

# market Shares Follow the Zipf Distribution

Hila Riemer

*University of Illinois at Urbana–Champaign*

Suman Mallik

*University of Illinois at Urbana–Champaign*

Devanathan Sudharshan

*University of Illinois at Urbana–Champaign*

## *Abstract*

The Zipf distribution is known to describe various natural phenomena, including city populations in the United States, frequency of English words in literature, immune system response in human beings, and certain aspects of Internet traffic. Using data from 70 markets, we show that the market shares by rank order follow the Zipf distribution. Our work makes a fundamental contribution in understanding the distribution of market shares, as we make no assumption about the order of entry and our results are valid for arbitrary number of competitors in market. We compare the predictions of our model with those from the analytical models in marketing literature. The comparison reveals that market share predicted by the Zipf model fits well with the predictions from the analytical models.

# **MARKET SHARES FOLLOW THE ZIPF DISTRIBUTION**

Hila Riemer  
hriemer@uiuc.edu

Suman Mallik  
mallik@uiuc.edu

D. Sudharshan  
sudharsh@uiuc.edu

Department of Business Administration  
University of Illinois at Urbana-Champaign  
350 Wohlers Hall  
1206 South Sixth Street  
Champaign, IL 61820

November 2002

# **MARKET SHARES FOLLOW THE ZIPF DISTRIBUTION**

## **ABSTRACT**

The Zipf distribution is known to describe various natural phenomena, including city populations in the United States, frequency of English words in literature, immune system response in human beings, and certain aspects of Internet traffic. Using data from 70 markets, we show that the market shares by rank order follow the Zipf distribution. Our work makes a fundamental contribution in understanding the distribution of market shares, as we make no assumption about the order of entry and our results are valid for arbitrary number of competitors in market. We compare the predictions of our model with those from the analytical models in marketing literature. The comparison reveals that market share predicted by the Zipf model fits well with the predictions from the analytical models.

**Keywords: Market Share, Market Structure, Zipf distribution**

## 1. INTRODUCTION

A recent report by Forrester Research about the consumer packaged goods (CPG) notes that the top 25 retailers sell more than half (53%) the volume generated by manufacturers like Kraft and General Mills, the balance is sold through thousands of other retailers. The report further notes that the top 10 retailers had a 43% share of the CPG market in the year 2000, while the top 11-25 retailers had a total of only 10% market share (Source: *Retailers Reshape CPG Trade*, Forrester Report, April 2001). Top two manufacturers of personal computers (HP and Dell) held about 30.4% share of US personal computer market for the second quarter of 2002 while the rest was divided among many different firms. IBM was the third largest manufacturer of personal computer with only 6.6% market share for Q2, 2002 (*Wall Street Journal*, October 15, 2002). So, is there a predictable distribution of market shares within a market? We seek to answer this question in this paper. Using data from 70 markets we show that market shares within a market are Zipf distributed and that the market share is a Zipf's law-form function of the rank order in the market. Since the answer to the question we posed is affirmative the result has managerial relevance. It allows a calculation of the expected share consequences of a gain or loss in rank as well as a calculation of how much the defense of a brand is worth.

The Marketing literature provides two broad insights about the market share distribution by relating market share to order of entry (cf. Kalyanaram et al. 1995). First, there exists a negative relationship between the order of entry and market share. This relationship has been shown to hold for both consumer and industrial products (e.g. Kerin et al. 1992; Urban et al. 1986; Kalyanaram and Urban 1992; Kalyanaram and Wittink

1994; Golder and Tellis 1993; Robinson and Fornell 1985; Robinson 1988; Parry and Bass 1990; Berndt et al. 1994; Brown and Lattin 1994; Huff and Robinson 1994; Glazer 1985; Lieberman 1989; Sullivan 1992; Mitchell 1991). The popular business press often documents similar findings. For example, Palm Inc., a pioneer in manufacturing and marketing personal digital assistants (PDA), has retained dominance in the hand-held-device hardware and software markets, despite attempts by rivals such as Microsoft to dislodge it. Research firm *International Data Corporation* estimates that Palm's devices had a leading 32.2% share in the world-wide hardware market for the second calendar quarter of 2002, while its software had about a 50% market share (*Wall Street Journal*, September 20, 2002). Second, the literature specifies the relationship between any entrant's market share and the pioneer's market share. This relationship can be written as  $MS_i = MS_1 / \sqrt{i}$ , where  $MS_i$  is the market share of the  $i^{\text{th}}$  entrant and  $MS_1$  is the market share of the pioneer (the first entrant in the market). The theoretical arguments in support of this relationship are that the expected incremental benefit from a new brand declines as the number of brands increases (Hauser and Wernerfelt 1990), and that the opportunity to serve unmet consumer needs declines as the number of brands increases (Prescott and Visscher 1977). This relationship has been proven to be correct for consumer packaged goods and prescription anti-ulcer drugs (Urban et al. 1986; Kalyanaram and Urban 1992; Berndt et al. 1994).

