

Young and Graff – 20 Years Later
An accessible and practical implementation method for heavy tail modeling

Roger J. Brown*
University of San Diego
PO Box 1146
Alpine CA 91903
619-445-3405
rjb21@cox.net

Beate Klingenberg
Marist College
c/o LdM
via del Melarancio 6/r
50123 Firenze, Italy
Beate.Klingenberg@marist.edu

*corresponding author

Abstract:

Young and Graff (1995) postulated that real estate is not normal. Since that time nearly two decades ago two phenomena have been observed. One is considerable evidence to support their thesis; the other is that, notwithstanding theoretical and empirical foundation, the use of heavy tails in modeling has not been widely adopted in the field. The reason for the latter could be traced to the somewhat daunting mathematics involved in representing non-normal data and estimating parameters under stable laws. Although much of that obstacle has been overcome with symbolic logic software offering numerically intensive methodology, the absence of an application built on a popular spreadsheet platform has been problematic.

This paper will present a simple, Excel-based workbook in a familiar discounted cash flow setting that incorporates heavy tail modeling. Included are a stable pseudo random number generator, the ability to vary the four stable parameters, robust simulation requiring no additional add-ins or custom programming, and a reconciliation of the bias inherent in simulating an IRR due to Jensen's Inequality. The goal is to provide practitioners with a familiar and flexible tool that allows explicit incorporation of a broad range of uncertainty into the cash flow model.

Keywords: Heavy tails; modeling risk; spreadsheet methods; stable distributions

Introduction

It has been said that if Levy-stable laws are found to be at the source of return generation that the world is a fundamentally riskier place than we once thought. Recent high profile news events involving the international financial system lend credence to this notion. The most recent crisis found real estate at its center. If it is time to embrace non-normal events as more frequent than previously thought, modeling distributions involving extreme values becomes key.

Markowitz (1959) doubled the number of parameters used in choosing assets by explicitly modeling risk as variance. His approach has been finance orthodoxy for more than half a century. But advances in mathematics and computer modeling have eclipsed the requirement to use a simple, static model that crucially depends on distributions having a finite variance. The use of variance as a risk measure has been under fire for some years.¹ Rather than retread that debate, it seems more useful to advance a method which permits the user to decide whether to include extreme events in the model and how extreme to make those events. Stable laws offer the opportunity to increase the number of risk parameters from one to three by explicitly adding two parameters describing tail behavior.²

Stable distributions are fully characterized by four parameters: $\alpha \in (0,2]$, is known as the *characteristic exponent* or *index of stability* and defines the thickness of the tails and corresponding “peakedness” of the center of the distribution. When $\alpha = 2$ the distribution is normal; $\beta \in [-1,1]$, the *skewness parameter*, is symmetric around the mode when $\beta = 0$; $\gamma > 0$, is a *scale parameter* expanding and contracting the distribution around its mode; and $\delta \in (-\infty,\infty)$, is the *location parameter*, shifting the distribution left and right along the x-axis. The last two are shape preserving under linear transformation allowing the distribution to be normalized in the usual fashion. Thus, our main interest is in how thick the tails are (α) and to what degree the distribution is skewed (β).

One might reasonably ask why not use the Student t distribution which has tails that can be modeled with varying thickness. Also, since the third and fourth moments always can be measured for finite samples, what is it about Levy-stable laws which are so special? To the first, an immediate response is the Student t is symmetric. We have a legitimate and genuine interest in how returns might be skewed and Student t does not inform in this area. The most important value of Levy-stable laws has to do with the important property of additivity. Classic economic theory postulates that prices move with information. If information arrives randomly it is important that the random variable represented by those arrivals may be incorporated in a shape-preserving manner. Levy-stable laws are shape preserving.³ Moreover, just as the normal finds strong foundation in the Central Limit Theorem, Levy-stable laws appeal to the Generalized Central Limit Theorem which, in essence, holds that if a distribution has a limiting distribution that distribution must be a member of the stable class. Hence

¹ Other risk measures have been proposed. Cf. Brown and Young (2011)

² Stable laws have a scale parameter similar to but not the same as standard deviation, constituting the third.

³ Provided the value of α does not change. This is an important and controversial restriction which goes beyond the scope of this effort.

