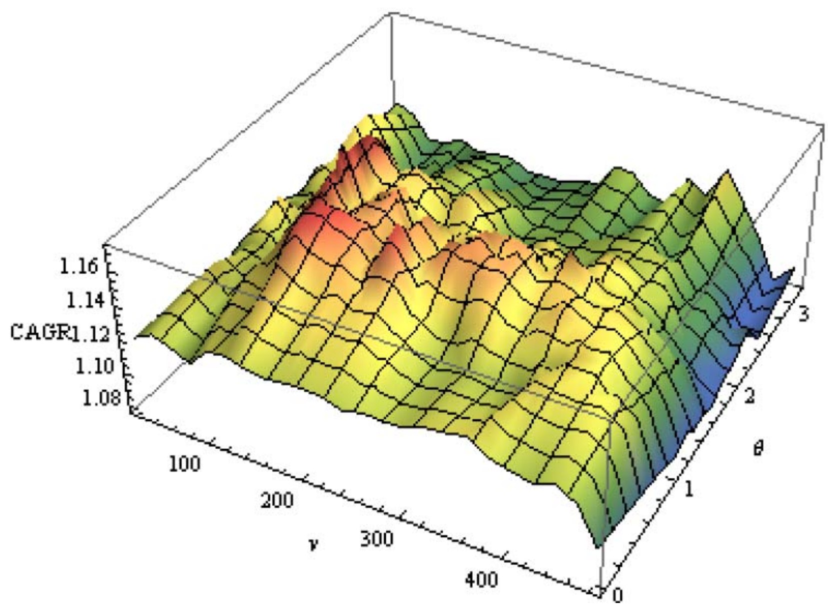

Figure 17: Returns obtained by varying lookback period in days and power means p (reparameterized as θ)

5. Tukey Transformational Ladder

For investors, there are lots of alternatives to equal or market value weights. The gamut runs from random to sophisticated computationally derived, a la the Simugram™. Another approach advocated by Professor Thompson was weights derived through Tukey's "transformational ladder of powers," investigated in Ernst, Thompson and Miao (ETM, 2017). This general method of linearization first proposed by Tukey is seen as $T(x) = x^p$, augmented by ln(x). This results in the range of transforms *T* as follows:

$$1/x^2 \quad 1/x \quad 1/\sqrt{x} \quad \ln(x) \quad 1/n \quad \sqrt{x} \quad x \quad x^2 .$$

Let X be the market capitalization of a stock (a common "size" measure) as the price times the number of shares outstanding. Then the Tukey weights for each of *N* stocks are

(4) $$w_i = \frac{T(x_i)}{\Sigma T(x_i)} \text{ for } i=1,2,\ldots N$$

Consider an example, for three stocks with market capitalizations (in billions of dollars) of (400, 100, 10), we have Tukey weights as:

$1/x^2$	$w = (0.000618, 0.009895, 0.989487)$	
$1/x$	$w = (0.022222, 0.088889, 0.888889)$	
$1/\sqrt{x}$	$w = (0.107244, 0.214487, 0.678269)$	
$\ln(x)$	$w = (0.464483, 0.357012, 0.178506)$	
$1/n$	$w = (0.333333, 0.333333, 0.333333)$	*EW*
$\sqrt{x}$	$w = (0.603095, 0.301547, 0.095358)$	
x	$w = (0.784314, 0.196078, 0.019608)$	*MW*
x^2	$w = (0.940623, 0.058789, 0.000588)$	

Note for positive powers *p* the weights favor large capitalization stocks, and vice versa with negative *p*. Powers $p=1$ and $p=0$ correspond to the market and equal weights, respectively.

The results for the Tukey weights backtests based on quarterly rebalancing are presented in tables 9 below. We note the monotonicity in *p* for the CAGR and volatility. Some of the volatilities are indeed high, although they are similar to other quantitative portfolios such as the MaxMeasures results presented above. The ETM analysis includes both dividends AND frictions such as commissions and slippage.

Table 9a: CAGR for Tukey weight portfolios, 1958 through 2015

$1/x^2$	$1/x$	$1/\sqrt{x}$	$\log(x)$	EQU	$\sqrt{x}$	MKC	x^2
18.00%	17.53%	15.23%	13.80%	13.32%	11.73%	10.43%	8.69%

Table 9b: Volatility for Tukey weight portfolios, 1958 through 2015

$1/x^2$	$1/x$	$1/\sqrt{x}$	$\log(x)$	EQU	$\sqrt{x}$	MKC	x^2
39.54%	26.44%	22.29%	20.01%	19.30%	17.52%	16.98%	18.05%