In addition to the relationships described above, research in marketing has shown (1) a higher survival rate for the pioneers than the early followers (Robinson and Min 2002) and (2) a negative relationship between the stages of entry by product life cycle and survival rate (Agarwal and Gort 1996; Agarwal 1996, 1997). These premises have

been supported in many studies and have been attributed to first mover advantage due to consumer risk aversion (Schmalensee 1982), pioneer's prototypicality (Carpenter and Nakamoto 1989), consumer learning (Kardes and Kalyanaram 1992), opportunity to gain intensive distribution (Porter 1980), scale economies (Schere and Ross 1990), and development of a broader product line (Prescott and Visscher 1977). In a recent study, Bohlmann et al. (2002) describe the conditions under which pioneers are more likely or less likely to have an advantage in the market.

The literature, however, does not provide any guidance about the market share distribution when the order of entry is unknown or is difficult to obtain. Using data from 70 markets we show that market shares within a market is a Zipf's law-form function of the rank order in the market. Our paper, thus, makes a fundamental contribution in understanding the distribution of market shares. Modeling the market share distribution is of importance to both the theoreticians and the practitioners. At a theoretical level, it allows us to study the dynamics of competition within a market. In practice, a model may assist a manager in predicting his/her expected market share.

The remainder of this paper is organized as follows. The next section provides a brief review of the Zipf distribution and describes our hypothesis. Section 3 describes our test data set and the statistical methodology and measures used to test our hypothesis. We present and discuss the results in Section 4. Section 5 contains concluding remarks.

## **2. HYPOTHESIS DEVELOPMENT**

We begin this section with a brief review of the Zipf distribution and then proceed to develop our hypothesis.

## 2.1 The Zipf Distribution

Zipf's Law (cf. Read 1988), named after the Harvard linguistic professor George Kingsley Zipf, describes a relationship between the size (i.e., magnitude) of a data value and its rank (i.e., its order in the series of numbers, when ordered by value). Zipf's law can be stated as

$$rx_{(r)} = \text{constant}, \quad (1)$$

where,  $r$  is the rank of an observation (or an event) and  $x(r)$  is its magnitude (or the frequency of occurrence). Zipf's law can be generalized into following form:

$$r^q x_{(r)} = \text{constant}, \quad q > 0. \quad (2)$$

The most famous example of Zipf's law is the frequency of English words. A study involving the count of top 50 words in 423 TIME magazine articles (total 245,412 occurrences of words) found "the" to be the most frequently-occurring word with rank one (appearing 15861 times), followed by "of" as number two (appearing 7239 times), and "to" as number three (6331 times), etc. When the number of occurrences is plotted as the function of the rank (1, 2, 3, etc.) the functional form is a power-law function with exponent close to 1 (Miller and Newman 1958; Miller et al. 1958).

Empirical studies have found that Zipf's law describes phenomena in various fields, including cities in the United States by population (Hill 1970), immune system response (Burgos and Moreno-Tovar 1996; Li 2001), city sizes (Zipf 1949, Gell-Mann 1994), and aspects of Internet traffic (Breslau et al. 2000). A recent study (Axtell 2001) shows that firm sizes (by number of employees) are also Zipf distributed, and suggests that the probability that a firm is larger than size  $s$  is inversely proportional to  $s$ . This point adds to the generalization regarding pioneer survival, first mover advantage, and the

square root form of the order of entry-market share function, and leads us to hypothesize that rank order-market share relationship will follow a Zipf distribution.

## 2.2 The Hypothesis

The Zipf curves represented by equation (2) have a tendency to hug the axes of the diagram when plotted in linear scale (the curves look linear on a double-logarithmic scale). A simple description of data that follow Zipf distribution is that they have:

- a few elements that score very high (for example, the usage of words “the”, “of” “to” in English language),
- a medium number of elements with middle-of-the-road scores (for example, the usage of words “dog”, “house” etc.),
- a huge number of elements that score very low (for example, the usage of words “Zipf”, “double-logarithmic” etc.).

Based on the intuition described in the previous Section and the research in Marketing, we hypothesize that:

**H1:** *market shares of products in a market are Zipf distributed; market share is a Zipf’s law-form function of rank order.*

Therefore, the market share of a brand/firm will be determined by the equation

$$MS_r = \frac{1}{\sum_{i=1}^N \sqrt{i}} \cdot \frac{1}{\sqrt{r}} \quad (3)$$

Where,

$MS_r$  = Market share of the  $r^{\text{th}}$  ranked brand or firm in the market;

$r$  = the rank order (by market share, when the leader’s rank is 1 etc.);

$N$  = Number of brand in the market.

According to this hypothesis the market share of a brand or a firm in the market is determined by its rank order and by the number of brands in the market. Moreover, the market shares distribution will be similar in markets with equal number of brands. The hypothesized market share distribution, as a function of the number of brand in the market is described in Figure 1.

