

Vol. 14 No. 2

# J.B.M.

## Journal of Business and Management

*Editors*

Cristina M. Giannantonio, Ph.D.

Amy E. Hurley-Hanson, Ph.D.

Published by Chapman University's Argyros School of Business and Economics  
Sponsored by the Western Decision Sciences Institute





## WESTERN DECISION SCIENCES INSTITUTE

The Western Decision Sciences Institute is a regional division of the Decision Sciences Institute. WDSI serves its interdisciplinary academic and business members primarily through the organization of an annual conference and the publication of the *Journal of Business and Management*. The conference and journal allow academicians and business professionals from all over the world to share information and research with respect to all aspects of education, business, and organizational decisions.

### PRESIDENT

Vijay R. Kannan  
*Utah State University*

### PRESIDENT-ELECT

G. Keong Leong  
*University of Nevada Las Vegas*

### PROGRAM CHAIR/VICE PRESIDENT FOR PROGRAMS/PROCEEDINGS EDITOR

Mahyar Amouzegar  
*California State University, Long Beach*

### VICE PRESIDENT FOR PROGRAMS-ELECT

Nafisseh Heiat  
*Montana State University-Billings*

### VICE PRESIDENT FOR MEMBER SERVICES

John Davies  
*Victoria University of Wellington*

### SECRETARY/TREASURER

Sheldon R. Smith  
*Utah Valley State College*

### DIRECTOR OF INFORMATION SYSTEMS

Abbas Heiat  
*Montana State University - Billings*

### IMMEDIATE PAST-PRESIDENT

Bruce Raymond  
*Montana State University*

### REGIONAL VICE PRESIDENT

Karen L. Fowler  
*University of Northern Colorado*

---

# Journal of Business and Management

---

Volume 14 Number 2

2008

## EDITORS

Cristina M. Giannantonio, *Chapman University*

Amy E. Hurley-Hanson, *Chapman University*

# J.B.M.

## Journal of Business and Management

### EDITORS

Cristina M. Giannantonio, *Chapman University*  
Amy E. Hurley-Hanson, *Chapman University*

### EDITORIAL BOARD

Nancy Borkowski  
*Florida International University*

Krishna S. Dhir  
*Berry College*

Sonia M. Goltz  
*Michigan Tech University*

Richard L. Jenson  
*Utah State University*

Miles G. Nicholls  
*RMIT University*

Terri A. Scandura  
*University of Miami*

Jeffrey A. Sonnenfeld  
*Yale University*

Victor H. Vroom  
*Yale University*

### PAST EDITORS

Burhan Yavas, *California State University Dominguez Hills* 1993-1999  
Raymond Hogler, *Colorado State University* 2000-2004

### EDITORIAL STAFF

Rosalinda Monroy, *Chapman University Publications*  
Heidi Stearns, *Editorial Assistant*  
Jaclyn Witt, *Editorial Assistant*

# Contents

Teaching Undergraduates About Dynamic Systems . . . . .	93
<i>Bruce Raymond and Laura J. Black</i>	
Data Mining Methods in the Detection of Spam . . . . .	117
<i>Dong-Her Shih, Hsiu-Sen Chiang, Chia-Shyang Lin and Ming-Hung Shih</i>	
A Resource-Based Look at Compensation Strategy: Application and Implementation of Competitive Advantage . . . . .	131
<i>James A. Carey</i>	
On the Predictability of Stock Market Returns: Evidence from Industry-Rotation Strategies . . . . .	149
<i>Robert R. Grauer</i>	
Brand Response-Effects of Perceived Sexual Harassment in the Workplace . . .	175
<i>Jeremy J. Sierra, Nina Compton and Kellilynn M. Frias-Gutierrez</i>	



# Teaching Undergraduates About Dynamic Systems

Bruce Raymond

Laura J. Black

*Montana State University*

*This study expands our understanding of how undergraduate business students learn about the behavior of dynamic systems, since the ability to assess and intervene in changing systems is increasingly important to effective business decision-making. A computer simulation is tested in a lower-division social sciences calculus class as a vehicle for improving students' effective understanding of dynamic systems. Pre and post assessments were compared to test the knowledge of the students. Results suggest that the simulation approach did improve student knowledge of dynamic systems.*

## Introduction

Ample research has indicated accelerating paces of change in business, both at the industry and firm levels (Fine, 2000). With innovation rates averaging nearly 10 percent annually and even faster in technological sectors (Mendelson & Pillai, 1999), business graduates who are capable of assessing, interpreting, and making effective decisions in dynamics situations are more likely to succeed than those who cannot recognize the accumulated implications of varying rates of change. A survey among top-ranked U.S. graduate business programs revealed that three-quarters of faculty view systemic thinking as an "essential" part of business education (Atwater, Kannan & Stephens, 2008). Yet abundant research also suggests that decision-makers perform poorly when facing tasks characterized by complex dynamics such as delays in cause and effect further obscured by separation in space and time (Serman, 1989; Dörner, 1996; Moxnes, 1998). Consequently, many teachers of business mathematics and other business courses are seeking improved methods of instruction and assessment regarding dynamic systems.

Some researchers of cognition and learning assert that conveying abstract rules, or context-free operations and formulae, help students learn more effectively than concrete instantiations. Kaminski, Sloutsky and Heckler (2008) provide evidence that, when presented with a novel-situated problem, undergraduate students who received mathematical instruction using only generic rules, significantly outperformed students who received instruction on the same principles using concrete examples or even concrete instantiations, plus a generic statement of the rules. It is unclear, however, if abstract mathematical statements describing dynamic systems, such as integration and differential equations, are effective at helping students grasp the consequences of relationships that may change nonlinearly over time. Sweeney and Serman (2000) contend that students, even those with graduate-level mathematics training at elite institutions, have a poor ability to extrapolate and interpret the situated consequences of dynamic relationships.

Given the pervasiveness of delays, feedback, and nonlinear influences in day-to-day business (Fine, 1998; Serman, 2000), it becomes critical to understand how to teach students effectively about dynamic systems. The literature providing models and theories of learning are both broad and diverse. One categorization proposed by Fenwick (2000) suggests that educators relinquish the premise that “experiential learning” can be uniquely defined, since every moment is one of experience and thus all learning is experiential. Additionally, Fenwick (2000) argues that cognition cannot be contained by a particular theory, but that multiple perspectives add to the overall understanding of learning and thinking. One of these perspectives, participation, suggests that the learner is an active participant and that learning is situated and all knowledge contextual. According to this idea, effective teaching leverages the contextual knowledge that students bring with them into the classroom setting.

A review of literature defining intelligence, knowledge, and learning (Raymond & Black, 2004) suggested the utility of using computer simulation as a teaching tool to link the academic conceptual development of mathematical abstraction skills to personal prior latent knowledge from common experience. To explore this possibility, the authors developed a simulation tool to link the concepts of system dynamics to the common experience of filling a bathtub. Pre and posttests were conducted to evaluate the effectiveness of the simulation method of supplemental instruction.

### **The Simulation Exercise**

A computer simulation was created with Visual Basic to represent a typical bathtub, as shown in Figure 1, and was used by the students of an introductory calculus course (designed for social science majors, rather than engineering majors) in concert with a structured simulation exercise (shown in Appendix I).

The purpose of the simulation and exercise was to build the student's knowledge of dynamic systems in a step-by-step fashion. This approach was based on the notion that, by “seeing” the mathematical concepts of a familiar situation dynamically simulated, the students would gain a deeper understanding of the mathematical concepts.



**Figure 1:** *Screen Capture of the Bathtub Simulation*

## Experimental Method

To test the hypothesis that the simulation and the structured exercise would improve student understanding of system dynamics, a beginning-of-semester pretest and an end-of-semester posttest were conducted using a set of assessment instruments provided in Sweeney and Sterman (2000). The four assessment problems are reproduced as Appendix II. These four problems assessed student mastery of two concepts of dynamic systems, the accumulation of stocks and rates of change, commonly known as stocks and flows. These two ideas form the conceptual heart of integral (accumulation) and differential (rates of change) calculus.

In each of these four problems, students were provided with the initial value of an accumulated resource, along with a graph delineating the rate at which the resource flowed into and exited from the system. The students were asked to provide a corresponding graph of the resource accumulation over time. These four problems included elementary cash flow and bathtub situations. Each student received only a single version of the problem. The four problems were distributed randomly among the students in the pretest, and again at the end of the course following the completion of the structured learning exercise, as a posttest. No procedures were used to assign the same problem to a particular student on the pretest and posttests, and the assessments were conducted with only the students in attendance on a particular date.

The first set of problems (CashFlow1 and Bathtub1) provided a square wave pattern of inflow (constant inflow with step changes in the rate) with a constant outflow, while the second set of problems (CashFlow2 and Bathtub2) provided a saw-tooth inflow pattern (variable inflow) coupled with a constant outflow. The four problems assumed knowledge of basic arithmetic, Cartesian graphing, and an intuitive understanding of rates/flows and accumulations/stocks. Algebraic abstraction and formulation were not required to solve the problems.

The four problems, CashFlow1, Bathtub1, CashFlow2, and Bathtub2, were assigned randomly across eleven sections of the calculus course. After discarding responses, due to both scoring and response irregularities, 254 instruments were coded from the pretest and 187 were coded from the posttest. Demographics including age, gender, class standing and major were collected from university records to check

the randomness of the assignment of the four problems. Demographics were not available for students who added the class after the date of the demographic data collection.

The experiment's null hypotheses stated that the students' performance would be equal when comparing the pretest results with the results of the posttest. Performance was measured as the fraction,  $p$ , of students who answered each of the problems correctly. It was assumed that an increase in performance on the posttest could be explained at least in part by the application of the structured simulation exercise. This assumption was verified by comparing the pretest and posttest performance with the performance of an equivalent group of students from the same course in a prior semester that completed the system dynamics assessments without doing the simulation and the structured exercise. This comparison is provided in the following section.

*Null Hypotheses:* Student performance was equal on the pretest and the posttest:

$$H_0: p_{PRE} = p_{POST}$$

*Alternative Hypothesis:* Student performance was lower on the pretest than on the posttest:

$$H_a: p_{PRE} < p_{POST}$$

## Results

A summary of the demographics is provided in Appendix III. There appeared to be no biased assignment of the four problems based on these demographic variables. The random method of distributing the four problems to students attending class on a particular day prevented ensuring that the same problem was administered to the same student on both the pretest and the posttest and that the same sample of students performed both the pretest and the posttest (not all of the students completed the simulation exercise). Because of this, we limited the hypothesis tests to a subset of students. Specifically, we focused only on students who completed the online simulation exercise and the posttest. Table 1 summarizes the results.

**Table 1:** Results for students who completed the simulation exercise and took the posttest

	PreTest			PostTest			Hypotheses Tests		
	Num	Correct	%	Num	Correct	%	$\chi^2$	p	Sig?
Bathtub1	30	3	10.00%	37	10	27.03%	3.07	0.04	Yes
CashFlow1	26	0	0.00%	31	7	22.58%	6.69	0.01	Yes
Bathtub2	24	0	0.00%	31	0	0.00%	0.00	1.00	No
CashFlow2	23	0	0.00%	35	0	0.00%	0.00	1.00	No

Hypothesis Test:  $H_0: p_{PRE} = p_{POST}$ ,  $H_a: p_{PRE} < p_{POST}$

The results suggest that the performance difference between the pretest and the posttest for students who completed the simulation exercise was significant at the  $\alpha = 0.05$  level on the Bathtub1 and CashFlow1 problems. Student performance did not improve on the Bathtub2 and CashFlow2 problems. To ascertain that the performance

improvements were due, at least in part, to the completion of the simulation exercise rather than to other factors (including the completion of the basic calculus course), the pre and posttest results were also compared to the performance results collected at the end of a prior semester. This group consisted of 165 similar calculus students who completed the Bathtub and CashFlow problems but did not complete the simulation exercise (Table 2). We assumed that completion of a basic integral and differential calculus course would improve student understanding of dynamic systems.

**Table 2:** Comparison of Pretest Results to Prior Semester Performance

	Pretest			Prior Semester			Hypotheses Tests		
	Num	Correct	%	Num	Correct	%	$\chi^2$	p	Sig?
Bathtub1	30	3	10.00%	43	9	20.93%	1.54	0.11	No
CashFlow1	26	0	0.00%	44	4	10.00%	2.51	0.06	No
Bathtub2	24	0	0.00%	38	1	2.63%	0.64	0.21	No
CashFlow2	23	0	0.00%	38	1	2.63%	0.62	0.22	No

Hypothesis Test:  $H_0: p_{PRE} = p_{Prior}$ ;  $H_a: p_{PRE} < p_{Prior}$

These results indicate that completion of the basic calculus class alone did lead to improvements in knowledge regarding dynamic systems on the Bathtub1 and CashFlow1 problems, but the performance increases were not significant at  $\alpha = 0.05$ . Table 3 shows the posttest results for students completing the basic calculus course and the simulation exercise, compared to the performance results for students completing the basic calculus course in the prior semester (with no simulation exercise).

**Table 3:** Comparison of Posttest Results to Prior Semester Performance

	Prior Semester			Posttest			Hypotheses Tests		
	Num	Correct	%	Num	Correct	%	$\chi^2$	p	Sig?
Bathtub1	43	9	20.93%	37	10	27.03%	0.41	0.26	No
CashFlow1	44	4	10.00%	31	7	22.58%	2.64	0.05	Yes
Bathtub2	38	1	2.63%	31	0	0.00%	0.83	0.18	No
CashFlow2	38	1	2.63%	35	0	0.00%	0.93	0.17	No

Hypothesis Test:  $H_0: p_{Prior} = p_{POST}$ ;  $H_a: p_{Prior} < p_{Post}$

For students completing the Bathtub1 problem, posttest performance was higher than in the prior semester, but the difference was not significant at the  $\alpha = 0.05$  level. Student performance on the posttest was significantly higher on the CashFlow1 problem when compared to students from the prior semester.

## Discussion and Conclusion

Introductory calculus texts often introduce integration by having students calculate the area under a graphical function on a Cartesian graph and also introduce differentiation by calculating the slope of tangential lines to a graphical function.

Apparently this common approach did make some difference in student understanding of dynamic systems, as noted in *Bathtub1* and *CashFlow1* results. However, the concepts, exercises, and exams provided in an introductory social sciences calculus course did not significantly improve student knowledge of dynamic systems in our study. What prevented these students from connecting their knowledge of elementary calculus to observations of state variables and the rates at which they change?

Furthermore, more students completed the *Bathtub1* problem accurately than completed the *CashFlow1* problem accurately, even though the mathematical structure underlying the exercises is identical. Why is this? It may be that in traditional calculus education, integration and differentiation are often not explicitly related to physical activities and states until students reach more advanced engineering-focused courses. But, if students were unable to relate the abstract language of mathematics to any real-world scenarios, we would expect that the data would suggest equivalent struggles with the bathtub and cash flow scenarios.

In the learning literature, we see evidence of situational knowledge that may help us interpret these results. Schliemann (1998) described the mathematical capabilities of street vendors and cooks and the use of scalar arithmetic in their every day activities. Their mathematical capabilities were sophisticated within the context of their work, but not generalizable without additional training. Another contextual anecdote involved the voting behavior and perceptions of a Brazilian woman. In this situation, the woman's interpretation of graphical information was influenced by her voting preference. Lave (1988) studied the mathematical sophistication of ordinary people living their day-to-day lives and found that in many instances, mathematical capabilities were uniquely situated. Lave described individuals who performed poorly on basic arithmetic tests given in a school environment but performed the same math calculations accurately and without apparent difficulty in the bowling alley. Shoppers were also found to demonstrate mastery of basic arithmetic in stores but they performed poorly in school settings on the same arithmetic operations.

We speculate that more students enrolled in first-semester calculus (usually undertaken during the first three semesters of college) have tacit, or automatic and situated knowledge of the dynamics of filling a bathtub with water than do have tacit knowledge of cash flows through a bank account. Probably very few students have translated their tacit understanding of stock-and-flow bathtub dynamics into the abstract language of mathematics. But those students who have developed an explicit understanding of first-derivative calculus may, with the tips offered in the simulation exercise, be able to relate their explicit understanding of first-derivative dynamics to their implicit understanding of bathtub dynamics.

The learning literature also documents the importance of expert/apprentice mentoring particularly in regards to "tips" or "rules of thumb" that are shared within a discipline or a knowledge community (Henning, 1998; Gick & Holyoak, 1980). Individuals do not readily transfer knowledge to new scenarios unless guided or tipped in advance. Reed, Ernst and Banerji (1974) examined the possibility that individuals use applicable learning and knowledge from a solved problem to solve an additional problem. They examined the transferability of the solution of the missionary-cannibal problem to the solution of the jealous-husband problem. The two problems are similar.

They found variable support for the hypothesis of learning transfer. In a similar study, Gick and Holyoak (1980) studied the learning transfer between two analogous problems, the radiation problem and the attack-dispersion problem. Once again they found that learning was not transferred from one problem to the next unless the subjects were prompted to make the connection between the two problems. However, once prompted, nearly all subjects were able to make the connections.

We propose that completion of the bathtub simulation exercise not only led to improvement of student performance on the Bathtub1 problem but also led to improvement on the Cashflow1 problem. Since the simulation exercise used the same graphical and structural cues, the students were “tipped off” regarding the structure of the cash flow problem even though the simulation exercise did not mention cash flow examples. We hypothesize that a wide range of experiences, and many translations among various levels of abstractions, are required for students to develop a robust understanding of generalized structures, such as the consequences of accumulating a rate that changes over time. Students may understand Cartesian coordinates and may be able to interpret graphs within the context of the graph itself, but translating that understanding, both into the concrete and physical realm as well as into the pure abstractions of math equations, apparently poses significant challenges.

Since many students do not appear to make connections readily between the abstractions of their math courses to their real-world experiences, we propose that simulations can offer intermediate abstractions. By providing interactive and visual experiences with rapid feedback, simulations can help students translate their tacit understanding of the dynamics in various circumstances into more general (and mathematically representable) structures that facilitate analytical understanding to complement intuitions that may or may not prove accurate. The learning literature supports this approach of “dual coding” provided through visual and interactive experiences (James & Galbraith, 1985; Carnevale et al., 1990; Gagne, 1977; Johari, 1998; Shu-Ling, 1998).

Student performance on the Bathtub2 and CashFlow2 problems, with linearly changing rates of inflow, suggests that even very simple dynamics systems are not well understood. It is uncommon in businesses that outflows will remain constant while only inflows change, or vice-versa. To be effective, business decision-makers must understand the difference between flows (such as annual profits) and stocks (such as retained earnings). In our experience, many undergraduate upper-division business students do not understand that income statement entries (sales, profits, expenses) are measured per time period (monthly, quarterly, or annually) and so are akin to the flow of water into and out of a bathtub. Similarly, they do not comprehend that balance sheet entries are stocks (snapshots of the accumulated business resources at a particular point in time) analogous to a bathtub's accumulated water. A resource-based view of business requires that decision-makers distinguish between stocks and activities measured per time period. Focusing on accumulated customers for a word-of-mouth marketing campaign will yield different and better results than focusing only on the rate of customer acquisition. Likewise, a business manager who does not understand that the accumulated customer base, rather than the new-customer acquisition rate, drives demands for after-sales service and maintenance risks losing

customers dissatisfied by staffing inadequate to meet customers' needs.

This study provides some support for the assertion that targeted simulation exercises can improve student understanding of stocks and flows in dynamic systems. The results were tempered, however, by concerns regarding the student sampling procedures. Future studies are needed to refine the simulation exercises and the experimental methods. Even so, the experiment's results raise provocative questions about how students learn (and do not learn) about dynamic systems:

- Why did students who completed a basic integral and differential calculus course perform poorly on tests of understanding regarding dynamic systems?
- How did the simulation exercise administered in this study improve student performance on tests of understanding regarding dynamic systems?
- Why did the performance results vary between the mathematically equivalent Bathtub and CashFlow problems?
- Why did performance not improve on the Bathtub2 and CashFlow2 problems (the ones with variable input flows)?

It has been demonstrated (Sterman, 1989; Dörner, 1996) that people's decision-making abilities deteriorate as system complexity (as indicated by the number of relationships among variables, and the delays between action and consequence) increases. Since the pace of change in businesses and markets is increasing over time, our educational endeavors will benefit from further research on how students learn, and fail to learn, to assess dynamic situations, so that we can improve the effectiveness of the instructional activities we provide.

## References

- Atwater, J.B, Kannan, V.R. & Stephens, A.A. (2008). Cultivating Systemic Thinking in the Next Generation of Business Leaders. *Academy of Management Learning & Education*, 7: 9-25.
- Carnevale, A.P., Gainer, L.J., Villet, J. & Holland, S.L. (1990). *Training Partnerships: Linking Employers and Providers*. Alexandria, VA: American Society for Training and Development.
- Dörner, D. (1996). *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*. New York: Metropolitan Books.
- Fenwick, T.J. (2000). Expanding Conceptions of Experiential Learning: A Review of the Five Contemporary Perspectives on Cognition, *Adult Education Quarterly*, 50: 243-272.
- Fine, C.H. (2000). Clockspeed-based Strategies for Supply Chain Management. *Production and Operations Management*, 9: 213-221.
- Fine, C.H. (1998). *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*. Reading, MA: Perseus Books.
- Gagne, R.M. (1970). *The Conditions of Learning*. New York: Holt, Rinehart and Winston.

- Gick, M.L. & Holyoak, K.J. (1980). Analogical Problem Solving. *Cognitive Psychology*, 12: 306-355.
- Henning, P.H. (1998). Ways of Learning: An Ethnographic Study of the Work and Situated Learning of a Group of Refrigeration Service Technicians. *Journal of Contemporary Ethnography*, 27: 85-136.
- James, W. & Galbraith, M.W. (1985). Perceptual Learning Styles: Implications and Techniques for the Practitioner. *Lifelong Learning*, 8: 20-23.
- Johari, A. (1998). Effects of Inductive Multimedia Programs Including Graphs on Creation of Linear Function and Variable Conceptualization. *Proceedings of Selected Research and Development Presentations at the National Convention of the Association for Educational Communications and Technology (ACET)*.
- Kaminski, J.A., Sloutsky, V.M. & Heckler, A.F. (2008). The Advantage of Abstract Examples in Learning Math. *Science*, 320: 454-455.
- Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. New York: Cambridge University Press.
- Mendelson, H. & Pillai, R.R. (1999). Industry Clockspeed: Measurement and Operational Implications. *Manufacturing & Service Operations Management*, 1: 1-20.
- Moxnes, E. (1998). Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics. *Management Science*, 44: 1234-1248.
- Raymond, B.C. & Black, L.J. Teaching and Learning: A Systems Perspective, *Proceedings of the 33rd Annual Meeting of the Western Decision Sciences Institute*, 2004: 197-199.
- Reed, S.K., Ernst, G.W. & Banerji, R. (1974). The Role of Analogy in Transfer between Similar Problem States. *Cognitive Psychology*, 6: 436-450.
- Schliemann, A.D. (1998). Logic of Meanings and Situated Cognition. *Learning and Instruction*, 8: 549-560.
- Shu-Ling, L. (1998). The Effects of Visual Display on Analogies Using Computer-Based Learning. *International Journal of Instructional Media*, 25: 151-159.
- Sterman, J.D. (1989). Misperceptions of Feedback in Dynamic Decision Making. *Organizational Behavior and Human Decision Processes*, 43: 301-335.
- Sweeney, L.B. & Sterman, J.D. (2000). Bathtub Dynamics: Initial Results of a Systems Thinking Inventory. *System Dynamics Review*, 16: 249-286.

## Appendix I

### Bathtub Stocks and Flows Structured Learning Assignment

Consider the bathtub shown below. Water flows in at a certain rate, and exits through the drain at another rate. If the water in the tub exceeds 252 liters, then the excess inflow runs out through an overflow drain.



The purpose of this exercise is to build your conceptual understanding of rates of change (flows) and accumulations (stocks). You will use a computerized simulation of a bathtub to assist you in this exercise.

**Familiarity** – To start you need to become familiar with the controls and data displays of the simulator.

Start	Start the simulation clock, water inflow must be $\geq$ zero
Pause	Pause the simulation clock, resume after pausing
Exit	Exit the simulation
Reset	Reset the simulation to the beginning zero values

**Input Parameters** – are used to set initial values before starting the simulation clock.

Each of the input parameters is increased/decreased by clicking the left button of the mouse on the up arrow or down arrow of the spinner control.

**Water Inflow Rate** (liters/min) – Use the spinner control to increase/decrease the rate at which water will flow into the bathtub; this control is a real-time control that can be used anytime during the simulation to change the rate at which water is flowing into the bathtub. The inflow rate must be greater than zero to start the simulation clock.

**Initial Water in Tub** (liters) – Use this control to start the simulation with up to 200 liters of water in the tub.

**Drain Flow Rate** (liters/min) – Set the initial drain rate of water from the tub.

**Output Variables** – display the current state of various aspects of the simulation.

These values provide feedback about what is happening in the bathtub.

**Water In Tub** (liters) – Displays the amount of water in the bathtub in liters.

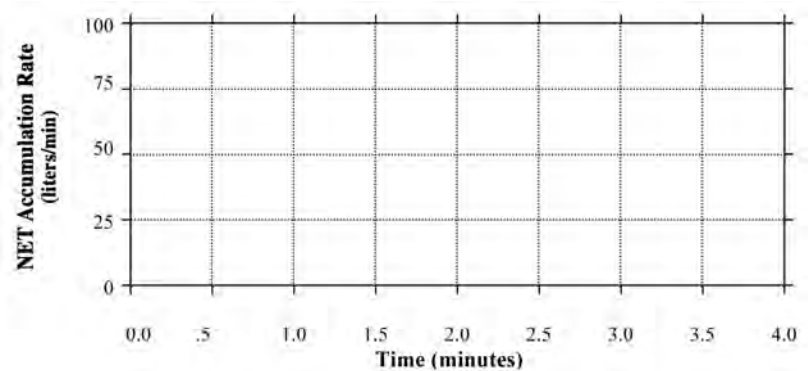
**Water Depth** (cm) – Displays the depth of the water in the bathtub in centimeters.

**Overflow Rate** (liters/min) – Displays the rate at which water is flowing through the overflow of the bathtub in liters per minute.



**Activity I: Constant rates flowing into and out of the tub**

- Before starting the simulation, use the spinner controls to set the inflow rate at 50 liters/min; the water in the tub to 75 liters, and the drain flow rate to 25 liters/min.
- Given the information above, at what rate will water be accumulating in the tub? \_\_\_\_\_
- Plot the constant net accumulation rate on the graph below over a 4 minute time period.



- Calculate the amount of water in the bathtub at the times shown below and enter your calculated results in the indicated row of the table. The first value is provided for you.

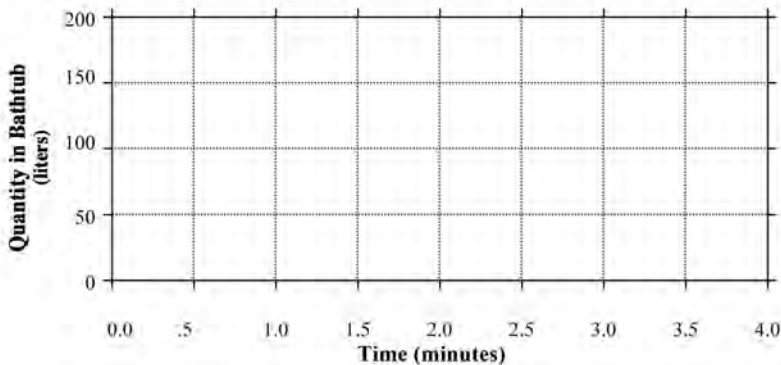
Time	1 min: 00 sec	1 min:30 sec	2 min: 45 sec	3 min: 00 sec
Calculated Water in Tub	100 liters			

*Hint:* accumulation at TimeNew = initial accumulation in *liters* + [(rate of inflow in *liters/minute*) \* (TimeNew – TimeStart in *minutes*)]  $y = f(x) = 75 + 25x$  where  $y$  = the liters of water in the tub and  $x$  = the elapsed time in minutes.

- Now start the simulation and pause the simulation at the times shown in the table to verify your calculations. Enter the values for the liters of water in the tub in the row for actual water in the tub. Are the actual values the same as the values you calculated? Why or why not?

Time	1 min: 00 sec	1 min:30 sec	2 min: 45 sec	3 min: 00 sec
Calculated Water in Tub	100 liters			

- If the Actual Water in Tub differed from your calculated amounts at any of the 4 times above, why is that? Re do your calculations and the simulations as often as you need until you understand how to calculate the effects of a constant rate on the accumulated water in the tub. You can experiment with changing the values of the inflow and drain rate to check your understanding.
- Using the graph below plot the four values for the actual amount of water in the tub at the four times noted above and the changes between those four points, along with the beginning value at time zero.



Is the graph linear or curved? Why?

Is it increasing or decreasing? Over what time periods?

If your graph is linear, calculate the slope of each line segment by considering how much the y value changes for each unit of x (i.e., use the traditional “rise/run” calculation). The x and y values for the first segment along with the required calculations are provided below as an example. Fill in the rest of the table.

Time Period	X1	X2	Y1	Y2	Slope = (Y2 – Y1)/(X2 – X1) =	Result
1. 00:00 – 01:00	0.0	1.0	75	100	$(100 - 75)/(1.0 - 0.0) =$	25
2. 01:00 – 01:30	1.0	1.5				
3. 01:30 – 02:45	1.5	2.75				
4. 02:45 – 03:00	2.75	3.0				

Note that the slope indicates the rate of change in the water in the tub. If water is accumulating, the slope is positive (and the line on the graph is increasing). If water is draining from the tub, the slope is negative (and the line you graphed is decreasing).

Are the slopes of the line segments you calculated the same value as the rate at which water is accumulating in the tub when you ran the simulation? You can run the simulation again, to make sure.

If the simulation continued to run would the bathtub eventually be full i.e. excess water would run out the overflow drain? If so, calculate the time at which the tub would be full. Run the simulation to verify your calculated answer.

Hint: Total Volume in liters = depth \* length \* width = (30 cm \* 140 cm \* 60 cm)/1000 = 252 liters

### Activity II: Changing accumulation of water in the tub

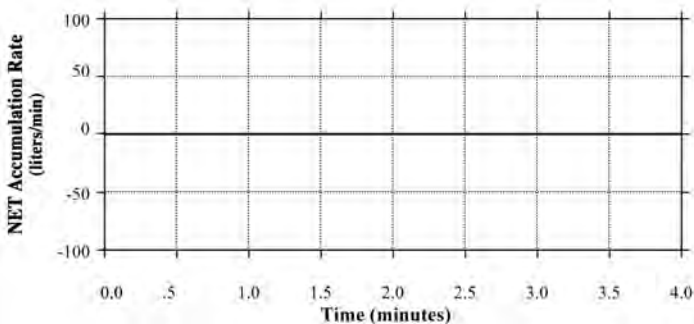
In the previous exercise, the inflow and outflow from the tub were constant. This time you will be changing the flows of water into, and draining from, the tub as time passes.

- The table below lists the initial values (0 min: 00 sec) and the changing inflow and outflow values. Before running the simulation calculate the net flow of water accumulating in or dispersing from the tub as well as the water in the tub.

Calculated Values (do these calculations prior to starting the simulation)

Time	Inflow Rate	Drain Rate	Net Accumulation Rate	Ending Water in Tub
0 min: 00 sec (initial values)	75 lit/min	25 lit/min		100 liters
0 min: 00 sec - 1 min: 00 sec	75 lit/min	25 lit/min	50 liters/minute	150 liters
1 min: 00 sec - 2 min: 00 sec	25 lit/min	50 lit/min		
2 min: 00 sec - 3 min: 00 sec	100 lit/min	0 lit/min		
3 min: 00 sec - 4 min: 00 sec	0 lit/min	50 lit/min		

- Plot the net accumulation rate on the graph below over the 4 minute time period.



- Now start the simulation and pause the simulation at the times shown in the table to adjust the inflow and drain rates; enter the simulation actual values at each time to verify your calculations.