Table 9c: Sharpe Ratios for Tukey weight portfolios, 1958 through 2015

$1/x^2$	$1/x$	$1/\sqrt{x}$	$\log(x)$	EQU	$\sqrt{x}$	MKC	x^2
56.07%	70.35%	70.21%	69.31%	68.81%	65.24%	59.25%	47.09%

It can be seen from table 9 that the cumulative values of the portfolios exactly follow the ordering inherent in the Tukey ladder. ETM (2017) are believed to be the first to have discovered this behavior. A graphical representation of the time series is presented in figure 18 below. Their investigation concluded on a practical basis that monthly rebalancing was optimal, owing to the terminal values for the daily, monthly, quarterly and annual options. It should be noted that daily rebalancing results in *ruin* because of transaction costs, which are usually ignored for simplicity's sake in studies such as this. Annual rebalancing, widely used in backtesting scenarios, is suboptimal, paying off 32% less than quarterly rebalancing.

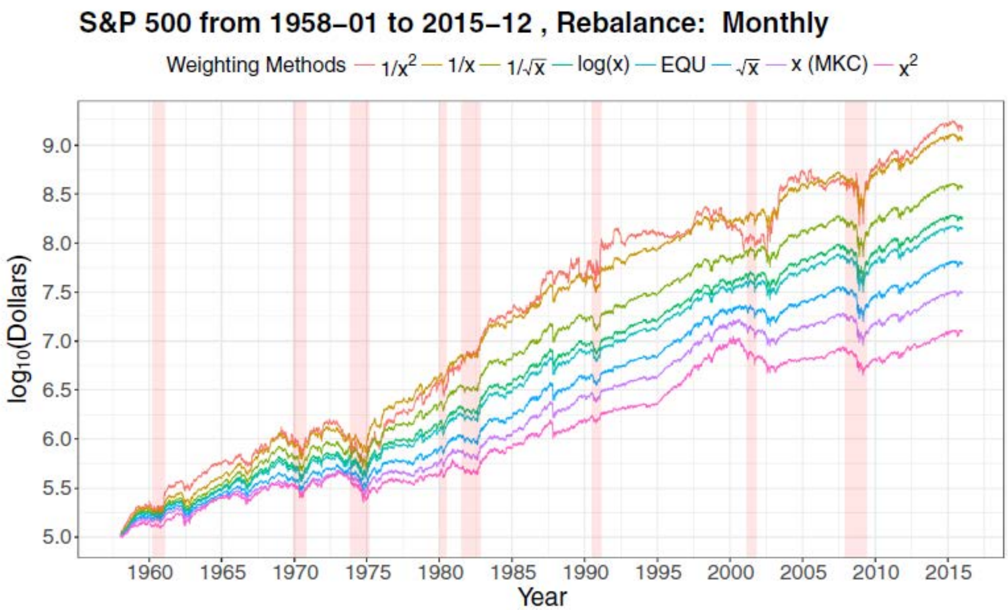

Figure 18: Cumulative $\log_{10}$ returns for 1958 to 2015 for the eight Tukey transformational ladder portfolios. This calculation includes dividends and frictions and assumes that $100,000 is invested on 1/2/58 and left to grow until 12/31/15. The major avowed recessions are depicted in the shaded zones

The terminal values at the end of this 58-year study period beginning with an investment of one dollar are:

$$
\begin{array}{lll}
1/x^2 & TV = 14,762 & \\
1/x & TV = 11,712 & \\
1/\sqrt{x} & TV = 3,722 & \\
\ln(x) & TV = 1,804 & \\
1/n & TV = 1,412 & EW \\
\sqrt{x} & TV = 622 & \\
x & TV = 315 & MW \\
x^2 & TV = 125 &
\end{array}
$$

This phenomenon is currently being actively investigated, both for additional validation, and for possible theoretical explanation. Regarding this last point, ETM observe that "our empirical results neither contradict nor support the small-firm effect hypothesis and therefore results concerning Tukey's transformational ladder for portfolio management must be viewed as their own distinct phenomena."

6. Conclusion

James Thompson is best known for his work in simulation, statistical process (quality) control, cancer and AIDS modeling, and other statistical methodology. He is the progenitor of 43 statistical descendants according to the American Mathematical Society and North Dakota State University's Mathematics Genealogy Project. He was also tireless in quantitative finance and was responsible for establishing the Statistics Department's computational finance specialization, which we are continuing today. Many of his approaches, while conceptually simple, are seen to be valid in the real world. As Warren Buffet has stated in various letters to Berkshire Hathaway shareholders,

> "If you need to use a computer or a calculator to make the calculation, you shouldn't buy [the company]…There's this holy writ, the efficient market theory. How do you teach your students everything is priced properly? What do you do for the rest of the hour? ... The more symbols they could work into their writing the more they were revered..."

Jim was a practitioner and put his money where is mouth was. Consequently. his precious widow is survived with an uncommonly healthy nest egg. May we all be so courageous, patient and wise.

