Hollywood Economics: Dealing with 'Wild' Uncertainty in the Movies and Pharmaceuticals

Arthur De Vany
Professor Emeritus of Economics and Mathematical Behavioral Sciences
University of California, Irvine

October 19, 2004

The Movies: An Extraordinary Business

Research strategy: dynamics → distributions.

Understand patterns – stochastic evolutions of the system – rather than the details of particular realizations.

It is a theory of process, not the accidental details of each unfolding of the process.

Model decentralized, dynamical information processes and the statistical attractors to which they converge.

The dynamics are complex, with elements of chaos, and the statistical attractors are "strange."

Yet, the dynamics retain a deeper order of stability and self-similarity.

The statistics are "wild" and far from Gaussian; it's not a Normal Business.

If the movies were Gaussian, *Gone with the Wind*, *Home Alone*, *Titanic*, or *Waterworld* would never occur.

There would be small differences among movies and stars—a mediocracy.

It is a kurtocracy; extraordinary movies and artists abound.

Modeling the Movies

Model decentralized, dynamical information processes and the statistical attractors to which they converge.

The dynamics converge on the "wild" statistics of the stable Paretian distribution.

The (Lévy) stable distribution is a model of extraordinary events.

The α-stable distribution represents with great fidelity the statistics of the movies on many levels.

Implications of the α-stable Paretian model:

- Complex dynamics, bifurcations, 'wild' attractors.
- The curse of the superstar.
- The angel's nightmare.
- "No body knows anything".
- Nobody knows and the option principle.
- Momentum; success is proportional to success.
- The HHI is a bad measure of competition.
- High kurtosis and non-finite moments.
- Self-similarity.

Dynamics

Recursive, non-linear dynamics can “go anywhere.”

Convergence in distribution during the run to a skewed, heavy tailed distribution.

Complex dynamics, bifurcations and contractions.

Hits have non-linear, expansive dynamics and rising conditional expectation.

Pictures.

Stable Distributions

- Stability: a random variable and its sums have the same probability distribution.
- The Gaussian distribution is stable, but it is the only one with finite variance.
- Pareto, Lévy, and Cauchy are the other (named) stable distributions.
- Generalized Central Limit Theorem: if the sum of a large number of independent identically distributed random variables has a limiting distribution (after appropriate scaling) the limiting distribution must belong to the stable class.
- Basin of attraction is characterized by tail weight (α).
- There are no closed forms for the whole class of distributions. They are characterized by the characteristic function.

Parameterization: $X \sim \mathbf{S}(\alpha, \beta, \gamma, \delta)$.

- When $\alpha \neq 1$ the characteristic function of X is $\left\{\exp\{-\gamma^{\alpha}|t|^{\alpha}[1 + i\beta(\tan\frac{\pi\alpha}{2})(\text{sign } t)((\gamma|t|)^{1-\alpha} - 1)] + i\delta t\}\right\}$
- The exponent α is a measure of the probability weight in the upper and lower tails of the distribution; it has a range of $0 < \alpha \leq 2$ and the variance of the stable distribution is infinite when $\alpha < 2$.
- The skewness coefficient $-1 \leq \beta \leq 1$ is a measure of the asymmetry of the distribution.
- The scale parameter γ must be positive. It expands or contracts the distribution in a non-linear way about
- the location parameter δ which is the center of the distribution.
- Its tails are Paretian $(P(x \geq K) = Kx^{-\alpha})$ and moments of order ≥ 2 do not exist when $\alpha < 2$.
- Its mean need not exist for values of $\alpha < 1$.
- The α-stable distribution becomes the Cauchy distribution when $\alpha = 1$ and $\beta = 0$,
- the Lévy distribution when $\alpha = 0.5$ and $\beta = \pm 1$.
- When $\alpha = 2$ the skew parameter ceases to have any impact and the stable distribution has only two parameters $(\gamma,\ \delta)$ the mean and variance of the symmetric Gaussian distribution.

Diagnostics and estimation

Look at mean, mode, median statistics.

Look for increasing variance.

Look for self-similar fractal paths (DFA program).

Look for independence of opening to total sales.

Look for a non-stationary mean.

Look for log-log linear regions in tails of pdfs (Mandelbrot's regression procedure).

Look at rank/size plots for log-log linearity at high ranks (regression estimates $\frac{1}{\alpha}$).

Estimate general stable model (Nolan's Stable program and McCulloch's quantile estimator) and test parameter values against the distributions discussed above.

The Stable Distribution of Profit

The industry's rate of return is from 0 to 4%, not high for such a risky business.

Seventy-eight percent of movies lose money and only twenty-two percent are profitable.

35% of *profitable* movies earn 80% of total profit (the Gini coefficient of profit, a measure of inequality, is 0.608).

Among *unprofitable* movies, 50% accounted for 80% of total losses (Gini coefficient = 0.461).

Just 6.3% of movies earned 80% of Hollywood's total profit over the past decade.

This concentration of profit is not unlike stock market returns where 6% of trading periods earn 80% of returns.

By these measures, the movies is a "winner-take-all" business.

Profit is stable-distributed. [Picture]

Table 1: Maximum Likelihood Parameter Estimates

	α Index	δ Location	γ Scale	β Skewness	Log-Likelihood
All Movies					
Normal	2	-3.351	8.442	0	-7855.37
symmetric α-stable	1.268	-4.079	4.032	0	-7279.87
α-stable	1.259	-4.042	4.020	0.043	-7279.46
Movies with Stars					
Normal	2	-2.083	14.186	0	-1439.69
symmetric α-stable	1.582	-4.568	10.555	0	-1419.16
α-stable	1.624	-6.385	10.805	0.768	-1410.82
Movies without Stars					
Normal	2	-3.595	6.789	0	-6216.46
symmetric α-stable	1.358	-3.932	3.507	0	-5739.47
α-stable	1.335	-3.827	3.441	-0.122	-5737.95

The Angel's Nightmare.

