ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 49/50 — 1998

The Size Distribution of Profits from Innovation

Frederic M. SCHERER*

ABSTRACT. – This paper reports on research seeking to determine how skew the distribution of profits from technological innovation is – *i.e.*, whether it conforms most closely to the Paretian, log normal, or some other distribution. The question is important, because high skewness makes it difficult to pursue risk-hedging portfolio strategies, and it may have real business cycle consequences. Data from several sources are examined: the royalties from U.S. university patent portfolios, the quasi-rents from marketed pharmaceutical entities, the stock market returns from three large samples of high-technology venture startups, and preliminary results from a survey of German patents on which renewal fees were paid until full-term expiration in 1995. The evidence reveals a mixture of distributions, some close to log normality and some Paretian. Preliminary hypotheses about the underlying behavioral processes are advanced.

La distribution des profits d'innovation

RÉSUMÉ. – L'article présente les résultats d'une recherche cherchant à déterminer le degré d'asymétrie de la distribution des profits d'innovation technologique, – autrement dit si cette distribution s'ajuste plus étroitement à une distribution de Pareto, log-normale ou autre. La question est importante, car une forte asymétrie rend difficile les stratégies de diversification et couverture de risque, et peut avoir des conséquences sur le cycle réel de conjoncture. L'analyse porte sur des données de sources variées : les redevances de portefeuilles de brevets d'universités américaines, les quasi-rentes de produits pharmaceutiques, les rendements boursiers de trois grands échantillons d'entreprises de haute technologie nouvellement créées, et les premiers résultats d'une enquête relative à des brevets allemands ayant été renouvellés jusqu'à leur terme complet d'expiration en 1995. Les estimations mettent en évidence un mélange de distributions, certaines plus proches de la loi de Pareto, d'autres de la loi log-normale. Des hypothèses de comportement pouvant rendre compte de telles distributions sont suggérées.

* F. M. SCHERER: Harvard University. The support from a Sloan Foundation grant is gratefully acknowledged, as are comments from three anonymous referees.

1 Introduction

It is now widely recognized that the size distribution of profit returns from technological innovation is skewed to the right. The most profitable cases contribute a disproportionate fraction of the total profits from innovation. The exact form the distribution function takes remains a mystery. In an early analysis (SCHERER [1965]), I reported that the profits measured through a small survey of U.S. patents conformed tolerably well to the Pareto-Levy distribution:

$$N = kV^{-\alpha}, \tag{1}$$

where V is the value of profits from an innovation, N is the number of cases with value V or greater, and k and α are positive parameters. The equation is linear in the logarithms. For the sample analyzed, the linear slope coefficient α appeared to be less than 0.5. Since then economists have sought to discern the size distribution's form by analyzing the rate at which patents have been allowed to expire before full term due to non-payment of the periodic maintenance fees imposed by many national patent offices. Early evidence analyzed by PAKES and SCHANKERMAN (1984, p. 78) favored the Pareto-Levy distribution, but later work by SCHANKERMAN and PAKES [1986] found mixed but stronger support for the log normal distribution [1].

The difference between distributions is important. When the distribution is Pareto-Levy and α is less than 2.0, the variance is not asymptotically finite, and for $\alpha < 1$, the mean is also not asymptotically finite. What this implies is that as one draws ever larger samples, there is an increasing chance that some unprecedentedly large value (e.g., an extraordinarily large profit) will be included, overwhelming the observations drawn previously and forcing the mean and variance upward, contrary to conventional expectations under the law of large numbers. With finite variances and means, log normal and similar skew distributions are better behaved statistically. Still the more rightward-skewed the distribution is, whether Pareto-Levy, log normal, or some related form, the more difficult it is to hedge against risk by supporting sizable portfolios of innovation projects. The potential variability of economic outcomes with Pareto-Levy distributions is so great that large portfolio draws from year to year can have consequences for the macroeconomy. In a simulation experiment, NORDHAUS (1989, p. 324) discovered that aggregated Pareto-distributed productivity effects from samples approximating in size the number of patents issued annually in the United States mimicked the long-term productivity fluctuations actually experienced by the U.S. economy between 1900 and 1985 [2].

Crucial to the portfolio properties of large invention samples is the value distribution of observations in the right-hand (most valuable) tail.

1. See also PAKES (1986, p. 777), SCHANKERMAN [1991] and LANJOUW [1992], all of whom find distributions less skew than the Pareto-Levy.
2. On the aggregation properties of Pareto-Levy distributions, see MANDELBROT (1963).

On this, studies of patent renewals provide only limited insight. Even in nations with relatively high maintenance fees that rise over the patent's life span, only 10 to 20 percent of the issued patents survive to full term (*i.e.*, 18 to 20 years) after paying all fees. Such patents are clearly of relatively high value. However, the distribution of values within the full-term cohort is ascertained in renewal studies only by extrapolation, not by direct measurement. Because, as we shall see, it is difficult under even the best of circumstances to discriminate among size distributions on the basis of right-hand tail characteristics, extrapolation is hazardous. Also, the mapping from patents to innovations is far from simple. Many innovations are covered by numerous patents, some with a crucial imitation-blocking role, some not. Although patents cannot add significant value to a worthless technology, they enhance the rewards appropriated from certain valuable innovations, but in other cases are unimportant because there are alternative barriers to competitive imitation. See SCHERER [1977] and LEVIN *et al.* [1987].

This paper reports the first results from an ongoing attempt to surmount the limitations of prior research. It probes the right-hand tail, analyzing detailed innovation data across the full spectrum of positive payoff outcomes (while discarding negative outcomes). It examines not only individual patents, but technological innovations construed more broadly. Specifically, evidence will be presented from three patent samples, in two or which complementary patents are bundled together; two nearly exhaustive samples of new pharmaceutical entities introduced into the U.S. market; and two large samples of high-technology venture capital investments.

The approach pursued here is inductive, seeking not to impose an a priori theory upon the data, but merely to see what general patterns the data reveal as a first step toward sorting out plausible from implausible theories [3]. Parallel research is exploring alternative models (such as Gibrat and Einstein-Bose processes) in an attempt to identify stochastic regimes most consistent with the observed distribution of profit outcomes [4].

Preliminary insight is maximized by plotting the data graphically, although statistical tests will also be reported. The Pareto-Levy distribution will be emphasized in part because of its simplicity, plotting linearly on doubly logarithmic coordinates, and partly because it has the most radical implications for risk-pooling under a large-numbers strategy. When the data permit, the log normal distribution (which emerges from Gibrat stochastic processes) and the exponential and Weibull distributions will be considered as alternative hypotheses. The focus, again, will be on the high-value tails of the observed distributions.

