Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2003 Proceedings

International Conference on Information Systems (ICIS)

December 2003

Firm Size and Information Technology Investment: Beyond Simple Averages

Tianyi Jiang
New York University

FIRM SIZE AND INFORMATION TECHNOLOGY INVESTMENT: BEYOND SIMPLE AVERAGES

Tianyi Jiang
Leonard N. Stern School of Business
New York University
New York, NY USA
tjiang@stern.nyu.edu

Abstract

We attempt to gain a better perspective on evolving firm-size in the past 20 years across industries by combining the empirical framework of Brynjolfsson et al. (1994) for measuring the effect of coordination cost reduction due to information technology investment, and the synopsis of theories of the firm by Kumar et al. (2001). We find that although in general Brynjolfsson et al.'s result holds for new firm data from COMPUSTAT, the firm size of the professional service sector grows as IT investment increases. The paper's potential contributions to empirical methods include (1) a different focus on the measurement of firm size by utilizing the weighted average employee-measure of firm size adopted by Kumar et al. work to replicate Brynjolfsson et al.'s findings with a new dataset, and (2) refinement of Kumar et al.'s weighted average employee-measure of firm size using entropy partition techniques from the machine learning literature, to fully account for the effect of larger firms within each industry.

Keywords: Firm size, IT investment, minimal entropy partitioning

Introduction

The study of the firm has been the focus of economics research for the past 50 years. Transaction cost (Coase 1937; Williamson 1979, 1985) and incomplete contract theories (Hart 1995) help explain the existence of the firm; information processing (Radner 1993) theories help explain the decision structure of the firm; and agency theory (Milgrom and Roberts 1992) help explain the cost of the firm. But what exactly defines the size of the firm? Should firm size be modeled by the number of employees, value-added, or average sales per firm? And how is firm size changing, especially in the context of evolving information technologies? IT decreases the decision costs, agency costs and coordination costs between and across firms (Gurbaxani and Whang 1991), which could make firms smaller (Malone et al. 1987), and outsource more from fewer suppliers (Bakos and Brynjolfsson 1993; Clemons et al. 1993). As IT becomes increasingly prevalent in today's business environment, how are the various definitions of firm size evolving? Are firms becoming smaller across all industries, or are there certain industries in which this trend is reversed?

This paper attempts to build upon earlier work on the correlation of firm size and IT investments, and takes an empirical approach to examining the evolving patterns of firm size across multiple industries for the period of 1982 to 2001. Specifically, this paper focuses on Brynjolfsson et al.'s (1994) seminal work on the relationship between IT and firm size and Kumar et al.'s (2001) work on factors that influence firm size in order to derive a better perspective on the study of firm size. We integrate the theories behind both studies in our effort to empirically study and understand the evolving industry trends in the past 20 years. To aid our empirical study, we extend Kumar et al.'s weighted average employee-measure of firm size by implementing unsupervised discretization algorithms (Dougherty et al. 1995) from the realm of machine learning.

This paper is organized as follows. The two studies by are introduced in more detail and our integrated perspective on the study of firm size is presented. The data and methodologies are then presented. The paper concludes with a discussion of the preliminary results.

Background

Brynjolfsson et al.'s (1994) seminal work on the empirical examination of the relationship between IT investment and firm size studied IT investment data from 1976 to 1989 and found that IT investments correlate highly with reduction in firm size across all industries. Firm size is defined as employees per establishment, employees per company, sales per firm, and value added per firm. Firm boundary is defined as the legal entity of firm, in accordance with Hart's (1995) definition of the firm. The findings are definitely provocative, although the authors point out that their findings do not imply any direct causal links between IT investment and the various definitions of firm size. The study is based on theoretical literature in transaction costs and coordination costs. While Brynjolfsson et al. acknowledge that there are numerous factors influencing the changes in firm size, their empirical results demonstrate that IT's effects on between-firms coordination costs dominate its effects on within-firm coordination costs.

Our criticism for the Brynjolfsson et al. study is two fold. First, we observe that the negative correlation in IT investment and firm size is not consistent across all industries, even allowing for time lag effects. Categorizing the industries based on the North American Industry Classification System (NAICS), our preliminary results show that IT investment has a significant positive correlation with firm size in the professional, scientific, and technical services industry. We believe that one possible reason for this discrepancy is that the Brynjolfsson study had a much coarser categorization of industry sectors. By breaking the industries into 17 categories versus the previous study's 6 industry sectors, we have a finer resolution in studying the relationship of firm size and IT investments. We postulate that in the case of professional services, IT's influence on coordination cost within firms dominates its influence on coordination cost between firms. Lower internal coordination cost increases the span of control, which increases firm size (Radner 1993).