**Insert Figure 1 about here**

### **3. TEST DATA SET AND STATISTICS**

#### **3.1 The Data**

Data about 70 markets were chosen from the *Market Share Reporter*<sup>1</sup>-2001 (Lazich 2001). The data is a compilation of market shares reports from business periodicals between 1997 and 2000. All market share values are revenue shares. The selected data includes various market segments, such as agricultural production crop, metal mining, oil and gas extraction, food and kindred products, tobacco, textile, lumber and wood, paper and allied, printing and publishing, and transportation.

In some of the cases the data about the market included an “other” categories, and “private label” categories. Since it was impossible to know how many brands are included in these categories several alternative assumptions have been made:

---

<sup>1</sup> Lazich, Robert S. Ed (2001), *Market Share Reporter – an annual compilation of reported market share data on companies, products, and services*, The Gale Group.

1. All the “others” / ”private labels” were considered as *one* brand (each group separately) whose rank is *according to their market share*.
2. All the “others” / “private labels” were considered as *two* brands whose rank is as their relevant market share (the market share in case was the total market share of the “others” divided by two).
3. All the “others” / “private labels” were considered as *three* brands whose rank is *as their relevant market share* (the market share in case was the total market share of the “others” divided by three).

The forecasted market shares as well as the goodness of actual versus predicted fit coefficients have been calculated for each of the above-mentioned assumption.

### 3.2 Test Statistics

To evaluate the fit of the Zipf distribution, two goodness-of-fit measures were calculated: the correlation coefficient between the predicted and the actual market share, and the Theil’s  $U^2$  statistic (Theil 1966, p. 28; Bliemel 1980). We follow the standard definition of correlation coefficient. Theil’s  $U^2$  is a measure of forecast accuracy and is defined as the ratio of mean squared error (MSE) from the forecasting model under consideration and the mean squared error from a naïve forecasting model (i.e. a model that does no real forecasting). That is:

$$U^2 = \frac{\text{MSE of the forecasting model}}{\text{MSE of Naive forecast}} = \frac{\sum_i (P_i - Y_i)^2}{\sum_i Y_i^2}, \quad (4)$$

Where:

$P_i$  = predicted value of observation  $i$ ,

$Y_i$  = actual value of observation  $i$ .

If the forecasts are perfect, Theil's  $U^2$  is 0. If the naïve model is used in forecasting, then  $U^2$  equals 1. The closer  $U^2$  is to zero, the better is the forecasting model. Typically a value of  $U^2$  less than or equal to 0.3025 ( $U \leq 0.55$ ) is considered an excellent fit. The  $U^2$  statistic has been used in many fields to test the accuracy of forecasting methods, including Marketing (Urban et al. 1986; Fader and Hardie 2002) securities analysis (El-Galfy and Forbes 2000), and statistics (Mizrach 1992).

#### **4. RESULTS AND DISCUSSION**

This Section organized into two sub-sections. We evaluate the goodness-of-fit of the Zipf distribution in predicting the market share in Section 4.1. We present a comparative analysis in Section 4.2, which compares our work with other models of predicting market shares.

##### **4.1 Evaluation of Goodness-of-Fit**

Table 1 describes the goodness-of-fit of the Zipf distribution in predicting the market share distributions. We have computed the average correlation coefficients between the actual and predicted market share and average  $U^2$  using the aggregate data of 70 markets.

**Insert Table 1 about here**

As mentioned in Section 3.1 the “private labels” and/or “others” categories were treated as 1, 2, or 3 brands. Our results are encouraging. As an be shown from Table 1, in all

cases the correlation coefficients are above 0.90 and the  $U^2$  statistic is 0.24 or lower. This indicates an excellent fit between the Zipf model and the actual market share distribution. Moreover, as can be seen from the coefficients in Table 1, splitting the “others” and “private labels” to several brands did not change the goodness-of-fit measures significantly. The fit is very good irrespective of whether we treat these categories as 1, 2, or 3 brands. Hence, the “others” / “private brands” categories can be treated as 1 brand in their relevant unified rank. This implies on the power of these brands. It may be that these brands have a power of one.

One of the possible causes for the unified power of the “others” and the “private labels” might be the common positioning of these brands. According to studies on private labels (e.g. Richardson et al. 1994), while the retailers try to position private labels as the lower price brands, the consumers use the low price as a cue to lower quality. This argument may apply to the “others” category too. Usually the “others” are associated with relatively small firms, who direct their offers to a particular niche that larger firms are not interested in (such as special needs, special tastes, geographic locations), and/or enjoy price umbrella in the market. It is possible, therefore, that the common positioning of either the “others” or the “private labels” may provide them a unified power, which lead the other players in the market to relate to them as one firm or brand.

Can we identify other patterns in the goodness-of-fit? In order to answer this question we compared the goodness-of-fit coefficients between markets with different numbers of brand. Table 2 presents the average  $U^2$  values for the markets by the number of brands.

