NOTES

A STOCHASTIC MODEL OF SUPERSTARDOM: AN APPLICATION OF THE YULE DISTRIBUTION

Kee H. Chung and Raymond A. K. Cox*

Abstract—This study employs a stochastic model developed by G. Udny Yule and Herbert A. Simon as the probability mechanism underlying the consumer's choice of artistic products and predicts that artistic outputs will be concentrated among a few *lucky* individuals. We find that the probability distribution implied by the stochastic model provides an excellent description of the empirical data in the popular music industry, suggesting that the stochastic model may represent the process generating the superstar phenomenon. Because the stochastic model does not require differential talents among individuals, our empirical results support the notion that the superstar phenomenon could exist among individuals with equal talent.

Recently rigorous economic analyses have been applied to the so-called "superstar phenomenon," wherein a relatively small number of people dominate the activities in which they are engaged and earn enormous amounts of income. Extraordinary incomes earned by superstars may be driven by an allocative equilibrium in which markets reward talented people with increasing returns to ability. Or perhaps, the superstar phenomenon has nothing to do with the differential talent of individuals. For instance, the phenomenon may emerge as a result of certain consumer behavior. If enormous incomes earned by superstars are the markets' reward for their superior talent, the superstar phenomenon may be socially admissible. If, on the other hand, the source of their high incomes is not their talent, the skewness in income distributions caused by the phenomenon may be perceived as unequitable by society. It is the purpose of this paper to shed further light on the possible sources of the superstar phenomenon.

Rosen (1981) suggests that much of the superstar phenomenon can be explained by convexity of sellers' revenue functions since the convex revenue function implies that the distribution of rewards is more skewed than the distribution of talent (i.e., small differences in talent are magnified into disproportionate levels of success). Rosen shows that the convexity of revenue functions and the extra skew it imparts to the distribution of earnings can be obtained by imperfect substitution (i.e., lesser talent is a poor substitute for greater talent) among different sellers. Rosen also demonstrates that the joint consumption technology (i.e., a performer puts out more or less the same effort in front of audiences of ten or one thousand), combined with imperfect substitution, can explain the marked concentration of output on those who have the most talent.

In a similar vein, MacDonald (1988) presents a dynamic version of Rosen's superstar model. He shows that in equilibrium only the young enter the occupation and earn low incomes playing to small crowds, and only the successful stay on. Overall, there are few stars in the industry but as a group they serve a large fraction of the audience and earn an even larger share of the rewards. In order to test the empirical significance of the theory of superstar, Hamlen (1991) examined the relationship between talent (proxied by voice quality) and success (measured by record sales) in the popular music industry while controlling for other factors such as gender, race, the type of music, and the duration of career. Although empirical results show that consumers recognize quality, the estimated elasticity of record sales with respect to voice quality is less than unity, repudiating the implication of the Rosen-MacDonald theory of superstar.

This paper examines the phenomenon of superstar from a perspective which is significantly different from that of the above studies. Specifically, this study employs a stochastic model of Yule (1924) and Simon (1955) as the probability mechanism underlying the consumer's choice of artistic products (e.g., records or motion pictures) and predicts that artistic outputs will be concentrated among a few *lucky* individuals. We find that the probability distribution implied by the stochastic model provides an excellent description of the empirical data in the popular music industry. Because the stochastic model does not require differential talents among individuals, our empirical results suggest that the superstar phenomenon could exist among individuals with equal talent. Hence our results are, in spirit, similar to those of Adler (1985).

Received for publication March 25, 1993. Revision accepted for publication February 7, 1994.

* University of Memphis and Central Michigan University, respectively.

The authors are grateful to two anonymous referees for their very valuable comments. The authors would also like to thank James Irwin, R. Gene Stout, and M. Richard Sussman for their useful comments on earlier versions of the paper.