Our second criticism of the study comes from the observation that firm size, as defined by Brynjolfsson et al.'s four different measurements, could decline for a given industry due to the entrance of a large number of small-sized firms. Thus, large firms with heavy IT investment could have increased in size, but a decrease in simple industry-average firm size could still lead to the conclusion that the firms are getting smaller in relation to IT investment. The richness of data on the distribution of firm size within a particular industry is missed by the simple averaging measure of employees per firm. These two observations argue for a better measure of firm size.

The paper by Kumar et al. (2001) on factors that determine firm size investigates the issue of firm size from the lens of three theories that were classified as technological, organizational, and institutional. The technological theory argues that market size could limit the "extent of specialization, and indirectly the size of the firm." Thus market size must be taken into consideration for any empirical analysis of firm size. The authors also argue that "greater capital intensity, proxied by investment per worker or R&D intensity" will have a positive correlation with firm size. Organization theory looks at the contract cost and argues that better patent rights protection will induce growth in firms that rely on "forms of critical resources, such as brand names, intellectual property, or innovative processes." Institutional theory postulates that regulatory forces and judicial efficiency influences firm size via barriers to entry and financial market development. Kumar et al.'s study looks at empirical data from 15 European countries and found that greater capital intensity correlates positively with firm size across industries; better legal systems correlates positively with R&D intensive industry sizes and negatively with non-R&D intensive industry sizes across countries; and lastly, as predicated by critical resource theory, better judicial efficiency correlates negatively with capital intensive firms.

While Brynjolfsson et al.'s analysis of firm size is based on production coordination costs, Kumar et al. offer a richer view of firm size from three different theories. When combined, these two streams of research give us a better perspective on the study of firm size. In practical terms, we adopt Brynjolfsson et al.'s definition of IT investment and Kumar et al.'s definition of weighted average firm size in our empirical study of the factors behind the evolving patterns of firm size in the past 20 years.

- Brynjolfsson et al. define IT investment as the office, computing, and accounting machinery (OCAM) category from Bureau of Economic Analysis (BEA)'s capital flow tables. The focus on a narrower definition of information technology underscores the emphasis of coordination theory to that of office and coordination uses.

- Kumar et al.'s definition of weighted average firm size addresses our earlier criticism of the simple average employees per firm scheme utilized in the Brynjolfsson et al. study. Kumar et al. argue that the simple averaging scheme not only ignores the distribution of firm size in an industry, but also gives a measure of average firm size that discounts the firms with greater share in the industry's production.

In the section on data and methodology, we introduce Kumar et al.'s measure based on the weighted average number of employees per firm, and an extension of this measure using the recursive minimal entropy partitioning (RMEP) technique from machine learning literature. Since this is a report of research in progress, we will hold off our interpretations in light of the aforementioned grounding theories until the completion of our analysis.

Data and Methodology

Our approach uses U.S. industry data from COMPUSTAT and Bureau of Economic Analysis input-output (IO) benchmark tables to directly examine the relationship between various factors, including IT investment, and average firm size as defined by Kumar et al.. The data is divided into 17 industries according to the NAICS code: Agriculture; Mining; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, Technical Services; Education Services; Health Care and Social Assistance; Arts, Entertainment and Recreation; and Accommodation and Food Services. These 17 industries represent nearly all manufacturing and services industries in the United States. We did not cover four other NAICS industries—Management of Companies and Enterprises; Administrative and Support and Waste Management and Remediation Services; Other Non-Public Administration Services; and Public Administration—because the method we used to derive IT investments did not give us appropriate mappings into these industries.