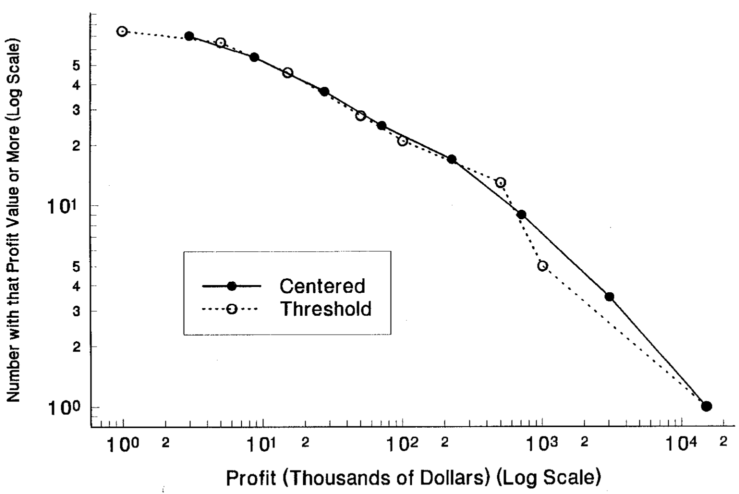
3. This methodology is more typical of the natural sciences than economics. See IJIRI and SIMON (1977, pp. 4-5 and 109-111); SCHERER [1986] (on Kepler, Einstein, Watson, and Crick), and more generally HANSON [1961].

4. See e.g. GIBRAT [1931], De VANY and WALLS [1996], and IJIRI and SIMON [1977].

2 Patent Data

Figure 1 plots on log-log coordinates the profitability data analyzed in my 1965 article [5]. The profit estimates were drawn from a survey of 600 U.S. patents by SANDERS *et al.* (1958, p. 355; 1964, p. 53), among which 74 useable responses with positive profits were received. (For 149 additional patents, net losses, typically small, were quantified.) The data were reported as patent counts over seven profit intervals. Consistent with the Pareto distribution assumptions of equation (1), the dotted line plots the data points at lower interval bounds. Since the lowest (under $5,000) interval was bounded at zero, for which no logarithm exists, an arbitrary lower bound of $1,000 is assumed. The distribution function's implied left-hand shape is sensitive to that assumption. The highest profit value interval, "over $1 million", included five patents, but a later (1964) article revealed the maximum value to be $15 million. Since that largest observation is a point rather than a lower bound, the solid line plots for consistency the data points at the geometric means of class intervals, assuming a value of $3,000 for the lowest-profit interval.

FIGURE 1



For the dotted lower-interval-bound plot, a straight line fitted by least squares yields an α value of 0.454, with standard error of 0.050 and r^2 of 0.931. For the solid geometric mean plot, $\alpha = 0.497$, with standard error of

5. For ease of plotting, all logarithms used in this article are taken to the base ten.

0.040 and $r^2 = 0.963$ [6]. If the distribution is in fact Paretian, both α values lie within the range where neither means nor variances are asymptotically finite. However, both plots reveal a modest degree of concavity inconsistent with the Pareto-Levy hypothesis. Adding a quadratic term to the centered value regression raises R^2 by 0.0343, which, despite retaining only five residual degrees of freedom, is highly significant in an F-ratio test.

For a first extension of the patent value analysis, a novel data source was tapped. The Bayh-Dole Act of 1980 changed U.S. law, permitting researchers in universities and other non-profit institutions supported by federal government grants to receive and assign exclusive rights to inventions resulting from their government-funded research. Many universities established technology licensing offices to apply for patents on such inventions and to negotiate licenses with private sector enterprises for their commercial exploitation. By 1993, the royalty revenues received by 117 U.S. universities from their outstanding technology licenses had reached an annual rate of $242 million [7]. Among those 117, the top ten institutions had royalty revenues of $171 million, or 70.6 percent of the total.

One of the top ten on this list was Harvard University, my emplyer. The Harvard Office of Technology Licensing kindly provided to me a detailed confidential tabulation of the total royalties received between 1977 and May 1995 on 118 technology "bundles" with non-zero royalties whose patents had been applied for by the end of 1990. Among the 118 bundles, 27 included more than one patent, and six included five or more patents. The twelve bundles with the largest cumulative royalties originated nearly 84 percent of total portfolios royalties – characteristic evidence of high royalty distribution skewness.

Figure 2 plots the royalty income (multiplied by a constant disguise parameter) from individual invention bundles. The plot is clearly not linear, as would be expected with a Pareto-Levy distribution, but shows considerable downward concavity. If a straight line (in the logarithms) is forced by least squares to the data, the indicated slope is 0.41, with standard error of 0.015 and r^2 of 0.865 [8]. Adding a quadratic term to the regression raises R^2 to 0.983; the increment is significant at the 0.01 level. The slope is plainly steeper in the higher-value tail of the distribution. A linear regression on the 39 most valuable invention bundles (*i.e.*, the top third of the distribution) yields a slope of 0.71, with the standard error of the coefficient being 0.016 and $r^2 = 0.981$. A further analysis of all 118 Harvard technology bundles revealed the fitted Pareto-Levy line to be insignificantly influenced by the number of patents in the bundles and (surprisingly) by the age of the patent bundles.

---

6. Since each group has different numbers of observations, one might prefer to use weighted least squares, where the weights are the square roots of the number of observations. This places relatively more weight on the lower-value observations, which are more numerous. The estimated α values (and standard errors) are 0.377 (0.048) and 0.433 (0.036) for the bounded and centered observations respectively.
7. Aggregated university data were provided by the Association of University Technology Managers.
8. Despite the poor fit, double-log slopes are reported here becaused they will be used in a later paper as part of a benchmark for evaluating alternative stochastic processes.

FIGURE 2

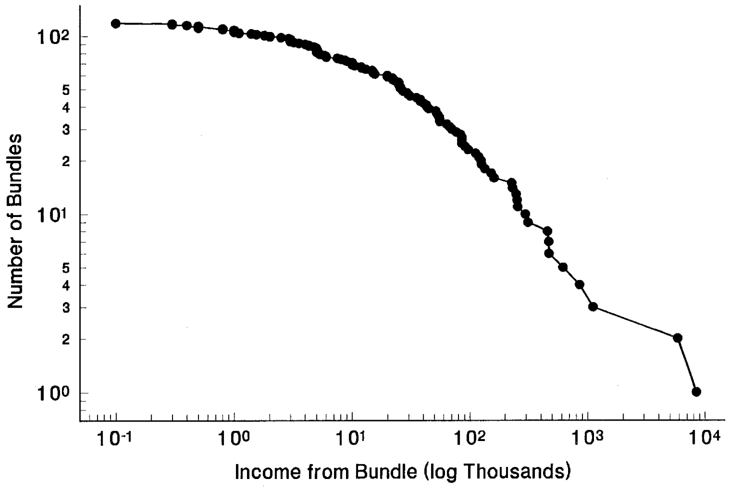
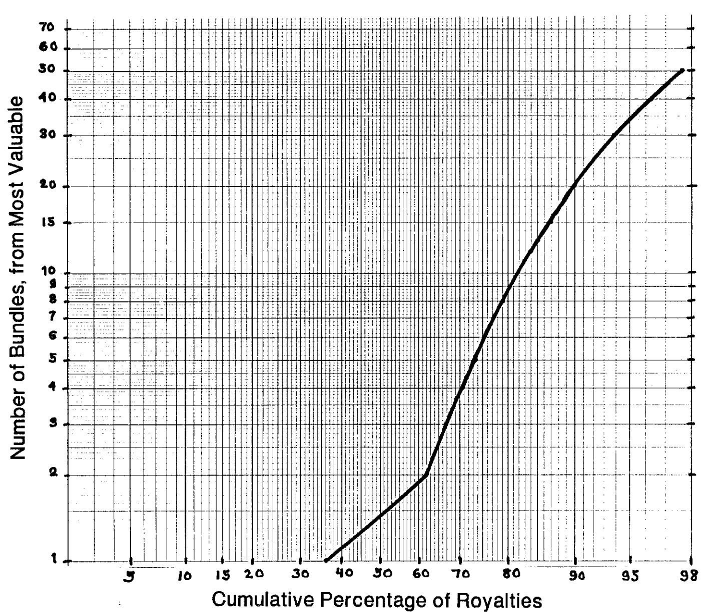