I. A Stochastic Model of Superstardom

The Rosen-MacDonald theory of superstar centers on an implicit comparison of success relative to the differences in talent. In this section we show that the phenomenon of superstar does not require differential talents among individuals using the stochastic model of Yule (1924) and Simon (1955) as a representation of the consumer's choice behavior. Simon suggests that a variety of sociological, biological, and economic phenomena are driven by certain probability mechanisms. Specifically, he shows that a wide range of empirical data (e.g., distributions of incomes by size, distributions of cities by population, distributions of biological genera by number of species, and distributions of scientists by number of papers published) conforms well to a class of distributions which can be obtained from stochastic processes similar to those yielding negative binomial or log series distributions. This class of distributions is given by (see Simon (1955), p. 426)

$$f(i) = \psi B(i, \rho + 1), \tag{1}$$

where ψ and ρ are constants and $B(i, \rho + 1)$ is the beta function of i and $\rho + 1$, i.e.,

$$B(i, \rho + 1) = \int_0^1 \lambda^{i-1}(1 - \lambda)^{\rho} d\lambda = \Gamma(i)\Gamma(\rho + 1)/\Gamma(i + \rho + 1) \quad (0 < i; 0 < \rho < \infty). \tag{2}$$

Since the class of distributions represented by the expression (1) was first derived by G. Udny Yule (1924), the distribution carries his name.

In essence, the stochastic process that would lead to the Yule distribution can be characterized as follows. Since the main thrust of this paper is to examine whether the Yule distribution can describe the relative record sales of artists in the popular music industry, we portray the process in such a context. For simplicity and without loss of generality suppose that each consumer buys the same number of records, n, and that the records are bought in the following order: All consumers first buy sequentially one record each. After the last consumer has bought her first record, the process repeats itself with the second record, and so on. Of course for each consumer, $\text{record}_s \neq \text{record}_t$, $s,t = 1, 2, \ldots, n$. Then, the following two assumptions depicting the probability mechanism underlying the consumer's choice of her next record yield the Yule distribution:[1]

ASSUMPTION I: *The probability that consumer $k + 1$ chooses a record that was already chosen by exactly i of the k previous consumers is proportional to i.*

ASSUMPTION II: *There is a constant probability, δ, that consumer $k + 1$ chooses a record that was not yet chosen by any of the previous k consumers.*

In spirit, the process implied by these assumptions is similar to the superstar generating process suggested by Adler (1985). Adler suggests that the superstar phenomenon exists where consumption requires knowledge. He claims that the need to discuss with other knowledgeable individuals to become familiar with an artist's work as a prerequisite to the ultimate consumption (appreciation) of the artist's work is an essential element in understanding the phenomenon. He argues that consumers minimize the cost of searching for knowledgeable discussants by choosing the most popular artist. Adler suggests that consumers are better off by patronizing the star when either other artists are not cheaper by more than the savings in search costs or other artists are not sufficiently better than the star. Probability mechanisms underlying the superstar generating process proposed by Adler can be summarized as follows: Suppose that consumers believe at first that all artists are equally likely to become stars, and that each consumer picks one artist at random. Assume further that consumers live n periods and revise their prior distributions after each period. If there were a slight majority of consumers that select an artist as their choice, that artist would snowball into a star because after each period the majority would increase. In other words, if at any period of time an artist had a market share only marginally larger than everybody else, this share would increase steadily, and ultimately the artist becomes a star.

Notice the close proximity between the assumptions underlying the Yule distribution and the superstar model proposed by Adler. It is not clear whether the above assumptions are a realistic representation of the process creating superstars in the popular music industry. Ultimately, the reasonableness of these assumptions can only be judged by the prescriptive power of their implication, i.e., the Yule distribution. Considering the ubiquity of the distribution in a wide range of social and economic data, however, we conjecture that it may have some predictive content in describing the superstar phenomenon. In the following sections, we examine whether the distribution can describe the cross-sectional distribution of artistic output (measured by the number of gold-records) among contemporary performers of popular music.