Acknowledgements

The author expresses his sincere appreciation to Professor Katherine B. Ensor for presenting an earlier version of this paper to a special session in honor of Jim Thompson at the 2018 Joint Statistical Meetings in Vancouver, BC.

References

Affinito, Ricardo. "*The Coordinated Max-Median Rule for Portfolio Selection*." OpenStax CNX. Web. 10/19/2009 http://cnx.org/content/m32523/1.1/

Ali, H. and D. Brodkey, et al., *Extensions on CAPM Validation and Exploration of Portfolio Growth*, STAT 486 Final Project, Rice University, Spring 2018.

Chakraborti, A., et al., "Econophysics Review: I. Empirical Facts," *Quant. Finance*, Vol 11, No. 7, July 2011, 991-1012.

Dobelman, John A., ed. Models and Reality: Festschrift for James Robert Thompson, Chicago: T&NO Company, 2017.

Ernst, P.A., Thompson, J.R., and Miao, Y. (2017) "Tukey's transformational ladder for portfolio management." *Financial Markets and Portfolio Management* 31(3): 317-355.

Ernst, P.A., Thompson, J.R., and Miao, Y. (2017). "Portfolio selection: the power of equal weight". In *Models and Reality: A Festschrift for James R. Thompson*, pp. 225-236.

Focardi, S.M. and F.J. Fabozzi, "*The Mathematics of Financial Modeling and Investment Management*," New Jersey: John Wiley & Sons, 2004.

Greg, Walter Wilson, ed. *Everyman from the edition by John Skot*. Louvain: Uystpruyst, 1904.

Lintner, J., “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *Review of Economics and Statistics*, February 1965, pp. 13-37.

Markowitz, H., "Portfolio Selection," Journal of Finance, March, 1952, 77- 91. (A).

Mossin, J., “Equilibrium in a Capital Asset Market,” *Econometrica*, October 1966, pp. 768-783.

Roberts, H., "Stock Price 'Patterns' and Financial Analysis: Methodological Suggestions," *Journal of Finance*, March 1959, pp. 1-10.

Sewell, M., *Characterization of Financial Time Series*, UCL Department of Computer Science Research Note RN/11/01, pp. 1-35.

Sharpe, W.F., “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk,” *Journal of Finance*, September 1964, pp. 425-442.

Thompson, J.R., Williams, E. E., and Findlay, M. C., *Models for Investors in Real World Markets*, John Wiley & Sons, New York, 2003.

Thompson, J.R., *Methods and Apparatus for Determining a Return Distribution for an Investment Portfolio*, U.S. Patent 7,720,738, Filed 1/3/2003, and issued 7/8/2004.

Thompson, J.R. and Williams, E. E., “A Post Keynesian Analysis of the Black-Scholes Option Pricing Model,” *The Journal of Post Keynesian Economics*, Winter, 1999, pp. 251-267.

Thompson, J.R. and L. Scott Baggett, “Everyman's MaxMedian Rule for Portfolio Management,” *The Proceedings of the U.S. Army Conference on Applied Statistics* (2007) http://www.armyconference.org/

Thompson, J.R., *The Little Formula That Beats The Market*." National Institute of Statistical Sciences, Durham, North Carolina. (October 2006)

Thompson, J.R., Empirical Model Building: Data, Models, and Reality. Wiley, 2nd edition, 2011, p.356

Tooth, Sarah M. “*Design and Validation of Ranking Statistical Families for Momentum-Based Portfolio Selection*.” Master's Thesis. Rice University, 2012 http://hdl.handle.net/1911/71697

Tooth, S.M. and Dobelman, J. (2016) A New Look at Generalized Means. *Applied Mathematics*, 7, 468-472.

Traflet, J. and M.P. Coyne, "Ending a NYSE Tradition: The 1975 Unraveling of Broker's Fixed Commissions and its Long Term Impact on Financial Advertising", *Essays in Economic and Business History*, Volume 25, 2007, p.131-141.

Treynor, J., “*Towards a Theory of Market Value of Risky Assets*,” originally an unpublished manuscript (1961) but published recently in *Asset Pricing and Portfolio Performance*, edited by Robert A. Korajczyk (London: Risk Publication, 1999), pp.15-22.

Tukey, J.W., “The Future of Data Analysis,” *Ann. Math. Statist*., Vol. 33, No. 1 (1962), p. 1-67.

Tukey, J.W., “*Exploratory Data Analysis, Limited Preliminary Edition*,” Addison-Wesley Publishing Co., Reading, MA, 1970.

Wojciechowski, W.C. and James R. Thompson, “Market Truths: Theory Versus Empirical Simulations,” *Journal of Statistical Computation and Simulation*, Vol. 76, No. 5, May 2006, 385-395.