The Data

As stated in the previous section, we adopted Brynjolfsson et al.'s capital flow OCAM category as our source for IT investment. Unfortunately, Bureau of Economic Analysis only publishes capital flow data for 1982 and 1992 for free and the industries were all listed under Standard Industrial Classification (SIC) codes. Figure 1 shows a graph of IT investment percentages for 17 NAICS industries based on 1992 capital flow data. To circumvent this lack of IT investment data, we extracted IT investment data from the four Bureau of Economic Analysis IO benchmark year (1982, 1987, 1992, and 1997) use tables, and interpolated IT investment data in the between IO benchmark years. While the interpolation step does introduce biases into the data, it is still a good first step in analyzing major trends. We will be purchasing annual IT investment data from Bureau of Economic Analysis to improve our analysis in the near future. The input commodity we used to calculate IT investment is the SIC classification of Computers and Office Equipment.

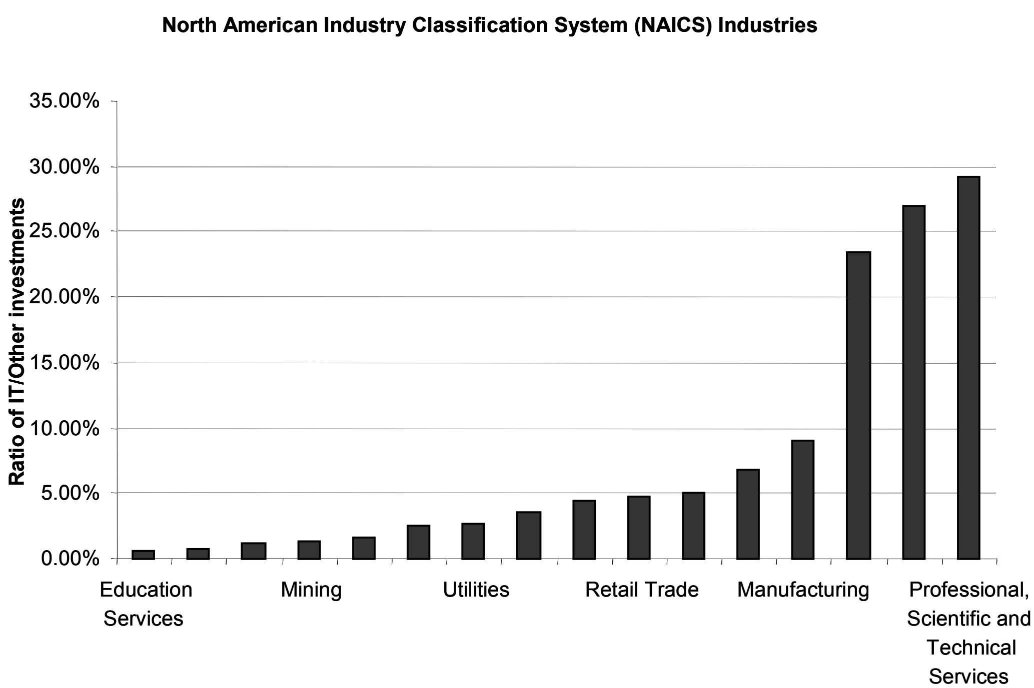

Figure 1. IT Investment Percentages

To study the evolving patterns of firm size across industries, 443,507 COMPUSTAT industry data points were collected for the period of 1982 to 2001. Data attributes used in this paper include industry classification code (SIC), CUSIP issuer code, stock ticker symbol, company name, industry name (NAICS), state, country, year, net sales (DATA12), and employees (DATA29). Since we are interested in firm size, all data points with missing employee attributes were dropped. In order to conduct any meaningful regression analysis, we also dropped any data points that have missing values for all other DATA type attributes. The resulting data set consisted of 143,812 data points.

The Methodology

Similar to the work done by Brynjolfsson et al., our basic technique for analyzing the data is a two-stage least-squares regression estimate of correlation between our selected attributes and firm size. Three different firm sizes are tested: simple employees per firm averages and two weighted averages derived from Kumar et al.'s weighted employee measure. The weighted employee measure was based on Davis and Henrekson's (1997) suggestion of using "coworker mean as a measure of size to emphasize the number of employees at the average worker's place of employment." Thus firms are partitioned into bins based on the homogeneity of firm employee sizes, and a weighted average of firm size is computed as the sum of average employees per bin multiplied by the percentage of total employees contained within the perspective bins.

The equation for employee-weighted average of firm size is

$$\textit{Weighted Average Number of Employees} = \sum_{1}^{n}\left(\frac{N_{bin}^{Emp}}{N_{Sector}^{Emp}}\right)\left(\frac{N_{bin}^{Emp}}{N_{bin}^{Firms}}\right)$$

N_{bin}^{Emp} ≡ the total number of employees in a bin,

N_{Sector}^{Emp} ≡ the total number of employees in the sector, and

N_{bin}^{Firms} ≡ the total number of firms in a bin.