**Insert Table 2 about here**

No specific pattern could be identified from the data in Table 2. This implies that the goodness-of-fit is not contingent on the number of brands in the market. In addition, the  $U^2$  values indicate a very good fit for the Zipf distribution at the market-level. We were also interested in determining whether the goodness-of-fit depends upon the type of the industry. In order to do this we calculated average goodness of fit for each of the market categories in our sample. Table 3 presents the results. Here again, no specific pattern could be identified, implying that the goodness-of-fit is not contingent upon the type of industry.

**Insert Table 3 about here**

To conclude, the results show that the Zipf distribution provide a good prediction for the market share distribution. It has been shown that the goodness-of-fit depends neither on the number of brand in the market nor on the type of products that are sold in the markets.

## **4.2 Comparative Analysis**

Several studies have dealt with the analysis of the competition in markets and its implications to the positioning decision of firms in the markets (e.g. Hauser and Shugan 1983; Lane 1980; Kumar and Sudharshan 1988). These studies have shown that various assumptions under which the market operates may lead to different equilibrium in terms of prices, number of players in the market, positioning, revenue, profit, and markets shares. In this section, we compare the market share results of the Zipf model with those

from other studies in the field. This comparison will assist in determining whether the underlying assumptions of other models lead to good markets share predictions. Our discussion will focus on the following studies.

- Hotelling (1929), for simultaneous entry,
- Lane (1980), for sequential entry,
- Kumar and Sudharshan (1988), for decupled response in defensive marketing, and
- Lilien et al. (1992), for sequential product positioning.

We discuss the underlying assumptions of these models and compare and contrast our results with those from these models. Sections 4.2.1 through 4.2.4 provide brief description of each of these models, while the discussion is presented in Section 4.2.5.

#### **4.2.1 Hotelling (1929)**

Hotelling (1929) first introduced a formal model of product differentiation set up as a two-stage oligopolistic competition in location and price. Customers are different in their tastes, which is captured by the assumption that customers are uniformly distributed along a linear line of length 1. Given the two firms' decisions about product locations and prices, each customer selects and buys a unit of product whose position is close to her preferences after adjusting for its price. The competition is characterized by the decisions made by two firms: first, the position of their product on the line and then, prices of the product. Of particular interest in the game is how the two firms would differentiate the products. Hotelling shows that when the competitors enter the market simultaneously, each firm will choose a central position along the line that will lead to a 50-50 market share distribution. When the two competitors enter the market sequentially,

the equilibrium positions and prices need to be determined computationally. This analysis has been discussed in detail in Lane (1980), which we discuss next. We provide a summary of results and assumptions of Hotelling (1929) in Table 4.

#### **4.2.2 Lane (1980)**

Lane (1980) provides a descriptive model of a market with differentiated consumers and firms. Firms enter the market sequentially and choose location and price. The price and location are endogenous to the model, while the number of firms entering the market can be either endogenous or exogenous. For a given set of product specification, Nash equilibrium in prices is shown to exist and to be unique. The locational equilibrium is then calculated using numerical techniques. The assumptions in the model are: fixed market size, uniform distribution of consumer tastes, perfect information and product availability to all consumers, identical cost structure for all firms in the market, positions that have been chosen by firms do not change. Lane determined the equilibrium state parameters when the number of brands in the market is determined either exogenously or endogenously for several fixed cost scenarios. Table 4 provides a summary of results and assumptions of Lane (1980).

#### **4.2.3 Kumar and Sudharshan (1988)**

Kumar and Sudharshan (1988) analyzed Lane's (1980) model by looking at attack-defense situations, rather than complete foresight. The entering brand in a market is designated as the attacker, while the existing brands are designated as the defenders. They analyze the defensive strategy based on a decoupled response function models of

advertising and distribution, and under the assumption of sequential entry, perfect foresight on subsequent entry, uniformly distributed tastes, constant market size, and exogenously determined demand. They use numerical technique to solve the model and calculate market shares for 1 attacker and 1 or 2 defenders. We provide a summary of results and assumptions of Kumar and Sudharshan (1988), along with those from other models, in Table 4.

#### **4.2.4 Lilien et al. (1992)**

Lilien et al. (1992) analyze exogenous fixed-size market with sequential entry, and uniformly distributed consumer tastes to model defensive product strategy. They show that the equilibrium market shares for the first two entrants in the market are  $2/3$  and  $1/3$  respectively. They highlight the differences between their model and the models of simultaneous entry (Hotelling 1929) and of a monopolist who does not foresee new competition (Moorthy 1988).

Table 4 summarizes the underlying assumptions and market share predictions of each of the models discussed in Sections 4.2.1 through 4.2.4 along with those from the Zipf model. We have used the revenue data from Tables 1 and 2 (p. 252 and p. 254 respectively) in Lane (1980) to calculate the market shares for the Lane model. The market shares for the Kumar and Sudharshan model has been taken directly from Tables 1 and 2 (p. 812) of Kumar and Sudharshan (1988).