**Actual Values (read from simulation)**

Time	Inflow Rate	Drain Rate	Ending Water in Tub
0 min: 00 sec (initial values)	75 lit/min	25 lit/min	100 liters
0 min: 00 sec - 1 min: 00 sec	75 lit/min	25 lit/min	
1 min: 00 sec - 2 min: 00 sec	25 lit/min	50 lit/min	
2 min: 00 sec - 3 min: 00 sec	100 lit/min	0 lit/min	
3 min: 00 sec - 4 min: 00 sec	0 lit/min	50 lit/min	

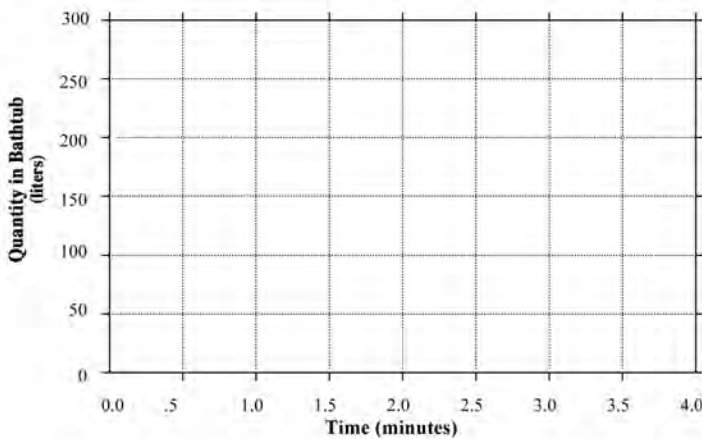
Did you calculate the volume of water in the tub at each point in time correctly? If not, continue experimenting with the controls on the simulator until you have an understanding of how

Inflow Rate – Drain Rate = Net Accumulation Rate,

and how, for each segment of time,

Net Accumulation Rate \* Duration + Previous Volume in Tub = New Volume in Tub

- Using the graph below, plot the actual amount of water in the tub.



Calculate the slope of each line segment by considering how much the y value changes for each unit of x (i.e., use the traditional “rise/run” calculation). In each time segment (0 – 1 min, 1 – 2 min, 2 – 3 min, 3 – 4 min) verify that the slope of your graph is equal to the Net Accumulation rates noted in the table above.

Note that the slope indicates the rate of change in the total volume of the water in the tub. If water is accumulating, the slope is positive (and the line on the graph is increasing). If water is draining from the tub, the slope is negative (and the line you graphed is decreasing). Are the slopes of the line segments you calculated the same value as the rate at which water is accumulating in the tub when you ran the simulation? You can run the simulation again, to make sure.

Time Period	X1	X2	Y1	Y2	Slope = $(Y2 - Y1)/(X2 - X1)$ =	Result
1. 00:00 – 01:00	0.0	1.0	100	150	$(150 - 100)/(1.0 - 0.0) =$	50
2. 01:00 – 02:00	1.0	2.0				
3. 02:00 – 03:00	2.0	3.0				
4. 03:00 – 04:00	3.0	4.0				

In the last time period if the simulation continued to run would the bathtub eventually be empty? Is so, calculate the time at which the bathtub would be empty. Run the simulation to verify your calculated answer.

### Activity III: Constantly changing accumulation of water in the tub

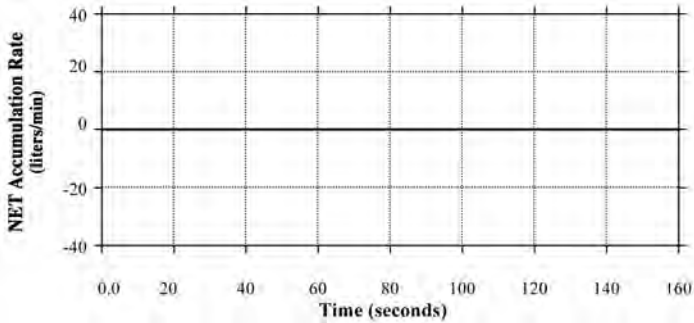
This time you will be changing the flows of water into, and draining from, the tub in very short time segments (every ten seconds). After working through a few time periods, you should be able to predict the shape of the curve for additional time periods.

- The table below lists the initial values (0 min: 00 sec) and the changing inflow and outflow values. Before running the simulation, calculate the net flow of water accumulating in or dispersing from the tub as well as the water in the tub. Remember to convert liters/min to liters/sec when calculating the water volume.

### Calculated Values

Time	Inflow Rate	Drain Rate	Accumulation Rate (liters/minute)	Accumulation Rate (liters/second)	Water in Tub
0 min: 00 sec (initial values)	50 lit/min	50 lit/min			100 liters
0 min: 00 sec - 0 min: 10 sec	50 lit/min	50 lit/min			
0 min: 10 sec - 0 min: 20 sec	60 lit/min	50 lit/min			
0 min: 20 sec - 0 min: 30 sec	70 lit/min	50 lit/min			
0 min: 30 sec - 0 min: 40 sec	80 lit/min	50 lit/min			
0 min: 40 sec - 0 min: 50 sec	90 lit/min	50 lit/min			
0 min: 50 sec - 1 min: 00 sec	80 lit/min	50 lit/min			
1 min: 00 sec - 1 min: 10 sec	70 lit/min	50 lit/min			
1 min: 10 sec - 1 min: 20 sec	60 lit/min	50 lit/min			
1 min: 20 sec - 1 min: 30 sec	50 lit/min	50 lit/min			
1 min: 30 sec - 1 min: 40 sec	40 lit/min	50 lit/min			
1 min: 40 sec - 1 min: 50 sec	30 lit/min	50 lit/min			
1 min: 50 sec - 2 min: 00 sec	20 lit/min	50 lit/min			
2 min: 00 sec - 2 min: 10 sec	10 lit/min	50 lit/min			
2 min: 10 sec - 2 min: 20 sec	20 lit/min	50 lit/min			
2 min: 20 sec - 2 min: 30 sec	30 lit/min	50 lit/min			
2 min: 30 sec - 2 min: 40 sec	40 lit/min	50 lit/min			

- Plot the net accumulation rate in liters/minute on the graph below over the 160 second time period.



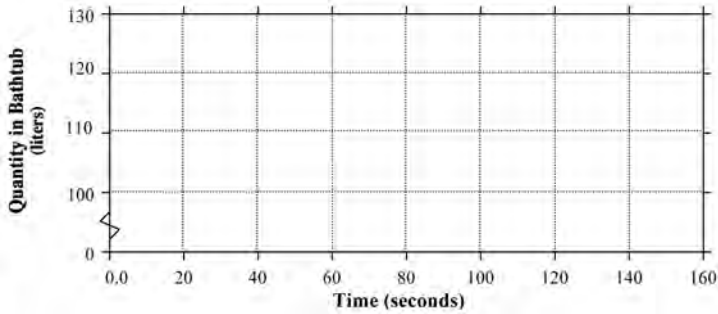
- Now run the simulation pausing at the times shown in the table to verify your calculations.

**Actual Values (read from simulation)**

Time	Inflow Rate	Drain Rate	Water in Tub
0 min: 00 sec (initial values)	50 lit/min	50 lit/min	100 liters
0 min: 00 sec - 0 min: 10 sec	50 lit/min	50 lit/min	
0 min: 10 sec - 0 min: 20 sec	60 lit/min	50 lit/min	
0 min: 20 sec - 0 min: 30 sec	70 lit/min	50 lit/min	
0 min: 30 sec - 0 min: 40 sec	80 lit/min	50 lit/min	
0 min: 40 sec - 0 min: 50 sec	90 lit/min	50 lit/min	
0 min: 50 sec - 1 min: 00 sec	80 lit/min	50 lit/min	
1 min: 00 sec - 1 min: 10 sec	70 lit/min	50 lit/min	
1 min: 10 sec - 1 min: 20 sec	60 lit/min	50 lit/min	
1 min: 20 sec - 1 min: 30 sec	50 lit/min	50 lit/min	
1 min: 30 sec - 1 min: 40 sec	40 lit/min	50 lit/min	
1 min: 40 sec - 1 min: 50 sec	30 lit/min	50 lit/min	
1 min: 50 sec - 2 min: 00 sec	20 lit/min	50 lit/min	
2 min: 00 sec - 2 min: 10 sec	10 lit/min	50 lit/min	
2 min: 10 sec - 2 min: 20 sec	20 lit/min	50 lit/min	
2 min: 20 sec - 2 min: 30 sec	30 lit/min	50 lit/min	
2 min: 30 sec - 2 min: 40 sec	40 lit/min	50 lit/min	

Were the amounts you calculated for each cell validated by the simulation? For any cells that differ, please make sure you understand the source of discrepancy and make sure you can calculate the correct value for each cell.

- Using the graph below plot the amount of water in the tub (time in seconds). Note the broken vertical axis.



In this final exercise the rate of change of water in the tub varies between each time segment. We can therefore make some observations about the rate of change in the rate of change. When the Net Accumulation Rate is increasing, but by less each time segment (we call that “increasing at a decreasing rate,” the graph approaches a peak, or local maximum. Similarly, when the Net Accumulation Rate is decreasing, but by less each time segment (decreasing at a decreasing rate), the graph approaches a local minimum.

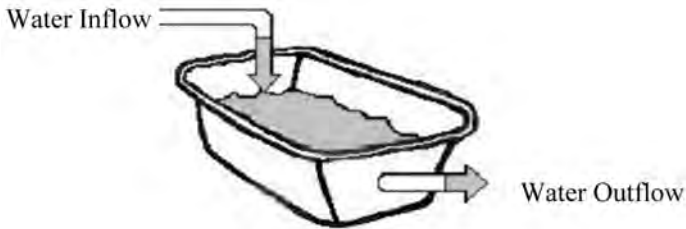
## Appendix II

### Four problems for assessing dynamic system knowledge.

(Used by permission of the author.)

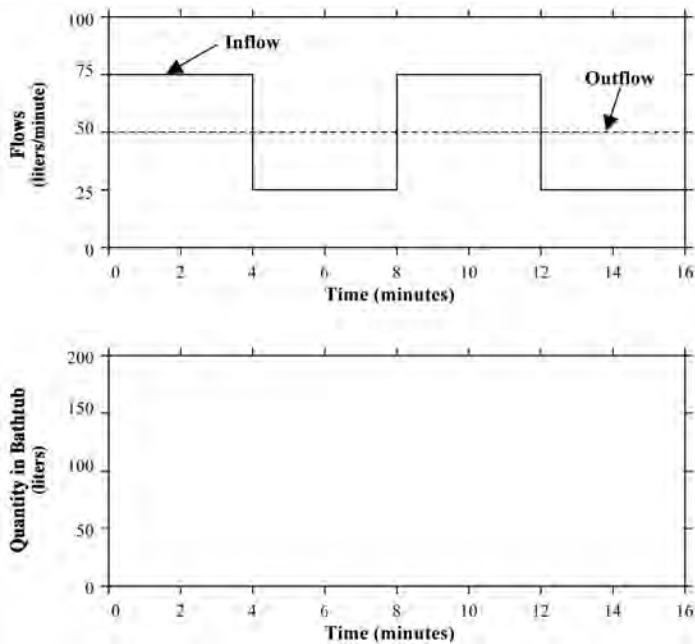
#### Bathtub 1: Square Wave Pattern<sup>1</sup>

Consider the bathtub shown below. Water flows in at a certain rate, and exits through the drain at another rate:



The graph below shows the hypothetical behavior of the inflow and outflow rates for the bathtub. From that information, draw the behavior of the quantity of water in the tub on the second graph below.

Assume the initial quantity in the tub (at time zero) is 100 liters.

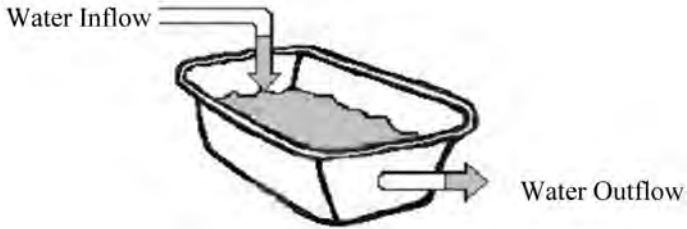


<sup>1</sup>Adapted from Sweeney, L.B. and Sterman, J.D. See references.



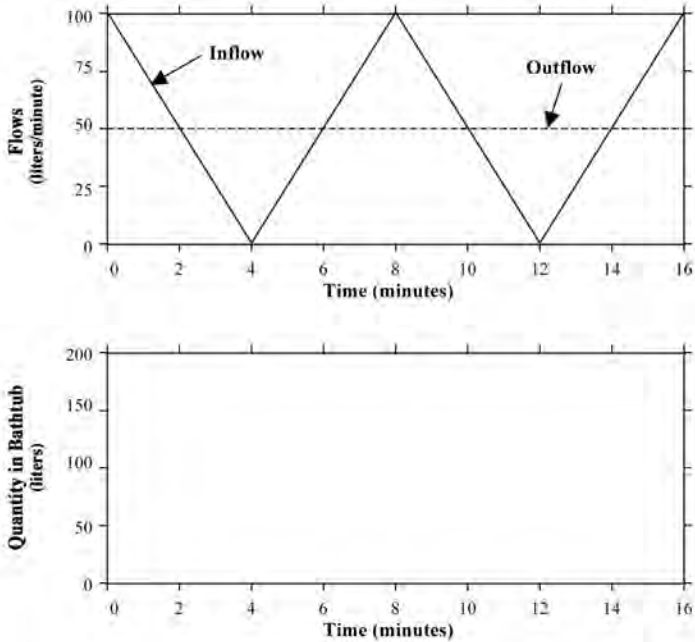
**Bathtub 2: Sawtooth Pattern<sup>2</sup>**

Consider the bathtub shown below. Water flows in at a certain rate, and exits through the drain at another rate:



The graph below shows the hypothetical behavior of the inflow and outflow rates for the bathtub. From that information, draw the behavior of the quantity of water in the tub on the second graph below.

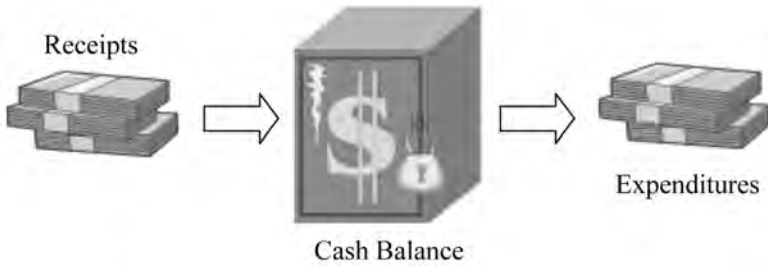
Assume the initial quantity in the tub (at time zero) is 100 liters.



<sup>2</sup>Adapted from Sweeney, L.B. and Sterman, J.D. See references.

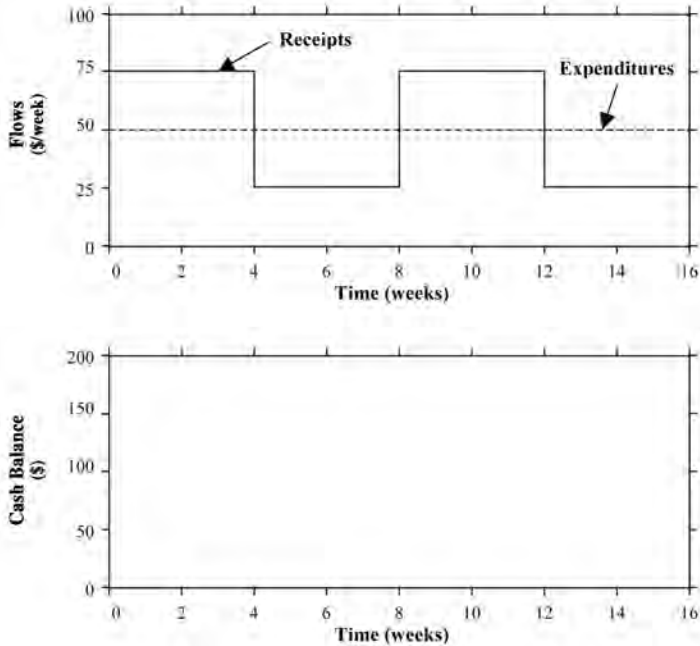
**Cash Flow 1: Square Wave Pattern<sup>3</sup>**

Consider the cash balance of a company. Receipts flow into the balance at a certain rate, and expenditures flow out at another rate:



The graph below shows the hypothetical behavior of receipts and expenditures. From that information, draw the behavior of the firm's cash balance on the second graph below.

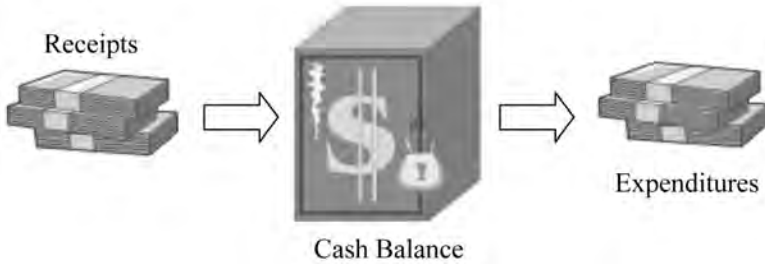
Assume the initial cash balance (at time zero) is \$100.



<sup>3</sup>Adapted from Sweeney, L.B. and Sterman, J.D. See references.

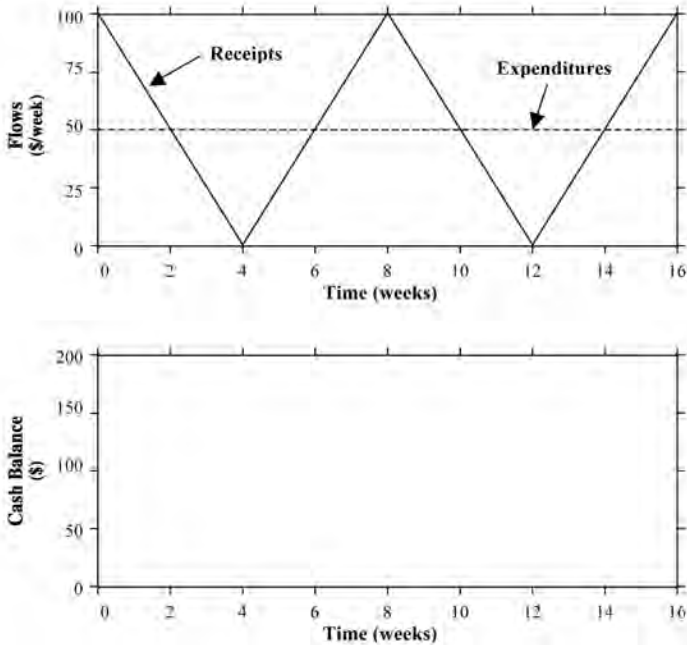
**Cash Flow 2: Sawtooth Pattern<sup>4</sup>**

Consider the cash balance of a company. Receipts flow into the balance at a certain rate, and expenditures flow out at another rate:



The graph below shows the hypothetical behavior of receipts and expenditures. From that information, draw the behavior of the firm's cash balance on the second graph below.

Assume the initial cash balance (at time zero) is \$100.



<sup>4</sup>Adapted from Sweeney, L.B. and Sterman, J.D. See references.

### Appendix III

#### Demographics of the Student Population:

Demographics are provided for the entire course population, along with the demographics for the four problem assignment groups on both the pretest and the posttest. The course population was defined as the students enrolled on the 15th class day, and the sample of students who actually took the pretest and the posttest depended on attendance on that particular class day.

Self-reported ethnicity of the respondents was obtained from university records; as shown below the majority of the population was Caucasian.

- Asian: 5
- Hispanic: 7
- Native American: 4
- Other: 3
- Missing Data: 24
- Caucasian: 340

**Table AIII-1: Age Distribution**

Age	Population		Pretest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
<=18	92	24.02%	70	26.79%	21	33.33%	16	28.57%	22	26.44%	11	18.64%
<=20	125	32.64%	74	27.92%	17	26.98%	21	37.50%	17	19.54%	19	32.20%
<=22	82	21.41%	55	20.75%	12	19.05%	10	17.86%	15	17.24%	18	30.51%
<=24	35	9.14%	17	6.42%	5	7.94%	4	7.14%	5	5.75%	3	5.08%
<=26	17	4.44%	10	10.94%	3	4.76%	0	0.00%	2	24.14%	5	8.47%
<=28	11	2.87%	6	2.26%	2	3.17%	1	1.79%	3	3.45%	0	0.00%
>28	21	5.48%	13	4.91%	3	4.76%	4	7.14%	3	3.45%	3	5.08%
Total	383		245		63		56		67		59	

Age	Population		Posttest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
<=18	92	0.2402%	40	0.2286%	12	0.2727%	9	0.2093%	8	0.1905%	11	0.2391%
<=20	125	0.3264%	60	0.3429%	14	0.3182%	17	0.3953%	10	0.2381%	19	0.413%
<=22	82	0.2141%	35	0.2%	8	0.1818%	9	0.2093%	10	0.2381%	8	0.1739%
<=24	35	0.0914%	16	0.0914%	2	0.0455%	4	0.093%	6	0.1429%	4	0.087%
<=26	17	0.0444%	7	0.04%	2	0.0455%	1	0.0233%	1	0.0238%	3	0.0652%
<=28	11	0.0287%	5	0.0286%	1	0.0227%	0	0%	4	0.0952%	0	0%
>28	21	0.0548%	12	0.0686%	5	0.1136%	3	0.0698%	3	0.0714%	1	0.0217%
Total	383		175		44		43		42		46	

Table AIII-2: Gender of Respondents

Gender	Population		Pretest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Female	178	46.48%	107	43.67%	28	44.44%	25	44.64%	27	40.30%	27	45.76%
Male	205	53.52%	138	56.33%	35	55.56%	31	55.36%	40	59.70%	32	54.24%
Total	383		245		63		56		67		59	

Gender	Population		Posttest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Female	178	46.48%	81	46.29%	23	52.27%	26	60.47%	14	33.33%	18	39.13%
Male	205	53.52%	94	53.71%	21	47.73%	17	39.53%	28	66.67%	28	60.87%
Total	383		175		44		43		42		46	

Table AIII-3: Class standing of respondents

Class	Population		Pretest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Freshman	190	49.61%	126	51.22%	34	53.97%	32	56.14%	36	53.73%	24	40.68%
Sophomore	98	25.59%	62	25.20%	16	25.40%	14	24.56%	15	22.39%	17	28.81%
Junior	56	14.62%	36	14.63%	10	15.87%	4	7.02%	10	14.93%	12	20.34%
Senior	32	8.36%	16	6.50%	3	4.76%	5	8.77%	3	4.48%	5	8.47%
*Other	7	1.83%	6	2.44%	0	0.00%	1	3.51%	3	4.48%	1	1.69%
Total	383		245		63		56		67		59	

Class	Population		Posttest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Freshman	190	49.61%	87	49.71%	21	47.73%	22	51.16%	17	40.48%	27	58.70%
Sophomore	98	25.59%	47	26.86%	15	34.09%	9	20.93%	13	30.95%	10	21.74%
Junior	56	14.62%	23	13.14%	4	9.09%	11	25.58%	4	9.52%	4	8.70%
Senior	32	8.36%	15	8.57%	4	9.09%	1	2.33%	5	11.90%	5	10.87%
*Other	7	1.83%	3	1.71%	0	0.00%	0	0.00%	3	7.14%	0	0.00%
Total	383		175		44		43		42		46	

\*Other includes international and non-degree students

Table AIII-4: Academic major of the respondents

Major	Population		Pretest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Ag	42	10.97%	97	39.59%	5	7.94%	32	57.14%	36	53.73%	24	40.68%
Bus	131	34.20%	64	26.12%	18	28.57%	14	25.00%	15	22.39%	17	28.81%
Sci/Tech	102	26.63%	46	18.78%	20	31.75%	4	7.14%	10	14.93%	12	20.34%
Other	108	28.20%	38	15.51%	20	31.75%	6	10.71%	6	8.96%	6	10.17%
Total	383		245		63		56		67		59	

Major	Population		Posttest		Bathtub1		CashFlow1		Bathtub2		CashFlow2	
	#	%	#	%	#	%	#	%	#	%	#	%
Ag	42	10.97%	17	9.71%	7	15.91%	3	6.98%	4	9.52%	3	6.52%
Bus	131	34.20%	60	34.29%	16	36.36%	15	34.88%	12	28.57%	17	36.96%
Sci/Tech	102	26.63%	48	27.43%	11	25.00%	16	37.21%	10	23.81%	11	23.91%
Other	108	28.20%	50	28.57%	10	22.73%	9	20.93%	16	38.10%	15	32.61%
Total	383		175		44		43		42		46	



# Data Mining Methods in the Detection of Spam

Dong-Her Shih

Hsiu-Sen Chiang

Chia-Shyang Lin

*National Yunlin University of Science and Technology*

Ming-Hung Shih<sup>1</sup>

*National Chiao Tung University*

*The spam problem has generated enormous costs for companies and users of the Internet. Internet users not only pay for the bandwidth to bring in volumes of spam mail but also pay for its storage. In this paper, we propose a modified Naïve Bayesian classifier and compare it with three data mining methods for identifying whether incoming mail is spam or legitimate automatically. The experimental results show that although there is no dominant algorithm to the spam problem, generally the decision tree has the better performance. Our proposed modified Naïve Bayesian classifier has the potential for further investigation as well.*

## Introduction

Unsolicited Bulk Email (UBE), also referred to as Unsolicited Commercial Email (UCE), is commonly called spam or “junk mail.” Spamming is the practice of sending mass mailings to large numbers of people who have no relationship with the sender and who didn’t request the mail. According to research conducted by Microsoft and published by the Radicati Group, the percentage of spam mail in the total number of emails sent daily has been consistently growing since 2005. As a result, spam is expected to represent 77% of emails sent worldwide by the end of 2009, amounting to almost 250 billion unsolicited emails delivered every day.

---

<sup>1</sup>The authors would like to give thanks to the National Science Council of Taiwan for their grant (NSC93-2213-E-224-038) to perform part of this research.

Competitive anti-spam products, legislation, and efforts towards a better user education have all been used in an attempt to stop spam. However, unsolicited emails keep consuming the space and time of all email users. Moreover, spam messages can be the cause of serious virus and spyware outbreaks, while others “phish” for sensitive information like bank accounts and passwords. SPAMHAUS says that spammers are carrying out a dictionary attack on hotmail.com. The spammers connect to victims’ mail servers and submit millions of random email accounts in common words and names (e.g. michaelFxy2@\_.com, marla1892@\_.com), recording which addresses succeed and add these automatically to their list. They also send spam to the variation of the account name (e.g. Lidia@\_.com to lidia@\_.com) without collecting the victims’ email accounts (Spammers Grab MSN Hotmail addresses, 2007).

As email becomes an important medium of communication, with direct impact on human relations and business, spam email causes considerable damage to the users and the entire internet foundation. For instance, organizations such as those in the financial and healthcare industries are required by law to archive email for up to seven years. They not only pay for the bandwidth to bring in volumes of spam mail but also pay for its storage (Paulson, 2003).

Due to the worsening spam problem, several studies have been conducted, ranging from technical to legal. In the technical respect, filtering is currently the most widely used method. The filter can be implemented on either the server end (mail transport agent, MTA) or the user’s end (mail user agent, MUA). Motivated by those previous works, in this study, we propose an enhanced Naïve Bayesian classifier method and compare it with three data mining methods in order to identify whether incoming mail is spam or legitimate automatically: ID3 Decision Tree, Naive Bayesian Filter, and R.A Fisher’s Probability Combination Method. The performance-measuring result shows that the ID3 decision tree has better performance, in general.

In the next section, we will briefly describe some related methods used to fight spam. Then, we will outline the three approaches: ID3 Decision Tree, Naive Bayesian, and R.A Fisher’s Probability combination method, and our enhanced Naive Bayesian methods in section 3. The experimental method and the performance measures will be provided in section 4. We will show and analyze the experiment results in the last section of this paper, and also give a brief conclusion.

## Related Methods

During these years, the CAN-SPAM Act of 2003 became the most well known out of all the regulatory solutions. The CAN-SPAM Act seeks to control rather than outright ban spam by filtering and forbidding deceptive email messages, which include either misleading email headers (routing information), often referred to as “header forging,” fake return addresses, or misleading subject lines.

Legal scholars also note that CAN-SPAM’s greatest deficiency is that it supersedes and nullifies much stricter state laws. The CAN-SPAM Act may be an important step for the spam regulation, but must be amended in order to provide further protection. In a world where spam holds such an important position, methods of preventing it should also be given increasing importance. To combat spam, multiple filtering technologies



have been developed that weed out most, but not all of the unsolicited email.

*Blacklist:* A blacklist spam filter can be a Domain Name System-based (DNS-based) or email-address-based blacklist. A DNS-based blacklist is a means by which an Internet site may publish a list of IP addresses that some people may want to avoid, in a format that can be easily queried by computer programs on the Internet. The technology is built on top of the Internet DNS. DNSBLs are chiefly used to publish lists of addresses linked to spamming. Most mail transport agent (mail server) software can be configured to reject or flag messages which have been sent from a site listed on one or more such lists. The email address-based blacklist utilizes the full email address of the user (i.e. a domain-based identifier plus the local part). Since the local part can also be spoofed, validation mechanisms must be in place. Examples of those are Lightweight Directory Access Protocol (LDAP) and Active Directory (Alperovitch, Judge & Krasser, 2007).

Blacklist is very useful at the ISP level, but it has several weaknesses also. First, more than half of the spam mail servers are not on the blacklist. Second, the effect of the blacklist depends on the administrator of the blacklist. If the blacklist is wrong, it is possible that legitimate emails may also get filtered in the process.

*Signature-based Filtering:* The method of signature-based filtering compares incoming email with the spam that has already been received. In order to know whether two emails are the same, the filter calculates “signatures” for them. Signature-based filters rarely block legitimate mails, but its weakness is that spammers can add elements to each email and give it a distinct signature, thereby tricking the signature-based filters.

*Rule-Based Filtering:* Rule-based filters try to discover the patterns found in many spam messages (e.g. words or phrases, malformed headers and misleading dates). SpamAssassin, a popularly used open-source spam filter, uses a large set of heuristic rules. But the main disadvantage of rule-based filters like SpamAssassin is that they tend to have high false-positive rates (O'Brien & Carl, 2003).

*Text Classification Filtering:* A text classification filter uses the text classification technique to filter spam. There have been several studies done on this application, including keyword-based, phrase-based, and character-based studies. The Naïve Bayes-based method is also another efficient approach of keyword and phrase-based studies which use features extracted from emails. Additionally, Support Vector Machine (SVM), Rocchio, and decision tree filtering based on the ID3, C4.5, or C5 algorithms can be identified as the representative methods to analyze keywords in email (Schapire & Singer, 2000; Drucker, Wu & Vapnik, 1999; Joachims, 1997; Quinlan, 1993).

More recently, text categorization techniques are being applied in anti-spam research. As mentioned earlier, the Naïve Bayesian classifier is a widely used method. This paper attempts to demonstrate the performance of the Naïve Bayesian classifier method by using the concept of integrated multi-attribute and also by incorporating information Gain (IG) techniques in extracting and computing the weights of feature terms.

## Methodology

In this section we will describe three different data mining methods that are used to generate classifiers that detect whether incoming mail is spam or legitimate. These algorithms include Naïve Bayes, R. A. Fisher's probability combination method, and ID3 decision tree. The Naïve Bayesian method has been used several times before to filter spam email (Androutsopoulos et al., 2000; Mehran et al., 1998) and has overall been very effective. On the other hand, when a Bayes-like method is proposed, it can release the independent assumption by combining Graham and Fisher's method of filtering spam (Robinson, 2003). We took Robinson's approach for generating probabilities associated with words, altered it slightly, and proposed a Bayesian calculation for dealing with words that hadn't appeared very often in the spam messages. Then, we took the approach based on the chi-square distribution for combining the individual word probabilities and turned it into a combined probability representing an email. In order to distinguish spam from useful email efficiently, we adopted an ID3 decision tree to produce some rules. Using those rules is an easy way to ensure that only valid emails reach recipients and can educate users to help in preventing spam distribution.

### Naïve Bayes Classifier

A Naïve Bayes classifier computes the likelihood that an email is spam given the features that are contained in the email itself. Assuming that there were similar contents in spam emails that differentiated them from legitimate emails, the class of legitimate emails had similar patterns that differentiated them from the spam emails. The model output by the Naïve Bayes algorithm labels emails based on their contents.

The Naïve Bayes algorithm computes the probability that a given feature is spam and the probability that a feature is legitimate by computing statistics on the set of training data. Then, to predict whether a mail is spam or legitimate, those probabilities are computed in the classifier and the Naïve Bayes independence assumption is used. The independence assumption is then applied in order to efficiently compute the probability that an email was spam or legitimate.