Figure 3 plots the Harvard royalty data on log normal probability coordinates, with the cumulative probability given on the horizontal axis and the cumulative number of invention bundles required to reach that

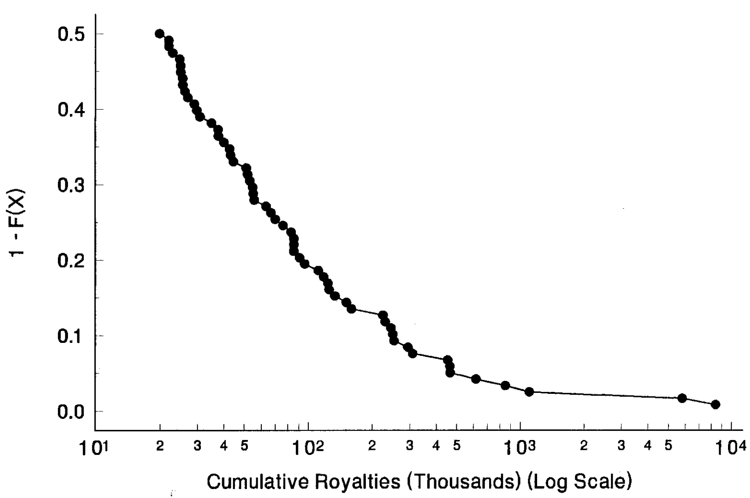
FIGURE 3



probability, starting from the most valuable bundle (*i.e.*, from the right-hand tail), on the vertical axis. The linear fit one would expect if the distribution were exactly log normal is absent. Letting μ_i be the i-th moment of a distribution, we define the coefficient of skewness to be $\sqrt{\beta_1} = \mu_3/\mu_2^{3/2}$. Taking logarithms of the royalty distribution, we find $\sqrt{\beta_1} = +0.17$, which rejects the hypothesis of log normality only at a confidence level short of 80 percent. See D'AGOSTINO and STEPHENS [1986], p. 380. The coefficient of kurtosisis is defined as $\beta_2 = \mu_4/\mu_2^2$. Its value for the log royalty distribution is 2.99, which is not significantly different from the 3.00 value associated with log normality.

Figure 4 subjects the Harvard data to a further graphic test attributable to M. C. BRYSON [1974] [9]. The horizontal axis measures royalties logarithmically. The vertical axis is scaled as one minus the cumulative probability integral, beginning with the lowest-royalty observation, but truncating at the midpoint of the distribution so that only the highest-value observations are plotted. If the underlying distribution were negative exponential, the plot would be linear; if the distribution were two-parameter Weibull, the plot would be concave downward. Both hypotheses are clearly rejected. The downward convexity exhibited in Figure 4 is consistent with log normal, Paretian, or other highly skew distributions. The analysis of the preceding paragraph favors an inference of log normality.

FIGURE 4



Eleven university technology licensing offices, including the top ten royalty recipients of 1993, were asked to provide information on the distribution of their technology license royalties, divided into nine value ranges, and on total royalty income, for each of their fiscal years 1991

---

9. See also D'AGOSTINO and STEPHENS [1986], pp. 18-23.

through 1994 [10]. Six of the eleven, with total royalties of nearly $83 million in 1993 on 466 positive-royalty cases, responded favorably. All six accounted for their licensed technologies as bundles of patents and disclosures rather than single patents, although, as in the Harvard case, most of the bundles contained only one patent. The largest royalty interval, "more than $5 million", was open-ended. Because data on total royalty receipts per year were obtained, however, it was possible to approximate the values of individual bundles in that tail of the distribution. The approximation was carried out by assigning to closed intervals the mean value of Harvard patents in that interval (which in every case was close to the geometric mean of the interval extremes). Interval totals were found by multiplying the number of patents in an interval by the mean values, and the sum of such totals for all closed intervals was subtracted from total royalties to estimate royalties in the right-hand tail (with at most one observation per university per year).

The fractions of total sample royalties contributed by the top six technology bundles in each year were as follows:

	1991	1992	1993	1994
Percent of royalties	66.2%	70.9%	75.6%	74.4%
Total number of royalty-paying bundles	350	408	466	486

The most lucrative bundle licensed by any university contained the process and product patents on gene splicing methods granted in 1980, 1984, and 1988 to Stanley Cohen of Stanford University and Herbert Boyer of the University of California (and administered by the Stanford Technology Licensing Office). At the end of fiscal year 1994, 290 non-exclusive licenses to that bundle had been issued. Over the four years covered by our sample, the bundle yielded royalty payments of roughly $75 million [11]. Since licenses to the Cohen-Boyer patents, which had a revolutionary impact on the biotechnology industry's development, carried only modest royalty payments [12], the social surplus contributed by the inventions was vastly in excess of royalties appropriated by the patent-holding institutions.

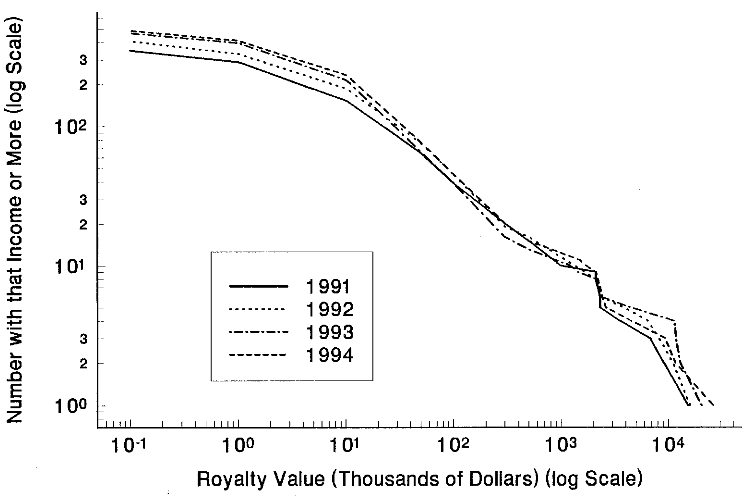
Figure 5 arrays the six universities' royalty distributions on double logarithmic coordinates. Over the four years sampled, the distributions are reasonably stable. They are clearly not linear as predicted under the Pareto-Levy law; considerable concavity is evident. However, it would be premature to reject the linearity hypothesis for the most valuable tail. Fitting log-linear regressions to bundles with annual royalties of $50,000 or more, the results are as follows:

---

10. In interpreting data published by the Association of University Technology Managers, one must be careful to eliminate royalties from trademark licensing, e.g., from firms printing university seals on their garments.
11. See WINSTON-SMITH [1996] and ALDRIDGE [1996], pp. 104-108. Aldridge's account confuses annual with total royalties.
12. The original terms called for a $10,000 advance payment plus royalty rates ranging from 0.5 percent (on the sale of end products such as injectable insulin) to 1-3 percent on bulk products and 10 percent on basic genetic vectors and enzymes.