Emphasizing the larger firms reduces the effect of entry and exit. This emphasis also has support from Sutton's (1997) work that most industry churn has little effect on the largest firms. We note that Kumar et al.'s weighted average scheme assumes a pre-partitioned set of bins, which if performed manually for all industries could be a time-consuming process prone to arbitrary biases and errors. We also note that proxies for large firms could be firm employee size as well as firm output. A firm that makes a significant contribution to an industry's total output should be considered just as important as the firm with the most employees. To that end, we improve Kumar et al.'s binning process by utilizing the RMEP technique (Fayyad and Irani 1993) from the realm of machine learning. Entropy, as defined in information theory, is a measure that "characterizes the (im)purity of an arbitrary collection of examples" (Mitchell 1997, p. 55). RMEP recursively partitions any attribute (in this case, employee) into two separate bins based on maximum entropy differences calculated among all possible separation of bins until a stopping criteria (based on the minimal description length principle) is reached. The intuition here is that RMEP continues to split the firm data into smaller bins of more homogenous firm employee sizes until the differences between two bin averages is too small to warrant a further division (for RMEP algorithm implementation details, see Dougherty et al. 1995).

Given the two proxies for large firms, we modified the RMEP algorithm to output two weighted average employee sizes: employee distribution based weighted average and sales (output) distribution based weighted average. To calculate the sales weighted averages, we use the firm's total annual sales as our entropy criteria. The resulting weighted average is then calculated by summing the average employees per bin multiplied by the percentage of industry annual sales contained within the respective bins. This use of unsupervised discretization of continuous features technique is novel in the context of empirical economics research. Unlike the predefied bins approach of Kumar et al., utilizing entropy partitioning techniques fully succeeds in emphasizing the largest firms since we now have prior knowledge on the distribution of firm size within the bins. Our approach in effect enhances Kumar et al.'s weighted average technique.

With three different average firm sizes defined, our model for measuring the relationship between firm size and other factors is as follows:

$$SIZE_t = \beta_0 + \beta_1 IT_t + \beta_2 IT_{t-1} + \beta_3 IT_{t-2} + \beta_4 IT_{t-3} + \beta_4 IT_{t-4}$$
$$\sum \beta_{7i} INDUSTRY_{it} + \beta_8 DATA12 + \varepsilon_i$$

SIZE ≡ the natural log of three different employee measures in year t
IT_t ≡ the natural log of the IT investment ratio per industry in year t
DATA12 ≡ the natural log of the net sales per industry per year
$INDUSTRY_{it}$ ≡ a dummy for each of the 17 industry sectors (for all three employee size measures)
ε_i ≡ an i.i.d error term with zero mean

We followed Hitt's (1999) use of IT investment ratio rather than Brynjolfsson et al.'s IT investment and other investment measures in order to normalize the size of various industries. We added net sales to control the effect of business cycle. The five-year delay helps us track the relationship between IT investments and firm size over time. The lag measurements will be more useful after we replace the interpolated data with more accurate annual IT investment data.

Preliminary Results

We have finished extracting industry IT investment data from the IO benchmark use for six tables: Professional Services,;Finance and Insurance; Manufacturing; Education; Whole Sale Trade; and Retail Trade. From Table 1, we see that overall IT investment ratio is negatively correlated with firm size. However, when we split the data into six separate industry sectors, the Professional Services industry sector has a significant positive correlation in firm size and IT investment ratio (Table 2).

In terms of our new entropy-based weighted employee average measures, we see that they are consistent with simple employee averages in some industries (Table 2) while they differ significantly in other industries (Table 3). At the all-industry level, we replicated previous studies' results with all three measures of firm size, indicating that our new measures are good proxies for average firm employee numbers. However, plots of weighted averages over 21 years of COMPUSTAT data show dramatic vertical movements and interesting lagging effects that are not present in simple employee averages (Figure 2). We will continue to evaluate the usefulness of our new measuring technique as we extract more industry level IT investment data across all non-IO benchmark years.