In the Naïve Bayes anti-spam method, each mail is represented by a vector  $\vec{x} = \langle x_1, x_2, x_3, \dots, x_n \rangle$ , where  $x_1, x_2, x_3, \dots, x_n$  are the values of attributes  $X_1, X_2, X_3, \dots, X_n$ . As shown previously, we following Sahami, et al. (1998) and using binary attributes, i.e.  $X_i = 1$  if the email has the property represented by  $X_i$  (in our case, a specific word), and  $X_i = 0$  otherwise.

Given the vector  $\vec{x} = \langle x_1, x_2, x_3, \dots, x_n \rangle$ , of email, and where  $k \in \{spam, legitismate\}$ , the probability that an email belongs to category  $c$  is:

$$P(C = c | \vec{X} = \vec{x}) = \frac{P(C = c) \cdot P(\vec{X} = \vec{x} | C = c)}{\sum_{k \in \{spam, legitismate\}} P(C = k) \cdot P(\vec{X} = \vec{x} | C = k)} \quad (1)$$

Androutsopoulos, et al. (2000) notes that the probabilities  $P(\vec{X} | C)$  are almost impossible to calculate, because the possible values of vector  $X$  are too many and there

are also data sparseness problems. The Naïve Bayes filter assumes that  $X_1, X_2, X_3, \dots, X_n$  are conditionally independent given the category  $C$ , which allows us to calculate  $P(C = c | \vec{X} = \vec{x})$  as:

$$P(C = c | \vec{X} = \vec{x}) = \frac{P(C = c) \cdot \prod_{i=1}^n P(X_i = x_i | C = c)}{\sum_{k \in \{\text{spam}, \text{legitimate}\}} P(C = k) \cdot \prod_{i=1}^n P(X_i = x_i | C = k)} \quad (2)$$

$P(X_i | C)$  and  $P(C)$  are easy to estimate from the frequencies of the training data. A large number of empirical studies have found the Naïve Bayes filter to be surprisingly effective, despite the fact that the assumption that  $X_1, X_2, X_3, \dots, X_n$  are conditionally independent is usually overly simplistic (Domingos & Pazzani, 1996; Langley, Wayne & Thompson, 1992).

#### *Fisher's Probability Combination Method*

Robinson (2003) proposed a Bayes-like method that can release the independent assumption through R. A. Fisher's method to combine probability. For each word that appears in the training data we calculate:

$$b(w) = \frac{\text{number of spam containing the word } w}{\text{total number of spam}} \quad (3)$$

$$g(w) = \frac{\text{number of legitimate mail containing the word } w}{\text{total number of legitimate mail}} \quad (4)$$

$$p(w) = \frac{b(w)}{b(w) + g(w)} \quad (5)$$

$p(w)$  can be interpreted as the probability that a randomly chosen email address containing the word "w" will be spam. There is a problem with the probabilities calculated above as when some words are very rare in the training set. For instance, if a word appears in exactly one email and is a spam, the value of  $p(w)$  is 1.0. Clearly, it is not a good idea to classify all future emails that contain that same word as spam. In fact, the situation is such that we simply don't have enough data to know the real probability.

Virtually any word can appear in either a spam or non-spam message, and those of data points are not enough to be completely certain that we know the real probability. The Fisher's probability combination approach lets us combine our general background information with the data we have collected for a word in such a way that both aspects are given their proper importance. In this way, we determine an appropriate degree of belief about whether, when we see the word again, it will be in a spam message. We calculate this degree of belief,  $f(w)$ , as follows:

$$f(w) = \frac{(s \times x) + (n \times p(w))}{s + n} \quad (6)$$

where:

$s$ : the strength we want to give to our background information

$x$ : our assumed probability, based on our general background information, that a word we don't have any other experience of will first appear in a spam

$n$ : the number of emails we have received that contain word

In practice, the values for  $s$  and  $x$  are found through testing to optimize performance. Reasonable starting points are 1 for  $s$  and 0.5 for  $x$ .

In the proposed method, we should first calculate  $(-2)\ln(p_1 \times p_2 \times \dots \times p_n)$ . Then, consider the result to have a Chi-square with  $2n$  degrees of freedom, and use Chi-square table to compute the probability. The "spamness" probability of an email that contains specific  $w$  is:

$$H = C^{-1}[-2 \ln \prod_w f(w), 2n] \quad (7)$$

where:

$H$ : the "spamness" probability of a mail

$C^{-1}$ : the inverse Chi-square function, used to derive a p-value from a Chi-square distributed random variable.

### *ID3 Decision Tree*

A decision tree is similar to a flow chart. In order to classify an unknown sample, the attribute values of the sample are tested against the decision tree. A path is traced from the root to a leaf node that holds the class prediction for that sample. Decision trees can easily be converted to classification rules. Decision trees have been used in many application areas ranging from medicine, to game theory and business. The basic algorithm used in decision tree induction is the greedy algorithm which constructs decision trees in a top-down, recursive, and divide-conquer manner. The algorithm summarized in Figure 1 is ID3, a well-known decision tree induction algorithm (Han & Kamber, 2001).

The information gain measure is used to select the test attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. This attribute minimizes the information needed to classify the samples in the resulting partitions and reflects the least randomness or "impurity" in these partitions. Such an information-theoretic approach minimizes the expected number of tests needed to classify an object and guarantees that a simple (but not necessarily the simplest) tree is found.

**Figure 1:** Basic algorithm for inducing a decision tree from training samples

**Algorithm:** Generate\_decision\_tree. Generate a decision tree from the given training data.

**Input:** The training samples, represented by discrete-valued attributes, the set of candidate attributes, attribute-list.

**Output:** A decision tree.

**Method:**  
 create a node N;  
 if samples are all of the same class, C then  
   return N as a leaf node labeled with the class C;  
 if attribute-list is empty then  
   return N as a leaf node labeled with the most common class in samples; // majority voting  
 select test-attribute, the attribute among attribute-list with the highest information gain;  
 label node N with test-attribute;  
 for each known value  $a_i$ ; of test-attribute // partition the samples  
   grow a branch from node N for the condition test-attribute =  $a_i$ ;  
   let  $s_i$  be the set of samples in samples for which test-attribute =  $a_i$ ; // a partition  
   if  $s_i$  is empty then  
     attach a leaf labeled with the most common class in samples;  
   else attach the node returned by Generate\_decision\_tree ( $s_i$ , attribute-list-test-attribute);

Let  $S$  be a set consisting of  $s$  data samples. Suppose the class label attribute has  $m$  distinct values defining  $m$  distinct classes,  $C_i$  (for  $i=1, \dots, m$ ). Let  $s_i$  be the number of samples of  $S$  in class  $C_i$ . The expected information needed to classify a given sample is given by:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (8)$$

Where  $p_i$  is the probability that an arbitrary sample belongs to class  $C_i$  and is estimated by  $s_i/s$ . Note that a log function to the base 2 is used since the information is encoded in bits. Let attribute  $A$  have  $v$  distinct values,  $\{a_1, a_2, \dots, a_v\}$ . Attribute  $A$  can be used to partition  $S$  into  $v$  subsets,  $\{S_1, S_2, \dots, S_v\}$ , where  $S_j$  contains those samples in  $S$  that have value  $a_j$  of  $A$ . If  $A$  were selected as the test attribute (i.e., the best attribute for splitting), then these subsets would correspond to the branches grown from the node containing the set  $S$ . Let  $a_{ij}$  be the number of samples of class  $C_i$  in a subset  $S_j$ . The entropy, or expected information based on the partitioning into subsets by  $A$ , is given by

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \quad (9)$$

The term  $\frac{s_{1j} + \dots + s_{mj}}{s}$  acts as the weight of the  $j$ th subset and is the number of

samples in the subset (i.e., having value  $a_j$  of  $A$ ) divided by the total number of samples in  $S$ . The smaller the entropy value is, the greater the purity of the subset partition. Note that for a given subset  $S_j$ ,

$$I(s_{1j}, s_{2j}, \dots, s_{mj}) = - \sum_{i=1}^m p_{ij} \log_2(p_{ij}) \tag{10}$$

Where  $p_{ij} = \frac{S_{ij}}{|S_j|}$  and is the probability that a sample in  $S_j$  belongs to class  $C_i$ .

The encoding information that would be gained by branching on  $A$  is

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \tag{11}$$

In other words,  $Gain(A)$  is the expected reduction in entropy caused by knowing the value of attribute  $A$ . The algorithm computes the information gain of each attribute. The attribute with the highest information gain is chosen as the test attribute for the given set  $S$ . A node is created and labeled with the attribute, branches are created for each value of the attribute, and the samples are partitioned accordingly. When a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Tree pruning methods address this problem of over fitting the data. Such methods typically use statistical measures to remove the least reliable branches, generally resulting in faster classification and an improvement in the ability of the tree to correctly classify independent test data.

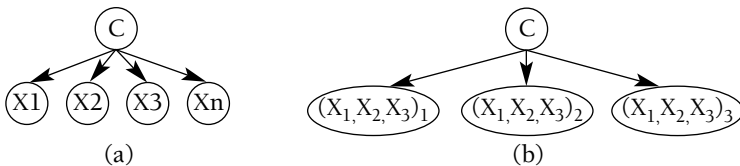
*A Novel Modified Naïve Bayesian Classifier*

As we mentioned previously, the probabilities  $P(\vec{X} | C)$  in equation (1) are almost impossible to calculate, because the possible values of vector  $X$  are too many and we may not have enough data to derive the value. Under the independent assumption, the Naïve Bayesian filter allows us to compute  $P(\vec{X} | C)$  as  $\prod P(X = x | C = c)$ . For example, we change our view to the vector  $\vec{x} = \langle x_1, x_2, x_3, \dots, x_n \rangle$  into  $\vec{x}' = \langle (x_1, x_2, x_3)_1, \dots, (x_1, x_2, x_3)_k \rangle$ , then we calculate  $P(C = c | \vec{X}' = \vec{x}')$  as:

$$P(C = c | \vec{X}' = \vec{x}') = \frac{P(C = c) \cdot \prod_{j=1}^k P[(x_1, x_2, x_3)_j | C = c]}{\sum_{k \in \{spam, legitimate\}} P(C = k) \cdot \prod_{j=1}^k P[(x_1, x_2, x_3)_j | C = c]} \tag{12}$$

The main difference is the posterior probability needed by the classifier. Figure 2 contrasts the structure of the Naïve Bayesian Classifier with our modified Bayesian Classifier.

**Figure 2:** Naïve Bayesian model (a), and the modified Naïve Bayesian model (b)



The logic behind our method is that the more dimensions of the feature space, the more accuracy of conditional probability we can get. It will also give us the chance to derive more information from the email content.

## Experiments

In this section, we will use the Spam Email Database from the UCI Machine Learning Repository to train and test the algorithms previously described.

### Data Set

The Spam Email Database was created by Hewlett-Packard Labs (UCI, 2004). It had been used for the HP internal-only technical report and other spam detection studies. The database contains 4601 instances. Each instance has 58 attributes (57 continuous, 1 nominal class label for the identification of spam and legitimate mail). In order to test the algorithms, we choose the 48 word attributes and transform the continuous attributes into 1 (has the specific word of the attribute) or 0 (has not). We randomly choose 50% instances (2282 instances) for the algorithm training and 2319 instances for the testing. The data set and its usage in our study is summarized as follows:

**Table 1:** Descriptive statistics and the usage of the data set

Total	
4601 instances	
Spam: 1813 (39.4%)	
Legitimate: 2788 (60.6%)	
Training Data	Testing Data
2282 instances	2319 instances
Spam: 905 (39.7%)	Spam: 908 (39.15%)
Legitimate: 1377 (60.34%)	Legitimate: 1411 (60.85%)

### Naïve Bayes

In order to calculate the “spamness” through Naïve Bayes method, first, we have to calculate the posteriori probability with the training data, as shown in Table 2.

**Table 2:** The posteriori probability (For Naïve Bayes)

$X_i$	make	address	free	business	credit	money	...	order
$P(X_i = 0   \text{SPAM})$	64.5%	66.4%	45.0%	62.3%	78.8%	61.5%	...	58.4%
$P(X_i = 1   \text{SPAM})$	35.5%	33.6%	55.0%	37.7%	21.2%	38.5%	...	41.6%
$P(X_i = 0   \text{Legitimate})$	86.0%	89.8%	91.1%	90.5%	98.2%	88.4%	...	98.8%
$P(X_i = 1   \text{Legitimate})$	14.0%	10.2%	8.9%	9.5%	1.8%	11.6%	...	1.2%

After the posterior was derived from the training data, we can calculate the probability of an email being spam through its feature vector and equation (2). The performance of the Naïve Bayes filter will be compared with other filters below.

*Fisher’s Method*

In Robinson (2003), the posterior probability needed for the R.A Fisher’s method is different from the Naïve Bayes. It has two more benefits than Naïve Bayes: (1) it can deal with the rare word problem we mentioned in 3.2; (2) the probability combination needs no attribute independent assumption, and is more reasonable for the reality. As shown in Table 3, we calculate the posteriori probability through equation (3), (4), and (5). And the degree of belief,  $f(w)$ , is derived from the equation (6).

**Table 3:** *The posteriori probability  $p(w)$  and degree of belief  $f(w)$*

$w_i$	make	address	free	business	credit	money	...	order
$p(w_i)$	71.7%	76.7%	86.1%	79.8%	92.1%	95.8%	...	81.3%
$f(w_i)$	71.7%	76.6%	86.0%	79.8%	91.9%	95.6%	...	81.2%

The combination of probability for a specific feature vector of email is derived through equation (7) with the degree of belief,  $f(w)$ .

*Decision Tree*

Instead of using a complicated Bayesian calculation to extract a simple rule set from our data, we have introduced the decision tree induction method for the classification of spam and legitimate emails. Given the generalized and relevant data relations, the information gain for each candidate attribute can be computed using the algorithm in Figure 1. The candidate attribute that gives the maximum information gain as the decision attribute at this current level is selected and the current set of objects are partitioned accordingly. For each subset created by the partitioning, it is necessary to repeat each step to further classify data until either (a) all or a substantial proportion (no less than the classification threshold) of the objects are in one class, (b) no more attributes can be used for further classification, or (c) the percentage of objects in the subclass (with respect to the total number of training samples) is below the exception threshold. The decision tree for spam email detection has been developed. The knowledge represented in the decision tree can be extracted and represented in the form of IF-THEN rules. One rule is created for each path from the root to a leaf node. Figure 3 shows the sample of the rules.

**Figure 3:** *The IF-THEN rule of the decision tree output*

<p>IF “remove”=“YES”, “money”=“YES”, “hp”=“YES”, “project” =“NO” THEN LEGITIMATE</p> <p>IF “remove”=“NO”, “money”=“YES”, “hp”=“NO”, “business”=“YES” THEN SPAM</p> <p>IF “remove”=“NO”, money=“YES”, “hp”=“NO”, “business”=“NO”, “edu”=“YES” THEN LEGITIMATE</p> <p>IF “remove”=“NO”, money=“YES”, hp=“NO”, business=“YES”, edu=“NO”, george=“YES” THEN SPAM</p>
--



### Comparisons and Analysis

Some very promising results were returned from the three algorithms, as can be seen in Table 4. To evaluate our system we were interested in several quantities typically used in measuring the query result of information retrieval. These are: (1) True Positives (TP): the number of spam email classified as spam, (2) True Negatives (TN): the number of legitimate email classified as legitimate, (3) False Positives (FP): the number of legitimate emails falsely classified as spam and (4) False Negatives (FN): the number of spam emails falsely classified as legitimate.

The performance of the algorithms can be measured in terms of accuracy rate. The precision rate,  $TP / (TP + FP)$ , denotes the portion of spam in the filtered mail, while the recall rate,  $TP / (TP + FN)$ , answers the question what portions of spam can the algorithm filter. The accuracy rate,  $(TP + TN) / (TP + FP + FN + TN)$ , represents the overall correct decision of the filtering.

**Table 4:** Results of the experiments

	True Positives	True Negatives	False Positives	False Negatives	Precision	Recall	Accuracy
Naive Bayes	31%	56%	2%	11%	94%	74%	87%
Fisher's Method	36%	53%	5%	6%	89%	87%	90%
Decision Tree	38%	55%	3%	4%	92%	91%	93%

The Naïve Bayes has the highest precision rate, but the recall and accuracy rates are not as good as others and often suffer from the false negatives rate. The Fisher's method has better recall and accuracy rates than Naïve Bayes, though the precision rate is the lowest of the three. The decision tree method generally has better performance than the others.

### Proposed Novel Modified Naïve Bayesian Classifier

To introduce cost-sensitive evaluation to our method, we employ the weighted accuracy ( $WAcc$ ) which was proposed by Androutsopoulos (2000). We treat each legitimate email as if it is  $\lambda$  emails. When a legitimate email is wrongly blocked, we will count it as  $\lambda$  errors. When a legitimate email passes the filter, we will count it as  $\lambda$  successes. These lead to the definition of the  $WAcc$  and weighted error rate ( $WErr$ ). We assume  $\lambda = 9.99$ . The result of our experiment is presented in Table 5. As can be seen from Table 6, the proposed method has better performance than Naïve Bayesian Filter.

$$WAcc = \frac{\lambda \cdot Tn + Tp}{\lambda \cdot (Tn + Fp) + Tp + Fn} \quad (13)$$

**Table 5:** Results of the extra experiments

Experiment	Method	TP	FP	TN	FN	Detection	False Positive	Overall Accuracy	Precision
a	Naive Bayesian	348	55	642	79	81.5%	7.9%	88.08%	86.35%
	Modified Naive Bayesian	349	43	655	77	81.92%	6.16%	89.32%	89.03%
b	Naive Bayesian	364	53	678	77	82.54%	7.25%	88.91%	87.29%
	Modified Naive Bayesian	364	44	687	77	82.54%	6.02%	89.68%	89.22%

**Table 6:** Weighted results of the extra experiments

$\lambda = 9$	Experiment	Methodology	WAcc	WErr = 1 - WAcc
9	a	Naive Bayesian	87.06%	12.94%
		Modified Naive Bayesian	88.52%	11.48%
	b	Naive Bayesian	87.68%	12.32%
		Modified Naive Bayesian	88.7%	11.3%
99	a	Naive Bayesian	91.62%	8.38%
		Modified Naive Bayesian	93.33%	6.67%
	b	Naive Bayesian	92.26%	7.74%
		Modified Naive Bayesian	93.47%	7.53%

## Conclusion

The spam problem has generated enormous costs for companies and other users of the Internet, and it continues to worsen. In order to deal with the huge amount of spam people receive daily, powerful email filters with high reliability are needed. In this study, we examined various ways to stop spam. Three data mining methods in the detection of spam were described and examined. From the results, we have found that it is possible to train the filter automatically through data mining algorithms. In our experiment results, the decision tree method generally had better performance than the other two methods. The main contribution of this study was to provide a clear performance measure of three data mining methods and propose a novel modified Naive Bayesian classifier in advance. Although spam email detection and protection is to be viewed as necessary for all Internet users, there is no single method that can dominate this work. We also explored several methods for spam filtering, ranging from social to technical approaches. One of the most important areas of future work is the development of more efficient algorithms. The current data mining methods require a significant amount of memory and computing resource. We would like to make these learning algorithms more efficient overall.

## References

- Alperovitch, D., Judge, P. & Krasser, S. (2007). Taxonomy of Email Reputation systems. *27th International Conference on Distributed Computing Systems Workshops: 27*.
- Androutsopoulos, I., Koutsias, J., Chandrinou, K. V. & Spyropoulos, C. D. (2000). An Experimental Comparison of Naïve Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages. *In Proc. of the 23rd Annual International ACM SIGR Conference on Research and Development in Information Retrieval*. Athens, Greece: 160–167.
- Cheng, J. & Greiner, R. (1999). Comparing Bayesian Network Classifiers. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI-99)*. San Francisco, CA: 101-108.
- O'Brien, C. & Vogel, C. (2003). Spam filters: Bayes vs. chi-squared; letters vs. words. *In Proceedings of the International Symposium on Information and Communication Technologies*. Dublin, Ireland: 291-296.
- Drucker, H., Wu, D. & Vapnik, V. N. (1999). Support Vector Machines for Spam Categorization. *IEEE Transaction on Neural Networks*, 10(5): 1048-1054.
- Domingos P. & Pazzani M. (1996). Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier. *In Proc. of the 13th International Conference on Machine Learning*. Bari, Italy: 105-112.
- Han, J. & Kamber, M. (2001). Data mining concepts and techniques. USA: Morgan Kaufmann: 284–287.
- Joachims, T. (1997). A probabilistic analysis of the Ricchio algorithm with TFIDF for text categorization. *In Proc. of 14th International Conference on Machine Learning*.
- Langley, P., Wayne, I. & Thompson, K. (1992). An Analysis of Bayesian Classifiers. *In Proc. of the 10th National Conference on Artificial Intelligence*. San Jose, California: 223-228.
- Metzger, J., Schillo M. & Fischer, K. (2003). A multi-agent based peer-to-peer network in java for distributed spam filtering. *In Proc. of the CEEMAS*, Czech Republic: 616-625.
- Paulson, L. D. (2003). Group Considers Drastic Measures to Stop Spam. *IEEE Computer*, 36(7): 21- 22.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. *Morgan Kaufmann series in machine learning*.
- Robinson, G. (2003). A Statistical Approach to the Spam Problem. *Linux Journal*, 107: 3.
- Sahami, M., Dumais, S., Heckerman, D. & Horvitz, E. (1998). A Bayesian Approach to Filtering Junk Email. *In Learning for Text Categorization – Paper from the AAAI Workshop*, Madison Wisconsin. AAAI Technical Report WS-98-05: 55-62.
- Schapire, R.E. & Singer, Y. (2000). BoosTexter: A Boosting-based System for Text Categorization. *Machine Learning*, 39: 135–168.
- Spammers Grab MSN Hotmail addresses. 2007 [Online]. Available: <http://www.spamhaus.org/news.lasso?article=6>.
- UCI Machine Learning Repository. 2004 [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>.



# A Resource-Based Look at Compensation Strategy: Application and Implementation of Competitive Advantage

James A. Carey

*Colorado State University - Pueblo*

*While resource-based views of the firm implicitly assume heterogeneity with respect to resource endowments and the efficacy of firms to acquire additional resources, competing firms can still produce competitive advantage when such differences are negated by industrial policy. This study examines salary and performance data over a ten-year period and integrated theory and application of compensation policy on competitive advantage within the National Football League. Path analysis is used to examine a model of league salary cap components and reveals that strategic management of specific components of this compensation system does result in more team wins.*

## Introduction

Industrial relations policies and human resource practices are organizational activities directed at managing human capital and human resources toward the fulfillment of organizational goals (Wright, McMahan & McWilliams, 1994). In the context of the resource-based view of the firm, these activities are undertaken to achieve competitive advantage within an industry. Research on strategic human resource management (HRM) or resource-based strategy often makes broad categorizations of such practices without isolating compensation policy and implementation as competitive tools. For instance, Schuler (1992) generically discusses the necessity of human resource (HR) practices and policies to match business needs, but makes no

specific mention of the compensation system. On the other hand, Wright and McMahan (1992) stress the importance and research value of HR practices that reinforce role behaviors important to organizational success. Perhaps no other system is implemented more directly with the intent of reinforcing role behaviors consistent with organizational strategy than is a firm's compensation system. Strategic implementation of compensation management is not a new idea (Milkovich, 1988; Gomez-Mejia & Balkin, 1992a). However, the specific application of compensation management and a test of the resultant competitive advantage is the impetus of this study.

### Competitive Advantage

Schulze (1994) observes two schools of thought in the resource-based strategy literature – a structural school and a process school – and asserts several assumptions made by these groups in explaining competitive advantage. Generically, Schulze (1994) cites: 1) Conner (1991) – all resource-based perspectivist assume that differences in product/service attributes (and therefore performance) are related to the differences in resources possessed or controlled by the firm; 2) Barney (1991) – resources are heterogeneous across competing firms; and 3) Rumelt (1987) – firms are rent seekers. Such firms seek returns in excess of the “normal” profit seeking perspective, and these rents can be differentiated by their source. For instance, the difference between managerially produced rents versus those produced by labor.

Schulze (1994) posits two additional assumptions which differentiate each school from the other. The structural school is characterized by the assumption that competitive advantage is sustainable from resources that are rare, valuable, imperfectly mobile, and inimitable through substitution (Barney, 1991). The process school assumes that rents are available to the firm through managerial learning, development of new resources, and achieving a better match between the competitive environment and the capabilities of the firm resources (Schoemaker, 1990); aspects which directly implicate the managerial quality and discretion in the implementation of strategy.

An implicit assumption of Conner (1991) is that competitive firms have differing inherent levels of efficacy with respect to their abilities to acquire critical resources. Firm-idiosyncratic resources are a central concept to the resource-based view and create a fundamental paradox with respect to generalizability of theory (Gibbert, 2006). Finkelstein and Peteraf (2007) also note that some environments and organizations limit managerial discretion. Some organizations or industries offer varying amounts of discretion depending upon their characteristics. A question arises then, if through chance or external policy, the ability of firms to acquire scarce resources was effectively equivalent, or non-idiosyncratic, would firm-by-firm managerial and developmental effects still account for all differences in rent production, or can the system of acquisition of resources account for these differences? Such an industry exists where, as a matter of policy, both the number of resources available to each “firm” is equivalent, as is the total amount of compensation each firm can provide.

### *Parity in Resource Acquisition*

In 1993, the National Football League (NFL) entered into a collective bargaining agreement (CBA) with the National Football League Players Association (NFLPA) that changed the economic and social structure of their industry, as well as the ability of each team to acquire and retain its critical human resources. Two of the major changes provided by the CBA were the provision of free agency for most players with three or more accrued seasons, and an annual hard “cap” on annual aggregate team payroll. Free agency removed the long-standing protection of resource immobility that owners enjoyed, and which prohibited application of basic economic theory. This revolutionary move was prototypical of what Barney (1991) calls “structural revolutions in an industry,” which may redefine or erode competitive advantage of individual firms.

Perhaps nowhere is the concept of competitive advantage more apparent than in the multi-billion dollar industries of professional sports. Here, resource acquisition, effective strategy, and management spell the difference between competitive success and mediocrity, making these industries prototypes for studies of industrial policy and resource-based strategy. Data from professional and collegiate athletics have proven to be a useful and insightful venue for many academics (Abelson, 1985; Bloom, 1999; Hofmann, Jacobs & Gerras, 1992; Mazur, 1994; Wright, Smart & McMahan, 1995) for at least two reasons. First, the use of sports data has been demonstrated to be a valid measure of performance (Hofmann et al., 1992), or used more generically as productivity indicants or outcomes (Mazur, 1994). There is some disagreement over specifically which statistics provide the most appropriate information (Henry & Hulin, 1987) for the given research circumstances and objectives, and it seems apparent that this information is valid, useful, and readily available. Second, athletics provides a fertile ground for the study of many of the issues of scholarly interest today, such as team formation and development, the interdependence of individuals and groups, effective leadership, resource management, and the implementation of strategic goals through compensation. The purpose of this article is to examine the function and effect of the salary cap in the NFL, its application to the resource-based view of the firm, and to empirically examine its structural components to determine how teams can win consistently despite using tools that are both available to and constrained to all competitors.

### *Why a Salary Cap?*

The salary cap was arguably necessary from a league standpoint on a number of grounds. First, as evidenced in other major sports, particularly baseball, free agency begets ever-escalating player salaries (Staudohar, 1996). The salary cap meant that the subsequent increases in player salaries would reflect, rather than overextend, the revenues earned by the league, maintaining a viable economic structure. In order for free agency to be effective, particularly from the viewpoint of the players’ association, salary levels must relate to player value or marginal revenue product (MRP). From a management viewpoint, an association between salary and value is validated by performance. In general, better performing players should have higher salaries, otherwise labor economics are being violated.

Second, it further demonstrated efforts at achieving competitive parity between the teams in the league. It was reasoned in part that, with each team's player expenses capped at the same amount, and in accordance with generated league revenues, teams had theoretically equal opportunity and economic means to purchase competitive resources in the newly created free agent market. Third, since the salary cap was tied directly to revenues which were divided equally among teams, it mitigated the concerns of small market franchises that bigger market, or more cash rich teams, could consistently outbid and more easily compensate valuable player resources. Thus, it was supposed to provide a means of achieving competitive and economic parity by institutionalizing compensation policy within the industry.

### *Specifics of the Cap*

The salary cap is determined as a percentage of defined gross revenues (DGR), which is outlined in the CBA as the aggregate revenues from all sources relating to the performance of NFL games. This includes gate receipts and most significantly, television revenues. Excluded from calculation of DGR are revenues from concessions, parking, local advertising and promotion, programs, and those revenues from NFL Films and NFL Properties, the licensing branch of the NFL. The percentage of DGR that comprises the salary cap varies slightly over the term of the CBA. In its inception in 1994, the amount was 64%.

The actual team salary cap is determined on a *pro rata* basis by dividing the dollar amount arrived at as a percentage of DGR by the number of teams, less a portion of that figure (about \$5 million) that is allocated annually to collective benefits, such as the player pension fund. The remaining portion represents the amount each team has available to spend on player salaries for that contract year. Every one of the thirty-two NFL teams has exactly the same amount as the limit they may spend, or more correctly, allocate collective player compensation. Each team also has the same number of active players. Thus, a situation exists whereby the number of critical human resources and the total amount of compensation are the same for all firms in the industry.

At the intuitive level, the salary cap seems to be a straightforward addition problem: do not spend more than x dollars in a given year on player salaries. Teams have found that implementation according to the rules agreed upon in the CBA is a much more complex issue. Since its realization in 1994, successful teams have had the same salary cap figure each year as losing teams have had. They appear to have been much more willing to incur significant risk, both from an economic and absolute standpoint, to do what they believe was necessary to sustain an advantage and win, either immediately or consistently. They chose to allocate salary cap dollars to future contract years. Although the cumulative dollars contained in players' contracts cannot exceed the cap number (without substantial financial penalty), the calculation of dollars spent for any given year may not reflect the actual dollars paid to a team's players. Even though the salary cap creates a hypothetical ceiling on the amount of money that teams can spend in a given year, the ceiling is frequently exceeded from a cash flow standpoint, depending upon a team's orientation toward risk.

Allocation is the key. The primary method by which teams routinely push cap dollars into future contract years is through the signing bonus. Signing bonuses,



regardless of when they are paid to a player, are prorated over the life of the contract. Thus, only that portion prorated to any given contract year counts against the total team salary cap, rather than the entire amount paid to the player at the time of signing. For example, if a player signs a five-year contract for \$500,000 per year and receives a \$5 million signing bonus, only \$1,500,000 in salary counts against the cap each year, despite the fact that the player actually received \$5,500,000 in his first year, and \$500,000 each year thereafter.

### *Individual Compensation*

In the presence of a competitive labor market, a player's maximum value would theoretically be equal to his marginal revenue product (MRP). From the team's perspective, in the competitive labor market, the most it should be willing to pay a player would be equal to his MRP. This aspect of compensation theory has applicability in many areas, not just professional sports. For instance, Gomez-Mejia, Tosi, and Hinkin (1987), in a study of executive compensation, provided the same explanation for determining the "upper limit" a firm would be willing to pay an executive. In this case, MRP was defined as the profit realized by a firm in excess of the amount estimated under the alternative or next best executive. The lower limit, or least amount an executive should accept, would be the next best offer or the amount he/she could obtain in their second-best employment situation. They also note that in a competitive market for executive talent, the MRP and second-best amounts converge and executive compensation will equal their MRP.

From the player perspective, the least he should be willing to accept in salary should be his next-best employment value, or reservation wage (Quirk & Fort, 1992). The reservation wage for some players may very well approach the minimum salary set forth in a collective bargaining agreement. The more competitive the free agent market, the more convergent the reservation wage and the MRP should be. The more substitutes that exist for a player's talent and the more players of equitable value available in the market, the higher the bargaining power of the team can be. In such situations, the consequent salary will be closer to the player's reservation wage. The greater the perceived uniqueness of ability or drawing power of the player (rarity and inimitability), the greater his individual bargaining position, resulting in a salary closer to his MRP. However, the dynamics of the salary cap can manipulate this situation. Without a cap, in a truly free market, the reservation wage and the MRP would be expected to converge. The actual effect is incongruent from what we expect in a free market exchange because the salary cap, as an absolute limit to (aggregate) market spending, can reduce the bargaining power of a player with respect to that of a team at or near that year's limit. This may serve to create and maintain a significant differential between his reservation wage with a team that has more room under the cap limit – a wage which may in fact be higher and closer to the player's true market value.