	1991	1992	1993	1994
Estimated α	0.665	0.660	0.583	0.649
Standard error	(.038)	(.036)	(.039)	(.029)
r^2	0.963	0.971	0.949	0.974
Observations	14	12	14	15
Bundles included	64	75	65	76

FIGURE 5



Again, the slope values lie in the range within which, asymptotically, Pareto-Levy distributions have neither finite means nor variances. Adding quadratic royalty value terms to the regressions effected variance reductions statistically significant (at the 0.05 level) only for 1991, with $F(1, 11) = 7.58$. Incremental F-ratios for the other years were 2.69, 0.3, and 3.64.

3 The Profitability of Approved New Pharmaceutical Entities

New chemical entities for use as pharmaceuticals in the United States must undergo a rigorous series of clinical tests before being approved by the Food and Drug Administration. On average, 17.5 new chemical entities

(NCEs) received FDA approval per year between 1970 and 1986 [13]. Only about 23 percent of the new chemical entities entered into human trials emerged with marketing approval from the FDA. Counting both successes and failures, but ignoring the time value of invested funds, the average pre-clinical and clinical research and development cost of new drugs appearing on the market during the late 1970s and early 1980s was nearly $100 million (in 1987 dollars). See DIMASI *et al.* [1991].

Henry GRABOWSKI and John VERNON [1990, 1994] used detailed data on drug sales to estimate the gross profitability (before deduction of R&D costs) of new chemical entities (other than cancer drugs) approved by the FDA during the 1970s and early 1980s whose development was carried out by industrial companies in the United States. Subtracting estimated production and marketing costs from sales revenues, domestic and foreign, they obtained for each drug what are best described as Marshallian quasi-rents to R&D investment [14]. These quasi-rents were discounted at a real discount rate of 9 percent to the date at which the drugs were first marketed. The drugs were divided into deciles in descending discounted quasi-rent order, leading to the value distribution shown in Figure 6 for average quasi-rents of drugs introduced during the 1970s. A similar analysis with nearly identical results was conducted for drugs introduced between 1980 and 1984. Also estimated was the average research and development investment per approved new chemical entity, including the cost of failed experiments, brought forward at compound interest to the time of marketing – $81 million (in 1986 dollars) per new drug introduced during the 1970s. Drugs in the top decile – the so-called blockbusters – generated 55 percent of total 1970s NCE sample quasi-rents, *i.e.*, 5.6 times the average R&D costs underlying their market entry [15]. Drugs in the second decile yielded double their R&D investments, drugs in the third decile essentially broke even, and drugs in the seven lowest deciles brought in discounted quasi-rents less than their average R&D investments.

A high degree of skewness is evident in Figure 6. To permit a more detailed analysis, GRABOWSKI and VERNON supplied the quasi-rent data for individual NCEs, multiplied to maintain confidentiality by an undisclosed disguise parameter. (Multiplication by a constant does not distort the size distribution parameters in which we are interested.) Figure 7 plots on

---

13. Pharmaceutical Manufacturers Association, Statistical Fact Book (August 1988), Table 2-4. The typical new drug is protected by one product patent and sometimes by a few process patents. In 1976, U.S. pharmaceutical manufacturers obtained at least 868 patents and, allowing for incomplete sample coverage, as many as 1,000. See SCHERER (1983, p. 110). Thus, there is far from a one-to-one correspondence between new patent counts and new product counts.

14. Grabowski and Vernon assumed the same ratio of production and distribution costs relative to sales for each NCE. If the better-selling NCEs had lower variable cost ratios, the resulting quasi-rent size distribution could be biased toward concavity. LU and COMANOR [1997] found that more important therapeutic advances had higher prices and hence, all else equal, higher profit margins. However, in an analysis of 28 NCEs introduced during the 1980s, I was unable to find any systematic relationship between sales and the degree of therapeutic advance, as characterized by the U.S. Food and Drug Administration.

15. One might expect R&D costs to be higher for the most lucrative drugs, but the available evidence fails to provide support. See DIMASI *et al.* [1995], p. 169.

FIGURE 6

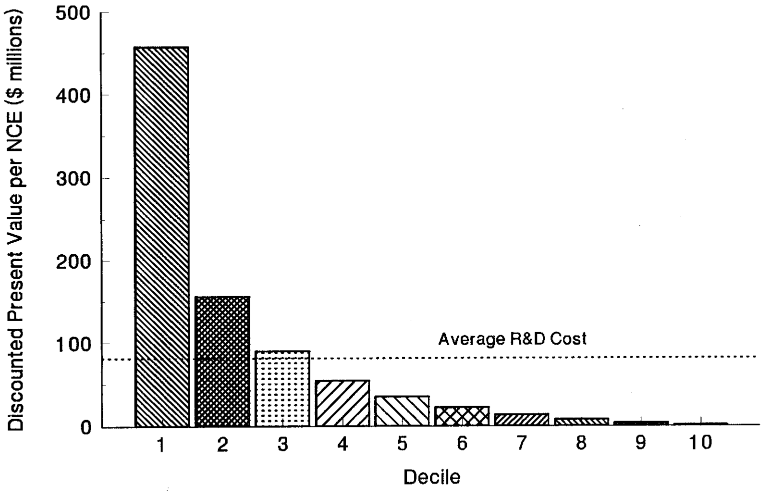
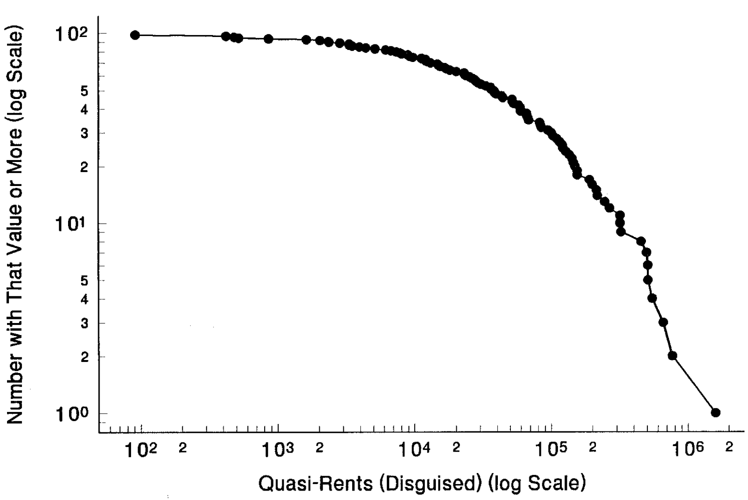
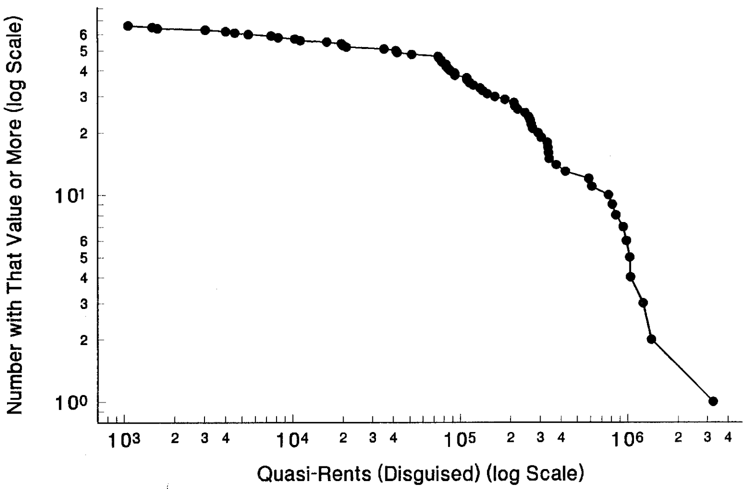