II. Empirical Results

A. Data Description

Data for the present study are from the Gold-Record Awards given by the Recording Industry Association of America (RIAA) Inc. compiled and reported

[1] For a detailed description of this stochastic process, see Simon (1955), pp. 427–433.

in Stambler (1989). We use the number of gold-records by performers as the measure of their artistic success.[2] This measure is used because the designation as gold-records by the RIAA Inc. ensures certain minimum sales volumes and hence our measurement unit is comparable across artists, and further, because the number of gold-records would closely proxy the monetary success of artists.[3,4]

Table 1 shows the frequency distribution of performers by the number of gold-records for the period 1958–1989.[5] Among 1,377 performers who earned at least one gold-record, 668 performers (48.5%) have one gold-record, 244 performers (17.7%) have two gold-records, 119 performers (8.6%) have three gold-records, and only 213 performers (15.5%) have more than five gold-records. In total, these performers produced 4,408 gold-records. Finally, it is interesting to note that although only 149 performers (10.8%) have more than seven gold-records, these performers collectively earned 1,900 (43.1%) gold-records, revealing a high degree of output concentration among top performers.

TABLE 1.—FREQUENCY DISTRIBUTION OF PERFORMERS BY THE NUMBER OF GOLD-RECORDS FOR THE PERIOD 1958–1989

Number of Gold-Records	Number of Performers		Percentage of Performers	
	Actual	Predicted[a]	Actual	Predicted[a]
1	668	689	48.51	50.00
2	244	230	17.72	16.67
3	119	115	8.64	8.33
4	78	69	5.66	5.00
5	55	46	4.00	3.33
6	40	33	2.90	2.38
7	24	25	1.74	1.79
8	32	19	2.32	1.39
9	24	15	1.74	1.11
10	14	13	1.02	0.91
11	16	10	1.16	0.76
12	13	9	0.94	0.64
13	11	8	0.80	0.55
14	5	7	0.36	0.48
15	4	6	0.29	0.42
16	4	5	0.29	0.37
17	2	5	0.15	0.33
18	7	4	0.51	0.29
19	2	4	0.15	0.26
20	3	3	0.22	0.24
21	1	3	0.07	0.21
22	3	3	0.22	0.20
23	1	2	0.07	0.18
24	1	2	0.07	0.17
29	1	2	0.07	0.11
34	1	1	0.07	0.08
36	1	1	0.07	0.08
37	1	1	0.07	0.07
45	1	1	0.07	0.05
46	1	1	0.07	0.05

[a] These columns give the predicted number (percentage) of performers in each category based upon the distribution (3), i.e., $f(i) = 1/i(i + 1)$.

B. *Empirical Testing*

Simon (1955) suggests that the Yule distribution provides a good fit to various empirical data particularly when the value of ρ is equal to one. Hence in this paper we assume, as an empirical approximation, that the probability that consumer $k + 1$ chooses a record which was not yet chosen by any of the previous k consumers is small ($\delta \approx 0$), so that ρ is close to 1 since $\rho = 1/(1 - \delta)$. For this case the distribution (1) can be approximated by the following form (see Simon (1955), p. 426):

$$f(i) = 1/i(i + 1), \quad \Sigma f(i) = 1, \tag{3}$$

where $f(i)$ may, in the context of this study, be labeled as the proportion of performers with i gold-records and Σ denotes the summation over $i = 1$ to ∞. Hence the proportion of performers with one gold-record should be:

$$f(1) = 1/1(1 + 1) = 0.500. \tag{4}$$

Likewise, the proportions of performers with two, three,..., and i gold-records should be:

$$f(2) = 1/2(2 + 1) = 0.167, \tag{5}$$

$$f(3) = 1/3(3 + 1) = 0.083, \tag{6}$$

$$\vdots$$

[2] If an individual performer switched to another group, or became a solo, the counting of gold-records begins anew. For example, Paul McCartney was a member of the Beatles and is currently a member of Wings. Each of these is treated as a unique performer.

[3] From 1958 until January 1, 1975, the requirement for a gold-album certification was a minimum of $1 million in manufacturer's dollar volume based on 33 1/3% of the list price of each LP and/or tape sold. Since 1975, album certification has been based on a minimum sale of 500,000 units with a multirecord or tape package counting as one unit. Since late 1983, sales of compact discs also have been combined with LP and/or tape sales. In addition, manufacturer's dollar volume must be at least $1 million based on 33 1/3% of the list price of each unit sold. For a gold-single certification, a minimum sale of one million copies is required with disco/dance-music records (12-inch 33s or 45s with one selection per side) counted as two units. Sales of 12-inch singles may be combined with counterpart 7-inch discs if the repertoire on both sides is identical as to artist and title.