Levy-stable may be viewed as a rich family of distributions that accommodate addition of random variables in the form of information and include, as a special case, the normal distribution.

Useful as Levy-stable distributions might be theoretically, they are anything but user friendly. Other than the normal and two others which are not applicable to finance, there are no closed-form solutions to describe the probability density function. In the limit, no moments after the first exist (hence no variance, no co-variance, no efficient set portfolios, no skewness and no kurtosis in the classic sense). The saving grace is the fact that while the density function may not exist, the characteristic function does. Here, in all of their resplendent glory, are two parameterizations of that function:⁴

$S_1(\alpha, \beta, \gamma, \delta) = \begin{cases} \exp \left\{ i u \delta - \gamma^\alpha u ^\alpha \left[1 + i \beta (u \gamma ^{1-\alpha} - 1) \operatorname{sgn}(u) \tan \left(\frac{1}{2} \pi \alpha \right) \right] \right\} & \text{for } \alpha \neq 1 \\ \exp \left\{ i u \delta - \frac{\gamma u \left[\pi + 2 i \beta \ln (u \gamma) \right] \operatorname{sgn}(u)}{\pi} \right\} & \text{for } \alpha = 1 \end{cases}$	Eq (1)
$S_2(\alpha, \beta, \gamma, \delta) = \begin{cases} \exp \left\{ i u \delta - \gamma^\alpha u ^\alpha \left[1 - i \beta \operatorname{sgn}(u) \tan \left(\frac{1}{2} \pi \alpha \right) \right] \right\} & \text{for } \alpha \neq 1 \\ \exp \left\{ i u \delta - \gamma u \left(1 + \frac{2 i \beta \ln [u] \operatorname{sgn}(u)}{\pi} \right) \right\} & \text{for } \alpha = 1 \end{cases}$	Eq (2)

Most will agree that Equations (1) and (2) are anything but appetizing.⁵

So, what is the conscientious practitioner to do? He realizes

- (a) that modeling in the restricted air of the normal distribution is the act of placing blinders on his view, blocking out the occasional life-threatening disaster;
- (b) he also recognizes that merely projecting a fixed return to occur in an uncertain future is the act of playing God, something his client will likely question; and
- (c) real estate suffers from a paucity of data under the best of conditions, placing benefits of the law of large numbers out of reach if one requires data composed of field-gathered realizations.

The answer to the last of these has long been to conduct a simulation. Popular software packages and Excel Add-Ins offer this facility, virtually of them based on an assumption of normality. What is missing is a simple, Excel based workbook incorporating stable laws that requires no special programming or add-in packages. This paper is intended to serve as an operating manual for just such a workbook, named simply "ModelingRisk.xlsx".

Overview

Technology and inspiration for this work comes from diverse sources. Chambers, Mallow and Stuck (1976) showed us how to generate stable random numbers. Recent enhancements of Excel allow for named ranges that become the base for simulation exercises. Finally, a web site,

⁴ The multiple parameter problem is explained in Nolan (in Press). The $\alpha \neq 1$ issue is handled by restricting the value of α to slightly more than 1 in the model used here.

⁵ Even less appetizing is the necessity of an Inverse Fourier Transform to work with the density function.

ProbabilityManagement.org offers a suite of tools, using the concept of SIPMath™, where SIP stands for “Stochastic Information Packet”, which embed macros and simulation routines directly into an Excel workbook. The artful combination of these lead to ModelingRisk.xlsx. It can be downloaded on the homepage of www.mathestate.com under Site News and Updates at date September 1, 2013.

The workbook is composed of several worksheets (Figure 1):



Figure 1

ReadMe worksheet provides a general description of the model. Note that the illustration here is only one of many possibilities. The tool is very robust. While the initial beta version models change in capitalization rate, the capable Excel user may change it to introduce uncertainty to another variable or multiple variables.⁶ Here the user can describe in words his own model and add the appropriate exculpatory language around limitations in its use, various interpretations and other descriptive information. Also on this sheet is the representation that the author is still improving it. Thus, while the general structure is not expected change, the version you download may be slightly different than the one described here.

Model worksheet is the heart of the tool. We will return to it at length later.