### *Monopsony Power*

In resource-based terms, the new CBA created resource mobility. Until the advent of the agreement, a free market in the NFL had not existed. The monopsonistic position of each team with regard to the negotiation rights to a player effectively represented

concrete immobility. Such an economic position leads to a situation where player salaries have no reason to equal or perhaps even to approach their MRP (Hamilton, 1995). Yet, even without free agency, there had been variance in the individually negotiated player salaries. NFL owners have long maintained that players were paid for their performance, while the NFLPA had long asserted that this was not the case.

Ahlburg and Dworkin (1991), in a study of 1982 (pre-free agency) NFL player salaries, found that player's salary was determined by each player's draft round, position, and years in the league, all nonperformance related factors, as was then contended by the NFLPA. Also significant in the analysis was a performance measure. There were differences in salary correlated with individual performance by position, but this relationship was not significant for every position and was not very strong overall. Ahlburg and Dworkin's (1991) study provides evidence that player salary determination had been made primarily on something other than performance factors, a significant boost to the labor position in its own right. Football player salaries had been reflective of a skill-based compensation system, rather than a performance or merit-based system. Many of the skill-based system criteria outlined by Gomez-Mejia and Balkin (1992b) are reflective of the NFL. Some of this criteria includes frequent changes in technology and organizational structure, frequent employee exchanges, new learning opportunities, high turnover, and worker values consistent with teamwork and participation – the latter being the very embodiment of team sports.

With the CBA, there is some evidence that the new labor market brought performance into the equation. In a pilot sample ( $n=110$ ) of the highest paid players (by position) in the first year of the salary cap, the 1994 playing season, Carey (1994) found performance to be highly significant ( $p<.001$ ) in a regression model of player compensation, as well as position and the interaction of performance and draft round. In contrast to the Ahlburg and Dworkin (1991) study, neither seniority, nor draft round (as a main effect) was significant. While the sample for this study is somewhat biased, examining the highest paid players at the time, performance was a clear indicator of compensation for these individuals at the margin in the first capped season.

As a fundamental managerial principal in HR strategy, Boxall (2003) notes that, apart from the prospect for HR advantage, firms have no incentive to pay above market clearing wages. That performance is a factor in compensation and is a vital aspect in compensation strategy and implementation under the salary cap. If performance does not matter, owners have no incentive to pay above league minimums, and their competitive advantage or their ability to win games is a random effect. Essentially, player productivity is homogeneous. This would not seem to be true, because with or without a salary cap, some teams are consistently able to win games, yet others consistently lose. Variance in team wins would be much greater if they were the result of random effects. The salary cap creates a zero-sum tournament in which cost/benefit utility becomes the overriding issue. Teams at or near the cap level must pay less to one or more players in order to meet the market salary level for just one premium player. Thus, performance is a critical distinction. Since the salary cap produces an exercise in cost/benefit utility for team owners, one in which they know there is an absolute limit in their ability to purchase resources, a difference in rent production (defined as wins beyond the mean) should be illustrated consistently by structural

differences in the aggregation and distribution of salary cap components at the team level. If it can be shown that a relationship exists between some (or all) salary cap components and team performance, then the presumption of resource heterogeneity is also affirmed. Or, differences in team salaries reflect differences in player performance, and the economic theory holds.

Sirmon, Hitt and Ireland (2007) assert that the role of managers in a dynamic environment is to structure resources and bundle capabilities, creating a causal relationship between resource management and value creation. A demonstration that specific managerial activities result in competitive advantage is necessary to support theoretical aspects of the implementation of the resource-based view and also HR strategy. The questions of interest regarding management of the salary cap in the NFL are: a) whether teams can ultimately win more games through effective management of salary cap components, and b) whether differences in player salaries reflect differences in performance.

Hypothesis 1: Salary cap components, as defined by the NFL, in either their distribution or levels or both, are indicators of team success (wins).

- H1a. Team Total Salary is an indicator of team wins.
- H1b. Team Total Base Salary is an indicator of team wins.
- H1c. Team Total Cap Value is an indicator of team wins.
- H1d. Team Total Signing Bonus is an indicator of team wins.
- H1e. Team Mean Salary is an indicator of team wins.
- H1f. Team Mean Base Salary is an indicator of team wins.
- H1g. Team Mean Cap Value is an indicator of team wins.
- H1h. Team Mean Signing Bonus is an indicator of team wins.
- H1i. Team Salary Cap Percentage is an indicator of team wins.

Hypothesis 2: Player salaries have a positive relationship with performance.

**Table 1:** Description of Variables

Team Salary	Cap Components designated by the NFL.
Wins	Team regular season wins each year.
Salary Cap	Specified total amount of salary for each team each year.
Total Salary	Total of all player compensation paid by a team in a given year. (Can and does exceed the salary cap for the year.)
Total Base Salary	Aggregate contract year-specific base salaries, exclusive of all bonuses and incentives.
Total Cap Value	Aggregate amount of base salary, signing and other bonuses and/or incentives accruing to the salary cap each year.
Total Signing Bonus	Aggregate year-specific amount of signing bonuses paid each year.
Mean Salary	Mean amount of total salary for team each year. (Total salary / number of players under contract.)
Mean Base Salary	Mean amount of base salary, exclusive of all bonuses and incentives each year.
Mean Cap Value	Mean amount of base salary, signing and other bonuses and/or incentives accruing to the salary cap each year.
Mean Signing Bonus	Mean amount of signing bonuses paid each year.
Additional Variable Used in Analysis.	
Cap Percentage	Percentage of accrued cap value. (Cap Value/Cap)

### Method

Salary data and components were collected for thirty NFL teams for ten seasons (1995-2004; n=300). The components of team salary that are defined, differentiated, and reported by the league and player’s association were the variables used as the basis for this study. They include: Cap Value (total amount of player salary accruing to that year’s salary cap), Total Salary (total amount of compensation paid), Total Base Salary (base pay rate, not including bonuses and incentives, for that season), Total Sign Bonus (amount of signing bonus accruing to that year’s cap), and yearly means (the mean values for each of the four components above). An additional measure, Cap Percentage, was formulated. Cap Percentage is an indicator of the percentage of total cap dollars a team uses each year, as opposed to the absolute amount noted above. Salary data from two teams, the expansion Cleveland Browns and Houston Texans were omitted because they did not operate over this entire ten-year period. Data were combined for the pre expansion Cleveland Browns (a team that later became the Baltimore Ravens) who operated uninterrupted over the time period.

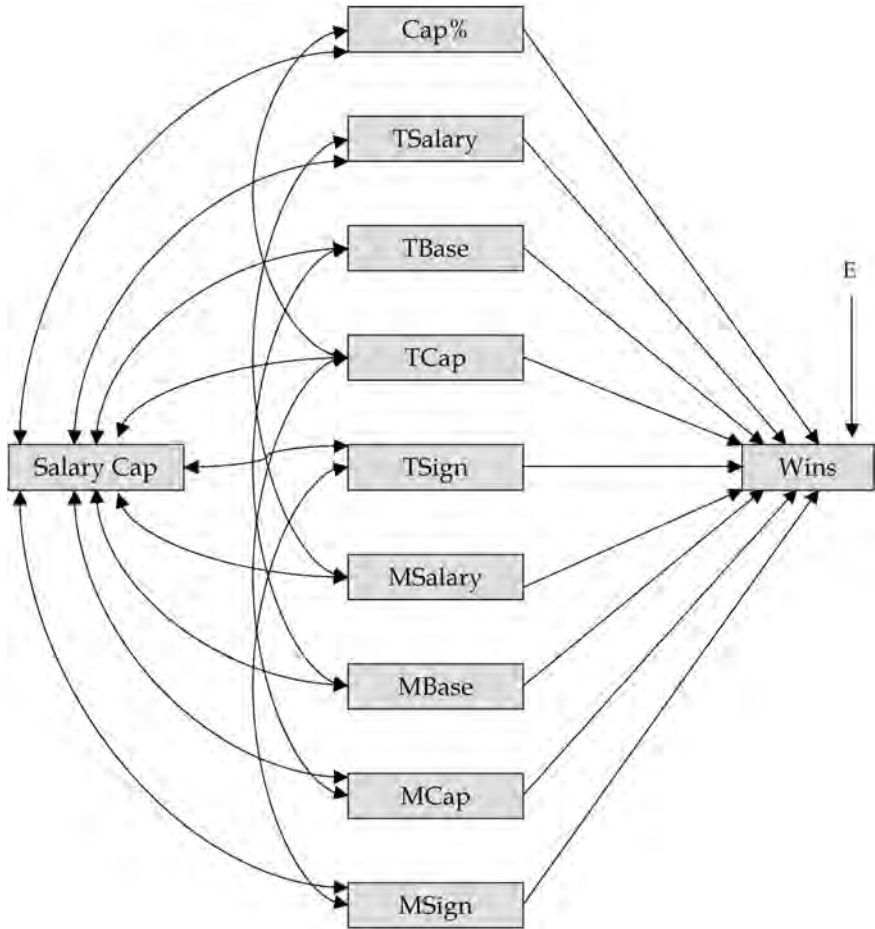
Path analysis was chosen as the appropriate method of testing a theoretical model of the effect of each variable on team win production and to produce tests for incremental model reduction (Hatcher, 1994). The original causal model was constructed according to convention as shown in Figure 1. Each salary component variable is illustrated using an arrow indicating the theoretical effect it exerts on team-win production. Curved, double-pointed arrows illustrate covariance between variables, including and especially between the year’s actual salary cap and each of the individual components. The model is over-identified, allowing for goodness-of-fit testing (Hatcher, 1994). Because some—but not all—of the salary components may be significant, an additional objective of the study is to determine the best fitting causal model through model modification tests procedures.

**Table 2: Correlation Matrix**

	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11
Cap	58633	14555	-										
Tsalary	57451	16080	0.84***	-									
Tcap	50674	13011	0.89***	0.84***	-								
Thase	31954	8205	0.69***	0.66***	0.85***	-							
Tsign	34134	29565	0.66***	0.78***	0.64***	0.52***	-						
Msalary	909.4	237.1	0.72***	0.91***	0.67***	0.44***	0.60***	-					
Mcap	802.8	173.8	0.82***	0.77***	0.87***	0.63***	0.46***	0.82***	-				
Mbase	502.3	102.0	0.63***	0.61***	0.73***	0.85***	0.37***	0.59***	0.76**	-			
Msign	746.7	454.7	0.13**	0.41***	0.16***	0.15***	0.70***	0.30***	0.01	0.03	-		
Cap%	0.869	0.089	-0.33***	-0.08	0.12*	0.21***	-0.09	-0.18***	-0.003	0.13**	0.07	-	
Wins	8.19	3.09	-0.02	0.05	0.04	-0.02	0.03	0.11*	0.11*	0.04	0.03	0.18***	-

\*p<.10 \*\*p<.05 \*\*\*p<.01

**Figure 1:** Full Theoretical Model of Salary Cap Components on Team Win Production



## Results

The theoretical model was examined using the LINEQS input in the CALIS procedure of the SAS® system to identify variables that have a direct effect on the team wins. The initial examination of the CALIS output indicates the model fits the data, although model reduction is also indicated (Hatcher, 1994). The chi-square for the model is 1.21 with a p-value of .27, which is outside the rejection range, but not strongly indicative of model fit. Bentler and Bonett's (1980) normed-fit index (NFI) also indicates a fit at .9980. Despite these indications, only the path estimate for Cap Percentage indicates a significant t-test. Stepwise model reduction is indicated by Wald tests, with Mean salary ( $p=.9698$ ), Mean cap ( $p=.4514$ ), and Mean sign bonus ( $p=.2180$ ) suggesting removal.

**Table 3: Tests of Full and Reduced Path Models**

	Full Model		Final Model	
	Path Coefficient <sup>a</sup>	t-value <sup>b</sup>	Path Coefficient <sup>a</sup>	t-value <sup>b</sup>
Cap%	0.282	3.91	0.265	4.19
Tsalary	0.142	0.17		
Tbase	-1.412	-1.77	-0.531	-3.70
Tcap	0.880	-1.08		
Tsign	0.163	1.07	0.196	2.60
Msalary	-0.029	-0.04		
Mbase	0.980	1.59	0.404	3.22
Mcap	-0.668	-0.99		
Msign	-0.080	-0.75		

<sup>a</sup> Standardized coefficients

<sup>b</sup> T-test significant where absolute value > 1.96

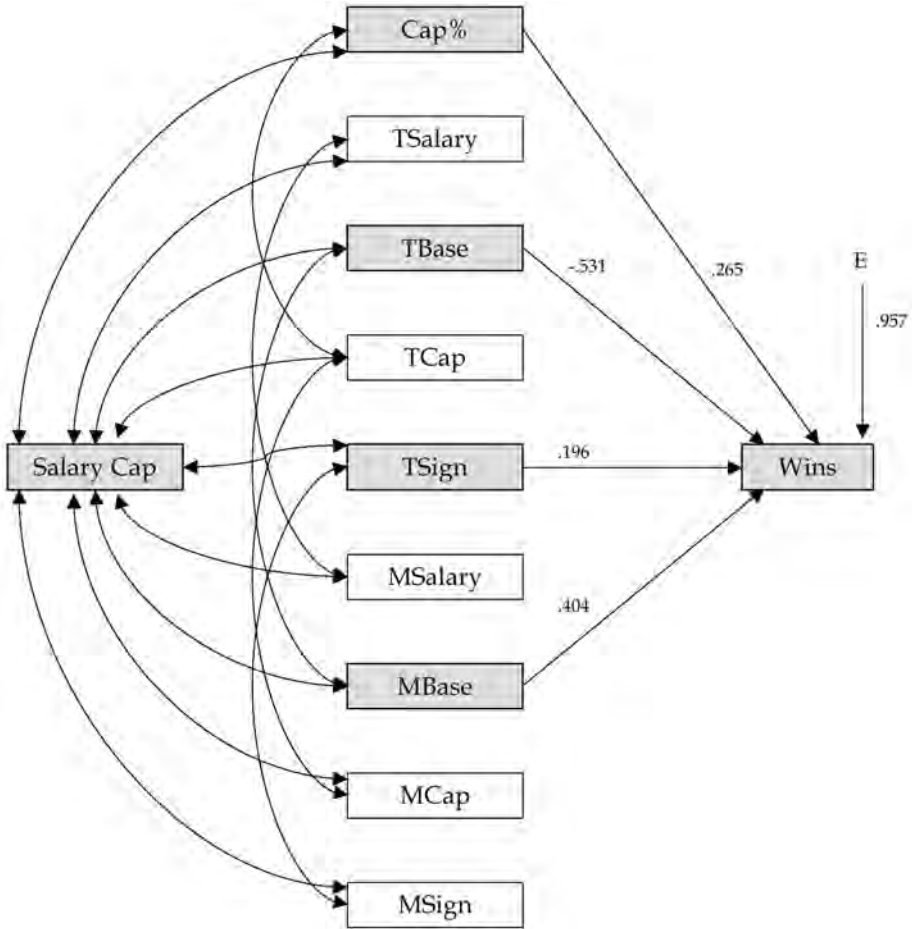
An incremental reduced model was tested eliminating Mean salary, Mean cap, and Mean sign bonus from the model and produced stronger results. The chi-square for the reduced model was 3.28 ( $p=.5114$ ) and the NFI is .9994, indicating a much better fit. Several path coefficients again are marginal or non-significant, however. Total salary ( $t=0.5095$ ), Total cap ( $t=0.8638$ ), and Total sign ( $t=1.1528$ ) each remain suspect. Wald tests indicate that Total salary is the only variable to remove, with the probability of the test at  $p=.6104$ .

Incremental reduction provides a better model when Total salary is eliminated. Chi-square for this reduced model is 3.54 ( $p=.6168$ ) and the NFI of .9998 indicates fit, but t-tests of the remaining path coefficients raise additional issues. T-tests for Cap Percentage ( $t=4.19$ ), Total base ( $t=-3.99$ ), and Mean base ( $t=3.12$ ) are all significant. Total cap ( $t=1.50$ ) and Total sign bonus ( $t=1.81$ ) are marginal. LaGrange Indices show no other variables should be included in the model, but Wald tests make removal of Total cap a possibility at  $p=.1345$ . Since its t-test was not significant, and the Wald test indicates removal, an addition modification was done.

Chi-square for a model of Cap percentage ( $t=4.19$ ), Total base ( $t=-3.70$ ), Total sign ( $t=2.60$ ), and Mean base ( $t=3.22$ ) effects on wins is 5.77 ( $p=.4490$ ). With NFI .9990, collectively this final model appears to fit the data well according to Hatcher's requisites. LaGrange Indices and Wald tests show no other variables should be included or eliminated from the model. Collectively, these tests provide primary support for Hypothesis 1 and indirect support for Hypothesis 2. Standardized path estimates for the final model are shown in Table 3. Specifically, hypotheses 1b, 1d, 1f, and 1i are supported. Hypotheses 1a, 1c, 1e, 1g, and 1h are not supported.

The findings also demonstrate that player compensation is related to performance on a broad scale, in support of Hypothesis 2. Because a model fits the data (i.e. the distribution of team wins is reflected in their salary and bonus distributions), it is an additional indication that salary distributions exhibit the performance abilities of those teams' players. Holding true to the MRP argument, teams cannot stockpile talented resources under a mask of economic equity, but must pay for them according to market forces, and allocate expenses according to operational or strategic needs.

Figure 2: Final Path Model



### Discussion

Results indicate that teams in the NFL that choose to remain as far under the salary cap as possible each year are doing their organizations and their fans a disservice, assuming that their goal is to win more games. The results of incremental path analysis using salary cap components as reported by the NFL indicate that a team’s ability to win consistently is possible despite resource equivalence. A general indication is that the closer your aggregate team compensation is to the salary cap, the more likely you are to win. Results of the model tested show that cap percentage, total base salary, total signing bonus, and mean base salary all contribute to win production in general support of Hypothesis 1.

As expected, the allocation of signing bonuses does affect win production and is shown to be a valuable tool in acquiring or stockpiling rare talent. Cap percentage

indicates that some teams win because they consistently use all of the financial resources available to them each year, rather than get by at the league minimums. It may also indicate that team owners or managers that do not push the limits of the cap may have a utility acceptable to them other than winning. Several possibilities exist, but profit may be a likely goal overriding wins. Total and mean base salaries are significant, yet the number of players accruing salary to the cap have little variance, due more commonly to injury or suspension. Because of this, these significant differences in wins likely come not from the base payments, but from player productivity, or heterogeneity of resources. Because the ultimate criterion of this study is wins, this is a stronger indication that players are in fact paid for differences in their performance than the simple fact that the data fits the model. Hypothesis 2 is thus supported and echoes basic labor economic principles. While some individual players may be overpaid or underpaid, there is a distribution of competencies among teams that is explicated by aggregate pay levels.

Essentially, resource heterogeneity is the norm for human resources when it comes to performance. Hunter and Schmidt (1989) showed that the standard deviation of performance is always at least 20% of the mean. Further examination by Hunter, Schmidt and Judiesch (1990) demonstrated this difference to be even larger within high complexity jobs (as much as 46%). It could be argued then, that part of the reason teams pay more for specialists is to reduce the possibility of variation in performance, particularly when there is a specific competitive intent in mind. Highly specialized performance that is interdependent, (i.e. in a team-based work environment) makes “skill and resource deployments” (Reed & DeFillippi, 1990, pg. 92) ambiguous to the competition, which raise barriers to imitation. Thus, teams help themselves by seeking and acquiring asset specificity (Williamson, 1985), and are willing to pay for those aspects. This would appear to be equally valid for managerial positions. This is very much in line with the resource-based view of the firm. Generalizing from the perspective of the structural school, this hints that managers might consider variable compensation systems that preserve the immobility of those higher producing human resources. The process school posits that rents can be and are produced by managerial resources. In this context, matching specific player personnel with a well-defined strategy, and this interpretation may account for much of the error variance in the model.

Some teams are somewhat superior in their ability to implement a strategic plan in confluence with the resources at their disposal. Thus, it is possible that the ability not only to select and compensate “better” players effectively, but also the ability to match these players’ abilities, or more importantly, to develop personnel who can do many things specifically toward the strategic intentions of the organization (or vice versa) are the mechanisms by which an organization produces rents. Some teams win more consistently because they are coached and/or managed better than others. This was the findings of Wright, Smart and McMahan (1995) using basketball data and also the finding of Kahn (1993) using professional baseball data. Better managers lead to more wins, and individual player performance improves as the managerial quality improves. Managerial quality includes the management and implementation of industrial relations and human resource practices such as compensation, as well as coaching



decision-making. While the literature examined sports industries, authors such as Castanias and Helfat (2001) effectively generalized the concept of managerial quality to apply in most industries.

### *Implications for Practice*

The results of this study demonstrate several principles useful for managers in broad contexts. First, competitive advantage can come from many parts of an organization, not just the planning and strategy formation processes. Implementation is key and this is particularly true when considering the activities and support functions that human resources perform. This study supports the notion that effective human resource policies and procedures such as selection and compensation can be critical to an organization's success.

Because this study illustrates that resource-based views of the firm also soundly apply to critical human resources, it lends support conceptually and economically to merit-based compensation systems over more traditional methods, such as seniority-based compensation, where these are appropriate. When firms benefit economically or competitively from idiosyncratic human and social capital, it may be in their best interests to share some of that benefit as a motivation and retention device.

Of even broader applicability is the demonstration that a structural change to an industry should not be viewed as merely changing the rules of competition. In such circumstances, a manager's tasks should focus on understanding the impact that structural change has on his or her organization and the relative effects such constraints have on their competitors. This study demonstrates that sustained advantage results from doing so. Because successful strategy is a dynamic process rather than a choice (D'Aveni, 1994), the effective manager understands that it is what he or she can do, not what new and complex constraints prevent doing, that produces a competitive advantage. While it is simple to suggest that managers merely reframe structural changes as new opportunities, competitive success depends upon proactively and positively responding to constraints when they occur.

### *Limitations*

As always, some limitations are evident. First, because of the construction and delineation of variables as they are defined by the NFL, there is multicollinearity between several of the variables. Second, while the purpose of this study was to examine those variables reported by the league, effects that are not included in these data may represent a significant clue to specifically understanding the management of salary in the league. For instance, average contract length would be useful in determining over how long a period the signing bonus amounts are spread. It is possible that contract lengths are gradually extending each calendar year as a method of accruing less to each cap year. If true, this might account for a non-trivial amount of error, and further illustrate a principle of effective management. This variable is not one of those reported by the league, and hence was not used in this study.

### *Recommendations for Future Research*

A starting point for continuing research in this area would be examination of

variables which may have additional explanatory value. As noted above, it might be useful to examine the absolute amount of initial signing bonus each year and the average lengths of the contracts. Not only do these amounts vary between teams and players, but such variation might also help explain differences in compensation strategy between successful and unsuccessful teams. Team-specific effects are also unaccounted for, and clearly have an impact on the criterion of wins. However, these effects can potentially be operationalized in a number of ways. For example, variables such as coaching turnover and even the construction or renovation of new facilities have been suggested, at least in popular media, as possible sources of team success.

Another key in further understanding the dynamics of competitive advantage would be to examine the effect of individual performance on team performance. In the context presented in this study, it would be valuable to determine how successful or unsuccessful teams distribute their salary cap among which of their players or positions. Individual performance could be both gauged against absolute standards and compared league-wide. A comprehensive understanding of the interaction of individual and team performance may provide a theoretical foundation that could have broad applicability, given the proliferation of team-based organizations and decision-making.

Although this study looked specifically at compensation policy, an additional area of research may be the effects of managers and coaches on team rents. This is particularly true given that other studies (e.g. Kahn, 1993) have found positive connections between management ability and firm success. Determining these and other such factors that contribute to firm- or team-specific effects is an important area for future research. Fundamentally, it would be useful to consistently demonstrate that effective management practices lead to competitive advantage, rather than convolute the association by assuming that a firm's success is a direct result of a manager's charisma, without first understanding what he or she actually did.

## Conclusion

It should be noted that it was not the intent of this study to pit Schulze's (1994) structural and process perspectives on resource-based strategy in an attempt to validate one or the other. These perspectives were used as a theoretical basis for exploring a wholly unique industry and circumstance where the ability of competitive firms to acquire critical resources is non-disparate. This study demonstrated that even under conditions of parity, competitive advantage can be attained by firms that effectively manage the tools available to them. The generalizability of operational human resources and managerial abilities as critical aspects of a firm's success, and the ability of managers to the compensation system to produce performance, should not be undervalued. Authors such as Koch and McGrath (1996) have asserted that human resource policies affect firm productivity. Thus, effective management of compensation and human resource practices may result in significant productivity gains for the

organization and improvements to effectiveness and efficiency objectives. One must accept that HR policy such as effective compensation does not exist in a vacuum. It is inextricably linked to other organizational processes and outcomes, such as effective selection and appraisal of performance. The end objective is competitive advantage.

In addition to selecting high performing human resources, firms would do well to take care to implement strategic planning by linking business needs and HR practices. One of the tools available to achieve competitive advantage is through effective management of compensation policy. This is in concurrence with Barney's (1991, 1995) theoretical perspectives on resource-based strategy.

## References

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97: 129-133.
- Ahlburg, D. A. & Dworkin, J. B. (1991). Player compensation in the National Football League. In Staudohar, P. D. & Mangan, J. A. (Eds.), *The business of professional sports*, 61-70. Champaign, IL: University of Illinois.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17: 99-120.
- Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Executive*, 9: 49-61.
- Bellante, D. & Jackson, M. (1983). *Labor economics: Choice in labor markets*. New York: McGraw-Hill.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88: 588-606.
- Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, 42: 25-40.
- Boxall, P. (2003). HR strategy and competitive advantage in the service sector. *Human Resource Management Journal*, 13: 5-20.
- Carey, J. A. (1994). Determinants of player compensation in the National Football League in the age of free agency and the salary cap. Unpublished working paper, Arizona State University.
- Castanias, R. P. & Helfat, C. E. (2001). The managerial rents model: Theory and empirical analysis. *Journal of Management*, 27: 661-78.
- Conner, K. R. (1991). A historical comparison of resource-based theory and five schools of thought within industrial organizational economics: Do we have a new theory of the firm? *Journal of Management*, 17: 155-171.
- D'Aveni, R. A. (1994). *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press.
- Finkelstein, S. & Peteraf, M. A. (2007). Managerial activities: A missing link in managerial discretion theory. *Strategic Organization*, 5(3): 237-248.
- Gibbert, M. (2006). Generalizing about uniqueness: An essay on an apparent paradox in the resource-based view, *Journal of Management Inquiry*, 15: 124-134.
- Gomez-Mejia, L. R. & Balkin, D. B. (1992a). *Compensation, organizational strategy, and firm performance*. Cincinnati, OH: South-Western.

- Gomez-Mejia, L. R. & Balkin, D. B. (1992b). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, 35: 921-955.
- Gomez-Mejia, L. R., Tosi, H. & Hinkin, T. (1987). Managerial control, performance, and executive compensation. *Academy of Management Journal*, 30: 51-70.
- Hamilton, J. C. (1995). Salary determination in professional sports. Doctoral dissertation, University of California at Berkeley, 1995.
- Hatcher, L. (1994). *A step-by-step approach to using the SAS® system for factor analysis and structural equation modeling*. Cary, NC: The SAS Institute.
- Henry, R. A. & Hulin, C. L. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology*, 72: 457-462.
- Hofmann, D. A., Jacobs, R. & Gerras, S. J. (1992). Mapping individual performance over time. *Journal of Applied Psychology*, 77: 185-195.
- Hunter, J. E. & Schmidt, F. L. (1989). Meta-analysis: Facts and theories. In Robertson, I. T. & Smith, M. (Eds.), *Advances in selection and assessment*. Oxford: John Wiley & Sons.
- Hunter, J. E., Schmidt, F. L. & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75: 28-42.
- Kahn, L. M. (1993). Managerial quality, team success, and individual player performance in Major League Baseball. *Industrial and Labor Relations Review*, 46: 531-547.
- Koch, M. J. & McGrath, R. G. (1996). Improving labor productivity: Human resource management policies do matter. *Strategic Management Journal*, 17: 335-354.
- Mazur, M. J. (1994). Evaluating the relative efficiency of baseball players. In Charnes, A., Cooper, W.W., Lewin, A.Y. & Seiford, L.M. (Eds.), *Data envelopment analysis: Theory, methodology and applications*. Boston: Kluwer.
- Milkovich, G. T. (1988). A strategic perspective on compensation management. *Research in Personnel and Human Resource Management*, 6: 263-288.
- National Football League & National Football League Players Association. (1993). *NFL collective bargaining agreement 1993-2000*. (Appended.) New York.
- Quirk, J. & Fort, R. D. (1992). *Pay dirt: The business of professional sports teams*. Princeton, NJ: Princeton University Press.
- Reed, R. & DeFillippi, R. J. (1990). Causal ambiguity, barriers to imitation, and sustainable competitive advantage. *Academy of Management Review*, 15: 88-102.
- Rumelt, R. P. (1987). Theory, strategy, and entrepreneurship. In Teece, D. (Ed.), *The competitive challenge*. New York: Harper Row.
- Schuler, R. S. (1992). Strategic human resource management: Linking the people with the strategic needs of the business. *Organizational Dynamics* (Summer): 18-31.
- Schulze, W. S. (1994). The two schools of thought in resource-based theory: Definitions and implications for research. In Shrivastava, P., Huff, A. S., & Dutton, J. E., (Eds.) *Advances in strategic management*. Greenwich, CT: JAI Press.
- Shoemaker, P. J. H. (1990). Strategy, complexity, and economic rent. *Management Science*, 4: 1178-1192.
- Sirmon, D. G., Hitt, M. A., Ireland, R. D. (2007). Managing firm resources in dynamic environments to create value: Looking inside the black box, *Academy of Management Review*, 32: 273-292.

- Staudohar, P. D. (1996). *Playing for dollars*. Ithaca, NY: Cornell University Press.
- Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: Free Press.
- Wright, P. M., McMahan, G. C. (1992). Theoretical perspectives for human resource management. *Journal of Management*, 18: 295-320.
- Wright, P. M., McMahan, G. C., & McWilliams, A. (1994). Human resources and sustained competitive advantage. *International Journal of Human Resource Management*, 5: 301-326.
- Wright, P. M., Smart, D. L., & McMahan, G. C. (1995). Matches between human resources and strategy among NCAA basketball teams. *Academy of Management Journal*, 38: 1052-1074.



# On the Predictability of Stock Market Returns: Evidence from Industry-Rotation Strategies<sup>1</sup>

Robert R. Grauer  
Simon Fraser University

*This paper evaluates historic, Bayes-Stein, Capital Asset Pricing Model (CAPM) and dividend-yield riskfree-rate estimators of asset means using statistical and economic criteria. None of the estimators exhibit much in the way of out-of-sample predictive ability when judged by statistical criteria. Yet, when combined with a discrete-time power-utility portfolio selection model, all the estimators generate economically significant returns judged in terms of compound return – standard deviation plots and accumulated wealth. Even so, the portfolios generated from dividend-yield riskfree-rate estimators perform by far the best and portfolios generated from traditional CAPM estimator perform the worst. For the most part, commonly accepted statistical measures of investment performance support these rankings.*

How do we judge whether returns are predictable? We could regress returns on past returns or on information variables that might include accounting data, dividend yields, riskfree interest rates and other macroeconomic indicators. Within this framework we could judge predictability in terms of in-sample slope coefficients and R-square values. Of course, out-of-sample measures of statistical significance—R-

---

<sup>1</sup>Earlier versions of the paper were presented at the VII International Conference on Stochastic Programming at the University of British Columbia in Vancouver, the Northern Finance Association meetings in Calgary, the Pacific Northwest Finance Conference in Seattle, the University of Victoria, the Western Decision Sciences Meetings in Vancouver, and the Decision Sciences Meetings in San Francisco. The author thanks the Social Sciences Research Council of Canada for financial support; seminar participants, Nils Hakansson, Kris Jacobs, Peter Klein, Anthony Lynch, Ieuan Morgan, Mark Rubinstein, William Ziemba and an anonymous reviewer for helpful comments; and Reo Audette, Steven Bovencamp, Poh Chung Fong, Christopher Fong, John Janmaat and William Ting for their most capable assistance.

squares or mean square errors—would lend more credence to any claims of predictability. However, Leitch and Tanner (1991) point out that statistical measures of predictability may not shed much light on the economic value of a forecast. They show that commercial interest rate forecasts do not perform any better than naïve forecasts when evaluated in terms of out-of-sample statistical criteria. But, the commercial (naïve) forecasts generate economically profitable (unprofitable) trading strategies. This paper evaluates historic, Bayes-Stein, CAPM and dividend-yield riskfree-rate estimators of asset means using statistical and economic criteria. None of the estimators exhibit much in the way of out-of-sample predictive ability when judged by statistical criteria. Yet, when combined with a discrete-time power-utility portfolio selection model, all the estimators generate economically significant returns. The results are in agreement with Leitch and Tanner's: evaluating the estimators in terms of out-of-sample statistical criteria sheds little light on their economic value.