**Insert Table 4 about here**

#### 4.2.5 Discussion

Unlike the models described in the previous sections, the Zipf model does not make any assumption about the mode of entry or the number of competitors. Thus, the analysis in this Section will be helpful in determining the validity of the assumptions made in the models described. Understanding the validity of the assumptions behind the market share models will contribute to the formulation of a theory that might explain the logic behind the fit of the Zipf model to the actual data.

We note that each of the models described earlier makes a different set of assumptions. As a result, it is difficult to assess the extent to which the numerical values obtained from the models are close to each other. A rough comparison, however, can still be made. With the exception of the Hotelling model that implies simultaneous entry, all other models lead to similar forecasts. To analyze the predictions of these analytical models more formally, we calculated the Theil  $U^2$  values for each of the models with respect to the Zipf model. Table 5 describes our results.

#### **Insert Table 5 about here**

To obtain the  $U^2$  values, we calculated of the sum of the squared deviations as a proportion of the sum of squares of the Zipf forecasted values (i.e. the  $Y_i$ 's in equation 4 come from the Zipf model). We can see from Table 5 that the  $U^2$  values range between 0.3% and 6%, which represents excellent fit. Therefore, we conclude that the forecasts from the analytical models (Lane 1980; Kumar and Sudharshan 1988; and Lilien et al. 1992) are very close to that of the Zipf model. This may imply that the assumptions

made in these models are acceptable. However, while the sequential entry describe a common situation, the other assumptions regarding the fixed market size, uniform distribution of consumer tastes, and differential cost structure may or may not describe a realistic situation. Therefore, we suggest that future research should assess the sensitivity of the forecasts with respect to these assumptions.

One such example is the study by Tyagi (2000). Tyagi (2000) considers sequential product positioning in a two-competitor market, and under differential cost. He shows that under the assumption of differential cost structure (in terms of production cost), the market share of the second competitor can be described as a function of the his relative cost advantage. Interestingly analysis of Tyagi's market share equation lead to the conclusion that the market share results will be as predicted by the Zipf model when the second mover has a cost advantage.

## **5. CONCLUSION**

We addressed a fundamental question in this paper: what is the distribution of market shares within a market? Using data from 70 markets we show that market shares within a market are Zipf distributed and that the market share is a Zipf's law-form function of the rank order in the market. Our empirical analysis show excellent fit of Zipf model to the market share data. Our analyses involved data from a wide range of industries, including agricultural production crop, metal mining, oil and gas extraction, food and kindred products, tobacco, textile, lumber and wood, paper and allied, printing and publishing, and transportation. The results show similar values of goodness-of-fit measures across industry types. We compare and contrast our work with other market

share models in the marketing literature. In particular, we show that market share predictions from our model are similar to those from Lane (1980), Kumar and Sudharshan (1988), and Lilien et al. (1992), which assume sequential entry to the market. Krugman (1996) in writing about a Zipf distribution regularity observed in the distribution of sizes of cities said (p. 399): *“The usual complaint about economic theory is that our models are over-simplified – that they offer excessively neat views of complex, messy reality... while there must be a compelling explanation of the astonishing empirical regularity in question, I have not found it.”* In contrast, in the case of the distribution of market shares within a market we not only have found an empirically observed regularity but have also shown it to be in consonance with derivations from existing analytical models.

Our work makes a fundamental contribution in understanding the distribution of market shares. Our model does not make any assumption about the order of entry and is valid for arbitrary number of competitors in the market. To the best of our knowledge, no other study in the literature predicts market share at such a fundamental level. Thus, our work can be used for prediction of market shares even when the order of entry is unknown. We make another interesting observation. Studies in order of entry literature in marketing relate the market share of an entrant to that of the pioneer using a square root relation of the following form:  $MS_i = MS_1 / \sqrt{i}$ , where  $MS_i$  is the market share of the  $i^{\text{th}}$  entrant and  $MS_1$  is the market share of the pioneer. We note that this form of relationship satisfies the condition of Zipf distribution described in equation (2) of Section 2.1. We, therefore, conclude that market share by *order of entry* also follows a Zipf distribution. Therefore, the Zipf distribution can be used to predict market shares

under several different scenarios. Our results highlight the importance of Zipf distribution in marketing, which is known to describe important phenomena in several other fields including linguistics and genetics.

We show that the market share predictions from the Zipf model and other analytical models are similar. However, the analytical models make several assumptions about the market characteristics, as described earlier. An interesting idea for extension will be to relax these assumptions and compare the findings of the relaxed models with that of the Zipf model.