CDF worksheet displays a Cumulative Distribution Function. This permits you to provide a probability measure with your projection of financial gain. Note that each draw requires resorting column B. Automating this is an opportunity for the advanced Excel user.

StablePRNG worksheet provides Chambers, Mallow and Stuck’s algorithm for generating pseudo random numbers.

Uniforms worksheet provides the two uniform random variables that are at the base of the StablePRNG.

PMTable worksheet generates the values for the variables affected by uncertainty that are delivered to the model. A conventional graph of the probability density function is also shown on this sheet.

The last three sheets require no user involvement and are protected from accidental modification. There is no password required so the experienced Excel user wishing to make further modifications or enhancement may do so.⁷

⁶ Taking the appropriate care to consider interdependencies and multi-collinearity. An obvious example is the connection between rental rates and vacancy which clearly relate to one another.

⁷ Caution is urged with respect to the StablePRNG sheet as it includes cells with very detailed input that would not be easy to repair.

The Jensen problem

In essence, the model displays a purchase, five year holding period and sale of an investment property under assumptions of fixed percentage increases in income over time. The usual performance metrics, Net Present Value and Internal Rate of Return, are computed. However, an important distinction between these two must be made in the context of simulation. Brown (2006) describes the problem associated with simulation of the IRR arising from Jensen's Inequality⁸. Simply, since IRR is a non-linear function of cash flow, taking its expectation is at least bad form if not just plain bad mathematics. Also, as the range of uncertainty increases, the precise condition when extreme values populate the distribution, the gap between the expectation of the function and the function of the expectation rises. That gap should be disclosed and ModelingRisk.xlsx makes the calculation attendant to that disclosure. Of course, since Net Present Value *is* a linear function of cash flow it does not suffer from similar distortion.

The model

We will take the worksheet containing the model in small bites.

Beginning with the most familiar, below is a portion of **Model** worksheet which "hides" columns D-H so that you see only the familiar operating statement information that leads to cash flow.⁹ Only the Acquisition and Disposition years are shown in Figure 2.

	A	B	C	I
1	Pre-Tax Real Estate Yield Model			
2	(See Notes below)			
3			Acquisition	Disposition
4	Value	100.00%	\$950,000.00	\$ 1,615,501.76
5	Loan	65.00%	(\$617,500.00)	(\$570,944.23)
6	Equity	35.00%	\$332,500.00	\$ 1,044,557.53
7				
8				
9	Gross Potential Income	100.00%	\$100,000.00	\$115,927.41
10	Vacancy	5.00%	(\$5,000.00)	(\$5,796.37)
11	Effective Gross Income	95.00%	\$95,000.00	\$110,131.04
12	Expenses	40.00%	(\$38,000.00)	(\$44,052.41)
13	Net Operating Income	55.00%	\$57,000.00	\$66,078.62
14				
15	Capitalization Rate		6.000%	4.090%
16				
17	Net Operating Income		\$57,000.00	\$66,078.62
18	Less: Debt Service		(\$42,073.16)	(\$42,073.16)
19	Cash Flow	4.49%	\$14,926.84	\$24,005.46

⁸ An interactive demonstration may be found at <http://demonstrations.wolfram.com/SimulatingTheIRR/>

⁹ Note that different accounting conventions, reporting requirements, property types and even regions affect the concept of "net income". The example in this model assumes an apartment building in private ownership. Commercial property in institutional ownership might distinguish "net income before capital expenditures", "net cash flow before debt service" and other variations intended to highlight subtle differences in how one arrives at the income stream which rewards capital investment.

Figure 2

Another portion at the top, showing only columns D-H in Figure 3, reveals the loan terms and holding period. This input information can be changed by the user and drives the debt service and loan balance calculations elsewhere.¹⁰

D	E	F	G	H
Loan				
Term	Interest Rate	Income Growth	Hold (yr)	
30	5.50%	3.00%	5	

Figure 3

Below on the same page, one finds the hurdle rate and five year projection of yearend cash flows (including the pre-tax sales proceeds net of loan payoff and sales costs in the final year) which drive the NPV and IRR calculations, all shown in Figure 4.