Grauer and Hakansson (1986, 1987) and Grauer, Hakansson and Shen (1990), apply a discrete-time power-utility model in conjunction with the empirical probability assessment approach (EPAA) in domestic, global and industry-rotation asset-allocation settings. The results are noteworthy for two reasons. First, the model often generates economically and statistically significant abnormal returns. Second, no attempt is made to correct for estimation error, which is clearly present in the EPAA. It would seem prudent, therefore, to examine the effects of making corrections for estimation error.

One approach adjusts for estimation error by estimating the means based on statistical, financial or forecasting models. The results of applying Stein and CAPM instead of historical estimators of the means are mixed. Evidence based on mean-variance (MV) portfolio selection, simulation analysis and out-of-sample portfolio performance suggests that Stein and CAPM estimators of the means can improve investment performance substantially (Jobson, Korkie & Ratti, 1979; Jobson & Korkie, 1980, 1981; Jorion, 1985, 1986, 1991). On the other hand, Grauer and Hakansson (1995) find that although the Stein estimators outperform the sample (historic) estimator in an industry-rotation setting, the gains are not as great as those reported by others. Moreover, in a global setting just the opposite is true: the sample estimator outperforms the Stein estimators. In all cases, the CAPM estimator exhibits the worst performance, which is just the opposite of what Jorion (1991) finds in an industry setting using MV analysis that allows short sales. In light of these contradictory results, this paper examines the effects of adding two dividend-yield riskfree-rate estimators to the mix.

A second approach adjusts for estimation risk by constraining portfolio weights. The results are again somewhat mixed. Employing simulation and MV analysis, Frost and Savarino (1988) report that imposing upper bounds both reduces estimation bias and improves performance. In a companion paper to this one, Grauer (2007) shows that the portfolios of less risk-averse MV investors generated from dividend-yield riskfree-rate estimators of the means bankrupt in an out-of-sample industry-rotation setting with short sales permitted. Yet, with short sales precluded, the dividend-yield riskfree-rate portfolios of these less risk-averse MV investors exhibit the best performance. Moreover, portfolios generated from the CAPM estimator, which display



the best performance with short sales permitted, exhibit the worst performance with short sales precluded. On the other hand, Grauer and Shen (2000), while employing the discrete-time power-utility model in an out-of-sample setting with short sales precluded, reported that constraining the portfolio weights further led to appreciably more diversification and less realized risk. But the cost is a less realized return.

Mean estimators trace their origins to different parts of the statistics and finance literature. The mean square error properties of the historic estimator make it an obvious choice for an estimate of a mean. Stein estimators are based on purely statistical arguments that minimize the mean square error of a vector of means and completely ignore risk-return tradeoffs that may be helpful in predicting stock returns. CAPM estimators fill this void by drawing on the best-known financial model of asset pricing. Dividend-yield riskfree-rate estimators, as well as estimators based on other information variables, trace their origin to the return predictability or weak form efficient markets literature (Fama, 1991). Return predictability is of interest not only because of its fundamental implications for market efficiency, but also because it is steeped in controversy. There is disagreement about whether returns are predictable and, if they are, whether predictability implies market inefficiency or is a result of rational variation in expected returns.

In the early to mid-1990s there was a consensus, based on statistical criteria, that stock market returns could be predicted from informational variables—at least over one-year to four-year decision horizons. (Fama & French, 1988, 1989; Fama, 1991; Hawawini & Keim, 1995). The importance of this evidence extended beyond statistical considerations, as it helped renew the interest in continuous-time portfolio choice, hedging demand and non-myopic investment decisions discussed below. Fama and French (1988) show that the power of dividend yields to forecast stock returns increases with the return horizon. The monthly and quarterly results are unimpressive, with R-squares on the order of 0.01. But, with four-year returns, the R-squares range from 0.13 to 0.64. More impressive, out-of-sample R-squares from forecasts made with coefficients estimated from 30-year rolling regressions are close to the in-sample R-squares for all return horizons.

The consensus began to crack through the 1990s and into the new century. Lo and MacKinlay (1990) and Foster, Smith and Whaley (1997) are concerned with data mining. Others (Hodrick, 1992; Goetzmann & Jorion, 1993; Goyal & Welch, 2003; Ang & Bekaert, 2003) question the long-horizon results on statistical grounds. Goetzmann and Jorion (1993) and Stambaugh (1999) study the biases due to dependent stochastic regressors. Ferson, Sarkissian and Simin (2003) question whether there is a spurious regression bias in predictive regressions. Bossaerts and Hillion (1999) examine the statistical significance of a variety of informational variables using monthly data in an international setting. They confirm the presence of in-sample predictability, but report that even the best prediction models have no out-of-sample forecasting power. Goyal and Welch (2003) confirm Bossaerts and Hillion's evidence, while Pesaran and Timmermann (1995) report contradictory results. Perhaps surprisingly in light of the early evidence on long-horizon predictability and the out-of-sample evidence in Bossaerts and Hillion, Goyal and Welch and this paper, Ang and Bekaert (2003) and Torous, Valkanov and Yan (2005) report that the

predictive power of dividend yields is best visible at short horizons—with the short rate as an additional regressor in Ang and Bekaert's case.

A number of authors (Breen, Glosten & Jagannathan, 1989; Fuller & Kling, 1990; Pesaran & Timmermann, 1994; Larsen & Wozniak, 1995; Pelaez, 1998; Schwert, 2003) investigate the economic value of trading rules based on predictive regressions, when the predictions are not combined with a MV or power-utility portfolio selection model. The results are mixed and sample-period dependent. Schwert, for example, examines a strategy of investing in short-term bonds when a dividend yield model predicts stock returns are lower than interest rates. The model predicts poorly during the 1990s. Schwert (2003: 953) concludes: "In short, the out-of-sample prediction performance of this model would have been disastrous." But, this paper shows that when dividend-yield riskfree-rate forecasts are combined with the discrete-time power-utility model, the results through the 1990s are anything but disastrous.

The single-period MV model, the discrete-time power-utility model and the continuous-time power-utility model either are or have been combined with forecasts based on information variables in order to determine the economic value of the forecasts. Contributors to the MV literature include: (Solnik, 1993; Klemkosky & Bharati, 1995; Connor, 1997; Beller, Kling & Levinson, 1998; Ferson & Seigel, 2001; Marquering & Verbeek, 2001; Fletcher & Hillier, 2002; Avramov, 2004; Avramov & Chordia, 2006). They examine the portfolio returns of MV investors who exhibit "average" degrees of risk aversion and revise their portfolios monthly. While these papers report economically significant returns in U.S. bond-stock, U.S. industries and international settings, none report results from before 1960. The benefits of the MV model include familiarity, ease of estimation—only the means, variances and covariances need to be estimated—and ease of computation.

This paper employs a discrete-time power-utility model that embodies a broad range of risk-aversion characteristics, quarterly decision horizons, borrowing and lending at different rates and an industry dataset that spans the 1934-99 period. The primary benefit of this model is the formal justification of a myopic decision rule. If returns are independent (but not necessarily stationary) from period to period, the use of a stationary myopic power-utility decision rule in each period is optimal. That is, the optimal policy only depends on next period's joint return distribution. (The single-period MV model simply assumes myopic policies are optimal.) The costs include increased complexity in estimation—the entire joint return distribution must be specified—and in computation.

Merton (1971, 1973) introduced a stochastically changing opportunity set that leads to hedging demand and non-myopic investment decisions. Recently, the (weak) evidence of in-sample return predictability based on information variables led to a resurgence of interest in the continuous-time model. A rich literature investigates hedging, the question of whether a long-horizon investor should allocate his wealth differently from a short-horizon investor, the effects of parameter and model uncertainty, the effects of transactions costs and the effects of conditioning on asset pricing models when returns are predictable, see (Kandel & Stambaugh, 1996; Kim & Omberg, 1996; Brennan, Schwartz & Lagnado, 1997; Balduzzi & Lynch, 1999; Brandt,

1999; Campbell & Viceira, 1999; Barberis, 2000; Pastor, 2000; Pastor & Stambaugh, 2000; Lynch, 2001; Lynch & Balduzzi, 2001; Avramov, 2002; Brennan & Xia, 2002). Much of the computational analysis calibrates the importance of hedging demand in simulated settings where the stochastic process is consistent with a regression of returns on informational variables. And, with the exception of Brennan, Schwartz and Lagnado (1997), there is little in the way of out-of-sample results. The approach provides a great deal of insight into the multiperiod investment problem, but does not come without the added costs of predicting returns beyond the current period and still further computational complexity. In many instances, specific distributional assumptions are needed to make the model tractable, which causes problems in computing expected utility. See, for example, Barberis (2000) and Kandel and Stambaugh (1996), who constrain investors from short selling and from buying on margin to insure that the expected utility problem has a feasible solution.

Clearly, the original in-sample evidence of return predictability generated from information variables calls into question the assumption that returns are intertemporally independent and explains the resurgence of interest in hedging and the continuous-time model. But, Bossaerts and Hillion (1999) and Goyal and Welch (2003) find no evidence of out-of-sample predictability based on information variables and in this paper, little, or no, evidence of out-of-sample predictability for *any* of the mean estimators is found. In light of this evidence and questions about in-sample predictability involving regressions of returns on information variables, this study employs inter-temporal independence and the myopic behavior of the discrete-time power-utility model as working assumptions in this paper.

The paper proceeds as follows. The next section outlines the basic multiperiod investment model and the method employed to make it operational. The data, the estimators of the means, and the statistical measures employed to evaluate the investment performance of the portfolios generated from five mean estimators are described in the following three sections. The results based on statistical criteria, economic criteria, and statistical measures of investment performance, and the robustness of the results are discussed in the second to last section. The final section contains a summary and conclusions.

### The Discrete-Time Power-Utility Model

The discrete-time model is the same as the one employed in Grauer and Hakansson (1986) and the reader is referred to that paper (specifically pages 288-291) for details. It is based on the pure reinvestment version of dynamic investment theory. In particular, if  $U_n(w_n)$  is the *induced* utility of wealth  $w$  with  $n$  periods to go (to the horizon) and  $r$  is the single-period return on the portfolio, the important convergence result:  $U_n(w_n) \rightarrow (1/\gamma)w^\gamma$  for some  $\gamma < 1$ , holds for a very broad class of terminal utility functions  $U_o(w_o)$  when returns are independent (but non-stationary) from period to period. (See Hakansson, 1974; Mossin, 1968; Hakansson, 1971; Leland, 1972; Ross, 1974; Huberman and Ross, 1983). Convergence implies that the use of the stationary *myopic* decision rule:  $\max E(1/\gamma)(1+r)^\gamma$ , for some  $\gamma < 1$ , in each period is optimal.

At the beginning of each period  $t$ , the investor chooses a vector of portfolio weights

$x_t$  on the basis of some member  $\gamma$  of the family of utility functions for returns  $r$  given by

$$\max_{x_t} E \pi \left[ \frac{1}{\gamma} (1 + m_t(x_t))^\gamma \right] = \max_{x_t} \sum_s \pi_{ts} \frac{1}{\gamma} (1 + (x_t))^{\gamma} \tag{1}$$

subject to

$$x_{it} \geq 0, \text{ all } i, x_{Lt} \geq 0, x_{Bt} \leq 0, \tag{2}$$

$$\sum_i x_{it} + x_{Lt} + x_{Bt} = 1, \tag{3}$$

$$\sum_i m_{it} x_{it} \leq 1, \tag{4}$$

$$1 + r_{ts}(x_t) > 0, \text{ for all } s, \tag{5}$$

where:

$r_{ts}(x_t) = \sum x_{it} r_{its} + x_{Lt} r_{Lt} + x_{Bt} r_{Bt}^d$  is the (*ex ante*) return on the portfolio in period  $t$  if state  $s$  occurs,

$\gamma \leq 1$  = a parameter that remains fixed over time,

$x_{it}$  = the amount invested in risky asset category  $i$  in period  $t$  as a fraction of own capital,

$x_{Lt}$  = the amount lent in period  $t$  as a fraction of own capital,

$x_{Bt}$  = the amount borrowed in period  $t$  as a fraction of own capital,

$x_t = (x_{1t}, \dots, x_{nt}, x_{Kt}, x_{Bt})'$ ,

$r_{it}$  = the anticipated total return (dividend yield plus capital gains or losses) on asset category  $i$  in period  $t$ ,

$r_{Lt}$  = the return on the riskfree asset in period  $t$ ,

$r_{Bt}^d$  = the interest rate on borrowing at the time of the decision at the beginning of period  $t$ ,

$m_{it}$  = the initial margin requirement for asset category  $i$  in period  $t$  expressed as a fraction,

$\pi_{ts}$  = the probability of state  $s$  at the end of period  $t$ , in which case the random return  $r_{it}$  will assume the value  $r_{its}$ .

Constraint (2) rules out short sales and ensures that lending (borrowing) is a positive (negative) fraction of capital. Constraint (3) is the budget constraint. Constraint (4) serves to limit borrowing (when desired) to the maximum permissible under the margin requirements that apply to the various asset categories. Constraint (5) rules out any *ex ante* probability of bankruptcy. The solvency constraint is not binding for the power functions, with  $\gamma < 0$  and discrete probability distributions with a finite number of outcomes, because the marginal utility of zero wealth is infinite.

Nonetheless, it is convenient to explicitly consider equation (5) so that the nonlinear programming algorithm used to solve the investment problem does not attempt to evaluate an infeasible policy as it searches for the optimum.

The inputs to the model are based on the “empirical probability assessment approach” (EPAA) with quarterly revisions. At the beginning of quarter  $t$ , the portfolio problem consisting of equations (1)-(5) for that quarter uses the following inputs: the (observable) riskfree return for quarter  $t$ , the (observable) call money rate +1% at the beginning of quarter  $t$  and the (observable) realized returns for the risky asset categories for the previous  $k$  quarters. Each joint realization in quarters  $t-k$  through  $t-1$  is given probability  $1/k$  of occurring in quarter  $t$ . Thus, under the EPAA, estimates are obtained on a moving basis and used in raw form without adjustment of any kind. On the other hand, since the whole joint distribution is specified and used, there is no information loss; all moments and correlations are implicitly taken into account. It may be noted that the empirical distribution of the past  $k$  periods is optimal if the investor has no information about the form and parameters of the true distribution, but believes that this distribution went into effect  $k$  periods ago.

With these inputs in place, the portfolio weights  $x_t$  for the various asset categories and the proportion of assets borrowed are calculated by solving equations (1)-(5) via nonlinear programming methods, (see Best (1975)). At the end of quarter  $t$ , the realized returns on the risky assets are observed, along with the realized borrowing rate  $r_{Bt}^r$  (which is calculated as a monthly average and may differ from the decision borrowing rate  $r_{Bt}^d$ ). Then, using the weights selected at the beginning of the quarter, the realized return on the portfolio chosen for quarter  $t$  is recorded. The cycle is repeated in all subsequent quarters. Note that if  $k = 32$  under quarterly revision, then the first quarter for which a portfolio can be selected is  $b+32$ , where  $b$  is the first quarter for which data is available.

All reported returns are gross of transaction costs and taxes and assume that the investor in question had no influence on prices. There are several reasons for this approach. First, as in previous studies, we wish to keep the complications to a minimum. Second, the return series used as inputs and for comparisons also exclude transaction costs (for reinvestment of interest and dividends) and taxes. Third, many investors are tax-exempt and various techniques are available for keeping transaction costs low. Finally, since the proper treatment of these items is nontrivial, they are better left to a later study.

## Data

The data used to estimate the probabilities of the next period's returns on risky assets and to calculate each period's realized returns on risky assets come from several sources. The returns for Standard and Poor's 500 Index come from the Ibbotson Associates database. The returns for the value-weighted industry groups are constructed from the returns on individual New York Stock Exchange firms contained in the Center for Research in Security Prices' (CRSP) monthly returns database. The firms are combined into twelve industry groups on the basis of the first two digits of the firms' SIC codes. (Grauer, Hakansson & Shen (1990) contains a detailed

description of the industry data.) The riskfree asset is assumed to be 90-day U.S. Treasury bills maturing at the end of the quarter. *The Survey of Current Business* and the *Wall Street Journal* are the sources. The borrowing rate is assumed to be the call money rate +1% for decision purposes (but not for rate of return calculations). The applicable beginning of period decision rate,  $r_{Bt}^d$ , is viewed as persisting throughout the period and thus as riskfree. For 1934-76, the call money rates are obtained from the *Survey of Current Business*. For later periods, the *Wall Street Journal* is the source. Finally, margin requirements for stocks are obtained from the *Federal Reserve Bulletin*. There is no practical way to take maintenance margins into account in our programs. In any case, it is evident from the results that they would come into play only for the more risk-tolerant strategies and for them only occasionally and that the net effect would be relatively neutral.

### Estimators of the Means

Under the historic approach means are not used directly but are implicitly computed from the realized returns in the estimation period. The  $n$ -vector of historic means at the beginning of period  $t$  is

$$\boldsymbol{\mu}_{Ht} = (\bar{r}_{1t}, \dots, \bar{r}_{nt})', \tag{6}$$

where  $\bar{r}_{it} = \frac{1}{k} \sum_{\tau=t-k}^{t-1} r_{i\tau}$ . This EPAA approach implicitly estimates the means one at

a time, relying exclusively on information contained in each of the time series.

Stein's (1955) suggestion that the efficiency of the estimate of the means could be improved by pooling the information across series leads to a number of so-called "shrinkage" estimators that shrink the historical means to some grand mean. A classic example is the James-Stein estimator (Efron & Morris 1973, 1975, 1977). It was first employed in the portfolio selection literature by Jobson, Korkie and Ratti (1979). However, this paper focuses on a Bayes-Stein (BS) estimator (Jorion, 1985, 1986, 1991)

$$\boldsymbol{\mu}_{BSit} = (1 - w_t) \boldsymbol{\mu}_{Ht} + w_t \bar{r}_{Gt} \mathbf{1}, \tag{7}$$

where  $w_t = \lambda_t / (\lambda_t + k)$  is the shrinking factor,  $\lambda_t = (n + 2) / ((\boldsymbol{\mu}_{Ht} - \bar{r}_{Gt} \mathbf{1})' S_t^{-1} (\boldsymbol{\mu}_{Ht} - \bar{r}_{Gt} \mathbf{1}))$ ,  $n$  is the number of risky assets,  $S_t$  is the sample covariance matrix calculated from the  $k$  periods in the estimation period,  $\bar{r}_{Gt} = \mathbf{1}' S_t^{-1} \boldsymbol{\mu}_{Ht} / (\mathbf{1}' S_t^{-1} \mathbf{1})$  is the grand mean and  $\mathbf{1}$  is a vector of ones. The  $\lambda$  does not contain an adjustment for degrees of freedom in estimating the covariance matrix as in Jorion (1985, 1986, 1991). We chose to model the problem this way to allow for the possibility of combining a Stein estimator with a set of non-equal probabilities for the states of nature used to estimate the joint distribution of security returns. In this case, the grand mean is the mean of the global minimum-variance portfolio generated from the historical data. Having calculated the Bayes-Stein and historic means for asset  $i$ , we add the difference  $(\bar{r}_{BSit} - \bar{r}_{it})$ , where  $\bar{r}_{BSit}$  and  $\bar{r}_{it}$  are the Bayes-Stein and historic means for asset  $i$  at time  $t$ , to each actual return

on asset  $i$  in the estimation period. That is, in each estimation period, we replace the raw return series with the adjusted return series  $r_{i\tau}^A = r_{i\tau} + (\bar{r}_{BSit} - \bar{r}_{i\tau})$ , for all  $i$  and  $\tau$ . No adjustment is made to the EPAA variance-covariance structure or to the other moments. Thus, the mean vector of the adjusted series is equal to the Bayes-Stein means of the original series; all other moments are unchanged. The same procedure is followed for the CAPM and dividend-yield riskfree-rate estimators discussed below.

A third estimator of the means is based on the CAPM (Sharpe, 1964; Lintner, 1965). The CAPM estimator is

$$\boldsymbol{\mu}_{CAPMt} = r_{Lt}\mathbf{1} + (\bar{r}_{mt} - \bar{r}_{Lt})\hat{\boldsymbol{\beta}}_t, \tag{8}$$

where  $\bar{r}_{mt} = \frac{1}{k} \sum_{\tau=t-k}^{t-1} r_{m\tau}$ ,  $\bar{r}_{Lt} = \frac{1}{k} \sum_{\tau=t-k}^{t-1} r_{L\tau}$ , and  $\bar{r}_{mt} - \bar{r}_{Lt}$  is an estimate of the expected

excess return on the “market” portfolio and  $\hat{\boldsymbol{\beta}}_t$  is the vector of estimated betas or systematic risk coefficients. At each time  $t$ ,  $\hat{\boldsymbol{\beta}}_t$  is estimated from the market model regressions

$$r_{i\tau} = \alpha_{it} + \beta_{it}r_{m\tau} + e_{i\tau}, \text{ for all } i \text{ and } \tau, \tag{9}$$

in the  $t-k$  to  $t-1$  estimation period, where the CRSP value-weighted index is employed as the proxy for the market portfolio. This method of estimating CAPM means, employed by Jorion (1991) and Grauer and Hakansson (1995), assumes the excess return on the market is constant over the estimation period. Alternatively, the ratio of the excess return on the market to the market’s standard deviation or variance might be assumed to be constant as in Merton (1980) and Best and Grauer (1985).

The next two estimators use dividend yields and riskfree interest rates to forecast the means. To construct the dividend-yield riskfree-rate estimators, the following regression is run at each time  $t$ :

$$r_{i\tau} = a_{0i} + a_{1i}dy_{\tau-1} + a_{2i}r_{L\tau} + e_{i\tau}, \text{ for all } i \text{ and } \tau, \tag{10}$$

in the  $t-k$  to  $t-1$  estimation period, where the  $i$  subscript denotes an industry,  $dy_{\tau-1}$  is the annual dividend yield on the CRSP value-weighted index lagged one month so that it is observable at the beginning of quarter  $t$ , and  $r_{L\tau}$  is the (observable) beginning-of-quarter Treasury bill rate. Both independent variables are “de-meanned.” Hence,  $a_{0i}$  is the historic average rate of return on asset (industry)  $i$ . The traditional one-period ahead forecast of the mean of industry  $i$  is

$$\bar{r}_{DRit} = \hat{a}_{0i} + a_{1i}dy_{t-1} + \hat{a}_{2i}r_{Lt}, \tag{11}$$

where  $\hat{a}_{0i}$ ,  $\hat{a}_{1i}$  and  $\hat{a}_{2i}$  are the estimated coefficients and  $dy_{t-1}$  and  $r_{Lt}$  are observable at the beginning of period  $t+1$ . That is, the quarterly variable  $dy_{t-1}$  is lagged one month and there is no need to lag  $r_{Lt}$  as it is observable at the beginning of the quarter. The vector of dividend-yield riskfree-rate (DR) estimators is

$$\mu_{DRt} = (\bar{r}_{DR1t}, \dots, \bar{r}_{DRnt})'. \quad (12)$$

However, this forecast is extremely variable. Therefore, we adopt the Bayesian framework advocated by Black and Litterman (1992) and shrink these mean forecasts to CAPM means. The Black-Litterman framework attempts to overcome three problems: the difficulty in estimating means; the extreme sensitivity of the solutions to slight perturbations in equilibrium means documented by Best and Grauer (1991, 1992) and Green and Hollifield (1992) and need for a set of means that would clear the market in an equilibrium setting. When the DR and CAPM means are assumed to be equally likely, the vector of dividend-yield riskfree-rate - CAPM (DRCAPM) mean estimators is

$$\mu_{DRCAPMt} = (\mu_{DRt} + \mu_{CAPMt})/2. \quad (13)$$

The CAPM means are estimated from equation (8). Black and Litterman (1992) estimate the means in a slightly different way. They estimate what they call equilibrium (and Best & Grauer (1985) call  $(\Sigma, \mathbf{x})$  - compatible) means from the equation  $\mu = r_L \mathbf{1} + \delta \Sigma \mathbf{x}$  where  $\Sigma$  is the covariance matrix of asset returns and  $\mathbf{x}$  is a vector of portfolio weights. When these means and  $\Sigma$  are inputs to a MV problem, subject only to a budget constraint,  $\mathbf{x}$  is the optimal solution.

### Statistical Measures of Investment Performance

In the results section we will see that the compound return – standard deviation plots and cumulative wealth values provide convincing evidence that: (1) all the mean estimators provide economic value when combined with the discrete-time power-utility model and (2) the two dividend-yield riskfree-rate estimators perform better than the traditional CAPM estimator. Although the figures and wealth values get to the heart of the matter, they do not give us a sense of how much of the difference can be attributed to randomness. In order to shed light on this issue the results from a number of commonly accepted statistical measures of performance are reported. Unfortunately, none is without problems. First, the industry-rotation strategies examined here are neither the pure selectivity strategies implicit in Jensen's (1968) test, nor the pure market-timing strategies embodied in Treynor and Mazuy's (1966) and Henriksson and Merton's (1981) tests of market timing. Second, Roll (1978) argues that Jensen's test is ambiguous because the choice of the benchmark (market) portfolio affects both systematic risk (beta) and abnormal return (alpha), also see (Dybvig & Ross, 1985; Grauer, 1991; Green, 1986). Third, expected returns and risk measures may vary with economic conditions.

In light of these problems this study employs an eclectic mix of performance measures that include conditional and unconditional versions of the Jensen, Henriksson-Merton and Treynor-Mazuy tests as well as the portfolio change measure, see Grinblatt and Titman (1993) which gauges performance without reference to a proxy for the market portfolio. For each of the measures the null hypothesis is that there is no superior investment performance. The alternative hypothesis is a one-tailed



test that there is superior performance. The Jensen, Henriksson-Merton and Treynor-Mazuy regressions are corrected for heteroskedasticity using White's (1980) correction.

The unconditional Jensen (1968) test is based on the regression

$$R_{pt} = \alpha_p + \beta_p R_{mt} + u_{pt}, \quad (14)$$

where  $R_{pt} = r_{pt} - r_{Lt}$  is the excess return on portfolio  $p$  over the Treasury bill rate,  $R_{mt} = r_{mt} - r_{Lt}$  is the excess return on the CRSP value-weighted index,  $\alpha_p$  is the unconditional measure of performance and  $\beta_p$  is the unconditional measure of risk.

However, expected returns and betas almost certainly change over time. Therefore, Ferson and Schadt (1996) and Ferson and Warther (1996) among others, building on the earlier work of Shanken (1990) advocate conditional performance measures. This study follows their suggestion that a portfolio's risk is related to dividend yields and short-term Treasury yields postulating that

$$\beta_p = b_{0p} + b_{1p} dy_{t-1} + b_{2p} r_{Lt}, \quad (15)$$

where  $dy_{t-1}$  is the CRSP value-weighted index annual dividend yield at the beginning of period  $t$  and  $r_{Lt}$  is the (observable) beginning-of-quarter Treasury bill rate, both measured as deviations from their estimation-period means. Substituting equation (15) into equation (14), yields the conditional Jensen test

$$R_{pt} = \alpha_{cp} + b_{0p} R_{mt} + b_{1p} [dy_{t-1} R_{mt}] + b_{2p} [r_{Lt} R_{mt}] + \epsilon_{pt}, \quad (16)$$

where  $\alpha_{cp}$  is the conditional measure of performance, and  $b_{1p}$  and  $b_{2p}$  measure how the conditional beta varies with dividend yields and Treasury bill rates.

The unconditional regression specification for the Treynor and Mazuy (1966) test is

$$R_{pt} = \alpha_p + \beta_p R_{mt} + \gamma_p R_{mt}^2 + u_{pt}, \quad (17)$$

where  $\alpha_p$  is the measure of selectivity,  $\beta_p$  is the unconditional beta and  $\gamma_p$  is the market-timing coefficient. Substituting for  $\beta_p$ , the conditional regression specification is

$$R_{pt} = \alpha_{cp} + b_{0p} R_{mt} + b_{1p} [dy_{t-1} R_{mt}] + b_{2p} [r_{Lt} R_{mt}] + \epsilon_{pt}, \quad (18)$$

where  $\alpha_{cp}$ ,  $b_{1p}$ ,  $b_{2p}$  and  $\gamma_p$  are defined above.

The unconditional Henriksson and Merton (1981) test is given by

$$R_{pt} = \alpha_p + b_{dp} R_{mt} + \gamma_p \max(0, R_{mt}) + u_{pt}, \quad (19)$$

where  $\alpha_p$  is the measure of selectivity,  $\beta_p$  is the down-market beta,  $\gamma_p$  is the market-timing coefficient, in this case the difference between the up- and down-market beta, and  $\max(0, R_{mt})$  is the payoff on a call option on the market with an exercise price equal to the riskfree rate of interest. Following Ferson and Schadt (1996) the

conditional Henriksson-Merton test is

$$R_{pt} = \alpha_{cp} + b_{dp}R_{mt} + b_{1p}[dy_{t-1}R_{mt}] + b_{2p}[r_{Lt}R_{mt}] + \gamma_p R_{mt}^* + b_{1p}^*[dy_{t-1}R_{mt}^*] + b_{2p}^*[r_{Lt}R_{mt}^*] + e_{pt}, \quad (20)$$

where  $R_{mt}^*$  is the product of the excess return on the CRSP value-weighted index and an indicator dummy for positive values of the difference between the excess return on the index and the conditional mean of the excess return. (The conditional mean is estimated by a linear regression of the excess return of the CRSP value-weighted index on  $dy_{t-1}$  and  $r_{Lt}$ .) The most important coefficients is  $\gamma_p$ , the market-timing coefficient, which in this case is the difference between the up- and down-market conditional betas.

In contrast to most other performance measures, Grinblatt and Titman's (1993) portfolio change measure employs portfolio holdings as well as rates of return and does not require an external benchmark (market) portfolio. In order to motivate the portfolio change measure, assume that uninformed investors perceive that the vector of expected returns is constant, while informed investors can predict whether expected returns vary over time. Informed investors can profit from changing expected returns by increasing (decreasing) their holdings of assets whose expected returns have increased (decreased). The holding of an asset that increases with an increase in its conditional expected rate of return will exhibit a positive unconditional covariance with the asset's returns. The portfolio change measure is constructed from an aggregation of these covariances. For evaluation purposes, let

$$PCM_t = \sum_i r_{it} (x_{it} - x_{i,t-j}),$$

where  $r_{it}$  is the quarterly rate of return on asset  $i$  time  $t$ ,  $x_{it}$  and  $x_{i,t-j}$  are the holdings of asset  $i$  at time  $t$  and time  $t-j$ , respectively. This expression provides an estimate of the covariance between returns and weights at a point in time. Alternatively, it may be viewed as the return on a zero-weight portfolio. The portfolio change measure is an average of the  $PCM_t$ 's

$$\overline{PCM} = \sum_t \sum_i [r_{it}(x_{it} - x_{i,t-j})/T], \quad (21)$$

where  $T$  is the number of time-series observations. The portfolio change measure test itself is a simple  $t$ -test based on the time series of zero-weight portfolio returns, i.e.,

$$t = (\overline{PCM} / \sigma(PCM))\sqrt{T}, \quad (22)$$

where  $\sigma(PCM)$  is the standard deviation of the time series of  $PCM_t$ 's.