## 6. REFERENCES

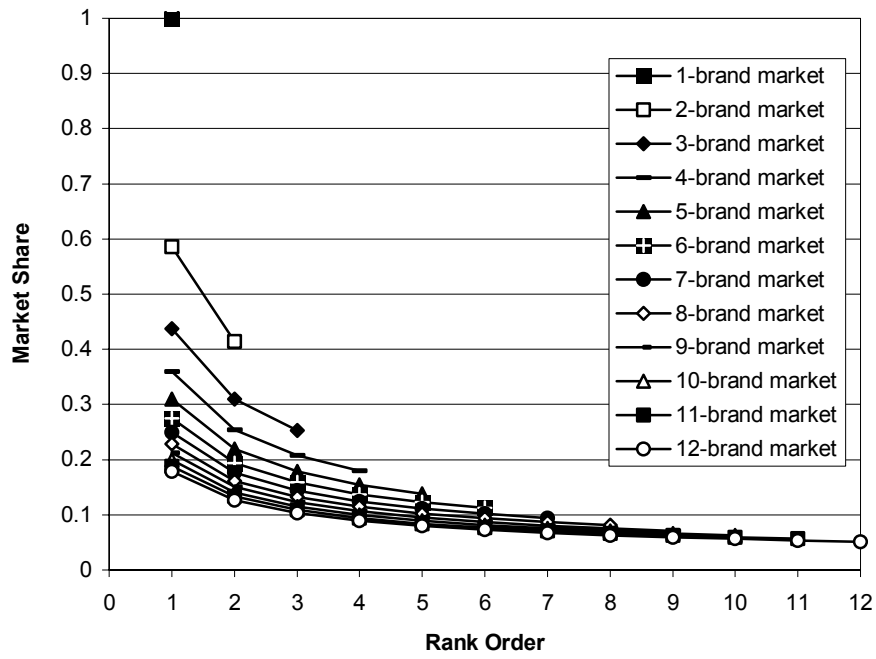
- Agarwal, R. 1996. Technological Activity and Survival of Firms. *Economic Letters*. 52 (July) 101-108.
- Agrawal, R. 1997. Survival of Firms over the Product Life Cycle. *Southern Economic Journal*. 63(3), 571-584.
- Agrawal, R., M. Gort. 1996. The Evolution of Markets and Entry, Exit, and Survival of Firms. *Review of Economics and Statistics*. 78 (November), 489-498.
- Axtell, R. 2001. Zipf Distribution of U.S. Firm Sizes, *Science Magazine*. 293 (September), 1818-1820.
- Berndt, E., L.T. Bui, D.H. Reiley, G.L. Urban. 1994. The Roles of Marketing Product Quality, and Price Competition in the Growth and Composition of the U.S. Anti-Ulcer Drug Industry. Working Paper #19-94 (May), Sloan School, MIT.
- Bliemel, R. 1980. Theil's Forecast Accuracy Coefficient: A Clarification. *Journal of Marketing Research*. 10 (November), 444-446.
- Bohlmann, J.D., P.N. Golder, D. Mitra. 2002. Deconstructing the Pioneer's Advantage: Examining Vintage Effects and Consumer Valuations of Quality and Variety. *Management Science*. 48(9), 1175-1195.
- Breslau, L., P.Cao, L. Fan, G. Phillips, S. Shenker. 2000. Web caching and Zipf-like distributions: evidence and implications. *Proceedings of INFOCOM'99*. IEEE Press.
- Brown, C.L., J.M. Lattin. 1994. Investigating the Relationship Between Time in the Market and Pioneering Advantage. *Management Science*. 40 (October), 1361-1369.
- Burgos J.D., P. Moreno-Tovar. 1996. Zipf-Scaling Behavior in the Immune System. *Biosystems*. 39(3), 227-232
- Carpenter, G.S., K. Nakamoto. 1989. Consumer Preference Formation and Pioneering Advantage. *Journal of Marketing Research*. 26 (August), 285-298.
- El-Galfy, A.M., W.P. Forbes. 2000. An Evaluation of US Security Analysts Forecasts 1983-1997. Working Paper, University of Manchester, England.
- Fader, P.S., B.G.S. Hardie. 2002. A Note on an Integrated Model of Customer Buying Behavior. *European Journal of Operational Research*. 139 (3), 682-687.
- Gell-Mann, M. 1994. *The Quark and the Jaguar*. New York: Freeman.
- Glazer, A. 1985. The Advantages of Being First. *American Economic Review*. 75 (June), 473-480.