[4] One may use the dollar volume of record sales as the measure of success. Since the Yule distribution is defined over discrete variables, however, we use the number of gold-records as the measure of success. Had we used the dollar sales volume, we would have been forced to make certain arbitrary categorization of dollar sales volume into different levels of success.

[5] See also appendix A for the list of performers who have earned the most gold-records. Some artists on the list are deceased and cannot release new records. Some performers are young with plenty of time to increase their output and disbanded groups may get back together to increase their output. Nevertheless, the list is of interest to see who were the superstars in popular music during the 1958–1989 period.

and

$$f(i) = 1/i(i+1). \tag{7}$$

The last two columns in table 1 compare actual and predicted proportions of performers with different numbers of gold-records. The results show that the distribution (3) provides an excellent description of the actual frequency distribution. Notice that the actual proportion of performers with just one gold-record is 48.5% which is remarkably close to the prediction (i.e., 50%) made by the Yule distribution. To test whether the Yule distribution describes the observed data, we perform the Chi-square goodness-of-fit test using the actual and predicted number of performers in table 1. Since the Chi-square test requires that the predicted (i.e., theoretical) number of observations in each category should be at least five, we used only the relevant sample observations in table 1 (i.e., the number of gold-records ≤ 17) in calculating the Chi-square statistic, Q:

$$Q = \sum_{j=1}^{17} (\text{Actual}_j - \text{Predicted}_j)^2/\text{Predicted}_j = 30.2.$$

Since the Chi-square statistic is less than $\chi^2_{1-\alpha}(k) = \chi^2_{.99}(16) = 32.0$, we cannot reject the hypothesis that the Yule distribution with $\rho \approx 1$ represents the process underlying the superstar phenomenon in the recorded popular music industry.

Finally, it is interesting to note that the values of $f(2)/f(1)$ and $f(1)$ in the popular music industry are 0.365 and 0.485, respectively, which are surprisingly similar to the findings of earlier studies that the values of $f(2)/f(1)$ and $f(1)$ are generally in the neighborhood of one-third and one-half, respectively, in the cases of word frequencies, publications, and biological genera (see Simon (1955) for details). It is hardly imaginable that there exists any commonality among word storage in human minds, blood cells on a microscope slide, and the fatal attraction of Sgt. Pepper's Lonely Hearts Club Band. On this account, we simply join Herbert Simon's observation that "Its appearance is so frequent, and the phenomena in which it appears so diverse, that one is led to the conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms (Simon (1955), p. 425)."

III. An Alternative Test of the Yule Distribution

This section presents an alternative test of the Yule distribution as the underlying probability mechanism of the superstar phenomenon. Note first that $\Gamma(i)/\Gamma(i + c) \approx 1/i^c$ for any constant c when i is much greater than c (Titchmarsh (1939), p. 58). Thus the distribution (1) can be approximately written as

$$f(i) = \psi\Gamma(\rho+1)i^{-(\rho+1)}. \tag{8}$$

Since $f(1) = \psi\Gamma(\rho+1)1^{-(\rho+1)} = \psi\Gamma(\rho+1)$, the distribution (8) can be rewritten as:[6]

$$f(i) = f(1)i^{-(\rho+1)}, \tag{9}$$

which upon rearrangement yields

$$f(i)/f(1) = i^{-(\rho+1)}. \tag{10}$$

Finally, taking the log of both sides of (10), we obtain

$$\log[f(i)/f(1)] = -(\rho+1)\log(i). \tag{11}$$

This modified specification of the Yule distribution is tested by applying the following regression model to the frequency distribution data in table 1:

$$\log[f(i)/f(1)] = \alpha + \beta\log(i) + \epsilon. \tag{12}$$

If the Yule distribution with $\rho \approx 1$ is a reasonable representation of the relative success of performers, we would expect that empirical estimates of the intercept α and the slope β in (12) should not be significantly different from zero and negative two, respectively. When we apply the above regression model to the empirical frequency distribution in table 1, we obtain the following results:

$$\log[f(i)/f(1)] = \underset{(1.89)}{0.474} - \underset{(-21.23)}{1.966}\log(i)$$

Adjusted-$R^2 = 0.939$, F-value $= 450.81$,

where values in parentheses are t-statistics.