Summary

By combining Brynjolfsson et al.'s empirical framework of measuring the effect of coordination cost with Kumar et al.'s different lens of firm size, we attempt to gain a better perspective on evolving firm size in the past 20 years. Our contributions to this stream of research in terms of empirical technique are that we

1. Utilize weighted average employee sizes in our attempt to duplicate Brynjolfsson et al.'s findings

2. Perfect Kumar et al.'s weighted average employee scheme via entropy partition techniques to fully account for the effect of large firms within a firm size bin

Table 1. All Industry Regression Table

Variable	SIZE 1 (Simple Average)	SIZE 2 (Employee Weighted Average)	SIZE 3 (Sales Weighted Average
Constant	1.1498	9.203***	5.253
IT Investment Ratio by Year			
ITINVRATIO(0)	-0.03755*	0.01644	-0.04677
ITINVRATIO(-1)	0.04905*	0.00909	-0.0268
ITINVRATIO(-2)	-0.0054	-0.044446	-0.0124
ITINVRATIO(-3)	0.01384	-0.04134	-0.0589
ITINVRATIO(-4)	-0.04005***	-0.04445*	-0.04774
Total Sales	0.10655**	-0.3263***	-0.1278
Industry Dummies			
Professional Services	-1.3162***	-0.5818**	0.5994
Manufacturing	-0.97705***	-1.7965***	-1.6223***
Finance and Insurance	-1.25363***	-1.24530***	-2.0865***
Education	-1.8108***	-5.8528***	-4.537***
Wholesale Trade	-1.84806***	-2.25321***	-2.0053***
R-Squared	0.9920	0.990	0.962
Durbin-Watson	1.26	0.73	0.60
F Statistic	709.90	539.00	136.76
Number of observations	72	72	72

Key: *Significant at 90% level; **Significant at 95% level; ***Significant at 99% level

Table 2. Professional Services Regression Table

Variable	SIZE 1 (Simple Average)	SIZE 2 (Employee Weighted Average)	SIZE 3 (Sales Weighted Average
Constant	-2.4469**	1.177	2.055
IT Investment Ratio by Year			
ITINVRATIO(0)	0.21593*	0.5574*	0.5242*
ITINVRATIO(-1)	0.0398	-0.0605	-0.1021
ITINVRATIO(-2)	-0.0075	-0.0346	0.1065
ITINVRATIO(-3)	-0.1985	-0.8244**	-0.8663**
ITINVRATIO(-4)	-0.02131	0.15920	0.16896*
Total Sales	0.30680***	0.2180	0.1829
R-Squared	0.872	0.984	0.984
Durbin-Watson	2.59	2.30	2.72
F Statistic	5.66	52.23	52.32
Number of Observations	12	12	12

Table 3. Education Regression Table

Variable	SIZE 1 (Simple Average)	SIZE 2 (Employee Weighted Average)	SIZE 3 (Sales Weighted Average
Constant	-2.2323**	2.1487**	1.9102**
IT Investment Ratio by Year			
ITINVRATIO(0)	-0.03304	0.01772	0.00945
ITINVRATIO(-1)	0.03366	0.00004	-0.00723
ITINVRATIO(-2)	-0.02356	0.04457	0.00674
ITINVRATIO(-3)	0.05264*	0.05213*	0.08056**
ITINVRATIO(-4)	0.00226	0.01950	0.05140*
Total Sales	0.36576***	-0.02896	0.00929
R-Squared	0.864	0.851	0.911
Durbin-Watson	2.41	2.42	2.63
F Statistic	5.3	4.77	8.51
Number of Observations	12	12	12