- Golder, P.N., G.J. Tellis. 1993. Pioneer Advantage: Marketing Logic or Marketing Legend?. *Journal of Marketing Research*. 30 (May), 158-170.
- Hauser, J.R., S.M. Shugan. 1983. Defensive Marketing Strategies. *Marketing Science*. 2(4), 319-360.
- Hauser, J.R., B. Wernerfelt. 1990. An Evaluation Cost Model of Consideration Sets. *Journal of Consumer Research*. 16 (March), 393-408.
- Hill, B.M. 1970. Zipf's Law and Prior Distributions for the Composition of a Population. *Journal of the American Statistical Association*. 65, 1220-1232.
- Hotelling, H. 1929. Stability in Competition. *The Economic Journal*. 39(153), 41-57.
- Huff, L.C., W.T. Robinson. 1994. The Impact of Leadtime and Years of Competitive Rivalry on Pioneer Market Share Advantages. *Management Science*. 40 (October), 1370-1377.
- Kalyanaram, G., W.T. Robinson, G.L. Urban. 1995. Order of Market Entry: Established Empirical Generalizations, Emerging Empirical Generalizations, and Future Research. *Marketing Science*. 14(3), 212-221.
- Kalyanaram, G., G.L. Urban. 1992. Dynamic Effects of the Order of Entry on Market Share, Trial Penetration, and Repeat Purchases for Frequently Purchased Consumer Goods. *Marketing Science*. 11(3), 235-250.
- Kalyanaram, G., D.R. Wittink. 1994. Heterogeneity in Entry Effects Between Nondurable Consumer Product Categories. *International Journal of Research in Marketing*.
- Kardes, F.R., G. Kalyanaram. 1992. Order-of-Entry Effects on Consumer Memory and Judgment: An Information Integration Perspective. *Journal of Marketing Research*. 29 (August), 343-357.
- Kerin, R.A., P.R. Varadarajan, R.A. Peterson. 1992. First-Mover Advantage: A Synthesis, Conceptual Framework, and Research Propositions. *Journal of Marketing*. 56 (October), 33-52.
- Krugman, P. 1996. Confronting the Mystery of Urban Hierarchy. *Journal of the Japanese and International Economies*. X, 399-418.
- Kumar, K.R., D. Sudharshan. 1988. Defensive Marketing Strategies: An Equilibrium Analysis Based on Decoupled Response Function Models. *Management Science*. 34(7), 805-815.

- Lane, W.J. 1980. Product Differentiation in a Market with Endogenous Sequential Entry. *The Bell Journal of Economics*. 11(1), 237-260.
- Lazich, R.S. 2001. *Market Share Reporter: An Annual Compilation of Reported Market Share Data on Companies, Products, and Services*. Farmington Hills, Michigan: The Gale Group.
- Li, W. 2001. Zipf's Law in Importance of Genes for Cancer Classification Using Microarray Data", *arxiv.org* e-print, physics/0104028 (April 2001).
- Lieberman, M.B. 1989. The Learning Curve, Technology Barriers to Entry, and Competitive Survival in the Chemical Processing Industries. *Strategic Management Journal*. 10 (September-October), 431-447.
- Lilien, G. L., P. Kotler, K. S. Moorthy. 1992. *Marketing Models*. Prentice Hall, Englewood Cliffs, NJ.
- Miller, G.A., E.B. Newman. 1958. Tests of a statistical explanation of the rank-frequency relation for words in written English. *American Journal of Psychology*. 71, 209-218.
- Miller, G.A., E.B. Newman, E.A. Friedman. 1958. Length-frequency statistics for written English. *Information and Control*. 1, 370-389.
- Mitchell, W. 1991. Dual Clocks: Entry Order Influences on Incumbent and Newcomer Market Share and Survival When Specialized Assets Retain Their Value. *Strategic Management Journal*. 12 (January-February), 85-100.
- Mizrach, B. 1992. The Distribution of the Theil-U Statistic in Bivariate Normal Populations," *Economics Letters*. 38, 163-167.
- Moorthy, K. S. 1988. Product and Price Competition in a Duopoly. *Marketing Science*. 7(2), 141-168.
- Parry, M., F.M. Bass. 1990. When to Lead or Follow? It Depends. *Marketing Letters*. 1 (November), 187-198.
- Porter, M.E. 1980. *Competitive Strategy*. New York: The Free Press.
- Prescott, E.C., M. Visscher. 1977. Sequential Location of Firms with Foresight. *Bell Journal of Economics*. 8 (Autumn), 378-393.
- Read, C.B. 1988. Zipf's Law. In Kotz, S., N.L. Johnson, C.B. Read (eds.). *Encyclopedia of Statistical Sciences*. New York: Wiley.
- Richardson, P. S., A. S. Dick, A. K. Jain. 1994. Extrinsic and Intrinsic Cue Effects on Perceptions of Store Brand Quality. *Journal of Marketing*. 58 (October), 28-36.