FIGURE 7



double log coordinates the data for 98 NCEs introduced during the 1970s. (Two observations with negative quasi-rents are omitted.) Figure 8 does the same for 66 NCEs introduced between 1980 and 1984. As with the quite differently constituted university patent bundle data, the distribution is much too concave to be Pareto-Levy. However, if one focuses only on

FIGURE 8



the most successful third of all the new products, a log linear regression fits tolerably well:

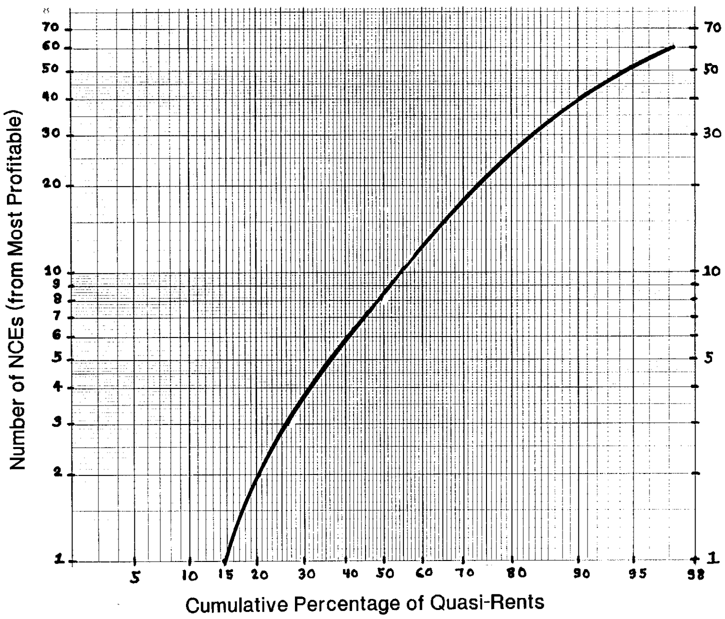
	1970s NCEs	1980s NCEs
Estimated slope	1.14	1.18
Standard error	(.040)	(.075)
r^2	0.964	0.926
Number of NCEs	33	22

Here, for the first time, with samples that cover almost exhaustively the relevant population of domestically developed and approved new chemical entities in their time frames, we find slope values in the tail exceeding the unit threshold below which Pareto-Levy first moments are asymptotically infinite.

Figure 9 plots the NCE data for the 1970s on log normal probability coordinates. As with the Harvard invention bundle sample, the plot deviates visibly from linearity. Its consistent downward concavity suggests more equality among the observations, and hence less skewness, than one would expect if the data conformed to a log normal process. The computation of moments for logarithms of the quasi-rent values for both the 1970s and 1980s samples (including all positive values) leads to the following coefficients:

	1970s NCEs	1980s NCEs
$\sqrt{\beta_1}$ (skewness)	–0.54	–123.16
β_2 (kurtosis)	3.27	777.31

FIGURE 9



The skewness coefficient for the 1970s differs from the zero value associated with log normality at the 0.05 level; the kurtosis coefficient is significantly different from 3.0 only at the 0.20 level. For the 1980s, the significance thresholds are strongly exceeded. Both distributions exhibit skewness to the *left* relative to the log normal, and the 1980s distribution is highly leptokurtic, *i.e.*, tightly bunched but with long thin tails.

As one would expect, a Bryson graph (not reproduced here, but comparable to Figure 4) of the 1970s quasi-rent data yields a plot clearly convex downward. A Bryson graph of the 1980s data, on the other hand, has two cycles of concavity followed by convexity. The most one can conlclude on this basis is that the 1980s data are ill-behaved, possibly combining two quite different distribution functions. See D'AGOSTINO and STEPHENS [1986], pp. 15-18.

4 High-Technology Venture Firm Startups

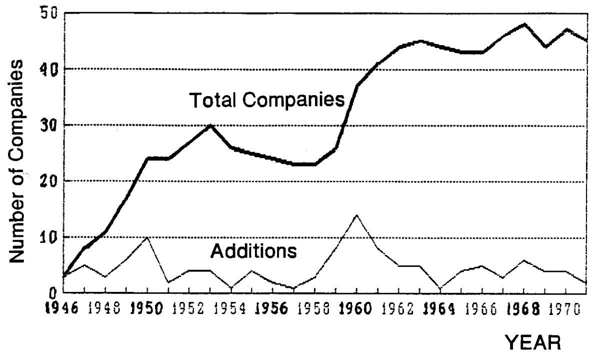
An institution that contributes enormously to America's prowess in high-technology fields is its venture capital industry. Hundreds of new firms are founded each year to develop and commercialize promising ideas emerging from university laboratories, independent inventors, and industrial corporations that for some reason chose not to pursue the opportunities internally. See ROBERTS [1991]. Typically, initial experiments and bench model development are supported using the technically trained entrepreneur's own funds and seed money raised from acquaintances (who as high-technology "angels" may sustain many such early investments). When this low-cost preliminary activity yields promising results, the fledgling enterprise turns to a high-technology venture fund for financial support, which ranges from a few hundred thousand to several million dollars. The venture capital fund raises money from an array of investors – in the industry's early history, from wealthy individuals, but more recently, from pension funds and university endowments – and attempts to pool its risks by investing in dozens of startup enterprises. If an individual venture succeeds in marketing one or more new products with good prospects, it "goes public" – *i.e.*, it floats an initial public offering (IPO) of its common stock; or (somewhat more frequently) its investors sell out their shares to a well-established company. The venture fund investors then "cash in" their proceeds or reinvest them in the new publicly-traded company shares.