In their empirical analysis of mutual fund performance, Grinblatt and Titman work with two values of  $j$  that represent one- and four-quarter lags. They report that four-quarter lag portfolio change measures are statistically significant. Hence, this paper will focus on the four-quarter lag portfolio change measure. The portfolio change measure is particularly apropos in the present study because the portfolio weights are

chosen according to a pre-specified set of rules over the same quarterly time interval as performance is measured. Thus, one does not have to worry about possible gaming or window-dressing problems that face researchers trying to gauge the performance of mutual funds.

## Results

### *Results Based on Statistical Criteria*

Table 1 depicts the labels used in the figures and tables. The out-of-sample statistical forecasting results are reported in Table 2. The findings confirm Bossaerts and Hillion's (1999) and Goyal and Welch's (2003) results that there is little, or no, out-of-sample short-horizon forecasting ability with information variables. There is little to distinguish between any of the forecasts and even some of the small differences are counterintuitive. The total mean square error of the DR means is the largest even though the results in the next section show that the performance of the portfolios generated from them dominates the performance of all but the DRCAPM portfolios. The average out-of-sample R-squares are below 0.01 except for the CAPM estimator. Yet, when this estimator is combined with the power-utility model the resulting portfolios exhibit the worst economic performance. It is also somewhat surprising that the Bayes-Stein estimator, noted for minimizing the total mean square error, exhibits the smallest average out-of-sample R-square.

**Table 1:** *Definitions of the Labels in Tables and Figures*

<b>Benchmarks</b>	
<b>RL</b>	Riskfree lending at the three month Treasury bill rate
<b>VW</b>	Market (value-weighted CRSP index)
<b>V5</b>	50% in VW, 50% in lending
<b>V15</b>	150% in VW, 50% in borrowing at the call money rate plus 1%
<b>V20</b>	200% in VW, 100% in borrowing at the call money rate plus 1%
<b>Estimators of the Menu</b>	
<b>Historic</b>	Historic means
<b>Bayes-Stein</b>	Bayes-Stein means
<b>CAPM</b>	CAPM means
<b>DR</b>	Dividend – yield riskfree – rate means
<b>DRCAPM</b>	Dividend – yield riskfree – rate means shrunk to CAPM means

There are 264 quarterly forecasts of the means for each industry in the 1934-1999 period. At a point in time, each estimator bases its forecast on data from the previous 32-quarters. R-square values are formed by squaring the correlation coefficient between the time series of 264 mean forecasts and the 264 realized return. Mean squared errors are in units of percent squared per quarter. See Table 1 for definitions of the labels.

**Table 2:** *Out-of-Sample R-squares and Mean Squared Errors for Five Estimators of Industry Returns*

	Historic		Bayes-Stein		CAPM		DR		DRCAPM	
	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE
<b>Petroleum</b>	0.00	92	0.00	92	0.00	92	0.00	99	0.00	91
<b>Finance &amp; Real Estate</b>	0.01	102	0.01	102	0.02	103	0.00	121	0.00	104
<b>Consumer Durables</b>	0.00	124	0.00	124	0.02	125	0.02	131	0.01	120
<b>Basic Industries</b>	0.00	85	0.00	85	0.01	85	0.01	93	0.00	84
<b>Food &amp; Tobacco</b>	0.00	65	0.00	65	0.00	65	0.01	73	0.01	65
<b>Construction</b>	0.00	142	0.00	139	0.02	142	0.02	149	0.01	137
<b>Capital Goods</b>	0.01	98	0.00	98	0.01	98	0.01	107	0.01	97
<b>Transportation</b>	0.02	142	0.00	140	0.01	142	0.00	163	0.00	142
<b>Utilities</b>	0.00	52	0.00	53	0.01	53	0.00	64	0.00	55
<b>Textiles &amp; Trade</b>	0.00	112	0.00	111	0.00	111	0.01	125	0.01	112
<b>Services</b>	0.00	214	0.00	213	0.00	211	0.01	237	0.01	212
<b>Leisure</b>	0.03	179	0.00	176	0.03	178	0.01	201	0.00	177
<b>Average R<sup>2</sup></b>	0.01		0.00		0.01		0.01		0.00	
<b>Sum MSE</b>		1407		1398		1405		1563		1397

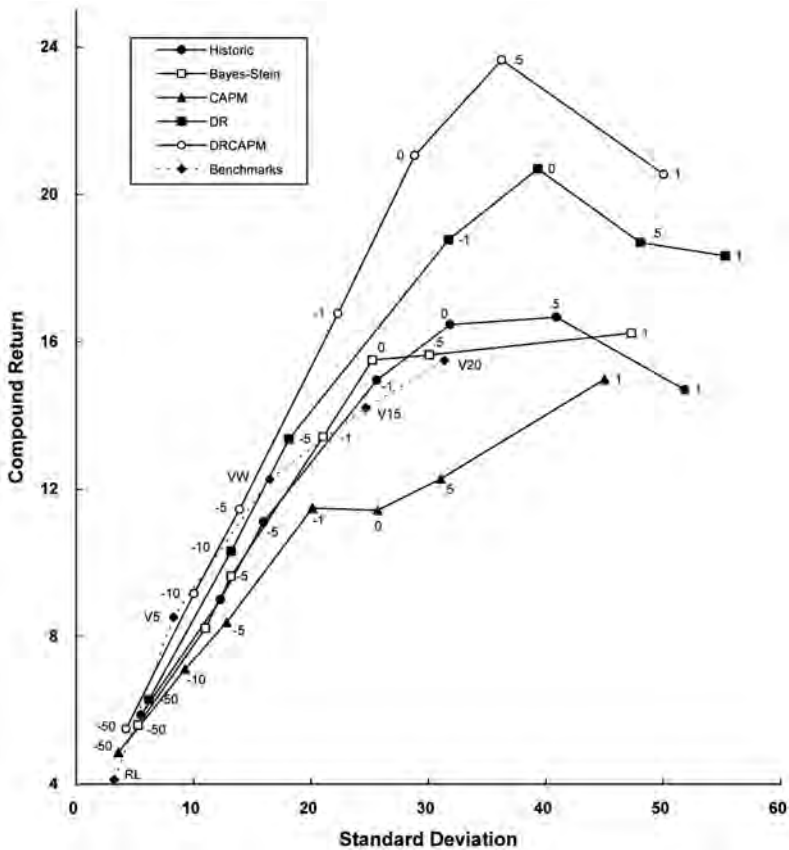
#### *Results Based on Economic Criteria*

Figure 1 plots the annual compound return (obtained by compounding the quarterly realized returns) and standard deviation of the realized returns for five sets of ten power-utility strategies, based on  $\gamma$ 's in equation (1) ranging from -50 (extremely risk averse) to 1 (risk-neutral), for the 66-year period from 1934-99. (For consistency with the compound return the standard deviation is based on the log of one plus the rate of return. This quantity is very similar to the standard deviation of the rate of return for levels less than 25 percent.) Portfolios are chosen each quarter, employing a 32-quarter estimation period, from an investment universe that includes the twelve value-weighted U.S. industry indices, lending and borrowing. The first set of strategies (black circles) shows the portfolio returns generated from historic means. The second set (open squares) displays the portfolio returns based on Bayes-Stein means. The third set (black triangles) depicts those based on CAPM means. The fourth set (black circles) presents the portfolio returns obtained employing the DR forecasts of the means. The fifth set (open circles) presents the portfolio returns obtained employing the DRCAPM forecasts. The figure also shows the benchmarks: RL, V5, VW, V15 and

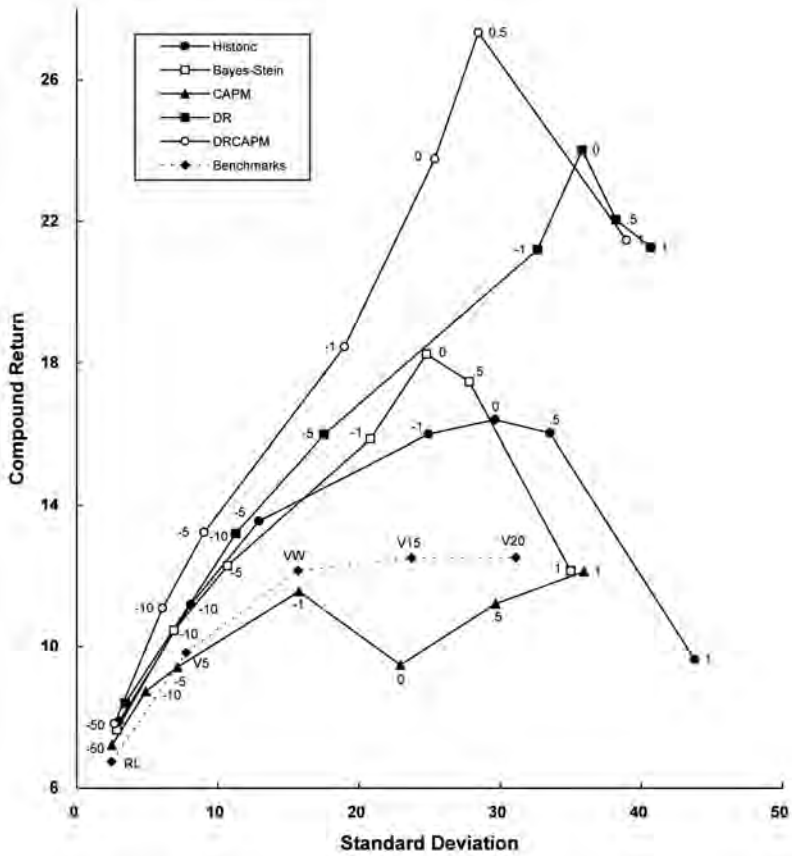
V20 as black diamonds. Figure 2 plots the corresponding results for the 30-year sub-period from 1966-99. Finally, Figure 3 displays the results for the (inflationary) 16-year sub-period from 1966-81, a period which experienced a one-half percent per year *negative* realized risk premium on the value-weighted portfolio of risky assets.

The figures show three main results. The portfolios generated from the DR and DRCAPM estimators “outperform” portfolios that employ the historic, Bayes-Stein and CAPM estimators. The portfolios based on the CAPM estimator are “dominated” by the other four and by the benchmark portfolios. In the 1966-81 period, however, portfolios employing historic and Bayes-Stein estimators do not so obviously dominate portfolios based on the CAPM estimator.

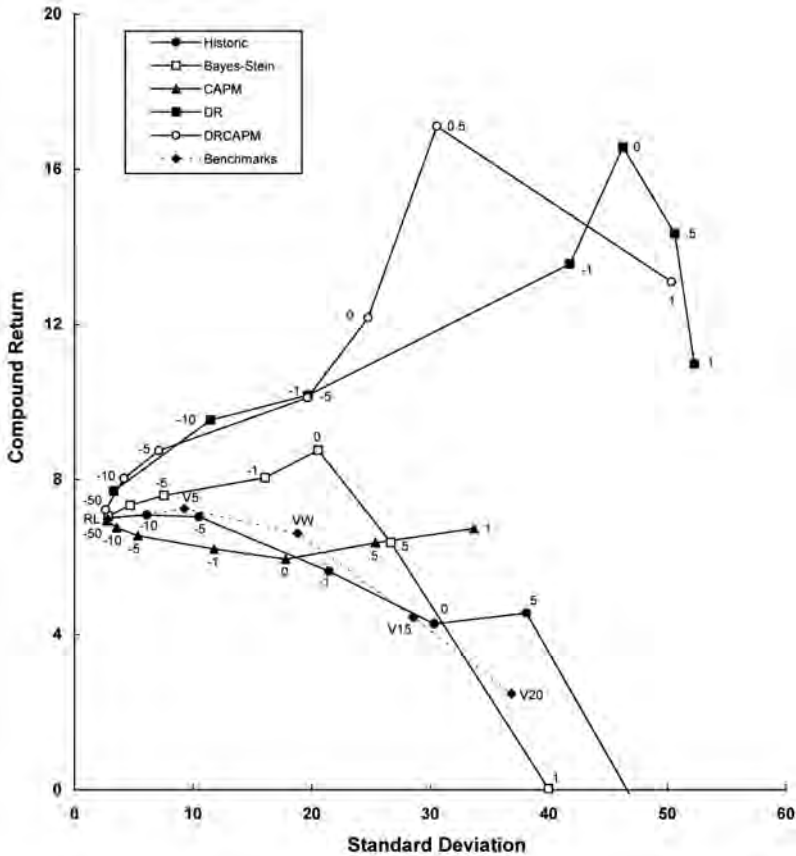
**Figure 1:** Annual compound return versus the standard deviation (of the log of one plus return) for five benchmarks and five sets of power-utility portfolios constructed from historic, Bayes-Stein, CAPM, dividend – yield riskfree – rate (DR), and dividend – yield riskfree – rate-CAPM (DRCAPM) estimators of the means in the 1934-1999 period



**Figure 2:** Annual compound return versus the standard deviation (of the log of one plus return) for five benchmarks and five sets of power-utility portfolios constructed from historic, Bayes-Stein, CAPM, dividend – yield riskfree – rate (DR), and dividend – yield riskfree – rate-CAPM (DRCAPM) estimators of the means in the 1966-1999 period



**Figure 3:** Annual compound return versus the standard deviation (of the log of one plus return) for five benchmarks and five sets of power-utility portfolios constructed from historic, Bayes-Stein, CAPM, dividend – yield riskfree – rate (DR), and dividend – yield riskfree – rate-CAPM (DRCAPM) estimators of the means in the 1966-1981 period



*Results Based on Statistical Measures of Investment Performance*

Table 3 summarizes the results obtained from the conditional and unconditional Jensen tests and from the Grinblatt-Titman portfolio change measure tests in the same three time-periods examined in the figures. For the most part, the results of the tests are consistent with conclusions drawn from the figures and the cumulative wealth values. The alphas and portfolio change measures of the portfolios generated from the DR and DRCAPM estimators are much larger and for the most part more statistically significant than those of the other three mean estimators, particularly in the 1934-99 and 1966-99 periods. The conditional Jensen test uniformly ranks the performance of the historic, Bayes-Stein and CAPM portfolios higher than the unconditional test, which is consistent with Ferson and Schadt's (1996) and Ferson and Warther's (1996) mutual fund results. But this pattern is less obvious for the DR and DRCAPM

portfolios. Furthermore, in the 1966-81 period the DR and DRCAPM portfolios perform poorly according to the portfolio change measure test, which is just the opposite of what is shown in Figure 3.

The alphas and portfolio change measures are averages calculated over ten power portfolios. Both are measured in units of percent per quarter. The portfolio change measures are based on a four-quarter lags. For both measures the null hypothesis is that there is no superior investment performance. The alternative hypothesis is a one-tailed test that there is superior performance. The Jensen regressions are corrected for heteroskedasticity using White's (1980) correction. Number  $\leq 0.05$  refers to the number of portfolios out of ten whose coefficients are statistically significant at the 5% level. See Table 1 for label definitions.

**Table 3: Unconditional and Conditional Jensen Alphas and Grinblatt-Titman Portfolio Change Measures for Ten Power Portfolios Estimated from Five Sets of Means**

	Unconditional Jensen			Conditional Jensen			Grinblatt-Titman		
	Alpha	Number Negative	Number $\leq 0.05$	Alpha	Number Negative	Number $\leq 0.05$	PCM	Number Negative	Number $\leq 0.05$
<b>Panel A: Historic Means</b>									
1934-1999	0.50	0	3	0.48	1	8	0.82	0	10
1966-1999	0.57	1	5	0.76	0	7	1.00	0	10
1966-1981	-0.20	4	0	0.62	0	0	0.72	0	6
<b>Panel B: Bayes-Stein Means</b>									
1934-1999	0.54	0	4	0.54	1	9	0.75	0	9
1966-1999	0.70	0	6	0.81	0	9	1.03	0	10
1966-1981	0.06	1	0	0.62	0	3	0.96	0	9
<b>Panel C: CAPM Means</b>									
1934-1999	0.17	0	0	0.13	1	0	0.05	8	1
1966-1999	0.11	1	0	0.20	1	0	0.21	1	0
1966-1981	0.06	8	0	0.86	0	3	0.26	1	0
<b>Panel D: DR Means</b>									
1934-1999	0.80	0	4	0.73	0	6	1.55	0	10
1966-1999	1.40	0	8	0.90	0	1	2.16	0	9
1966-1981	1.75	0	4	0.53	0	0	1.18	4	1
<b>Panel E: DRCAPM Means</b>									
1934-1999	0.90	0	9	0.88	0	9	1.00	0	10
1966-1999	1.28	0	10	1.16	0	9	1.33	0	10
1966-1981	1.07	0	3	0.89	0	2	0.85	0	0



Table 4 contains the results of the unconditional and conditional Henriksson-Merton and Treynor-Mazuy market-timing tests. There is little or no evidence of market-timing ability according to both unconditional tests. In the 1934-99 period the majority of the historic, Bayes-Stein and CAPM portfolios show negative timing ability. In the 1966-99 period there is more evidence of market-timing ability especially according to the conditional tests. But the tests indicate more market-timing ability for the historic, Bayes-Stein and CAPM portfolios than for the DR and DRCAPM portfolios. In the 1966-81 period the results are anomalous. The conditional tests, especially the Treynor-Mazuy conditional test, show strong evidence of market-timing ability for the historic, Bayes-Stein and CAPM estimators, and no evidence of market-timing ability for any of the dividend-yield riskfree-rate estimators, which seems to be at variance with the results reported in Figure 3 and Table 3.

The reported gammas (e.g. the timing coefficients) are averages calculated over ten power portfolios. The null hypothesis is that there is no timing ability. The alternative hypothesis is a one-tailed test that there is positive timing ability. The regressions are corrected for heteroskedasticity using White's (1980) correction. Number  $\leq 0.05$  refers to the number of portfolios out of ten whose timing coefficients are statistically significant at the 5% level. See Table 1 for label definitions.

**Table 4:** Unconditional and Conditional Henriksson-Merton and Treynor-Mazuy Timing Coefficients for Ten Power Portfolios Estimated from Five Sets of Means

	Henriksson-Merton						Treynor-Mazuy					
	Unconditional			Conditional			Unconditional			Conditional		
	Gamma	Number Negative	Number $\leq 0.05$	Gamma	Number Negative	Number $\leq 0.05$	Gamma	Number Negative	Number $\leq 0.05$	Gamma	Number Negative	Number $\leq 0.05$
<b>Panel A: Historic Means</b>												
1934-1999	0.04	6	0	0.09	1	0	-0.001	9	0	0.003	1	0
1966-1999	0.18	1	0	0.29	1	8	0.005	1	0	0.008	0	7
1966-1981	-0.10	3	0	0.17	1	6	-0.002	2	0	0.007	0	9
<b>Panel B: Bayes-Stein Means</b>												
1934-1999	0.01	8	0	0.08	1	0	-0.002	9	0	0.003	0	0
1966-1999	0.19	1	0	0.63	0	7	0.005	1	0	0.007	0	2
1966-1981	-0.06	2	0	0.14	1	3	0.000	1	0	0.005	1	9
<b>Panel C: CAPM Means</b>												
1934-1999	0.01	8	0	0.05	0	0	-0.002	9	0	0.001	1	0
1966-1999	0.08	1	0	0.15	0	0	0.002	1	0	0.004	0	0
1966-1981	-0.02	3	0	0.27	0	9	0.000	2	0	0.010	0	10
<b>Panel D: DR Means</b>												
1934-1999	0.33	0	2	0.20	0	1	0.009	0	1	0.005	1	8
1966-1999	0.43	0	5	0.19	0	3	0.008	0	4	0.003	3	1
1966-1981	0.38	0	0	0.13	0	0	0.004	2	0	-0.011	10	0
<b>Panel E: DRCAPM Means</b>												
1934-1999	0.15	0	1	0.18	0	1	0.001	8	1	0.003	0	0
1966-1999	0.28	0	4	0.24	0	0	0.006	0	0	0.005	0	6
1966-1981	0.08	2	0	0.20	0	1	-0.001	9	0	-0.005	10	0

### *Robustness of the Results*

Four additional estimators of the means were examined. First, because of the variability in the dividend-yield riskfree-rate means, these forecasts were shrunk to historic means. The portfolios based on the dividend-yield riskfree-rate – historic (DRH) means performed better than those based solely on dividend-yield riskfree-rate means, but not as well as those based on dividend-yield riskfree-rate means shrunk to CAPM (e.g., DRCAPM) means. Second, in order to more fully investigate the Henriksson-Merton and Treynor-Mazuy market-timing test results, which indicate that there is little or no market timing ability in portfolios generated from different mean forecasts, two timing means were developed. The mean on the market (rather than the means on industries) was forecast using the dividend-yield riskfree-rate – historic mean method. These forecasts were called  $\mu_{DRMKT}$  means. Then, a second set of CAPM means was developed using the  $\mu_{DRMKT}$  means to set the slope of the SML, i.e.,  $\mu_{DRMCAPM_i} = r_{Li} + (\mu_{DRMKT} - r_{Li})\beta_i$ . The DRMKT portfolios (that consisted of the market and either borrowing or lending only) and the industry-rotation DRMCAPM portfolios timed the market according to the Henriksson-Merton and Treynor-Mazuy market-timing tests. The industry-rotation portfolios based on the DRMCAPM means accumulated more wealth than the pure market timing DRM portfolios, but not as much as the industry rotation strategies based on the DRH and DRCAPM means. Finally, the robustness of the 32-quarter moving window approach was examined by comparing it with an expanding window approach that employed an all-of-history dividend-yield riskfree-rate – historic mean forecast of industry means. The portfolios based on this expanding-window method of forecasting the means performed much worse than portfolios based on the other dividend-yield riskfree-rate mean forecasts.

On a different dimension, a one-quarter portfolio change measure test was conducted. Like Grinblatt-Titman's (1993) results for mutual funds there was no statistically significant abnormal performance according to this measure. Moreover, the DR and DRCAPM portfolios performed abysmally in all three periods according to the one-quarter lag portfolio change measure, which is at complete odds with the other evidence presented in the paper.

### **Summary and Concluding Comments**

This paper evaluates historic, Bayes-Stein, CAPM and two dividend-yield riskfree-rate estimators of asset means employing statistical and economic criteria in an industry-rotation setting using quarterly data. None of the estimators exhibit much in the way of out-of-sample predictive ability when judged by statistical criteria. The average out-of-sample R-squares are below 0.01 for all but the CAPM estimator. Moreover, there is little difference in the mean square errors of the five industry-rotation mean estimators. Yet, when the mean estimators are combined with a discrete-time power-utility portfolio selection model, the resulting portfolios earn economically significant returns. Even so, judged in terms of the compound return, standard deviation plots in Figures 1-3, or in terms of accumulated wealth, some of the resulting portfolios perform appreciably better than others. Specifically, the two dividend-yield riskfree-rate portfolios perform by far the best and the traditional CAPM portfolios

perform the worst. For the most part, unconditional and conditional Jensen and Grinblatt-Titman tests support the compound return - standard deviation and wealth rankings, especially in the 1934-99 and 1966-99 periods. The unconditional and conditional Treynor-Mazuy and Henriksson-Merton market-timing tests indicate that the results did not arise from any market-timing ability.

So, are returns predictable? Clearly, the answer depends on the length of the decision horizon examined and metric chosen. With quarterly returns, the out-of-sample statistical answer is a clear “no”—and the economic answer is a resounding “yes.” In this case, one can argue that the economic answer is compelling. You can’t spend a slope coefficient, a t-statistic, an R-square, or a mean square error. But, four of the more risk-tolerant power-utility portfolios generated from DRCAPM means grew from one dollar to \$28,000, \$301,000, \$1,213,000 or \$228,000 over the 1934-99 period. These portfolios provided investors with real spending opportunities—especially when compared to an investment in the market that grew to only \$3,600, or an investment of 200% in the market financed by 100% borrowing that grew to \$9,700 over the same period. In this case, Leitch and Tanner (1991: p. 580) were correct in suggesting that: “... least-squares regression analysis may not be appropriate for many studies of economic behavior.”

## References

- Ang, A. & Bekaert, G. (2003). Stock Return Predictability: Is it There? Manuscript, Columbia University.
- Doron, A. (2002). Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics*, 64: 423-458.
- Doron, A. (2004). Stock Return Predictability and Asset Pricing Models. *Review of Financial Studies*, 17: 699-738.
- Doron, A. & Tarun, C. (2006). Predicting Stock Returns. *Journal of Financial Economics*, 82: 387-415.
- Balduzzi P. & Lynch, A.W. (1999). Transaction Costs and Predictability: Some Utility Cost Calculations. *Journal of Financial Economics*, 52: 47-78.
- Barberis, N. (2000). Investing for the Long Run when Returns are Predictable. *Journal of Finance*, 55: 225-264.
- Beller, K.R., Kling, J.L. & Levinson, M. (1998). Are Industry Stock Returns Predictable? *Financial Analysts Journal*, 54: 42-57.
- Best, M.J. (1975). A Feasible Conjugate Direction Method to Solve Linearly Constrained Optimization Problems. *Journal of Optimization Theory and Applications*, 16: 25-38.
- Best, M.J. & Grauer, R.R. (1985). Capital Asset Pricing Compatible with Observed Market Value Weights. *Journal of Finance*, 40: 85-103.
- Best, M.J. & Grauer, R.R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Review of Financial Studies*, 4: 315-342.
- Best, M.J. & Grauer, R.R. (1992). Positively Weighted Minimum-Variance Portfolios and the Structure of Asset Expected Returns. *Journal of Financial and Quantitative*

*Analysis*, 27: 513-537.

- Black, F & Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 48: 28-43.
- Bossaerts, P. & Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies*, 12: 405-428.
- Brandt, M.W. (1999). Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach. *Journal of Finance*, 54: 1609-1645.
- Brennan, M.J., Schwartz, E.S. & Lagnado, R. (1997). Strategic Asset Allocation. *Journal of Economic Dynamics and Control*, 21: 1377-1403.
- Brennan, M.J. & Xia, Y. (2002). Dynamic Asset Allocation under Inflation. *Journal of Finance*, 57: 1201-1238.
- Campbell, J. & Viceira, L. (May 1999). Consumption and Portfolio Decisions When Expected Returns are Time Varying. *Quarterly Journal of Economics*, 114: 433-495.
- Connor, G. (1997). Sensible Return Forecasting for Portfolio Management. *Financial Analysts Journal*, 53: 44-51.
- Dybvig, P.H. & Ross, S.A. (1985). The Analytics of Performance Measurement Using a Security Market Line. *Journal of Finance*, 40: 401-416.
- Fama, E.F. & French, K.R. (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics*, 22: 3-25.
- Fama, E.F. & French, K.R. (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, 25: 23-49.
- Ferson, W.E. & Schadt, R.W. (1996). Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance*, 51: 25-462.
- Ferson, W.E. & Warther, V.A. (1996). Evaluating Fund Performance in a Dynamic Market. *Financial Analysts Journal*, 52: 20-28.
- Ferson, W.E. & Siegel, A.F. (2001). The Efficient Use of Conditioning Information in Portfolios. *Journal of Finance*, 56: 967-982.
- Ferson, W.E., Sarkissian, S. & Simin, T. (2003). Spurious Regressions in Financial Economics? *Journal of Finance*, 58: 1393-1413.
- Frost, P.A., & Savarino, J.E. (1988). For Better Performance: Constrain Portfolio Weights. *Journal of Portfolio Management*, 15: 29-34.
- Goetzmann, W.N. & Jorion P. (1993). Testing the Predictive Power of Dividend Yields. *Journal of Finance*, 48: 663-680.
- Grauer, R.R. (1991). Further Ambiguity When Performance Is Measured by the Security Market Line. *Financial Review*, 26: 569-585.
- Grauer, R.R. (2007). Extreme Mean-Variance Solutions: Estimation Error versus Modeling Error? Manuscript, Simon Fraser University.
- Grauer, R. R. & Hakansson, N.H. (1986). A Half-Century of Returns on Levered and Unlevered Portfolios of Stocks, Bonds, and Bills, with and without Small Stocks. *Journal of Business*, 59: 287-318.
- Grauer, R.R. & Hakansson, N.H. (1987). Gains from International Diversification: 1968-85 Returns on Portfolios of Stocks and Bonds. *Journal of Finance*, 42: 721-739.
- Grauer, R.R. & Hakansson, N.H. (1995). Stein and CAPM Estimators of the Means in Asset Allocation. *International Review of Financial Analysis*, 4: 35-66.
- Grauer, R.R., Hakansson, N.H. & Shen, F.C. (1990). Industry Rotation in the U.S.

- Stock Market: 1934-1986 Returns on Passive, Semi-passive, and Active Strategies. *Journal of Banking and Finance*, 14: 513-535.
- Grauer, R.R. & Shen, F.C. (2000). Do Constraints Improve Portfolio Performance? *Journal of Banking and Finance*, 24: 1253-1274.
- Green, R.C. (1986). Benchmark Portfolio Inefficiency and Deviations from the Security Market Line. *Journal of Finance*, 41: 295-312.
- Green, R.C. & Hollifield, B. (1992). When Will Mean-Variance Efficient Portfolios Be Well-Diversified? *Journal of Finance*, 47: 1785-1810.
- Grinblatt, M. & Titman, S. (1993). Performance Measurement Without Benchmarks: An Examination of Mutual Fund Returns. *Journal of Business*, 66: 47-68.
- Hakansson, N.H. (1971). On Optimal Myopic Portfolio Policies, With and Without Serial Correlation of Yields. *Journal of Business*, 44: 324-34.
- Hakansson, N.H. (1974). Convergence to Isoelastic Utility and Policy in Multiperiod Portfolio Choice. *Journal of Financial Economics*, 1: 201-24.
- Hawawini, G. & Keim, D. (1995). On the Predictability of Common Stock Returns: World Wide Evidence. In Jarrow, R.A. Maksimovic, V. & Ziemba, W.T. editors: *Finance, Handbooks in Operations Research & Management Science*, Amsterdam: North Holland (9): 497-544.
- Henriksson, R.D. & Merton, R.C. (1981). On Market Timing and Investment Performance II. Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business*, 54: 513-533.
- Hodrick R.J. (1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *Review of Financial Studies*, 5: 357-386.
- Huberman, G. & Ross, S.A. (1983). Portfolio Turnpike Theorems, Risk Aversion and Regularly Varying Utility Functions. *Econometrica*, 51: 1104-19.
- Jensen, M.C. (1968) The Performance of Mutual Funds in the Period 1945-1964. *Journal of Finance*, 23: 389-416.
- Jobson, J.D., Korkie, B. & Ratti, V. (1979). Improved Estimation for Markowitz Portfolios using James-Stein Type Estimators. *Proceedings of the American Statistical Association*, Business & Economics Statistics Section, 41: 279-284.
- Jobson, J.D. & Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75: 544-554.
- Jobson, J.D. & Korkie, B. (1981). Putting Markowitz Theory to Work. *Journal of Portfolio Management*, 7: 70-74.
- Jorion, P. (1985). International Portfolio Diversification with Estimation Risk. *Journal of Business*, 58: 259-278.
- Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis* 21: 279-292.
- Jorion, P. (1991). Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection. *Journal of Banking and Finance*, 15: 717-727.
- Klemkosky, R.C. & Bharati, R. (1991). Time-Varying Expected Returns and Asset Allocation. *Journal of Portfolio Management*, 21: 80-88.
- Leitch, G. & Tanner, J.E. (1991). Economic Forecast Evaluation: Profits versus the Conventional Error Measures. *American Economic Review*, 81 (June 1991), 580-599.
- Leland, H. (1972). On Turnpike Portfolios. In *Mathematical Methods in Investment and*

- Finance*. Shell, K. & Szego, G.P., editors., Amsterdam: North-Holland.
- Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47: 13-47.
- Lynch, A.W. (2001). Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability. *Journal of Financial Economics*, 62: 67-130.
- Lynch, A.W. & Balduzzi, P. (2000). Predictability and Transaction Costs: The Impact on Rebalancing Rules and Behavior. *Journal of Finance*, 55: 2285-2309.
- Marquering, W. & Verbeek, M. (2001). The Economic Value of Predicting Stock Index Returns and Volatility. Manuscript, Erasmus University
- Merton, R.C. (1971). Optimum Consumption and Portfolio Rules in a Continuous Time Model. *Journal of Economic Theory*, 3: 373-413.
- Merton, R.C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, 41: 867-888.
- Merton, R.C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8: 323-361.
- Michaud, R.O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*, 45: 31-42
- Mossin, J. (1968). Optimal Multiperiod Portfolio Policies. *Journal of Business*, 41: 215-229.
- Pelaez, R.F. (1998). Economically Significant Stock Market Forecasts. *Financial Review*, 33: 65-76.
- Pastor, L. (2000). Portfolio Selection and Asset Pricing Models. *Journal of Finance*, 55: 179-223.
- Pastor, L. & Stambaugh, R.F. (2000). Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics*, 56: 335-381.
- Roll, R. (1978). Ambiguity When Performance Is Measured by the Securities Market Line. *Journal of Finance*, 33: 1051-1069.
- Ross, S.A. (1974). Portfolio Turnpike Theorems for Constant Policies. *Journal of Financial Economics*, 1: 171-98.
- Schwert, G.W. (2003). Anomalies and Market Efficiency. In Constantinides, G.M., Harris, M. & Stultz, R. editors: *Handbook in the Economics of Finance*, Amsterdam: Elsevier Science B.V.: 939-972.
- Shanken, J. (1990). Intertemporal Asset Pricing: an Empirical Investigation. *Journal of Econometrics*, 45: 99-120.
- Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19: 425-442.
- Solnik, B. (1993). The Performance of International Asset Allocation Strategies Using Conditioning Information. *Journal of Empirical Finance*, 1: 33-55.
- Stein, C. (1955). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics I*, Berkeley: University of California Press.
- Stambaugh, R.F. (1999). Predictive Regressions. *Journal of Financial Economics*, 54: 375-421.
- Treynor, J.L. & Mazuy, K. (1966). Can mutual funds outguess the market? Harvard

*Business Review*, 44: 131-136.