- Robinson, W.T. 1988. Sources of Market Pioneer Advantages: The Case of Industrial Goods Industries. *Journal of Marketing Research*. 25 (February), 87-94.
- Robinson, W.T., C. Fornell. 1985. Sources of Market Pioneer Advantages in Consumer Goods Industries. *Journal of Marketing Research*. 22 (August), 305-317.
- Robinson, W.T., S. Min. 2002. Is the First to Market the First to Fail? Empirical Evidence for Industrial Goods Businesses. *Journal of Marketing Research*. 39 (February), 120-128.
- Scherer, F.M., D. Ross. 1990. *Industrial Market Structure and Economic Performance*, (third ed.). Boston: Houghton Mifflin.
- Schmalensee, R. 1982. Product Differentiation Advantages of Pioneering Brands. *American Economic Review*. 72 (June), 349-365.
- Sullivan, M.W. 1992. Brand Extensions: When to Use Them. *Management Science*. 38 (June), 793-806.
- Theil, H. 1966. *Applied Economics Forecasting*. Amsterdam: North-Holland. 28.
- Tyagi, R.K. 2000. Sequential Product Positioning Under Differential Costs. *Management Science*. 46(7), 928-940.
- Urban, G.L., T. Carter, S. Gaskin, Z. Mucha. 1986. Market Share Rewards to Pioneering Brands: An Empirical Analysis and Strategic Implications. *Management Science*. 32 (June), 645-659.
- Zipf, G.K. 1949. *Human Behavior and the Principle of Least Effort*. New York: Addison Wesley.

**Figure 1: Predicted Market Share Distribution for Market with Different Number of Brands**



**Table 1: Goodness-of-fit Coefficients.**

Goodness-of-Fit Measure	Treatment of the categories “other”/“private labels”		
	1 Brand	2 Brands	3 Brands
Average Correlation between predicted and observed	0.95	0.93	0.92
Average $U^2$	0.21	0.19	0.24

**Table 2: Average  $U^2$  for Markets by Number of Brands in the Market**

Number of brands	3	4	5	6	7	8	9	10	11
$U^2$ [%]	11.2	15.4	37.1	15.3	21.8	17.1	18.3	22.3	17.8

**Table 3: Goodness-of-fit Coefficients by Industry**

Category	Number of Markets in the sample from this category	Others/private labels as 1 brand		Others/private labels as 2 brands		Others/private labels as 3 brands	
		Av correlation coefficient	Av U <sup>2</sup>	Av correlation coefficient	Av U <sup>2</sup>	Av correlation coefficient	Av U <sup>2</sup>
Agriculture production - crops	3	0.96	0.14	0.96	0.17	0.93	0.22
Metal Mining	2	0.95	0.07	0.93	0.09	0.96	0.06
Oil and gas extraction	1	0.98	0.17	0.98	0.28	0.98	0.19
Food and kindred products	34	0.95	0.28	0.92	0.28	0.91	0.37
Tobacco products	2	0.94	0.32	0.96	0.30	0.94	0.31
Textile mill products	1	0.86	0.02	0.97	0.00	.94	0.04
Lumber and wood products	2	0.96	0.07	0.91	0.04	0.96	0.03
Paper and allied products	6	0.95	0.19	0.93	0.14	0.93	0.14
Printing and publishing	5	0.97	0.14	0.95	0.07	0.89	0.07
Chemicals and allied products	10	0.96	0.14	0.95	0.12	0.94	0.12
Petroleum and coal products	1	0.99	0.12	0.82	0.03	0.64	0.06
Rubber and miscellaneous plastic products	1	0.93	0.05	0.89	0.03	0.95	0.02
Transportation by air	2	0.95	0.02	0.92	0.04	0.93	0.06

**Table 4: Comparison of the Zipf Model with Other Models of Predicting the Market shares**

	A s s u m p t i o n s								Predicted Market Share
	Number of competitors (N)	Determination of number of brands in the market	Sequential/ simultaneous entry	Fixed market size	Uniform distribution of consumer's tastes	Perfect information	Cost structure	Re-position	
Hotelling (1929)	2	Exogenous	Simultaneous		✓				2 competitors: 0.5:0.5
Lane (1980)	N	Exogenous/ endogenous	Sequential		✓	✓	Identical to all	No	<u>Exogenous N:</u> 2 competitors: 0.68: 0.32 3 competitors: 0.4443:0.2714:0.2843 4 competitors: 0.3902:0.2517:0.2529:0.1052  <u>Endogenous N:</u> <ul style="list-style-type: none"> <li>Results depend on cost structure</li> <li>No used for comparison with the Zipf model</li> </ul>
Kumar and Sudharshan (1988)	N	Exogenous	Sequential	✓	✓	✓			2 competitors: 0.6362:0.3638 3 competitors: 0.4233:0.1562:0.4205
Lilien et al. (1992)	2	Exogenous	Sequential	✓	✓				2 competitors: 0.67:0.33
The Zipf model									2 competitors: 0.59: 0.41 3 competitors: 0.4377: 0.3095: 0.2527 4 competitors:0 .3591:0 .2540:0 .2074:0 .1796

**Table 5: Comparison of Forecasts from Different Models**

No. of brands In the market	Theil U <sup>2</sup>		
	Lane (1980)	Kumar & Sudharshan (1988)	Lilien et al. (1992)
2	0.0314	0.0083	0.0248
3	0.0029	0.0622	
4	0.0315		