The first modern U.S. high-technology venture capital fund was the American Research and Development Corporation (ARDC), founded in Boston shortly after the close of the Second World War. Figures 10a and 10b, drawn from WILLMANN [1991], trace ARDC's early portfolio history. Figure 10a shows the number of individual startup companies in which ARDC invested annually (light dotted line) and the total number of companies in its portfolio (solid line). During the 1950s, its portfolio contained from 23 to 30 companies. Its investment target count rose into the mid 40s by the 1960s. Figure 10b traces the net value of ARDC's investment portfolio. During the mid-1950s, a few successes (such as High Voltage Engineering Company and Airborne Instruments Inc.) fueled an appreciable portfolio value increase. In 1966, however, the portfolio value exploded. By decomposing the portfolio into two parts – Digital Equipment Company (DEC) and more than 40 other companies – Figure 10b shows that most of the increase was attributable to ARDC's $70,000 investment (in 1957) in DEC. DEC's great success came with the introduction of the first time-sharing computer, the PDP-6, in 1964, and a powerful but inexpensive minicomputer, the PDP-8, in 1965. An initial public offering of DEC's common stock was floated on the New York Stock Exchange in 1966.

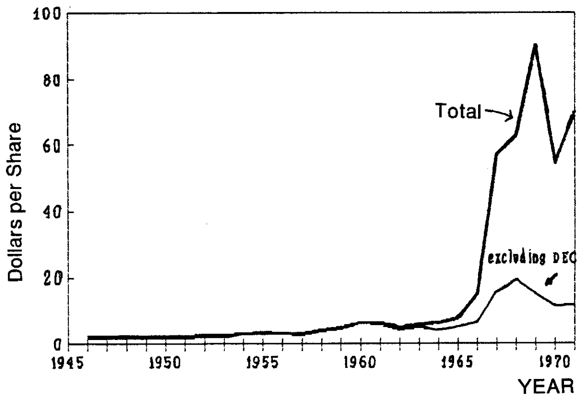
From the history of ARDC, we see considerable skewness in the returns from high-technology investments and the volatility such skewness can impart, despite venture investors' attempts to hedge against risk by forming sizeable portfolios. To determine how well the history of ARDC generalizes, two additional data sources were tapped.

One study of venture capital performance was conducted by the leading source of information on U.S. venture funds, Venture Economics, Inc. (1988, Chapter II). Venture Economics analyzed the success of 383 individual startup company investments made by 13 U.S. venture portfolios between 1969 and 1985 whose investment cycles had been largely completed by

FIGURE 10



(a)



(b)

1986 [16]. Figure 11 arrays the individual investments by the multiple of terminal value relative to original investment outlays. The 26 individual startups returning ten times or more the funds' initial stakes accounted for 49 percent of total terminal portfolio values. More than a third of the ventures returned less than their original fund investments. Figure 12 plots the distribution function on double log coordinates in two ways, one (solid line) using interval average values per portfolio investment as the horizontal axis variable (and assuming "total losses" to return 10 percent of the initialinvestment); the other using the lower threshold values for the intervals. Since the average return in the highest-return interval was reported to be 21.6 times original investment values, the threshold approach suppresses important information on the distribution's upper tail. With both methods, the distribution function is concave, not linear as implied by

16. During the early 1980s, venture capital funds began to invest in real estate deals, leveraged buyouts, and other targets as well as high-technology startups. The fraction of the investments attributable to genuinely high-technology startups in the portfolios analyzed here was not disclosed.

FIGURE 11

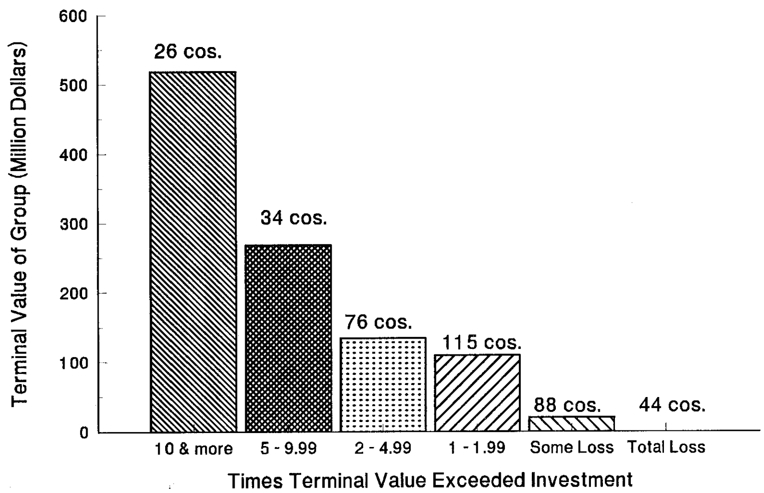
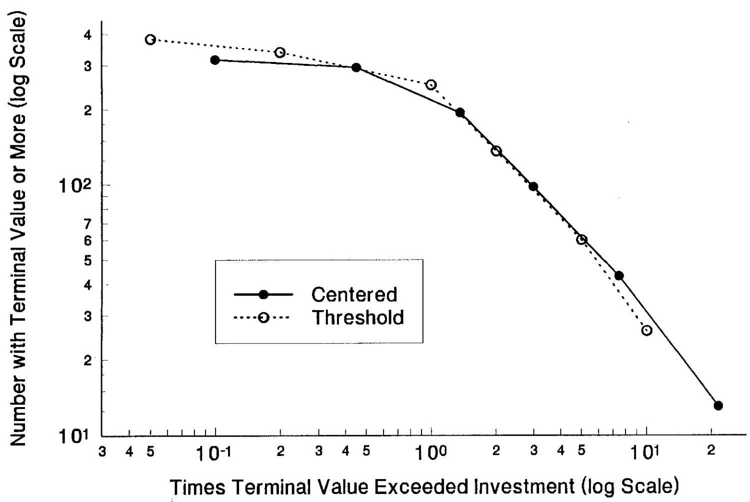


FIGURE 12



the Pareto-Levy hypothesis. If a linear regression is forced upon the full centered observation data set by least squares, the implied slope value is 0.60. However, if the regression is limited to the four highest-value (*i.e.*, right-hand tail) interval means, the fitted slope coefficient is 0.974, with standard error of 0.044 and r^2 of 0.996.

Figure 13 summarizes the results of a similar study by a San Francisco venture capital house, Horsley Keogh Associates (1990). Included in the analysis were 670 distinct investments (totalling $496 million) in 460 companies made between 1972 and 1983 by 16 venture capital partnerships.