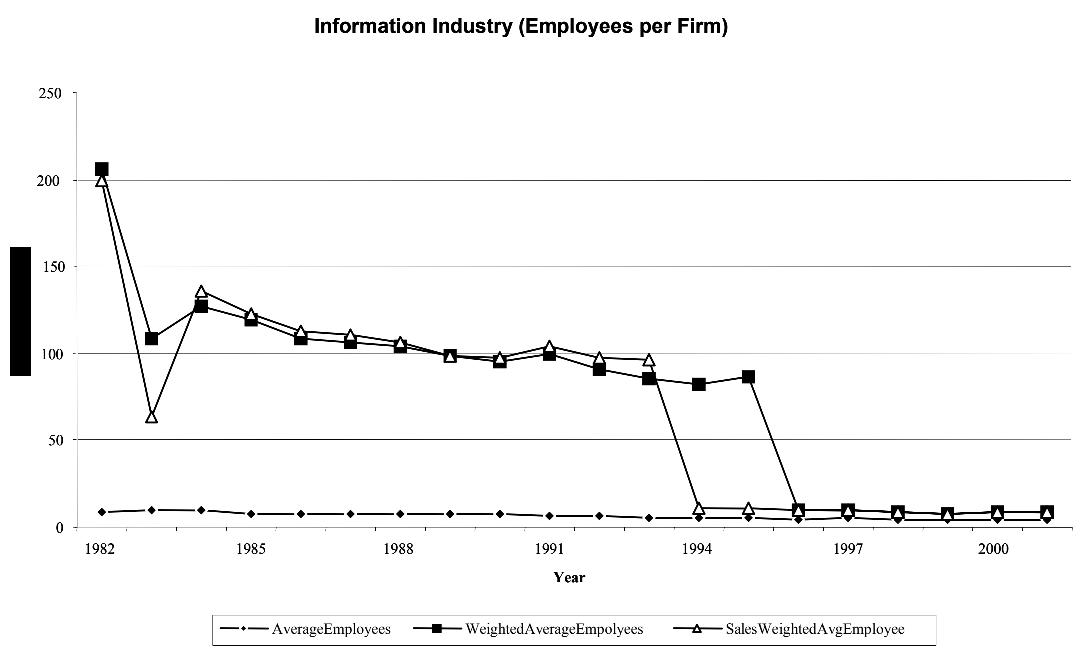

Figure 2. Employee Averages

References

Bakos, Y., and Brynjolfsson, E. "From Vendors to Partners: Information Technology and Incomplete Contracts in Buyer-Supplier Relationships," *Journal of Organizational Computing* (3:3), December 1993, pp. 301-328.

Brynjolfsson, E., Malone, T., Gurbaxani, V., and Kambil, A. (1994), "Does Information Technology Lead to Smaller Firms?," *Management Science* (40:12), December 1994, pp. 1628-1644.

Clemons, E. K., Reddi, S. P., and Row, M. C. "The Impact of Information Technology on the Organization of Economic Activity: The 'Move to the Middle' Hypothesis," *Journal of Management Information Systems* (10:2), , Fall 1993 , pp. 9-36.

Coase, R. H. "The Nature of the Firm," *Economica* (4:16), November 1937, pp. 386-405.

Davis, S., and Henrekson, M. "Explaining National Differences in the Size and Industry Distribution of Employment," Working Paper, The University of Chicago Graduate School of Business, 1997.

Dougherty, J., Kohavi, R., Sahami, M. "Supervised and Unsupervised Discreptization of Continuous Features," in *Proceedings of the 12th International Conference on Machine Learning*, Morgan Kaufmann, Tahoe City, CA, 1995, pp. 194-202.

Fayyad, U. M., and Irani, K. B. "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, Sidney, Australia, 1993, pp. 1022-1027

Gurbaxani, V., and Whang, S. "The Impact of Information Systems on Organizations and Markets," *Communications of the ACM* (34:1), January 1991, pp. 59-73.

Hart, O. *Firms, Contracts, and Financial Structure,* Oxford University Press Inc., New York, 1995.

Hitt, L. "Information Technology and Firm Boundaries: Evidence from Panel Data," *Information Systems Research* (10:2), June 1999, pp. 134-149.

Kumar, K., Rajan, R., and Zingales, L. "What Determines Firm Size?," Working Paper, The University of Chicago Graduate School of Business, 2001.

Malone, T. W., Yates, J., and Benjamin, I. "Electronic Markets and Electronic Hierarchies," *Communications of the ACM* (30;6), June 1987, pp. 484-497.

Milgrom, P., and Roberts, J. *Economics, Organization, and Management*, Prentice Hall, Englewood Cliffs, NJ, 1992.

Mitchell, T. *Machine Learning,* MIT Press and The McGraw-Hill Companies, Inc., Singapore, 1997.

Radner, R. "The Organization of Decentralized Information Process," *Econometrica* (61:5), September 1993, pp. 1109-1146

Sutton, J. "Gibrat's Legacy," *Journal of Economic Literature* (35:1), March 1997, pp. 40-59.

Williamson, O. E. *The Economic Institutions of Capitalism,* The Free Press, New York, 1985.

Williamson, O. E. "Transaction Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics* (22:2), 1979, pp. 233-261.