The Paretian distribution implies that a movie's expected production cost grows in proportion to the costs already expended.

Consider a movie whose expenditures are X. The expected budget conditional on X is $X\frac{\alpha}{(\alpha-1)} = X \times 1.609$.

Suppose \$20 million has already been spent on a movie that was expected to cost \$16 million (the average cost).

The expected cost of the movie, conditional on \$20 million having been expended is $20 \times 1.609 = 32$.

The movie is already \$4 million over budget and yet its expected cost is, \$32 million, twice the original planned budget!

Theory: production is a recursive, non-linear and complex process.

The Curse of the Superstar.

The superstar profit distribution is centered on a loss, has a narrow central region, and is asymmetric, with a heavier upper than lower tail. [Picture]

The shape of the stable distribution causes the sample average, expected, and most likely superstar profits different from one another.

The expectation of profit is 7.684, the sample average is -2.083, and the most probable profit (mode) is -7.500.

If bidding for a superstar raises her fee to equal expected profit, the movie almost surely will lose money (80% chance). Mode dominates in small samples.

Producers can guard against the curse by paying contingent compensation.

Self Similarity

- Cultural self-similarity. The size distribution of motion picture revenues in various countries is Paretian: United States, Hong Kong, the UK, Australia and Ireland.
- The concentration of revenues among movies is universal among cultures and film markets. The top film will earn about the same proportion of total revenues in each market where it plays.
- Temporal self-similarity. US motion picture revenue distributions have been stable for more than two decades.
- Audiences still rank movies in such a way as to give the same proportional revenues to the top movies.
- A striking form of temporal self-similarity is seen in the ragged profile of the time series of motion picture revenues.
- Budget self-similarity.
- Self-similarity in careers, movie gross revenues, and artist pay.
- In short, the movies are statistically self similar over time, genre, culture, rating, budget and other windows.

What is wrong about the Gaussian Distribution?

Option models use the Black-Scholes-Merton option pricing model to evaluate R&D expenditures.

The log-Gaussian distribution reflects a world view that is not faithful to the innovation process.

It assumes the existence of a mean value of innovations and the existence of a variance.

Positive deviations from the mean larger than two standard deviations should not occur more often than in 2.3% of innovations.

There would be no penicillin, Prozac, or Lipitor in a Gaussian world.

The Gaussian distribution cannot describe the huge range of impacts and potential benefits from rare breakthroughs or discoveries in drugs.

Growth is not proportional to firm size.

The leaps and falls of sales are Levy flights.

Similarities: movies and pharmaceuticals

- Innovation:
  - discovery versus copy cats
  - new films and artists versus new medical entities
  - sequels and derivatives versus "me too", generics and improved entities.
- Hit driven and extreme events dominate.
- Skew and concentration.
- Upfront costs and deferred income.
- Low marginal cost.
- Pickups.
- Non-linear dynamical systems have no natural scale.
- Revenue is just a number: a measure of information.
- They are both information industries.
- These imply the "nobody knows" principle.

Some Innovation Statistics

Scherer found the distribution of patent royalties to be Paretian (though he now tells me it is Log Normal).

He also adopted the BE process as a model of the dynamics.

Kurtosis in pharmacueticals.

IMEs from the top five categories jointly accounted for an estimated half of the total $16.9 billion increase in spending derived from IMEs from 1995 to 2000.

The top ten categories accounted for $11.5 billion or 68%, and the top 12 categories for an estimated $12.6 billion, or 75%.

The top five categories for NME spending accounted for over half (53%) of the total of $26.5 billion in increased spending attributed to NMEs in 1995–2000.

The top ten contributed 73% of the total from NMEs.

Mean, mode, and median comparisons indicate leptokurtosis in drugs and innovations.

Growth rates from a pure Gibrat process (growth in proportion to size).

A Gibrat growth process predicts a linear increase of the variance with time; this is not found.

Second, the standard deviation of the growth rates does not decrease according to $S^{1/2}$.

Gibrat process converges on a log-Normal distribution and this is rejected in the dynamics (they do not settle down and show longer range correlations).

The growth distributions are non-Gaussian and leptokurtotic.[Picture]

Thoughts and suggestions

Its not a Gaussian world out there.

Look for dependence on large events and unpredictability.

Look for self-similarity where "big" drug projects have the same cost and payout distribution as "small" projects.

Watch for the curse of the superstar, ie skewed profits and asymmetric tails.

Watch for the angel's nightmare, cost expectations grow with cost expended.

Expectation is different from average and averages tend to be unstable. (Forecasting problem.)

Be comfortable with infinite variance. Standard regression analysis is error-prone.

HHI may be meaningless, though often used in drug industry research.

A lot of success is just luck, eg. antibiotics and "crossover" drugs for heart disease.

So far, genetic engineering hasn't produced much — insulin is the most successful product (because they pulled pork-derived insulin off the market) and it is more "brittle" than pork-derived insulin.

How did the statins become so big in the absence of much evidence to show they are of value? Four massive trials show no mortality difference and big side effects.

It is still not known how they work; they are good anti-inflammatories and likely damp cardiovascular inflammation.

Medical statistics seldom acknowledge "heavy tails." Statins modest effect on cholesterol is secondary, mostly benefiting hyperchoserolaemics (the dominant upper tail of mortality) who do not have tight control on cholesterol.

This says something about the way drugs and information about them spread over the medical community.

Cardiovascular disease has the pattern of a novel pathogen invading humans. A rise followed by an accommodation and a decline.

Is it chlamydia which is found in damaged vascular tissue?