FIGURE 13

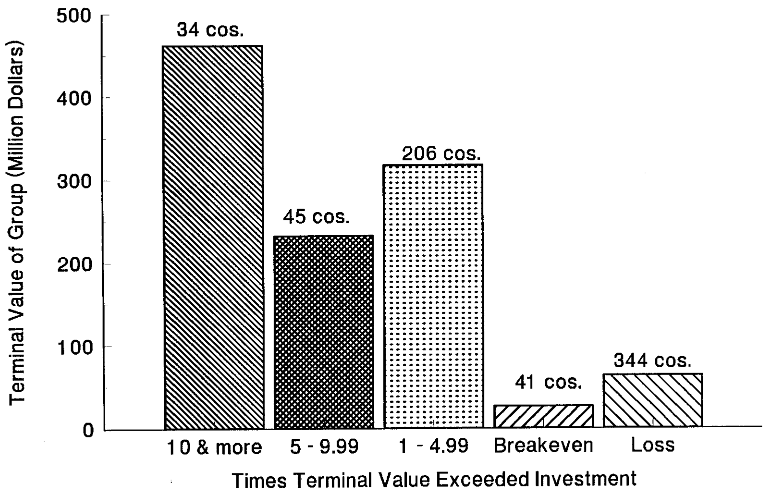
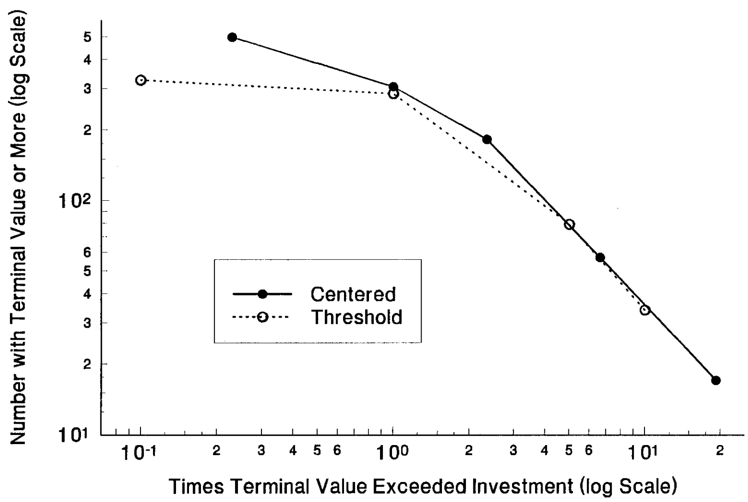


FIGURE 14



The ultimate portfolio value was calculated as of December 1988, at which time the funds had distributed to their partners $822 million and retained assets of $278 million. The 34 investments (*i.e.*, five percent of the total) that yielded ten times or more their original value contributed 42 percent of the portfolios' total terminal value. Slightly more than half of the investments entailed some loss. Figure 14 plots the distribution function on double-log coordinates, again using both mean values (solid line) and interval threshold values (dotted line). (Investments in the most lucrative interval returned on average 19.25 times their initial value.) Again we find concavity over the full range of observations, but near linearity in the right-hand tail. For all centered data points, the fitted least-squares line has an α of 0.77. However, for the four highest-value centered observations, the slope is 0.998, its standard error 0.083, and $r^2 = 0.986$.

5 Ongoing Empirical Research

To supplement the insights achieved through the studies summarized above, two more ambitious empirical research projects were in their initial stages at the time this paper was originally presented. Both remain incomplete at the time of revision. However, sufficient progress has been made to report preliminary insights.

One project, carried out by research associate Jörg Kukies, analyzes the stock price histories of 131 U.S. high-technology companies, initially backed by venture capital funds, that floated initial public stock offerings between 1983 and 1986. The 131 companies are believed to be an exhaustive sample of such IPOs in the relevant time frame. Monthly changes in their common stock values were tracked to the end of 1995 or to disappearance of the companies through merger or liquidation. The stock values for companies with a continuing presence evolve over a period of ten to twelve years in what was evidently a random walk, with frequent path crossovers and with a few big "winners" emerging among the much more numerous humdrum performances. When cross-sectional size distributions are plotted, they become increasingly skew over time, but after ten years remain concave downward on doubly logarithmic coordinates.

A second project, conducted jointly with Dietmar HARHOFF of the Center for European Economic Research in Mannheim, Germany, probes the tail of the distribution of values associated with patents applied for in 1977 and eventually issued by the German Patent Office. From that population 4,349 patents, including 1,435 patents of domestic German origin and 896 of U.S. origin, paid all maintenance fees and expired after running their full statutory 18-year term in 1995. The German maintenance fees, it is worth noting, are among the highest and most progressive in the world. Through mail and telephone surveys, rough preliminary discounted present value estimates were obtained from German and U.S. companies holding the full-term patents. Pareto plots of the survey response data showed only slight concavity relative to doubly logarithmic coordinates. The estimated slope value for the bounded German data (with 772 responses) was 0.404, with standard error 0.039 and r^2 of 0.973. For 223 U.S. patents linked to equivalent full-term German patents, the estimated slope value was 0.31, with standard error of 0.016 and r^2 of 0.984. The U.S. patents exhibit more skewness than the German sample, at least in part because they were three times filtered – once for application both in Germany and the United States, then for being issued in both jurisdictions, and then for renewal to full term in Germany. Interviews were held with 73 German companies holding patents reported in the first-stage survey to have had a value in 1980 exceeding DM 5 million. More precise profitability estimates obtained through these interviews ranged from less than DM 5 million (*i.e.,* where survey responses were exaggerated) to more than DM 1 billion. The size distribution of patent values in this "tail of tails" was concave downward, rather than exhibiting the linearity on doubly logarithmic coordinates associated with a Pareto-Levy distribution.

6 Implications

It seems clear that there are important regularities in the size distribution of profit returns from technological innovation. The distributions uniformly exhibit some downward concavity on doulby logarithmic coordinates. The Pareto-Levy hypothesis is not strongly supported. However, the concavity is greater – *i.e.,* there is less skewness – for whole innovations (such as new drug chemical entities) and for the fortunes of new high-technology firms as integrated entities than for individual patents. Since new high-technology firms often hold numerous patents and since innovations are commonly fenced in by a multiplicity of patents, complementarities among individual patents would appear to eliminate some of the skewness.

The variance of returns from portfolios of inventions, both at the level of the individual firm and on a broader macroeconomic plane, depends critically on the shape of the most valuable tail of the size distribution. It is for that reason that I have reported separately log-log slope values for the right-hand tail of the distribution. In most cases the tail plots on Pareto coordinates are very close to linear, with slope values that range from 0.41 (for our survey of U.S. patents with full-term German counterparts) to unity (for high-technology venture investments) and slightly more than unity (for new drug chemical entities). When $\alpha = 1$, the Pareto distribution coincides with the Yule distribution, a variant of the log normal distribution. See IJIRI and SIMON (1977, p. 75). Simulation analyses reveal that over a plausible range of α values, it is quite difficult to distinguish between Pareto and log normal distributions on the basis of right-hand tail observations. Another simulation analysis (to be reported in a separate paper) reveals that, given the quasi-rent size distribution found for new drug chemical entities introduced during the 1970s, which closely approximates log normality, aggregated profit returns are quite unstable even when portfolios as large as those held by all innovating pharmaceutical firms together are formed.

That there are persistent regularities in the distribution of profit outcomes from innovation suggests that more or less well-defined behavioral processes generate those outcomes. Subsequent research will attempt to identify those processes.

One possibility, emphasized inter alia by EATON and KORTUM [1994], is that some Supreme Power regularly strews about the industrial landscape a distribution of raw profit potentials for technological innovation that is highly skew, just as the distribution of petroleum reservoirs within a land mass, and hence the opportunity for profiting from exploratory well drilling, is believed to be log normal. See ADELMAN (1972, p. 35). The profit opportunities from innovation might be roughly proportional to the size of markets, which, we know from the distribution of sales or value added across conventionally defined industries, is skew-distributed.