	A	B	C	D	E	F	G	H	I
21	Hurdle Rate					13.0000%			
22									
23	Cash Flow Year		0	1	2	3	4	5	
24	Cash Flow Amount			\$ 14,926.84	\$ 16,636.84	\$ 18,398.14	\$ 20,212.27	\$ 22,080.84	
25	Sale Costs	6.00%							
26	Net Sale Proceeds (pre-tax)							\$438,390.51	Yield
27	All Flows		(\$332,500.00)	\$ 14,926.84	\$ 16,636.84	\$ 18,398.14	\$ 20,212.27	\$ 460,471.35	10.5963%
28								(\$27,600)	←NetPresVal→

Figure 4

The heart of what is new, therefore requiring the most exposition, is in the center of the **Model** worksheet. The best way to appreciate what is going on is to experiment with the worksheet input values.

	D	E	F	G	H
7					
8				2	
9					
10					
11					
12			0.15395%	CapRateIncrement	
13					
14					
15		rv_α=1	0.15395%	CapRateIncrement	
16	alpha	1.5		Avg	0.04478%
17	beta	0.5		Trial	4
18	gamma	1			
19	delta	0			
20					

Figure 5

¹⁰ Some changes are more painful than others. Excel requires a column for each year so a change in holding period (Cell G3) requires a structural change in the worksheet. Following all the other equations while doing so is critical.

Cells F16 and F17 in Figure 5 contain, respectively, up and down arrows. Clicking on one of these changes, in Cell F15, the value of the individual Trial number (shown in Cell H17) up or down among the 10,000 trials. The random variable of interest is CapRateIncrement. This number, depending on its sign, increases or decreases the going in capitalization rate to its new value at sale, the going out capitalization rate.

	A	B	C	D	E	F	G	H	I
15	Capitalization Rate		6.000%			0.15395%	CapRateIncrement		6.154%

Figure 6

In the example (Figure 6) the cap rate increased over the five year holding period. If income has remained constant during that time the value decreased. This leads to an important modeling question. Should the uncertainty be placed on the change in income or the change in how the market values income? While this is an interesting and unresolved question, the user may choose either or both. This monograph merely describes how one particular model was constructed. One may as easily change the random variable from CapRateIncrement in Cell F15 to Income Growth in Cell F3.¹¹

The block of cells highlighted in yellow in Figure 5 offer the opportunity to input the four stable parameters. Ignoring for the moment the last two which are scale and shift parameters, alpha permits the user to specify tail thickness (and center peakedness) and beta permits the user to specify skewness. On the right of **Model** worksheet (and on **PMTable** worksheet) is a graph of the Yield Histogram. A good exercise at this point is to notice that the default alpha value when you first download the workbook is 1.5, a number that has much empirical support in the literature. It is that number, with the beta value, that determines the shape of the histogram. That shape describes the risk.¹² When alpha = 2 the distribution is normal.¹³ Focus on the default shape while you change the value of alpha to 2. Notice how the distribution goes from the heavy right tailed shape to normal.

When the cursor is placed in a parameter cell you are shown a small flag telling you the value limits that parameter takes. Heed those warning messages. Impermissible values will nearly always generate an error message. For technical reasons, while alpha can theoretically take values from zero to two, alpha in ModelingRisk.xlsx is limited to values between 1.1 and 2.¹⁴ Excel has some limitations. For instance, gamma can take any positive number. Excel has no way to express infinity except in text form so the internal limitation in that cell is zero to a very large positive number not likely to represent anyone's actual gamma.

¹¹ Care must be taken with scale. Notice that CapRateIncrement is scaled such that it only changes significant digits on the far right. Wild results can be obtained by ignoring this caveat.

¹² This is a useful simplification. Viewing risk as not just a number but a shape can move the practitioner well along the learning curve of formal risk management.

¹³ Since normal is, by definition, symmetrical, this renders beta either zero or irrelevant, depending on your level of mathematical formality.

¹⁴ Those technical reasons are the subject of a number of other papers in the mathematical finance area. As a practical matter, real estate alpha values are rarely close to one. This is more than a mathematical convenience, as any aficionado of extreme values knows.

Figure 7 deals shows how we deal with the Jensen problem.