The above results show that indeed the Yule distribution is an excellent abstraction of the distribution of numbers of gold-records among different artists. It explains nearly 94% of the empirical distribution of gold-records among performers. Furthermore, estimated intercept and slope are not statistically different from zero and negative two, respectively, at the 1% significance level.[7] Therefore, we conclude that although there appears to be no common thread between the distribution of scientists by the number of papers published and the distribution of performers by the number of gold-records they earned, probability mechanisms underlying these phenomena may be quite similar.

IV. Summary and Concluding Remarks

Casual empiricism suggests that there exists a marked skewness in the distribution of output and earnings

[6] The prevalence of equation (9) as the underlying bibliometric distribution in various academic disciplines is well known. For example, Cox and Chung (1991) find that equation (9) provides an excellent fit in the economics literature with an exponent of 1.84.

[7] The standard error of the estimate of slope coefficient is 0.093. Hence the t-statistic for testing the null hypothesis H_0: $\beta = -2$ is $-(2 - 1.966)/0.093 = -0.366$.

among individuals in various social-economic fields. Several recent studies have examined this so-called superstar phenomenon, and suggested that much of this phenomenon can be explained by certain consumption technologies and imperfect substitutions among different sellers. Existing empirical evidence, however, appears to be inconsistent with the prediction of these studies.

This paper has examined the phenomenon of superstar from a perspective which is significantly different from that of earlier studies. Specifically, this study views the superstar phenomenon as an implication of the probabilistic mechanism underlying the record-buying behavior of consumers. Empirical results suggest that the stochastic process similar to that yielding negative binomial or log series distributions may represent the process generating the superstar phenomenon. Because the stochastic model does not require differential talent levels among individuals, our empirical results suggest that the superstar phenomenon could exist among individuals with equal talent. To the extent that very large incomes of superstars are driven by sheer fortune rather than by their superior (if any) talent, the superstar phenomenon may result in a socially unequitable wealth distribution.

REFERENCES

Adler, Moshe, "Stardom and Talent," *American Economic Review* 75 (Mar. 1985), 208–212.

Cox, Raymond A. K., and Kee H. Chung, "Patterns of Research Output and Author Concentration in the Economics Literature," this REVIEW 73 (Nov. 1991), 740–747.

Hamlen, William A. Jr., "Superstardom in Popular Music: Empirical Evidence," this REVIEW 73 (Nov. 1991), 729–733.

MacDonald, Glenn M., "The Economics of Rising Stars," *American Economic Review* 78 (Mar. 1988), 155–167.

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review* 71 (Dec. 1981), 845–858.

Simon, Herbert A., "On a Class of Skew Distribution Functions," *Biometrika* 42 (1955), 425–440.

Stambler, Irwin, *The Encyclopedia of Pop Rock and Soul* (New York: St. Martin's Press, 1989).

Titchmarsh, E. C., *The Theory of Functions* (London: Oxford University Press, 1939).

Yule, G. Udny, "A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S.," *Philosophical Transactions of the Royal Society* B. 213 (1924), 21–87.

APPENDIX A.—40 MOST SUCCESSFUL PERFORMERS

Artist	Number of Gold-Records
Beatles	46
Elvis Presley	45
Elton John	37
Rolling Stones	36
Barbra Streisand	34
Neil Diamond	29
Aretha Franklin	24
Chicago	23
Donna Summer	22
Kenny Rogers	22
Olivia Newton-John	22
Beach Boys	21
Bob Dylan	20
Earth, Wind, & Fire	20
Hall & Oates	20
Barry Manilow	19
Three Dog Night	19
Bee Gees	18
Carpenters	18
Creedence Clearwater Revival	18
John Denver	18
Kiss	18
Kool & the Gang	18
Rod Stewart	18
Andy Williams	17
Frank Sinatra	17
Billy Joel	16
Hank Williams, Jr.	16
Linda Ronstadt	16
Willie Nelson	16
Doors	15
Glen Campbell	15
Queen	15
REO Speedwagon	15
Anne Murray	14
Doobie Brothers	14
Jethro Tull	14
Johnny Mathis	14
O'Jays	14
AC/DC	13