An alternative or complementary hypothesis is that the distribution of returns from innovation results from some variant of a Gibrat process, under which numerous chance events interact multiplicatively, reducing the profits actually realized from an innovative potential which may or may not be skew initially. See GIBRAT [1931]. If P_0 is the initial potential and ε_i is the i-th stochastic multiplier affecting the amount of value that can be appropriated by an innovator, the ultimate innovator's quasi-rent is:

$$V = P_0\,\varepsilon_1 \,...\, \varepsilon_i \,...\, \varepsilon_n; \tag{2}$$

where the typical ε consists of an expected value less than unity plus an error component. The ϵ's reflect inter alia the initial probability of technical success, the time at which the innovator arrives on the market with its new product and hence the strength of first-mover advantages, the strength of the innovator's patent protection, the finesse with which initial marketing efforts are conducted (crucial e.g. in the competition between anti-ulcer drugs TAGAMET and ZANTAC), and the extent to which the market is fragmented by imitators in each subsequent year of commercial sales (which is likely to be correlated with the strength of first-mover advantages). Taking logarithms, we have:

$$\log V = \log P_0 + \log \varepsilon_1 + ... + \log \varepsilon_i + ... + \log \varepsilon_n; \tag{3}$$

which, by the central limit theorem, is normally distributed with sufficiently large n. Thus, under the logic of Gibrat's law, a log normal distribution of V might be anticipated. The "sufficiently large n" assumption may be crucial, since patents, new drugs, and the startup phases of new high-technology ventures are bounded in time – perhaps too much so for a Gibrat process to converge on log normality.

When, contrary to the standard Gibrat model's assumptions, the initial population is not fixed and entry is concentrated in the low-value tail, a Yule distribution emerges instead of a log normal distribution. See IJIRI and SIMON [1977]. At the other extreme, long high-value tails could result from stochastic processes with path dependence, *i.e.*, in which an early head start leads to increasing dominance over time. See e.g. DEVANY and WALLS [1996]. Month-to-month stock price movements for the 131 high-technology companies that floated initial public stock offerings are being analyzed for the deviations from a random walk that would reveal path dependence. Interviews with the holders of particularly valuable German patents will also provide evidence on the extent of path dependence.

Thus, there are plausible links between the stochastic search, experimentation, and market penetration processes associated with innovation and the kinds of profit return distributions found here. Further theory-based and simulation research will attempt to identify plausible stochastic processes generating the observed profit return distributions. It is hoped that in this way a deeper understanding of the economics of technological innovation will follow.

● References

ADELMAN, Morris A. (1972). – "*The World Petroleum Market*", Johns Hopkins University Press.

ALDRIDGE, S. (1996). – "*The Thread of Life: The Story of Genes and Genetic Engineering*", Cambridge University Press.

D'AGOSTINO, R. B., STEPHENS, M. A. (1986). – "*Goodness-of-Fit Techniques*", New York: Marcel Dekker.

BRYSON, M. C. (1974). – "Heavy-Tailed Distributions: Properties and Test", *Technometrics*, vol. 16, pp. 61-68.

DEVANY, A., WALLS, W. D. (1996). – "Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry", *Economic Journal*, vol. 106, pp. 1493-1514.

DIMASI, J., HANSEN, R. W., GRABOWSKI, H., LASAGNA, L. (1995). – "Research and Development Costs for New Drugs by Therapeutic Category", *PharmacoEconomics*, vol. 7, pp. 152-169.

EATON, J., KORTUM, S. (1994). – "International Patenting and Technology Diffusion", *National Bureau of Economic Research working paper no. 4931*.

GIBRAT, R. (1931). – *Les Inégalités Économiques*, Paris: Recueil Sirey.

GRABOWSKI, H., VERNON, J. (1990). – "A New Look at the Returns and Risks to Pharmaceutical R&D", *Management Science*, vol. 36, pp. 804-821.

GRABOWSKI, H., VERNON, J. (1994). – "Returns on New Drug Introductions in the 1980s", *Journal of Health Economics*, vol. 13, pp. 383-406.

HANSON, N. R. (1958). – "*Patterns of Discovery*", *Yale University Press.*

Horsley Keogh Associates, *Horsley Keogh Venture Study*, San Francisco: *Privately distributed.*

IJIRI, Y., SIMON, H. A. (1977). – "*Skew Distributions and the Sizes of Business Firms*", Amsterdam: North-Holland.

LANJOUW, J. O. (1992). – "The Private Value of Patent Rights in Post WWII West Germany", *working paper, Yale University.*

LEVIN, R. C., KLEVORICK, A., NELSON R. R., WINTER, S. C. (1987). – "Appropriating the Returns from Industrial Research and Development", *Brookings Papers on Economic Activity*, no. 3, pp. 783-820.

LU, Z. J., COMANOR, W. S. (1998). – "Strategic Pricing of New Pharmaceuticals", *Review of Economics and Statistics*, vol. 80, pp. 108-118.

MANDELBROT, B. (1963). – "New Methods in Statistical Economics", *Journal of Political Economy*, vol. 71, pp. 421-440.

NORDHAUS, W. D. (1989). – "Comment", Brookings Papers on Economic Activity: Microeconomics, pp. 320-325.

PAKES, A., SCHANKERMAN, M. (1984). – "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources", in Zvi Griliches, ed., *R&D, Patents, and Productivity (University of Chicago Press)*, pp. 73-88.

PAKES, A. (1986). – "Patents as Options: Some Estimates of the Value of Holding European 1986 Patent Stocks", *Econometrica*, vol. 54, pp. 755-784.

ROBERTS, Edward B. (1991). – *Entrepreneurs in High Technology* (New York: Oxford University Press).

SANDERS, B., ROSSMAN, J., HARRIS, L. J. (1958). – "The Economic Impact of Patents", *Patent, Trademark, and Copyright Journal*, vol. 2, pp. 340-363.

SANDERS, B. (1964). – "Patterns of Commercial Exploitation of Patented Inventions by Large and Small Corporations", *Patent, Trademark, and Copyright Journal*, vol. 8, pp. 51-93.

SCHANKERMAN, M., PAKES A., (1986) – "Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period", *Economic Journal*, vol. 97, pp. 1-25.

SCHANKERMAN, M. (1991). – "How Valuable Is Patent Protection? Estimates by Technology Field Using Patent Renewal Data", *National Bureau of Economic Research Working Paper no. 3780.*

SCHERER, F. M. (1965). – "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions", *American Economic Review*, vol. 55, pp. 1097-1123.

SCHERER, F. M. (1977). – The Economic Effects of Compulsory Patent Licensing (New York University Graduate School of Business Administration, *Monograph Series in Finance and Economics*).

SCHERER, F. M. (1983). – "The Propensity To Patent", *International Journal of Industrial Organization*, vol. 1, pp. 107-129.

SCHERER, F. M. (1986). – "On the Current State of Knowledge in Industrial Organization", in H. W. de Jong and W. G. Shepherd, eds. *Mainstreams in Industrial Organization*, Book I (Dordrecht: Kluwer), pp. 5-22.

Venture Economics, Inc., (1988) *Venture Capital Performance* (Boston).

WILLMANN, H. (1991). – "Innovation in the Venture Capital Industry : A Study of American Research and Development Corporation", *term paper*, John F. Kennedy School of Government, Harvard University.

WINSTON-SMITH S. (1996). – "The Cohen-Boyer Patent: A Case Study", *term paper*, John F. Kennedy School of Government, Harvard University.