Torous, W., Valkanov, R. & Yan, S. (2005) On Predicting Stock Returns with Nearly Integrated Explanatory Variables. *Journal of Business*, 78 (in press).

White, H. (1980). A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48: 817-838.





# Brand Response-Effects of Perceived Sexual Harassment in the Workplace

Jeremy J. Sierra<sup>1</sup>

Texas State University – San Marcos

Nina Compton

New Mexico State University

Kellilynn M. Frias-Gutierrez

The University of Arizona

*Although sexual harassment literature indicates significant relationships between sexual harassment and both individual and organizational outcomes, no published research has examined the effects of sexual harassment on attitude and behavioral brand-related factors of potential job applicants. To help close this research gap, a structural equation model that relates perceived sexual harassment in the workplace to attitudes toward the brand, brand image, and intentions to work for a firm, is developed and tested. Using data from undergraduate business students, the empirical results provide support for these relationships and the structural equation model.*

Business scholars and professionals agree that the negativity associated with workplace sexual harassment can impede strategy implementation and also damage stakeholder value. Although the study of work-related sexual harassment is not new (e.g., individual and organizational outcomes of sexual harassment have been examined), the simultaneous effects of such ill-advised behavior on brand perceptions and employee recruitment remains a mystery.

---

<sup>1</sup>The authors wish to thank Michael R. Hyman and Shaun McQuitty for their comments and suggestions on previous versions of this article, as well as, Robert S. Heiser and Harry A. Taute for their input and insight regarding the brand image scale developed and used in this study.

Thus, it is unclear how a firm's reputation of a sexual harassment culture influences prospective job applicant's responses toward the brand and their willingness to work for the firm. To help fill this research gap, we propose a conceptual model that examines the effects of perceived sexual harassment in the workplace (i.e., in a sales firm) on attitudinal (i.e., attitudes toward the brand, brand image) and behavioral (i.e., intentions to work) brand-response factors of potential employees.

### Sexual Harassment Literature

Sexual harassment is a serious social dilemma that negatively affects individuals, organizations, and society (Gelfand, Fitzgerald & Drasgow, 1995; O'Connell & Korabik, 2000). Since 1986, the U.S. Supreme Court has recognized sexual harassment as being illegal (e.g., *Harris v. Forklift Systems, Inc.*, 1993; *Meritor Savings Bank v. Vinson*, 1986). However, over the past ten years, reported incidents of workplace sexual harassment have increased 2,700 percent (Gilbert, 2005).

According to the Equal Employment Opportunity Commission (EEOC) (1980), the federal agency in charge of enforcing sexual harassment claims, "unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature constitute *sexual harassment* when (1) submission to such conduct is made either explicitly or implicitly a term or condition of an individual's employment, (2) submission to or rejection of such conduct by an individual is used as the basis for an employment decision affecting such individual, or (3) such conduct has the purpose or effect of unreasonably interfering with an individual's work performance or creating an intimidating, hostile, or offensive work environment" (29 CFR Part 1604).

Based on this definition, two types of sexual harassment are recognized, including *quid pro quo* and *hostile environment*. The former pertains to hiring, firing, promoting, and/or compensating based on an employee's submission to sexual demands; the latter occurs when a perpetrator creates an odious or daunting work environment for victims (Bennett-Alexander & Pincus, 1994; Robinson et al., 1998).

To qualify for *quid pro quo* sexual harassment, the following conditions must be evident: (1) there must be a request or demand for sexual favors; (2) there must be an expressed or implied threat to a "material" job benefit related to the employee's acceptance or rejection of the sexual activity; and (3) the supervisor or manager must be in a position to implement the threat (*Moylan v. Maries County*, 1986). In order for the harassment to be considered *hostile environment* sexual harassment, the victim must show that the verbal or physical conduct was sufficient, severe, or pervasive, which created an abusive environment that adversely affected his/her ability to work effectively (Bennett-Alexander & Pincus, 1994; Robinson, Fink & Lane, 1994).

From defining its domain and tabulating its frequency, research on sexual harassment has advanced to examining its antecedents and effects (Collins & Blodgett, 1981). Previous studies show that the organizational climate, job gender context (e.g., proportion of women in the organization), demographic factors like gender, and behavioral traits (e.g., aggressive and forceful behavior) affect sexual harassment (Fitzgerald et al., 1997; O'Connell & Korabik, 2000; Terpstra & Baker, 1986). Additionally, obtaining sexual favors or activities and/or abusing or increasing one's

power over someone else are significant determinants of workplace sexual harassment (O'Leary-Kelly, Paetzold & Griffin, 2000; Stringer et al., 1990).

Previous studies have examined the organizational and individual outcomes of sexual harassment in the workplace. In terms of organizational outcomes resulting from sexual harassment, negative correlates include business and team performance, workgroup productivity, and recruiting, retaining, and motivating employees (Langhout et al., 2005; Lengnick-Hall, 1995; Raver & Gelfand, 2005). Also, lawsuits brought against organizations (Popovich & Licata, 1987; Terpstra & Baker, 1986) and hostile work environments (Fine, Shepherd & Josephs, 1994) can result from workplace sexual harassment. Additionally, arguments suggest that a company's reputation and image (Terpstra & Baker, 1986), and its ability to attract and retain employees (Lengnick-Hall, 1995) can both be negatively affected by workplace sexual harassment.

In terms of individual effects, sexual harassment in the workplace relates negatively to job satisfaction and physical health, but relates positively to lawsuits, employee stress, absenteeism, on-the-job inefficiencies, self-blame, detrimental psychological states (e.g., degradation, depression), and employee turnover (Fitzgerald et al., 1997; Langhout et al., 2005; Robinson et al., 1998; Stedham & Mitchell, 1998; Terpstra & Baker, 1986; Terpstra & Baker, 1989; Willness, Steel, & Lee, 2007).

Although the aforementioned literature is useful in explaining antecedents and effects of workplace sexual harassment, additional research examining a broader range of attitudinal and behavioral consequences of sexual harassment is needed (Gilbert, 2005; O'Connell & Korabik, 2000; Schneider, Swan & Fitzgerald, 1997). For example, studying the effects of sexual harassment on brand-related factors should add meaningful insight into harassment inquiry (Foy, 2000). This may help firms devise strategies (e.g., modify policy statements, emphasize internal branding) to alleviate sexual harassment in the workplace. Thus, we propose for the first time, a conceptual model that examines the brand-related, attitudinal (i.e., attitudes toward the brand, brand image) and behavioral (i.e., intentions to work) effects of workplace sexual harassment of potential job applicants.

### **Brand Image Literature and Hypotheses**

Sexual harassment not only leads to employees feeling negatively about their work and workplace (Tangri, Burt & Johnson, 1982), it also adversely affects organizations (e.g., Raver & Gelfand, 2005) and individuals (e.g., Willness, Steel & Lee, 2007). For example, in casino organizations, government agencies, and public utility companies, sexual harassment either by supervisors or coworkers negatively affects work satisfaction and organizational commitment (Fitzgerald et al., 1997; Glomb et al., 1997; Morrow, McElroy & Phillips, 1994; Stedham & Mitchell, 1998). Likewise, in the U.S. Armed Forces, sexual harassment has a negative effect on job satisfaction, psychological well-being, health perceptions, workgroup productivity, and organizational commitment (Langhout et al., 2005). For Latinas who experienced sexual harassment, job dissatisfaction and life dissatisfaction, as well as organizational

withdrawal, increased (Cortina, Fitzgerald & Drasgow, 2002).

Additionally, workplace sexual harassment can be detrimental in terms of developing a brand, hiring, and retaining effective employees. For example, conceptual frameworks suggest that sexual harassment has a negative effect on company reputation and image (Foy, 2000; Terpstra & Baker, 1986) through direct (e.g., confrontation) and indirect (e.g., avoidance) effects on victims. It can also negatively affect attraction and retention of employees (Lengnick-Hall, 1995).

The financial ramifications of such hidden costs like demoralized employees, in addition to overt costs such as litigation, can prove difficult to overcome and detrimental to organizational development, sustainability, and growth (Foy, 2000; Terpstra & Baker, 1986). Thus, workplace sexual harassment leads to grim consequences for organizations, individuals, and society in general. Because of the negative perception associated with, and the damaging effects of sexual harassment, we expect workplace sexual harassment in a sales context to correlate negatively with prospective employee responses toward the brand and their willingness to work for the firm. Thus, we tested the following hypotheses:

H1: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their attitudes toward the brand.

H2: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their brand image of this firm.

H3: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their intentions to work for this firm.

Arguments suggest that consumers' attitudes toward brands, which are internal assessments of particular brands (Mitchell & Olson, 1981), can be influenced by their perceived image of such brands (Winters, 1986), where brand image pertains to, "perceptions about a brand as reflected by the brand associations held in consumer memory," (Keller, 1993, p.3). Thus, corporate-based brand representation via economic and non-economic factors, influences cognitive-related brand responses (Fombrun & Shanley, 1990). For example, corporate credibility has a positive effect on attitudes toward the brand (Goldsmith, Lafferty & Newell, 2000; Lafferty, Goldsmith & Newell, 2002). Also, in examining consumer behavior pertaining to low-involvement household goods (e.g., dish detergent) and high-involvement cosmetic goods (e.g., lotion), corporate image correlates positively with attitudes toward the brand (Suh & Yi, 2006). The attitudes toward the brand construct is considered a determinant of and thus, distinct from brand image (MacKenzie & Lutz, 1989; Suh & Yi, 2006). We anticipate then, that in a sales-oriented and recruitment-related context, brand image will relate positively with brand attitudes. Hence, we suggest the following hypothesis:

H4: Prospective employees' brand image of a sales firm is positively related to their attitudes toward the brand.

Generally, consumers' attitudes strongly influence their purchase behaviors (Holmes & Crocker, 1987; Pope & Voges, 2000; Priester et al., 2004; Whittler, 1989).

For example, across various advertising contexts, attitudes toward the brand are positively related with intentions to purchase the advertised brand (Brown & Stayman, 1992; Goldsmith et al., 2000; Lafferty et al., 1989). Because brand attitudes are a basis for consumer intentions and behaviors (Keller, 1993), we expect prospective employees' brand attitudes toward a potential employer to correlate positively to their willingness to work for this firm. Thus, we assessed the following hypothesis:

H5: Prospective employees' attitudes toward a sales firm's brand are positively related to their intentions to work for this firm.

Building on the notion that brand perceptions have the capacity to influence consumer behavioral responses toward the brand source, arguments suggest that customers' perceptions of brand image can influence their behaviors and intentions regarding that brand (Biel, 1992; Ferrand & Pages, 1999). For example, for sporting event sponsors, corporate image relates positively to purchase intention (Pope & Voges, 2000). Similarly, for print advertisers, brand image correlates positively with purchase intentions of the advertised brand (Batra & Homer, 2004).

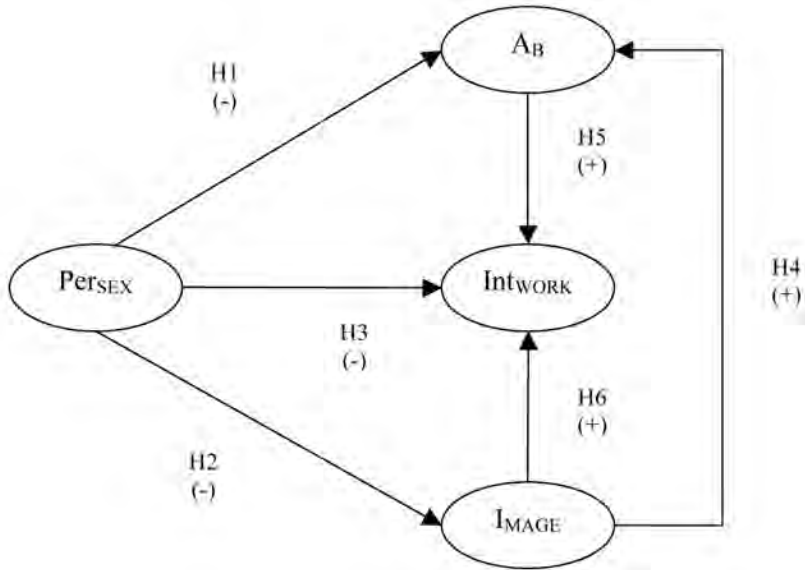
Applying these findings to employment and recruiting, Belt and Paolillo (1982) found a positive relationship between corporate image and reader response to a recruitment advertisement. Turban and Greening (1996) revealed a positive relationship between corporate social performance and attracting applicants. Additionally, in a recruiting context, Gatewood, Gowan and Lautenschlager (1993), found a positive relationship between corporate image and initial decisions about pursuing contact with organizations. Some arguments also suggest that employer image influences employee attraction (Backhaus & Tikoo, 2004). These findings and arguments suggest that an employee's willingness to work at a particular firm is influenced by that individual's perceived image of that firm (Sullivan, 2003). Hence, customers and employees appear to respond analogously to certain business facets, such as processes and brands. In terms of processes, customers and employees respond similarly to failed services when recovered effectively and service procedures when role expectations are understood (Chung-Herrera, Goldschmidt & Hoffman, 2004; Mohr & Bitner, 1991). In terms of brand perceptions, customer and employee responses did not vary in terms of firm competence, reliability, and prestige (Chun & Davies, 2006).

Although the aforementioned research offers insight into brand image effects on employee attraction to the firm, no research has specifically examined the direct effect of brand image, as influenced by perceived workplace sexual harassment, on prospective employee intentions to work for the firm. To explore if the brand image of a firm correlates positively with one's willingness to work for the firm, we propose the following hypothesis:

H6: Prospective employees' brand image of a sales firm is positively related to their intentions to work for this firm.

The conceptual model shown in Figure 1 summarizes these six hypotheses. The scales and methods used for data collection prior to evaluating the hypotheses with a structural equation model are described below.

**Figure 1: Conceptual Model**



## Methodology

### Scale Descriptions

The survey contained questions from four scales pertaining to perceived sexual harassment ( $Per_{SEX}$ ), attitudes toward the brand ( $AB$ ), brand image ( $I_{IMAGE}$ ), and intentions to work ( $Int_{WORK}$ ). Table 1 lists the 19 items comprising the four scales. We briefly describe each of these scales.

### Perceived Sexual Harassment

Sexual harassment is regarded as a matter of perception (Terpstra, 1996) and can be experienced personally, indirectly, and/or by third parties. Personal sexual harassment is experienced by one individual as a result of others' actions (e.g., being subjected to suggestive jokes or remarks about one's personal sexuality). Indirect or ambient sexual harassment pertains to indirect exposure to sexual harassment (Glomb et al., 1997) (e.g., being subjected and negatively affected by suggestive jokes or remarks about a coworker's sexuality). Third party sexual harassment entails sexual harassment of a firm's employees by third parties, such as customers and business partners (Aalberts & Seidman, 1994). Examples of this sexual harassment include being subjected to suggestive jokes or remarks about one's personal sexuality from

customers and/or business partners. For this study, our vignette design examines perceived *ambient* sexual harassment in the workplace.

Active measures of sexual harassment may incur substantial self-report bias (Arvey & Cavanaugh, 1995). To help alleviate such bias, a structured, indirect-question, multi-vignette-based scale, for  $Per_{SEX}$  was used. This measure is a modified and broadened version of the sexual harassment instruments used in Swift and Denton (1994) and York (1989). Originally, six items or vignettes were developed; however, factor analysis results revealed that only three of the original six items loaded on the same factor. Hence,  $Per_{SEX}$  was measured with a three-item, seven-point Likert scale, ranging from 1 (clearly no) to 7 (clearly yes). Previous research investigating sexual harassment supports the use of vignettes (Gervasio & Ruckdeschel, 1992; Gowan & Zimmerman, 1996; Hartnett, Robinson & Singh, 1989; Terpstra & Baker, 1989). To operationalize  $Per_{SEX}$ , respondents were asked to read a series of short vignettes pertaining to a fictitious U.S. sales firm and indicate whether the act described in the vignette could be considered sexual harassment. The vignettes (Wason, Polonsky & Hyman, 2002) included sufficient detail about conditions pertaining to sexual harassment, such as gender harassment, sexist hostility, unwanted sexual attention, sexual coercion, verbal remarks, and nonverbal displays (Fitzgerald et al., 1988; Gruber, 1992; Wright & Bean, 1993). Because women are more likely to be sexually harassed at work compared to men (Leap & Gray, 1980; Tangri, Burt & Johnson, 1982), each vignette depicted a female being sexually harassed by a male (see Appendix).

#### *Attitudes toward the Brand*

Defined as the internal assessment of a certain brand (Mitchell & Olson, 1981),  $A_B$  was measured with a four-item, seven-point, semantic differential scale based on the scale used in Grier and Deshpandé (2001), which was meant to assess general attitudes about an advertised brand. Thus,  $A_B$  is operationalized as a *general* overall assessment of a brand (e.g., degree of favorableness). The four items are anchored by the bipolar endpoints: unfavorable/favorable, bad/good, unpleasant/pleasant, and negative/positive.

#### *Brand Image*

Brand image is a multi-faceted construct comprised of multiple brand factors pertaining to an organization and its offerings (Biel, 1992; Keller, 1993; Smith, 2004). A review of the literature on brand image (Davies et al., 2004) inspired a new, nine-item measure. This measure was intended to accurately capture distinct corporate and product image dimensions of brand image (Biel, 1992), and thus, is operationalized as concrete or specific content assessments of a brand (Boush & Jones, 2006; Dobni & Zinkhan, 1990). Factor analysis results revealed that only seven of the original nine items loaded on the same factor. Hence,  $I_{IMAGE}$  was measured with a new, seven-item, seven-point, semantic differential scale, anchored by bipolar endpoints. For the corporate image dimension these endpoints are: not credible/credible (Blackston, 1992), not prestigious/prestigious (Hsieh, 2002), disreputable/reputable (Herbig & Milewicz, 1995), and indecent/virtuous (Hamilton, 2000). For the *product* image

dimension the endpoints are: not trustworthy/trustworthy (Driesener & Romaniuk, 2006), low quality/high quality (Völckner & Sattler, 2006), and unreliable/reliable (Hsieh, 2002).

### *Intentions to Work*

We argue that intentions to work pertain to the likelihood that a person will choose to work for a specific company, which is similar to existing recruitment conceptualizations such as organizational attractiveness as an employer and willingness to work for the union (Gordon et al., 1980; Liden & Parsons, 1986; Turban & Greening, 1996). To offer a sound assessment of intentions to work, and because intentions are a robust proxy of actual behaviors (Ajzen, 1991; Eagly & Chaiken, 1993), we use a five-item, seven-point, semantic differential scale based on the purchase intention scales used in Holmes and Crocker (1987) and Mackenzie, Lutz, and Belch (1986). The five items are anchored by the bipolar endpoints: would not seek out/would seek out, not very likely/very likely, improbable/probable, would not consider/would consider, and unwilling/willing.

### *Pretest*

The factor structure and reliability of the four scales were assessed. Forty-four undergraduate business students attending a large research university in the southwestern U.S. supplied the requisite data during a regularly scheduled class. Principal components analysis with varimax rotation was used to assess factor structure. Missing data were handled via pairwise deletion. The four and five items used to measure  $A_B$  and  $Int_{WORK}$  respectively, loaded appropriately. For the  $I_{IMAGE}$  and  $Per_{SEX}$  scales, cross-loadings among some items were evident. Except for the  $Per_{SEX}$  measure, all scale reliabilities were adequate (i.e.,  $>.70$ ) (Nunnally & Bernstein, 1994). Based on respondents' written feedback about the  $Per_{SEX}$  scale (e.g., this sentence is confusing), changes were made to the wording of some items used in the main study.

### *Procedure for Main Study*

Undergraduate students enrolled in the business college of a large research university located in the southwest U.S. were asked to complete a ten-minute questionnaire during regularly scheduled classes. Students received prior notice about survey administration and were informed that they were participating in a study about sexual harassment in the workplace. The survey administrator reminded students repeatedly that their responses were anonymous. To control for possible previous sexual harassment experiences, it was conveyed to students that their responses were based solely on the information regarding the sales firm described in the survey. Students were debriefed and offered course extra credit for their efforts.

A definition of sexual harassment, based on the Equal Employment Opportunity Commission's interpretation of what constitutes sexual harassment, was provided at the onset of the questionnaire; it read: "Sexual harassment can be defined as any unwanted or unwelcome sexual behavior. Such behaviors include: verbal (e.g., crude sexual comments), non-verbal (e.g., looking someone 'up-and-down'), physical (e.g., brushing up against someone 'accidentally'), and/or visual (e.g., displaying sexual



images).” Subsequently, respondents were asked to indicate (either yes or no), if after graduating from college, the prospect of working in a sales-related field was plausible. Respondents who answered no accounted for 2% of the original sample and were eliminated from the sample. Respondents who answered yes (N=217) were then asked to read a brief description, about a fictitious U.S. sales firm, which included data pertaining to number of employees, product offerings, gross annual revenue, and trading areas. Next, respondents read vignettes about the fictitious sales firm. Each vignette was embedded with potential sexual harassment situations, and respondents were asked to indicate the degree of sexual harassment described in each vignette. Subsequently, responses were given regarding attitudes toward the brand, brand image, and intentions to work for this firm.

### *Sample Profile*

The final sample size of 217 respondents meets the requirements for effective structural equation modeling (Hair et al., 2006). Males (54.7%) outnumber females, and the ethnicities represented are White (81.7%), Hispanic (8.9%), Asian (4.2%), American Indian (2.8%), and Black (2.3%). The mean age of respondents is 22.18 (SD = 4.39); the majority is either junior (54.7%) or senior (35.8%) level college students, and 74.9% of the sample are employed.

## **Results**

Independent sample t-test results for males (n=115) and females (n=95) and each of the three items used to measure ( $Per_{SEX}$ ) revealed non-significant mean differences at the  $P < 0.05$ ; that is, for *vignette 1*,  $M_{Males} = 3.46$ ,  $M_{Females} = 3.92$ ,  $t(208) = -1.75$ ,  $P = 0.08$ , for *vignette 2*,  $M_{Males} = 5.19$ ,  $M_{Females} = 5.13$ ,  $t(208) = 0.23$ ,  $P = 0.81$ , and for *vignette 3*,  $M_{Males} = 3.38$ ,  $M_{Females} = 3.71$ ,  $t(208) = -1.32$ ,  $P = 0.18$ . Because men and women responded similarly to perceived sexual harassment in the workplace (Hartnett, Robinson & Singh, 1989; York, 1989), data pooling was justified for the analyses.

### *Factor Structure*

Principal component analysis with varimax rotation was used to confirm the structure of the 19 items that comprised the four scales. Missing data was handled via pairwise deletion. The resulting four factor solution, in which each item loaded on the appropriate factor, accounted for 67.24% of the variance. Reliabilities for the four scales ranged from  $\alpha = .731-.928$ , which exceed the .70 threshold for preliminary research (Nunnally & Bernstein, 1994). Coefficient alphas and factor loadings are provided in Table 1. High factor loadings and alphas are desirable because they provide evidence for reliability and convergent validity (i.e., items attempting to measure the same construct are highly correlated). The lack of significant cross-loadings (see Table 1) provides evidence of discriminant validity.

**Table 1: Factor Loadings and Reliabilities**

Construct Alpha	Scale Items (AB, INTWORK, and IMAGE are scaled on seven-point semantic differential scales; PERSEX is scaled on a seven-point Likert scale (1-Clearly No to 7-Clearly Yes))	Factor Loading	Factor Loading	Factor Loading	Factor Loading
AB .854	(AB1) Unfavorable / Favorable	.264	.191	<b>.768</b>	.102
	(AB2) Bad / Good	.253	.267	<b>.737</b>	.099
	(AB3) Unpleasant / Pleasant	.263	.215	<b>.781</b>	.109
	(AB4) Negative / Positive	.221	.228	<b>.703</b>	.251
Intwork .928	(INTWORK1) Would not seek out / Would seek out	<b>.772</b>	.238	.165	.240
	(INTWORK2) Not very likely / Very likely	<b>.777</b>	.285	.315	.128
	(INTWORK3) Improbable / Probable	<b>.787</b>	.300	.235	.073
	(INTWORK4) Would not consider / Would consider	<b>.816</b>	.239	.226	.118
	(INTWORK5) Unwilling / Willing	<b>.794</b>	.219	.317	.127
IMAGE .868	(IMAGE1) Not credible / Credible	.283	<b>.594</b>	.129	.236
	(IMAGE2) Not prestigious / Prestigious	.201	<b>.733</b>	.120	.014
	(IMAGE3) Disreputable / Reputable	.265	<b>.708</b>	.251	.022
	(IMAGE4) Not trustworthy / Trustworthy	.366	<b>.630</b>	.155	.126
	(IMAGE5) Low quality / High quality	.033	<b>.781</b>	.217	.056
	(IMAGE6) Unreliable / Reliable	.169	<b>.749</b>	.139	.160
	(IMAGE7) Indecent / Virtuous	.381	<b>.546</b>	.251	.088
PerSEX .731	(PERSEX1) Is this a case of sexual harassment?	.182	.023	.110	<b>.841</b>
	(PERSEX2) Is this a case of sexual harassment?	.192	.053	.225	<b>.761</b>
	(PERSEX3) Is this a case of sexual harassment?	.028	.177	.057	<b>.724</b>

A measurement model was estimated with LISREL 8.50 and the 19 items comprising the four scales. The average variance extracted (AVE) values for each construct, except  $I_{\text{IMAGE}}$  (AVE=49.30%), exceed .50, which provides additional evidence of convergent validity. Also, the AVE values for each construct are greater than the squared correlations between each construct and the other constructs, except one correlation and the  $I_{\text{IMAGE}}$  value (see Phi and Phi<sup>2</sup> matrices in Table 2), which offers further evidence of discriminant validity (Fornell & Larcker, 1981; Hair et al., 2006). Estimation of the measurement model produced the following goodness-of-fit statistics:  $\chi^2$  (146)=319.61 (P<.00), comparative fit index (CFI)=.93, non-normed fit index (NNFI)=.92, goodness of fit index (GFI)=.87, and standardized root mean square residual (SRMR)=.049. In general, these fit statistics provide evidence of adequate model fit and the measures used to examine the studied constructs appear valid (Hair et al., 2006).

### Structural Equation Model

The relationships shown in Figure 1 were tested using a structural equation model with LISREL 8.50. A covariance matrix and maximum likelihood estimation were used to estimate model parameters. Missing data were handled via pairwise deletion. The four constructs—*perceived sexual harassment*, *attitudes toward the brand*, *brand image*, and *intentions to work*—with three, four, seven, and five items, respectively, were included in the model. One additional parameter, beyond that explained by the common factor, is included in the model.

Model estimation produced the following goodness-of-fit statistics:  $\chi^2$ (145)=270.26 (P<.00), CFI=.95, NNFI=.94, GFI=.88, and SRMR=.049. The ratio of  $\chi^2$  per degree of freedom is less than two, which indicates an acceptable fit of the model to the data (Hair et al., 2006). The CFI, NNFI, and SRMR statistics also indicate

a good model fit, but the GFI statistic implies only a marginal fit between the model and the data (Hair et al., 2006; Hu & Bentler, 1999). Therefore, overall model fit is interpreted as acceptable, and the model cannot be rejected based on these data.

**Table 2: Confirmatory Factor Analysis**

Constructs	PERSEX	AB	INTWORK	IMAGE	Item Reliabilities	Delta ( $\delta$ )
PERSEX1	.80				.640	.360
PERSEX2	.79				.624	.376
PERSEX3	.51				.260	.740
AB1		.77			.593	.407
AB2		.77			.593	.407
AB3		.80			.640	.360
AB4		.74			.548	.452
INTWORK1			.81		.656	.344
INTWORK2			.88		.774	.226
INTWORK3			.85		.723	.277
INTWORK4			.86		.740	.260
INTWORK5			.87		.757	.243
IMAGE1				.65	.423	.577
IMAGE2				.68	.462	.538
IMAGE3				.75	.563	.437
IMAGE4				.73	.533	.467
IMAGE5				.68	.462	.538
IMAGE6				.72	.518	.482
IMAGE7				.70	.490	.510
Average Variance Extracted	50.80%	59.35%	73.00%	49.30%		
Phi ( $\Phi$ ) Matrix						
PERSEX	1.00					
AB	-.46	1.00				
INTWORK	-.44	.71	1.00			
IMAGE	-.34	.66	.70	1.00		
Phi ( $\Phi$ ) <sup>2</sup> Matrix						
PERSEX	1.00					
AB	.21	1.00				
INTWORK	.19	.50	1.00			
IMAGE	.12	.44	.49	1.00		

The structural equation model's path coefficients are used to evaluate the hypotheses. The t statistic associated with all six path coefficients is significant at the  $p < .05$  level or better, which implies that the hypotheses cannot be rejected (see Table 3). Specifically, prospective employees' perceived sexual harassment in a sales workplace was negatively related to their: attitudes toward the brand (H1), brand image (H2), and intentions to work for the firm (H3); prospective employees' brand image of a sales firm was positively related to both their attitudes toward the brand (H4), and intentions to work for this firm (H6); also, prospective employees' attitudes toward the service firm's brand were positively related to their intentions to work for

this firm (H5). Overall, the data support all six hypotheses and the structural equation model.

**Table 3:** *Hypotheses Tests*

Hypothesis	Structural Coefficient	t -statistic
H1: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their attitudes toward the brand.	-.27	-3.55*
H2: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their brand image of this firm.	-.34	-3.91*
H3: For prospective employees in a sales context, perceived sexual harassment in the workplace is negatively related to their intentions to work for this firm.	-.13	-2.03**
H4: Prospective employees' brand image of a sales firm is positively related to their attitudes toward the brand.	.57	6.49*
H5: Prospective employees' attitudes toward a sales firm's brand are positively related to their intentions to work for this firm.	.37	4.19*
H6: Prospective employee's brand image of a sales firm is positively related to their intentions to work for this firm.	.42	4.81*

\*Significant at the  $P < .01$  level, \*\*Significant at the  $P < .05$  level

## Discussion

The severity of workplace sexual harassment may range from victims experiencing heightened levels of stress and uneasiness to firms incurring substantial financial litigation (Foy, 2000; Gervasio & Ruckdeschel, 1992; Schneider Swan & Fitzgerald, 1997). All factors in this range are detrimental to brand integrity, workplace efficiencies, and stakeholder value. Although the existing research on sexual harassment offers valuable insight into its consequences, additional research is needed to extend our understanding of attitudinal and behavioral outcomes of workplace sexual harassment (Gilbert, 2005). Thus, we examine, for the first time, the effects of perceived sexual harassment, in a sales-oriented workplace, on attitudes toward the brand, brand image, and intentions to work, of prospective job applicants. The significant and direct negative effects found for each of these relationships, offers preliminary evidence that sexual harassment undermines brand development and recruiting. Hence, firms seeking to grow their brand's value and hire effective employees must be equipped to prevent and solve their sexual harassment quandaries.