	A	I	J	K	L	M	N	O	P	Q	R	S
22			Expectation of the function								Function of the expectation	
23	Cash Flow Year				0	1	2	3	4	5		
24	Cash Flow Amount				\$ -	\$ 14,926.84	\$ 16,636.84	\$ 18,398.14	\$ 20,212.27	\$ 22,080.84		
25	Sale Costs											
26	Net Sale Proceeds (pre-tax)		Yield Average							\$ 839,146.85		Bias
27	All Flows		10.5963%	9.83%	(\$332,500.00)	\$ 14,926.84	\$ 16,636.84	\$ 18,398.14	\$ 20,212.27	\$ 861,227.69	24.19%	14.36%
28		←NetPresVal→	\$ 164,890							Net Present Val	\$164,890	\$0.00

Figure 7

Perhaps the best way to describe this is to start at the point where there is no Jensen problem, when NPV is the performance yardstick of choice. The worksheet provides two different NPVs. One is shown in Figure 4, Cell H28, and is based on the user’s pre-specified Hurdle Rate shown in Figure 4, Cell F21. The other, shown in Cells J28 and R28 (Figure 7) have the same value but are arrived at via two different calculations. The value in Cell J28 is the average of all 10,000 net present values arising from the 10,000 possible CapRateIncrement values resulting from the random draw (which the user specified having a particular shape due to careful selection of alpha and beta) shown in Column F of **PMTable** worksheet. *This is the expectation of the function.* In Cell R28 the NPV is calculated using values from Columns L-Q and the hurdle rate. Columns L-P are identical to Columns D-G in the same rows (Rows 24 and 27). What is different is Cell Q26 (not the same as Cell H26) which is the average of all the 10,000 sale prices generated in the random draw shown in Column E of **PMTable** worksheet.¹⁵ *This is the function of the expectation.* The two NPV values in Cells J28 and R28 are the same, despite being the result of different calculations because NPV is a linear function of cash flow.

The same is not true of IRR. In figure 7 we see three outputs, any of which might be categorized as a “yield” (an IRR?). In Cell I27, highlighted in green and beneath the word “Yield” is the IRR for the particular trial shown in Cell H17. Click on the up button in Cell F16 or the down button in Cell F17 to watch this value change. The other two percentage values, in Cells J27 and R27 represent different ways to portray IRR for the entire model given the specific 10,000 draws. The percentage value in Cell J27 is the average of all 10,000 IRR yields shown in Column D of **PMTable** worksheet, again *the expectation of the function.* Cell R27 reflects the IRR of the cash flows in the cells immediately adjacent to the left (Cells L-Q) which includes the average net sales proceeds, *the function of the expectation.*

Critical to this discussion is what may be termed “the bias” arising from Jensen’s inequality shown in Cell S27.¹⁶ The difference between the values in Cells J27 and R27 is a manifestation of Jensen’s insight that the function of the expectation is always less than or equal to the expectation of the function for convex functions.¹⁷

In the past the Jensen problem has been largely ignored as being relatively small in scale. The user can demonstrate that by again returning alpha to 2 making the draw normal. In the presence of extreme

¹⁵ Note that this series is constrained to have only non-negative values for the obvious reason.

¹⁶ Note that the bias is zero in Cell S28 for NPV because there is no bias.

¹⁷ Of course, the reverse is true for concave functions.

values the bias becomes significant (Cell S27). This calls into question the use of IRR at all or, at a minimum, reporting to the bias to the client.¹⁸

Figure 5 shows a box with buttons preceding labels “Avg Input” and “Simulated” which allow the user to toggle between two alternatives. Change the option to Avg Input and the average of 10,000 CapRateIncrements is used for the present yield calculation. Other changes should be noted. Cell F15 becomes the same as Cell H16, the average of CapRateIncrements. Yield and Expectation of the Function become the same (Cells I27 and J27). Also the Bias becomes zero for all yield measures. This is to be expected when relying on a point estimate as no simulation is being done. Note the Yield Histogram has a single bar at the average. The reason the bias in Jensen’s Inequality goes away is that all variation has been removed from the model as a result of relying on the average. Distinguish between this outcome and the prior when there was both much variation and a large bias and you understand why IRR simulation is only accurate when you need it least.