The contribution of our study to knowledge of sexual harassment is three-fold. First, our model and data support the notion that perceived sexual harassment in the workplace has a direct negative effect on distinct attitudinal (i.e.,  $A_B$  and  $I_{IMAGE}$ ) and behavioral (i.e.,  $Int_{WORK}$ ) brand-response constructs. By testing these effects, our findings meaningfully extend knowledge regarding the consequences to brands, caused by workplace sexual harassment (Gilbert, 2005; O'Connell & Korabik, 2000; Terpstra & Baker, 1986). Second, our model shows direct positive effects for both attitudes toward the brand and brand image on intentions to work. These findings

suggest that attitudinal responses to brands have the capacity to influence individuals' willingness to work for such companies. Third, we developed a scale that measures brand image and extend previously developed instruments used to measure perceived sexual harassment in the workplace.

Data collected when using these scales are reliable and, based on a factor analysis, evidence of convergent and discriminant validity exists for these measures. By using a new multi-item sexual harassment scale, our study provides a better understanding of individuals' perceptions of what constitutes sexual harassment (Arvey & Cavanaugh, 1995). As a result, employers may be better equipped to develop effective policy statements that dissuade workplace sexual harassment (York, 1989).

It is a well-documented fact that sexual harassment can lead to detrimental consequences for organizations (e.g., Wright & Bean, 1993). Our brand-specific results, which further corroborate these assumptions, show that  $Per_{SEX}$  negatively influences  $A_B$ ,  $I_{IMAGE}$ , and  $Int_{WORK}$  of potential employees. Thus, to help preserve brand status in the marketplace and to effectively recruit and retain employees, organizations should stress that any form of sexual harassment can lead to workplace inefficiencies, suboptimal financial results, reduced stakeholder wealth, and litigation, to name a few. Additionally, our findings indicate that both  $A_B$  and  $I_{IMAGE}$  correlate positively with  $Int_{WORK}$ . Thus, businesses seeking to acquire capable employees, should take the necessary steps to protect and ameliorate their brand's status in these individual's minds. By doing so, firms' recruiting and retention capacity should be strengthened (Sullivan, 2003). Lastly, our study finds a positive correlation between  $I_{IMAGE}$  and  $A_B$ . Because more favorable customer attitudes lead to more favorable intentions and behaviors toward brands (e.g., Pope & Voges, 2000), organizations should engage in strategies that positively affect its image.

### *Implications*

The way consumers respond to brands can have a significant effect on both business success and stakeholder value (Keller, 1993). Understanding factors that negatively affect brands should help businesses develop strategies to minimize such adverse consequences. Our study shows that perceived sexual harassment in the workplace has a negative effect on attitudes toward the brand, brand image, and intentions to work for prospective employees. Thus, an obvious implication of our findings is that businesses should engage in tactics that minimize sexual harassment behaviors within their organization. For example, a preventative, rather than a reactive (Woodford & Risetto, 2004; Wright & Bean, 1993) approach to developing and executing sexual harassment policies, may reduce the likelihood of sexual harassment in the workplace, which should lead to more favorable attitudinal and behavioral responses toward the company and brand.

Our results show positive effects between attitudes toward the brand and brand image on intentions to work, and a positive effect between brand image and attitudes toward the brand. These findings imply that as attitudinal responses toward brands become more favorable, individuals are more likely to work for such organizations. In this sense, protecting and building an organization's image through strategies aimed at alleviating workplace sexual harassment, may lead to the attraction and retention of

qualified employees (Sullivan, 2003).

Educators play a pertinent role in developing competent employees. By incorporating sexual harassment content (e.g., its domain, its effects on individuals, organizations, and society) into course curriculum, instructors can help mold business students into effective managerial prospects (Swift & Denton, 1994).

Just as customers should be encouraged to provide feedback about their service experience and employee performance (Bitner et al., 1997), employees should be encouraged to discuss, in confidence, personal, ambient, and third-party sexual harassment as witnessed and/or experienced within their organization. Reflections on such experiences may help firms to avoid liability under *respondent superior*, which means that the employer knew or should have known of the sexual harassment and took no effective remedial action (Bundy v. Jackson, 1981). Such reflections may also help firms better understand what constitutes sexual harassment and its frequency within business. As a result, organizations may be in a better position to develop effective policy statements, which should discourage sexual harassment behaviors.

Creating awareness, avoiding litigation, protecting the brand, and generating workplace efficiencies are goals of a sexual harassment policy; its effectiveness is determined by employee knowledge and understanding of its principles and procedures (Stokes, Stewart-Belle & Barnes, 2000). An organization's culture is pivotal in developing an anti-sexual harassment environment. For example, a firm's zero tolerance culture should offer training in harassment prevention, take sexual harassment complaints seriously, carry out fair investigations when complaints arise, and appropriately punish offenders (Cava, 2001; Johnson, 2004; Stokes et al., 2000). Through effective leadership (Vallaster & de Chernatony, 2006), these policies must be communicated overtly (Creyer & Ross, 1997). For example, firms could offer on-line training sessions, in-house seminars, and distribute memos summarizing recent court decisions, where termination of employees and financial damages were recovered by plaintiffs (Acken, St. Pierre & Veglahn, 1991; Collins & Blodgett, 1981; Dunne & Lusch, 2008). By not disseminating to employees information pertaining to a sexual harassment policy, liability on behalf of the employer may result (e.g., Faragher v. City of Boca Raton, 1998). As harassment knowledge proliferates via overt communication, a zero tolerance mindset will become an integral part of a firm's culture, which should minimize workplace sexual harassment and increase overall brand value.

#### *Limitations and Future Research Directions*

Our study is not without limitations. First, data was collected using college students at one university location in the southwest U.S. Additional data from actual job or experienced job seekers across different regions and cultures would be needed to establish the external validity of our findings (Winer, 1999). Second, the four scales we used for data collection may not be equally valid across all samples and exchange settings. This factor can affect the measurement properties of the constructs and their relationships with one another. Third, although the *perceived sexual harassment* and *brand image* scales we used demonstrated convergent and discriminant validity, mono-method bias (Cook & Campbell, 1979) may be evident based on our method of data

collection. Thus, additional quantitative and qualitative research is necessary to further validate these scales.

To broaden the scope of this study, other measures relevant to sexual harassment research, such as affect, coping, locus of control, moral intensity, self-blame, sex-role power, verbal sexual harassment, the Corporate Character Scale, and the Sexual Experiences Questionnaire (Bowes-Sperry & Powell, 1999; Davies et al., 2004; Fitzgerald et al., 1988; Gervasio & Ruckdeschel, 1992; Gutek & Morasch, 1982; Jensen & Gutek, 1982; Malamut & Offermann, 2001; Popovich & Licata, 1987), could be added to our model. Additional research tools, such as interpretive, canonical, and/or cluster analyses, could be used to examine individual and organizational effects of perceived sexual harassment in the workplace (Cortina & Wasti, 2005; Dan, Pinesof & Riggs, 1995; O'Connell & Korabik, 2000). Longitudinal studies could be implemented to examine the lasting effect of perceived sexual harassment in the workplace on brand-response constructs (Lengnick-Hall, 1995). Moods can affect consumer decision-making (Bakamitsos & Siomkos, 2004), and examining how moods such as anger or depression moderate the relationships in our model would add insight into attitudinal and behavioral responses toward workplace sexual harassment (Terpstra & Baker, 1986).

Moreover, alternative vignettes that measure perceived sexual harassment could be developed. For example, vignettes could include additional information regarding ethical dimensions (Bowes-Sperry & Powell, 1999), personal sexual harassment (Langhout et al., 2005), sex-role power (Gutek & Morasch, 1982), third party sexual harassment (Aalberts & Seidman, 1994; Fine et al., 1994), sexual hostility (Fitzgerald et al., 1988), different forms of harassment communication such as email or text messaging, and/or depict a male being sexually harassed. Lastly, factors related to corporate reputation pertain to more than financial performance measures. Additional research examining the effect of non-economic factors (e.g., corporate social performance, familiarity, personality) on brand and corporate image is needed (Backhaus & Tikoo, 2004; Fombrun & Shanley, 1990; Gatewood, Gowan & Lautenschlager, 1993; Turban & Greening, 1996). For example, because an employer's reputation matters to employees (Earle, 2003), prospective employees may be willing to work for an organization only if the image of the firm is consistent with their personality (Yi & La, 2006).

## Conclusion

Although the image of a firm influences prospective employees' interest in pursuing employment with the firm and workplace sexual harassment influences brand image perceptions of the firm, there remains a paucity of research simultaneously examining sexual harassment effects (Gilbert, 2005; O'Connell & Korabik, 2000), recruitment processes (Breaugh & Starke, 2000; Ryan & Tippins, 2004), and brand image influences and effects (Backhaus & Tikoo, 2004; Gatewood et al., 2001). To help fill this research void, our study examined if a firm's reputation of sexual harassment influences prospective employees' brand image perception of the firm and their willingness to seek employment with the firm.

Our findings show that perceived sexual harassment in the workplace negatively affects attitudinal and behavioral brand-response constructs. Thus, an overt implication of our findings is that businesses should engage in tactics (e.g., internal brand-building strategies) that minimize sexual harassment behaviors within their organization. For example, when internal stakeholders understand, embrace, and execute organizational brand values (e.g., a zero tolerance policy regarding sexual harassment), the company has an opportunity to gain a competitive advantage in the marketplace and the brand has an opportunity to flourish (Gapp & Merrilees, 2006). In this sense, internal brand strategies are critical for overall business success. When communicated effectively (e.g., during the interview process, recruiter behavior, recruitment strategies) (Liden & Parsons, 1986; Rynes & Barber, 1990; Rynes & Miller, 1983), internal branding may be used as a tool to attract qualified applicants by assuring prospective employees that the organization is a desirable place to work (Backhaus & Tikoo, 2004), thereby leading to a sustainable competitive advantage (Kickul, 2001).

Our results also show positive effects between attitudes toward the brand and brand image on intentions to work. They also show a positive effect between brand image and attitudes toward the brand. Both factors indicate a positive relationship between attitudinal responses toward brands and the likelihood that prospective employees will seek employment with the firm. Thus, fostering an organization's image through internal brand strategies aimed at alleviating workplace sexual harassment, may lead to the attraction and retention of qualified employees (Sullivan, 2003).

## References

- Aalberts, R.J. & Seidman, L.H. (1994). Sexual harassment by clients, customers and suppliers: How employers should handle an emerging legal problem. *Employee Relations Law Journal*, 20: 85-100.
- Acken, B.T., St. Pierre, K., & Veglahn, P. (1991). Limiting sexual harassment liability: The best defense is an effective policy against harassment. *Journal of Accountancy*, 171: 42-47.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50: 179-211.
- Arvey, R.D. & Cavanaugh, M.A. (1995). Using surveys to assess the prevalence of sexual harassment: Some methodological problems. *Journal of Social Issues*, 51: 39-52.
- Backhaus, K. & Tikoo, S. (2004). Conceptualizing and researching employer branding. *Career Development International*, 9: 501-517.
- Bakamitsos, G.A. & Siomkos, G.J. (2004). Context effects in marketing practice: The case of mood. *Journal of Consumer Behaviour*, 3: 303-314.
- Batra, R. & Homer, P.M. (2004). The situational impact of brand image beliefs. *Journal of Consumer Psychology*, 14: 318-330.
- Belt, J.A. & Paolillo, J.G.P. (1982). The influence of corporate image and specificity on candidate qualifications on response to recruitment advertisement. *Journal of Management*, 8: 105-112.



- Bennett-Alexander, D.D. & Pincus, L.B. (1994). *Employment law for business*. Chicago, IL: McGraw-Hill.
- Biel, A.L. (1992). How brand image drives brand equity. *Journal of Advertising Research*, 32: RC6-RC12.
- Bitner, M.J., Faranda, W.T., Hubbert, A.R., & Zeithaml, V.A. (1997). Customer contributions and roles in service delivery. *International Journal of Service Industry Management*, 8: 193-205.
- Blackston, M. (1992). Observations: Building brand equity by managing the brand's relationships. *Journal of Advertising Research*, 32: 79-83.
- Boush, D.M. & Jones, S.M. (2006). A strategy-based framework for extending brand image research. In Kahle, L.R. & Kim, C.H. (Eds.), *Creating images and the psychology of marketing communication*. Mahwah, NJ: Lawrence Erlbaum Associates: 3-29.
- Bowes-Sperry, L. & Powell, G.N. (1999). Observers' reactions to social-sexual behavior at work: An ethical decision making perspective. *Journal of Management*, 25: 779-802.
- Breugh, J.A. & Starke, M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management*, 26: 405-434.
- Brown, S.P. & Stayman, D.M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research*, 19: 34-51.
- Bundy v. Jackson (1981). 641 F.2d 934 (D.C. Cir.)
- Cava, A. (2001). Sexual harassment claims: New framework for employers. *Business & Economic Review*, 47: 13-16.
- Chun, R. & Davies, G. (2006). The influence of corporate character on customers and employees: Exploring similarities and differences. *Journal of the Academy of Marketing Science*, 34: 138-146.
- Chung-Herrera, B.G., Goldschmidt, N., & Hoffman, K.D. (2004). Customer and employee views of critical service incidents. *Journal of Services Marketing*, 18: 241-254.
- Civil Rights Act (1991). 42 U.S.C. § 1981A.
- Collins, E.G.C. & Blodgett, T.B. (1981). Sexual harassment...some see it...some won't. *Harvard Business Review*, 59: 76-94.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cortina, L.M., Fitzgerald, L.F. & Drasgow, F. (2002). Contextualizing Latina experiences of sexual harassment: Preliminary tests of a structural model. *Basic and Applied Social Psychology*, 24: 295-311.
- Cortina, L.M. & Wasti, S.A. (2005). Profiles in coping: Responses to sexual harassment across persons, organizations, and cultures. *Journal of Applied Psychology*, 90: 182-192.
- Creyer, E.H. & Ross Jr., W.T. (1997). The influence of firm behavior on purchase intention: Do consumers really care about business ethics? *Journal of Consumer Marketing*, 14: 419-432.
- Dan, A.J., Pinsof, D.A. & Riggs, L.L. (1995). Sexual harassment as an occupational hazard in nursing. *Basic and Applied Social Psychology*, 17: 563-580.

- Davies, G., Chun, R., da Silva, R.V. & Roper, S. (2004). A corporate character scale to assess employee and customer views of organization reputation. *Corporate Reputation Review*, 7: 125-146.
- Dobni, D. & Zinkhan, G.M. (1990). In search of brand image: A foundation analysis. In Goldberg, M.E., Gorn, G., & Pollay, R. (Eds.), *Advances in Consumer Research, Provo, UT: Association for Consumer Research*, 17: 110-119.
- Driesener, C. & Romaniuk, J. (2006). Comparing methods of brand image measurement. *International Journal of Market Research*, 48: 681-697.
- Dunne, P.M. & Lusch, R.F. (2008). *Retailing*. Mason, OH: Thomson South-Western.
- Eagly, A.H. & Chaiken, S. (1993). *The psychology of attitudes*. New York, NY: Harcourt Brace College Publishers.
- Earle, H.A. (2003). Building a workplace of choice: Using the work environment to attract and retain top talent. *Journal of Facilities Management*, 2: 244-257.
- Equal Employment Opportunity Commission (1980). Guidelines on discrimination because of sex. 29 CFR Part 1604, *Federal Register*, November 10: 45.
- Faragher v. City of Boca Raton (1998). 118 S.Ct. 2275.
- Ferrand, A. & Pages, M. (1999). Image management in sport organisations: The creation of value. *European Journal of Marketing*, 33: 387-401.
- Fine, L.M., Shepherd, C.D. & Josephs, S.L. (1994). Sexual harassment in the sales force: The customer is NOT always right. *Journal of Personal Selling & Sales Management*, 14: 15-30.
- Fitzgerald, L.F., Drasgow, F., Hulin, C.L., Gelfand, M.J., & Magley, V.J. (1997). Antecedents and consequences of sexual harassment in organizations: A test of an integrated model. *Journal of Applied Psychology*, 82: 578-589.
- Fitzgerald, L.F., Shullman, S.L., Bailey, N., Richards, M., Swecker, J., Gold, Y., Ormerod, M. & Weitzman, L. (1988). The incidence and dimensions of sexual harassment in academia and the workplace. *Journal of Vocational Behavior*, 32: 152-175.
- Fombrun, C. & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, 33: 233-258.
- Fornell, C. & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18: 39-50.
- Foy, N.F. (2000). Sexual harassment can threaten your bottom line. *Strategic Finance*, 82: 56-61.
- Gapp, R. & Merrilees, B. (2006). Important factors to consider when using internal branding as a management strategy: A healthcare case study. *Journal of Brand Management*, 14: 162-176.
- Gatewood, R.D., Gowan, M.A. & Lautenschlager, G.J. (1993). Corporate image, recruitment image, and initial job choice decisions. *Academy of Management Journal*, 36: 414-427.
- Gelfand, M.J., Fitzgerald, L.F. & Drasgow, F. (1995). The structure of sexual harassment: A confirmatory analysis across cultures and settings. *Journal of Vocational Behavior*, 47: 164-177.
- Gervasio, A.H. & Ruckdeschel, K. (1992). College students' judgments of verbal

- sexual harassment. *Journal of Applied Social Psychology*, 22: 190-211.
- Gilbert, J.A. (2005). Sexual harassment and demographic diversity: Implications for organizational punishment. *Public Personnel Management*, 34: 161-174.
- Glomb, T.M., Richman, W.L., Hulin, C.L., Drasgow, F., Schneider, K.T. & Fitzgerald, L.F. (1997). Ambient sexual harassment: An integrated model of antecedents and consequences. *Organizational Behavior and Human Decision Processes*, 71: 309-328.
- Goldsmith, R.E., Lafferty, B.A. & Newell, S.J. (2000). The impact of corporate credibility and celebrity credibility on consumer reaction to advertisements and brands. *Journal of Advertising*, 29: 43-54.
- Gordon, M.E., Philpot, J.W., Burt, R.E., Thompson, C.A. & Spiller, W.E. (1980). Commitment to the union: Development of a measure and an examination of its correlates. *Journal of Applied Psychology Monograph*, 65: 479-499.
- Gotsi, M. & Wilson, A. (2001). Corporate reputation management: 'Living the brand.' *Management Decision*, 39: 99-104.
- Gowan, M.A. & Zimmerman, R.A. (1996). Impact of ethnicity, gender, and previous experience on juror judgments in sexual harassment cases. *Journal of Applied Social Psychology*, 26: 596-617.
- Grier, S.A. & Deshpandé, R. (2001). Social dimensions of consumer distinctiveness: The influence of social status of group identity and advertising persuasion. *Journal of Marketing Research*, 38: 216-224.
- Gruber, J.E. (1992). A typology of personal and environmental sexual harassment: Research and policy implications for the 1990s. *Sex Roles*, 26: 447-464.
- Guterk, B.A. & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harassment of women at work. *Journal of Social Issues*, 38: 55-74.
- Hair J.F Jr., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hamilton, K. (2000). Project galore: Qualitative research and leveraging Scotland's brand equity. *Journal of Advertising Research*, 40: 107-111.
- Harris v. Forklift Systems, Inc. (1993). 114 S.Ct. 367.
- Hartnett, J.J., Robinson, D. & Singh, B. (1989). Perceptions of males and females toward sexual harassment and acquiescence. *Journal of Social Behavior and Personality*, 4: 291-298.
- Herbig, P. & Milewicz, J. (1995). The relationship of reputation and credibility to brand success. *Journal of Consumer Marketing*, 12: 5-10.
- Holmes, J.H. & Crocker, K.E. (1987). Predispositions and the comparative effectiveness of rational, emotional and discrepant appeals for both high involvement and low involvement products. *Journal of the Academy of Marketing Science*, 15: 27-35.
- Hsieh, M.H. (2002). Identifying brand image dimensionality and measuring the degree of brand globalization: A cross-national study. *Journal of International Marketing*, 10: 46-67.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6: 1-55.
- Jensen, I.W. & Guterk, B.A. (1982). Attributions and assignment of responsibility in

- sexual harassment. *Journal of Social Issues*, 38: 121-136.
- Johnson, M.W. (2004). Harassment and discrimination prevention training: What the law requires. *Labor Law Journal*, 55: 119-129.
- Keller, K.L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, 57: 1-22.
- Kickul, J. (2001). Promises made, promises broken: An exploration of employee attraction and retention practices in small business. *Journal of Small Business Management*, 39: 320-335.
- Lafferty, B.A., Goldsmith, R.E. & Newell, S.J. (2002). The dual credibility model: The influence of corporate and endorser credibility on attitudes and purchase intentions. *Journal of Marketing Theory and Practice*, 10: 1-12.
- Langhout, R.D., Bergman, M.E., Cortina, L.M., Fitzgerald, L.F., Drasgow, F. & Williams, J.H. (2005). Sexual harassment severity: Assessing situational and personal determinants and outcomes. *Journal of Applied Social Psychology*, 35: 975-1007.
- Leap, T.L. & Gray, E.R. (1980). Corporate responsibility in cases of sexual harassment. *Business Horizons*, 23: 58-65.
- Lengnick-Hall, M.L. (1995). Sexual harassment research: A methodological critique. *Personnel Psychology*, 48: 841-864.
- Liden, R.C. & Parsons, C.K. (1986). A field study of job applicant interview perceptions, alternative opportunities, and demographic characteristics. *Personnel Psychology*, 39: 109-122.
- Mackenzie, S.B. & Lutz, R.J. (1989). An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context. *Journal of Marketing*, 53: 48-65.
- Mackenzie, S.B., Lutz, R.J. & Belch, G.E. (1986). The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. *Journal of Marketing Research*, 23: 130-143.
- Malamut, A.B. & Offermann, L.R. (2001). Coping with sexual harassment: Personal, environmental, and cognitive determinants. *Journal of Applied Psychology*, 86: 1152-1166.
- Meritor Savings Bank v. Vinson (1986). 477 US 57.
- Mitchell, A.A. & Olson, J.C. (1981). Are product beliefs the only mediator of advertising effect on brand attitude? *Journal of Marketing Research*, 18: 318-332.
- Mohr, L.A. & Bitner, M.J. (1991). Mutual understanding between customers and employees in service encounters. In Holman, R.H. & Solomon, M.R. (Eds.), *Advances in Consumer Research*, Provo, UT: Association for Consumer Research, 18: 611-617.
- Morrow, P.C., McElroy, J.C. & Phillips, C.M. (1994). Sexual harassment behaviors and work related perceptions and attitudes. *Journal of Vocational Behavior*, 45: 295-309.
- Moylan v. Maries County, (1986). 792 F2d 746 (8th Cir.)
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- O'Connell, C.E. & Korabik, K. (2000). Sexual harassment: The relationship of personal vulnerability, work context, perpetrator status, and type of harassment to outcomes. *Journal of Vocational Behavior*, 56: 299-329.

- O'Leary-Kelly, A.M., Paetzold, R.L. & Griffin, R.W. (2000). Sexual harassment as aggressive behavior: An actor-based perspective. *Academy of Management Review*, 25: 372-388.
- Pope, N.K.U. & Voges, K.E. (2000). The impact of sport sponsorship activities, corporate image, and prior use on consumer purchase intention. *Sport Marketing Quarterly*, 9: 96-102.
- Popovich, P.M. & Licata, B.J. (1987). A role model approach to sexual harassment. *Journal of Management*, 13: 149-161.
- Priester, J.R., Nayakankuppam, D., Fleming, M.A. & Godek, J. (2004). The A2SC2 model: The influence of attitudes and attitudes strength on consideration and choice. *Journal of Consumer Research*, 30: 574-587.
- Raver, J.L. & Gelfand, M.J. (2005). Beyond the individual victim: Linking sexual harassment, team processes, and team performance. *Academy of Management Journal*, 48: 387-400.
- Robinson, R.K., Fink, R.L. & Lane, P.L. (1994). Developing an expert system as a sexual harassment investigative tool. *Human Resource Planning*, 17: 75-86.
- Robinson, R.K., Jackson, W.T., Franklin, G.M. & Hensley, D. (1998). U.S. sexual harassment law: Implications for small businesses. *Journal of Small Business Management*, 36: 1-12.
- Ryan, A.M. & Tippins, N.T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management*, 43: 305-318.
- Rynes, S.L. & Barber, A.E. (1990). Applicant attraction strategies: An organizational perspective. *Academy of Management Review*, 15: 286-310.
- Rynes, S.L. & Miller, H.E. (1983). Recruiter and job influences on candidates for employment. *Journal of Applied Psychology*, 68: 147-154.
- Schneider, K.T., Swan, S. & Fitzgerald, L.F. (1997). Job-related and psychological effects of sexual harassment in the workplace: Empirical evidence from two organizations. *Journal of Applied Psychology*, 82: 401-415.
- Smith, G. (2004). Brand image transfer through sponsorship: A consumer learning perspective. *Journal of Marketing Management*, 20: 457-474.
- Stedham, Y. & Mitchell, M.C. (1998). Sexual harassment in casinos: Effects on employee attitudes and behaviors. *Journal of Gambling Studies*, 14: 381-400.
- Stokes, P.P., Stewart-Belle, S. & Barnes, J.M. (2000). The Supreme Court holds class on sexual harassment: How to avoid a failing grade. *Employee Responsibilities and Rights Journal*, 12: 79-91.
- Stringer, D.M., Remick, H., Salisbury, J. & Ginorio, A.B. (1990). The power and reasons behind sexual harassment: An employer's guide to solutions. *Public Personnel Management*, 19: 43-52.
- Suh, J.C. & Yi, Y. (2006). When brand attitudes affect the customer satisfaction-loyalty relation: The moderating role of product involvement. *Journal of Consumer Psychology*, 16: 145-155.
- Sullivan, J. (2003). Measuring employment brand. *Strategic HR Review*, 2: 7.
- Swift, C.O. & Denton, L.T. (1994). Teaching about sexual harassment issues: A sales management approach. *Marketing Education Review*, 4: 34-40.
- Tangri, S.S., Burt, M.R. & Johnson, L.B. (1982). Sexual harassment at work: Three

- explanatory models. *Journal of Social Issues*, 38: 33-54.
- Terpstra, D.E. (1996). The effects of diversity on sexual harassment: Some recommendations for research. *Employee Responsibilities and Rights Journal*, 9: 303-313.
- Terpstra, D.E. & Baker, D.D. (1986). A framework for the study of sexual harassment. *Basic and Applied Social Psychology*, 7: 17-34.
- Terpstra, D.E. & Baker, D.D. (1989). The identification and classification of reactions to sexual harassment. *Journal of Organizational Behavior*, 10: 1-14.
- Turban, D.B. & Greening, D.W. (1996). Corporate social performance and organizational attractiveness to prospective employees. *Academy of Management Journal*, 40: 658-672.
- Vallaster, C. & de Chernatony, L. (2006). Internal brand building and structuration: The role of leadership. *European Journal of Marketing*, 40: 761-784.
- Völckner, F. & Sattler, H. (2006). Drivers of brand extension success. *Journal of Marketing*, 70: 18-34.
- Wason, K.D., Polonsky, M.J. & Hyman, M.R. (2002). Designing vignette studies in marketing. *Australasian Marketing Journal*, 10: 41-58.
- Whittler, T.E. (1989). Viewers' processing of actor's race and message claims in advertising stimuli. *Psychology & Marketing*, 6: 287-309.
- Willness, C.R., Steel, P. & Lee, K. (2007). A meta-analysis of the antecedents and consequences of workplace sexual harassment. *Personnel Psychology*, 60: 127-162.
- Winer, R.S. (1999). Experimentation in the 21st century: The importance of external validity. *Journal of the Academy of Marketing Science*, 27: 349-358.
- Winters, L.C. (1986). The effect of brand advertising on company image: Implications for corporate advertising. *Journal of Advertising Research*, 26: 54-59.
- Woodford, K.C. & Risetto, H.A. (2004). The best advice for sexual harassment victims: Don't complain, just quit and sue – take two. *Labor Law Journal*, 55: 213-221.
- Wright, P.C. & Bean, S.A. (1993). Sexual harassment: An issue of employee effectiveness. *Journal of Managerial Psychology*, 8: 30-36.
- Yi, Y. & La, S. (2006). Well-matched employees make customers happy: Effects of brand-employee congruence. In Kahle, L.R. & Kim, C.H. (Eds.), *Creating images and the psychology of marketing communication*. Mahwah, NJ: Lawrence Erlbaum Associates: 223-243.
- York, K.M. (1989). Defining sexual harassment in workplaces: A policy-capturing approach. *Academy of Management Journal*, 32: 830-850.

## Appendix

### Perceived Sexual Harassment Scale Items

Vignette 1 (PERSEX1):

On the inside of his office door, Mike, a sales manager, has a “revealing” women’s calendar. The women in the calendar are beautiful and are wearing “skimpy” bikinis. Each time a worker enters his office they can’t help but notice this calendar.

Is this a case of sexual harassment?

Clearly No

Clearly Yes

1      2      3      4      5      6      7

Vignette 2 (PERSEX2):

John is known to view exotic websites of women while at work. For “fun,” he will call over male colleagues to get their opinion of the website. Because of this activity, the women employees are beginning to feel awkward at work.

Is this a case of sexual harassment?

Clearly No

Clearly Yes

1      2      3      4      5      6      7

Vignette 3 (PERSEX3):

Ted, a sales manager, and a new sales rep, Yvonne, have started dating. Since they began seeing each other, Yvonne has received merit increases and a recent promotion to sales trainer. Until this relationship began, Jessica was considered the “star” of the sales department.

Is this a case of sexual harassment?

Clearly No

Clearly Yes

1      2      3      4      5      6      7





# J.B.M.

## Invitation To Review Manuscripts

The review process is a critical step in publishing a quality journal. The editors of the Journal of Business and Management invite you to participate in the ongoing activities necessary to make JBM a reputable scholarly outlet. If you would like us to send you manuscripts to review, please complete the form below or email us with the information at [jbm@chapman.edu](mailto:jbm@chapman.edu).

Name \_\_\_\_\_

Address \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

Email \_\_\_\_\_

Please list your major areas of interest:

Please indicate how many manuscripts you would be willing to review in an academic year:

1     2     3     4     5

*Please return the form to:*

Amy E. Hurley-Hanson, Ph.D.

Cristina M. Giannantonio, Ph.D.

*Editors, Journal of Business and Management*

Argyros School of Business and Economics

Chapman University

One University Drive

Orange, CA 92866

FAX (714) 532-6081



# J.B.M.

## Subscription Form

The Journal of Business and Management is published by the Argyros School of Business and Economics, Chapman University. It is sponsored by the Western Decision Sciences Institute.

The annual subscription fee is \$50.00 for individuals and \$100.00 for institutions. Your check or money order should be made out to CHAPMAN UNIVERSITY/JBM. Please complete your mailing information below.

Name \_\_\_\_\_

Department \_\_\_\_\_

Address Line 1 \_\_\_\_\_

Address Line 2 \_\_\_\_\_

Address Line 3 \_\_\_\_\_

Tel \_\_\_\_\_

Fax \_\_\_\_\_

*Mail this form with your check or money order to:*

Journal of Business and Management  
Argyros School of Business and Economics  
Chapman University  
One University Drive  
Orange, CA 92866

email: [jbm@chapman.edu](mailto:jbm@chapman.edu)



## NOTES

## NOTES

ISSN: 1535-668X  
Editorial Offices: Journal of Business and Management  
Argyros School of Business and Economics  
Chapman University  
One University Drive  
Orange, CA 92866

Fax: (714) 532-6081

E-mail: [jbm@chapman.edu](mailto:jbm@chapman.edu)

Submission information is available on our website: <http://jbm.chapman.edu>

*The Journal of Business and Management* is published by Chapman University's Argyros School of Business and Economics with the sponsorship of the Western Decision Sciences Institute. The primary editorial objective of the Journal of Business and Management is to provide a forum for the dissemination of theory and research in all areas of business, management, and organizational decisions which would be of interest to academics and practitioners.

The views expressed in published articles are those of the authors and not necessarily those of the editors, editorial board, WDSI or Chapman University.

Copyright © 2008 Argyros School of Business and Economics, Chapman University.



**CHAPMAN**  
UNIVERSITY

THE GEORGE L. ARGYROS  
SCHOOL OF BUSINESS AND ECONOMICS  
ONE UNIVERSITY DRIVE  
ORANGE, CALIFORNIA 92866

---

Non-Profit Organization  
U.S. POSTAGE

**PAID**

Orange, California  
Permit No. 58

---