More importantly, by toggling between the simple average point estimate and the shape of actual variation, we understand why the average outcome is wrong on average.¹⁹ The “Avg Input” option is what practitioners do when they represent that an investment will yield a fixed return over time. This is a proclamation that one can foretell the future. Even if the client does not explicitly object, that client reality tests by mentally forming a probability distribution around that fixed, single yield prediction. What this tool does is make explicit that reality test by giving it real numbers. The final result is typically shown in the form of a cumulative distribution function (CDF), which is the subject we discuss next.

The CDF

The empirical cumulative distribution function is about as simple as it can be. It makes no assumptions about the limiting distribution. It merely lines up the random variables in order, smallest to largest, and counts them, normalized to the y-axis. The inverse, known as Quantiles, answer the question: “How many observations are below this percentage of all observations?” Together they are handy tools for explain probability and risk to clients.

With alpha still set to 1.5 and beta set to .5, make sure that column B (and only Column B) in the **CDF** worksheet is sorted. Then note a few individual points on the far right (upper) part of the plot. These are the extreme values that cause the distribution to have a heavy right tail. Since our random variable drives yield, these are happy events.

Highlighted in yellow in Figure 8 are six common quantiles. The user can insert his own if he prefers an odd quantile like .325 or .6841. The 50% quantile is explained in the box to the right and of course all others follow. Now the model is furnishing useful information that can guide a decision. If the investor is

¹⁸ This will represent a special challenge to those who avoided NPV because it is too hard to explain.

¹⁹ A tip of the hat to Sam Savage for this pithy phrase.

happy with a 75% chance of a 16.1% return and believes the risk inherent in a distribution with the tails specified to produce this set of 10,000 returns, this project is for him.²⁰

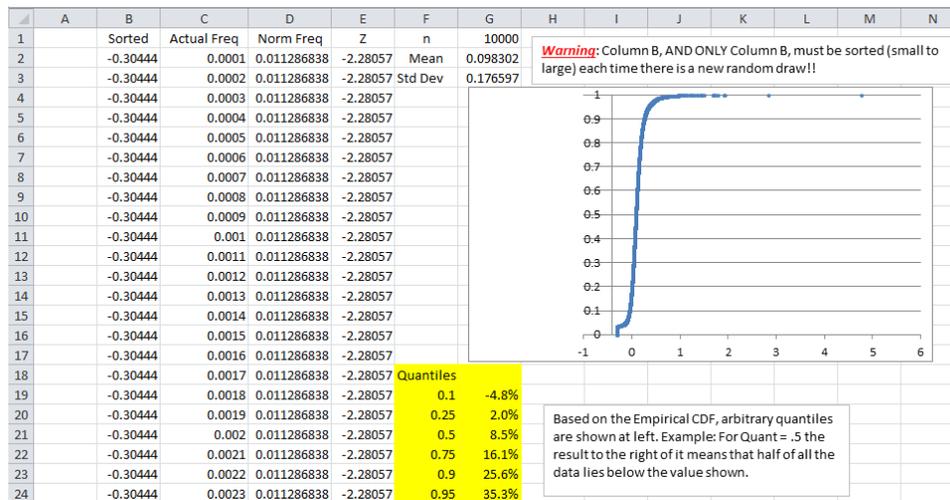


Figure 8

Conclusion

ModelingRisk.xlsx delivers a rich variety of outcomes based on sound distributional theory. Rather than tell a client the return is a deterministic x% “because I said so”, one may offer the client the opportunity to choose a risk comfort zone in which to contemplate a range of returns.

Most important to the process is to explicitly consider risk and what it means to real estate investment. Clients know you are not God. They always passively superimpose a probability spectrum over your future return projections. Now the client and his advisor may actively model uncertainty, arriving at quantitative conclusions. Even if those conclusions carry with them assumptions, qualifications and provisos, the open discussion and debate of how such hedging affects the computations is a worthwhile exercise.

²⁰ Jensen is not involved here as there is no expectation being taken. The empirical distribution takes the data as it comes. In this case, the 10,000 IRRs are in Column D of the **PMTable** worksheet. The 16.1% is just the return in the sorted list that appears at the point closest to return #7500 of the list. Check this by looking at Cell B7501 on the **CDF** worksheet.